Antenatal detection of fetal growth anomalies
a hybrid effectiveness-implementation randomised control trial of a complex intervention

Relph, Sophie

Awarding institution:
King's College London

Antenatal detection of fetal growth anomalies: a hybrid effectiveness-implementation randomised control trial of a complex intervention

Sophie Alexandra Relph

A thesis submitted to King's College London for the degree of Doctor of Philosophy

Submitted June 2021

Department of Women and Children's Health

School of Life Course Sciences, Faculty of Life Sciences and Medicine

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS AND GLOSSARY

AC	Abdominal circumference	Fetal biometric measurement taken at ultrasound, component part of the estimated fetal weight calculation.
ACOG	American College of Obstetricians and Gynaecologists	American professional body which produces clinical guidance, amongst other roles.
AGA	Appropriate for gestational age	Fetal or neonatal weight between the 10th and 90th centiles of that expected for gestational age.
aOR	Adjusted odds ratio	Ratio of the odds of an event occurring in one group to the odds of it occurring in another group, adjusted by other relevant characteristics to improve comparability.
AUROC	Area under the receiver operating characteristics (curve)	Statistical term used to describe overall test performance through combination of test sensitivity and specificity.
BMI	Body mass index	A measure of body size calculated by dividing the weight (in kilograms) by the squared height (in metres). Enables categorisation of individuals into underweight, healthy weight, overweight or obese groups.
BPD	Biparietal diameter	Fetal biometric measurement taken at ultrasound, component part of the estimated fetal weight calculation.
CAG	Confidentiality Advisory Group	Independent body in England which provides expert advice on the use of confidential patient information.
CHEERS	Consolidated Health Economic Evaluation Reporting Standards	Reporting standards for economic evaluation
CICI	Context and Implementation of Complex Interventions	Qualitative framework for studying the context and implementation processes of complex interventions
CINAHL	Cumulative Index to Nursing and Allied Health Literature	Academic literature database which specialises in publications relevant to nursing and allied health professions.
CNST	Clinical Negligence Scheme for Trusts	Handles all clinical negligence claims against member NHS bodies, who pay a fee for the service
CONSORT	Consolidated Standards of Reporting Trials	Reporting standards for randomised control trials
DESiGN trial	DEtection of the Small for GestatioNal age fetus trial	Randomised cluster control trial comparing the effect of the Growth Assessment Protocol and standard care on the rate of detection of the small for gestational age fetus.
DMC	Data Monitoring Committee	A group of independent experts with trial oversight who assess trial progress, safety and analysis of trial data.
DOB	Date of birth	In the context of this thesis, refers to the date of the pregnant woman's birth, not the date the baby was born (date of delivery).
EDC	Estimated date of conception	Calculated from the date of delivery, minus the gestational age at birth (in days) plus 14 days
EDF	End diastolic flow	Flow in a blood vessel during the diastolic phase (cardiac relaxation) as measured by Doppler studies.

EFW	Estimated fetal weight	Calculated estimate of the weight of the fetus made using established formulae which incorporate ultrasound measurements of fetal biometry.
EPR	Electronic patient record	Also known as electronic medical records. Store data that are routinely recorded during the course of normal clinical investigation and management, for the purpose of documenting the events of the patient visit.
FGR	Fetal growth restriction	A fetus for whom growth has been restricted during pregnancy. Consensus definition available in Table 1.1.
FH	Fundal height	Measurement of the size of the gravid uterus. Used to estimate fetal size during pregnancy
FL	Femur length	Fetal biometric measurement taken at ultrasound, component part of the estimated fetal weight calculation.
FN	False negative	A screening outcome that indicates that the condition does not exist when it actually does.
FP	False positive	A screening outcome that indicates that the condition exists when it does not.
GAP	Growth Assessment Protocol	A complex antenatal intervention involving site-wide training, risk-stratification and surveillance protocols, customised fetal growth assessments, audit and missed-case analysis and aiming to improve the rate of detection of the small for gestational age fetus and reduce rates of stillbirth.
GLM	Generalised linear model	A statistical model commonly used in economic evaluations which is similar to linear regression but flexibly allows for distributions which are not parametric.
GROW	Gestation-Related Optimal Weight	Customised fetal and neonatal growth or weight standards (see section 1.1.7.1 for further detail).
HC	Head circumference	Fetal biometric measurement taken at ultrasound, component part of the estimated fetal weight calculation.
HDU	High Dependency Unit (neonatal)	Also known as a local or level 2 neonatal unit, an intermediate level of neonatal care. Specification available at https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2015/01/e08-serv-spec-neonatal-critical.pdf
HIE	Hypoxic Ischaemic Encephalopathy	A heterogenous prenatal, intrapartum or neonatal brain injury caused by a hypoxic-ischemic event and affecting full term infants.
HMIC	Health Management Information Consortium	Academic literature database which specialises in publications relevant to hospital administrators and managers
ICD-PM	International Classification of Diseases – Perinatal Mortality	An international classification of causes for or diseases associated with perinatal mortality.
ICER	Incremental cost effectiveness ratio	The cost associated with each additional unit increase in a specified outcome
ID	Identifier	Identification number given to an individual during a process of anonymisation of medical records.
LGA	Large for gestational age	Fetal or neonatal weight above the 90th centile of that expected for gestational age (any growth chart).
LGA$_{both}$		Large for gestational age as defined by both customised and population growth charts
LGA$_{cust}$		Large for gestational age as defined by customised growth charts

LGA$_{pop}$		Large for gestational age as defined by population growth charts
LR	Likelihood ratio	Statistical comparison of test sensitivity and specificity. Estimates the likelihood of getting a given test result for patient with the disease compared to in a patient without the disease.
MBRRACE	Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK	Collaborative group conducting a national programme of work including surveillance and investigation of the causes of maternal deaths, stillbirths and infant deaths.
MCA	Middle cerebral artery	Doppler study of the middle cerebral artery is used in fetal growth restriction to determine whether there is evidence of cerebral redistribution of blood flow.
MICE	Multiple imputation through chained equations	A statistical method used when dealing with missing data.
MRC	Medical Research Council	UK body which invests in medical research on behalf of the UK taxpayer.
MRI	Magnetic resonance imaging	A medical imaging technique.
NHS	National Health Service	The publicly funded healthcare system of the UK.
NICE	National Institute for Health and Care Excellence	Public body that provides guidance, advice and information services for health, public health and social care professionals in England and Wales.
NICU	Neonatal intensive care unit	Also known as a level 3 neonatal unit. Provides the highest level of neonatal care. Specification available at https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2015/01/e08-serv-spec-neonatal-critical.pdf
NPV	Negative predictive value	Of those individuals who have a negative test result, the proportion without the studied disease/condition.
OR	Odds ratio	Ratio of the odds of an event occurring in one group to the odds of it occurring in another group.
PAPP-A	Pregnancy Associated Plasma Protein-A	Produced by the placenta during pregnancy and regulates the bioavailability of insulin-like growth factor that is essential for normal fetal development. Low levels are seen in trisomy disorders or placental-mediated disorders such as pre-eclampsia or fetal growth restriction.
PAR	Population attributable risk	The proportion of all cases of a disease or condition in a population that is attributable to a specific exposure.
Perinatal Institute	The Perinatal Institute	A not-for-profit organisation which provides training, protocols, GROW chart software, missed-case analysis and benchmarking tools as part of GAP.
PI	Pulsatility index	The difference between peak systolic and end diastolic flow velocity in a blood vessel, divided by the time-average flow velocity. Measured during Doppler studies. A high PI indicates high resistance to blood flow.
PlGF	Placenta-like Growth Factor	A growth-stimulating hormone produced by the placenta. Low levels of PlGF are associated with states of placental insufficiency.
PMR	Perinatal mortality rate	The rate of combined stillbirth and neonatal death, usually expressed per 1,000 registerable births.
PPV	Positive predictive value	Of those individuals with a positive test result, the proportion who actually have the studied disease/condition.

PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses	Reporting standards for systematic reviews and meta-analyses
QALY	Quality adjusted life year	One QALY is equal to one year of life in perfect health, calculated to estimate the standardised number of years of perfect health gained or lost following an intervention.
QHES	Quality of Health Economic Analyses	Checklist for quality assessment of reported economic evaluations
RCOG	Royal College of Obstetricians and Gynaecologists	UK professional body of obstetricians and gynaecologists which provides clinical guidance, amongst other roles
RCT	Randomised control trial	A research study in which participants or groups are randomly allocated to two or more interventions.
REC	Research Ethics Committee	English committees which review health and social care research proposals and advise on whether they consider the research to be ethical.
SCBU	Special Care Baby Unit	Also known as a level 1 neonatal unit. Provide a lower level of neonatal care. Specification available at https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2015/01/e08-serv-spec-neonatal-critical.pdf
SD	Standard deviation	A statistical measure of the variation or dispersion of values in a sample.
SGA	Small for gestational age	Fetal or neonatal weight below the 10th centile of that expected for gestational age (any growth chart).
SGA_{both}		Small for gestational age as defined by both customised and population growth charts
SGA_{cust}		Small for gestational age as defined by customised growth charts
SGA_{pop}		Small for gestational age as defined by population growth charts
StaRI	Standards for Reporting Implementation Studies	Reporting standards for studies of implementation
STROBE	The Strengthening the Reporting of Observational Studies in Epidemiology	Reporting standards for observational studies
TIDieR	Template for Intervention Description and Replication	A standardised template recommended for use when describing the components of an intervention.
TN	True negative	A screening outcome where it is correctly indicated that the condition does not exist.
TP	True positive	A screening outcome where it is correctly indicated that the condition does exist.
TSC	Trial Steering Committee	An independent oversight committee which provides overall supervision for a project on behalf of the sponsor and funder. Aims to ensure that the project is conducted to rigorous standards.
UK	United Kingdom	The United Kingdom of Great Britain (England, Scotland and Wales) and Northern Ireland.
WHO	World Health Organisation	International body which aims to direct international health within the United Nations' system and to lead partners in global health responses.

ACKNOWLEDGEMENTS

ACHIEVEMENTS DURING THIS PHD

I chose to pursue a postgraduate research degree in the field of perinatal epidemiology following my involvement in local observational research projects during my junior clinical training years. I had thoroughly enjoyed these early opportunities and found my skill set to be well-suited to epidemiology, basic statistics and academic writing, but I sought to further develop this learning and add new academic skills by training under the supervision of subject experts and whilst contributing to higher quality research. I hoped that such professional development would equip me with knowledge and proficiencies that I could translate into many aspects of my future obstetric career, including through contribution to further research and to local, regional or national quality improvement and clinical service development.

During my study for a Doctorate in Philosophy at King's College London, I have also been fortunate to work alongside inspiring academics and health service leaders who have offered me the opportunity to contribute to additional projects, thereby expanding my own experience and skillset within the field of perinatal research and service improvement. This has included additional research projects conducted with colleagues in the Department of Women and Children's Health at King's College London, audit projects conducted during a part-time role as Obstetric Fellow for the National Maternity and Perinatal Audit (NMPA) based at the Royal College of Obstetricians and Gynaecologists (RCOG), and a large contribution to the combined RCOG and Royal College of Midwives (RCM) guidance and patient information published to support national and international health services during the 2019/2020 coronavirus (COVID-19) pandemic.

I have detailed everything that I have achieved during my PhD studies over the following pages, categorised according to the context in which it was achieved and including a summary of my contribution to the work.

Achievements related to the work presented in this PhD

Publications

Relph S, Elstad M, Coker B, Vieira MC, Moitt N, Muruet Gutierrez W, Khalil A, Sandall J, Copas A, Lawlor DA, Pasupathy D on behalf of the DESIGN Trial Team. Using electronic patient records to assess the effect of a complex antenatal intervention in a cluster randomised controlled trial - Data management experience from the DESiGN Trial team. **Trials** 2021. 22:195. https://doi.org/10.1186/s13063-021-05141-8 *Planned and wrote this methodological paper on the data management processes of the DESiGN trial that I had developed with the data management team.*

Relph S, Delaney L, Vieira MC, Pasupathy D, Healey A. Costing the Impact of Interventions during Pregnancy in the UK: A Systematic Review of Economic Evaluations. **BMJ Open** 2020;10:e040022. http://dx.doi.org/10.1136/bmjopen-2020-040022. *Planned, conducted, analysed and wrote the manuscript for this systematic review.*

Vieira MC, **Relph S**, Persson M, Seed PT, Pasupathy D. Determination of birth-weight centile thresholds associated with adverse perinatal outcomes using population, customised, and Intergrowth charts: A Swedish population-based cohort study. **PLoS Medicine** 2019; 16(9): e1002902 10.1371/journal.pmed.1002902 *Involved in planning the project, reviewing results and providing feedback on the manuscript.*

Vieira MC, **Relph S**, Copas A, Healey A, Coxon K, Alagna A, Briley A, Johnson M, Lawlor DA, Lees C, Marlow N, McCowan L, Page L, Peebles D, Shennan A, Thilaganathan B, Khalil A, Sandall J, Pasupathy D. DEtection of Small for Gestational age Neonate (SGA) – a cluster randomised controlled trial to evaluate the effect of the Growth assessment protocol (GAP) programme: The DESiGN Trial Protocol. **Trials.** 2019 March 4;20(1):154 https://doi.org/10.1186/s13063-019-3242-6 *Adopted the existing trial protocol to write this published manuscript.*

Publications in draft form

Vieira MC, **Relph S**, Muruet-Gutierrez W, Elstad M, Coker B, Moitt N, Delaney L, Verger C, Healey A, Coxon K, Alagna A, Briley A, Johnson M, Page L, Peebles D, Shennan A, Thilaganathan B, Marlow N, McCowan L, Lees C, Lawlor DA, Khalil A, Sandall J, Copas A, Pasupathy D on behalf of the DESiGN Collaborative Group. Effect of the Growth Assessment Protocol (GAP) on the detection of small for gestational age: the DESiGN cluster randomised trial. *Managed the clinical trial, collected all data, contributed to analyses, reviewed and commented on manuscript drafts. Manuscript submitted.*

Relph S,* Coxon K,* Vieira MC, Copas A, Healey A, Alagna A, Briley A, Johnson M, Lawlor DA, Lees C, Marlow N, McCowan L, McMicking J, Page L, Peebles D, Shennan A, Thilaganathan B, Khalil A, Pasupathy D, Sandall J on behalf of the DESiGN Trial Team. Effect of the Growth Assessment Protocol on the DEtection of the Small for GestatioNal Age Fetus: Process evaluation from the DESiGN cluster randomised trial. *Authors contributed equally. *Collected data, conducted all analyses and co-wrote this manuscript with Dr Coxon. First draft complete, awaiting co-investigator comments.*

Relph S, Vieira MC, Copas A, Coxon K, Alagna A, Briley A, Johnson M, Page L, Peebles D, Shennan A, Thilaganathan B, Marlow N, McCowan L, Lees C, Lawlor DA, Khalil A, Sandall J, Pasupathy D and Healey A on behalf of the DESIGN Trial Team. Comparing the Growth

Assessment Protocol to standard care for detection of the small for gestational age fetus – a within trial, cost-effectiveness analysis. *Co-designed study, collected all data, co-conducted the analysis with Dr Healey, wrote this draft manuscript. Manuscript currently in late draft format.*

Presentations

Poster: Elstad M, Dhouiri A, Pasupathy D, **Relph S**, Vieira M, Coker B. Pseudo-anonymisation of routine maternity data from heterogeneous National Health Service systems: The DESiGN trial. **Young Statisticians' Meeting**, August 2019. Leeds, UK. *Worked closely with the data management team to plan the pseudonymisation process reported in this poster presentation.*

Poster: Relph S, Coxon K, Silverio SA, Vieira MC, Khalil A, Pasupathy D, Sandall J (on behalf of the DESiGN Trial Team). Evaluating the Growth Assessment Protocol (GAP) to reduce Stillbirths. **Continuity of care and managing complexity during and after pregnancy: what we have learnt: the CLAHRC South London reporting event,** 12th July 2019. *Planned, designed and presented this research poster on the process evaluation methods of the DESiGN trial.*

Poster: Relph S, Delaney L, Vieira M, Healey A, Pasupathy D (on behalf of the DESiGN Trial Team). Costing the Impact of Interventions during Intrapartum Care in the UK: A Systematic Review of Economic Evaluations. **King's Health Partners: Women and Children's Health Institute** launch event, 10th June 2019. *Planned, designed and presented this poster on the systematic review reported above.*

Poster: Vieira MC, **Relph S**, Persson M, Pasupathy D. Population, customised and Intergrowth charts: the importance of assessing different thresholds for abnormal fetal growth. **British Maternal and Fetal Medicine Society's 21st Annual Conference**, March 2019. Edinburgh, UK. *Involved in the planning and design of this poster which presented the manuscript cited above.*

Poster: Delaney L, **Relph S**, Melaugh A, Vieira M, Healey A, Pasupathy D. Attributing economic costs to intrapartum events: A literature review. **British Maternal and Fetal Medicine Conference**, April 2018. *Supervised the planning, design and presentation of this poster by a research assistant (findings of the systematic review cited above).*

<u>Achievements related to work conducted with colleagues at King's College London, but unrelated to this PhD</u>

Award

University of California, San Francisco and King's College London Designing Clinical Research course – **First prize (£750) for best protocol presentation**, presented by Professor Sir Robert Lechler (December 2017). *Planned and presented the winning research protocol.*

Publications

Dalrymple KV, Uwhubetine O, Flynn AC, Pasupathy D, Briley AL, **Relph S**, Seed PT, O'Keeffe M, Poston L. Modifiable determinants of postpartum weight loss in women with obesity – A secondary analysis of the UPBEAT trial. **Nutrients** 2021. 13: 1979. https://doi.org/10.3390/nu13061979 *Provided clinical insight for this project, reviewed and provided feedback on the draft manuscripts.*

Relph S, Guo Y, Harvey ALJ, Vieira MC, Corsi DJ, Gaudet LM, Pasupathy D. Characteristics associated with uncomplicated pregnancies in women with obesity: a population-based cohort study. **BMC Pregnancy and Childbirth** 2021. 21:182. *https://doi.org/10.1186/s12884-021-03663-2 Planned, reviewed the analysis, interpreted the results and wrote the manuscript for this population study.*

Relph S, Ong M, Vieira MC, Pasupathy D, Sandall J. Perceptions of risk and influences of choice in pregnant women with obesity. An evidence synthesis of qualitative research. **PLOS ONE** 2020; 15(1): e0227325. https://doi.org/10.1371/journal.pone.0227325 *Planned, collected data for, analysed and reported this qualitative evidence synthesis, including supervision of a medical student who assisted me with the data collection and analysis.*

Dalrymple KV, Flynn AC, **Relph SA**, O'Keefe M, Poston L. Lifestyle interventions in overweight and obese pregnant or postpartum women for postpartum weight management: A systematic review of the literature. **Nutrients** 2018 Nov 7;10(11), E1704; https://doi.org/10.3390/nu10111704 *Provided clinical insight for this project, reviewed and provided feedback on the draft manuscripts.*

Presentations

Poster: Relph S, Vieira MC, Ong M, Harvey A, Guo Y, Persson M, Nippita T, Gidaszewski B, Souter V, Sandall J, Gaudet L, Pasupathy D. Uncomplicated Pregnancy and Birth amongst Obese Women: An International Collaboration to increase maternal choice. **King's Health**

Partners: Women and Children's Health Institute launch event, 10th June 2019. *Planned, designed and presented this research poster describing the potential for an international research collaboration on this topic.*

Poster: Relph S, Ong M, Vieira MC, Pasupathy D, Sandall J. How do obese women perceive and make decisions regarding risks and choices presented during pregnancy? An evidence synthesis of qualitative literature. **British Maternal and Fetal Medicine Society's 21st Annual Conference**, March 2019. Edinburgh, UK. *Planned, designed and presented the poster describing the qualitative evidence synthesis cited above.*

Achievements related to work conducted whilst an Obstetric Fellow for the National Maternity and Perinatal Audit

Published reports

Relph S, NMPA Project Team. NHS Maternity Care for Women with a Body Mass Index of 30 kg/m² or above. London: RCOG; 2021. https://maternityaudit.org.uk/FilesUploaded/NMPA%20BMI%20Over%2030%20Report.pdf *Planned the project, worked with the statistician to execute it, set-up and worked with a lay advisory group, wrote the final report.*

Relph S, NMPA Project Team. NHS Maternity Care for Women with Multiple Births and Their Babies: A study on the feasibility of assessing care using data from births between 1 April 2015 and 31 March 2017 in England, Wales and Scotland. London: RCOG; 2020. https://maternityaudit.org.uk/FilesUploaded/NMPA%20Multiple%20Births%20Report%202020.pdf *Planned the project, conducted the statistical analysis under supervision, interpreted the results, wrote the final report.*

Presentations

Oral presentation: Relph S, on behalf of the National Maternity and Perinatal Audit; Maternal and neonatal birth outcomes in Britain for women according to body mass index and parity: A national audit. **RCOG 2021 Virtual World Congress,** 12 June 2021**.** *Invited speaker to present the work described in the report cited above.*

Oral presentation: Relph S, Blotkamp A on behalf of National Maternity and Perinatal Audit project team; The National Maternity and Perinatal Audit. Presented to the **Maternity and Children Quality Improvement Collaborative (Scotland)** meeting. 23 October 2019 in Sterling, Scotland. *Invited speaker and panel member to discuss the NMPA approach to quality improvement of maternity care in the UK.*

Achievements related to work conducted whilst working with the COVID-19 team at the RCOG/RCM

Award

RCOG Chief Executive Officer's Team of the Year – Presented to the COVID-19 guidance production team (December 2020). *I was one of two lead clinical fellows who co-developed the full suite of RCOG COVID-19 guidance, under the supervision of the RCOG President and Vice-Presidents, and working closely with RCOG clinical quality, digital, media, public relations and political staff between March 2020-February 2021.*

Publications

Relph S, Thangaratinam S. Maternal Medicine in the COVID-19 era. Invited review - **Best Practice & Research Clinical Obstetrics & Gynaecology** 2021. 73. https://doi.org/10.1016/j.bpobgyn.2021.03.003 *Invited by Professor Thangaratinam to write this narrative review based on the RCOG guidance on maternal medicine during the COVID-19 pandemic.*

Jardine J*, **Relph S***, Magee LA, von Dadelszen P, Morris E, Ross-Davie M, Draycott T, Khalil A. Maternity services in the UK during the coronavirus disease 2019 pandemic: a national survey of modifications to standard care. **BJOG** 2020; https://doi.org/10.1111/1471-0528.16547 *Authors contributed equally. *Planned, conducted, and co-wrote the final manuscript for this national survey study.*

Published guidance, patient information, frameworks and statements

NHS England, RCOG, RCM and Society and College of Radiographs. Framework to assist NHS trusts to reintroduce access for partners, visitors and other supporters of pregnant women in English maternity services.
https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/09/par001599-framework-for-the-reintroduction-of-visitors-throughout-maternity-services-sep-2020.pdf *Lead developer – consulted with stakeholders and drafted this framework on behalf of NHSE&I and RCOG.*

RCOG, RCM. Coronavirus (COVID-19) infection in pregnancy: Information for healthcare professionals v1.0-12.0.
https://www.rcog.org.uk/en/guidelines-research-services/guidelines/coronavirus-pregnancy/ *Co-developer of versions 1-12, planning, writing and reviewing in response to peer-review and stakeholder consultation.*

RCOG. Principles for the testing and triage of women seeking maternity care in hospital settings, during the COVID-19 pandemic v1.0-2.0. https://www.rcog.org.uk/globalassets/documents/guidelines/2020-08-10-principles-for-the-testing-and-triage-of-women-seeking-maternity-care-in-hospital-settings-during-the-covid-19-pandemic.pdf *Co-developer, planning, writing and reviewing in response to peer-review and stakeholder consultation.*

RCOG. Guidance for maternal medicine services in the evolving coronavirus (COVID-19) pandemic: Information for healthcare professionals v1.0-2.4. https://www.rcog.org.uk/globalassets/documents/guidelines/2020-07-10-guidance-for-maternal-medicine.pdf *Lead developer on behalf of RCOG, planning, working with clinical experts, writing, reviewing in response to peer-review and stakeholder consultation and updating.*

Presentations

COVID-19: Lessons learned and future plans. **National Trainees Conference: Bristol,** 27 May 2021. *Invited to speak on my experience working as a clinical fellow in the RCOG COVID-19 guidance team.*

Contributing to the global COVID-19 pandemic: An RCOG perspective. Presented at the **RCOG Annual Professional Development Conference**, 19 November 2020. *Invited to speak on the RCOG contribution to the COVID-19 global pandemic.*

FOREWORD

The World Health Organisation's ambitious Every Baby Action Plan (2014) aims to end preventable perinatal deaths by 2030. Internationally, many stillborn babies are small-for-gestational-age (SGA) following periods of fetal growth restriction. Rates of antenatal detection of SGA are often low but increases have been associated with a halving of stillbirth rates in high-income countries. Improvement of the rate of detection of SGA is, therefore, a global priority.

One strategy that has been recommended to improve the rate of antenatal detection of SGA is the 'Growth Assessment Protocol' (GAP). GAP is a complex antenatal intervention that includes staff training, risk stratification, screening, management protocols to improve detection and monitoring of SGA, and regular audit. GAP has been implemented in maternity units throughout the UK, Australia and New Zealand following positive results from observational research.

The DEtection of the Small for GestatioNal age fetus (DESiGN) trial was the first randomised cluster control trial that assessed the clinical-effectiveness, cost-effectiveness, and implementation of GAP in improving the antenatal detection of the SGA fetus. Whilst I was actively and heavily involved in the trial management, data collection, analysis, interpretation and presentation of the primary trial results, during this thesis I have concentrated on four secondary analyses of the DESiGN trial, three of which were pre-planned at the time of writing the trial protocol. In addition, I will present the results of a systematic review conducted to inform the secondary analysis on the cost-effectiveness of GAP.

STATEMENT OF CONTRIBUTION

The DESiGN trial protocol was developed by a team of co-investigators comprised of experts in the fields of maternal or fetal medicine, neonatal care, epidemiology and statistics, implementation science and health-economics, and a lay representative during 2014-2016. Ethical approval was obtained, and sites recruited prior to my involvement in the project in 2017. At the time that I commenced the project, major protocol amendments related to a change in sponsor and change in the primary outcome definition were ongoing, two further sites were being recruited and the sites already randomly allocated to implement the intervention were in the preparation phases for this.

In October 2017, I formally took on the DESiGN trial management role, acting under the supervision of Professor Dharmintra Pasupathy, the trial Chief Investigator. This role involved formalising protocol amendments, making major and minor amendment applications to the ethics committee and confidentiality advisory group, working with site research and development departments, principal investigators and clinicians, liaising with the implementation team at the Perinatal Institute (intervention provider) and at each clinical site, liaising with digital professionals at maternity units and with the trial data management team to collect and manage quantitative data, updating the trial co-investigators and the steering committee amongst many other tasks.

Specifically, for the work described in this thesis I made the following contributions, under the supervision of Professors Dharmintra Pasupathy and Jane Sandall:

- **Chapter 1:** I conducted a systematic literature search of PubMed and EMBASE to identify manuscripts relevant to the diagnosis of small-for-gestational-age or growth restricted fetuses. I independently synthesised the findings of this and added evidence from the grey literature to write the narrative review presented.
- **Chapter 2:** Whilst the trial protocol was written prior to my involvement in the project, I adapted it and personally wrote the published version of the protocol which has been further adapted for this chapter. I worked closely with the data management team to develop processes for this trial. I published these in a methodology paper which I have also modified for this chapter.
- **Chapter 3 – Qualitative analysis:** Working with Dr Kirstie Coxon and under the supervision of Professor Sandall, I planned and conducted semi-structured interviews through which we collected qualitative data on implementation process, context, acceptability and feasibility. This included writing topic guides and independently conducting 28 interviews with clinical midwives and sonographers

(Dr Coxon conducted the other 27 interviews with clinical managers/GAP implementing leads). We each coded our own interviews onto the implementation framework and split the synthesis of findings equally.

- **Chapter 3 – Quantitative analysis:** I identified an opportunity to conduct a more detailed study of the strength of implementation of the trial intervention. I presented this and my proposed plans to assess implementation strength to my supervisors and we agreed to proceed with a detailed review of notes for a sample of women exposed to GAP during the DESiGN trial. I independently planned, conducted and analysed this study under the supervision of Professor Sandall.

- **Chapter 4:** Whilst planning the health-economic evaluation, I identified multiple potential cost-estimates for the same maternity activities. I wanted to both describe this heterogeneity and also develop a strategy for costing the DESiGN trial intervention. I planned, interpreted, and analysed this report under the supervision of Dr Andy Healey. I collected the data in tandem with Dr Louisa Delaney and completed the assessments of study quality with both Dr Delaney and Dr Alexandra Melaugh.

- **Chapter 5:** For the cost-effectiveness analysis, I worked closely with Dr Andy Healey (trial health economist) to plan this analysis. I determined the costing framework, applied costs to all records, generated descriptive statistics and prepared the dataset for advanced analysis. Dr Healey completed the advanced stages of the analysis that I have presented in this thesis.

- **Chapter 6:** Given the overall low rate of SGA detection seen both in the DESiGN trial (with or without implementation of the intervention) and universally in the literature, I proposed to conduct a study aiming to understand factors associated with antenatally missing the diagnosis of SGA. I planned this study, conducted the analysis independently, interpreted and wrote the results under the supervision of Dr Matias Vieira, Professor Pasupathy and Professor Andrew Copas (trial statistician).

- **Chapter 7:** A study of the effect of the intervention on the rate of detection of large for gestational age fetuses was pre-planned as a secondary analysis in the trial protocol. I adapted the statistical code for the primary trial analysis (effect of the intervention on the rate of detection of small for gestational age), modifying as necessary, to conduct this analysis. I interpreted the results and wrote the report under the supervision of Professor Pasupathy.

ABSTRACT

Background

Stillbirth affects 2.6 million pregnant women worldwide each year. Approximately 43% of stillbirths are small for gestational age (SGA) at the time of death. Antenatal detection of SGA is low; less than half of SGA babies are detected when universal fundal height measurement and selective ultrasound screening are applied, this increases to 80-90% with universal ultrasound screening in trial populations but such rates have not been replicated in clinical practice. Increasing antenatal detection of SGA is associated with a reduction in the rate of perinatal mortality.

The Growth Assessment Protocol (GAP) is a complex antenatal intervention that aims to improve the rate of antenatal detection of SGA. It includes staff training, site-wide changes to screening protocols for SGA, customisation of fetal and neonatal growth charts, audit and missed case analysis. Estimates from observational studies suggest that GAP could increase rate of antenatal detection of SGA from 20% to 40 or 60%. Its implementation has also been associated with a reduction in rates of stillbirth. The DESiGN trial, part reported in this thesis, was the first randomised control trial comparing the effect of GAP and standard care on the rate of antenatal detection of SGA. The trial found no difference between the rate of SGA detected in clusters allocated to GAP or standard care (25·9% vs 27·7%; adjusted difference 2·4%, 95% CI -6·1% to 10·8%; p=0·58).

The UK National Institute for Health and Care Excellence (NICE) bases its recommendations on the implementation of new interventions according to both their clinical efficacy and cost-effectiveness. Process evaluation is key to understanding effectiveness when evaluating complex interventions in everyday practice. Both process and economic evaluation are therefore pertinent to the study of complex interventions such as GAP. GAP has not previously been studied using either formal method.

Improving antenatal detection of SGA is a global priority, but strategies are currently less effective than required. It has previously been demonstrated that women who are multiparous or with low body mass index are more likely to have SGA detected, and women without a third trimester scan are more likely to have SGA missed. A full understanding of factors associated with missed SGA is essential when improving screening programs intended to detect it.

Finally, implementation of interventions to screen for SGA have the unintended consequences of also detecting large for gestational age (LGA) fetuses. A systematic review demonstrated that inducing labour before 40 weeks of gestation for women with a fetus suspected to be LGA reduces the risk of shoulder dystocia and adverse neonatal outcomes

however screening is not currently recommended by national guidelines. The effect of GAP on the antenatal detection of LGA and associated maternal and perinatal outcomes is unknown.

Aim and objectives

The aim of the work contained within this thesis was to conduct a detailed evaluation of the implementation and cost-effectiveness of the Growth Assessment Protocol in the context of the DESiGN trial, including an assessment of its impact on both large and small-for-gestational-age babies.

The objectives were to:

1. Determine barriers and facilitating factors associated with the strength of implementation of the Growth Assessment Protocol.
2. Review the current evidence on costs incurred following implementation of maternity care interventions in the UK.
3. Evaluate the cost-effectiveness of the Growth Assessment Protocol in terms of its ability to improve antenatal detection of the SGA fetus.
4. Identify the clinical and ultrasound utilisation characteristics of pregnancies in which an antenatal diagnosis of SGA is missed, compared to those in which it is made, to understand how interventions can be better targeted to improve detection.
5. Assess for the presence of an unintended impact from implementing the Growth Assessment Protocol on the detection and management of pregnancies in which the baby is LGA.

Methods

The DESiGN trial was a randomised cluster control trial, conducted in the UK between November 2015 – February 2019. Clusters were randomly allocated to implement the Growth Assessment Protocol, or to continue standard care. The trial was powered to detect a difference in the primary outcome: antenatal detection of SGA in fetuses confirmed as SGA at birth. Quantitative data were also collected on a range of secondary maternal, fetal, neonatal and economic outcomes. Missing data for key characteristics and outcomes were multiply imputed.

A process evaluation was conducted. Clinical staff were recruited to semi-structured interviews and asked about intervention acceptability, feasibility, barriers, facilitators, and implementation context. Clinical guidelines and training records were collected, and a notes review conducted to determine implementation strength (composite of fidelity, dose and reach). A framework analysis of transcribed interviews was performed using the Context

and Implementation of Complex Interventions (CICI) framework. Data from the notes review and training records were analysed using summary statistics at the cluster level.

The economic costing framework was informed by a systematic review of national maternity costs. Costs were applied to all antenatal, intrapartum, postnatal and neonatal activities hypothesised to potentially be influenced by the intervention. The incremental cost-effectiveness ratio was calculated in terms of the cost per additional case of SGA detected.

Maternal and fetal characteristics or patterns of ultrasound use associated with missed SGA were studied. Associations were adjusted using multivariable regression for the other characteristics and for trial factors (cluster site, trial phase and the intervention group).

A cluster summary analysis was performed to determine the effect of GAP on the rate of detection of LGA, restricted to births which occurred at or after 36^{+0} weeks of gestation and adjusted for the pre-implementation rate, maternal parity, ethnicity, and age. The same method was applied to determine the effect of GAP on maternal and perinatal secondary outcomes, measures of ultrasound utilisation and screening outcomes.

Findings

Despite GAP being introduced into a political and epidemiological context which drove the implementation of interventions designed to reduce stillbirth, the program was implemented with variable fidelity (training targets, concordance to guidelines and risk stratification protocols), generally high reach (proportion of women with growth chart), but low dose (frequency of fundal height or ultrasound fetal weight measurements, as appropriate). Degree of implementation within each domain studied was highly variable between clusters. Resource unavailability was a key barrier to achieving high implementation strength, however the intervention was facilitated by staff who believed in it, collaborated with one-another and developed bespoke materials to facilitate its adoption.

Of 5,084 titles or full texts screened for the systematic review of published maternity costs, 37 papers were included in the final review (27 primary research articles, 7 review articles and 3 economic evaluations from NICE guidelines). Variation was noted in cost estimates for healthcare activities throughout the maternity care pathway: for midwife-led outpatient appointments the range was £27.34 – £146.25 (mean £81.78), emergency caesarean section range was £1056.44 – £4982.21 (mean £3508.93) and postnatal admission, range was £103.00 – £870.10 per day (mean £469.55).

A costing framework was developed, informed by the systematic review, to cost the activity observed in the DESiGN trial. Of the four cost-effectiveness outcomes, GAP was most

likely to be clinically effective with higher cost (44.1% chance) and was expected to cost an additional £34,559 per 1,000 pregnancies compared to standard care. This equated to an additional £19,525 per additional SGA baby detected antenatally.

Amongst the estimated 14,053 SGA$_{both}$ babies (SGA as defined by both population and customised centile charts) in the imputed dataset, the rate of missed SGA was 75.9%. Of these, 65.2% had no recorded risk factor that would have indicated serial fetal growth scans. Missed SGA was less likely in pregnancies with a risk factor for SGA (aOR 0.59, 95% CI 0.51-0.67, p<0.001) or a non-cephalic fetal presentation at birth (aOR 0.61, 95% CI 0.48-0.77, p<0.001), and more likely with a BMI of 25.0-29.9 kg/m^2 (aOR 1.15, 95% CI 1.01-1.32, p=0.04) or with less severe SGA (increase in aOR 1.19/centile increase between 0-10, 95% CI 1.17-1.22, p<0.001). Compared to women with detected SGA, those with missed SGA had a longer duration between their last scan and their birth (27.4 days vs. 9.8 days, adjusted difference 17.6 days, CI: 16.9-18.4, p<0.001); the duration increased with higher gestational age at birth.

Of the 80,856 women and babies included across the pre-randomisation and outcome comparison phases of both arms of the trial, 5.4% were LGA$_{both}$. GAP was not found to increase the rate of incidental antenatal diagnoses of LGA$_{both}$ (48.0% vs. 38.0%, adjusted effect size -4.9%, CI: -20.5 to 10.7, p=0.54), nor did it affect the rate of maternal or perinatal secondary outcomes amongst LGA$_{both}$ babies born at or after 36^{+0} weeks' gestation. GAP was however associated with a lower total number of scans for LGA$_{both}$ babies than standard care (3.8 versus 4.7 per pregnancy, adjusted effect size -0.9, CI: -1.3 to -0.5, p=0.002).

Conclusions

In the context of the DESiGN trial, GAP was found to be neither clinically- nor cost-effective when compared to standard care at increasing the rate of detection of SGA; neither did it result in an incidental change in the rate of detection of LGA. The implementation of GAP was challenged by resource availability, this may have been contributory to its lack of effect.

The search to find an intervention which improves SGA detection, without causing harm through false positive diagnoses continues. Interventions worthy of further research include those which seek to improve the dose (number) of fundal height measurements received by women at low risk of SGA, a universal offer of a fetal growth scan at term, or strategies targeted specifically to women with BMI in the overweight range.

1 INTRODUCTION

Anomalies of fetal growth have long been recognised to be associated with adverse perinatal and maternal outcomes, such as stillbirth, neonatal or infant morbidity, pre-eclampsia or gestational diabetes.[1,2] Timely recognition, appropriate surveillance and management of fetal growth anomalies, and their associated maternal conditions, is associated with an improvement of outcomes.[3] Specifically, stillbirth is one of the most severe adverse consequences of fetal growth anomalies.[3] Observational research has identified that an improvement in the rate of diagnosis of small for gestational age fetuses (SGA, defined below), has been associated in a halving of the rate of stillbirth.[4]

The weight of babies at birth has been measured since at least the late seventeenth century, although weighing scales themselves have been documented since ancient Babylon and Egypt.[5] Normal infant growth was first studied and documented in 1830 by a Belgian paediatrician, who recognised that anomalies of infant growth could not be identified without first understanding normal growth.[5] However, it was not until after the advent of diagnostic ultrasound in the 1940s, and Ian Donald's subsequent application to visualise the growing fetus in Glasgow in 1956, that parameters of normal and abnormal fetal growth could be documented.[6] Diagnostic ultrasound of the fetus became more widespread in the UK from the mid 1970s.[7]

Assessment of fetal growth continues to be a central pillar of antenatal care in the 21st century.[8] Pregnant women are routinely stratified by their risk of having a small baby and surveillance offered accordingly.[1] Whilst screening to detect large babies is not routinely offered,[9] large babies may be incidentally identified through screening for small babies, and it is becoming increasingly clear that antenatal diagnosis and management of pregnancies in which the baby is larger than normal is associated with an improvement in outcomes at birth.[9]

1.1 DEFINING FETAL GROWTH ANOMALIES

Fetal growth anomalies are first suspected following an assessment of either fetal or neonatal size at a single timepoint. Subsequent measurements are then used to further characterise the anomaly. The size of either a fetus or neonate is categorised into one of three groups, commonly using specific reference populations (section 1.1.5) and thresholds that are statistically constructed to describe their estimated or actual weight as small, appropriate, or large for gestational age. Surveillance or management strategies are then targeted according to growth status.

Fetal growth restriction (FGR) refers to an antenatal condition in which the fetus has had a pathological restriction on their growth potential. FGR is not synonymous with SGA.

Whilst some SGA infants are growth restricted, others are thought to be constitutionally small. Some growth restricted fetuses weigh within the range expected for gestational age but are smaller than their genetically-determined birthweight potential.[10] Low birthweight is a term that is also used to infer fetal growth restriction, but when usually defined by an absolute birthweight threshold (e.g. less than 2,500g),[11] its diagnosis depends on gestational age at birth, and commonly includes many preterm babies.

Similarly, large for gestational age (LGA) is not synonymous with macrosomia ('big baby'). Macrosomia commonly refers to a birthweight above a given threshold, usually above 4,000g or 4,500g.[2] Whilst many macrosomic babies will be LGA, some will be appropriately grown for gestational age (AGA) but born post-term.

In this section, I will describe and compare current strategies for categorising anomalies of fetal growth.

1.1.1 Measures of growth status at birth

In the UK, it is recommended that the growth status of all babies is assessed at birth.[12] This enables identification of babies that require close monitoring in the neonatal period because of a higher risk of neonatal complications (e.g. hypothermia, hypoglycemia).

Neonatal weight and head circumference (HC) are the standard measures of growth status at birth. Neonatal length was previously used but is subject to significant inter-observer variation.[13] Other anthropometric measurements, such as skinfold thickness, ponderal index (an index of weight in relation to length) and mid-arm circumference at birth can provide an assessment of the nutritional status of the neonate but are mostly limited to research uses.[14,15]

Neonatal weight is commonly compared to the expected range for a given gestational age, detailed on reference charts. When a baby's weight is outside the expected range (commonly outside the 10th-90th centile parameters), the baby is categorised as LGA or SGA, and monitoring or further care commenced as indicated. One challenge with this is the choice of reference chart. A review of published reference charts for neonatal birth weight or fetal biometric measurements identified wide variation between charts in the weight values reported at the same centile and gestational age.[16] This was likely related to the heterogeneity of the reference populations and of the statistical methods used to construct the charts (detailed further in sections 1.1.6 and 1.1.7).[17]

1.1.2 Measures of growth status during the fetal period

Uptake of ultrasound to visualise the fetus became more widespread in the UK during the mid 1970s.[7] Early ultrasound-derived estimated fetal weight (EFW) was based only on the biparietal diameter (BPD), but with poor accuracy particularly for fetuses who were under- or over-grown.[18] The addition of abdominal measurement improved accuracy,[19] and in 1984 Hadlock identified that the inclusion of femur length improved the model further.[20] Hadlock's EFW calculator uses the parameters HC, BPD, abdominal circumference (AC) and femur length (FL) - Figure 1.1, and has been shown to be the most accurate of studied formula, with an estimated error margin of less than 5%.[21] It is still in common practice worldwide.

Figure 1.1 Standard fetal biometry: sonographic measurements of (a) the biparietal diameter and head circumference, (b) the abdominal circumference and (c) the femur diaphysis length.
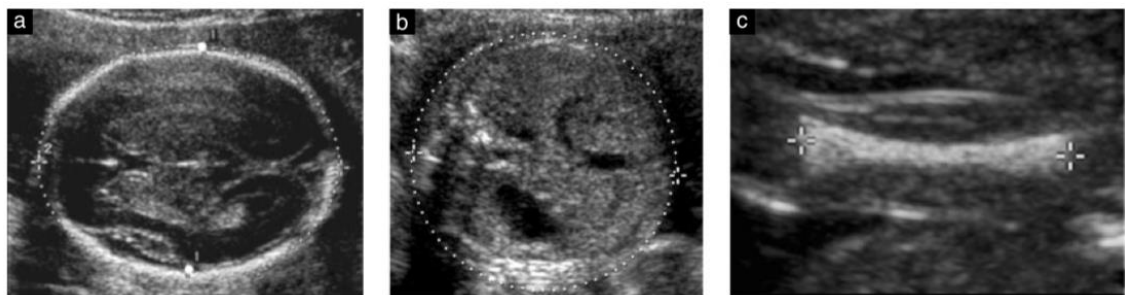
Copyright 2010 International Society of Ultrasound in Obstetrics and Gynaecology. Published by John Wiley & Sons, Ltd.[22] Reproduced with permission under the terms of the Creative Commons Attribution License.

Fetal charts can be used to assess growth status by plotting the biometric measurements or EFW and comparing to the 10th or 90th centiles, or to monitor growth velocity by plotting subsequent measurements. Fetuses who deviate from the expected growth trajectory are noted to have either accelerative or restrictive growth.

Doppler studies of fetal vasculature (typically umbilical artery, ductus venosus or middle cerebral artery) can also be conducted at the time of fetal ultrasound to measure resistance to blood flow. Abnormalities of Doppler measurement (e.g., high resistance in the umbilical artery or low resistance in the middle cerebral artery) provide additional information which can be used to further characterise fetal growth anomalies. These are further described in section 1.5.2.

1.1.3 Defining small for gestational age

Historically, the definition of SGA has varied in the scientific literature with published thresholds including the 5th or 10th centile of birthweight for gestational age, or one or two standard deviations (SDs) below the mean.[23-31] The variation in definitions creates difficulty

when summarising studies of perinatal outcomes for SGA fetuses, or the effectiveness of potential screening or management interventions such as increased surveillance or iatrogenic birth.[1] Whilst neither formal UK nor international consensus on the definition of SGA has been reached, most international guidelines for the detection of SGA do recommend a 10th centile threshold on a birthweight chart.[1,32-37] Conversely, consensus reached through Delphi methodology has been reached to define criteria for the diagnosis of FGR (Table 1.1); these have been adopted by the International Society for Ultrasound in Obstetrics and Gynaecology (ISUOG).[37] Using these criteria, FGR should be diagnosed using either a solitary criterion (e.g. EFW or AC<3rd centile) or combinations of contributory criteria (i.e. those that require other abnormalities to be present for the diagnosis of FGR).[38]

Table 1.1 - Parameters agreed by Delphi consensus for the diagnosis of fetal growth restriction

Parameters used to define early FGR (<32 weeks')		Parameters used to define late FGR (>=32 weeks')	
Solitary parameters	**Contributory parameters**	**Solitary parameters**	**Contributory parameters**
-AC < 3rd centile OR -EFW < 3rd centile OR: -Absent end-diastolic flow (EDF) in the umbilical artery	AC or EFW < 10th centile AND: a pulsatility index (PI) > 95th centile in either the umbilical or uterine artery	-AC < 3rd centile OR: -EFW < 3rd centile	EFW/AC < 10th centile or AC/EFW crossing centiles by > two quartiles on growth charts AND: cerebroplacental ratio < 5th centile or umbilical artery PI > 95th centile

Across the viable gestational age range, the weight defined by the 10th centile threshold has shifted since the 1980s.[39] This follows a rightward shift in the population birthweight distribution that is related, at least in part, to an increasing rate of intervention to prevent and treat growth restricted fetuses during this time. The weight at which SGA is diagnosed is now higher, meaning that bigger babies are defined as SGA compared to previously.

As with any screening test, the threshold at which clinicians are recommended to act has consequences for those women and babies either side of the threshold. Any threshold for the diagnosis of SGA is arbitrary; the severity of fetal growth anomalies and associated perinatal morbidity and mortality increases along a continuum.[40-42] With lower thresholds (i.e., using the 5th rather than the 10th centile), the sensitivity of the screening test decreases (i.e., the proportion of growth restricted fetuses identified by the test), but the specificity increases (i.e., the proportion of fetuses for whom growth was not restricted who are correctly not identified by the test). For SGA, it is most important that infants at risk of perinatal morbidity and mortality are identified and monitored, and so a high sensitivity is

preferred but at the cost of low specificity, or a high false positive rate (i.e., many fetuses or infants will receive interventions that are not required).

In a retrospective population-based cohort study conducted by members of my research group, we identified that the centile threshold by which either SGA- or LGA-associated morbidity and mortality could best be predicted differed by birthweight chart used.[43] This has also been shown by another group.[44] Furthermore, the performance of birthweight charts, in terms of their ability to identify babies at risk of growth-associated morbidity or mortality, differs by the population in which they are applied. Thresholds at which fetal growth anomalies are suspected should ideally be population-specific, derived by testing each chart in the population in which it is to be used.

As the most used definition by international guidelines,[45] birth or fetal weight less than the 10th centile for gestational age (and in some cases, other adjustment characteristics) will be used to define SGA for the remainder of this thesis.

1.1.4 Defining large for gestational age

The centile threshold by which LGA is defined also differs, with the 90th centile being most commonly cited,[2] but the 95th and 97th centiles also in common use.[46,47] The same considerations apply as for SGA, high diagnostic sensitivity is partnered with low specificity, or *vice versa*, and the centile threshold which best predicts adverse outcome differs according to the chart in use,[43] and also by the population studied.

As the most commonly cited definition of LGA,[2] birth or fetal weight above the 90th centile for gestational age will be used to define LGA in this thesis.

1.1.5 Categorising weight centile charts

There are many different perinatal weight charts in use, each with their own philosophy. They can broadly be categorised as:

- descriptive (reference) or prescriptive (standard) centile charts (section 1.1.5.1),
- fetal or neonatal weight centile charts (section 1.1.5.2),
- population-based or customised weight centile charts (sections 1.1.6-1.1.7).

These categories of perinatal weight chart are described in the sections which follow. UK guidelines on antenatal care from NICE do not recommend a specified chart type.[8] Guidelines from the RCOG on the detection and management of the SGA fetus also do not directly recommend a chart type but do acknowledge that:

'Use of a customised fetal weight reference may improve prediction of an SGA neonate and adverse perinatal outcome. In women having serial assessment of fetal size, use of a customised fetal weight reference may improve the prediction of normal perinatal outcome'.[1]

1.1.5.1 *Descriptive and prescriptive centile charts*

Descriptive weight charts simply describe the distribution of the weight in a given population. An example is the distribution of birthweights for babies born in the UK during the year 1990, presented by gestational age (in days) at birth. Such charts are also called reference charts.[48]

Prescriptive weight charts are developed to demonstrate how growth is expected to occur without pathological or adverse environmental influences i.e., the expected range of optimal growth. Prescriptive charts are therefore also called 'growth standards'. These are commonly made by excluding women with pathological influences on growth (e.g., hypertension, diabetes, or low socio-economic status) from the reference population.[48]

1.1.5.2 *Fetal and neonatal weight centile charts*

Fetal weight charts are constructed using the estimated fetal weights of a reference population, whereas birthweight (neonatal) charts are constructed using the actual weight of the baby shortly after birth.

Evidence suggests that the preterm portion of birthweight charts are negatively skewed.[49-52] This is a product of preterm birth which, whether spontaneous or iatrogenic, is more likely to occur in a pathologically small baby,[53] although babies born following spontaneous onset of preterm labour are bigger than those who are iatrogenically delivered before term.[54] This theory is difficult to prove without iatrogenic birth and assessment of birthweight in healthy fetuses who would otherwise be born at term, since intrauterine weight assessment is less accurate than birthweight measurement, and may artefactually change the weight centile of the fetus.[55]

When neonatal weight charts for preterm babies are designed using the estimated weights of babies who remain intrauterine, as well as those of babies who are born early, rather than charts developed using only the birthweight of babies born early, a greater proportion of preterm babies are classified as SGA than are term babies;[56,57] babies born at extreme preterm gestations are most likely to be classified as SGA.[58] Babies defined as SGA by fetal charts are more likely to be born by spontaneous preterm birth,[59] and fetal charts are better predictive of SGA-associated preterm perinatal mortality.[60]

1.1.6   Population weight centiles

Population *references* enable comparison of a birthweight to the centiles derived from a reference population of babies born at the same gestational age. Typically speaking, the reference population is descriptive (section 1.1.5.1).[61] All babies are compared to the same reference adopting the philosophy that all fetuses have a similar growth potential at conception. Whilst normal variation is expected, the growth potential is not expected to differ significantly by maternal or paternal characteristics such as height or ethnicity. Reference centiles exist for fetal weights,[62] and for neonatal birthweights for babies born at any viable gestation.[63,64] In some cases, birthweight references are presented separately for male and female neonates.[65]

Population charts in use in the UK include the UK-WHO birthweight reference (2009)[65] which was derived from births in 1990,[66] and the updated population birthweight centiles for England and Wales, derived from national data collected in 2018 for the MBRRACE-UK (Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries) programme.[67]

Population references are limited by their generalisability between populations because the same centile thresholds are used to diagnose SGA or LGA when the reference is applied in different populations. Whilst the philosophy dictates that all fetuses have the same growth potential, the mean birthweight in some populations (e.g., in Asia) is significantly lower than the mean birthweight in others (e.g., Scandinavia). Whilst this may be because of environmental and pathological rather than physiological influences, applying a population reference derived in one population to a different population has the effect of classifying a larger (or smaller) proportion of babies as SGA, thereby affecting the test sensitivity and specificity of SGA/FGR defined by the reference to predict morbidity and mortality.

Furthermore, some fetuses defined as SGA by population birthweight references are appropriately grown (i.e., the constitutionally small fetus). Other fetuses with impaired growth may still fall within the normal range by population references, particularly if expected to have birthweights at higher centiles under optimal conditions for growth.

In contrast to customised standards (section 1.1.7 below), population references often have higher specificity, but lower sensitivity to detect fetuses and neonates with abnormal growth,[61] although this can be modified by choosing a different diagnostic threshold.[43]

1.1.7   Customised weight centiles

Customised fetal growth standards were developed with the underlying hypothesis that the fetal growth trajectory is individualised according to the physiological intrauterine environment and genetic potential. These are usually growth *standards* (prescriptive charts, section 1.1.5.1) and therefore exclude pathological pregnancies from the population in which they are developed. Customised standards are intended to differentiate between pathologically or physiologically small fetuses or neonates.

Customised birthweight charts have been developed in varying formats since the late 1960s.[68] Most versions adjust for maternal characteristics including maternal height, weight, parity and ethnicity, and neonatal sex. These are considered to be 'physiological' influences on fetal growth by proponents of this methodology. The most common version in circulation in the UK is the Gestation-Related Optimal Weight (GROW) chart, first described in 1992 (detailed in section 1.1.7.1).[69] Other customised charts have been proposed,[70-72] but are not commonly used in the UK.

1.1.7.1   *Development of Gestation-Related Optimal Weight customised centile charts*

GROW charts were developed by Professor Jason Gardosi and colleagues and are available from the Perinatal Institute (www.perinatal.org.uk), a non-profit organisation set up by Professor Gardosi. By June 2021, GROW charts had been implemented in 78% of UK maternity units.[73]

GROW charts are based on three principles:

- they are individualised (adjusted for physiological factors that affect birthweight).
- they aim to predict the optimal growth potential by excluding babies from the development population for whom pathological factors (e.g., smoking or diabetes) may have affected the birthweight.
- they use fetal growth curves (intrauterine fetal weight estimates) for the preterm gestations and birthweight curves for term gestations.[74]

Gardosi and colleagues first published a proposal for a customised fetal growth chart in 1992.[69] These first charts were developed using data from live singleton births in Nottingham using the following steps:

1. Multiple regression analysis of candidate 'physiological' influences on growth showed maternal weight and height at booking appointment, maternal ethnicity, parity, gestational age at birth and fetal sex to be independent determinates of

birthweight. Maternal weight was most strongly associated. The evidence for the physiological nature of these characteristics is considered in section 1.1.7.2.

2. For any combination of maternal characteristics, the mean birthweight at 40 weeks of gestation was plotted onto a chart with gestational age in weeks on the x-axis and birthweight on the y-axis (Figure 1.2).

    a. The maternal physiological characteristics shifted the mean birthweight at 40 weeks' up or down the vertical axis.

3. Using the assumption that birthweight is normally distributed at a given gestational age, the 10th and 90th centiles were calculated by multiplying the standard deviation by ±1.28. These centile thresholds were also drawn on the chart, enabling identification of SGA and LGA fetuses at term.

4. The shape of the customised chart prior to 40 weeks' gestational age was determined using existing work from other authors:

    a. Thomson et al (1968) demonstrated that the shape of the fetal growth trajectory at term is the same for different maternal subgroups,[68] and so this was added between 37-42 gestational weeks.

    b. The preterm portion of the chart was extrapolated backwards from the birthweight centiles at 37 weeks' gestation using Deter's established mathematical formula (1982) for intrauterine fetal weight gain.[75]

    c. Estimated fetal weight was assumed to be equivalent to birthweight at the same gestational age.

5. A second y-axis was added for the uterine fundal height (described in section 1.4.1.2) at each gestation. This was originally produced using the same scale and intersect for all women, regardless of characteristics. Gardosi acknowledged that the predictive power of the tool could be improved by also customising the fundal height measurements, and this was changed in 1995.[76]

Subsequent to the development of the first chart, a series of improvements were made to the GROW charts:

- A dataset from a larger multi-ethnic population was used and babies excluded if they were born from multiple pregnancies, pregnancies where the estimated due date was not calculated until after 24 weeks' gestation and pregnancies in which care was transferred into the maternity unit late, or they were stillborn or had congenital anomalies, leaving a total population of 41,718 neonates.[77]

- At the time of the first reference to 'optimal weight', meaning desired weight in the absence of pathological factors which affect this, in a baby born to a mother with a

specified set of physiological characteristics,[76] babies born to mothers who smoked were also excluded from the development dataset.

- A later change in methodology meant that the preterm portion of the GROW chart was now derived using a 'proportionality' adaptation to Hadlock's formula for calculating expected fetal weight at a given gestational age (GA, measured in weeks):[62]

$$\% \text{ of birthweight} = 299.1 - 31.85GA + 1.094GA^2 - 0.01055GA^3$$

For illustrative purposes, a GROW chart for a pregnant woman of British European origin, with height of 1.77m, weight of 78kg and obstetric history of a previous SGA baby has been included in Figure 1.2. The fundal height scale is on the left y-axis and the scale for estimated fetal weight/birthweight is on the right. The lines shown on the chart are for the 10th, 50th and 90th centiles. An EFW is plotted for a 2,500g fetus at 37+0 weeks of gestation, this estimate falls below the 10th centile for the gestational age, meaning that it meets the criteria to diagnose SGA antenatally.

Customised centile calculators have since been expanded internationally, with calculation of centiles specific to local populations in Ireland,[78] New Zealand,[79] Australia,[80,81] Spain,[82] the USA,[83] Slovenia,[84] France,[85] & Sweden,[86] but with similar coefficients, particularly when the charts are reproduced in populations of largely white ethnicity.[83]

Figure 1.2 - Example of a GROW chart, produced for pregnant woman of British European origin, with height 1.77m, weight 78kg and obstetric history of a previous SGA baby. The estimated fetal weight (EFW) of her baby at 37⁺⁰ weeks of gestation is plotted on the chart below the 10ᵗʰ centile line.
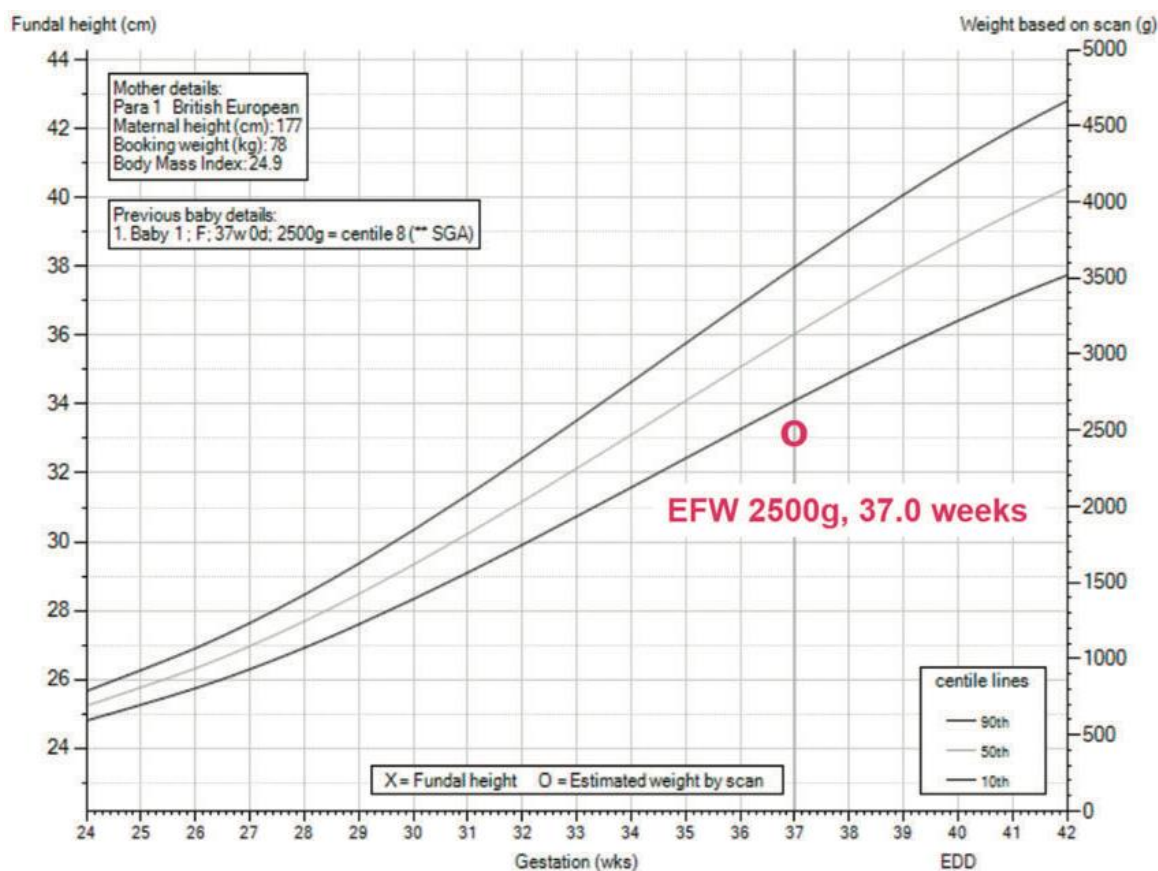
1.1.7.2 *Evidence for maternal characteristics used to calculate the customised centiles*

The published coefficients used in the GROW calculator for the English population (Table 1.2) adjust the expected optimal weight by maternal height, weight, parity and ethnicity, compared to a baseline 'standard woman' (European origin, height 163cm, weight 64kg, first ongoing pregnancy and with neonatal sex the average of male and female).[83] The evidence behind the adjustment of each characteristic is outlined below.

Table 1.2 - Coefficients to adjust optimal birthweight using maternal characteristics, in comparison to a baseline woman of European origin, height 163cm, weight 64kg, first ongoing pregnancy and with neonatal sex the average of male and female.

Variable		Coefficient
Constant		3455.6
Standard error of the model		389.0
Gestational age (calculated from 280 days)	Gestational age	20.7
	Gestational age^2	-0.213
	Gestational age^3	-0.00017
Sex	Male	48.9
	Female	-48.9
Maternal height (from 163cm)		6.7
Maternal weight (from 64kg)	Weight	9.18
	Weight^2	-0.151
	Weight^3	0.001
Parity	Para 1	101.9
	Para 2	133.7
	Para 3	140.2
	Para 4	162.7
Ethnic origin	African Caribbean	-127.5
	African	-218.5
	Middle Eastern	-89.9
	Bangladeshi	-79.3
	Indian	-149.4
	Pakistani	-187.3

1.1.7.2.1 Maternal height

A positive linear association between maternal height and neonatal birthweight was demonstrated by Thomson et al (1968).[68] More recently, Gardosi (2009) also described this relationship but with a plateauing of birthweight at the upper extreme of maternal height.[88]

In a comparison of population and customised growth charts applied to the babies born to Dutch women between 1992 and 1995 (n=220), the SGA fetuses identified through application of population charts were more likely to be born to women of short stature, when compared to SGA babies as defined by customised charts (i.e., the customised charts normalised the growth of small babies born to women of short stature).[89]

Shorter height is associated with lower socio-economic status, although the strength of this association has reduced over time (the strength of the positive association between body mass index and socio-economic status has simultaneously increased).[90] Maternal height is also associated with adverse perinatal outcomes, with shorter women at greater risk of gestational diabetes, nulliparous caesarean birth and preterm birth, but at lower risk of pre-eclampsia than taller women.[91] It is therefore evident that maternal height is not an entirely physiological influence on pregnancy.

1.1.7.2.2 Maternal weight

Neonatal birthweight increases with maternal weight.[68,69,92] This relationship also includes a pathological association with birthweight whereby underweight women (body mass index [BMI] of less than 18.5kg/m^2) have a higher risk of having an SGA baby,[1] and women with obesity (BMI of 30 kg/m^2 or above) are more likely to have an LGA or macrosomic baby, when compared to women with BMI in the healthy range (18.5-24.9 kg/m^2).[93]

Persson et al (1978) noted that whilst neonatal birthweight does increase with maternal weight, the increase in size occurs in soft tissue – there was no association between maternal weight and the size of bony infant structures such as the biparietal diameter.[92]

GROW centiles previously only adjusted for maternal weight where the mother's BMI fell within the range 20.0-30.0 kg/m^2, to avoid adjustment for pathological factors such as obesity, when determining fetal growth potential. This was later changed to adjust for maternal weight regardless of BMI.[86] This is shown in Table 1.3, where I have used the customised calculator to calculate birthweight centiles for women and their babies with the same characteristics to one another, except for maternal weight (and BMI). The change was justified by an observation that the rate of SGA amongst women with obesity increased when weight was adjusted for throughout the range, just as the rate of perinatal mortality increased. No evaluation was conducted to determine whether the babies who were now defined as SGA were those at higher risk of perinatal death.[86] The change has the effect that SGA is defined at a higher birthweight for women with higher weight, i.e. small babies are defined as such relative to maternal weight.

Table 1.3 - Change in neonatal birthweight centile with maternal weight (and BMI) for a woman of constant height (163cm), ethnicity (British European), parity (nulliparous), and male baby of birthweight 3514g at 40^{+0} weeks of gestation

Mother and baby pair	Weight (kg)	BMI	Gestational age (days)	Birthweight	Birthweight centile
1	45.2	17.0	280	3514	69.9
2	53.1	20.0	280	3514	60.7
3	66.4	25.0	280	3514	50.0
4	79.7	30.0	280	3514	42.5
5	93.0	35.0	280	3514	37.9
6	106.3	40.0	280	3514	35.1

Using customised centiles, Figueras et al (2009) demonstrated a U-shaped association between maternal weight and birthweight centile; both women who were underweight and those with obesity were at increased risk of having an SGA baby.[88] A similar association has also been shown by Zhang et al (2007) who compared customised standards to a population reference in a large Swedish cohort (782,303) and noted that babies in whom SGA was defined by only the customised standard (not by the population reference) were more likely to be born to overweight mothers than babies born SGA as defined only by population charts (and not by customised standards).[94] Whether these SGA babies were at higher risk of perinatal morbidity or mortality than larger babies born to overweight mothers was not reported.

Conversely, Sjaarda et al (2014) compared a customised standard which included customisation for maternal weight, to one which did not. LGA babies diagnosed by the model which didn't adjust for maternal weight were more likely to have neonatal morbidity (shoulder dystocia, admission to neonatal care, or respiratory complications) than LGA babies diagnosed by the model which customised for weight, meaning that adjustment for weight was reducing the ability for birthweight centiles to identify LGA at risk of adverse neonatal outcomes.[95]

1.1.7.2.3   Maternal ethnicity

The influence of ethnicity on fetal growth is complex due to the known association between ethnicity and socio-economic deprivation,[96-98] and the association between ethnicity and low birthweight or antepartum stillbirth.[99-101] Ethnicity is also associated with other maternal characteristics and behaviours, including differences in maternal vascular disease, diabetes, smoking, BMI and age.[102-104] Furthermore, ethnicity can be hard to define in a multi-ethnic population with individuals of increasingly mixed ethnicities.[105]

Many studies have demonstrated that women of white ethnicity have bigger babies than women of non-white ethnicity.[106-110] Even for women from Western Europe, where the majority are of white ethnicity, significant differences have been found in the birthweights of babies born in different countries.[111] In a Canadian systematic review, babies born to mothers from immigrant populations had higher birthweights if the mother resided in Canada, compared to if the mother resided in her respective native country.[112] The extent to which these differences represent a physiological improvement (i.e. improved nutrition as opposed to a maternal increase in high calorie diets) is unknown.

Babies born with low birthweight, or SGA (defined by references not adjusted for ethnicity) are more likely to survive when born to non-white mothers, than if they are born

to white mothers.[109,113-117] In a large cohort study by Hanley et al (2013), newborn infants classified as SGA by ethnic-specific birthweight reference charts were over twice as likely to have adverse perinatal outcomes than infants classed as SGA by population references, suggesting that use of population-specific references missed non-white neonates at risk of perinatal morbidity.[118] Conversely, in a large international cohort study (237,025 babies born in 24 countries) adjustment for maternal country of origin made little difference to the prediction of adverse perinatal outcome (area under the receiver-operator, AUROC curve of 0.679 vs. 0.699).[70]

1.1.7.2.4 Maternal parity

Mean birthweight increases with each subsequent birth, although the difference is greatest between the first and second born babies.[119] The extent to which a lower birthweight of the first-born baby is pathological is disputed.[120] Nulliparous women are more likely to experience adverse outcomes such as pre-eclampsia, neonatal morbidity and perinatal mortality, than women having their second or third birth. The risk of adverse pregnancy or birth outcomes increases again with the fourth or subsequent birth.[121] Since all of these outcomes are also associated with SGA, it follows that the baby born to a nulliparous woman that is smaller than her subsequent babies may be pathological.

Using a Swedish population-based cohort, Gardosi et al (2009) identified that the pattern with which the rate of SGA varied with parity was better reflective of the pattern with which the rate of perinatal mortality varied with parity if SGA was defined using a customised rather than a population birthweight curve (Figure 1.3).[86] However, perinatal mortality was not restricted to babies born SGA, and the perinatal mortality rate may differ by parity for reasons other than SGA, or be confounded by the pathological influence of socio-economic status or other factors.[122] It is yet to be established that these patterns should be the same.

Figure 1.3 Perinatal mortality rate (PMR) and SGA by customised (SGA$_{cust}$) and population based (SGA$_{pop}$) centiles, according to maternal parity at the start of pregnancy.
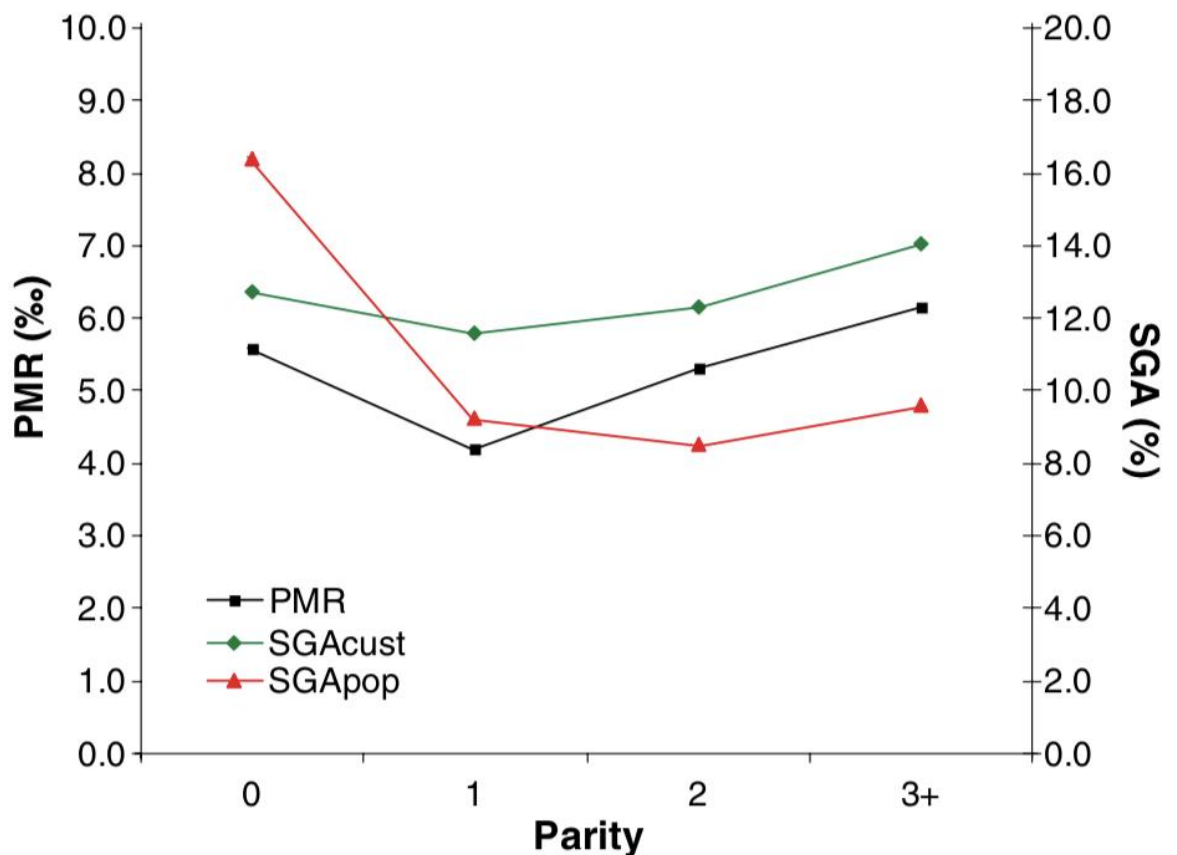
1.1.7.2.5   Neonatal sex:

It is universally accepted that male babies weigh approximately 100-150g more than female babies at term. Given that sex is a physiological difference, it can be assumed that this is a true difference which should not be attributed to pathological mechanisms.[68,110] Some population reference charts also present centiles separately for male and female neonatal birthweights.[65]

1.1.7.3   *Paternal factors which influence fetal growth*

It is notable that paternal factors are not included in the list of customisation factors used in GROW charts, despite fathers contributing half of the fetal genes. Both paternal height and paternal birthweight have been shown to have a significant influence on neonatal birthweight,[123-125] although these are often associated with the same maternal characteristics. Difficulty also arises from uncertainty surrounding attribution of paternity.

1.1.7.4 *Critique of customised growth charts*

In a cohort study, Owen et al (2002) assessed the relationship between customised birthweight centiles and neonatal anthropometric markers of fetal growth restriction. Customised birthweight centiles were only moderately useful in identifying neonates with subscapular or triceps skinfold thickness <10th centile or ponderal index <25th centile (positive likelihood ratios of 4-5 and kappa statistics of 0.3-0.4), and not useful in identifying mid-arm circumference to occipito-frontal circumference ratios of 1SD below the mean.[126]

Hutcheon et al (2008) conducted a population-based cohort study using data on Swedish births to compare the performance of a customised standard, a population standard based on birthweights and a population standard based on fetal weights in the identification of SGA fetuses at risk of stillbirth and early neonatal mortality.[127] The population fetal weight standard performed with similar accuracy to the customised standard (fetal weights for preterm babies, birthweights used for term babies), and both performed better than the population birthweight standard, leading the authors to conclude that customisation for maternal characteristics added little to the identification of fetuses at risk of perinatal mortality and that the accuracy of customised charts was related to the design which incorporated fetal weight charts (rather than birthweight charts) for preterm gestations. Similar findings have also been replicated in other studies.[128,129]

Gaillard and colleagues (2011) argue that Gardosi's strategy for generating customised fetal weight standards, which assumes that the effect of maternal characteristics on fetal weight throughout pregnancy is proportional, is flawed. In their study of 5,473 infants with fetal weight serially estimated by ultrasound scans, they identified that the influence of maternal height, weight, ethnic origin, parity and fetal sex increased with increasing gestational age.[130] This has been neither confirmed nor refuted.

1.1.8 International fetal or neonatal weight centiles

There are two other fetal or birth weight charts worthy of description in this thesis; both are high profile charts which have been recommended for international uptake:

- The INTERGROWTH-21st fetal and neonatal weight centiles (section 1.1.8.1).
- The WHO neonatal birthweight and growth centiles (section 1.1.8.2).

1.1.8.1 *The INTERGROWTH-21st fetal and neonatal weight centiles*

The INTERGROWTH-21st centiles were derived from a study which recruited low-risk, well-nourished, multi-ethnic women of optimal health, education, and socioeconomic status from eight urban populations (in Brazil, Italy, Oman, UK, USA, China, India and Kenya) in

which the health and nutritional needs of individuals were met and adequate antenatal care was provided. Two component studies ran in parallel:

- 4,607 women were recruited to the Fetal Growth Longitudinal Study (FGLS) in which ultrasound measurements of fetal biometry were taken including crown-rump length at <14 weeks' gestation and fetal HC after 14 weeks' gestation,
- 59,137 women were recruited to the Newborn Cross-Sectional Study, in which birth length was measured in all newborn babies. Only 20,486 of these women met the criteria for FGLS (educated, affluent and healthy women with adequate nutritional status).

The authors acknowledged that differences were identified between countries for crown-rump length and fetal HC but, they determined that only 1.9-3.5% of variation could be attributed to between-population differences, concluding that this was not a meaningful difference and that fetal growth and newborn length are similar across diverse settings if the nutritional and health needs of the mothers are met. They therefore pooled their growth data for the construction of a single international fetal and neonatal weight standard.[131]

A key criticism of this work is the focus on fetal skeletal measurements. SGA definitions are based on fetal and neonatal weight, which takes into account both skeletal and soft tissue.[132] The developers of the INTERGROWTH-21st centiles have since also developed tables for estimated fetal weight centiles, finding these to be consistent with the term birthweights in the INTERGROWTH-21st Newborn Size Standards, but higher than the birthweights of the same standards at preterm gestations.[133]

When these charts are applied to healthy populations in high-income countries, they result in a smaller proportion of women being diagnosed with SGA, and a higher proportion being diagnosed with LGA.[43,134] In a study by my research group, INTERGROWTH-21st charts were compared to GROW charts and an internally-produced population chart using data from a Swedish population (212,101 singleton births between 2006-2015).[43] Fewer SGA babies were diagnosed using the INTERGROWTH-21st charts, but these babies had a higher risk of adverse outcomes than the larger groups of SGA infants identified by the other two charts (i.e. the INTERGROWTH-21st charts were more specific but less sensitive for predicting adverse outcomes).

1.1.8.2 *WHO neonatal growth centiles*

More recently, the WHO published its updated version of fetal growth charts.[135] These were produced during a multinational longitudinal study of fetal growth, similar to that of the INTERGROWTH group, which recruited multi-ethnic women with low-risk singleton

pregnancies, of high or middle socioeconomic status and without known environmental constraints on fetal growth from ten countries (Argentina, Brazil, Democratic Republic of the Congo, Denmark, Egypt, France, Germany, India, Norway, and Thailand). Participants were invited for seven ultrasound assessments of the fetus during the pregnancy. In total, the serial ultrasound sets of 7,924 women were used to generate reference growth curves for fetal biometric measurements. Like the authors of the INTERGROWTH-21st study, the authors of the WHO centiles started with the assumption that data from different countries would be similar enough to be pooled. In contrast to the INTERGROWTH-21st study authors, the WHO team did acknowledge differences in fetal growth between women of different origin countries, and that fetal growth differed by maternal height, weight, age & parity. The authors of the WHO centiles concluded that adjustment of their fetal growth charts may be necessary to optimise local applicability.

1.1.9 Comparison of fetal or neonatal weight centile chart types

Different fetal and birthweight charts have been compared repeatedly in terms of their ability to identify babies most at risk of adverse perinatal outcome. However, a Cochrane systematic review conducted in 2014 to compare the performance of customised and population-based growth charts in screening for SGA in low-risk pregnant women in trial settings found that there were no randomised trials which met the inclusion criteria, nor were there an trials in progress, and none have been reported since.[136]

Research comparing the performance of different fetal or birthweight centile charts in terms of their ability to identify SGA babies at risk of stillbirth is summarised in Table 1.4, including calculated values for sensitivity, specificity, positive and negative predictive values (comparable test performance statistics were not reported in any paper and so these have been calculated using the available data). These observational studies show that stillborn babies are more likely to be identified as SGA on customised than population charts. Calculated sensitivity for this outcome ranges between 32.7-59.3% on customised charts and 28.6-52.2% on population reference charts.[85,89,94,137-139] Babies defined as SGA by both population and customised centiles (SGA$_{both}$, rather than one chart or neither) had the highest risk of stillbirth in five of the eight studies which assessed this, but with positive predictive values below 4%. Similar findings also exist for the performance of customised and population charts when used to identify SGA babies at risk of perinatal death,[85,140,141] neonatal death,[94,97,137,138,142] low Apgar score at 5 minutes after birth,[137,142-144] and neonatal unit admission.[85,89,139,141,144,145]

Table 1.4 - Comparing test performance for each chart type in terms of ability to identify SGA babies at risk of stillbirth

Chart types compared in studies	Study reference	Study population	Definition of SGA	Sensitivity[+]	Specificity[+]	PPV[+]	NPV[+]
Customised charts versus INTERGROWTH-21st charts	Francis (2018)[146]	1,250,000 term pregnancies from 10 countries (1,667 stillbirths)	SGA - customised only			0.2%	
			SGA – customised & INTERGROWTH			0.4%	
			SGA – INTERGROWTH only			0.1%	
Customised charts versus a population-based reference	Clausson (2001)[137]	326,377 births in Sweden (908 stillbirths)	SGA – population only	2.3%	97.3%	0.2%	99.7%
			SGA – customised only	11.7%	97.3%	1.2%	99.7%
			SGA – population & customised	23.9%	93.3%	1.0%	99.8%
			SGA - population (all)	26.2%	90.6%	0.8%	99.8%
			SGA - customised (all)	35.6%	90.6%	1.0%	99.8%
	Zhang (2007)[94]	782,303 singleton births >28 weeks' born to Nordic mothers (2,354 stillbirths)	SGA - population only	2.2%	97.6%	0.3%	99.7%
			SGA - customised only	13.1%	97.4%	1.5%	99.7%
			SGA - population & customised	27.4%	92.6%	1.1%	99.8%
			SGA - population (all)	29.6%	90.3%	0.9%	99.8%
			SGA - customised (all)	40.4%	90.1%	1.2%	99.8%
	Ego (2006)[85]	75,306 singletons births without congenital anomalies, occurring at 22 gestational weeks or later in 5 French hospitals (232 stillbirths)	SGA - population only	1.3%	98.2%	0.3%	99.6%
			SGA - customised only	6.0%	97.4%	0.9%	99.6%
			SGA - both	50.9%	88.2%	1.7%	99.8%
			SGA - population (all)	52.2%	86.4%	1.5%	99.8%
			SGA - customised (all)	56.9%	85.6%	1.6%	99.8%
	De Jong (1998)[89]	High-risk (for utero-placental insufficiency) Dutch population (2 stillbirths)	SGA - population (all)	50.0%	86.0%	3.2%	99.5%
			SGA - customised (all)	50.0%	69.3%	1.5%	99.3%
	Rowan (2009)[139]	212 singleton births to women with T2DM (5 stillbirths)	SGA - both	40.0%	94.7%	15.4%	98.5%
			SGA - customised only	40.0%	91.8%	10.5%	98.4%

Chart types compared in studies	Study reference	Study population	Definition of SGA	Sensitivity[+]	Specificity[+]	PPV[+]	NPV[+]
Customised charts versus a population standard	Figueras (2007)[143]	13,661 non-anomalous singleton babies (80 stillbirths)	SGA – population only	2.5%	98.5%	1.0%	99.4%
			SGA – customised only	12.5%	95.9%	1.8%	99.5%
			SGA – population & customised	21.3%	90.5%	1.3%	99.5%
			SGA - population (all)	23.8%	89.0%	1.3%	99.5%
			SGA - customised (all)	33.8%	86.4%	1.4%	99.6%
	Gibbons (2013)[147]	54890 singleton term births in Australia (67 stillbirths)	SGA - population only	1.5%	98.9%	0.2%	99.9%
			SGA - customised only	7.5%	95.7%	0.2%	99.9%
			SGA - both	31.3%	92.5%	0.5%	99.9%
			SGA - population (all)	32.8%	91.3%	0.5%	99.9%
			SGA - customised (all)	38.8%	88.2%	0.4%	99.9%
	Anderson (2012)[97]	25,976 singleton pregnancies born in NW 2006-2009 (150 stillbirths)	SGA – population only	2.0%	97.4%	0.4%	99.4%
			SGA – customised only	6.7%	96.6%	1.1%	99.4%
			SGA – population & customised	52.7%	91.9%	3.6%	99.7%
			SGA - population (all)	54.7%	89.2%	2.9%	99.7%
			SGA - customised (all)	59.3%	88.5%	2.9%	99.7%
Customised charts developed in the Korean population versus population reference charts	Cha et al (2012)[144]	9052 non-anomalous, low-risk Korean babies born at 28-42 weeks'. (7 stillbirths)	SGA - population only	0.0%	96.4%	0.0%	99.9%
			SGA - customised only	0.0%	97.2%	0.0%	99.9%
			SGA – population & customised	28.6%	92.3%	0.3%	99.9%
			SGA - population (all)	28.6%	88.7%	0.2%	99.9%
			SGA - customised (all)	28.6%	89.6%	0.2%	99.9%

Chart types compared in studies	Study reference	Study population	Definition of SGA	Sensitivity[+]	Specificity[+]	PPV[+]	NPV[+]
Customised centiles versus a birthweight population standard and an intrauterine population standard	Hutcheon (2008)[127]	782303 singleton births from Swedish registry born at or after 28/40 (2,354 stillbirths) – same population as Zhang et al (2007).[94]	SGA - birthweight (all)	29.2%	90.2%	0.9%	99.8%
			SGA - intrauterine (all)	38.5%	90.7%	1.2%	99.8%
			SGA - customised (all)	40.4%	90.1%	1.2%	99.8%
	Smith (2014)[128]	49 singleton stillborn babies born after 24/w who had received an USS within 1 months of birth, 4 matched controls to each stillbirth (197 livebirths)	SGA - birthweight (all)	12.2%	97.0%	50.0%	81.6%
			SGA - intrauterine (all)	28.6%	90.4%	42.4%	83.6%
			SGA - customised (all)	32.7%	86.3%	37.2%	83.7%

[+] All values were calculated from the raw numerators and denominators in the published manuscripts (test performance statistics not published).

Research comparing the performance of different fetal or birthweight centile charts in terms of their ability to identify LGA babies at risk of LGA-associated morbidity and mortality are summarised in Table 1.5, including reported sensitivity, specificity, positive and negative predictive values, or calculated values where these were not reported but data were available for the calculation. In contrast to studies on SGA, no studies reported the risks for babies that are LGA by both population and customised centiles (LGA$_{both}$). With regards to the type of weight centile charts most suited to identification of LGA babies at risk of perinatal death, there were conflicting findings. One study compared customised to population centiles for risk of stillbirth, finding customised centiles to have higher sensitivity (16.3% versus 12.2%).[128] Customised centiles were not superior to population centiles for the identification of LGA fetuses at highest risk of perinatal death in a different study (sensitivity 8.3% versus 9.7%),[95] or in identifying babies likely to be born with low Apgar scores or hypoglycemia.[95,145] Mothers with babies identified as LGA by customised centiles(LGA$_{cust}$) were more likely to experience adverse intrapartum outcomes (postpartum haemorrhage, shoulder dystocia and severe perineal tear) when compared to mothers giving birth to babies defined as LGA by population centiles (LGA$_{pop}$, positive predictive values detailed in Table 1.5).[95,145,148]

Table 1.5 - Performance of each type of weight centile chart in identifying LGA babies' or their mothers' risk of adverse outcome

Outcome	Study reference	Study population	Definitions of LGA compared	Sensitivity	Specificity	PPV	NPV
Stillbirth	Smith (2014)[128]	49 singleton stillborn babies born after 24/w who had received an USS within 1 months of birth, 4 matched controls to each stillbirth	LGA birthweight (all)	12.2%	81.7%	14.3%	78.9%
			LGA intrauterine (all)	12.2%	88.3%	20.7%	80.2%
			LGA customised (all)	16.3%	77.7%	15.4%	78.9%
Perinatal death	Sjaarda (2014)[95]	110,447 singleton term births	LGA population (all)	9.7%[+]	91.5%[+]	0.2%[+]	99.9%[+]
			LGA customised (all)	8.3%[+]	94.4%[+]	0.2%[+]	99.9%[+]
	Vieira (2019)[43]	212,101 singleton births in Sweden	LGA population (all)	8%[+]	90%[+]		
			LGA customised (all)	7%[+]	92%[+]		
			LGA INTERGROWTH (all)	15%[+]	75%[+]		
Low Apgar	Gonzalez-Gonzalez (2015)* [145]	1,921 with DM/GDM	LGA pop (all)	66.7%	80.4%	1.1%	99.9%
			LGA customised (all)	50.0%	84.9%	1.0%	99.8%
	Sjaarda (2014)[+][95]	110,447 singleton term births	LGA pop (all)	10.4%[+]	91.5%[+]	0.4%[+]	99.7%[+]
			LGA customised (all)	6.5%[+]	94.5%[+]	17.5%[+]	84.9%[+]
Hypoglycemia	Sjaarda (2014)[95]	110,447 singleton term births	LGA pop (all)	18.8%[+]	91.6%[+]	1.9%[+]	99.2%[+]
			LGA customised (all)	13.2%[+]	94.5%[+]	2.1%[+]	99.2%[+]
	Costantine (2013)[149]	2083 neonates born to mothers included in an RCT on treatment for mild GDM	LGA population (all)	10.9%[+]	90.9%[+]	18.5%[+]	84.4%[+]
			LGA customised (all)	16.6%[+]	87.0%[+]	19.5%[+]	84.7%[+]
Postpartum haemorrhage	Sjaarda (2014)[95]	110,447 singleton term births	LGA population (all)	9.8%[+]	91.5%[+]	6.6%[+]	94.3%[+]
			LGA customised (all)	6.3%[+]	94.4%[+]	6.5%[+]	94.3%[+]
	Pasupathy (2012)[148]	2668 infants born to nulliparous women in New Zealand and Australia.	LGA population (all)	22.5%	89.4%	9.7%	95.8%
			LGA customised (all)	20.9%	90.3%	9.9%	95.7%

Outcome	Study reference	Study population	Definitions of LGA compared	Sensitivity	Specificity	PPV	NPV
Shoulder dystocia	Sjaarda (2014) [95]	110,447 singleton term births	LGA population (all)	35.9%[+]	93.4%[+]	9.4%[+]	90.6%[+]
			LGA customised (all)	22.7%[+]	96.1%[+]	10.0%[+]	90.0%[+]
	Cha (2012)[150]	8279 singleton Korean women who gave birth between 37-41 weeks. Excluded congenital anomalies and maternal co-morbidities	LGA population (all)	32.8%	93.6%	3.5%	99.5%
			LGA customised (all)	43.1%	89.4%	2.8%	99.6%
	Gonzalez-Gonzalez (2015) [145]	1,921 with DM/GDM	LGA population (all)	85.0%	81.0%	4.5%	99.8%
			LGA customised (all)	85.0%	85.5%	5.8%	99.8%
Severe perineal tear	Sjaarda (2014) [95]	110,447 singleton term births	LGA population (all)	13.6%[+]	93.1%[+]	6.5%[+]	93.6%[+]
			LGA customised (all)	10.7%[+]	96.0%[+]	8.5%[+]	91.5%[+]

Apgar score <5 at 5 minutes, +Apgar score<5 at 4 minutes

⁺ These values were reported in the published manuscripts, all other values were calculated from the raw numerators and denominators.

1.2 AETIOLOGY OF FETAL GROWTH ANOMALIES

With many biological measurements, it is generally accepted that there will be normal variation around a mean. For birthweight, this means that some babies will be healthy and small, others will be healthy and large. Most babies born SGA or LGA have normal perinatal outcomes. The challenge comes in distinguishing healthy small (or large) babies, whose growth continues along its expected trajectory, from those in whom intrauterine growth was abnormal (growth did not meet its expected trajectory) and from those who are at risk of perinatal morbidity and/or mortality. In this section, I will summarise current understanding of the aetiology of pathological SGA and LGA.

1.2.1 Aetiology of pathologically small for gestational age growth status

Pathological SGA can be attributed to a variety of pathological processes; the prevalence of each cause differs by the gestation at onset, and by country. The most common causes are:

- placental insufficiency (section 1.2.1.1) including that which occurs in multiple pregnancy,
- intrauterine infection (section 1.2.1.2),
- genetic disorders, including chromosomal and single gene anomalies (section 1.2.1.3),
- structural anomalies (section 1.2.1.4).

It is controversial as to whether the cause affects the FGR phenotype.[42,151-154] Fetuses with early-onset growth restriction, particularly if caused by genetic changes, intrauterine infection or associated with structural anomalies have been described as symmetrically small i.e., all fetal biometric measurements are small. Late onset fetal growth restriction, most often secondary to placental insufficiency, is expected to cause a more asymmetrically small phenotype i.e., relative sparing of the fetal head size in comparison to a small abdominal circumference. The different aetiologies are described in further detail below.

1.2.1.1 *Placental insufficiency*

Placental insufficiency is a wide-ranging term that refers to a situation in which the materno-fetal interface is unable to supply the fetus with its metabolic requirements. Macroscopic and microscopic placental lesions can broadly be categorised into maternal stromal-vascular lesions, fetal stromal-vascular lesions, infectious inflammatory lesions (see section 1.2.1.2), immune/idiopathic inflammatory lesions and other placental processes.[155,156] Each of these will be described below.

Maternal stromal-vascular lesions are caused by errors in the development of the placenta, preventing adequate exchange of solutes across the maternal-fetal interface. Evidence of global or partial maternal malperfusion is one of the most common findings on placental histology following pregnancies affected by FGR.[155] Placental development commences during the early embryological phase during which the outer cell mass of the blastocyst differentiates into the cytotrophoblast and syncytiotrophoblast. The syncytiotrophoblast produces proteolytic enzymes that invade the uterine stroma forming projections and lacunae. The lacunae eventually coalesce into the single intervillous space. The cytotrophoblast cells then proliferate down the syncytiotrophoblast projections, invading the distal maternal spiral arteries and remodelling their endovascular surface. This remodelling converts the muscular spiral artery branches into inert, flaccid arteries incapable of vasoconstriction.[157] Deficient remodelling of these spiral arteries can result in:

- higher velocity flow of blood into the intervillous space causing less even perfusion and less time for solute exchange,
- pulsatile flow into the intervillous spaces causing mechanical damage to the placenta,
- smooth muscle artery vasoconstriction causing intermittent perfusion,
- atheroma formation within the arteries that narrows their lumen,
- oxidative stress (caused by a combination of the above processes) and subsequently, areas of infarction within the placenta.[158]

Fetal stromal-vascular lesions refer to a group of lesions in which the fetus is less able to perfuse the materno-fetal interface at the placenta. This category of pathological processes associated with FGR and SGA is less common than the maternal vascular category.[159] These include lesions of the fetal side of the placenta caused by erroneous development (such as delayed maturation of the villi or dysmorphic villi), velamentous insertion of the umbilical cord into the placenta, single umbilical artery, obstructive lesions of the umbilical cord (true knots, thrombosis) and large or small vessel rupture. These lesions are less likely to recur in subsequent pregnancies than maternal vascular lesions,[155] and are more common amongst women with hypertensive disorders including chronic or gestational hypertensive diseases.[160]

Villitis of unknown aetiology and chronic histiocytic intervillositis are both examples of immune or idiopathic-mediated placental inflammation. Villitis of unknown aetiology is more often seen in placentas of FGR babies born at term,[161] but can also be seen in 15% of placentas from uncomplicated pregnancies.[162] It is characterised by infiltration of maternal T-cells and fetal macrophages into the villous stroma and frequently hypothesised to be caused by a maternal immunological response to a semi-allogenic fetus.[161] Chronic

histiocytic intervillositis is a rare placental condition characterised by infiltration of maternal macrophages into the placental intervillous space. It is strongly associated with FGR, miscarriage and stillbirth, with recurrence rates between 25-100%.[155,163]

Finally, the unclassified category of placental causes for FGR includes an uncommon pathological process that is important because of the strength of its association with FGR and its recurrence risk. Massive peri-villous fibrin deposition (maternal floor infarction) is thought to be caused by inappropriate activation of the maternal clotting cascade, resulting in deposition of fibrin in the intervillous space, preventing exchange of solutes at the materno-fetal interface.[164] It was identified in up to 8.7% of FGR cases in an Italian cohort of 355 FGR pregnancies.[165] It has a recurrence risk of between 40-60%.[155]

1.2.1.2 *Infectious causes*

Prevalence of congenital infection as a cause for FGR varies across the world. In areas where malaria is endemic, approximately 50% of women who give birth to a low birthweight baby (as a proxy for SGA or preterm birth where pregnancies are not accurately dated) have malaria infection antenatally.[166] RCOG guidance on investigating SGA recommends screening for toxoplasmosis and cytomegalovirus in severely SGA fetuses,[1] despite the referenced systematic review noting no association between SGA and toxoplasmosis.[167] Whilst cytomegalovirus has been found in approximately 10% of cases of FGR in the USA,[168] a Japanese study found it in only 1.8% of FGR cases.[169] Where FGR is attributed to infectious causes, the pathophysiology is also placental with cytomegalovirus causing placental villitis and malaria being associated with an intervillositis.[155]

1.2.1.3 *Genetic anomalies*

In a systematic review of cohort studies including a total of 874 babies with apparently isolated FGR (no associated structural anomalies identified) diagnosed in any trimester of pregnancy, a mean of 6.4% of babies were found to have a chromosomal anomaly (22 numeric anomalies and 8 structural chromosome anomalies). Trisomies 21, 18 and 13 were most commonly identified. This study is however limited by the inclusion of pregnancies since 1983 and the subsequent advancement of ultrasound technology and genetic test accuracy, as well as the lack of differentiation between early and late FGR. More recently, a systematic review and meta-analysis including studies that examined the additional yield of chromosomal microarray following a normal karyotype in fetuses with growth restriction diagnosed copy number variants in an additional 4% of fetuses with isolated FGR and 10% of fetuses with associated congenital anomalies. The most frequently identified

copy number variants were 22q11.2 duplication, Xp22.3 deletion, and 7q11.23 deletion, particularly in isolated FGR.[170]

Early-onset SGA is also associated with some rare monogenic disorders. These either cause a phenotype of early FGR with AC and fetal long bones measuring below the 3rd centile (e.g. Silver Russel syndrome, Fanconi anaemia) or a phenotype in which the long bones are short but the AC and HC measure normally (the skeletal dysplasias, achondroplasia, Noonan syndrome).[171]

1.2.1.4 *Congenital structural anomalies*

Congenital anomalies, in the absence of genetic disease, are associated with FGR in approximately 1-2% of cases. Common examples include congenital heart disease, abdominal wall defects (e.g., gastroschisis), or anencephaly. The severity of FGR has also been found to be related to the number of structural defects.[168]

1.2.2 Aetiology of pathological large for gestational age growth status

Pathologically LGA babies have an excess of soft tissue with a high weight to length ratio,[172] which is likely to be caused by fetal hyperglycaemia, hyperinsulinemia and release of insulin-like growth factors. This is prompted by the placental passage of high concentrations of maternal glucose. It therefore follows that having an LGA baby is more common amongst women with metabolic dysfunction, most often caused by maternal obesity or diabetes.[2]

In a large UK cohort study of over 350,000 neonates, LGA was twice as common in babies born to women with a BMI above 30kg/m[2], three times as likely in babies born to women with gestational diabetes (GDM) and, seven times as common in babies born to women with pre-existing diabetes.[173] Having an LGA baby is also more common for women with high gestational weight gain or dyslipidemia.[2]

1.3 SEQUELAE OF FETAL GROWTH ANOMALIES

In this section, I will summarise the current knowledge about the potential antenatal, intrapartum, neonatal, infant and adult consequences for babies with fetal growth anomalies. These include consequences for babies in whom a diagnosis is made, and consequences for babies in whom the growth anomaly is not recognised.

1.3.1 Sequelae for the growth restricted fetus

Sequelae for the SGA fetus can be divided into antenatal, intrapartum, neonatal, intermediate (infant and child) and long-term (adult) consequences (Figure 1.4). The evidence for these sequelae is detailed in sections 1.3.1.1-1.3.1.5.

Figure 1.4 – Potential short, intermediate, and long-term sequelae for the small for gestational age fetus.

Antenatal
- Antenatal stillbirth

Intrapartum
- Operative birth for fetal distress
- Low Apgar score at birth
- Birth asphyxia
- Intrapartum stillbirth

Neonatal
- Admission to neonatal unit
- Need for respiratory support
- Hypoglycemia
- Hypothermia
- Major morbidity: intraventricular haemorrhage, hypoxic ischaemic encephalopathy, necrotising entrocolitis, bronchopulmonary dysplasis, pulmonary haemorrhage, infection
- Neonatal death

Infant and child
- Cerebral palsy
- Learning difficulties (speech disorders, attention deficit, behavioural problems)
- Shorter stature
- Infant death

Adult
- Coronary heart disease
- Stroke
- Hypertension
- Type 2 diabetes

1.3.1.1 *Antenatal sequelae*

Being SGA is strongly associated with a propensity for stillbirth and neonatal death.[3] Internationally, stillbirth affects approximately 2.6 million women every year.[174] Whilst stillbirth rates are highest in low- and middle-income countries, rates in the UK are higher than in similar Western European countries.[175,176]

Small for gestational age itself is not a cause for stillbirth, but up to 57% of babies who die in utero are SGA (dependent on birthweight chart applied).[3,177,178] This estimate is however associated with uncertainty because the gestational age at time of death (and therefore cessation of growth) is often not known and fetuses may lose up to 20% of body weight through intrauterine maceration and ex-utero dehydration prior to birthweight measurement.[179]

Some of the risk factors for SGA are also attributed to stillbirth, as are the causes for SGA. This explains why many stillborn babies are SGA at birth. In a meta-analysis of major risk factors for stillbirth in high-income countries, Flenady et al (2011) identified that overweight/obesity (population-attributable risk, PAR: 8-18%), age >35 years (PAR: 7-11%), smoking (PAR, 4-7%) and nulliparity (PAR: 15%) were all important risk factors for stillbirth, as they are for SGA (sections 1.2.1.1 and 1.4.1.1).[180]

The proportion of stillbirths attributed to potential causes varies dependent upon the classification system used and depth of investigation conducted. In a systematic review of systems aiming to identify stillbirth aetiology, Aminu et al (2017) demonstrated that the proportion of stillbirths with no known cause varied between less than 1% and up to 50% dependent on classification system applied.[181] In a different systematic review assessing global causes of stillbirth, 5 high quality reports including 6,194 stillbirths were mapped to the International Classification of Diseases Perinatal Mortality (ICD-PM) matrix. Amongst these, a fetal growth disorder was attributed in 1,080 (17.4%) of stillbirths (including 435 pregnancies in which a complication of the cord, placenta or membranes was also identified). Congenital malformations were attributed as the primary cause in a further 1,344 (21.7%) pregnancies, these may also be associated with SGA although small size is not necessarily the cause of the intrauterine death in these cases.[182]

The severity of FGR is directly related to the risk of intrauterine fetal death, regardless of gestational age.[40-42] The risk of stillbirth for SGA infants also increases during each intrauterine week after 37 weeks' gestation.[183] Retrospective cohort data has identified that antenatal detection of SGA reduces the rate of stillbirth.[184]

1.3.1.2 *Intrapartum sequelae*

Growth restricted fetuses are often chronically hypoxic before labour commences and are therefore less likely to tolerate transient reductions in placental blood flow caused by uterine contractions. This may cause suspected 'fetal distress' as defined by abnormalities in the fetal heart rate pattern on continuous cardiotocograph monitoring or, when most severe, the consequences of birth asphyxia including neonatal hypoxic ischaemic encephalopathy (HIE) or intrapartum stillbirth.[185] In cases of lesser severity, operative birth (caesarean or assisted vaginal birth), a low Apgar score at birth, need for neonatal respiratory assistance at birth and admission to the neonatal unit are all more common for SGA babies[42,186] The risk increases with increasing severity of SGA.[42,187]

1.3.1.3 *Neonatal sequelae*

SGA infants are more likely to be affected by minor neonatal morbidity (hypothermia, hypoglycaemia), major morbidity (intraventricular haemorrhage, necrotising enterocolitis, bronchopulmonary dysplasia, pulmonary haemorrhage, infection) and neonatal death than infants born AGA.[185,188-190] These risks are somewhat increased by management strategies for SGA which include iatrogenic preterm birth to offset the risk of stillbirth.

Culture-proven sepsis is also more common in SGA neonates, possibly related to a compromised immune system,[191] although there is no association at extremely preterm gestations, whereby prematurity is the most dominant risk factor.[189]

1.3.1.4 *Infant and child sequelae*

Babies born SGA are more likely to die in infancy from sudden infant death syndrome, sudden unexpected death or other causes.[192]

Babies born SGA at term are more likely than AGA babies to later be diagnosed with cerebral palsy. Cerebral palsy is a heterogeneous syndrome with complex aetiology. In a large Australian population study, approximately 50% of babies who were born at term and later diagnosed with cerebral palsy had a co-existent congenital structural anomaly. Preterm babies are more likely to be diagnosed with cerebral palsy than term babies, but growth status of babies born preterm does not increase this risk, suggesting that prematurity is a more important cause of cerebral palsy than growth status in this group.[193] However, children born after 32 weeks' with a birthweight <3rd centile for gestational age (using a fetal weight population reference) included in a European registry study were 8.4 times more likely to have cerebral palsy than children whose birthweight was between the 25th and 75th centiles, and children with birthweight between 3rd and 10th centiles were 2.5

times more likely.[194] SGA is also associated with higher rates of other learning difficulties including speech disorders, attention deficits or behavioural problems, which can affect school performance.[185,195]

Where FGR is antenatally detected, infants are more likely to be affected by gastrointestinal/urogenital disorders and respiratory disorders, although much of this may be caused by iatrogenic prematurity intended to prevent stillbirth, rather than SGA itself.[196]

Infants born SGA are generally shorter during childhood and as adults (typically achieving a final height under one SD below the mean). This is achieved following catch-up growth during the first 12 months of life and a slowing thereafter. Catch-up growth, although not birthweight, is associated with obesity in later life.[195,197]

1.3.1.5 *Sequelae in adult life*

Being SGA at birth has repeatedly been shown to be associated with higher rates of coronary heart disease, stroke, hypertension, and type 2 diabetes. This appears to occur independently of risk factors such as smoking, obesity and socio-economic class in adult life, although the extent to which this is independent from the same (passive) exposures in early life is unclear.[198] It is hypothesised that these associations occur because of developmental plasticity which is reactive to the early life environment. Furthermore, impaired growth in infancy or rapid childhood weight gain exacerbate these associations.[199,200]

1.3.2 Sequelae for the large for gestational age fetus

Sequelae of LGA can also be presented according to the time frame in which they occur (Figure 1.5). The risk of these adverse outcomes is associated with high birthweight in a dose-dependent manner.[201,202] The evidence for these associations is summarised in sections 1.3.2.1-1.3.2.4.

Figure 1.5 - Potential short-, intermediate-, and long-term sequelae for the large for gestational age fetus

Antenatal
- Stillbirth (uncertain)

Intrapartum
- Prolonged labour
- Assisted vaginal birth
- Emergency caesarean brith
- Shoulder dystocia
- Severe perineal trauma
- Postpartum haemrrhage
- Neonatal birth trauma (brachial plexus injury, clavicular fracture)

Neonatal
- Admission to the neonatal unit
- Hypoxic ischaemic encephalopathy
- Transient tachypnoea of the newborn
- Hyperbilirubinaemia
- Hypoglycemia

Infant and child
- Cerebral palsy
- Metabolic syndrome
- Overweight or obese

1.3.2.1 *Antenatal sequelae*

Evidence from large cohort studies is conflicting regarding whether macrosomia or LGA are associated with an increased risk of stillbirth. In a study conducted in a large London population (n=350,311), outcomes of LGA and macrosomia (>4,000g birthweight) were compared to those of normally grown babies (AGA or birthweight 2,500-4,000g respectively). There were no differences in stillbirth with either comparison.[173] Conversely, in a large population based study from the USA (n= 10,733,983 singleton births), the rate of fetal death increased with increasing birthweight for both non-diabetic and diabetic women (above a threshold of 4249g and 3999g, respectively).[203] Another population study from Scotland (n=979,912 term singleton births), birthweight above the 90th centile was associated with a higher rate of stillbirth, but not of perinatal death (includes stillbirth and infant death).[204]

1.3.2.2 *Intrapartum sequelae*

Fetal macrosomia and LGA are both associated with prolonged labour, assisted vaginal birth, emergency caesarean birth, shoulder dystocia, second- and third-degree perineal trauma, postpartum haemorrhage, low Apgar score and admission to the neonatal unit, independently of maternal characteristics (ethnicity, BMI, parity, age, pre-existing hypertension/pre-eclampsia, pre-existing or gestational diabetes and smoking).[173,205,206]

1.3.2.3 *Neonatal sequelae*

LGA infants are at risk of the consequences of complicated birth, including birth trauma (upper limb fractures, brachial plexus injuries), HIE and acidosis.[172,205,207] They also have higher rates of transient tachypnoea of the newborn, although this may be related to higher rates of caesarean birth, [206,207] and of neonatal hyperbilirubinemia or hypoglycaemia.[172,206]

1.3.2.4 *Infant, child, and adult sequelae*

Birthweight above the 97th centile for gestational age is also associated with cerebral palsy,[194] and LGA babies born to women with GDM or obesity are more likely to be diagnosed with childhood metabolic syndrome (compared to AGA babies, or those born to women without obesity or diabetes).[208] Larger birthweight is associated with being overweight or obese in young adulthood.[209]

1.4 SCREENING AND DIAGNOSIS OF FETAL GROWTH ANOMALIES

International guidelines for healthcare professionals providing maternity care, including those in the UK, highlight the importance of recognising fetal growth anomalies, and in particular of SGA.[1,32-36] In a series on stillbirth published in the Lancet, strategies to screen for and manage SGA were identified as important in the effort to tackle stillbirth rates in high income countries.[174,176] In this section, I will describe the evidence behind the currently available approaches to diagnosing fetal growth anomalies.

1.4.1 Screening for and diagnosing SGA

Studies have reported varying success for the detection of SGA through routine antenatal care. Detection rates range from 21% to 50%, varying according to the care provided, or inclusion of any of the following strategies (sections 1.4.1.1-1.4.1.6) in the care pathway.[4,210-216]

1.4.1.1 *Risk stratification*

In the UK, it is recommended that all pregnant women are assessed at their first antenatal appointment to determine whether their risk of having an SGA baby is high or low.[1] In England, national guidance on risk assessment is available from National Health Service (NHS) England's Saving Babies' Lives care bundle.[217] At the time of the DESiGN trial, the risk stratification protocol of version 1.0 of this care bundle was in circulation (Figure 1.6); this has since been updated (March 2019). The risk assessment tool was adapted from a strategy first recommended by the RCOG Green-top guideline on the investigation and management of the SGA Fetus.[1] Broadly speaking, the NHS England risk assessment tool categorises women into those who are at low or high-risk of having an SGA baby according to the presence of a major risk factors (e.g. maternal age >40 years, previous SGA baby, essential hypertension).

Figure 1.6 - NHS England's algorithm and risk assessment tool for screening and surveillance of fetal growth in singleton pregnancies

Low Risk
☐ No known risk factors

Increased Risk: one or more of the following:

Maternal Risk Factors
☐ Maternal age >40 years
☐ Ongoing smoker (at booking)
☐ Drug misuse
Previous Pregnancy History
☐ Previous SGA baby (<10th centile)
☐ Previous stillbirth
Maternal Medical History
☐ Chronic hypertension
☐ Diabetes
☐ Renal impairment
☐ Antiphospholipid syndrome
Unsuitable for monitoring by fundal height- e.g.
☐ Large fibroids
☐ BMI >35

Current Pregnancy Complications
Early Pregnancy
☐ PAPP-A <0.415 MoM
☐ Fetal echogenic bowel
Late Pregnancy
☐ Severe pregnancy induced hypertension
or pre-eclampsia (=PIH and proteinuria)
☐ Unexplained antepartum haemorrhage

No risk factors

One or more risk factors

Low Risk Care
Serial assessment (2-3 weekly) of symphysis fundal height (SFH) from 26-28 weeks until delivery SFH measurements plotted on chart

Suspected abnormal growth (SFH <10th centile or serial measurements which demonstrate slow or static growth)

Direct referral for assessment (as soon as practically possible and should be within 72 hours) for estimated fetal weight (EFW), liquor volume and umbilical artery Doppler

Abnormal growth or abnormal umbilical artery pulsatility index

High Risk Care
Serial assessment of fetal weight and umbilical Doppler from 26-28 weeks until delivery. EFWs plotted on chart

Normal

Refer to RCOG guidance on management of the SGA fetus

68

1.4.1.2 *Measurement of fundal height*

Women who are assessed to be at low risk of SGA commonly have fetal growth monitored throughout pregnancy using serial fundal height measurements.[1,32-36] A fundal height is measured from the top of the uterine fundus to the superior aspect of the symphysis pubis (Figure 1.7). Fundal height should not be measured more often than fortnightly because the increment in fetal growth at lower intervals is less than the measurement error.[1] The RCOG recommend measuring fundal height at each antenatal appointment from 24 weeks of pregnancy.[1]

Figure 1.7 - Method of measuring the fundal height.
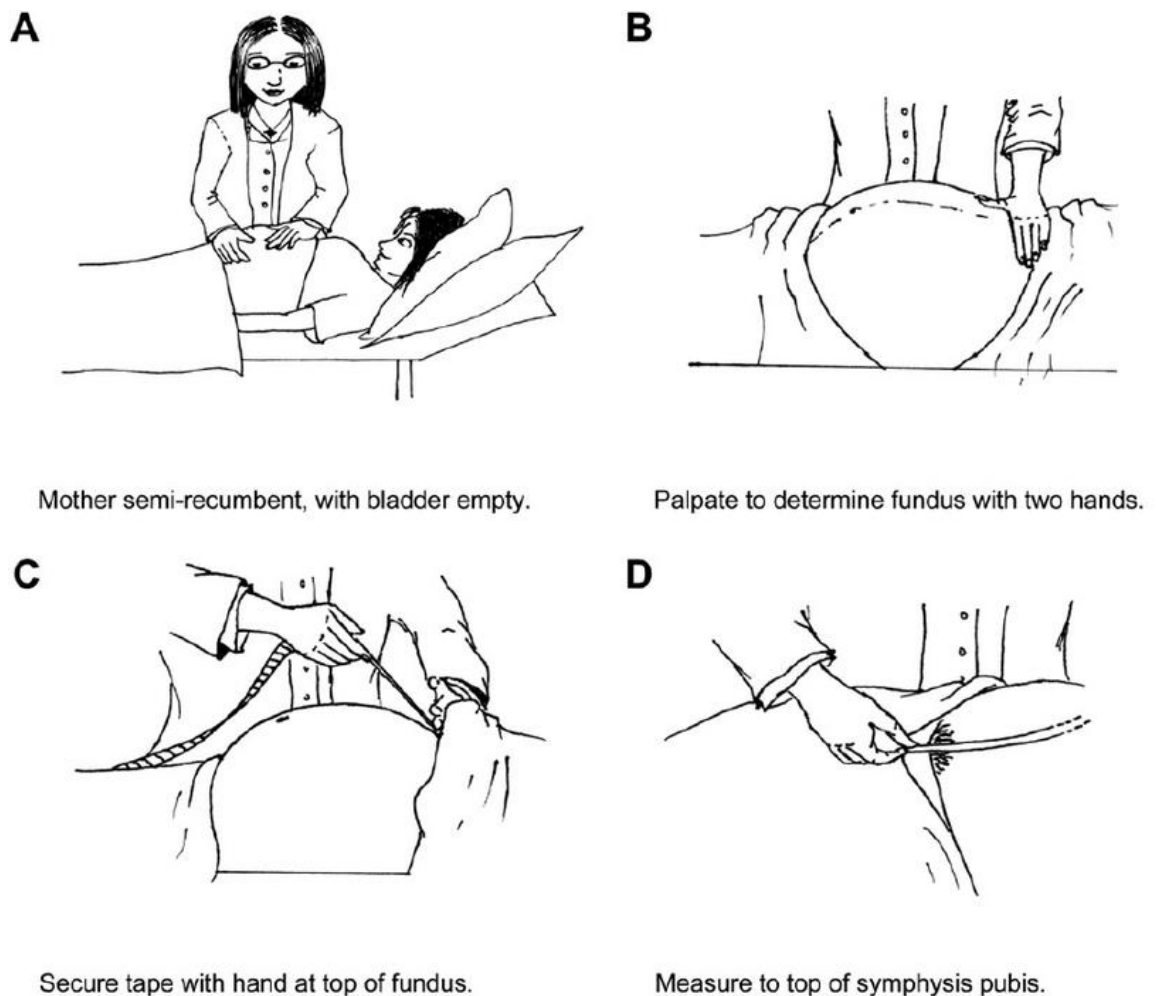
A

Mother semi-recumbent, with bladder empty.

B

Palpate to determine fundus with two hands.

C

Secure tape with hand at top of fundus.

D

Measure to top of symphysis pubis.

Fundal height measurements have been criticised because they neglect to account for variation in maternal size, normal variation in amniotic fluid and engagement of the presenting part into the pelvis.[1] There is also significant inter-observer variation that affects clinician agreement on whether a fetus is suspected to be small,[219] and measurements are

subject to clinician bias,[220] these limitations can be reduced by having the same clinician take serial measurements using a non-marked tape.[220]

A meta-analysis conducted to assess the diagnostic accuracy of fundal height measurements for SGA identified heterogeneity in the threshold at which SGA was suspected. The intervention was found to have an AUROC of 0.82, suggesting moderate accuracy.[221] In a Cochrane review (2000) of trials comparing fundal height measurement with serial ultrasound of fetal growth or clinical palpation to detect abnormal fetal growth, the authors identified only one trial (Lindhard, 1990)[222] which met the inclusion criteria.[223] In this study, there was no difference in the incidence of SGA, or in rates of perinatal mortality, neonatal hypoglycemia, admission to a neonatal unit, induction of labour or caesarean birth when serial fundal height measurement was compared to abdominal palpation. No prospective clinical trials have been conducted that compare assessment of fundal height using customised and non-customised charts.

Guidelines on serial fundal height measurement differ in terms of the expected measurement at each gestational week. Fundal height measurements are either assessed using the McDonald rule,[224] or compared to centile charts which may be based on population references or customised standards.[218] The 'McDonald rule' method assumes that fundal height should be equal to gestational age in weeks. An acceptable margin of variation and error is ±2cm or ±3cm.[30,31,224]

Gardosi and colleagues developed the first customised charts for fundal height measurements in 1995,[76] and compared these to assessment using abdominal palpation only in a non-randomised controlled trial (n=1,272).[225] The rate of SGA detection increased from 29% with abdominal palpation to 48% with serial fundal height measurement (2-3 weekly, starting at 26 weeks') plotted onto a customised chart. There was also an increase in the rate of detection of LGA (24% to 46%), but no differences in perinatal outcome. In an observational study, Roex et al (2012) compared serial fundal height measurements plotted onto a customised growth chart with documentation of measurements in the notes (from a historical control group) for nulliparous women, finding that plotting on a chart increased the rate of antenatal detection of SGA from 24.8% to 50.6% (P < 0.001).[226] Unfortunately, the observational or non-randomised nature of these studies limits the reliability of their results, as does the lack of comparison to plotting measurement on a population reference chart, however the use of customised fundal height charts is recommended by the RCOG on the basis of this observational evidence.[1]

Where the trajectory of the fundal height measurements differs from that expected, the woman is referred for a fetal growth ultrasound scan.[1] Using McDonald's rule, Cnattingius

and colleagues (1986) reported four types of fetal weight trajectory during pregnancy: normal (no measurement more than 2cm below the mean with a steadily increasing curve), static (no increase in the last 3 measurements), catch-up (at least one measurement was 3cm or more below the mean, but the last measurement was closer to the mean) and low (last measurement was 3cm or more below the mean). They identified that the low fetal weight trajectory pattern was associated with lower Apgar scores (less than 7) at 1 minute and more long-term stays in the neonatal intensive care unit than the other three patterns, although the latter was also associated with more prematurity.[29]

1.4.1.3 *Ultrasound assessment of fetal size*

In the UK, women assessed as being at high risk of SGA are recommended to have routine ultrasound monitoring of fetal growth. Women at low risk of SGA are recommended to have an ultrasound only where there are clinical concerns about the size or wellbeing of the fetus, or where their risk status changes because of antenatal complications.[1,8,217] In a review of guidelines on screening for SGA from six high income countries, McCowan et al (2018) identified that five of the six guidelines agreed that there was insufficient evidence to support routine third trimester fetal ultrasound for all pregnant women, only the French guideline recommended a single ultrasound at 32 weeks' gestation.[45]

Serial fetal growth surveillance by ultrasound for all (unselected) women has been associated with an 80-90% sensitivity for detection of SGA under trial conditions, but has not been associated with the same sensitivity in routine care (32% in a German study and 21.7% in a French study).[211,212] Universal ultrasound screening for SGA also has poor specificity, with approximately half of SGA antenatal diagnoses in a national French study actually being false positives.[211] The Hadlock A method of estimating fetal weight consistently has the lowest margin of error.[21,227]

For women at high risk of SGA, a Cochrane systematic review reported that EFW<10th centile on routine screening had sensitivity ranging between 40-82% and specificity ranging between 45-100% when predicting SGA at birth.[228]

For women at low risk of SGA, a meta-analysis including 21 studies (2019) found that a single screening ultrasound for fetal growth restriction after 32 weeks' gestation with EFW<10th centile has a sensitivity of 38% (95% CI: 31-46%) and specificity of 95% (95% CI: 93-97%) for predicting SGA at birth. AC<10th centile performed similarly. A significant increase in sensitivity was noted when the ultrasound scan was performed later in pregnancy.[229] Universal ultrasound for low-risk nulliparous women was shown to be of only borderline cost-effectiveness following a systematic review with economic modelling.[230]

For serial fetal ultrasound in low-risk or unselected women, a Cochrane systematic review (2008) assessing trials which examined the value of routine late pregnancy (after 24 weeks' gestation) serial ultrasound for a low-risk or unselected population, included eight trials recruiting 27,024 women found that there was no difference in antenatal, obstetric or neonatal morbidity, or intervention between the screened and control groups.[231] The review was limited by heterogeneity in the definition of a positive test. Routine late pregnancy ultrasound was not associated with a difference in perinatal mortality, although the review may have been underpowered to detect a difference in this outcome.[230]

1.4.1.4 *Biochemical predictive markers*

Biochemical markers have been identified as predictive of SGA, FGR or stillbirth. In a Cochrane meta-analysis of biochemical markers used to screen for SGA in mixed populations, human placental lactogen performed better (sensitivity 38%, 95% CI: 23-55% and specificity 88%, 95% CI: 78-94%) than serum or urinary oestriol however meta-analysis estimates could not be obtained for placental growth factor (PlGF) or uric acid due to heterogenous thresholds.[228] None of the tests performed as well as ultrasound assessment of fetal growth. In a different meta-analysis low human chorionic gonadotrophin was also associated with an increased risk of having an SGA fetus, but other biochemical markers used in early pregnancy to screen for fetal trisomies (alpha-fetoprotein, inhibin A, unconjugated estriol) had low predictive accuracy for an SGA fetus.[232,233]

There were insufficient data in the above-mentioned meta-analyses to test the predictive ability of PlGF or pregnancy-associated plasma protein A (PAPP-A). However, a cohort study of 213 neonates, found that PlGF<5th centile for gestational age was highly predictive of placental-mediated FGR with an AUROC of 0.96 (95% CI 0.93-0.98), sensitivity of 98.2% (95% CI 90.5-99.9) and specificity of 75.1% (95% CI 67.6-81.7).[234] Low PAPP-A (<0.415MoM) is also strongly associated with SGA (odds ratio, OR:2.7) and stillbirth, with a high negative predictive value (>90%).[233,235,236]

1.4.1.5 *Uterine artery Doppler*

Abnormal flow velocity ratio in the uterine arteries, measured by Doppler between 20-24 weeks' gestation (Figure 1.8), is associated with inadequate trophoblast invasion of the spiral arteries.[237]

Figure 1.8 – Normal transabdominal Doppler ultrasound examination of uterine artery.
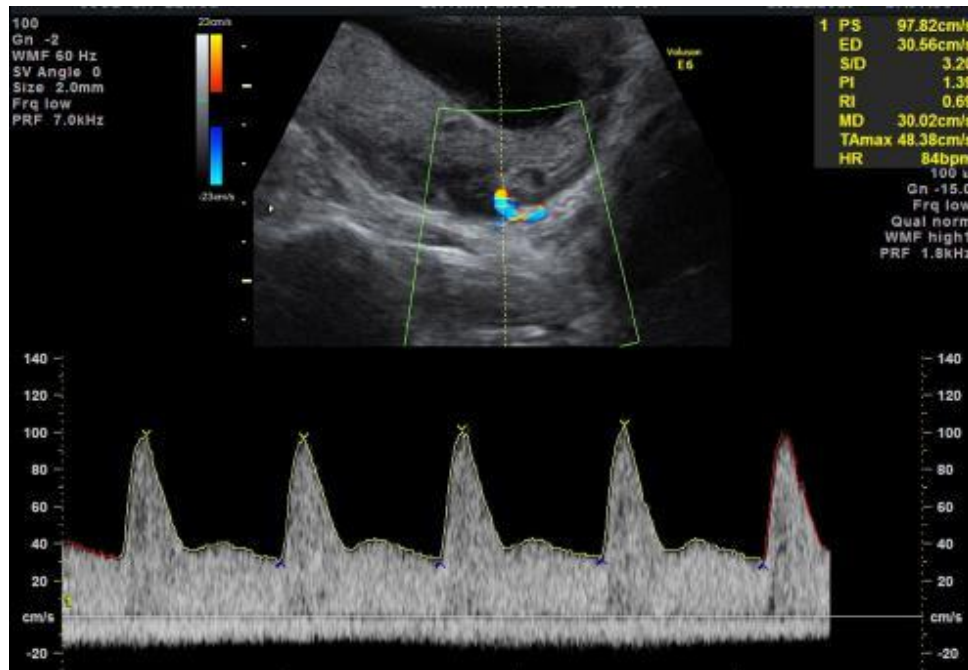
A meta-analysis examining the ability of uterine artery Doppler PI to predict SGA found that it performed best in a low risk population (pooled positive likelihood ratio, LR of 3.4 (95% CI: 1.7–5.1)), and more so when used to predict severe SGA (pooled positive LR: 3.7 (95% CI:10.3–16.9)).[239] In contrast, the pooled positive LR for women deemed at high risk of SGA was 2.3 (95% CI: 1.0–3.6) for SGA and 1.2 (95% CI: 0.5–3.2) for severe SGA. For this reason, uterine artery Dopplers are most useful for screening a low-risk population and should not be used to change monitoring frequency in women already identified to be high-risk of SGA following stratification using risk factors.

A randomised controlled trial compared routine uterine artery Doppler at the second trimester fetal anomaly scan to no Doppler in an unselected population, with follow-up for women with uterine artery mean PI>90th. The authors reported that 60% of cases of early-onset FGR had abnormal Doppler and that the rate of antenatal corticosteroid administration and induction of labour was increased in the women allocated to Doppler screening, but that there were no differences in the rates of perinatal or maternal complications between the two groups.[240]

1.4.1.6   *Other methods*

A systematic review examining the usefulness of ultrasound in predicting adverse neonatal outcomes amongst nulliparous women found that measurement of umbilical

artery Doppler flow, cerebroplacental ratio (ratio of umbilical artery to middle cerebral artery Doppler flow), or identification of oligohydramnios were only weakly predictive of the risk of giving birth to an SGA infant.[230]

Placental grading was studied in a Cochrane meta-analysis and was found to be poorly sensitive (38%, 95% CI: 23-55%) and moderately specific (79%, 95% CI: 62-90%) when used to screen for SGA, but slightly better if used to screen for risk of stillbirth (35% sensitivity and 94% specificity).[228]

1.4.2   Diagnosing large-for-gestational-age

Routine screening for the LGA fetus is not currently recommended in UK practice, despite LGA being associated with significant maternal and perinatal morbidity. This is partly because of the low predictive ability of fetal ultrasound scan when screening for macrosomia, and the historical uncertainty on which interventions should be offered to women who are diagnosed with an LGA fetus.[205] Whilst a Cochrane systematic review published in 2015 (dominated by one multi-centre randomised control trial)[241] showed that induction of labour indicated for suspected LGA or fetal macrosomia did reduce the risk of shoulder dystocia and fetal fractures without affecting mode of birth,[9] national guidance recommending induction does not currently exist in the UK.

1.4.2.1   *Ultrasound estimation of fetal size*

In the UK, the NICE guideline 'Antenatal Care for Uncomplicated Pregnancies' recommends that ultrasound estimation of fetal weight should not be performed for suspected LGA in the absence of another indication in low risk women.[8] Estimation of fetal weight through offer of routine ultrasound in a low risk population is poorly correlated with birthweight and LGA, although the negative predictive value is high.[242,243] A systematic review examining the diagnostic effectiveness of offering universal screening by ultrasound to detect macrosomia and predict adverse neonatal outcome found that an EFW >4000g or above the 90th centile had only 22% sensitivity to predict shoulder dystocia.[230]

1.4.2.2   *Other methods*

Other methods to predict LGA or macrosomia include ultrasound measurement of fetal adiposity (fetal thigh and abdominal wall)[244] or fetal thigh volume,[245] and measurement of EFW using magnetic resonance imaging (MRI).[246] In a single cohort study, fetal adiposity measurements above the 90th centile were more predictive of unplanned caesarean birth than EFW>90th centile. MRI-derived EFW performed significantly better than ultrasound-derived EFW for the prediction of LGA (AUROC 0.985 vs 0.900, P < 0.001 within 48 hours of

birth and 0.957 vs. 0.880 P < 0.001 for births >48h after investigation), although both measures performed relatively well.[247] The current cost of an MRI is approximately 3-4 times that of an ultrasound and the technique for fetal adiposity measurement has not been standardised nor is it routinely taught, so neither are currently adopted in routine clinical practice.

1.5 ANTENATAL MANAGEMENT OF FETAL GROWTH ANOMALIES

Following recognition that a woman is at risk of having a baby with a fetal growth anomaly, or following diagnosis of a growth anomaly, clinical pathways are required which aim to reduce the potential for associated short- and long-term morbidity or mortality. In this section, I will briefly summarise current strategies for prevention, surveillance, and management of fetal growth anomalies.

1.5.1 Preventing fetal growth restriction

Aspirin is currently recommended in UK practice for prevention of pre-eclampsia in women with one major risk factor or more than one moderate risk factor.[248] Many of the risk factors for pre-eclampsia are also common to SGA.[1] In a Cochrane systematic review, aspirin was reported to reduce the risk of having an SGA baby by 10% when given to women at high-risk of developing pre-eclampsia.[249]

Smoking is strongly associated with an increased risk of SGA and a Cochrane systematic review of smoking cessation in pregnancy found that it reduced the risk of both low birthweight and preterm birth, although SGA was not studied.[250]

Preconception commencement of folic acid has also been associated with a reduced rate of SGA in meta-analysis (adjusted, aOR 0.75, 95% CI: 0.61, 0.92), although antenatal commencement had no effect (aOR 0.82, 95% CI: 0.63, 1.06).[251] The FACT trial (a randomised, multicentre, controlled trial assessing the effect of high-dose folic acid supplementation of pre-eclampsia, published since the meta-analysis) has confirmed the latter finding. Women with at least one risk factor for pre-eclampsia (most of which are also risk factors for SGA) were randomised between 8 to 16 weeks' gestation to either high dose folic acid or placebo. There was no difference in the rate of stillbirth (RR 0.60, 95% CI 0.30-1.19, p=0.14) or SGA (RR 1.03, 95% CI 0.81-1.30, p=0.82) between trial arms. There is no good evidence that other dietary changes, progesterone or calcium supplementation prevent SGA.[252-254]

1.5.2 Antenatal management of the small for gestational age fetus

Timely delivery of an SGA fetus, usually guided by gestation-specific changes in fetal vascular flow, has been shown to reduce the risk of serious perinatal morbidity and mortality.[4] In the UK, the RCOG Green-top guideline 'Small for Gestational Age Fetus: Investigation and Management' is used to guide the management of the SGA fetus (Figure 1.9).[1] Briefly, this recommends:

- Primary surveillance by assessment of umbilical artery dopplers. These can be repeated fortnightly if they remain normal and SGA is not severe.
- When umbilical artery dopplers are abnormal, monitoring is dependent upon the presence of EDF (twice weekly where present, daily where absent or reversed)
- In the term SGA fetus, a PI in the middle cerebral artery (MCA) of <5th centile has moderate predictive value for fetal acidaemia and should be used to guide timing of birth.
- Ductus venosus Dopplers should be used to monitor the preterm SGA fetus with abnormal umbilical artery Dopplers.
- In an SGA fetus with absent or reversed EDF in the umbilical artery prior to 32 weeks' gestation, timing of birth should be guided by abnormalities in the ductus venosus Doppler pattern (provided the fetus is considered viable and after administration of corticosteroids).
- After 32 weeks' gestation, RCOG guideline does not specify how to make decisions regarding birth, except to say that the umbilical artery Doppler should be used, and that birth should be arranged by 37 weeks' gestation.

No antenatal interventions have been found to be effective at improving the intrauterine growth trajectory of fetuses with impaired growth during ongoing pregnancy. Randomised control trials comparing sildenafil or 100 mg aspirin to placebo given to women with pregnancies affected by FGR of varying severity found no difference in the birthweight or rate of perinatal mortality between the two groups.[255,256] A meta-analysis identified only two studies that investigated antenatal heparin for management of presumed placental-mediated FGR; heparin resulted in a significantly higher birthweight (mean difference: 365; 95% CI: 236 to 494; $P < 0.001$) but did not affect rates of low Apgar scores, neonatal admission, neonatal mortality, or adverse composite of neonatal morbidity.[257]

Figure 1.9 - Algorithm for the management of the small for gestational age fetus, reproduced from the RCOG Green-top guideline no. 31: The Investigation and Management of the Small–for–Gestational–Age Fetus
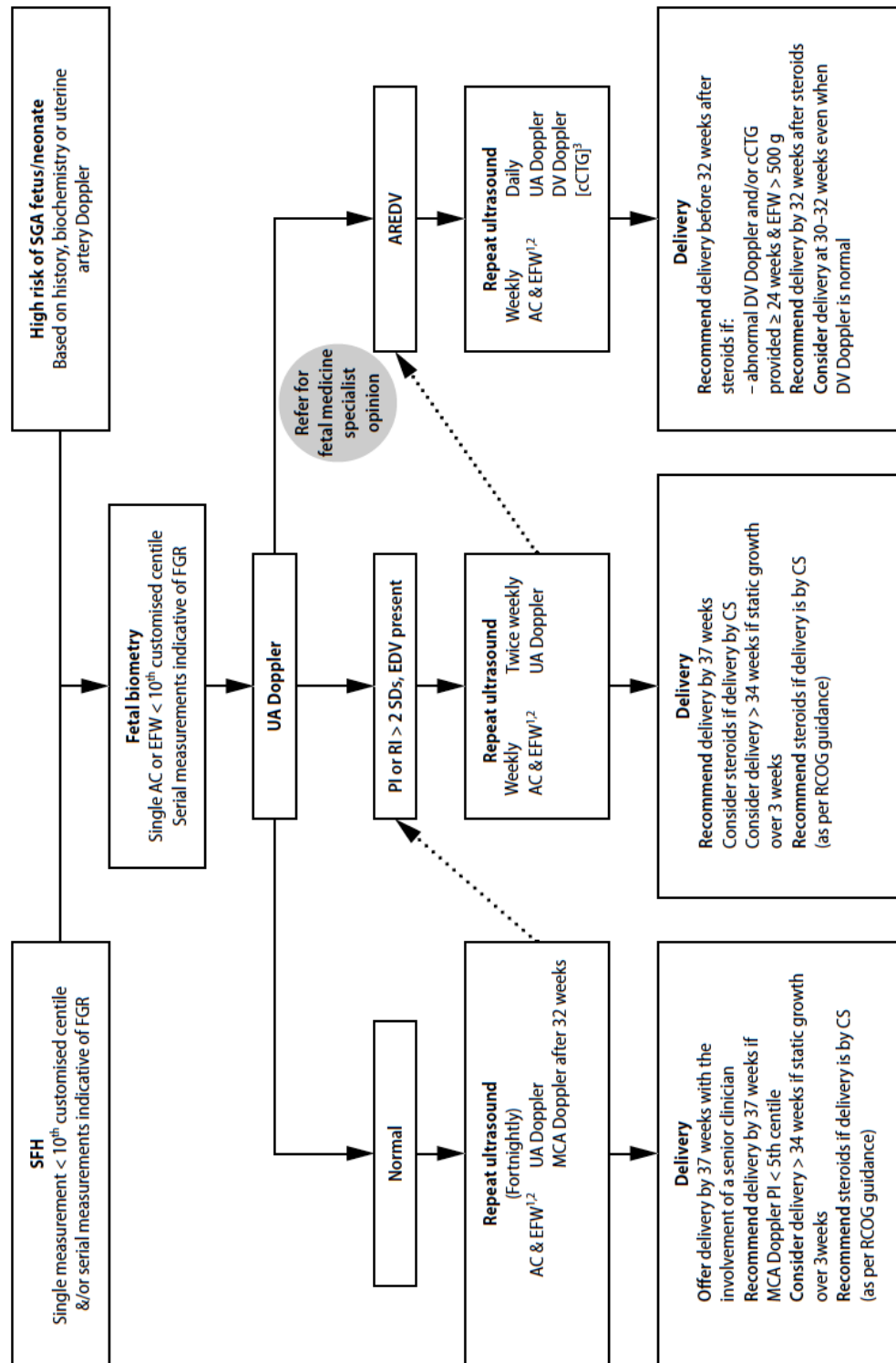
1 *Weekly measurement of fetal size is valuable in predicting birthweight and determining size-for-gestational age*

2 *If two AC/EFW measurements are used to estimate growth, they should be at least 3 weeks apart*

3 *Use cCTG when DV Doppler is unavailable, or results are inconsistent – recommend delivery if STV < 3 ms*

Abbreviations: AC, abdominal circumference; EFW, estimated fetal weight; PI, pulsatility index; RI, resistance index; UA, umbilical artery; MCA, middle cerebral artery; DV ducts venosus; SD, standard deviation; AREDV., Absent/reversed end–diastolic velocities; cCTG, computerised cardiotocography; STV, short term variation; SFH, symphysis–fundal height; FGR, fetal growth restriction; EDV, end–diastolic velocities.

1.5.3 Preventing large for gestational age fetal growth

The prevention of excessive fetal growth starts pre-conception, with lifestyle interventions targeted at reducing maternal BMI to within the healthy range (18.5-24.9 kg/m²) and treatment of metabolic dysfunction that predisposes to having an LGA baby. For women with a BMI of 35 kg/m² or above, pre-conception bariatric surgery has also been shown to reduce the risk of having an LGA baby although conversely, bariatric surgery increases the risk of SGA.[258] Meta-analyses have shown that antenatal exercise is effective at reducing the risk of macrosomia or LGA in the general pregnant population (women of any BMI),[259,260] however antenatal dietary interventions are only effective in reducing the risk of macrosomia for overweight women, or those with diabetes.[261] For diabetic women, control of hyperglycaemia either through diet or medication (e.g. metformin or insulin) also reduces the risk of excessive fetal growth.[262,263]

1.5.4 Antenatal management of the large for gestational age fetus

Most maternal and neonatal morbidity associated with having an LGA or macrosomic fetus occurs through intrapartum events (section 1.3.2), it therefore follows that earlier birth of the neonate, with a lower birthweight, is expected to reduce size-associated morbidity. Yet obstetric interventions and early term birth are not without risk and UK NICE guidance does not currently advise iatrogenic earlier birth for suspected LGA in non-diabetic mothers.[8,264] RCOG guidelines published since the NICE guideline do however recommend inducing labour for suspected macrosomia in mothers with diabetes,[264] offering elective caesarean birth to reduce the risk of shoulder dystocia for women in whom the neonatal birthweight is expected to be greater than 4,500g,[264] and considering induction of labour for women with obesity who are suspected to have an LGA baby.[93]

A Cochrane review (2016) assessed a strategy of inducing labour at or near term (defined as 37+0 to 40+0 weeks') for suspected fetal macrosomia.[9] Suspected macrosomia was defined differently in each of the four included trials, including estimated birthweight of 4,000-4,750g, EFW >95th or EFW >97th centile. Overall, a strategy of induction did not affect the rate of caesarean or assisted vaginal birth, neonatal brachial plexus injury or measures of poor neonatal condition at birth but did decrease the rate of shoulder dystocia and neonatal fractures. The review was dominated by a single multi-centre randomised control trial,[241] limited by poor quality of evidence on mode of birth and was underpowered to detect a difference in rare outcomes, such as brachial plexus injury.

1.6 COMPLEX ANTENATAL INTERVENTIONS TO IMPROVE ANTENATAL DETECTION OF SGA AND REDUCE STILLBIRTH

Complex interventions are commonly used in health and social care. They are defined as such because they have several interacting components.[265] In the UK, there are a number of national strategies which exist to reduce the rate of stillbirth,[266,267] which is higher than that of many other Western European countries.[175] Two of these initiatives are complex interventions: NHS-England's Saving Babies' Lives stillbirth care bundle,[217] and the Growth Assessment Protocol (GAP).[74]

1.6.1 The NHS-England Saving Babies' Lives care bundle

The NHS-England's Saving Babies' Lives care bundle, aims to support maternity care providers, commissioners and healthcare professionals to reduce perinatal mortality through implementation of 'elements' targeting: smoking cessation in pregnancy, antenatal surveillance for SGA, care of women with reduced fetal movements in pregnancy and intrapartum fetal monitoring.[217] The care bundle was revised and a fifth element added in 2019 to reduce perinatal morbidity and mortality from preterm birth.[268]

The antenatal surveillance for SGA element of the 2016 version of the Saving Babies' Lives care bundle (applicable at the time of the DESiGN trial), required maternity units to be compliant with five components:

- Use either the NHS-England (Figure 1.6) or RCOG algorithm,[1] to categorise women and arrange surveillance for SGA, which differs according to whether they are at low- or high-risk.
- For women at high risk of FGR, monitor fetal growth using serial fetal ultrasound. Record the estimated fetal weight on a customised or population-based growth chart.
- For women at low risk of FGR, assess fetal growth using fundal height measurement. All staff conducting these measurements must be trained in the measurement, plotting on centile charts, interpretation, and protocols for onward referral.
- Contemporaneous audit, reporting and publishing of the rates of SGA at birth, antenatal detection of SGA, false positives and false negatives.
- Investigation through case note review of SGA cases which were not antenatally detected, to identify learning and improve future rate of detection.

In a prospective observational evaluation of 19 NHS Trusts that each implemented the Saving Babies' Lives care bundle in April 2015, researchers identified that the stillbirth rate in the recruited Trusts decreased by 20% between April 2013 and April 2017, although this could not be attributed only to the implementation of the care bundle. Screening for SGA infants according to the care bundle protocols increased the rate of antenatal detection of SGA from 33.8 to 53.7%, there was also an increase in the number of ultrasound scans, the percentage of women having an induction of labour or caesarean birth, and the percentage of babies born prematurely or admitted to a neonatal unit.[269]

All components of the fetal growth surveillance element of the Saving Babies' Lives care bundle can be achieved through complete implementation of the GAP care package. There are no alternative complete packages available in the UK and so, maternity units who do not implement GAP need to make bespoke plans to ensure compliance with this element.

1.6.2 The Growth Assessment Protocol

GAP is a complex antenatal intervention developed by Professor Gardosi and colleagues. It was first described in published literature in 2013 by Clifford et al,[74] following the extensive work on customised antenatal growth charts conducted by Professor Gardosi (section 1.1.7.1). National dissemination of the GAP program has since been conducted by Professor Gardosi and his team, working under the umbrella of the Perinatal Institute (www.perinatal.org.uk), a not-for-profit organisation which aims to 'further enhancements in the quality and safety of maternity care'.[73] By June 2021, the GAP programme had been implemented in 78% (n=121) of UK NHS Trusts/Boards.[270]

The components of GAP are summarised in Table 1.6 and, using the TIDieR guidance (Template for Intervention Description and Replication) in Appendix section 10.1. Detail of each component follows below. At all stages of GAP implementation support and regular communication are available from the Perinatal Institute.

Table 1.6 - Expected components of GAP implementation

Implementation Stage	GAP requirements
Preparation and planning	• Nominated staff from each NHS Trust/Board to attend 'Train the Trainers' GAP workshop. • NHS Trust/Board to conduct a baseline audit of SGA detection (10% of annual births). • NHS Trust/Board to prepare local guideline for the 'Assessment of Fetal Growth' modelled on GAP recommendations.
Implementation	• NHS Trust/Board trainers to cascade face-to-face training to at least 75% of colleagues from each professional group (midwives, obstetricians, sonographers). • GAP e-learning module to also be completed by at least 75% staff members from each professional group.
Ongoing use of GAP	• Access to GROW chart online program provided by the Perinatal Institute after NHS Trust/Board compliant with above requirements. • Each pregnant woman assessed for risk of SGA at antenatal booking appointment using GAP tool. • Customised GROW chart printed for each pregnant woman at antenatal booking appointment and used to assess fetal growth by plotting fundal height measurements or estimated fetal weight on the chart. • Women at low risk of SGA expected to have a fundal height measured 2-3-weekly during pregnancy, commencing between 26 and 28 weeks. If plots deviate from that which is expected (first plot below 10th centile, slow/static/accelerative growth), the woman should be referred for a fetal growth scan. • Women at high risk of SGA expected to have an ultrasound scan to estimate fetal weight 3-weekly during pregnancy, commencing between 26 and 28 weeks. • Where GROW chart EFW plots deviate from the expected trajectory (as per fundal height deviations), RCOG protocols should be followed for further investigation of suspected SGA.[271]
Evaluation	• Regular audit of SGA detection rates • A tool to support investigation of 'missed cases' of SGA (undetected SGA)

1.6.2.1 *Comprehensive staff training*

GAP training is conducted through three different methods:

- Train the trainers workshops,
- Trainers cascading face-to-face training to all maternity staff,
- An e-learning module.[272]

Train the trainers:

The Perinatal Institute holds a rolling programme of 'train the trainers' workshops which should be attended by nominated staff members from each implementing NHS Trust/Board. The workshops are one day in length and can be arranged either at the Perinatal Institute in Birmingham (cost of £500 per NHS Trust), or locally (£500 plus expenses). I had the opportunity to attend a training day at the Perinatal Institute during the first year of my PhD study. Nominated staff members are those who are required to

cascade training to their colleagues following the workshop, prior to implementation of the intervention locally. The 'train the trainers' workshop covers the following topics:

- Pregnancy risk assessment for fetal growth restriction, including care pathways,
- Fetal growth surveillance, including summary of national guidelines,
- Clinical and practical application of customised growth charts,
- Standardised technique for fundal height measurement,
- The efficacy of the Growth Assessment Protocol and guide to local implementation,
- Explanation of routine data collection procedures (uploading data on growth surveillance to the Perinatal Institute),
- Guide to conduct of missed cases audit.

Cascade of face-to-face training:

Following attendance at the 'train the trainers' workshop, trainers are then expected to cascade face-to-face training to obstetric, midwifery and sonography colleagues practising within antenatal or intrapartum care. A coursebook and slide pack is provided to the trainers to assist in the cascade of training. It is expected that the following topics are covered:

- Medical, social and obstetric risk factors for fetal growth restriction and perinatal mortality,
- Principles and supporting evidence for the use of customised fetal growth charts,
- Hands-on demonstration of the standardised technique for measuring the symphysis-fundal height,
- Demonstration of the method to produce GROW charts and calculate the neonatal birthweight centile,
- Hands-on demonstration of the plotting of fundal heights or ultrasound-derived EFW onto the GROW chart, including assessment of competency with a test paper,
- Criteria for referral when a plot on the GROW chart deviates from the expected trajectory.

E-learning:

All obstetric, midwifery and sonography professionals are also expected to complete the GAP e-learning modules, which should then be repeated annually. There are two modules with corresponding tests, on which healthcare professionals are required to score 100% before receiving a certificate of completion. The modules are on:

- Theory of GAP and GROW – six chapters containing information on the evidence and background to GROW,
- Clinical practice – six chapters containing information on the implementation of GROW within clinical practice.

Training compliance:

Both trainer and trainee members of staff are required to update their face-to-face and e-learning training annually. Implementing sites are required to keep a training log and report this to the Perinatal Institute who stipulate in their service level agreement with NHS Trusts (Nov 2019) that *"75% of clinical staff engaged in maternity care are trained and accredited in GAP (minimum of face-to-face with test for year 1 and E-learning and competency assessment year 2 onwards)".[273]*

1.6.2.2 *GROW chart software*

The GROW chart software is accessed through a password-protected website. A login and password are issued to NHS Trusts/Boards, following compliance with the pre-implementation targets and payment of a fee to the Perinatal Institute (varies between £1,500-£5,000, dependent on number of annual births at the NHS Trust).

The GROW chart software has the following functions:

1. Generation of a GROW chart for each pregnant woman following entry of all required data items (see Figure 1.10 for a screenshot of the data entry form),[274]
2. Calculation of the birthweight customised centile for neonates who have been exposed to the intervention antenatally,
3. Access to software-generated reports on local compliance which report information for each financial quarter.

Figure 1.10 - GROW software - data entry form

1.6.2.3 *Guidelines for clinical practice*

The Perinatal Institute recommends that implementing trusts adopt the NHS England algorithm and risk assessment tool for screening and surveillance of fetal growth in singleton pregnancies (Figure 1.6).[217] Women who book for antenatal care at an implementing site should be risk assessed at the first appointment, following which, the women should be streamlined into low- or high-risk care.

The GAP programme recommends that serial standardised measurements of fundal height (low-risk women) and EFW measurements from ultrasound (high-risk women) should be plotted onto the GROW charts (section 1.1.7.1), starting from 26-28 weeks' gestational age. Fundal height measurements should be plotted using an 'X' symbol and EFW should be plotted using a small circle. For low-risk women, the fundal height should be measured and plotted 2-3 weekly throughout pregnancy. For high-risk women, fetal biometry should be measured at ultrasound and serial EFW calculated and plotted 3-weekly throughout pregnancy. Measurements of both should continue until birth. Serial fundal height measurements with a normal growth trajectory are illustrated in Figure 1.11. This could equally demonstrate serial EFW derived from ultrasound, but the 'X' symbol would be replaced with a circle.

Figure 1.11 – Example GROW chart with plots of fundal height growth progressing along a normally-expected trajectory.[275]
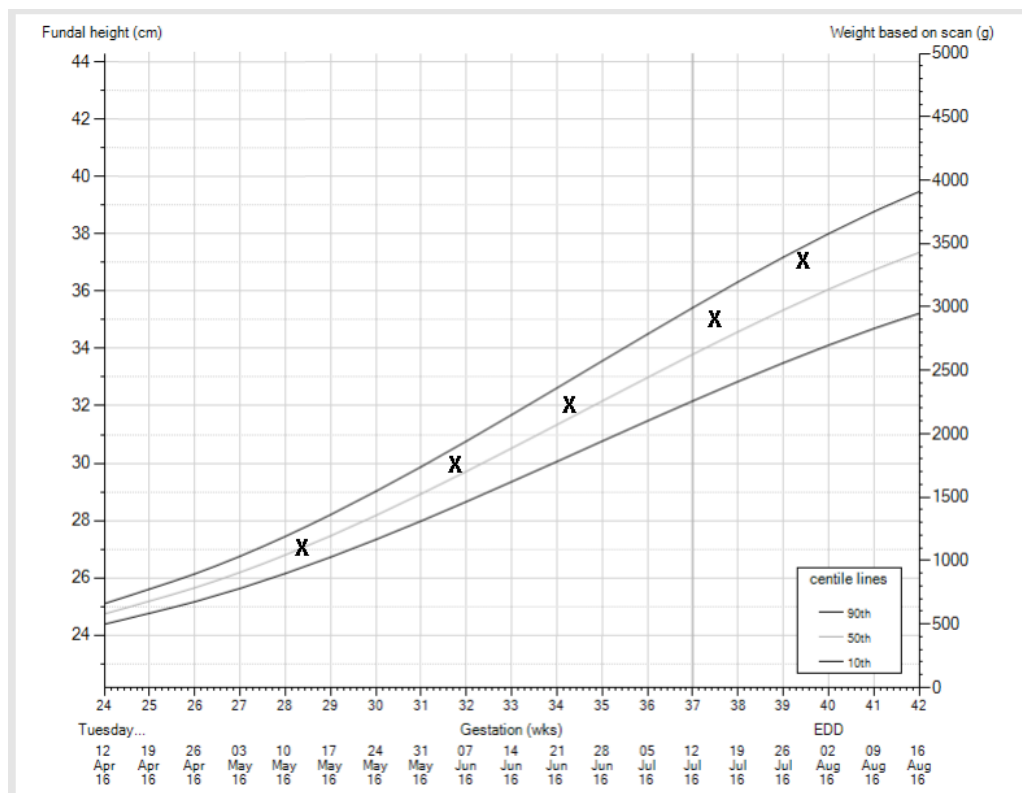
A deviation of the fundal height from the normal GROW chart trajectory in one of the four following ways, should prompt referral for an ultrasound growth of the fetus:

1. First plot below the 10th centile on the GROW chart (Figure 1.12a)
2. A serial plot which demonstrates a slowing of growth (Figure 1.12b)
3. A serial plot which demonstrates static growth (Figure 1.12c)
4. A serial plot which demonstrates accelerative growth (Figure 1.12d)

It is important to note that a first or subsequent plot above the 90th centile on the GROW chart should not prompt an ultrasound for growth, in the absence of the above criteria.

Figure 1.12 – Deviations of symphysis-fundal height on the GROW chart: a) First plot below the 10th centile, b) Slow growth, c) Static growth, and d) Accelerative growth.
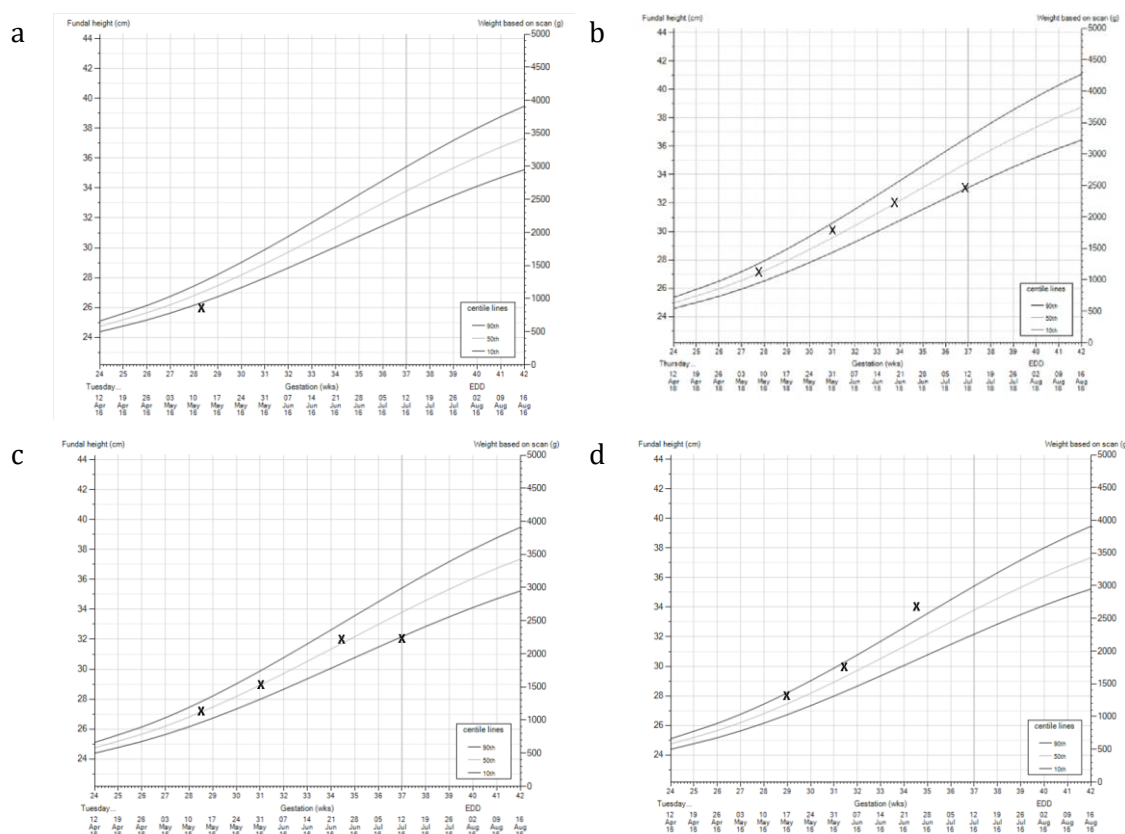
Deviations in the EFW plotted on the GROW chart, with the same patterns as per the deviation of fundal height (Figure 1.12a–d), should prompt management of suspected SGA as per the recommendations of the RCOG Green-top guideline on Investigation and Management of the Small-for-Gestational-Age Fetus (section 1.5.2).[1]

1.6.2.4 *Regular audit of SGA detection rates*

The GROW software has an integrated function to generate quarterly reports of performance statistics for each implementing site. The usefulness of these statistics depends on the quality of data entry. The statistics are:

- Number of charts generated (including percentage of expected),
- Percentage of generated charts where birthweight centiles were calculated using the software,
- Percentage of babies born who were SGA by customised standards,
- Percentage of SGA babies who were referred antenatally for suspected SGA,
- Percentage of SGA babies who were diagnosed as SGA antenatally,
- Comparison to the national GAP user average for all the above.

These reports allow NHS Trusts/Boards to monitor their compliance and performance, with the intention that this prompts quality improvement.

1.6.2.5 *A tool to support investigation of missed cases*

The Perinatal Institute recommends that a team of clinicians audit a randomly chosen selection of missed SGA cases every six months. Following completion of the case summary, the GAP-SCORE (Standardised Case Outcome Review and Evaluation) software summarises the learning points and prompts the reviewing team to make recommendations to improve care quality.

1.6.2.6 *Support and communication*

It is recommended that each trust nominates obstetric, midwifery and sonography GAP leads. These clinicians provide local leadership for the implementation of GAP and liaise with the Perinatal Institute regarding any GAP queries that cannot be solved internally.

1.6.2.7 *Clinical effectiveness of the Growth Assessment Protocol*

The DESiGN trial was the first randomised control trial (RCT) that compared the implementation of GAP to standard care.[276] DESiGN was conceived in 2014, at which time GAP had been implemented in less than 5% of London NHS Trusts but had already been implemented in much of the rest of England and Wales. National uptake of implementation had followed publication of multiple observational studies, all conducted by the team at the Perinatal Institute, and finding GAP to be associated with an increase in the rate of SGA detection and a reduction in stillbirth. Since GAP was already widespread in the rest of England and Wales, and higher quality research (in the form of an RCT) was needed, a London-based trial was seen as the last UK opportunity to achieve this.[276] At the time of writing, no other RCTs comparing GAP to an alternative strategy have been published. As the research reported in this thesis was primarily conducted as part of the DESiGN trial, the trial methods, progress, and findings have been detailed in Chapter 2 (Methods). Observational studies of the clinical effectiveness of GAP are detailed below.

A study in the UK comparing regions with 'high uptake' of GAP to regions with 'low uptake' during the period 2008 to 2012 used stillbirth data from the Office for National Statistics and found that there was a 22% lower stillbirth rate in the high uptake regions during the period analysed, compared to static rates in the low uptake areas. [277] Notably, the low uptake regions had lower rates initially (4.86 stillbirths/1000 births vs. 5.63/1000 births in the high uptake regions) and ecological comparisons such as these may reflect any number of changes during the time period.[225] The MBRRACE-UK Perinatal Mortality

Surveillance Report (2013) found a gradual decline in stillbirth rates across the UK throughout the decade 2003-13 and attributed this to Sands (Stillbirth and Neonatal Death) Charity in raising awareness of initiatives designed to reduce stillbirth rates, and/or to new guidance from the RCOG in redefining which stillborn babies require registration (i.e. not those known to have died in utero prior to the end of the 24th gestational week).[278]

Conversely, a population-based study compared the decline in the stillbirth rate in England and Wales where GAP uptake was high, to Scotland where GAP uptake was low. The authors found that the decline in stillbirths from the rate in January 2000-December 2009 to the rate in January 2010-December 2015 was greater in Scotland by 48 stillbirths per 100,000 births. They concluded that the stillbirth decline in England could therefore not be used to infer that GAP is efficacious.[279] Gardosi et al refuted this by demonstrating that the decline in the Scottish stillbirth rate was more similar to that in England when the years 2016-2018 are included in the analysis, and that whilst GAP was not implemented in Scotland pre-2015, there was a national patient safety initiative targeted towards management of pregnancies with fetal growth anomalies, standardised case note reviews being conducted in some units and a national RCT which increased the use of third trimester ultrasound and Doppler studies to investigate fetal wellbeing in pregnancies where women were concerned about fetal movements. Many of these initiatives overlap with the strategies provided with GAP.[280]

More recently (2020), Gardosi et al have published a further retrospective observational study comparing the fall in stillbirth rates between 2008 and 2017 in all English NHS Trusts, between those who did not, partially or completely implemented GAP.[281] Complete implementation was not clearly defined in the publication, but appeared to have been determined by individual Trust reporting of the birthweight and SGA outcomes in >75% of expected births. A greater reduction in the stillbirth rate was seen in Trusts who completely implemented GAP, compared to those who did not implement. The strength of association was greater amongst the 20 Trusts with strongest implementation (highest reporting rates). Critically, it is possible that there may have been other factors which have contributed to the fall in stillbirth in the 'complete implementer' Trusts, since it is quite likely that Trusts who have complied well with all components of GAP, are also likely to have complied well with components of other interventions, such as the Saving Babies' Lives care bundle,[217] intended to reduce that rate of stillbirth.

Elsewhere, Jayawardena et al (2019) compared the rate of SGA detection before and after implementation of GAP in Melbourne, Australia.[282] Fetal growth in the pre-implementation group was assessed without a growth chart, and SGA was defined for both

groups using customised birthweight centiles. This is a major limitation, given that fetal growth was not assessed using customised centiles in the pre-implementation group. The authors found that implementation of GAP was associated with an increase in the rate of SGA detection from 21% to 41% without a concomitant increase in false positive diagnoses. There was no difference in the rate of stillbirth, but the study was underpowered to assess this.

In New Zealand, Cowan et al (2020) studied the rate of SGA detection before and after implementation of GAP in Auckland midwifery-practices and found that the rate of SGA detection increased from 22.9% in 2012 to 57.9% in 2017/18.[283] Aside from its observational nature, this study was also limited by the absence of any SGA guideline before 2013, being limited only to women receiving low-risk care, and a change in the definition of SGA between the time periods (from population to customised centiles) which was not accounted for in the analysis.

1.6.2.8 *Cost-effectiveness of the Growth Assessment Protocol*

Given the low rates of antenatal detection of SGA in current practice, it is pertinent to ensure that implementation of any strategy to improve detection are subject to both clinical and cost-effectiveness studies. Cost-effectiveness studies are used in healthcare to ensure that scarce resources are allocated appropriately to ensure maximum health benefit amongst all potential and current service users.

In a retrospective cohort study to assess the cost benefit of serial fetal growth scans for women at high risk of SGA, the GAP team estimated that 25.5% of women giving birth (from a West Midlands cohort of 146,774 women) would require serial scans during pregnancy using the risk assessment as defined by the RCOG guideline on Detection and Management of the SGA Fetus. This was estimated to cost an additional £10 per pregnancy however, the cost calculation was based on a cost estimate of £15 per scan, which contravenes the national NHS reference cost at that time (2015-16) of £103.84 for a standard antenatal ultrasound scan.[284] This was compared to an estimated saving of £120 per pregnancy by reducing neonatal admissions, perinatal morbidity and mortality, cerebral palsy and litigation.[87]

1.7 THESIS AIMS AND HYPOTHESES

1.7.1 Thesis aim

To conduct a detailed evaluation of the implementation and cost-effectiveness of GAP in the context of the DESiGN trial, including an assessment of its impact on both large and small-for-gestational-age babies.

1.7.2 Hypothesis 1

The strength with which GAP is implemented is affected by pre-conceived beliefs about its effectiveness and by the resources available.

Aim of analysis 1 (Chapter 3): To determine barriers and facilitating factors associated with the strength of implementation of GAP.

1.7.3 Hypothesis 2

Implementation of GAP is associated with higher cost related to both the costs of implementation and increasing ultrasound use, but this leads to an improvement in the rate of detection of SGA (i.e., GAP is cost-effective).

Aim of analysis 2 (Chapter 4): Review the current evidence on costs incurred following implementation of maternity care interventions in the UK.

Aim of analysis 3 (Chapter 5): Evaluate the cost-effectiveness of the Growth Assessment Protocol in terms of its ability to improve antenatal detection of the SGA fetus.

1.7.4 Hypothesis 3

Pregnancies in which SGA occurs but is not diagnosed antenatally are less likely to have identifiable risk factors for SGA and therefore receive fewer scans, particularly at or near term.

Aim of analysis 4 (Chapter 6): Identify the clinical and ultrasound utilisation characteristics of pregnancies in which an antenatal diagnosis of SGA is missed, compared to those in which it is made, to understand how we can better target interventions to improve detection.

1.7.5 Hypothesis 4

GAP leads to an increase in the antenatal detection of LGA through an increase in ultrasound utilisation for large babies.

Aim of analysis 5 (Chapter 7): Assess for an unintended impact from implementing GAP on the detection and management of pregnancies in which the baby is LGA.

2 Methods

I have previously published sections of this chapter (full manuscripts are free to access through the links provided):

Vieira MC, **Relph S**, Copas A, Healey A, Coxon K, Alagna A, Briley A, Johnson M, Lawlor DA, Lees C, Marlow N, McCowan L, Page L, Peebles D, Shennan A, Thilaganathan B, Khalil A, Sandall J, Pasupathy D. DEtection of Small for Gestational age Neonate (SGA) – a cluster randomised controlled trial to evaluate the effect of the Growth assessment protocol (GAP) programme: The DESiGN Trial Protocol. Trials (March 2019);20(1):154 https://doi.org/10.1186/s13063-019-3242-6

Relph S, Elstad M, Coker B, Vieira MC, Moitt N, Muruet Gutierrez W, Khalil A, Sandall J, Copas A, Lawlor DA, Pasupathy D on behalf of the DESIGN Trial Team. Using electronic patient records to assess the effect of a complex antenatal intervention in a cluster randomised controlled trial - Data management experience from the DESiGN Trial team. Trials (2021); 22:195. https://doi.org/10.1186/s13063-021-05141-8

2.1 STUDY DESIGN AND POPULATION

The DESiGN trial (<u>DE</u>tection of the <u>S</u>mall for <u>G</u>estatio<u>N</u>al age fetus) is the only RCT conducted to date examining the clinical effectiveness of GAP. More specifically it was a UK-based hybrid effectiveness-implementation, pragmatic, randomised cluster control trial which aimed to examine the clinical-effectiveness, implementation and cost-effectiveness of the Growth Assessment Protocol (GAP),[74] comparing it to standard care, in antenatal screening and management of the SGA fetus.[276]

A hybrid-effectiveness implementation trial is one that plans a dual focus on both clinical-effectiveness and implementation. Curran et al (2012) proposed three types of such studies. The DESiGN trial best fits the description of a type 2 hybrid trial: rather than testing one of clinical effectiveness or implementation strategy while collecting data on the other (types 1 and 3 respectively), we planned to test both the clinical intervention and the implementation strategy simultaneously.[285]

DESiGN was planned as a pragmatic trial to capture the reality of implementing GAP in everyday clinical practice with standard support from the Perinatal Institute. A pragmatic trial is one that is undertaken in 'real world' conditions, with application of the intervention alongside usual care.[286] It is intended to help support decisions on whether to deliver interventions in practice. Through application of the PRECIS-2 tool (

Table **2.1**),[287] I have demonstrated that the DESiGN trial was a 'very pragmatic' trial including the reasons for forming this conclusion.

A cluster design was chosen in preference to the randomisation of individual women because of the nature of the intervention. GAP requires whole-site training and modification of clinical protocols that would likely cause contamination of interventions if women were individually randomised. All women giving birth to singleton, non-anomalous infants at the cluster sites during the study period were included.

The trial was not blinded because the nature of the intervention meant that it was not possible to conceal intervention allocation from women or staff in maternity clusters. Due to the odd number of clusters, small trial team, and statistician involvement at every point of the trial, it was also not possible to conceal the statistician from cluster allocation when performing the analysis.

Table 2.1 - Explanation of the extent to which the DESiGN trial was a pragmatic trial, by application of the PRECIS-2 tool

Domain	Score	Explanation
Eligibility: To what extent are the participants in the trial similar to those who would receive this intervention if it was part of usual care?	5 – very pragmatic	All pregnant women (i.e. all individuals to whom the intervention would be provided in usual care) were included in the trial, except for those with multiple pregnancies, fetal congenital anomalies or births prior 24 weeks, all of whom the intervention is less appliable to.
Recruitment: How much extra effort is made to recruit participants over and above what would be used in the usual care setting to engage with patients?	5 – very pragmatic	Recruitment occurred at the cluster level, although the intervention was measured at the individual level. This means that all individuals voluntarily attending the cluster for care were included.
Setting: How different are the settings of the trial from the usual care setting?	5 – very pragmatic	The trial clusters were recruited as the remaining (mostly London-based) maternity units that had not yet commenced implementation of the intervention but were considering it i.e., the setting was the same as that in which the intervention would be implemented if shown to be effective.
Organisation: How different are the resources, provider expertise, and the organisation of care delivery in the intervention arm of the trial from those available in usual care?	5 – very pragmatic	No additional resources or expertise were supplied to the recruited clusters.
Flexibility (delivery): How different is the flexibility in how the intervention is delivered and the flexibility anticipated in usual care?	4 – rather pragmatic	The intervention provider slightly changed the intervention for the purposes of the trial (adjusted requirements to consider clusters training compliant, see section 1.6.2.1 and 2.1.1). No other elements were different to how the intervention would be delivered outside of a trial setting.
Flexibility (adherence): How different is the flexibility in how participants are monitored and encouraged to adhere to the intervention from the flexibility anticipated in usual care?	4 – rather pragmatic	No additional monitoring was recommended for individual women or individual staff members. The intervention provider did provide additional implementation support during the pre-implementation phase, in the form of monthly group meetings with the cluster clinical leads. These stopped before clusters were 'going live' with GAP.
Follow-up: How different is the intensity of measurement and follow-up of participants in the trial from the typical follow-up in usual care?	5 – very pragmatic	There were no differences between the trial and intervention follow-up protocols. Data for the trial was downloaded from electronic patient records or retrospective notes review, as was recorded in the course of clinical care.
Primary outcome: To what extent is the trial's primary outcome directly relevant to participants?	3 – equally pragmatic and explanatory	The primary outcome was detection of small-for-gestational age as a proxy measure for stillbirth prevention. The latter is more directly relevant to participants, but it was not possible to power a cluster trial to test the effect of the intervention on this. This was a pragmatic decision and chosen because improving detection of SGA is the mechanism through which any effect on stillbirth was expected to act.
Primary analysis: To what extent are all data included in the analysis of the primary outcome?	5 – very pragmatic	Intention to treat analysis was conducted with multiple imputation of data.

The objectives of the DESiGN trial were to:

1. Determine whether GAP implementation improves the antenatal ultrasound detection of SGA (primary objective).
2. Investigate what effect GAP has on other short-term maternal and perinatal outcomes.
3. Conduct a process evaluation to assess fidelity, quality and acceptability of GAP implementation (part reported in Chapter 3).
4. Assess the cost-effectiveness of GAP through a health-economic evaluation (reported in Chapter 5).

2.1.1 Interventions

GAP was compared to standard care for antenatal screening and management of SGA. GAP has already been described in section 1.6.2 and summarised in Appendix section 10.1.

Before the trial, the Perinatal Institute altered the training requirement for clusters recruited to the trial, from that which is stipulated in the GAP-GROW protocol (section 1.6.2.1). The new requirement was that all sites must train at least 75% of members from each staff group (midwives, sonographers, and doctors) using both face-to-face and e-learning methods before release of GROW software to sites. During the trial, this requirement was modified again by the Perinatal Institute when it became evident that research clusters were finding it difficult to reach the e-learning training target. Whilst the target for over 75% staff to be trained using face-to-face methods before offering the intervention to women ('going live') remained, the timeline for the e-learning target was pushed back to be achieved by 3-months of 'going live'.

Standard care refers to the usual antenatal care that occurs in English maternity units that are not implementing GAP. There was no pre-specification of policies in the standard care arm of the DESiGN trial, except that they should not implement GAP or adopt customised perinatal growth charts. DESiGN clusters which were London-based (11 of 13 clusters) were also exempted by the London Perinatal Morbidity and Mortality group (NHS England) from compliance with the FGR requirements (element 2) of the 'Saving Babies' Lives' care bundle (section 1.6.1), widespread national implementation of which was recommended in 2016. Nevertheless, it was considered unethical to completely stop other local or national strategies which aimed to reduce the rate of stillbirth.

Since maternity unit guidelines have been informed by the RCOG Green-top guideline 'The Investigation and Management of the Small for Gestational Age Fetus' since its publication in 2002, with a major update in 2013, this was the basis of local guidelines in

standard care clusters during the DESiGN trial.[1] Guidelines were collected from standard care sites and statements from each local guideline were compared with statements of the key GAP components. This comparison has been included in Appendix section 10.2 and is summarised here. Two clusters provided no guidance on screening for SGA amongst low-risk women. In the other four standard care sites, low risk screening was conducted using serial fundal height measurement plotted onto population charts in two sites and using McDonald's rule (section 1.4.1.2) in the remaining two sites. To identify women at high risk of SGA, all but one site had guidelines that broadly followed the risk stratification protocol proposed either by the RCOG,[1] or the abbreviated version proposed by NHS-England.[217] Two of these sites employed uterine artery Doppler measurements at the fetal anomaly scan to aid in risk stratification. Frequency of serial fetal ultrasound was highly variable between sites, and within sites for different indications. Serial scan patterns included four-weekly scan policies, scans at set gestational ages (e.g. at 28 and 34 weeks, or at 28, 32 and 36 weeks) or single scan policies (e.g. at 36 weeks). None of the sites routinely conducted training sessions for staff on screening for SGA, nor were policies in place for auditing detection rates or missed case analyses, as are recommended by GAP.

2.1.2 Recruitment and randomisation procedures

Cluster randomisation into the trial commenced on 05 November 2016 and ended 05 July 2017. The number of clusters required to assess the primary objective was chosen following a power calculation (see section 2.2.1.3). Thirteen clusters were recruited and randomised at three time points (Table 2.2), determined by the date of study approval by the local NHS Research and Development team.

The first eight clusters were divided into two strata (clusters with the lowest birth rate during financial year 2013-14 in one stratum, those with the highest rates in the other stratum) and randomised to a trial arm on 03 November 2016. Three further clusters were randomised on 23 December 2016 (the allocation of two clusters to one arm rather than the other was also chosen randomly). The final two clusters were randomised to either trial arm on 05 July 2017. The birth rate did not affect the randomisation method for clusters allocated at the latter two time points. Allocation to the intervention or control group within each group was by random permutation using Stata v14 (StataCorp LP, College Station, Texas).

Table 2.2 - Random allocation of recruited clusters to trial arm

Date of Randomisation	Composition of strata		Allocation
03/11/2016	Small cluster strata*	Kingston Hospital NHS Foundation Trust (birth rate=5763)	Control arm
		Royal Surrey County Hospital NHS Foundation Trust (birth rate=3396)	
		West Middlesex University Hospital NHS Trust (birth rate=4774)	Intervention arm
		London North West Healthcare NHS Trust (birth rate=4863)	
	Large cluster strata*	Homerton University Hospital NHS Foundation Trust (birth rate=5877)	Control arm
		Guy's and St Thomas' Hospital NHS Foundation Trust (birth rate=6788)	
		Imperial College Healthcare NHS Trust+ (birth rate=8633)	Intervention arm
		University College London Hospitals NHS Foundation Trust (birth rate=6175)	
23/12/2016	The Hillingdon Hospitals NHS Foundation Trust		Control arm
	St George's University Hospital NHS Trust		
	Croydon University Hospital		Intervention arm
05/07/2017	North Middlesex University Hospitals NHS Trust		Control arm
	Chesterfield Royal Hospital		Intervention arm

*Cluster strata determined by the birth rate at these maternity units during the financial year 2013-14. Birth rate therefore provided for each unit. Birth rate did not affect randomisation methodology for clusters allocated at the latter two time points.

+This NHS Trust contains maternity units at two hospital sites; however, the units share leadership and clinical care guidelines; they have therefore been randomly allocated into the trial as a single cluster.

For the remainder of this thesis, cluster sites will be referred to by anonymised site numbers to ensure data from women and staff are treated confidentially, but still allow both between and within site comparisons of results throughout this thesis. Control sites are numbered 1-6; implementation sites are numbered 7-13.

2.1.3   Inclusion and Exclusion Criteria for Women and Babies

Women and their babies were included in the DESiGN trial if they gave birth during the trial period (from the date of cluster randomisation into the study until 28 February 2019, the end of the outcome period) or during the year prior to randomisation (pre-randomisation / baseline period). Women were given the opportunity to opt-out of data

sharing with the DESiGN trial, by the advertisement of this option on posters and leaflets in the antenatal departments of DESiGN clusters. We were not notified of any women requesting to opt-out of data sharing.

Women and babies were excluded from the primary DESiGN analysis, and all analyses reported in this thesis, if the birth occurred prior to 24^{+1} weeks of gestation, if there was a multiple birth, or the fetus had a significant congenital anomaly identified on fetal ultrasound screening.

2.1.4 Ethical Review

Ethical review of the trial protocol was conducted by the Bloomsbury Research Ethics Council (REC Ref: 15/LO/1632) and the UK Confidentiality Advisory Group (CAG Ref: 15/CAG/0195). Ethical review by CAG was required for the use of women's medical record data without direct consent (on an opt-out basis) and so that the DESiGN research team were able to access identifiable data at the clinical research sites (but not to remove this data from the sites). This was essential for accurate linkage of data from different electronic patient record (EPR) systems, through pseudonymisation processes.

2.1.5 Patient and public involvement

During the conception of the DESiGN trial, stakeholder groups including patient representatives (the London perinatal morbidity group, the RCOG Stillbirth Clinical Study group, Sands charity and Tommy's charity), were consulted for their opinion about the study relevance and the plan to use EPR data without direct patient consent. Patient or public representatives were provided with a lay summary of the study plan. The feedback from all stakeholder group participants was used to inform both the final study protocol and the application for ethical approval. A patient representative from the London perinatal group and Tommy's charity was invited to provide continued participation throughout the study, including through attendance to all co-investigator meetings and review of the trial manuscripts. He has aided in both interpretation of results and explanation for a lay audience.

2.1.6 Trial oversight

A joint Data Monitoring and Trial Steering Committee (DMC/TSC) was formed and first met in February 2018. Three meetings were held per year in 2018 and in 2019. A meeting was convened in January 2020 to share the initial trial results and a final meeting was convened in 2021, just prior to submission of the trial results manuscript for peer-review and publication.

The DMC/TSC reviewed and approved the trial protocol, the statistical, health-economic and process evaluation analysis plans, and the data management processes. They instigated a quarterly safety reporting process in the intervention arm of the trial – no serious adverse events were attributed to GAP implementation. The committee were informed of the trial progress at each meeting, including discussion of planned ethical amendments and trial extensions.

2.1.7 Trial registration

The trial was pre-registered on the ISRCTN Registry (Ref:67698474) on 02 November 2016. https://doi.org/10.1186/ISRCTN67698474

2.1.8 Funding Sources

2.2 Studies of Clinical Outcomes

Analyses conducted for the DESiGN trial and for this thesis have distinct methodological differences for the study design, exposures and outcomes, collection of data and techniques of analysis. The analyses can be grouped into three categories, further methods of which are detailed in separate sections:

1. Studies of clinical outcomes (this section)
2. Studies of implementation - process evaluation (section 2.3)
3. Studies of cost-effectiveness – health economic evaluation (section 2.4)

In this first section I will outline the clinical outcomes, statistical power, data collection strategy and data management methods of the DESiGN trial. These decisions and processes are also relevant to the secondary clinical analyses conducted for this thesis (Chapters 6 and 7), which use the same data. I will also outline the statistical analysis techniques employed for studies of clinical outcomes in this thesis.

2.2.1 Outcomes

The choice of clinical outcomes for the DESiGN trial primary analysis determined the data collection strategy, and available data. Whilst this limits the availability of data to that which was required for the primary analysis, DESiGN and both secondary clinical studies reported here were conducted to answer research questions relevant to the assessment and management of fetal growth anomalies; they therefore required very similar data.

2.2.1.1 *Primary clinical outcome*

Complex antenatal interventions such as GAP are intended to reduce the rate of stillbirth through the antenatal detection and appropriate management of SGA babies. Designing a cluster RCT with sufficient power to detect the anticipated difference in stillbirth (a rare outcome) was unfortunately not considered feasible because too many clusters would be required at a time when GAP uptake already covered over 60% of English and Welsh maternity units. Antenatal detection of SGA was therefore chosen as a proxy primary outcome.

The primary clinical outcome of the DESiGN trial was defined as '*the rate of antenatal ultrasound detection of SGA'*. Since a comparable outcome was required across both trial arms, the denominator for both control and intervention clusters was birthweight less than the 10th centile for gestational age as defined by both customised standards and population references (SGA_{both}). The numerator was defined by these babies for whom an antenatal

diagnosis of SGA had been made, using the definition of SGA in that arm of the trial at the time:

- During the trial outcome period, the intervention clusters defined suspected SGA as EFW less than the 10th centile for gestational age according to the customised standard.
- Intervention clusters defined suspected SGA as EFW less than the 10th centile for gestational age according to population references during the baseline (pre-randomisation) and pre-implementation periods.
- The control sites defined suspected SGA as EFW less than the 10th centile for gestational age according to population references during all trial periods.

Fetal ultrasounds are only offered in the third trimester of pregnancy if the woman is at high risk of having an SGA baby, if SGA is suspected following clinical assessment, or for other obstetric indications i.e., when a scan is required to check fetal presentation or placenta location (section 1.4.1). For this reason, we assumed that women who had no record of a scan in the ultrasound data, had not had a scan and therefore, that a fetal growth anomaly had not been identified antenatally.

The definition of antenatal ultrasound detection of SGA as described above, has also been used in the analysis described in Chapter 6 of this thesis. Furthermore, the same methods have been used to define antenatal ultrasound detection of LGA for the analysis described in Chapter 7, the only difference being that the numerator and denominator apply to EFW>90th centile and birthweight>90th centile respectively.

When planning the trial, estimates for the congruence of the SGA definition by customised and population centiles were obtained from previous studies for the purpose of calculating the required sample size.[85,94,137,288] From the pooled estimates, it was concluded that 75% of neonates who were defined as SGA by customised standards were also defined as SGA by population standards. Since the threshold for defining SGA in a population is <10th centile, it follows that 12.5% of infants were expected to be SGA by either definition and 7.5% were expected to be SGA by both definitions (Figure 2.1).

Figure 2.1 - Percentage of infants who are SGA by customised (cust), population (pop) or both centile definitions

2.5% babies are SGA as defined only by both customised customised charts

2.5% babies are SGA as defined only by both population customised charts

7.5% babies are SGA as defined by both population and customised charts

87.5% babies not SGA by either criteria

2.2.1.2 *Secondary clinical outcomes*

The pre-determined secondary clinical outcomes of the DESiGN trial are listed in Table 2.3. This informs the availability of data on clinical outcomes for the analyses described in Chapters 6 and 7. These outcomes can be categorised into three groups:

- The performance of the screening test, including performance when the definition of SGA is varied,
- Perinatal outcomes, the rates of which have the potential to change with GAP implementation. Where GAP increases detection of SGA, monitoring of these fetuses may lead to iatrogenic prematurity, which potentially risks neonatal morbidity,
- Maternal outcomes with particular emphasis on changes to the rate of interventions or adverse intrapartum outcomes in exposed mothers.

Table 2.3 - Pre-determined secondary clinical outcomes of the DESiGN Trial

Clinical Outcomes		
Performance of the screening test	**Perinatal outcomes**	**Maternal outcomes**
Rate of antenatal ultrasound detection of SGA at birth by customised standards and separately by population references.	**Basic parameters**: Gestational age at birth Birthweight Head circumference	**Antenatal:** Length of stay in hospital
Antenatal clinical detection* of SGA.	**Condition at birth:** 5-minute Apgar score <7 Arterial cord pH <7.1 Any respiratory support given at delivery	**Intrapartum:** Induction of labour Mode of delivery (including Caesarean section rates)
Analysis of GAP diagnostic test performance (specificity, sensitivity, negative predictive value, positive predictive value).	**Neonatal admissions:** Length of stay at each neonatal level of care	Postpartum haemorrhage Rates of 3rd or 4th degree perineal tear
Ultrasound assessment of SGA using a different threshold e.g., 5th centile. Growth trajectories (fetal biometry and EFW) and Doppler parameters in the detection of SGA**.	**Neonatal morbidity:** Major neonatal morbidity (any of neonatal brain injury, receipt of supplemental oxygen at 28 days of age, Bell stage 2+ necrotising enterocolitis, Culture-positive sepsis, retinopathy requiring ophthalmic intervention).	**Postnatal:** Length of stay in hospital Breastfeeding at discharge
Comparison of GROW ultrasound charts against standard population charts on classification of fetal growth (SGA, AGA or LGA).	Minor neonatal morbidity (any of: hypothermia, hypoglycaemia, nasogastric tube feeding)	
	Perinatal loss: Antepartum or intrapartum stillbirth Neonatal death (early or late) Death before neonatal discharge (after 28 days of birth) Cause of death.	

*Clinical detection of SGA is defined as 'antenatal acknowledgement that the fetus is expected to weigh below the 10th centile at birth, by charts appropriate to the study arm'.

Due to data availability, the secondary outcomes reported in the final DESiGN analysis are slightly modified from those which were pre-specified. Data were unavailable on Bell stage of necrotising enterocolitis, culture status of babies with sepsis and ophthalmic referrals for babies with retinopathy of prematurity. The definition of major neonatal morbidity was therefore changed to 'any of: neonatal brain injury, receipt of supplemental oxygen at 28 days of age, *any* necrotising enterocolitis, *clinically-defined* sepsis, retinopathy *of prematurity'*.

2.2.1.3 *Statistical Power*

The minimum number of clusters in the trial was calculated as follows:

1. The mean birth rate in London maternity units likely to participate in the trial was calculated as 5,053 births per year based on data provided by the trusts for births during the financial year 2013/14.

2. The assumptions for the number of births which meet the criteria for the numerator and denominator of the primary outcome are explained above (section 2.2.1).

3. During the trial outcome period (data collection planned for four months of births) in a standard cluster, it was therefore expected that 42 (2.5%) babies would be born SGA as per criteria for customised standards, 42 (2.5%) babies would be born SGA as per population references and 126 (7.5%) babies would be born SGA by both population references and customised standards. This is the number of births relevant to the primary outcome denominator.

4. With regards to the expected number of SGA fetuses who are suspected to be SGA following antenatal screening (the numerator of the primary outcome), published reports suggest that only 20% of SGA fetuses have suspected SGA when screened using standard antenatal practice.[74,214,289] However, in the intervention arm, it was expected that GAP would increase the antenatal detection rate from 20% to 33%. This increase from 20% to 33% reflects a doubling in the odds of detection (4/1 to 2/1) which reflects a modest yet clinically meaningful difference. Evidence from observational research suggests that GAP can result in a far higher increase in the rate of SGA detection.[282,283]

a. An intra-cluster correlation coefficient [*] for detection of SGA from published literature was required but could not be identified. A cluster coefficient for detection of fetal growth restriction was therefore used, as the most similar clinical outcome. This was identified as 0.019 in the only source.[290]

5. Using α=0.05 and β=0.2, it was determined that the minimum number of clusters to provide statistical power to the trial was 12 (6 per trial arm). Overall, this provides 84% power to demonstrate superiority of GAP at the 5% significance level, with the condition that the above assumptions hold as a minimum.

The intra-cluster correlation coefficient actually observed in the main DESiGN trial analysis was 0.008 (95% CI 0.002 to 0.039). This was lower than that which was estimated

---

[*] The intra-cluster correlation coefficient is a measure of how related observations within a cluster are to one another, as opposed to how related they are to observations in other clusters. Mathematically, it is a comparison of the within-cluster variance to the between-cluster variance. A value of 1 occurs when all observations within a cluster are identical. A value approaching 0 represents higher variance within the cluster.

from a proxy outcome. The observed coefficient would have provided 93% power (with α=0.05 and β=0.2) to detect a difference in the primary outcome.

2.2.2 Data collection

From the outset of this trial, data collection was planned to use datasets downloaded from established EPR systems, rather than bespoke data collection. This was necessitated by the moderate trial budget and number of individual women included. Entry of data on trial outcomes into a research database by research staff was not feasible for the 30,000 births expected in 12 maternity units during the 4- to 6-month outcome phase of the trial. In addition to this, data were required from the baseline period for an expected additional 60,000 women. Bespoke data collection was expected to be time-consuming and therefore not cost-efficient.

The data required for the clinical studies fall into four categories according to the expected source EPR system (Figure 2.2):

- Data on maternal demographics and clinical details of the pregnancy and birth events and outcomes (Maternity EPR),
- Data on fetal ultrasound scan findings (Ultrasound EPR),
- Data on the care of babies admitted to neonatal units (Neonatal EPR),
- Data on administrative aspects of maternity care e.g., clinic appointments, date of hospital admission or discharge (Administrative EPR).

Figure 2.2 - Types of data and source electronic patient records
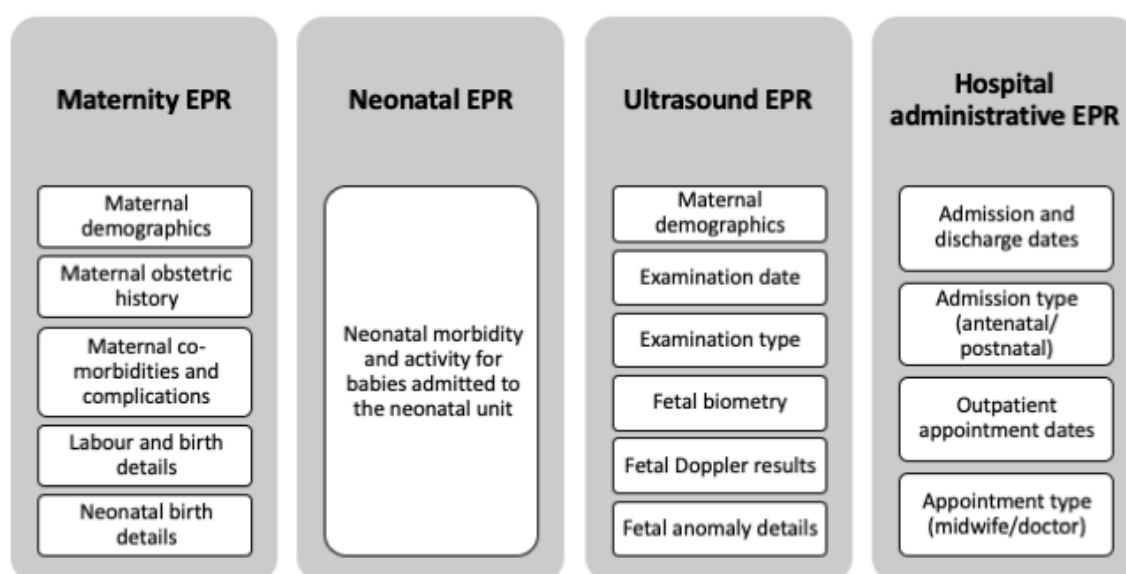
The maternity EPR was expected to act as a spine onto which the other three data types were to be linked (Figure 2.3).

Figure 2.3 - Planned linkage of the four dataset types
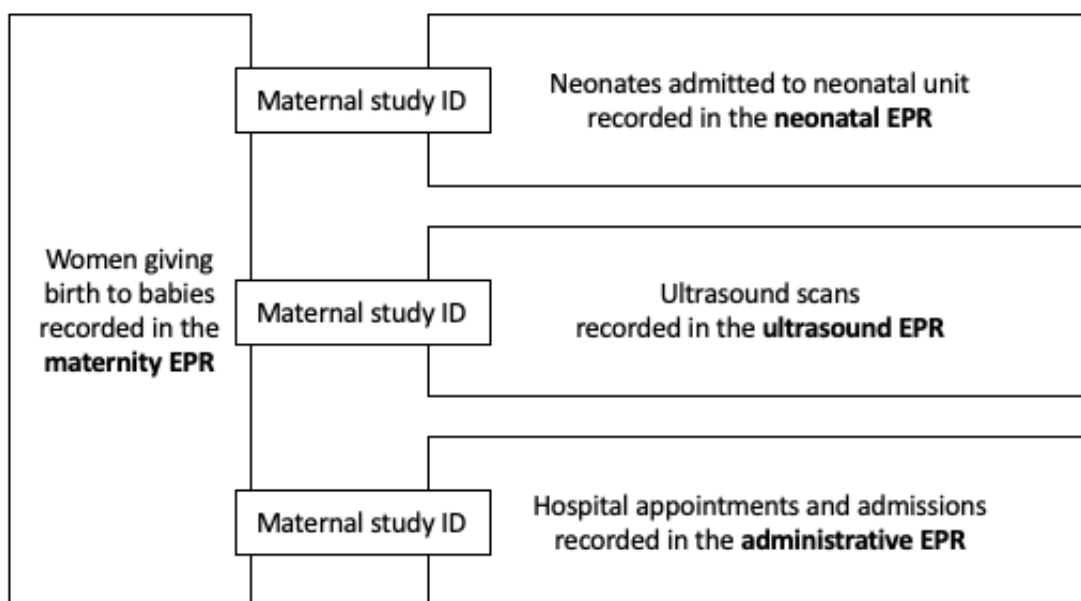
Data were required from three trial phases:

1. One year pre-randomisation (to study baseline characteristics and use these to adjust outcomes across clusters),
2. A variable implementation washout phase (during which the sites allocated to the intervention were preparing for and implementing the intervention),
3. The trial outcome phase (a 4–6-month period during which all births in a site contributed to the primary analysis of the trial outcomes and the care continued as allocated).
    i. 6 months of data from all control sites
    ii. 6 months of data from all intervention sites where possible, however if implementation of the intervention was late, the trial outcome phase started at least 7 months following implementation (6 months for the pregnant women to give birth at full term after their charts are printed at 12 weeks' gestation and 1 additional month to account for variation and late term births).
    iii. There is one exception for the trial data collection period. One cluster was allocated to the control arm of the trial but then experienced significant external pressure to implement GAP in July 2018. Furthermore, the outcome phase of the trial was delayed by the time taken for the last cluster allocated

to GAP, to complete implementation. Their trial data were therefore used from the 6 months prior to when they commenced implementation, which they had planned at the expected end date of the trial (before this was postponed).

The specification of a data request form (Appendix section 10.3) was developed by the trial team to ensure all data were provided that were essential to describe the trial population and assess the trial primary and secondary outcomes. The form was tested by the maternal and neonatal EPR administrators at the lead cluster site.

Since the neonatal data were expected to come from the same EPR software in all units (CleverMed Badgernet Neonatal), a data extraction tool was built, specific to that software (the code is available in Appendix section 10.4). This was intended to assist the site clinicians in quickly extracting the relevant neonatal data for the study. It was also possible to share pre-built data extraction tools for two of the ultrasound EPR software types (Viewpoint, GE Healthcare and Astraia).

These data were collected at two time points from a key point of contact (usually a research midwife) at the 13 cluster sites. Data were requested in Microsoft Excel spreadsheets. The first request was initiated in January 2018 for data on births during the pre-randomisation trial phase (phase 1), to allow for data quality checks and troubleshooting prior to the subsequent request. The last research site supplied data for this download request in February 2019. The second request was initiated in March 2019. This request was for a re-run of any data downloaded for the pre-randomisation phase that had been amended following resolution of quality issues from the first request and for data from the implementation washout and trial outcome phases (trial phases 2 and 3). The last dataset was collected in October 2019.

The data collection and management stages are summarised in Figure 2.4.

Figure 2.4 - Stages of data collection and data management

Stage	Description
First data request	• Data requested for all births during one year pre-randomisation from four types of EPR (maternity, neonatal, ultrasound and hospital activity) • Data pseudonymised on site and sent centrally using secure NHSmail
First data quality check	• Data quality checks conducted to include: availability of all four data types, availability and completeness of all requested variables, time period covered, duplicate study IDs, distribution of continuously reported values, linkage rate between data types. • Data problems discussed with cluster sites and data specification clarified
Second data request	• Data requested for: (i) amended data from pre-randomisation period following resolution of data quality issues, where possible; (ii) new data from implementation washout and trial comparison periods • Data pseudonymised on site and sent centrally using secure NHSmail.
Second data quality check	• Data assumed to have not deteriorated in quality since first data request • Data quality checks conducted to include assessment of: availability, completeness and distribution of data items key to calculating the trial primary outcome.
Data harmonisation	• Data harmonised by data management team using the data dictionary. The data dictionary detailed how to format all data items from each cluster site, so that they were the same across all thirteen cluster sites.
Data linkage	• Unique babies and pregnancies within cluster sites were identified and their records from each of neonatal, ultrasound and activity datasets were linked to the corresponding record in the maternity dataset.
Data management	• Outlier rules developed and applied. Values which lay outside outlier limits were converted to missing. • Rules for multiple imputation of key characteristics and outcomes were developed according to whether the variable was expected to vary between cluster sites, the trial arm or neither.

2.2.2.1 *Process of pseudonymisation*

Data provided by the maternity sites included patient identifiers: NHS patient number, hospital patient number, date of birth (DOB) and postcode. Patient identifiers were required so that a unique study identifier (ID) could be generated for each woman in the study across all linked datasets, using a pseudonymisation tool.

The pseudonymisation tool was developed by the research team as an Excel macro using Microsoft Visual Basic for Applications and refined and tested using simulated data in Stata v15 to prevent generation of duplicate study identifiers. The simulation dataset included fictitious NHS numbers, DOBs and dates of delivery for the infants. All simulated women had DOB between 1st January 1989 and 31st December 1990 and date of delivery between 1st January 2017 and 30th November 2018. Narrow date ranges were chosen to increase the number of duplicate dates and therefore test the risk that different women could be allocated the same study ID.

A simulation of 100,000 records using different maternal identifiers and random numbers produced a non-duplicated study ID with the fourth iteration (Table 2.4). Maternal NHS number and DOB were chosen as the identifiable variables due to low missingness and high reliability. The final tool (test round 4) created a 20-character study ID where the same pseudonym was always generated for the same patient. The algorithm for the pseudonymisation tool cannot be shared because of the risk of de-anonymisation.

Table 2.4 - Refinement of the pseudonymisation tool

Test round	Maternal data components used	Study ID format	Study ID length	Duplicates, n/N
1st	DOB, DOD, NHS number	########## e.g.,1234567890	10	18,382 /100 000
2nd	DOB, DOD, NHS number	############ e.g., 123456789012	12	18,382 /100 000
3rd	DOB, DOD, NHS number, Check digit	###############XXX e.g., 123456789012345ABC	18	21 /100 000
4th	DOB, NHS number, check digit, random component with seed	###################X e.g., 1234567891234567891A	20	0 /100 000

DOB – Date of birth; DOD – Date of delivery (of neonate); ID – identifier, NHS – National Health service

On site, I conducted the pseudonymisation procedure with a data manager, under the supervision of the cluster's key clinical contact, in keeping with ethical approval for data flow. A manually produced pseudonym was generated if a woman did not have both the variables needed by the pseudonymisation tool.

Of the 201,209 records in the final dataset, 1,261 (0.6%) did not have key information required to generate an automated study pseudonym and so were given a manually

produced ID. Whilst the study ID could identify the same woman across records from any site, it is important to note that this manually produced ID would not be given to the same woman if she had another birth elsewhere during the whole trial period.

2.2.2.2 *Data extraction and storage*

Following generation of a pseudonymised ID, the women's DOBs were also used to calculate their age at delivery of the neonate and their postcodes were used to generate measures of socioeconomic deprivation (index of multiple deprivation, lower layer and middle layer super output areas) using the National Statistics Postcode look-up tables.[291] Following this, all identifiable data (NHS/hospital numbers, maternal DOB and postcodes) were removed from the pseudonymised dataset.

The pseudonymised ID allowed early linkage between the four datasets. Women and babies who featured in the neonatal, ultrasound or activity datasets, but did not have a record of a birth in the maternity dataset were identified and removed from these three datasets. Later linkage was conducted to link unique pregnancies across the four datasets, for women who had more than one pregnancy during the study period.

All pseudonymised datasets were checked for absence of patient identifiers by a data manager, a local clinical contact and me before the data were electronically transferred back to the central research site using NHS Digital's secured email system, NHSmail,[292] and stored on the servers based at King's College London. The keys linking the newly generated study ID with the women's identifiers were left with the clinical contact, to be stored for a minimum of 5 years.

2.2.3 Data management processes

Data management processes were conducted by a specialist team, under the direction and specification of the clinical trial team (with me operating as the primary clinical contact).

2.2.3.1 *Assessment of data quality and completeness*

The data collected at the first download (for the pre-randomisation trial phase) underwent the following checks for completeness and plausibility:

- Presence of all four requested datasets (maternity, neonatal, ultrasound and activity data),
- Duplication of study IDs - includes assessment of whether these were true duplicates, or the same woman with multiple birth(s),

- Matching of study IDs across the four datasets - proportion of the study IDs from the maternal data which appeared in the linked datasets,
- Presence of the requested variables in the dataset,
- Level of completeness of the requested variables,
- Range, median, 5th, 25th, 75th and 95th centile of the continuously reported variables e.g., maternal age, height, weight,
- Date range of all reported births and hospital activities – checking these were within the requested timeframe.

Where data quality issues could be rectified, these were addressed with the cluster site before the second data download. If data quality issues could not be resolved, these were recorded.

The second data download was then subject to a limited data quality checklist (it was assumed that the data quality had not deteriorated since the first download):

- Resolution of any data quality issues raised from data quality checks on the data downloaded following the first request,
- Assessment for duplicate study IDs,
- Completeness of the variables which are key to the calculation of the trial primary outcome,
- Distribution of the continuous variables required to calculate the primary outcome (as per assessment of numerical distribution above),
- Value responses to categorical variables required to calculate the primary outcome,
- Date range of all reported births and hospital activity data.

2.2.3.2 *Harmonisation of data from multiple sources*

To harmonise the four datasets from the different EPR systems (Table 2.5), a data dictionary was developed using the list of requested variables. This was intended to guide the data management team in the harmonisation process by listing the abbreviated and full variable names, a description of the variable and, for each research cluster, a guide to the re-categorisation of textual data (e.g., mode of birth) or re-calculation of units for numerical data. A sample of the full data dictionary can be viewed in Appendix section 10.5.

Table 2.5 - EPR systems used at sites for each data type

Maternity EPR system	Ultrasound EPR system	Neonatal EPR system	Hospital administrative EPR system (appointments / admissions)
Medway Maternity	Astraia	Badgernet Neonatal (CleverMed)	Medway
K2	Viewpoint (GE Healthcare)		PAS
E3			CMIS
Cerner	CRIS		CareCast
Euroking	Solitorn		EPR
CMIS	RIS / PACS		APAS
EPR			OASIS
Badgernet Maternity (Clevermed)			iClip

A standardised nomenclature (Figure 2.5) was used for the abbreviated variable names to identify the source of the data (e.g., maternal, or neonatal EPR systems) and the degree of data management which had been done to the variable (e.g., raw, harmonised, or calculated variable).

Figure 2.5 - Standardised Nomenclature for Data Dictionary
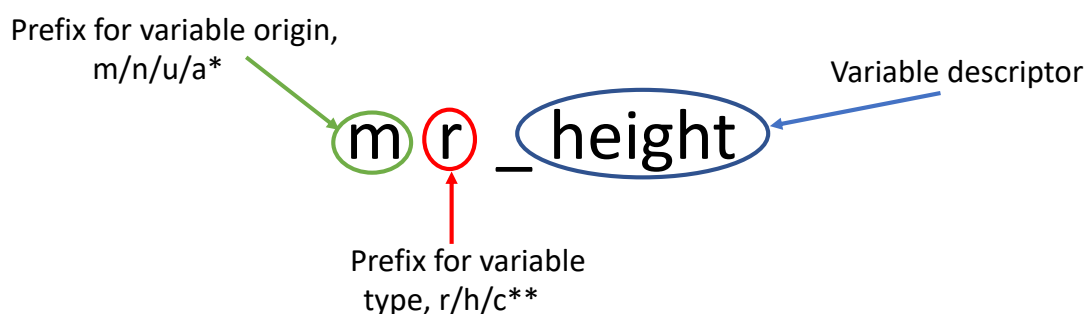
*Variable origin: m=maternity dataset, n=neonatal dataset, u=ultrasound dataset, a=activity dataset

**Prefix for variable type: r=raw (original), h=harmonised, c=calculated

For continuous numerical variables, guidance was provided on the appropriate units to use, and how to calculate where necessary (e.g., if height was provided in feet and inches, it was converted to metres and then used to calculate BMI).

For categorical variables, site-specific guidance was produced on how to change the variable to the harmonised version (e.g., how to merge the multiple descriptions of mode of birth into three categories: unassisted vaginal birth, assisted vaginal birth and caesarean section). The way in which text responses to categorical (or free text) variables were to be (re)categorised was decided in advance by consensus of the clinicians in the research team, following both familiarisation of the early datasets and consideration of what categories would be useful for the final planned analyses. Where possible, variables were re-categorised as binary e.g., pre-existing hypertension: 'yes'/'no'. In all cases, but particularly

important where it was not possible to re-categorise a variable according to the pre-planned categories, the raw data were also kept in the final dataset. The main clinical research team (Professor Dharmintra Pasupathy, Dr Matias Vieira, Professor Andrew Copas and me) familiarised ourselves with the available data formats and agreed rules for data harmonisation through consensus, in advance of the active data management processes.

- Within any cluster's dataset, if only an affirmative value was recorded for a binary variable (e.g., presence of chronic hypertension), the missing values were treated as negatives (i.e., no hypertension) and changed to values which reflected this. We assumed that only the affirmative option was available to the person who entered the data.
  - Where negative values were recorded as well as affirmative, missing values were left missing during data management. For some instances however the proportion of affirmative values in the cluster was close to the proportion expected following comparison with national audit results from the same period,[293] (e.g. for severe perineal trauma) suggesting missing values would *almost* always be negative. These missing values were not imputed and were subsequently treated as negative in analysis.
- Within any dataset, where data for clinical diagnoses or events were available from more than one raw source, any record of the value was regarded as it being positive, even if it was not recorded as positive elsewhere. For example, where a woman was recorded as having an epidural in the 'labour anaesthesia' variable but this was not recorded in the 'birth anaesthesia' variable (both from the maternity dataset), she was regarded as having had an epidural.

Harmonisation was an iterative process. Initially, only data derived from maternity EPR were harmonised. I then checked the harmonised maternity dataset (one per hospital) for adherence to the data dictionary and data management rules. Where errors were made, I sent a written list of required edits to the data management team, who effected the changes. This meant that a clear audit trail was in place for changes to the data. The same process was then repeated for the updated versions of the maternity datasets and for the ultrasound, neonatal and activity datasets, until I was satisfied with the final datasets.

2.2.3.3 *Harmonising the ethnicity and country of birth data*

The Perinatal Institute software for calculating the customised weight centile of a fetus or neonate requires granular information regarding the mother's ethnicity, preferable to the level of country of family origin. The level of detail required is detailed in Table 2.6.

Table 2.6 - Ethnic origins, as specified by the GROW online fetal/birthweight centile calculator

Ethnic Origin	Examples of Country of Origin
British European	England, Northern Ireland, Scotland, Wales
Irish European	Ireland (Republic of)
North European	Denmark, Finland, Iceland, Norway, Sweden
East European	Albania, Bulgaria, Hungary, Latvia, Poland, Romania, Slovakia, Ukraine
South European	Cyprus, Greece, Italy, Malta, Portugal, Spain
West European	Austria, Belgium, France, Germany, Netherlands, Switzerland
North African	Algeria, Egypt, Libya, Morocco, Sudan, Tunisia
East African	Eritrea, Ethiopia, Kenya, Madagascar, Mauritius, Somalia, Tanzania, Uganda
Central African	Cameroon, Central African Republic, Chad, Congo, South Sudan
South African Black	Angola, Botswana, Mozambique, Namibia, South Africa, Zambia, Zimbabwe
West African	Ghana, Ivory Coast, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone
Middle Eastern	Iraq, Israel, Jordan, Lebanon, Saudi Arabia, Syria, Turkey
Bangladeshi	Bangladesh
Indian	India, Sri Lanka
Pakistani	Pakistan
Chinese	China, Hong Kong, Taiwan
Other Far East	Japan, Korea, Mongolia
South East Asia	Brunei, Cambodia, Indonesia, Malaysia, Myanmar, Philippines, Thailand
Caribbean	Barbados, Cuba, Jamaica, Puerto Rico
Mixed African-European	
Mixed Asian-European	
Mixed Caribbean-European	
Other	Australasia, Americas, Asia - Other
Unclassified	

Such granular ethnicity was not always available from the raw hospital data, and so a set of assumptions were agreed amongst the trial team to guide the development of the harmonised, granular ethnicity variable.

1. Where ethnicity provided by the hospital maternity EPR was as granular as that required by perinatal institute calculator, we kept hospital-determined ethnicity
2. Where ethnicity provided by the hospital was less granular than that required for customisation, we referred to the country of origin (if available).
   1. We assumed that the country of birth = ethnic origin (and therefore granular ethnicity) IF the common ethnicity in the country of birth matched the ethnicity provided by the hospital maternity/ultrasound system.
   2. If there was a discrepancy in ethnicity and common ethnicity of country of birth:
      1. We looked for ethnicity from the ultrasound EPR and applied rules 1, followed by 2.

2. If there remained a discrepancy in ethnicity from both the maternal and ultrasound EPR and country of birth, we allocated the ethnic origin according to the most common white, Black, Asian or mixed ethnicity in the borough in which the cluster is based.

3. If ethnicity was 'other/mixed' (therefore not possible to determine if matched or discrepant, compared to country of birth), we used country of origin where available

4. If it was not possible to assign ethnic origin using the rules above, we assigned 'Other' ethnicity if ethnicity was known, or 'Unknown' if not known.

2.2.3.4 *Data linkage for the linkage of unique pregnancies*

Data linkage to match data for the same pregnancy or infant from the neonatal, ultrasound and activity datasets (Figure 2.3) with the correct pregnancy and birth in the maternal dataset (some women had more than one pregnancy or baby during the trial period) was conducted using the pseudonyms generated at the time of data collection and the following rules, by adding a '_n' suffix to the study ID (where 'n' refers to the nth pregnancy at a particular site during the whole trial period for the woman with that study ID):

- The neonates in the neonatal EPR were matched to the correct mother and pregnancy in the maternal EPR system using (i) the maternal study ID (present in both datasets) and (ii) the neonatal DOB within seven days of the mother's date of delivery.

- The hospital activity within the administrative dataset was matched to the appropriate mother and pregnancy using (i) the maternal study ID (present in both datasets) and (ii) the timing of the appointment or admission: this was required to fall between the estimated date of conception (EDC = date of delivery - gestational age at birth + 14 days) and the date of delivery. For this trial, we were not collecting data on postnatal readmissions.

- The ultrasound scans within the ultrasound dataset were matched to the correct mother and pregnancy using (i) the maternal study ID (present in both datasets) and (ii) the timing of the ultrasound scan: this was required to fall between the EDC and the date of delivery.

All (100%) study IDs from the linked (neonatal, ultrasound and activity) datasets matched with study IDs from the maternal dataset because the IDs in the linked datasets were derived from the maternal dataset. The percentage of pregnancy IDs in the maternity dataset which were linked with the other three datasets are presented in Figure 2.6. The linkage rate for the maternal data with the neonatal dataset was low because babies only

have neonatal records when admitted to neonatal care (e.g., preterm or born in poor condition). The linkage rate of the maternal and activity dataset was affected by one trial site which was unable to provide activity data for any of their women. The linkage rate between the maternity and ultrasound or activity datasets was not expected to be 100% because not all women who give birth in a maternity unit had antenatal care in the same maternity unit (they may have received either no antenatal care, or antenatal care elsewhere).

Figure 2.6 - Linkage rates between the maternity and linked datasets
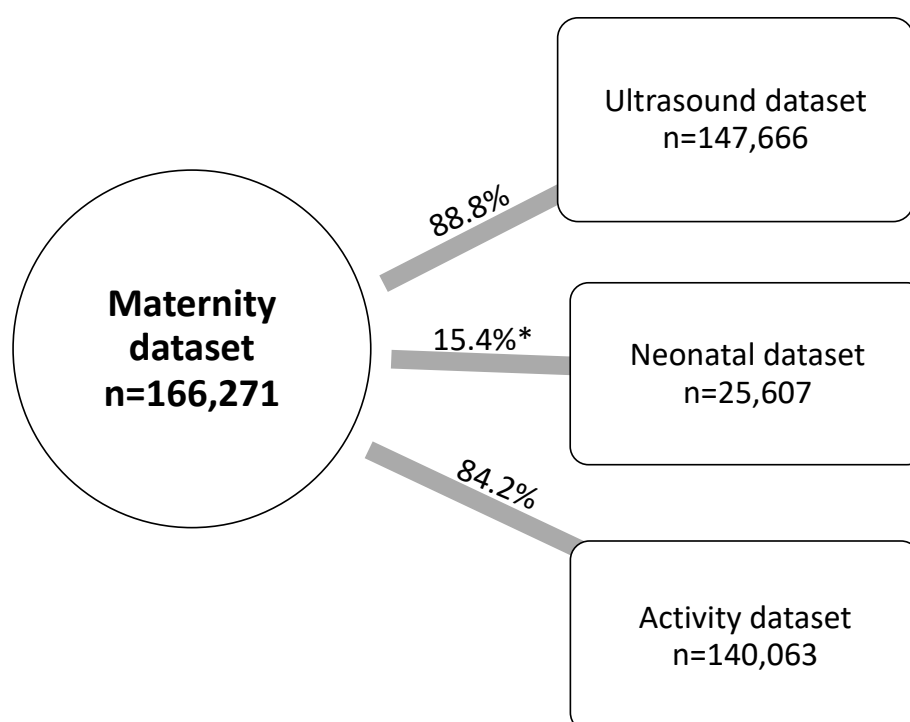
Linkage rates represent the percentage of study IDs from the maternity dataset which were also found in the linked dataset.

*Only those infants who were admitted to neonatal care are expected in both datasets, the linkage rate reflects this.

2.2.3.5 *Assessing data quality in the linked dataset*

Quantitative variables were assessed according to an outlier policy. We calculated the 3rd, 4th and 5th standard deviation limits and the 1st and 99th centiles in the data distribution for each continuous variable. These were used to highlight possible implausible values that were likely erroneous data entries. However, these distributional cut-points are arbitrary and are also influenced by whether the variable has an (approximate) normal distribution. Simply removing variables beyond these limits was expected to result in accurate data being removed. These limits were therefore used as indicators of potential erroneous values and the final outlier limits were derived following trial clinician consensus on values which were sensible (Table 2.7). Values outside the outlier limits were converted to missing.

Table 2.7 - Outlier limits derived following clinical consensus

	Lower limit	Upper limit
Age (years)	13y	60y
Height (cm)	120cm	200cm
Weight (kg)	30kg	200kg
BMI (kg/m²)	13kg/m²	70kg/m²
EBL (mL)	1mL	15,000mL
Birthweight (g)	100g	6000g

Levels of missing data for key variables required to calculate the trial primary outcome were compared across the linked dataset, following application of the outlier policy, comparing trial arms at two trial phases (pre-randomisation and trial outcome phases). The findings are summarised in Table 2.8.

Table 2.8 - Levels of missing data for key variables after the second data download, comparing trial phases

	Pre-randomisation phase		Trial outcome phase	
	Control	Intervention	Control	Intervention
Maternal height	25.0%	11.0%	20.7%	3.0%
Maternal weight	15.3%	21.6%	12.0%	12.9%
Ethnicity	3.7%	14.1%	3.1%	7.1%
Parity	11.1%	15.2%	15.1%	7.4%
Neonatal sex	0.01%	0.1%	0.1%	0.5%
Neonatal birthweight	0.1%	0.3%	0.3%	0.5%
Gestational age	0.2%	4.8%	1.3%	3.8%

For neonatal sex and birth weight there were very few missing data, there were more missing data for maternal BMI, despite infilling missing values in the maternity dataset with data from the ultrasound dataset. For variables with a notable level of missing data there was variation between trial arms and the period of data collection. Those allocated to the control arm were more likely to have missing BMI data than those in the intervention arm

with this difference being more marked in the trial outcome phase as the extent of missingness decreased more in the intervention than control arm over time. By contrast ethnicity was more likely to be missing in the intervention arm than control arm, but the proportion with missing data for both variables decreased between the pre-randomisation and trial outcome phases. In the pre-randomisation phase missing parity data were high (11-15%) in both trial arms, this was because we requested parity as an integer (as required to calculate our primary outcome), but it was often only supplied as categorical (nulliparous/multiparous). Nulliparity can easily be converted to parity=0, but multiparity cannot.

2.2.4   Management of missing data

2.2.4.1   *Infilling of maternal anthropometric data from other datasets*

It was expected that maternal data on height, weight and BMI might arise from either the maternity EPR (where midwives are expected to enter this data at the time of booking for antenatal care), or from the ultrasound information system (because such data is entered for the calculation of a woman's risk for fetal trisomy during the appointment for ultrasound dating scan and combined screening test). The following steps were therefore developed to determine which data to use, and to infill where data was missing from one source:

1. Assess for impossible values and change them to missing.
2. Assess degree of data completeness for women with both height and weight recorded in hospital reports from the maternal data and the ultrasound data.
3. The report with higher levels of completeness would be used as the primary source of data for this cluster site's harmonised height and weight.
4. Missing data on height and weight in the harmonized variable after step 3 above were completed with the data available in the other dataset (the one that was less complete in point number 2).
5. Calculate missing height (from weight and BMI) and weight (from height and BMI) in each dataset, where possible.

2.2.4.2   *Multiple imputation of missing data*

Missing values were multiply imputed through chained equations (MICE) with 10 imputations under the missing-at-random assumption.[294] A common set of predictors (Table 2.9) was chosen to predict missing values in each variable, each chosen because it was expected to be a good predictor of most if not all the variables. Predictors also included

the primary outcome of the DESiGN trial and trial phase (pre-randomisation, washout, or outcome).

Table 2.9 - Data variables imputed and their predictors

Variable	Level of imputation	Predictors
Index of Multiple Deprivation (IMD)	Within cluster	Age*, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Maternal age at estimated conception	Within cluster	IMD, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Maternal ethnicity (customised groups)	Within cluster	IMD, age*, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Maternal Height	Across clusters	IMD, age*, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Maternal Weight	Across clusters	IMD, age*, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Parity (ordinal)	Across clusters	IMD, age*, ethnicity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Gestational age at birth	Within cluster	IMD, age*, ethnicity, parity, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, time period
Birthweight	Within cluster	IMD, age*, ethnicity, parity, gestational age at birth, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Onset of labour	Within cluster	IMD, age*, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Mode of birth	Within cluster	IMD, age*, ethnicity, parity, gestational age at birth, birthweight, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Estimated blood loss	Within cluster	IMD, age*, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Number of antenatal appointments after 24 weeks' gestation	Within cluster	IMD, age*, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Number of ultrasound scans after 24 weeks' gestation	Within cluster	IMD, age*, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
Total length of antenatal stay after 24 weeks' gestation (per pregnancy)	Within cluster	IMD, age,* ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss,

		number of ultrasound scans >24 weeks', SGA detected, trial phase
Total length of postnatal stay (per pregnancy)	Within cluster	IMD, age*, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
SGA detected (SGA detected, SGA not detected, no SGA)	Within cluster	IMD, age*, ethnicity, parity, gestational age at birth, birthweight, mode of birth, estimated blood loss, number of ultrasound scans >24 weeks', SGA detected, trial phase
*Maternal age at estimated conception (38 weeks prior to the estimated due date of the baby)		

During the imputation process the primary outcome was captured by a three-category variable:

- baby born SGA and detected by antenatal ultrasound,
- baby born SGA and not detected by ultrasound,
- baby not born SGA.

This three-category variable was imputed like any other during the imputation process, but the primary outcome finally used for analysis was calculated from its imputed components (e.g., maternal weight, neonatal birthweight, ethnicity). This approach was taken because we needed to use the primary outcome as a predictor of other variables in the imputation process, but it was not feasible to passively impute the primary outcome repeatedly and instantly because identification of a baby as SGA by customised centiles at birth, and by antenatal ultrasound, requires calculations of centiles that could only be done manually using an Excel macro (no direct formula was available).

Neonatal admission data were not imputed because neonates without a record of admission to a neonatal care unit were assumed to have not been admitted to neonatal care. Admission to neonatal care is almost universally recorded in Badgernet Neonatal EPRs (from where we sourced the trial data), and it is therefore reasonable to assume that babies without a neonatal record were not admitted to the neonatal unit.

Variables were imputed within cluster wherever possible, as characteristics of women and clinical processes were expected to vary between clusters. Parity, maternal height, and maternal weight were imputed across clusters, because some clusters had high levels of missing data (and for each factor one site had no data) and the rates were not expected to vary widely between clusters. For one site, parity was only available as a binary variable (nulliparous or multiparous), but an ordinal variable was required. Parity was imputed using the common set of predictors and, at this site, predicted parity values of zero for women known to be multiparous were replaced with a pre-specified value of 'one'. Ethnicity

had high rates of missing data at some sites but was imputed within cluster nevertheless because this factor was known to vary strongly across sites.

Summary statistics from the imputed dataset were compared to the equivalent summary statistics from the observed data. Table 2.10 shows the distributions for variables that are required for the calculation of the primary outcome, comparing observed data and imputed data by each of the intervention arms. Notably, the proportion of women coded as 'white' ethnicity increased during the pre-randomisation phase in the intervention arm; this was expected because one cluster randomised to the intervention, at which most of the pregnant women are white, had high levels of missing data on ethnicity during this period, this was corrected by the time the trial comparison phase data were collected. The proportion of multiparous women increased in the imputed data for the control arm of the trial during both time periods, this was also expected because one control arm cluster only provided information on nulliparity (converted to parity=0) or multiparity, as a binary value. Parity was therefore only imputed for multiparous women because we knew they had had a baby previously, but not how many.

Table 2.10 - Comparison of summary statistics between trial phases and intervention allocation

	Pre-randomisation phase				Trial outcome phase			
	Control sites		Intervention sites		Control sites		Intervention sites	
	Observed data	Imputed dataset	Observed data	Imputed dataset	Observed data	Imputed dataset	Observed data	Imputed dataset
Maternal height (median (IQR))	1.64 (1.60, 1.68)	1.64 (1.60, 1.69)	1.63 (1.58, 1.68)	1.63 (1.58, 1.68)	1.64 (1.60, 1.69)	1.64 (1.60, 1.69)	1.64 (1.60, 1.69)	1.63 (1.59, 1.68)
Maternal weight (median (IQR))	66.0 (58.8, 76.0)	66.0 (58.5, 76.0)	64.5 (57.0, 74.0)	64.7 (57.3, 74.4)	67.2 (59.9, 78.0)	67.0 (59.5, 77.9)	67.2 (59.9, 78.0)	65.4 (58.0, 76.0)
Ethnicity*: White (%)	62.4%	62.8%	46.9%	50.2%	62.4%	62.7%	50.2%	50.6%
Black (%)	16.6%	16.2%	12.4%	11.8%	15.4%	15.1%	10.8%	10.9%
Asian (%)	13.4%	13.3%	27.1%	24.8%	13.5%	13.5%	25.2%	24.6%
Mixed (%)	2.2%	2.1%	1.6%	1.6%	2.6%	2.6%	1.3%	1.3%
Other (%)	5.4%	5.5%	12.1%	11.6%	6.1%	6.1%	12.5%	12.6%
Parity (% multiparous)	49.4%	53.7%	46.8%	46.5%	48.8%	52.5%	48.8%	51.6%
Neonatal birthweight in grams (median (IQR))	3,380 (3,050, 3,700)	3,380 (3,050, 3,700)	3,340 (3,010, 3,660)	3,340 (3,010, 3,660)	3,360 (3,040, 3,670)	3,360 (3,035, 3,670)	3,320 (3,000, 3,645)	3,320 (2,996, 3,645)
Neonatal gestational age at birth in weeks (median (IQR))	40 (39, 41)	40 (39, 41)	40 (39, 41)	40 (39, 41)	40 (39, 41)	40 (39, 41)	40 (39, 41)	40 (39, 41)

*Five categories of ethnicity presented here for brevity, but ethnicity imputed according to categories used by the fetal/neonatal weight customised centile calculator e.g. British European, Central African, Chinese, Mixed Asian-European.

124

2.2.5   The final dataset

The final data resource comprised data on 209,314 pregnancies of which 4,385 were multiple pregnancies. It was not possible to obtain one data type (hospital administrative data) at one site for either time period, hence the final data resource was derived from 102 of a potential 104 datasets (four datasets across two time periods at 13 research sites). Following exclusions there were 65,959 women and their singleton babies included during the pre-randomisation trial phase, 106,061 during the washout phase and 29,189 during the outcome phase.

2.2.6   Statistical Analyses

Statistical techniques used in this thesis are detailed in the sections below. Quantitative analyses were all conducted using Stata software version 16 (StataCorp LP, College Station, Texas).

2.2.6.1   *Summary statistics*

For continuous numerical data, mean and standard deviation (SD) were used to summarise parametric data and median and interquartile range (IQR) were used to summarise non-parametric data. An exception to this is hospital activity data in economic analyses, where use of mean and standard deviation is standard in the calculation of individual and overall costs, even for non-parametric data.

For categorical or binary data, number and percentage were used to describe frequencies and proportions.

2.2.6.2   *Calculation of fetal and neonatal weight centiles*

Customised centiles, as used in GAP, are suitable for use for both estimated fetal weights and birthweights. These were calculated by exporting maternal (height, weight, ethnicity and parity) and fetal (weight and gestational age) parameters into the standardised calculator issued annually by the Perinatal Institute. Version 8.0.4 of the UK Bulk Centile Calculator was used to calculate customised weight centiles for both fetal and neonatal weights in the DESiGN trial.

There are no population weight centiles which are suitable for application to both fetal and neonatal weights, this is because existing centiles have only been calculated from one, or the other population. For reasons discussion in section 1.1.5.2, population birthweight centiles at preterm gestations, cannot be automatically translated to estimated fetal weights. To calculate estimated fetal weight population centiles from ultrasound scans, the Hadlock fetal weight centile formula was used (section 1.1.7.1).[62] To calculate birthweight

population centiles, the 'zanthro' user-made Stata program was used, applying UK 1990 birthweight references.[63]

2.2.6.3 *Comparative statistics*

The women and babies included in the DESiGN trial were treated as two different population types during this thesis, each requiring a different statistical approach:

- Population recruited to a randomised cluster control trial; outcomes of pregnancies in this population type were analysed using cluster summary statistics.
- Cohort population, treated as such because no statistically significant differences were identified in the rates of the primary, or most secondary outcomes of the trial; predictors and outcomes in this population type were analysed using classical univariate and multivariate methods (with adjustment for cluster site included in the latter methods).

2.2.6.3.1 Cluster summary statistics

To acknowledge the clustered nature of the data, whereby the characteristics and outcomes of women and babies included in the study are more likely to be alike to those of women and babies from the same cluster, as opposed to women and babies from another cluster, clustering has been considered.

Clustering in trials is usually accounted for by including individual patient characteristics and cluster characteristics in random-effects regression models, however this is a method best suited to trials with a large number of clusters. In the DESiGN trial and the secondary analyses presented in this thesis, a cluster summary approach was preferred because of the moderate number of clusters recruited (n=13). The ANCOVA approach allows individual predictors to still be accounted for, using steps as recommended by Hayes and Moulton.[295] ANCOVA is a statistical technique whereby an analysis of variance is conducted to examine the difference in means between groups, but with adjustment for the effect of other exposure variables (covariates).[296]

For the DESiGN trial and analyses reported in this thesis, ANCOVA was also used to estimate the mean difference between groups for which the data were not parametrically distributed. Whilst the parameters in the sample are not normally distributed, the normal distribution is a good approximation of the sampling distribution of the studied parameter, regardless of its underlying distribution, provided that the sample is large (central limit theorem) therefore suggesting that it is statistically reasonable to use parametric comparative statistics for non-parametric data.[296]

This approach was conducted in two stages:

1. The cluster summary values for a given outcome (e.g., rate of SGA detection) during the pre-randomisation and trial outcome periods were adjusted to account for individual women's ethnicity, age and parity in the cluster, using a logistic regression model. This generated adjusted cluster summaries (one summary value for each outcome, within each cluster, at both time points) to be used in step 2.

2. The adjusted cluster summary values from the outcome period where then compared in the ANCOVA (linear regression) model, which included the intervention arm as the exposure of interest and accounted for the randomisation strata, the baseline adjusted cluster summary value and cluster summary values of ethnicity, parity & age.

   a. Three distinct randomisation strata were considered: large cluster randomised in first group, small cluster randomised in first group, cluster randomised in second or third group.

The ANCOVA model estimates linear regression lines for the adjusted pre-randomisation and outcome period values, comparing these for the standard care and intervention arms of the trial. The model computes a coefficient that represents the mean adjusted difference and standard error for the studied outcome, either as a difference in proportions (e.g., rate of detection of SGA) or means (e.g., mean birthweight). These are then presented with a 95% confidence interval.

Where imputed data were used, the steps described were conducted on each of the ten imputed datasets and then combined using Rubin's rules to produce a single treatment effect estimate and 95% confidence interval.[297]

2.2.6.3.2 Univariate comparisons

Comparison of categorical data between groups were summarised using odds ratios and chi-squared test for non-paired parametric data, or univariate logistic regression. Comparisons between continuously reported data were summarised using student's t-test for non-paired parametric data.

2.2.6.3.3 Multivariable analysis and adjustment for confounding factors

For adjusted analyses, dependent categorical variables were converted to binary and studied using logistic regression. Continuously reported variables were assessed using linear regression. All regression models were conducted using each imputed dataset and

combined with the principles of Rubin's rule through the Stata *mi estimate* command.[297] The following steps were used to assess for confounding and effect modification:

1. An a priori list of possible confounding variables was determined.
2. Each potential confounder was introduced one by one into the regression model to determine whether it affected the estimate of association between the dependent and independent variables. Potential confounders which affected the effect estimate by more than 10% were considered for further analysis.
   a. A confounder must be associated with both the independent and dependent variables but should not sit on the causative pathway. To confirm if a variable was a likely confounder, a univariate comparison was conducted between the potential confounder and the independent variable, followed by the dependent variable.
3. The effect of association between the independent and dependent variable was then stratified by levels of the potential confounder to determine whether the association effect differed by stratum. A test for heterogeneity was applied. Where strata were heterogenous, the data were stratified. Where strata were homogenous, the confounding effect was combined.
4. When all risk factors and a priori variables had been assessed for confounding, they were combined in the model to determine the independent effect of each on the dependent variable.
5. Finally, likelihood ratio tests were conducted to compare models which included a small number of interaction parameters, for variables with heterogenous effects between strata.

2.2.6.4 *Significance level of tests*

All confidence intervals are at the 95% significance level and two-sided. Statistical tests have been interpreted using a two-sided *p* value of 0.05.

2.3 PROCESS EVALUATION

The DESiGN trial was planned as a hybrid effectiveness-implementation trial.[285] Medical Research Council UK (MRC) guidance on developing and evaluating complex interventions advises that process evaluation is key to understanding effectiveness in everyday practice.[298] Process evaluation can be used to assess fidelity of implementation, generate causal hypotheses, and identify factors of local or national context which are associated with differences in outcomes.[299] Evaluation of implementation is necessary to prevent type III error, the dismissal of an intervention because of failure to implement it as intended.[300]

The process evaluation conducted during the DESiGN trial aimed to understand the functioning of the intervention by examining implementation, mechanisms of impact, and contextual factors. Components of this process evaluation are reported in this thesis.

2.3.1 Study design

The process evaluation drew on the MRC framework for trials of complex interventions, Steckler and Linnan's framework for process evaluation of public health interventions and research and the trial logic model.[298,301] The logic model was developed during the design of the trial and assisted in the hypotheses around the mechanisms by which GAP is expected to achieve impact. It depicts the context, resources, implementation strategies, moderating factors and hypothesised outcomes of the GAP intervention, informing data items for the evaluation (Figure 2.7).

Figure 2.7 - Logic model for the GAP intervention

Context	Resources	Intervention	Implementation strategies	Moderating factors	Outcomes
National concern:	Trial funding	**GAPGROW components:**	Multidisciplinary local GAP team identified	UK/London political context	(see protocol)
• Poor AN detection of SGA	Professional guidance to inform improved SGA detection	3m Retrospective audit	GAP champion identified	*Acceptability*	**Clinical primary outcome:**
• UK high stillbirth rate compared to other high income countries	Clinical support for trial from London SCN	SGA risk assessment at booking	Staff training by PI & cascade training (75% coverage)	Professional norms	AN USS detection of SGA by customised and population standards
• London regional stillbirth higher than other areas of the UK	Access to Perinatal Institute implementation expertise and training	Low/high risk care pathway	Retrospective notes audit	Fit with established practice	Clinical secondary outcomes & neonatal and maternal outcomes
Scientific uncertainty:	PPI support (Tommy's charity and JLA)	Customised chart at first USS	Revised and agree Trust protocols	Staff engagement	Implementation process outcomes (Context, acceptability, fidelity, reach, dose)
• GAPGROW protocol is a complex intervention		Standardised SFH measurement plotted on CGC from 26 weeks	From 'Go live':	Trust QI resources and openness to change	Health economics outcomes:
• Good quality observational evidence supports its use		Referral to USS for suspected SGA, urgent referral for suspected growth concerns	AN risk assessment at booking, customised chart, plotting on CGC from 26/40, USS referral protocol, GROW obstetric management, cust centile at birth, missed case audit	*Fidelity*	(Costs of USS, ADU use, IOL, CS, LOS, NICU/SCBU admissions)
• Uncertainty about effectiveness and costs/benefits of intervention		GROW obstetric management protocol		Staff training	
				Support from Perinatal Institute	
				Funding for implementation leads within trusts	
				Scanning resources	

Abbreviations ADU: Antenatal Day Unit, AN: Antenatal, CGC: Customised growth chart, CS: Caesarean section, GAP: Growth Assessment Protocol, GROW: Gestation-related optimal weight chart, IOL: Induction of Labour, LOS: Length of Stay, m: month, NICU: Neonatal Intensive Care Unit, PI: Perinatal Institute, PPI: Patient and public involvement, QI: Quality Improvement, SCBU: Special Care Baby Unit, SCN: Strategic Clinical Network, SFH: symphyseal fundal height, SGA: Small for gestational age, USS: Ultrasound Scan.

2.3.2 Objectives

The objectives of the process evaluation of the DESiGN trial were to:

1. Assess the strength of implementation at each of the implementing clusters.
2. Consider the context of implementation of GAP, and how it affected the implementation process.
3. Describe the process of GAP implementation from the perspectives of frontline staff and lead clinicians.
4. Identify the barriers and facilitators to GAP implementation described by frontline staff and lead clinicians.
5. Determine the views of frontline staff and lead clinicians regarding the acceptability and feasibility of GAP implementation. *Objectives 5 is not reported in this thesis.*

2.3.2.1 *Implementation strength*

Implementation strength is a measure that incorporates the quantity and quality of an intervention protocol which is carried out in practice; its measurement intends to capture the amount of the protocol which is delivered.[302] In a review of implementation literature, Schellenberg et al (2012) aimed to synthesise existing literature on measurements of implementation strength. The authors were unable to identify a widely accepted definition, measurement technique or reporting method for implementation strength.

For objective 1, strength of implementation was defined using Steckler and Linnan's implementation outcomes: fidelity, reach, dose delivered, and dose received (Table 2.11). The 'implementation' dimension of this framework is comparable to 'implementation strength', as defined above. Whilst this framework also includes outcomes on context and recruitment, recruitment was not considered relevant to process evaluation of this intervention, since individual participants were not approached nor attracted. Context was measured for this process evaluation but was interpreted to inform mechanisms behind strength of implementation and so was not included as a measure of implementation strength.

Table 2.11 - Implementation dimensions as defined by Steckler and Linnan (2002)

Dimension	Description
Context	Aspects of the social, political or and economic environment that may influence implementation
Reach	The proportion of intended target audience that participates in an intervention. Often measured by attendance.
Dose delivered	The number or number of intended units of each intervention or component delivered or provided.
Dose received	The extent to which participants actively engage with, or use, recommended materials or resources.
Fidelity	The extent to which the intervention is delivered as conceived
Implementation	A composite score that indicates the extent to which the intervention has been implemented and received by the intended audience.
Recruitment	Procedures used to approach and attract participants.

2.3.2.2 *Context of implementation*

Plans to assess the context of implementation for objective 2 of the process evaluation were informed by the Context and Implementation of Complex Interventions (CICI) analytical framework.[303] The authors of this framework define context as:

> *'A set of characteristics and circumstances that consist of active and unique factors, within which the implementation is embedded. As such, context is not a backdrop for implementation, but interacts, influences, modifies and facilitates or constrains the intervention and its implementation. Context is usually considered in relation to an intervention, with which it actively interacts. It is an overarching concept, comprising not only a physical location but also roles, interactions and relationships at multiple levels.'* [303]

In the CICI framework, context is separated into seven domains: geographical, epidemiological, socio-cultural, socio-economic, ethical, legal and political (Figure 2.8) and each domain is interpreted on a macro (regional/national), meso (organisational) or micro (individual person or team) level. These domains were used to analyse the context of implementation for the DESiGN trial.

Figure 2.8 - The Context and Implementation of Complex Interventions (CICI) framework
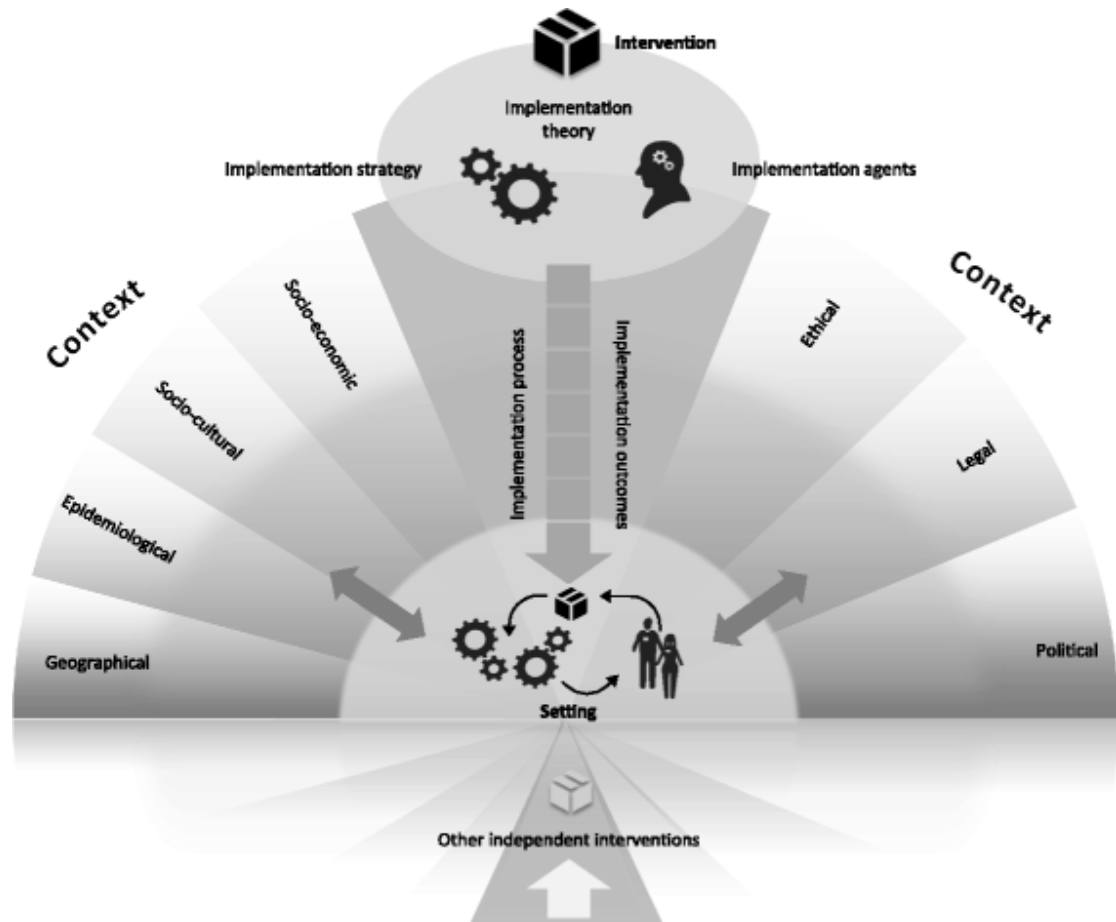
2.3.2.3 *Implementation processes*

The CICI framework also defines distinct processes of implementation. For objective 3 of the DESiGN analysis, we categorised implementation processes at each site into CICI domains as described in Table 2.12.

.

Table 2.12 - GAP processes as applied to CICI implementation process domains

	GAP processes
Exploration	As defined in CICI framework: *'exploration of organizational needs, intervention-organizational fit as well as capacity and readiness assessment in a given setting'*[303]
Decision to adopt	Considerations that led to the adoption of GAP, or recruitment into the DESiGN trial.
Planning and preparation	Baseline audit of SGA detection (6 months of births prior to GAP implementation) Incorporate GAP into local guidelines Identification of a GAP team Train the trainers – Perinatal Institute training session
Initial implementation	Cascade training to >75% staff from each major staff group >75% staff to complete GAP e-learning Prepare to 'Go Live'
Full implementation	Risk assessment of all pregnant women prior to 24 weeks gestation Production of GROW charts for women Serial scans for high risk women (3-weekly from 26-28 weeks) Serial fundal height measurement for low risk women (3-weekly from 26-28 weeks) EFW and fundal height plotted on GROW chart Growth scans within 3 days if triggered by fundal height plot on GROW chart Management of SGA as per RCOG guidance
Evaluation and reflection	Missed case audit
Sustainment	Annual face-to-face and e-learning training for >75% staff members from each major group.

2.3.2.4 *Other implementation outcomes*

Objective 4 was planned to assess barriers and facilitators to implementation. Objective 5 was planned to assess acceptability and feasibility. These outcomes are all defined in Table 2.13.

Table 2.13 - Definition in implementation outcomes for objectives 4 and 5.

Outcome	Definition
Barrier	Unanticipated negative factors that influence an organisation's efforts to implement change.[304]
Facilitator	Unanticipated positive factors that influence an organisation's efforts to implement change.[304]
Acceptability	The perception among implementation stakeholders that a given treatment, service, practice, or innovation is agreeable, palatable, or satisfactory.[305]
Feasibility	The extent to which a new treatment, or an innovation, can be successfully used or carried out within a given agency or setting.[306]

2.3.3 Data collection

The process evaluation of the DESiGN trial was a mixed-methods study. Different types of data were collected for each type of process or outcome. These are summarised in Table 2.14. Further details are described in Chapter 3.

Table 2.14 - Sources of data collected for each implementation outcome

Implementation outcome	Application to implementation of GAP	Data source
Context	Macro (regional/national), meso (organisational) and micro (individual team or person) context of GAP implementation.	Semi-structured interviews with lead clinicians and frontline staff.
Fidelity	Adherence to the Perinatal Institute GAP training requirements.	Staff training records from the Perinatal Institute.
	Concordance with GAP guidelines.	Cluster guidelines for the screening and management of SGA fetuses.
	Proportion of women correctly risk stratified (according to GAP strategies)	Notes review of the maternity records
Reach	Proportion of women with a GAP-GROW chart in the notes.	
Dose delivered and received	Proportion of low-risk women* who had at least the minimum expected fundal height measurements performed and plotted on the chart	
	Proportion of low-risk women* referred for growth scan when indicated.	
	Proportion of high-risk women* who had at least the minimum expected growth scans performed and plotted on the chart	
Acceptability	Acceptability of GAP implementation from the perspectives of clinicians.	Semi-structured interviews with lead clinicians and frontline staff.
Feasibility	The degree to which GAP implementation is feasible, from the perspectives of interview participants.	
Barriers	Unanticipated influences which negatively impacted on GAP implementation.	
Facilitators	Unanticipated influences which positively impacted on GAP implementation.	

2.3.4 Analyses of the process evaluation

The main methods of analysis in the process evaluation can broadly be categorised into quantitative methods used to analyse the maternity notes review, and qualitative methods used to analyse the semi-structured interviews.

2.3.4.1 *Quantitative analysis*

Data extracted from paper maternity notes were analysed using summary statistics as described in section 2.2.6.1.

2.3.4.2 *Qualitative analysis*

Interview findings were synthesized using framework analysis. The framework method is a qualitative analysis technique whereby data are summarised into an existing framework of codes and categories, which has been designed to help researchers structure their data and therefore to support analysis. It is useful when multiple researchers are working on the same project, and also when there is already a substantial body of implementation theory which supports the analysis.[307]

The CICI framework (Figure 2.8) was chosen because it is a recent addition to the implementation literature and designed to be used for process evaluation of complex interventions.[303] This framework builds on and incorporates knowledge from previous frameworks and provides an opportunity to assess context at different levels (micro, meso and macro). It also incorporates extensive evaluation of implementation context, which is often not evaluated in process evaluations, but is vital in understanding both the quality of implementation and its generalisability. When planning the analysis, there was no published evidence that the framework had been tested in a trial setting; we therefore also planned to comment on the usefulness of the framework for this purpose.

2.4 HEALTH ECONOMIC EVALUATION

2.4.1 Study design

The economic evaluation of the DESiGN trial was planned as a cost-effectiveness analysis. A cost-effectiveness analysis is one which compares both the costs and effects of alternative strategies, such as GAP and standard care. The incremental changes in cost and effect between the alternative strategies are then combined into a single metric (incremental cost-effectiveness ratio), such as the cost incurred per stillbirth avoided, or cost incurred per additional case of SGA detected.

2.4.2 Clinical effectiveness

Antenatal detection of SGA neonates was as defined for the primary outcome of the DESiGN trial (2.2.1.1).

2.4.3 Calculation of costs

2.4.3.1 *Economic perspective*

Economic (cost) perspective refers to the level at which the costs are assessed, the most common examples are patients and families, a single healthcare provider, the health and/or social services, and wider society.[308] This economic evaluation took the cost perspective of the average UK hospital provider rather than that of the NHS, which is generally the preferred perspective in national Health Technology Appraisals.[309] This is because maternity care within the NHS is funded by regional commissioners using bundled payments for antenatal, intrapartum, postnatal and neonatal care.[310] As discussed in chapter 4, bundled payments represent the average price of similar care provided by the hospital, their use in cost-effectiveness analyses is limited by not being sensitive to small changes in resource use, e.g., need for additional fetal ultrasound surveillance.[311] This was key to the estimation of incremental costs for this intervention.

2.4.3.2 *Resource use*

To calculate the cost of implementing GAP, and the cost of continuing standard care, different types of activity had to be considered:

- Activities relevant to the implementation of GAP e.g., costs of training staff. These are only relevant to the cluster sites allocated to implement GAP, since other cluster sites are simply continuing normal care.
- Activities which are hypothesised to potentially be affected following implementation of a new intervention, this includes activities in the patient pathway which follow initial exposure to the intervention.

To calculate the cost of implementing GAP, data were therefore required on the following types of resource use:

- Number of staff members who completed training, and the time taken to complete the training.
- Additional time allocated to assess women at the time of antenatal booking for the risk of SGA, and time allocated to generate customised GROW charts.
- Clinical resource use during antenatal, intrapartum, and postnatal care for the woman, which has the potential to be impacted by GAP, or by outcomes when screening for SGA (Table 2.15).
- Clinical resource use during neonatal care for the baby that has the potential to be impacted by GAP, or by outcomes when screening for SGA (Table 2.15).

Table 2.15 - Activities within the maternity and neonatal care pathway which were hypothesised to potentially vary with implementation of GAP

Antenatal care (after 24 weeks' gestation)	Intrapartum care	Postnatal care	Neonatal care
• Antenatal appointments • Attendance to day assessment unit/triage • Antenatal inpatient admission • Ultrasound scan for fetal growth • Ultrasound scan for fetal growth	• Induction of labour • Epidural • Unassisted vaginal birth • Assisted vaginal birth • Elective Caesarean section • Emergency Caesarean section • Repair 3rd/4th degree tear • Treatment of postpartum haemorrhage (500mL-1500mL) • Treatment of major obstetric haemorrhage (>1500mL)	• Maternal stay in postnatal ward (with/without baby)	• Admission to neonatal intensive care unit • Admission to neonatal high dependency unit • Admission to special care baby unit • Neonatal transitional care

These data were collected, and the data managed using the same methods of the clinical outcomes study (sections 2.2.2-2.2.4).

2.4.3.3 *Valuation of resource use*

Valuation of resource use targeted the list of maternal and neonatal care activities listed in Table 2.15. The valuation was informed by a systematic review of resource unit costs quoted in published economic evaluations of interventions, in maternity and economic evaluations published in national guidelines, or health technology assessments during the last decade (reported in Chapter 4), and review of the available costs published

by the Department of Health as part of the national maternity tariff from 2015-16 and 2017-18.[284,310]

For each item of activity, the most recent and nationally-appliable cost was identified, which was sufficiently granular to cost the single activity item (rather than a bundled cost for a composite of activity items). Cost sources were prioritised using the strategy detailed in Figure 2.9. Costs were then inflated where appropriate to 2018-19 prices, using the Department of Health's Pay & Price Series for financial years 2008/09 - 2015/16 and the NHS Improvement Economic Assumptions for years 2016/17 to 2018/19. [312,313] Costs were not discounted because all costs were expected to occur within a single year.

Figure 2.9 - Approach to sourcing unit costs for maternal and neonatal care activities

Preference 1: National Reference costs within the last 5 years.

If not available

Preference 2: Costs sourced from **economic evaluations conducted for or used by National guidelines**.

Neither of above available

Preference 3: Unit costs taken from **published economic evaluations maternity interventions**, with preference given to those conducted in the UK within the last 5 years.
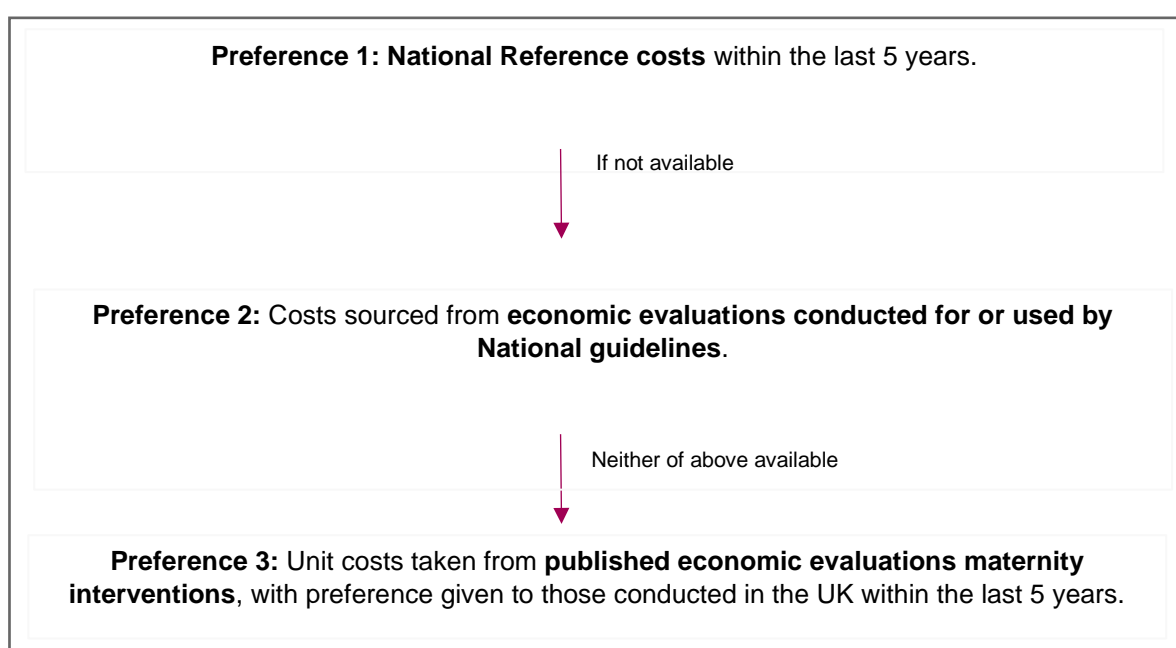
2.4.4 Economic analyses

The following methods were used to describe and assess cost-effectiveness.

2.4.4.1 *Summary statistics*

The proportion of women or babies requiring each type of resource, the intensity of resource use for each woman or baby within a trial arm and phase (pre-randomisation and outcome comparison) and composite costs for implementation, antenatal, intrapartum, postnatal and neonatal care were summarised using summary statistics (section 2.2.6.1).

2.4.4.2 *Comparative statistics*

Resource use was compared between trial arms using unadjusted univariate comparisons and the cluster summary statistic approaches described in section 2.2.6.3.

Costs were also compared using the unadjusted univariate comparisons, but generalised linear models were used rather than the cluster summary statistic approach for adjusted comparisons of antenatal, intrapartum, postnatal and neonatal costs. The generalised linear model is a type of linear regression model in which a transformation of the outcome variable is modelled rather than the outcome itself, this is another method that enables modelling of non-parametrically distributed outcomes.[296]

2.4.4.3 *Cost-effectiveness analysis*

A cost-effectiveness analysis compared the costs of implementing GAP to the clinical effectiveness of GAP in the antenatal detection of SGA and reported using an incremental cost effectiveness ratio (ICER) – cost per additional case of SGA detected.

The cost-effectiveness analysis was conducted using a Monte Carlo simulation model in Microsoft Excel. A Monte Carlo model is a mathematical technique that is used to estimate the probability of an outcome by taking into account the probability distributions of the predictor variables. It recalculates the outcome repeatedly, each time using a different set of random numbers from each of the independent parameters' probability distributions.

2.5 REPORTING GUIDELINES

Chapters of this thesis have been written using the guidelines specific to the study design:

- Standards for Reporting Implementation Studies (STaRI) statement for reporting implementation studies (Chapter 3),[314]
- Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for reporting systematic reviews (Chapter 4),[315]
- Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement for reporting economic evaluations (Chapter 5).[316]
- The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (Chapter 6),[317]
- Consolidated Standards of Reporting Trials (CONSORT) cluster extension guideline for reporting results of randomised control trials (Chapter 7),[318,319]

2.6 FINDINGS OF THE DESiGN TRIAL

As detailed in section 1.7, the aims of this thesis include secondary and subgroup analyses of the DESiGN trial, but do not cover the primary trial clinical outcome results. The primary findings of the trial have been submitted for publication.[320] Results that are important to the chapters of this thesis are summarised below for context.

2.6.1 Recruitment

Of 15 maternity units approached for the DESiGN trial, 13 were recruited and randomly allocated to the intervention or standard care. Following reported financial pressures of two clusters allocated to the intervention arm, these clusters felt unable to implement the trial intervention for which there was no financial support and withdrew shortly after randomisation. This left five clusters implementing the intervention and six clusters randomised to the control arm of the trial. The withdrawing clusters agreed to continue in the trial, providing data on the women giving birth in their maternity units so that an intention-to-treat statistical analysis could still be supported. However, since the practice in these clusters sites is not expected to be informative of the clinical effectiveness of GAP, the primary analysis for all trial outcomes was conducted using a modified intention to treat principle (i.e., these two cluster sites were excluded), the full intention to treat analysis was published as a sensitivity analysis.

2.6.2 Characteristics of women and babies included in the trial analyses

Data from 24,906 women and babies were included in analyses for the outcome phase, using adjustments from 55,950 women and babies included in the pre-randomisation (baseline) phase. The consort diagram is included in Figure 2.10.

Figure 2.10 - CONSORT diagram detailing cluster and individual inclusion and exclusion from the DESiGN trial
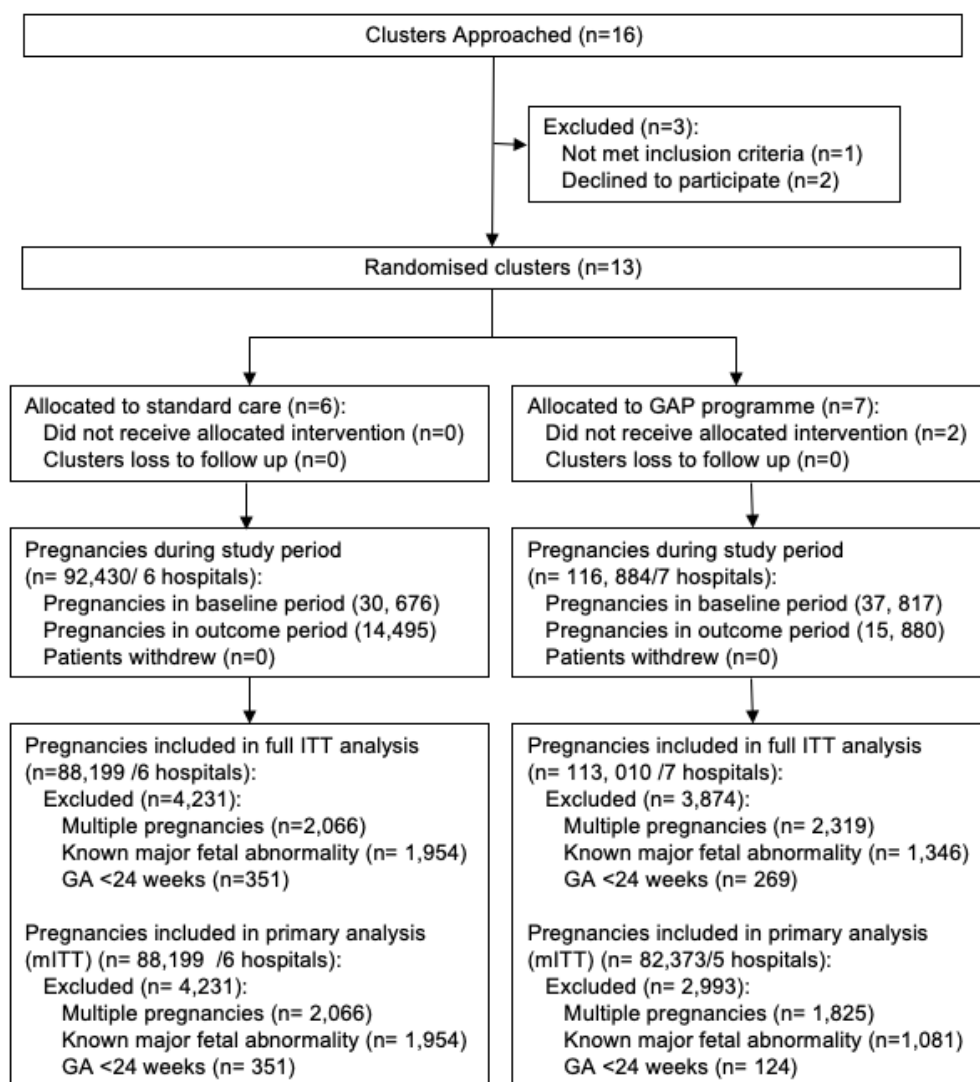
*CONSORT diagram taken directly from primary DESiGN trial report (unpublished at the time of writing).[320]

The characteristics of women included during the pre-randomisation and outcome phases within each arm of the trial are summarised in Table 2.16 and Table 2.17.

Table 2.16 – Imputed characteristics of mothers and babies included in the DESiGN trial, presented by trial arm for the pre-randomisation (baseline) and outcome phases

		Pre-randomisation phase		Outcome phase	
		Standard Care (n=29,404)	Intervention (n=26,546)	Standard Care (n=13,810)	Intervention (n=11,096)
Imputed data					
Age at conception, years, median (IQR)		31.6 (27.5, 35.2)	31.5 (27.6, 35.2)	32.0 (27.9, 35.4)	31.8 (27.9, 35.5)
Ethnicity, n (%)	White	62.8%	55.9%	62.7%	56.2%
	Black	16.2%	12.7%	15.1%	12.6%
	Asian	13.3%	19.4%	13.5%	20.3%
	Mixed	2.1%	1.9%	2.6%	1.6%
	Other	5.5%	10.1%	6.1%	9.2%
Index of Multiple Deprivation Quintiles, n (%)	1 (Least deprived)	17.4%	7.6%	16.5%	7.5%
	2	12.5%	10.8%	12.7%	10.6%
	3	16.1%	23.2%	16.6%	23.6%
	4	28.5%	34.7%	28.7%	35.4%
	5 (Most deprived)	25.4%	23.7%	25.5%	22.9%
Maternal Height, median (IQR)		1.64 (1.60, 1.69)	1.64 (1.59, 1.68)	1.64 (1.60, 1.69)	1.64 (1.60, 1.68)
Maternal Weight, median (IQR)		66.0 (58.5, 76.0)	64.6 (57.4, 74.0)	67.0 (59.5, 77.9)	65.4 (58.0, 76.0)
Body Mass Index Categories, n (%)	<18.5	3.9%	4.1%	3.4%	3.4%
	18.5-24.9	50.1%	53.9%	47.2%	51.6%
	25.0-29.9	28.0%	26.3%	29.5%	27.2%
	30.0-34.9	11.9%	10.5%	13.1%	11.3%
	35.0-39.9	4.2%	3.5%	4.6%	4.4%
	≥40	2.0%	1.7%	2.2%	2.1%
Parity, n (%)	Nulliparous	46.4%	59.0%	47.5%	51.6%
	1	33.8%	26.3%	34.0%	30.3%
	2	11.6%	9.4%	11.0%	11.1%
	3	4.6%	3.2%	4.2%	4.2%
	≥4	3.7%	2.2%	3.3%	2.9%

Table 2.17 - Non-imputed characteristics of mothers and babies included in the DESiGN trial, presented by trial arm for the pre-randomisation (baseline) and outcome phases

	Pre-randomisation phase		Outcome phase	
	Standard Care (n=29,404)	Intervention (n=26,546)	Standard Care (n=13,810)	Intervention (n=11,096)
Non-imputed data				
Smoking in pregnancy, n (%)	5.8% (1,646/28,252)	5.2% (1,090/21,149)	5.2% (698/13,466)	5.7% (569/10,010)
Missing smoking	*1,152*	*5,397*	*344*	*1,086*
Pre-existing comorbidities, n (%)				
Hypertension	2.0% (379/19,324)	1.5% (303/20,162)	1.3% (119/9,276)	1.4% (130/9,189)
Missing hypertension	*10,080*	*6,384*	*4,534*	*1,907*
Diabetes	0.9% (162/18,511)	2.5% (497/20,162)	1.0% (94/9,153)	3.4% (299/8,862)
Missing diabetes	*10,893*	*6,384*	*4,657*	*2,234*
Systemic Lupus Erythematous (SLE)	0.2% (35/19,344)	0.03% (7/20,154)	0.2% (16/9,294)	0.02% (2/8,521)
Missing (SLE)	*10,060*	*6,392*	*4,516*	*2,575*
Antiphospholipid Syndrome (APS)	0.05% (9/19,285)	0.00% (0/11,629)	0.05% (5/9,294)	0.00% (0/4,904)
Missing APS	*10,119*	*14,917*	*4,516*	*6,192*
Pregnancy comorbidities, n (%)				
Gestational diabetes (GDM)	3.5% (833/23,957)	6.2% (1,242/20,087)	6.3% (713/11,416)	8.1% (707/8,699)
Missing GDM	*5,447*	*6,459*	*2,394*	*2,397*
Gestational hypertension (Gest HT)	1.7% (308/18,506)	2.6% (401/15,215)	1.2% (136/11,418)	3.4% (219/6,498)
Missing Gest HT	*10,898*	*11,331*	*2,392*	*4,598*
Pre-eclampsia	0.7% (132/18,504)	1.8% (368/20,150)	1.2% (100/8,663)	2.4% (216/9,185)
Missing Pre-eclampsia	*10,900*	*6,396*	*5,147*	*1,911*
Eclampsia	0.3% (54/18,504)	0.1% (10/11,372)	0.3% (26/8,663)	0.1% (4/4,827)
Missing Eclampsia	*10,900*	*15,174*	*5,147*	*6,269*
Infant sex, male, n (%)	51.3% (15,086/29,397)	51.3% (13,586/26,494)	51.1% (7,053/13,798)	50.7% (5,590/11,023)
Missing Infant sex	*7*	*52*	*12*	*73*

2.6.3 Primary and secondary clinical outcomes

The DESiGN trial found no difference in the rate of antenatal detection of SGA between the intervention and standard care arms of the trial (25.9% vs 27.7%; adjusted mean difference 2·4%, 95% CI -6·1% to 10·8%; p=0·58). There were also no differences in 24 out of 26 pre-specified maternal and perinatal clinical outcomes. The only differences identified were for stillbirth, for which there was evidence of a greater reduction in the clusters implementing GAP than those allocated to standard care (0·31% vs 0·36%, adjusted difference -0·07%, 95% CI -0·14% to -0·01%; p=0·03) and similarly for perinatal mortality (0·37% vs 0·41%, adjusted difference -0·09%, 95% CI -0·17% to -0·004%). Whilst these may be chance findings which should be interpreted with caution, it was interesting to note that the only differences were found in the clinical outcomes that the intervention is designed to affect.

2.6.3.1 *Subgroup and sensitivity analyses*

In a pre-specified subgroup analysis in which only infants born SGA were included, infants in implementing clusters were born 2 days earlier (gestational age at birth 38.6 weeks vs 38.8, adjusted difference -0·3, 95% CI -0·5 to-0·1), had a lower mean birthweight (2,436g vs 2,482g, adjusted difference -58g, 95% CI -99g to -18g) and were less likely to be stillborn (1.39% vs 2.19%, adjusted difference -0·76%, 95% CI -1·50% to -0·03%) than infants born in clusters continuing standard care.

In a sensitivity analysis in which the population was restricted to women who had received an ultrasound scan for fetal anomalies (defined as any scan between 18- and 24-weeks' gestation) in the cluster site in which they gave birth, there were no differences in any of the primary or secondary clinical outcomes between trial arms.

3 ASSESSING FACTORS ASSOCIATED WITH THE STRENGTH OF IMPLEMENTATION OF THE GROWTH ASSESSMENT PROTOCOL, IN THE CONTEXT OF THE DESiGN TRIAL

3.1 INTRODUCTION

The Growth Assessment Protocol has been implemented in over 70% of maternity units in the UK,[270] following observational evidence suggesting that its implementation may be associated with an improvement in the rate of detection of SGA, and a reduction in the rate of stillbirth.[321] At the time of planning and conducting the DESiGN trial, implementation of GAP in clinical practice had not been independently studied. In 2020, Gardosi et al published further observational evidence supporting implementation of GAP; this included a comparison of NHS Trusts/Boards who had 'completely' or 'partially' implemented the protocol.[281] Those who had completely implemented the protocol (defined as reporting the birthweight centile and outcome of at least 75% of babies born) were found to have a significantly sharper decline in the rate of stillbirth than those with partial (lower levels of reporting) or no implementation. Whilst reporting of SGA detection rates has also been shown elsewhere to improve the rate of detection,[322] it represents only one component of this complex intervention and does not explain the mechanism by which GAP impacts on SGA detection and stillbirth.

The DESiGN trial was planned as a hybrid effectiveness-implementation trial. Such trials and associated frameworks or guidance have been driven by the need to understand if lack of effect is attributable to the intervention itself or to low strength of implementation.[285] In the context of the DESiGN trial finding that GAP was no more clinically effective than standard care, the results of a process evaluation are essential to understand the possible causes of no effect. Process evaluation can also aid in understanding how an intervention works (or why it does not work), what the full range of effects are, how the effectiveness varies between groups or implementation sites and what might be causing the variation.[265] Detailed process evaluations of intervention implementation can help to plan adjustments where poor implementation is assessed as being responsible for low or lacking clinical effect, or scale up and spread for interventions which are found to be clinically effective.

The measurement of implementation strength is an important component of a process evaluation. Following a systematic review of studies of implementation strength, Schellenberg et al (2012) identified frequently used but heterogenous strategies to report overall strength of implementation. These included qualitative assessments based on attainment of specific conditions, or quantitative assessments, often reported on a 0-100 scale (including percentage score assessment). There was no consensus on how to present an overall score using quantitative assessment, with research groups using either presentation of mean scores of all or subgroup components, and other groups suggesting

methods to weight scores. In studies where domains were weighted, the weights applied were determined by expert consensus.[323,324] The linkage of intervention effectiveness with implementation strength is an evolving methodology.[325-327]

A study of the context in which GAP was implemented, the processes, acceptability and feasibility of implementation, and the interplay between context and GAP implementation is currently being prepared for publication. In this manuscript, Dr Coxon and I report that clinicians were willing to implement GAP and most believed in its ability to improve detection of SGA; although the evidence base supporting its effectiveness was questioned by some senior members of maternity staff. Staff also found that it was not always feasible to implement GAP - the success of its implementation was dependent on the availability of resources. The findings of this study on context and processes will also be summarised in detail within this chapter because they are essential to understanding the other findings.

The aim of this analysis was to assess the strength of implementation of GAP in the context of the DESiGN trial, and to understand the reasons behind inter-cluster variation.

3.1.1 Objectives

The objectives of this study were to:

- Assess the strength of implementation of GAP in each cluster site allocated to implement it.
- Consider the reasons for inter-cluster variation in implementation, including a study of context, processes, and barriers or facilitators of implementation.

3.2 METHODS

3.2.1 Study Design

A process evaluation of implementation nested within the DESiGN trial; a hybrid effectiveness-implementation randomised cluster control trial.

The study design and general data collection and analysis strategies have already been described in section 2.3.

3.2.2 Reporting checklist

This study has been reported according to the recommendations of the STaRi guidelines and checklist.[314] The completed checklist is included in the Appendix (section 10.6).

3.2.3 Measuring implementation strength

For this assessment of implementation strength, a mixed methodology approach was required. The implementation domains of fidelity, reach, dose delivered and dose received were applied to the processes inherent to GAP. These are summarised in Table 3.1 and detailed below.

Table 3.1 - Implementation domains as applied to GAP processes

Implementation outcome	GAP process to be measured
Fidelity	Degree of adherence to Perinatal Institute guideline
	Proportion of staff trained within each professional group (face-to-face and e-learning methods)
	Proportion of women correctly risk stratified (according to GAP)
Reach	Proportion of women with a GAP-GROW chart in the notes.
Dose delivered and received	Proportion of low-risk women* who had at least the minimum expected fundal height measurements performed and plotted on the chart
	Proportion of low-risk women* referred for growth scan when indicated.
	Proportion of high-risk women* who had at least the minimum expected growth scans performed and plotted on the chart

Risk status as determined by clinician. Risk assessment is expected to consider the risk stratification protocol specified in the GAP guidelines but may be modified for local practice.

3.2.3.2 *Implementation outcomes*

Fidelity:

Fidelity was assessed by:

- comparing recommendations made in cluster protocols to those made in the GAP guideline,
- assessing how adherent clusters were to the Perinatal Institute requirement that over 75% of staff members from each professional group were trained using both face-to-face and e-learning modalities
- comparing the extent to which the GAP-adopted NHS England risk stratification tool (Figure 1.6)[217] was applied when determining which SGA screening pathway pregnant women should be offered.

In assessing the degree of concordance of cluster protocols to the GAP guidelines, development of a quantitative scoring system by the trial team which accurately reflected the importance of (and therefore weight carried by) each recommendation, and the degree of any modification was expected to be complex, particularly since it is not known which are the most important components of the intervention. For this reason, the degree of adherence to GAP guidelines was assessed qualitatively using statements developed with reference to Schellenberg's review. These statements are:

- Low strength: Poorly adherent with partial or no inclusion of GAP recommendations throughout the guidelines, affecting over half of the recommendations.
- Medium strength: Moderately adherent with partial or no inclusion of GAP recommendations in less than half of the recommendations
- High strength: Very adherent with only occasional differences where GAP recommendations were partially included.

It was pre-specified in the trial protocol that the software required for GAP implementation (used to generate individual GROW charts) would only be released by the Perinatal Institute to cluster sites following attainment of the training targets. Since the Perinatal Institute recommended that at least 75% of staff from each professional group were trained, I also used the 75% threshold to identify sites which had achieved high implementation strength for this process.

For pregnant women, the proportion of women who were risk assessed correctly was judged according to the GAP-adopted NHS-England protocol (Figure 1.6).[217]

Reach:

Reach was assessed by the proportion of women giving birth, for whom there was evidence that the intervention had been implemented. This was evidenced by the presence of a GROW chart in the maternity record because without a GROW chart, the elements of fetal growth surveillance cannot be employed.

Dose delivered and received:

The GAP intervention is a point-of-care intervention. Whilst the woman has a choice to accept the components of the intervention offered to her (measurement of fundal height or estimation of fetal weight by ultrasound), we were not able to ascertain the proportion of cases in which the intervention was offered but not accepted. We found no evidence from qualitative interviews with staff that this was happening. For this reason, the dose domains have been measured together, with respect to the dose delivered to and received by pregnant women.

Dose was assessed by measuring the proportion of low- or high-risk women included in the notes review, who had received at least the minimum number of SGA screening assessments during pregnancy (fundal height measurement for low-risk women, fetal growth scans for high-risk women, both plotted on the GROW chart, Table 3.1).

Implementation strength:

Most GAP processes could be measured quantitatively (percentage scores), but adherence to guidelines was assessed qualitatively; this presented a challenge in appropriately combining the results for each process assessment into a single reported measure of implementation strength. Whilst an arbitrary point score could be allocated to each qualitative statement (i.e., high=90%, medium=60% and low=30%), a further challenge was also identified in calculating an overall score from all the quantitative results. Previously used strategies have been detailed in section 3.1. As the degree to which each domain contributes to the effectiveness of the intervention was unknown, the implementation of this intervention has not previously been studied and it was not possible to examine the separate clinical effect of each part of the process, it was concluded inappropriate to develop an arbitrary weighting system or assume equal weight for each domain by applying a system of mean scores. For this reason, the results of each assessment of implementation strength as applied to a specific domain and process are only reported individually. An overall impression of implementation strength can be derived by comparing scores for domains within and between cluster sites.

3.2.3.3 *Data collection*

The overall process of data collection for the assessment of implementation strength has already been described in section 2.3.3.

Briefly, for the assessment of fidelity, clinical guidelines on screening pregnant women for SGA were requested from each site principal investigator and training records were requested from both the GAP leads at implementing cluster sites and the Perinatal Institute.

Data for the remaining components of implementation strength were collected from a review of 600 maternity records. The number of notes to be reviewed (40 per month at each of five sites, over a three month period during the trial outcome phase, n=600 total), was determined following discussion with the trial Chief Investigator (Professor Pasupathy), the statistician (Professor Copas) and the lead co-investigator for the process evaluation (Professor Sandall). The final decision for the number of notes considered a subjective assessment of the number of notes required to draw robust conclusions of implementation strength and a pragmatic decision regarding feasibility, taking into consideration staffing resource. Paper maternity records were randomly selected from the postnatal notes stores at each cluster site, for births during the last three months of the trial comparison period (December 2018-February 2019). Maternity records were excluded according to the same exclusion criteria of the DESIGN trial (section 2.1.3).

Data collected during the notes review included individual items from each clinical record (data collection form included in the Appendix – section 10.7):

- NHS number and woman's DOB, required to allocate a pseudonymised study ID (section 0)
- Risk factors for SGA, as per the risk stratification tool from NHS England (Figure 1.6),[217]
- Risk status as assessed/managed by the clinical team,
- Presence of a GROW chart,
- Number of fundal height measurements plotted on the GROW chart (plots only included if separated by at least two weeks and occurring at or after 26 weeks' gestation),
- Number of fetal growth scans conducted, with a minimum interval of two weeks and minimum gestational age of 26 weeks', including number plotted on the GROW chart,
- Presence of fundal height or EFW plots which deviate from expected centile lines on the GROW chart,

- Clinical acknowledgement of a chart plot deviation
- Evidence of referral for a fetal growth scan if deviating fundal height GROW chart plots,
- Change in risk status during pregnancy, including gestational age at time of change.

It is common practice to only document risk status where risk is identified. Those women for whom a high-risk assessment or management protocol was documented were therefore assessed as being identified as 'high-risk'. Where risk assessment was documented and a woman judged low-risk for SGA, or where risk was not documented but the woman did not receive serial fetal ultrasound scans, the women were assumed to have been identified as 'low-risk'. This risk status was compared to that which was recommended both by GAP, and to that which was recommended by the cluster-specific protocol, according to the documented maternal characteristics, previous obstetric history, and co-existing morbidities.

Where a GROW chart was not identified in a woman's maternity record, it was assumed that she had not been provided with one during the pregnancy. The possibility of a misplaced chart is recognised as a limitation to this assumption.

To ensure that inappropriate risk stratification was not measured twice in the assessment of implementation strength, further assessment by risk status allocation was assessed according to the risk status determined by the woman's clinical team (not that which would be recommended by GAP).

The collection of data on deviations of GROW chart plots from the expected growth curve was acknowledged to be a subjective assessment. Whilst GROW chart plots below than the 10th centile and subsequent plots which are static (i.e., no change in the value of the y axis between consecutive plots) are easily identified, GAP guidelines do not provide definitions for slow or accelerative growth. Some degree of normal variation away from the true fetal growth line is expected due to inter- or intra-observer variation in measurement both of fundal height and of EFW at ultrasound.[218,219,328] No guidance is available from the Perinatal Institute on the distinction between this and a concerning deviation that warrants further investigation. This was dealt with during data collection in two ways:

- Two senior obstetric training grade doctors conducted the first two site visits for notes review together and discussed each GROW chart to agree whether the plotted measurements featured a deviation from the normal/expected fetal growth curve.
- Any deviations in fetal growth were classified as either 'definite' or 'possible' deviations. A 'definite' deviation represents a subjectively acute change in the fetal

growth curve (either positive or negative deflection), a 'possible' deviation is a deviation which is only slightly deviated from the normal curve, this may therefore represent normal variation. This classification of definite and possible deviations continued for notes reviews at all remaining sites and months.

3.2.3.4 *Data analysis*

In assessing the concordance of locally modified guidelines to those pre-specified by GAP, statements of recommendation were extracted from GAP and matched to similar statements in each of the cluster guidelines. Deviance from GAP was noted as either absence or modifications of GAP recommendations. A subjective assessment was made of the overall degree of concordance to the GAP recommendations, and the assessment was fitted to the descriptive statements (as detailed in Section 2.3.4.1).

Summary statistics were used to describe maternal and neonatal characteristics for pregnancies included in the notes review (section 2.3.4.1). Ethnicity was presented by the regions used by the GROW customised centile charts for the UK. For babies, the growth status at birth (SGA/AGA/LGA) was calculated as described in section 2.2.6.2. The remaining measures of implementation strength, stratified by site, were summarised using number and percentage of maternal records in which there was evidence that the care was compliant with the recommendation or target.

The expected number of fundal height or EFW plots on the GROW chart was only calculated for women who had a chart in their handheld notes and using the following considerations:

- Recommendation that first fundal height or EFW measurement should be taken and plotted at 26-28 weeks' gestation.
- Number of gestational weeks between 28 weeks' gestation and birth
- Recommendations that fundal height or EFW should be measured at a maximum interval of three-weekly (recommended two to three weekly measurements of fundal height). For example, a woman who gives birth at 38^{+5} weeks of gestation should have a minimum of four fundal height plots or fetal growth scans for surveillance (at 28w, 31w, 34w, 37w).

NHS guidance on antenatal care recommends that low risk multiparous women are seen less frequently than low risk nulliparous women.[8] This measure of dose for low-risk women was therefore also stratified by parity. A chi-squared test was conducted to determine whether there was a difference by parity overall.

The quantitatively measured components of implementation strength were summarised using percentage scores and presented together in a single table. Since the total number of implementing clusters was small (n=5), it was not considered appropriate to conduct a sensitivity analysis comparing the primary outcome of the trial only in sites with high implementation strength, nor was it possible to explore correlation between strength and outcome statistically. The relationship was therefore reviewed descriptively, a method which has been described previously.[329]

3.2.4   *Assessing reasons for inter-cluster variation in implementation strength*

3.2.4.1   *Data collection*

A study of GAP processes, the context of implementation and evaluation of barriers and facilitators that may affect implementation strength was conducted as part of the comprehensive qualitative evaluation of GAP implementation.

Qualitative data were collected at the five implementing clusters through semi-structured interviews with purposively sampled clinical leads and frontline staff. Sampling targeted recruitment of one clinical GAP lead from each professional group (obstetricians, midwives, and sonographers) and five to seven frontline midwives and sonographers working at each cluster.

Semi-structured interviews were also conducted with clinical leads at non-implementing sites (including the two sites allocated to implement the intervention who chose not to do so early in the trial). These interviews collected data on context and practice in the non-implementing sites, including data on reasons for not implementing in the two sites who chose not to.

Interviews with all frontline staff (n=28), were conducted by me, a senior obstetric training-grade doctor; interviews with all clinical leads (n=27), were conducted by Dr Kirstie Coxon, an experienced qualitative researcher with a clinical background in midwifery.

Where possible, interviews were conducted face-to-face. Interviews were arranged at times and in locations which were convenient with participants, including a minority which were conducted over the phone. Interviews were recorded electronically and professionally transcribed. The quality of each transcript was checked and accordingly edited by the responsible interviewer.

The topic guides for the interviews (Appendix section 10.8-10.10) were designed around the process evaluation objectives and in reference to the logic model (Figure 2.7), implementation strategy (Box 3.1) and protocols of the Perinatal Institute.[303] All interviews and the analysis were conducted before the results of the main trial were known, to prevent this from biasing the question strategy and interpretations.

Box 3.1 – GAP implementation strategy of the Perinatal Institute

- Selected staff to attend GAP 'train the trainers' workshop, led by midwives from the Perinatal Institute.
- Identify local GAP team and administrative leads (Midwife, Sonography, Obstetric leads, IT liaison for chart generation & software).
- Deliver both face-to-face and e-learning GAP training to 75% of staff from each professional group: midwives, sonographers, and obstetricians (face-to-face training cascaded by trainers).
- Complete baseline audit of rate of SGA, referral for suspected SGA and confirmed SGA detection (3 months' records).
- Update site-wide fetal growth assessment protocol in line with guidance issued by the Perinatal Institute.

3.2.4.2 *Data analysis*

Transcribed copies of the interviews were uploaded into and analysed using NVivo v12.0 (QSR International Pty Ltd, Victoria, Australia). Interview data were deductively coded onto the domains and sub-domains of the CICI framework (see sections 2.3.2.2 and 2.3.4.2). The data were coded by two independent researchers (Dr Kirstie Coxon and me) who regularly discussed and documented coding decisions to ensure inter-researcher consistency. Where the data did not fit clearly into the available codes, we discussed with Professor Sandall, a senior qualitative researcher, and, if required, added sub-codes within existing CICI domains e.g., for data on feasibility or leadership. Following coding of the full dataset, a matrix was set-up in NVivo to allow comparison of codes between sites and healthcare professional groups.

Context was summarised for each implementing cluster to describe the areas in which there was thematic saturation of data (as described by Saunders et al),[330] either across cluster sites (macro context), within cluster sites (meso context) or across individuals (micro context), using the CICI context and setting domains. We judged that saturation had occurred where detailed, in-depth data were available to confirm a finding from a range of different participants and sites provided.[330]

To describe site-based implementation processes in each cluster for all stages of implementation (Table 2.12), including concordance with GAP or adaptations made, process domains were summarised descriptively for each site.

There is no CICI framework domain specifically for the coding of barriers and facilitators to implementation. Pre-existing barriers or facilitators were identified by reviewing data coded in the context dimension. Those which arose during implementation and affected specific implementation processes were identified by reviewing the data coded in the implementation dimension. The findings were described according to the frequency that they appeared in the data, including the number of clinicians who spoke about them and the number of sites at which they were mentioned. These were compared descriptively at the site with highest and the site with the lowest implementation strength.

3.3 RESULTS

3.3.1 Evaluation of implementation strength

Clinical guidelines were collected from all five GAP-implementing sites and training records were collected for all five sites from the Perinatal Institute. For the notes review, 595 notes were reviewed across the five research clusters. The characteristics of the women included in the notes review are detailed in Table 3.2, including the number/percentage of women with risk factors for SGA, or for inaccurate fundal height measurement during pregnancy.

Table 3.2 - Characteristics of women included in the notes review for implementation strength

Characteristic		Mean / Median / Number	SD / IQR / %
Age at 12 weeks' gestation	Age (mean/SD)	31.2y	5.6y
	Women with age>40y at 12/40 (n/%)	28	5.0%
	Missing (n/%)	39	6.6%
Ethnicity	United Kingdom	233	39.2%
	Other European	114	19.3%
	Middle East	23	3.9%
	African	72	12.1%
	Caribbean	15	2.5%
	Asian	98	16.5%
	North America	2	0.3%
	South America	2	0.3%
	Australian	1	0.2%
	Mixed	16	2.7%
	Unclassified	17	2.9%
	Missing	2	0.3%
Body mass index	BMI (mean/SD)	25.6 kg/m^2	5.4 kg/m^2
	BMI>35 (n/%)	38	6.4%
	Missing (n/%)	0	0.0%
Parity	Median/range	1	0-7
	Nulliparity (n/%)	289	48.6%
	Missing (n/%)	1	0.17%
Social risk factors	Smoker at booking	45	7.6%
	Smoking status missing	1	0.2%
	Illicit drug users	6	1.0%
	Drug use status missing	7	1.2%
Previous obstetric history	Previous SGA baby – on GROW chart	57	9.6%
	Previous SGA unknown	49	8.2%
	History of stillbirth	3	0.5%
Medical co-morbidities	Chronic hypertension	8	1.3%
	Pre-existing diabetes	6	1.0%
	Chronic renal impairment	1	0.2%
	Antiphospholipid syndrome	0	0.0%
Antenatal risk factors	Low PAPP-A* <0.415MoM	18	3.0%
	Low PAPP-A <0.30MoM	3	0.5%
	PAPP-A missing	240	40.3%
	Echogenic bowel	1	0.2%
	Scan history missing	42	7.1%
Large Fibroids (>4cm)	Large fibroids (n/%)	9	1.5%
	Missing scan history	14	2.5%

*PAPP-A: Pregnancy-associated plasma protein A. PAPP-A often missing because it is a biochemical marker usually reported on a loose sheet of paper or via the electronic ultrasound records system (and not transferred into maternal handheld records).

The characteristics of babies born to women whose notes were included in the review of implementation strength are summarised in Table 3.3.

Table 3.3 - Characteristics of babies born to women included in the notes review for assessment of implementation strength

Birthweight	
Median (IQR)	3370g (3025-3680g)
Missing (n/%)	4 (0.7%)
SGA by customised centile (n/%)	80 (13.4%)
LGA by customised centile (n/%)	60 (10.1%)
Gestational age at birth	
Median (IQR)	39^{+5} ($38^{+5} - 40^{+4}$)
Range	$26^{+5} - 42^{+3}$
Born preterm <37^{+0}/40 (n/%)	28 (4.7%)
Neonatal sex	
Male (n/%)	279 (46.9%)
Female (n/%)	314 (52.8%)
Missing (n/%)	2 (0.3%)

Fidelity:

Clinical guidelines from two cluster sites were judged to have high fidelity, two sites medium and one site low fidelity to the guidelines recommended by GAP. Areas of concordance or deviation are described in (Table 3.4). All sites achieved the target of >75% staff from each staff group to be trained by face-to-face methods. Only one site achieved this for the e-learning training target, even after extension of the deadline to 3-months after 'going live' with GAP.

When all indications for serial fetal growth scans in pregnancy were considered together, 177 women (29.7%) had at least one risk factor. Of the remaining 418 women with no known risk factors, 197 (47.1%) had missing information on at least one risk factor (Table 3.5), most commonly missing data on pregnancy-associated plasma protein A (PAPP-A). These proportions varied across study sites. Of the 419 women with no known risk factors in early pregnancy, 26 developed risk factors which upgraded their risk status antenatally, warranting late referral for serial growth scans.

Table 3.4 - Assessment of fidelity of local guideline to that recommended by GAP

Fidelity score	Site number	Comments
Low fidelity	**Site 7**	Splits risk factors into major 1, major 2 and minor. Growth scans recommended for major 1 risk factors at 30 and 34 weeks only. For major 2 risk factor or ≥3 of the minor risk factors, women are referred for uterine artery Dopplers at 20-24 weeks and then have 'serial' scans at 30 and 34 weeks if UADs abnormal or one growth scan at 34 weeks if UADs normal. Management of low risk women and management of SGA is the same as recommended by the Perinatal Institute.
Medium fidelity	**Site 9**	**Differences in high-risk definition:** Smoker >10 rather than any smoker, BMI<18 additional category, misses group with BMI 35-40, misses group with PAPP-A 0.3-0.415MoM, scans women with any PIH rather than severe PIH, significant APH rather than unexplained APH. Additional high risk groups: Heavy bleeding 1st TM similar to menses, low lying placenta, GDM, PPROM, fetal hydronephrosis, polyhydramnios >30cm, previous PET, sickle cell disease. **Differences in scan protocols:** Recommends growth scans at 28 & 32/40, rather than 3-weekly scans for smokers, drug users, high BMI, age>40y, low PAPP-A, fibroids Recommends scans at 28,32,36 rather than 3-weekly for previous FGR (plus additional scan at 34/40 for early onset FGR), previous stillbirth, diabetes, twins. Obstetric medicine team to determine scan frequency for obstetric medical conditions (no guidance provided). For PET, recommends two weekly scans (rather than 3-weekly), for PIH recommends 4-weekly (rather than 3-weekly). **Does not clearly specify to plot the scan EFW onto the customised chart**
	Site 10	Management of **low-risk women** is the same (except if the low risk woman has abnormal uterine artery Doppler on routine screening, in which case she is managed as medium risk). **For high risk women:** Women missed off risk screening: Drug misuse, previous stillbirth, APLS, PAPP-A 0.4-0.415, fetal echogenic bowel, unexplained APH Additional growth scan indications: Previous SGA<10th centile on population chart, previous 2.5kg baby at term, previous early onset PET, single umbilical artery, current hyperthyroidism on medication, new onset PIH (any severity), new GDM. Introduced a medium risk category for women with no risk factors but high uterine artery Dopplers OR women with risk factors but normal uterine artery Dopplers. Serial growth scans are 28/32/36 for high-risk women rather than 3 weekly. Medium risk women get scans at 28 and 36 weeks only.
High fidelity	**Site 8**	Almost no differences (the only difference recommends to also refer the fetus with AC<5th centile to consultant, as well as EFW<10th centile on customised chart).
	Site 11	Recommends SFH measurement in low-risk women and EFW measurement in high risk women to be conducted 2-4 weekly (rather than 2-3 weekly)

Table 3.5 - Overall risk factor assessment per site

	Site 7	Site 8	Site 9	Site 10	Site 11	All
Women with at least one known risk factor for SGA (n/%)	39 (32.5%)	36 (30.8%)	30 (24.8%)	35 (29.9%)	37 (30.8%)	177 (29.7%)
Women with no known risk factors and complete data (n/%)	29 (24.1%)	59 (50.4%)	42 (34.7%)	49 (41.9%)	42 (35.0%)	221 (37.1%)
Women with no known risk factors but missing information (n/%)	52 (43.3%)	22 (18.8%)	49 (40.5%)	33 (28.2%)	41 (34.2%)	197 (33.1%)
Total (n)	120	117	121	117	120	595

The assessment of risk stratification as concordant with GAP guidelines and, where practice was deviant, concordance with local policies is summarised in Table 3.6. 84.9% of women were correctly risk stratified as per GAP. In total, 90 women had discordant risk stratification when compared to that recommended by GAP; for 18 of these women, the risk assessment by the clinician was correct according to local protocols which had modified GAP recommendations.

Where new onset risk factors for SGA arose during pregnancy, 5/26 women (19.2%) did not have documentation of renewed assessment of risk for SGA, nor were they referred for serial growth scans following the change is risk status.

Table 3.6 - Outcome of the assessment of risk stratification, comparing clinician assessment to GAP and local recommendations

	Risk status (by GAP)	Site reference					
		Site 7	Site 8	Site 9	Site 10	Site 11	All
Agreement between GAP and clinician	**High risk (n)**	32	24	21	32	24	133
	Low risk (n)	73	68	87	68	76	372
	Both n(%)	105 (87.5%)	92 (78.6%)	108 (89.3%)	100 (85.5%)	100 (83.3%)	505 (84.9%)
Clinician has not classified risk as recommended by GAP	**High risk (n)**	9	13	9	3	14	48
	Low risk (n)	6	12	4	14	6	42
	Both n(%)	15 (12.5%)	25 (21.4%)	13 (10.7%)	17 (14.5%)	20 (16.7%)	90 (15.1%)
If misclassified by GAP standard, classified correctly as per local policy?	**n(%)**	2 (13.3%)	0 (0.0%)	7 (53.8%)	7 (41.2%)	3 (15.0%)	19 (21.1%)

Reach:

Overall, 88.7% of women had a GROW chart in their maternity record. This differed by site; the percentage of women with a GROW chart ranged from 61.7% at site 7 to 98.3% at site 8 (median 94.2%). Site specific percentages are summarised in Table 3.10.

Dose delivered and received:

Of women who were at low risk of SGA in early pregnancy and who had a GROW chart in their notes, the proportion with at least the minimum expected number of fundal height plots on their chart is presented per site in Table 3.7. Overall, 30.7% of women (n=114/371) had at least the minimum expected number of fundal height plots. Nulliparous women were overall more likely to have had at least the minimum expected number of fundal height measurements plotted than multiparous women (p<0.001).

Table 3.7 - Proportion of low-risk women with at least the minimum expected number of fundal height plots on GROW chart

Site identifier	Women with at least the minimum expected number of fundal height chart plots	
	Number	**Percentage**
Site 7 (n=49)	4	8.2%
Nulliparous (n=25)	3	12.0%
Multiparous (n=24)	1	4.2%
Site 8 (n=79)	42	53.2%
Nulliparous (n=43)	28	65.1%
Multiparous (n=36)	14	38.9%
Site 9 (n=90)	31	34.4%
Nulliparous (n=55)	22	40.0%
Multiparous (n=35)	9	25.7%
Site 10 (n=70)	22	31.4%
Nulliparous (n=43)	15	34.9%
Multiparous (n=27)	7	25.9%
Site 11 (n=83)	15	18.1%
Nulliparous (n=36)	9	25.0%
Multiparous (n=47)	6	12.8%
Total (n=371)	**114**	**30.7%**
Nulliparous (n=202)	**77**	**38.1%***
Multiparous (n=169)	**37**	**21.9%***

*Chi-squared test comparing proportion of nulliparous to multiparous women
with the expected number of fundal height plots, p<0.001

The proportion of low-risk women referred for a fetal growth scan when indicated by a deviation in the fundal height growth trajectory was categorised into those women with any deviation in the trajectory (possible or definite deviation) and those women with a definite deviation. Overall, 42.3% of women had any evidence of a deviation in the fundal height and 27.5% had a definite deviation in the fundal height growth trajectory. 56.1% of the women with any deviation and 67.6% of the women with a definite deviation were referred appropriately for a fetal growth scan. Further breakdown by site is presented in Table 3.8.

Table 3.8 - Proportion of low-risk women referred for a growth scan when indicated by deviations (definite/possible) in the fundal height plots

	Site reference	n	%
Women with <u>any</u> fundal height centile deviation (n/%)	Site 7 (n=49)	14	28.6%
	Site 8 (n=79)	33	41.8%
	Site 9 (n=90)	41	45.6%
	Site 10 (n=70)	29	41.4%
	Site 11 (n=83)	40	48.2%
	All (n=371)	**157**	**42.3%**
Women with <u>any</u> fundal height centile deviation who were correctly referred for a growth scan – overall (n/%)	Site 7 (n=14)	4	28.6%
	Site 8 (n=33)	26	78.8%
	Site 9 (n=41)	25	61.0%
	Site 10 (n=29)	11	37.9%
	Site 11 (n=40)	22	55.0%
	All (n=157)	**88**	**56.1%**
Women with a <u>definite</u> fundal height centile deviation (n/%)	Site 7 (n=49)	10	20.4%
	Site 8 (n=79)	24	30.4%
	Site 9 (n=90)	22	24.4%
	Site 10 (n=70)	15	21.4%
	Site 11 (n=83)	31	37.4%
	All (n=371)	**102**	**27.5%**
Women with a <u>definite</u> fundal height centile deviation who were correctly referred for a growth scan (n/%)	Site 7 (n=10)	4	40.0%
	Site 8 (n=24)	19	79.2%
	Site 9 (n=22)	17	77.3%
	Site 10 (n=15)	10	66.7%
	Site 11 (n=31)	19	61.3%
	All (n=102)	**69**	**67.6%**

For women who were assessed as high risk by the clinician, the median number of growth scans conducted overall was 3 (IQR 2-4). Amongst all sites, 76.9% of the EFWs generated from these growth scans were plotted onto the GROW charts. Of high-risk women with a GROW chart (n=201), 12.4% overall received at least the minimum expected number of fetal growth surveillance scans and 8.5% had at least the minimum number plotted on their GROW chart. This varied by site – women at site 7 received the lowest intervention dose, women at site 10 received the highest dose (Table 3.9).

Table 3.9 - Proportion of high-risk women with at least the minimum expected number of fetal growth scans conducted and plotted on GROW chart

	Site reference	Number / Median	Percentage / IQR
Median number of growth scans received by high-risk women	Site 7 (n=33)	2	2-3
	Site 8 (n=48)	4	3-4
	Site 9 (n=35)	3	3-4
	Site 10 (n=47)	3	2-4
	Site 11 (n=38)	3	2-4
	All (n=201)	3	2-4
Proportion of all EFW measurements in high-risk women which were plotted on a GROW chart (n/%) (n=number of EFW measurements per site)	Site 7 (n=110)	40	36.4%
	Site 8 (n=169)	163	96.4%
	Site 9 (n=106)	77	72.6%
	Site 10 (n=156)	117	75.0%
	Site 11 (n=104)	99	95.2%
	All (n=645)	**496**	**76.9%**
Proportion of high-risk women with a GROW chart who had at least the minimum expected number of fetal growth scans (n/%)	Site 7 (n=33)	0	0.0%
	Site 8 (n=48)	8	16.7%
	Site 9 (n=35)	4	11.4%
	Site 10 (n=47)	11	23.4%
	Site 11 (n=38)	2	5.3%
	All (n=201)	**25**	**12.4%**
Proportion of high-risk women with a GROW chart who had at least the minimum expected number of EFWs plotted on the chart (n/%)	Site 7 (n=33)	0	0.0%
	Site 8 (n=48)	8	16.7%
	Site 9 (n=35)	1	2.9%
	Site 10 (n=47)	6	12.8%
	Site 11 (n=38)	2	5.3%
	Total (n=201)	**17**	**8.5%**

Implementation strength:

The overall assessment of implementation strength brings together all measures for fidelity, dose and reach in Table 3.10, and compares these alongside the rate of detection of SGA at each site. The baseline rate of SGA is included for context, demonstrating the degree to which sites increased their detection from baseline. Site 7 consistently scored the lowest for implementation fidelity, dose and reach. With regards to the site which consistently achieved higher compliance for each of the measures than the other units, it is likely to have been site 8, but this is less clear. This ranking is important because it leads onto a comparison on barriers and facilitators to implementation strength in units that had low and high implementation strength.

This ranking of sites in terms of their implementation strength is also useful for a comparison of implementation strength with intervention effectiveness. Site 8 is likely to be the site with greatest implementation strength and did achieve the highest rate of SGA detection (41.4%), although it is not known whether this represents an increase from its baseline rate of detection (the EFWs from ultrasound data were not available at this site during the baseline period). Site 7 was the weakest in terms of implementation strength, and also achieved one of the lowest rates of SGA detection (20.3%) which changed only slightly from its baseline rate of detection (18.3%). In terms of the other sites, site 9 was

also one of the strongest in terms of implementation strength but achieved the lowest rate of SGA detection (19.6%), thereby being an outlier of the trend seen in the other sites between strength of implementation and rate of SGA detection.

Table 3.10 - Overall assessment of implementation strength

			Site 7	Site 8	Site 9	Site 10	Site 11
Detection of SGA	Proportion of SGA babies who were antenatally detected*	Baseline phase	18.3%	-⁺	21.5%	31.7%	24.2%
		Trial comparison phase	20.3%	41.4%	19.6%	33.3%	32.1%
Fidelity	Proportion of staff trained within each professional group	Face-to-face target (%)	>75%	>75%	>75%	>75%	>75%
		E-learning target (%)	<75%	<75%	>75%	<75%	<75%
	Degree of concordance with the Perinatal Institute guideline	% (n/N)	Low fidelity	High fidelity	Medium fidelity	Medium fidelity	High fidelity
	Proportion of women risk stratified according to GAP	% (n/N)	87.5% (105/120)	78.6% (92/117)	84.2% (105/121)	83.2% (99/119)	84.4% (98/116)
Reach	Proportion of women with a GAP-GROW chart in the notes.	% (n/N)	62.2% (74/119)	98.3% (115/117)	93.3% (131/121)	96.6% (115/119)	94.2% (113/120)
Dose	Proportion of low-risk women who had at least the minimum expected number of fundal height measurements performed and plotted on GROW	% (n/N)	8.2% (4/49)	53.2% (42/79)	34.4% (31/90)	31.4% (22/70)	18.1% (15/83)
	Proportion of low-risk women referred for growth scan when definite plot deviation.	% (n/N)	40.0% (4/10)	79.2% (19/24)	80.9% (17/21)	66.7% (10/15)	61.2% (19/31)
	Proportion of high-risk women who had at least the minimum expected number of growth scans performed and plotted on GROW	% (n/N)	0.0% (0/33)	16.7% (8/48)	2.9% (1/35)	12.8% (6/47)	5.3% (2/38)

*Proportion of SGA babies calculated using imputed dataset, so numerator and denominator not applicable.
⁺Estimated fetal weights from ultrasound were not available for any birth during the baseline phase at site 8.

167

3.3.2 Qualitative analysis of barriers and facilitators to effective implementation

The interviews were conducted between February 2018 and May 2019, throughout the process of GAP implementation (Figure 3.1). In total, 55 interviews were conducted; 28 with frontline staff at the implementation sites and 27 interviews with the GAP leads (22 at implementation sites and 5 at control sites). All but three interviews were conducted face-to-face, the remaining were conducted over telephone, at the request of the participant. Only interviews from implementing sites were used for this analysis, the professional roles of these interviewees are summarised in Table 3.11.

Table 3.11 - Professional roles of participants from implementation sites in semi-structured interviews

	GAP Leads			Frontline staff		Total
Implementation site	Obstetrician	Sonographer	Midwife	Midwife	Sonographer	
7	1	0	2	5	0	8
8	1	1	2	3	2	9
9	1	1	2	4	2	10
10	2	0	2	2	3	9
11	1	1	1	5	2	10

Figure 3.1 - Timeline of implementing GAP within the context of the DESiGN trial

Intervention sites*	2016			2017													2018												2019		
	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb

Key:

Key:	
Randomisation	
Training	
Go Live	
Exposed term babies are born	

Sites: A, B, C, D, E

Interview data collected

Site Numbers not used to prevent identification from randomisation and implementation dates

169
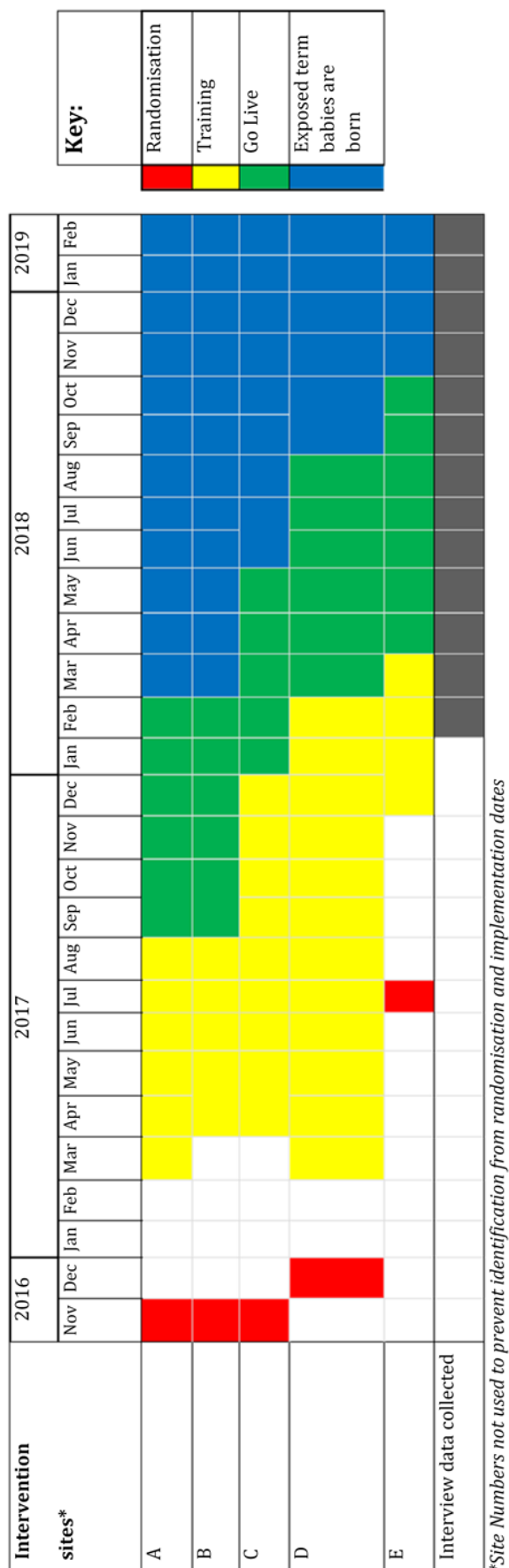
3.3.2.1 *Describing the context in which GAP was implemented*

The context and setting in which GAP was implemented is key to understanding the barriers to and facilitators of high strength implementation. Summary findings are presented sequentially for the CICI setting dimension and, following this, for each domain within the context dimension, with examples provided of the data which led to the findings. All data underpinning each summarised conclusion can be read in the Appendix (section 10.11).

Setting:

The setting dimension refers more specifically to the physical space in which implementation occurs and occupational aspects which affect it. At all sites, participants frequently spoke about problems with computer or printer access:

"My biggest problem is getting onto the system at the beginning of the day because the computer system here is very slow" (HP51, midwife, site 8)

"So when we go to a clinic or outreach area, we have to take the information from the computer, so blood results, scan results and document them in their…in their hospital notes which we take to the clinic" (HP5, midwife, site 11)

"I think where we are going paper-light, we have become light on printers, so I think then we are stuck, isn't it, because we want the printers because we want to print this and the trust is like, "We no longer need that many printers because we are going paper-light"."
(HP78, midwife, site 7)

Many participants spoke about staffing shortages.

"'Cause I think, we just had a period where we couldn't recruit, because, I think we have like about eight people, all on maternity leave at the same time. So they were trying to back-recruit, but the paperwork was going through. People… we had people leaving."
(SC20, GAP lead, site 7)

"There is a national and international shortage of sonographers. And so, trying to staff departments properly, and trying to expand at the rate that you have to do all this extra scanning, is very very difficult, because there just aren't the sonographers out there."
(SC24, GAP lead, site 11)

Epidemiological:

There were variable beliefs expressed regarding the local rate of stillbirth and of the local effectiveness of antenatal SGA detection.

"We definitely have a problem in terms of adverse outcomes like stillbirths, miscarriages and neonatal deaths." (HP37, midwife, site 7)

"I think in this trust particularly we are doing well. Again, I would say GAP would be more beneficial in other trusts or other areas." (HP93, sonographer, site 9)

"Again, I'm not sure of our exact percentage, but I'm pretty sure that we are on par, if not below national" (HP26, sonographer, site 10)

Participants from all trusts commented on characteristics of their service users which make their population at higher risk of having an SGA or stillborn baby.

"We deal with a really high-risk population which includes asylum seekers, refugees, many women who don't speak English at all who come from countries where they haven't had any health care, who haven't got full medical history." (HP37, midwife, site 7)

"[...] a lot of patients with complex conditions and some of these conditions come with a higher risk of stillbirth, so are seen here. So yes, our population might have a slightly higher risk of stillbirth that others because it's slightly more complex, a bit older than average." (SC05, GAP lead, site 10)

"Yeah...yeah, because in our population here, we do have lots of Asian [women]... and I think that our constitutionally... small babies are fine as opposed to the babies that really ought to have more intervention and observation." (SC17, GAP lead, site 11)

Ethical:

Participants from several of the sites commented on their desire to perform better at work, with respect to individual responsibility to detect SGA and prevent stillbirth.

"occasionally things are missed just because of the pressure of work, if they are both feeling rushed and all of that, that might be a reason for it being missed." (HP78, midwife, site 7)

"I have a vested interest as a community midwife in identifying the low-risk ladies whose babies unfortunately don't make it. I felt sort of a responsibility that you know, as part of the low-risk midwifery role to actually be better at doing this..." (SC17, GAP lead, site 11)

Many participants felt distress when thinking about past adverse fetal events, in which they had been professionally involved in caring for women.

"it's an awful thing that happens at the time" (HP46/58, sonographers, site 8)

"It is me being super cautious because of past experience." (HP47, midwife, site 9)

There were also several descriptions of conflicting interests with regards to the opportunity cost of arranging additional ultrasound scans for fetal growth.

"It's how you then balance the need for that woman to have the scan and also with the availability of slots.." (SC33, GAP lead, site 7)

"Our growth scan rate has gone through the roof and that in turn has affected our gynae waiting list, so we have been breaching our gynae waiting list and I've had one permanent member of staff resign and that was because... well one of the reasons, was mentioned when she told me, was GAP." (HP23, sonographer, site 11)

Geographical:

There were only two references to geography which came from one interviewee. The DESiGN trial was mostly set in a highly developed environment (central and suburban London) in which infrastructure barriers and issues of physical access to resources were rarely problematic, or relevant to this research question. At the one site where geography was deemed important, community clinics were far from the primary hospital site and the interviewee perceived that women didn't like to travel to the hospital if they required investigation of a pregnancy complication.

"well, community in [maternity unit] it's quite, it's quite broad, it goes up to [location border]. So it's, we've got midwives with their own clinics are quite far from the hospital anyway. […] So sometimes the women are not that keen on travelling."
(SC03, GAP lead, site 9)

Legal:

The legal domain includes sub-domains on professional autonomy and guidelines. Many participants spoke about professional autonomy, it was particularly interesting to note how the degree to which individuals perceived their own autonomy differed between sites.

"so extra growth scan, er, the midwife just refers to the, to the scan department, and they just book their own appointment." (SC03, GAP lead, site 9)

"just to make sure that the scans are absolutely required because, as you can imagine, a lady can be referred a scan for any reason. So to make sure that the scans are appropriate, to make sure that they are happening at the right time. Normally, it will go for our [Maternal and Fetal Assessment Unit] lead who can triage and just make sure that the scan is appropriate and happening in a timely manner." (HP26, sonographer, site 10)

"from an ultrasound perspective, I vet every single form that comes in as superintendent. And I make sure that they're valid requests. If they're not, I'll bounce them back"
(SC24, GAP lead, site 11)

GAP was being implemented within the context of the DESiGN trial at the same time as other national initiatives to reduce the rate of stillbirth. This included a recommendation for trusts to introduce new guidelines for the management of women attending for urgent care following a concern about reduced fetal movements.

"So, if they have reduced foetal movements, they are encouraged to come into the day assessment unit and if they are less than, I think, twenty-eight weeks, the midwife just listens in. If they are more than twenty-eight weeks, then they do the CTG and if they come for their second presentation of reduced foetal movements, then they get referred for a scan."
(SC31, GAP lead, site 9)

Political:

This domain covers public policies and influential people. Participants were all prompted to discuss other initiatives to reduce stillbirth which had been implemented locally alongside GAP. Other initiatives with national roll-out included the Saving Babies' Lives care bundle, of which element two is particularly relevant to the GAP intervention, MBRRACE-UK perinatal morbidity and mortality registry and the RCOG's Each Baby Counts program.[217,267,331]

"So Jeremy Hunt [Secretary of State for Health] made up a figure and then we all tried to achieve it....But actually, I do think it was wonderful because it was such an ambitious figure which resulted in some really ambitious work to achieve it. So I think he said he wanted to reduce it by 25% or 50% by 2025, the number of stillbirths and neonatal deaths, then we've had RCOG with Each Baby Counts, the GAP protocol." (HP37, midwife, site 7)

In addition to Jeremy Hunt, former Secretary of State for Health, other influential people were also discussed:

"I remember a CQC assessor coming who had been the lead midwife at an organisation and she said to me 'Don't rush into it, you will not realise what the impact is on your service', and that echoed with me" (SC04/07, GAP leads, site 8)

"then there was the Perinatal Institute's GAP GROW, and our commissioners were very keen" (SC04/07, GAP leads, site 8)

"Well going back a few years ago I remember seeing a Panorama programme, I think it was about Gardosi, about his findings and the implementation or the beginning of the GAP charts in the West Midlands and its purpose to reduce the stillbirth rate." (HP52, midwife, site 11)

Sociocultural:

Midwifery staff often spoke about workplace culture, describing that they were comfortable to approach a range of colleagues for help when they were unsure with their practice:

"we have got a group WhatsApp and we can message and somebody who is generally at a computer will look it up for you. If not, just ring the clinic and they'll advise you". (HP74, midwife, site 7)

"if we were unsure, we'd just ask one of our colleagues, "Is this the way we do it?" or, "Can you tell me...?" That kind of thing". (HP99, midwife, site 8)

Staff were aware of practice in neighbouring maternity units because of contacts with their professional or social networks. Some felt that they were being left behind without implementing GAP.

"I can remember when I was training, other midwives would speak about different trusts that had more personalised growth charts and that maybe at our hospital we had a poor detection rate of SGA babies or [FGR] because we were just using the standardised chart and we were quite behind on other trusts by doing that." (HP44, midwife, site 9)

"We had heard quite a bit about it, yes, because we've been to quite a few study days where hospitals had implemented it and they were giving us their feedback on it."
(HP23, sonographer, site 11)

Socioeconomic:

Four out of five sites spoke about financial concerns related to the implementation of GAP.

"if the workload increases, […] we need more sonographers, and we don't have the money!"
(SC28, GAP lead, site 7)

"I believe it's definitely more expensive to have that electronic version because I know that we have looked into it but I think financially it's just not something that this trust I think is willing to do at the moment." (HP26, sonographer, site 10)

3.3.2.2 *Describing the implementation processes adopted at each site*

Components of GAP implementation were grouped into implementation process categories as defined by the CICI framework. The practice at each research site was compared to that which was recommended and processes which were either deviant or additional to GAP recommendations were documented (Table 3.12). In terms of the overarching processes of implementation, sites were not compliant only with elements of 'Evaluation and Reflection', namely three sites did not interact with the GAP-provided missed case audit tool and none of the sites provided evidence for ongoing training of staff, although assessments were conducted within the first few months of implementation.

For each process component of implementation (excluding sustainment for which assessment was too early), sites described additional methods they adopted to assist local implementation (Table 3.12).

Table 3.12 - Implementation processes at research sites, compared to recommendations of the GAP

	Perinatal Institute expectation	Deviant sites	Additional processes identified
Exploration	N/A	Nil	Site 9 - Worked with scan department to reduce unnecessary scans in preparation for expected increase in workload.
Decision to adopt	N/A	Nil	External pressures from management/commissioning organisations
Planning and preparation	Baseline audit of SGA detection (6 months of births prior to GAP implementation) Incorporate GAP into local guidelines Identification of a GAP team Train the trainers – Perinatal Institute training session	Nil	Site 8&9 – extension of scan capacity Site 10 – commenced specialist FGR clinic Site 11 – seconded lead midwife Site 7, 8, 9 & 11– developed posters for staff Site 8,11 – risk assessment aids
Initial implementation	Cascade training to >75% staff from each major staff group >75% staff to complete GAP e-learning Prepare to 'Go Live'	Nil	Site 7,8&10 – training incorporated into mandatory training Site 9& 10 arranged 'drop in' training sessions Site 11 – ad hoc 'on the job' training Variation between and within sites re. staff released to do e-learning at work
Full implementation	Risk assessment of all pregnant women prior to 24 weeks gestation Production of GROW charts for women Serial scans for high risk women (3-weekly from 26-28 weeks) Serial fundal height measurement for low risk women (3-weekly from 26-28 weeks) EFW and FH plotted on GROW chart Growth scans within 3/7 if triggered by FH on GROW Management of SGA as per RCOG guidance	Nil	Variation in when GROW charts printed and by whom Site 7 – commenced additional scan slots Site 7 – direct line to request urgent growth scans Site 8 – sticker to identify women on scan pathway Site 9 – sonographers arrange serial scans at time of anomaly scan (according to risk assessment sheet) Site 9 – escalation policy in case no scan availability Site 10 – sonographers do risk assessment and arrange scans Site 11 – all urgent scan requests vetted prior to slots being allocated
Evaluation and reflection	Missed case audit	Sites 7, 8 & 10 did not access tool	Site 7, 8 & 11 – local audits on GAP compliance (all women) Site 7 – all GROW charts produced to be recorded centrally
Sustainment	Annual face-to-face and e-learning training for >75% staff members from each major group.	No data collected in DESiGN trial	

175

3.3.2.3 *Barriers to implementation*

The common and important barriers to implementation of GAP are summarised in Table 3.13, including a distinction about whether these were pre-existing issues or arose during implementation, and an indication of which implementation process they affected.

The most identified pre-existing barrier to implementation was a problem with access to the computer hardware required to use the GAP software package (for generating GROW charts and birthweight centiles). 15 participants spoke about this 20 times across all five implementing sites.

"so when you are in clinic and you've got a computer that is attached to a printer, you can do it while the woman is there and put it in the notes. The problem we have when there isn't a printer or you haven't got a computer and you've got to generate it later, then what's happening, we are having to send them through the post to the woman and ask her to put them in the notes and that, I think, is where it can fall down." (HP74, midwife, site 7)

"in the community we got like a very difficult system, where they have to er, er go out in the Children's Centre, so they don't have the printer, they don't have this, they don't have that, so they were like no, we can't do, we can't do the regular chart." (SC03, GAP lead, site 9)

As was noted in section 3.3.2.1 (Setting), staff also spoke about shortages in midwives and sonographers, with national shortages of sonographers being of particular concern.

Staff at two of the implementing sites spoke about challenges presented by the characteristics of their local population in either ensuring women have full exposure to the intervention (dose),

"Because of our clientele, we get a lot of very vulnerable ladies, a lot of ladies who are asylum seekers who have had very little antenatal care" (HP71, midwife, site 7)

or, in their ability to offer the intervention to all women for whom it was recommended (reach). For example, one site perceived that they had an above average prevalence of women with a body mass index about 35kg/m^2 and felt unable to offer serial fetal growth scans to all the women:

"if we do routine scans for [sigh] serial scans for all women who've got a BMI of over 35, again, that's most of our patients, they said. Yeah. And that's the sad reality of life really, now, is that, you know, a lot of our patients are overweight when they get pregnant. And a lot are obese. A lot will fall into that category of BMI of 35, and I don't think there's anything wrong with it, the women. They might have always been big, and just a bit bigger than, you know, so, you know, I can understand why they were so stressed about it, about us implementing that, because, definitely our patients are getting heavier. Yeah, they're getting bigger." (SC06, GAP lead, site 9)

With regards to barriers which arose during the process of implementation, there were several difficulties raised with rolling out training to >75% of staff from each professional

group. At four sites, it was noted that midwives and sonographers were not routinely being given time away from their clinical work to do their e-learning training and that this had caused challenges in achieving the training targets:

"The fact they have to do an e-learning, they're like, we are already too much going on, and they're working really hard, so you can kind of explain the absence, it's very challenging."
(SC03, GAP lead, site 9)

"So it's mostly their excuse for not doing it is usually having time to do it."
(HP78, midwife, site 7)

Some groups of staff were noted to be more difficult to train than others. This included staff who work on temporary contracts, staff from groups with high turnover and staff who work part-time:

Interviewer: "were there any other sort of barriers to you being able to cascade that training?"
HP5: "Well time as always [laugh] and obviously, you know, when you have got people who work part time…" (HP5, midwife, site 11)

"I think the other issue is that they've had a lot of agency sonographers as well, which doesn't help." (HP52, midwife, site 11)

Perhaps unique to the trial setting, but one participant spoke about the time delay from receiving training and using the charts in practice. By the time the charts were in use, some clinicians had forgotten some of the principles:

"I would say there was a bit of a slow uptake at the beginning, people were a little bit confused. There had been quite a big time gap between us doing the actual original training and then actually implementing it. I think I mentioned before that we had had a few delays so that had meant that people had become a bit rusty with the training."
(HP23, sonographer, site 11)

Lots of interviewees spoke about problems they were having with generating and using customised GROW charts. This included not having the information to hand when generating the GROW chart, particularly if this was being done before the clinical appointment, either to save time or because of lacking computer facilities at the clinic:

"It's hard to say really. Not that long if you've got all of the information there, but if you've got somebody and we are looking and their birth weight is not documented from a previous baby and had that baby at another hospital, that sort of thing, you are trying to find out that information and input it. And sometimes it might only be two, three, four minutes but if you are doing that ten times a day then obviously that's increased workload."
(HP 99, midwife, site 8)

Although some colleagues expressed concern over the accuracy of the information, even where it was available:

"Is that women sometimes aren't … they will just pull, you know… I have sat there and sometimes and they just pull random figures out of nowhere and you kind of sit there and just think…well…..not trying to put numbers in their head, but you are kind of just thinking, this is just (laughter) so wrong because I am trying not to influence you when I….but I am, if you know what I mean." (HP7, midwife, site 9)

Table 3.13 - Barriers to implementation of GAP

	Implementation process affected	Barrier	References	Sites
Pre-existing barriers	Full implementation	Problems with access to IT hardware	20 references from 15 participants	7,8,9,10,11
		Staffing shortages	6 references from 5 participants	7,11
		High prevalence of obesity	1 reference	9
		Population of women who don't access antenatal care	3 references from 3 participants	7
Barriers which arose during implementation	Decision to adopt	Lack of clinical support/interest	1 reference	7
	Initial implementation	Colleagues whose working patterns mean they are less available to train	4 reference from 4 participants	7,10,11
		Staff not being dedicated time to do training	9 references from 9 participants	7,8,9,11
	Full implementation	Availability of scan appointments	12 references from 11 participants	7,8,9,10,11
		Limited clinical time – one extra thing to do	9 references from 7 participants	7,8
		Subjective interpretation of chart plots	7 references from 6 participants	7,9,11
		Incomplete understanding of the protocol	4 references from 4 participants	7,8,11
		Unable to find the GROW chart in the notes	8 references from 7 participants	7,9,10,11
		Not having information to hand to generate GROW chart	5 references from 5 participants	7,8,9,11
		Charts in ultrasound software which are different to GROW	5 references from 4 participants	9,11
		Change in practice	7 references from 6 participants	10,11
		Disagreement with customisation	4 references from 3 participants	9,10,11
		GROW charts not integrated into local software	1 reference	7
		Time delay between training and implementation	1 reference	11

Seven participants across all five implementing sites explained common problems in finding the GROW chart amongst the other paperwork in the handheld maternity records:

"as the sonographers every time we do a growth scan we need to plot the weight and the weeks and then we found very difficult to find that piece of paper because we need to go page by page to find exactly …" (HP41, sonographer, site 9)

Including problems where women inadvertently left the charts at home:

"If it's not hole punched and you don't tell them where to put it they could stick it anywhere and you do end up with patients saying, "Oh yes, it's at home". It's not any use at home."
(HP71, midwife, site 7)

More senior colleagues spoke about the junior team members who had not entirely understood the new protocols,

"In the beginning we had quite a lot of midwives that were comparing the fundal height with a scan. So a scan might be on the 50th and then they do a fundal height and it's on the 10th and they're saying look, it's dropped off. No, you can't compare the two; they're two completely different lines." (HP46/58, sonographers, site 8)

"sometimes the midwives are getting confused whether they should be putting a cross or a dot [laughs] and so you are looking at them and you are thinking was that their ultrasound or was that their fundal height? So that gets a bit confusing sometimes."
(HP23, sonographer, site 11)

or about colleagues who were struggling to change their practice or thinking, having conducted it in a different way for many years:

"you just have to get this whole 28 centimetres in 28 weeks out of your head, and often people will measure, and immediately say to the woman, oh that seems fine, and then plot, and there's an issue. Or, ooh, this seems a bit, you know, you know, maybe, maybe there's an issue with the growth, and then plot and it's normal, and then, and then the woman's worried." (SC21, GAP lead, site 10)

"It was interesting when I was doing my training because I did have one um.. midwife who had been taught a different way... um, so to measure from the symphysis rather than fundal um, so that was quite interesting, and she found it particularly different...difficult to change her practice. She tells me she has." (HP5, midwife, site 11)

Interview participants noted two difficulties with using the GROW charts. Midwives commented on how the assessment of serial charts plots was often subjective, as to whether a woman required a fetal growth scan:

"I don't think I particularly understand very well, like I say, how to interpret if somebody... if a plot drops off, how to interpret whether somebody needs a growth scan. I've had experiences... Unless there is an upward trajectory in all of the growth scans I've done, I've always queried it with people and in the instances that I've queried it I've usually had to go through at least one or two people until we get complete clarification on whether or not this person requires a growth scan.
(HP27, midwife, site 11)

Sonographers also had difficulties in disregarding the population-based centile charts which are inbuilt in their ultrasound software, used to assess the growth status of the fetus on the basis of the overall (estimated) fetal weight and on singular fetal biometry e.g., abdominal circumference:

"I had a case today where, on [ultrasound software] charts, the baby is normal, the weight is normal or the measurements are normal, but on the GAP it's above the 95th centile and was a big discrepancy and very difficult for them to understand two different charts with the different measurements and one is trial, one is what we use to base on the measurements and, yes, this is a little bit difficult to explain to patients." (HP41, sonographer, site 9)

"Also, it's sometimes difficult because you can think that the growth is normal during the scan and then input the measurement onto the GAP chart and actually see that it's not within normal range and then you have to bring the lady back in to do Dopplers so that is adding extra time."
(HP23, sonographer, site 11)

Finally, an important barrier to implementation was described by three clinicians at three sites. They described fundamental disagreement with the principles of customisation of the GAP which had hindered buy-in from colleagues:

"But we all knew to be honest it's something that the stakeholders, especially, you know, at the level of fetal medicine consultant, er, we do know about the issues around customisation, and how it has been investigated so far." (SC18, GAP lead, site 10)

"I feel it is a bit of a racist concept to have a customised growth chart that is based on your country of origin when so much mixing nowadays and you do not even take into consideration the husband, the partner's ethnic identity. I just question the whole premise of using the basis of being whether a baby is small or whether a baby is big, so an INTERGROWTH has some evidence for that as well challenging this idea that certain races of people are naturally smaller, so it means... and my point is also that whatever small racial differences there is, how big are the differences? Is it really worth us paying all this money for this package?" (SC32, GAP lead, site 9)

"[Abdominal circumference] is the most important on a growth scan. We still report if that's dropped and we still send them to the day assessment unit and we still do Doppler studies. [...] So we were plotting it, then we had to go back and do Dopplers and we were avoiding the AC which we were trained or to know that that's the most important measurement for predicting fetal growth restriction, but then it was [consultant obstetrician] who told us no, we definitely report our scan as well and then just write 'see GAP chart'."
(HP12, sonographer, site 11)

All barriers which arose during implementation, and most barriers which were present before the decision to adopt the intervention, were identified as impacting on the full implementation phase i.e., during application of the intervention in clinical practice. In contrast, facilitators impacted throughout the implementation process and often arose in response to barriers identified.

3.3.2.4 *Facilitators of implementation*

The most common and important facilitators of implementation are summarised in Table 3.14, including a description of the number of references in the interview data on the theme, the number of participants who spoke about it and the number of sites at which this featured. As for barriers, facilitators are identified as being pre-existing or strategies which

arose during implementation to facilitate the process, and an indication of which implementation process they affected is provided.

Table 3.14 - Facilitators for GAP implementation

	Implementation process affected	Facilitator	References	Sites
Pre-existing	**Decision to adopt**	Belief that GAP was needed	5 references from 4 participants	7,8,9,11
		External pressures to implement	5 references from 4 participants	8,9,10,11
	Full implementation	Collaborative colleagues	9 references from 7 participants	7,8,9,11
Arose during implementation	**Initial implementation**	Provision of the time needed to train	7 references from 6 participants	7,8,11
		Enjoyable, informative training	3 references from 3 participants	7,10,11
	Initial and full implementation	Hands-on project lead	4 references from 4 participants	9,11
	Full implementation	Materials developed to support implementation	11 references from 8 participants	7,8,9,11
		Arranging additional scan slots	4 references from 4 participants	7,8,9,10
		Clear, consistent guidance	6 references from 5 participants	8,9,10,11
		Quick responses to queries from the Perinatal Institute	1 reference	9

With regards to facilitators which pre-existed GAP, interview participants spoke about collaboration between team members to implement, understand and make decisions regarding GAP:

"I think there is some, you know, discussion amongst ourselves about, well, I plotted it here what do you think? 'Cos its…you know you don't when you have got anything new, you don't want it to be wrong and it is different from what we've been doing before…" (HP32, site 11)

"Failing that, we have got a group WhatsApp and we can message and somebody who is generally at a computer will look it up for you." (HP74, midwife, site 7)

There were factors which assisted teams in gaining buy-in for GAP implementation, this included external pressure from neighbouring hospital sites, national incentive schemes and influential people:

"There was already a bit of pressure for us to get on with it and so it was a bit of a no-brainer really." (SC17, GAP lead, site 11)

"then there was the Perinatal Institute's GAP GROW, and our commissioners were very keen." (SC4/7, GAP leads, site 8)

181

"So they are offering a 10% reduction in your CNST † or a maximum of 10% it's going to be graded um… uh.. um and so one of the ten elements is about being compliant with the same baby loss care bundle" (SC23, GAP leads, site 11)

Conversely to the participant quotations noted in the section on barriers, there were also participants at four of the sites who felt that the GAP intervention was needed:

"The UK… so the way we have sold this, so the UK has a problem with the stillbirth rate, London in particular, you know, we are not identifying the babies as we should, we've got to do something, and this is a way to do it. Um, let's get on board. Um... and I think the people that have shown an interest and looked into it a bit more deeply, have asked the right questions and have got on board with it really." (SC17, GAP lead, site 11)

Further facilitators were identified when participants spoke about their training experiences. They noted the usefulness of the engaging and informative one-day face-to-face training session and clinicians at one site spoke about how they had seconded one colleague to roll-out GAP training as her primary role across the site:

"The midwife presenting, she was just quite engaging, and I think there was lots of pictures on the presentation as well and the information was presented in quite a clear manner as well as the evidence. So, I remember coming away from that talk feeling very much in favour of customised charts." (HP26, sonographer, site 10)

"Yeah, I, yeah, we did it very rigorously. I mean I think [Trust GAP lead] was quite impressed. But it's been quite rigorous, so, you know, we've, […] almost daily if not, […] at least once a week, some email coming through with updates, and when we go live, and, and she was very hands-on, so she would come around to the department to see how we were coping, and, that, if we felt that there were any areas with the GAP charts, that she would then take those back to the requesting midwives" (SC24, GAP lead, site 11)

Participants also spoke regarding initiatives taken at a local level to aid GAP implementation. These included the materials which were developed locally to support implementation. Eight participants spoke about these at four of the sites:

"a piece of work that I was doing, […] was, um, making sure we were screening all of our women who were at risk of SGA, and making sure that they had the right growth scans in place. And unfortunately, um, it wasn't being done particularly well. 'Cause there was no mechanism, essentially, in place, um, it was down to individuals that, did they notice? […] So, one of the things that I, I did, er, with, um, one of the consultants is that we designed a request form, that lists all of the um reasons for a lady needing to have those growth scans, er, and that was implemented. […] And, um, that had a significant increase, which we knew it would, because we knew we weren't identifying the right women. Um. And that, I'd say, that single thing has made a huge difference, in terms of, um, identifying, er the ladies who need regular growth scans." (SC19, GAP lead, site 9)

---

† CNST: Clinical Negligence Scheme for Trusts. An organisation which handles clinical negligence claims for all English NHS Trusts. The costs of the scheme are met by membership contributions, which are discounted through compliance with components of the maternity incentive scheme.

"Yes, because we have a flowchart, so that's very helpful, particularly in the beginning because obviously you've got that and then if there's a situation which perhaps doesn't come up that often, we've obviously got the flowchart to look at." (HP99, midwife, site 8)

Another facilitator that participants discussed was the provision of the time need for staff to train on GAP:

Interviewer: "What did you do to encourage people to do the online training? How did you encourage people?"
HP23: "Basically, giving them the time to do it because it's difficult to find the time. Obviously, it's really important and the training is quite long [laughs] and you need to get 100% at the end of the course. So yes, it was basically giving them the time and the opportunity to do it…"
(HP12, sonographer, site 11)

The availability of scan slots was discussed above as a barrier to implementation, but at some sites work was done to extend the availability of scan slots by asking colleagues to work additional hours:

"So currently we are doing early morning scan lists to clear any backlog and we always have an urgent scan list, so an extra list five days a week which helps." (HP71, midwife, site 7)

Participants at four sites spoke about the importance of clear and consistent guidance:

"I think doing the GAP and GROW, it offers us more consistency so that we don't have, theoretically, some consultants having scans at, say, 30 and 34 weeks and others offering scans at 28 and 32, so there seems to be a little bit more structure with it, so I think that's very helpful." (HP99, midwife, site 8)

"Yes. Because at the bottom of each chart it does explain who needs referral, so I definitely used that as a reference point sometimes if I was unsure. So, when I first started using the charts, for instance, if there was a first plot over the 90th centile, at first I was thinking she needs a referral, but actually if you look at the form she doesn't." (HP44, midwife, site 9)

Importantly, one participant thanked the Perinatal Institute for prompt responses to queries during implementation and explained that this was important in helping them to get GAP into practice:

"So, it's just good having the Perinatal Institute and they always come back to us really really quickly. [GAP representative] does anyway, she's amazing! She comes back so quickly, I think she must just live on her phone, or computer! [laughs] […] But it's been really good in that respect, there's been lots of support for us, for, for me and [colleague]."
(SC06, GAP lead, site 9)

3.3.2.5 *How do barriers and facilitators affect strength of implementation*

For reasons described earlier, it was not possible to determine an appropriate overall ranking system for implementation strength which compared each implementing site. However, site 7 persistently scored lower than the other sites on each of the implementation measures and site 8 was probably the highest scoring site, or at the very least one of the most compliant sites.

With regards to the interviews held at these sites, eight interviews were held at site 7 with three lead clinicians and five midwives. The trial team were unable to arrange interviews with sonographers at site 7, despite multiple attempts to do so. Interviews were held with four lead clinicians, two midwives and three sonographers (nine clinical colleagues) in six interviews at site 8, this was because participants were keen to be interviewed in pairs at site 8 and so the trial team facilitated this. As a result, the frequency of references may be reduced – because where topics were covered by one participant in the interview, the other tended to either agree or at least, not also speak on the same topic.

All of the barriers and facilitators identified at these two sites are compared between the two sites, including a comparison of the frequency of references (Table 3.15 and Table 3.16). At site 7 (site with lowest implementation strength), 36 references to implementation barriers were identified compared to 13 references at site 8 (site with highest implementation strength). Interview participants at site 7 spoke more frequently regarding problems with computer access and staff shortages. Having a local population of women who didn't always access antenatal care was also a common issue discussed at site 7. These were not mentioned as issues at site 8. Participants at site 7 also frequently spoke about the low availability of scan slots (despite their local guidance stipulating those high-risk women be offered a total of two to three fetal growth scans, compared to three-weekly scans throughout the 3rd trimester at site 8), existing pressures on clinical time which were exacerbated by GAP and that staff were not given dedicated time to train in GAP. Finally, lead clinicians at site 7 spoke about an unwillingness or lack of interest to implement GAP amongst their senior colleagues, whereas lead clinicians at site 8 generally spoke positively about the intervention. Whilst staff at site 8 did speak about lacking clinical space, including sonography rooms, this did not translate into a difficulty in arranging ultrasound scans. The latter two barriers (existing pressures and time to train) were only mentioned in one interview each at site 8.

With regards to facilitators, interview participants at site 7 spoke about these 22 times, compared to 10 references at site 8. Participants at site 7 spoke frequently regarding materials which had guided the use of GAP although only one participant spoke about this at site 8. Participants at both sites also spoke about collaborative colleagues, provision of the time to train and arrangement of additional scanning capacity, although references to these were infrequent in both cases.

Table 3.15 - Comparison of barriers at the sites with the lowest (site 7) and highest (site 8) implementation strength

		Site 7 (lowest implementation strength)	Site 8 (highest implementation strength)
Barriers which existed pre-implementation	Computer problems	8 references, 6 interviews	1 reference
	Staff shortages	3 references, 3 interviews	0 references
	High staff turnover	1 reference	0 references
	Existing high paperwork load	1 reference	0 references
	Lack of clinical support/interest	1 reference	0 references
	Early concerns regarding potential impact of GAP on clinical service	0 references	1 reference
	Women who don't access antenatal care	3 references, 3 interviews	0 references
	Insufficient clinical rooms (including scan rooms)	0 references	2 references, 2 interviews
Barriers which arose during implementation	Low availability of scan slots	4 references, 3 interviews	0 references
	Limited clinical time	6 references, 6 interviews	2 references, 1 interview
	Staff not given dedicated time for training	4 references, 4 interviews	1 reference
	Incomplete understanding of protocol	1 reference	2 references, 2 interviews
	Unable to find GROW chart in notes	1 reference	0 references
	GROW charts which take time to generate	1 reference	1 reference
	Colleagues who haven't trained	1 reference	0 references
	Lack of system prompts to generate neonatal centiles	1 reference	0 references
	Midwives not allowed to arrange serial growth scans (doctors' role)	0 references	1 reference
	Subjective interpretation of charts	0 references	1 reference
	Lack of continuity of carer	0 references	1 reference

Table 3.16 - Comparison of facilitators at the sites with the lowest (site 7) and highest (site 8) implementation strength

		Site 7 (lowest implementation strength)	Site 8 (highest implementation strength)
Facilitators which existed pre-implementation	Collaborative colleagues	2 references, 2 interviews	1 reference
	Belief that its needed	1 reference	1 reference
	Sufficient computer access	2 references, 2 interviews	0 references
	Feeling passionate about GAP	1 reference	0 references
	Performance of neighbouring units	1 reference	0 references
	Having administrative days at work	1 reference	0 references
	Lack of consistency with neighbouring hospitals	1 reference	0 references
	External motivators	0 references	1 reference
Facilitators which arose during implementation	Materials which ease the use of GAP	6 references, 2 interviews	1 reference
	Provision of time to train	2 references, 2 interviews	2 references, 1 interview
	Additional scan slots	1 reference	2 references, 2 interviews
	Local promotional materials	1 reference	1 reference
	Enjoyable / useful training	1 reference	0 references
	Team leader feedback	2 references, 2 interviews	0 references
	Structured protocol	0 references	1 reference

3.4 DISCUSSION

3.4.1 Summary of the key findings

The GAP intervention was implemented through the DESiGN trial, in the context of the UK having a national policy to tackle a higher rate of stillbirth than that seen in other Western European countries. Implementing staff were aware of the national stillbirth problem but didn't always express belief that it was relevant to them locally. Politically, tackling stillbirth was high up on the agenda of the UK Secretary of State for Health,[332] and numerous national initiatives were in place alongside GAP in an attempt to investigate and reduce the high stillbirth rate.[217,267,331] Staff felt a pressure to implement GAP to 'keep up' with practices nationally, or in neighbouring sites. Despite stillbirth being prominent on the political agenda, representatives from most sites spoke about financial pressures and staff shortages within their service which made it difficult to fully implemented initiatives aimed at tackling stillbirth.

Through an assessment of implementation strength, I have identified elements of both high and low implementation quality, including a significant variation between cluster sites. The overall impression of implementation strength appears to be associated with the trial primary outcome, antenatal detection of the SGA fetus, although this is difficult to confirm with so few clusters.

With regards to implementation fidelity, there was significant variation noted particularly in the comparison of local guidelines to that recommended by GAP. One cluster site had low guideline concordance that was mostly relevant to women at high risk of SGA, with significant differences in how the site defined risk and in how the site provided care for these women. A common reason for cluster sites making adaptations to the recommended guidelines was concern about the ability to provide scans to all the women that GAP suggested. These concerns arose either from a local lack of sonographers or physical scan rooms, conflicting interests with regards to balancing the needs of women accessing maternity versus gynaecology services, or from assessments of the local population which suggested that risk factors were common, and the site didn't have the resources to provide serial ultrasounds for all. These concerns occurred in the wider context of a national shortage of sonographers (and of other clinical staff), a national policy which also recommended an offer of additional ultrasound for women concerned about fetal movements and universal concerns that local populations were high risk for SGA and stillbirth. Despite these, two included sites did develop guidelines which were highly concordant to those recommended by GAP.

The second measure of fidelity examined how local risk assessment guidelines were applied in practice. 84.9% of all women were appropriately risk stratified according to GAP recommendations and 3.2% additional women were appropriately risk stratified by local protocols (where risk stratification was not correct by GAP). This relatively high score is likely to have been aided by the almost universal introduction of risk assessment aids and guidance that was noted to have been clear and consistent across all sites.

It was reassuring that all sites were able to achieve the training target for face-to-face training, however only one site achieved the e-learning target. The qualitative interviews identified that barriers to achieve this target included a lack of dedicated time to conduct the training and occupational characteristics which made some staff groups harder to reach (part-time workers, temporary staff and high staff turnover). Conversely, staff reported enjoying training and found it useful; those who had been provided with time to train or had provided their junior colleagues with time to train, noted this as a facilitating factor.

Nearly 9 in 10 of all women included in the notes review had evidence of being reached by the intervention, assessed by the presence of GROW charts in their maternity records, although this measure did vary highly between sites. Given that staff described problems with finding GROW charts in the records or examples where women had removed the charts from their notes and left them at home, it is quite possible that this represents an underestimate. Staff frequently reported barriers to the production and printing of GROW charts, including problems with access to computers or printers, and difficulty in instantaneously accessing the information needed to generate a customised chart, but this has not had a major detrimental impact on reach, which may have been aided by staff members' belief that GAP was needed, or initiatives implemented to aid uptake, including a frontline GAP lead and materials (e.g. posters) developed to prompt clinicians.

Dose is the domain of implementation strength in which sites performed least well, although again there was definite variation seen between sites. Overall, fewer than one in three women deemed low risk by the clinician had at least the minimum number of fundal height measurements plotted on their GROW charts. Women who give birth at term are expected to have a minimum of four fundal height measurements, the median number was lower than this at most sites. Since one significant way in which GAP differs from standard care is the screening policy for women at low risk of SGA (both policies employ similar screening strategies for women deemed high risk of SGA), this finding may explain the lack of difference identified in the primary outcome between the sites allocated to intervention or standard care. I was correct in hypothesising that this dose target might be more commonly achieved in nulliparous women because the frequency of recommended fundal height measurements fits better with the NICE antenatal care schedule for nulliparous

women than multiparous women. Interview participants reported that barriers to compliance with plotting fundal heights on the charts GROW charts were charts either missing or left at home, socially vulnerable women who attended for less antenatal care and the difficulty in changing behaviour for midwives who had been practising differently for years.

The second measure of dose covers referral for a fetal growth scan in low-risk women who have a deviation from the expected trajectory of fundal height plots. Half of women had evidence of a possible or definite fundal height deviation, and just over half of these were referred for a growth scan in response to this deviation. This low rate of referral may explain in part the trial finding of no difference in the rate of the primary outcome between sites allocated to the intervention or standard care. This measure of dose is complicated by two factors. Firstly, that identification of a deviation in the trajectory is subjective, this was also noted by participants in interviews who spoke about referring to colleagues for a second opinion when the interpretation was unclear. This is demonstrated also by the fact that a higher percentage of the women with a definite fundal height deviation on their GROW charts (compared to a possible or definite deviation) were referred for a growth scan (67.6%). Staff collaboration in this area was noted as a facilitating factor. Secondly, women were referred during pregnancy for a fetal growth scan for other indications during the trial, for example if they were concerned about a change in fetal movements. In some cases, it is possible that clinicians were aware of another referral for a scan and so did not refer twice, but there was no documented evidence in the antenatal notes that this was the case for any of the women who were not referred for a fundal height deviation.

The final measure of implementation dose pertains to the frequency of ultrasound scans that were plotted on GROW charts in women at high risk of SGA. Less than one in ten high-risk women had at least the minimum recommended number of fetal growth scans in which an EFW was measured and plotted on the GROW chart. Only three-quarters of EFWs were plotted on GROW charts, but the proportion of women who had at least the minimum number of growth scans with/without plotting of EFW on the GROW chart was only marginally greater than the proportion with plotted scans. This is partly because three of the five cluster sites had local policies recommending that scans were conducted at specified time points with a maximum of three scans per pregnancy, compared to the GAP recommendation of scanning a minimum of three-weekly from 28 weeks' until birth. These policies were implemented because of local GAP leadership concerns that the sites did not have the physical, staffing or financial resources to support a policy of 3-weekly scans. Not surprisingly, opening additional sonographer appointment slots was noted to be a facilitating factor at four of the five sites. Sonographers reported problems in finding the

GROW charts, but also concerns about the flow of the plotting process, particularly where they were used to automated chart generation in their inbuilt ultrasound reporting software.

3.4.2 Interpretation of the findings

GAP was implemented through the DESiGN trial at a time where there was the political motivation, but not the financial resource to support it. It was also implemented at a time where all sites in the UK were motivated through national audit and financial incentives (in the form of reduced insurance premiums),[333] to commence surveillance of fetal growth and target reductions in stillbirth; this also affected the DESiGN trial clusters allocated to standard care.

The DESiGN trial identified that GAP was not more clinically effective than standard care at the antenatal detection of SGA fetuses.[320] There are two possible explanations for this finding: that the intervention is truly not more effective than standard care, or that the intervention was not implemented with sufficient strength to impact the outcome. Whilst the cluster sites allocated to the intervention were implementing GAP, the cluster sites allocated to standard care were externally motivated to implement local policies aimed at fetal growth surveillance, by publication of the Saving Babies' Lives Care Bundle.[217] The key difference was that the standard care sites did not use the GAP-standardised protocol (although may have used the same risk stratification tool which was published by NHS-England and adopted by GAP) and did not use customised fetal growth charts (but did use alternative fetal growth charts for high-risk women, and alternative strategies for assessing fundal height growth for low-risk women).

Because GAP is a complex intervention, the components of which are implemented together, it is difficult to conclude which specific elements of the implementation strength measure, and therefore which intentional adaptations or aspects of low strength, contribute with most weight to the overall effectiveness. One common modification to GAP guidelines was for sites to reduce the frequency with which they offered fetal growth surveillance ultrasound scans from the recommended 3-weekly scans for women at high risk of SGA. Whilst RCOG guidelines on the Investigation and Management of the Small–for–Gestational–Age Fetus recommend that scans not be performed more frequently than 3-weekly to minimise the false positive diagnosis rate,[1] there have been no randomised trials comparing a scan policy of 3-weekly or 4-weekly surveillance for the detection of SGA or avoidance of stillbirth. Theoretically, the less frequently that scans are conducted, the longer it will take to identify a growth restricted fetus and the higher the chance of stillbirth. Hawe (2009) argues that allowing standardised intervention approaches to change across local contexts

may aid in achieving fidelity to its intended functions.[334] GAP was commonly adapted at cluster-level according to local context, a method used by sites to target the areas in which fidelity is optimised given limited resources. As May's general theory of implementation (2013) states, successful implementation is the result of the action of agents, but that their actions are shaped by capacity, or the resources available to them.[335] This is likely to have been a factor in this trial.

There is further evidence that guidelines for the detection of SGA at other maternity services in the UK also have low fidelity to GAP recommendations. In the SPiRE study, 19 maternity units were recruited to assess outcomes and implementation of the Saving Babies' Lives care bundle, using a pre- and post-implementation comparison.[336] Fifteen of the 19 included sites were enrolled in GAP, none of which participated in the DESiGN trial. In a review of the guidelines on screening for and management of SGA in the SPiRE study, Lau et al (2020) identified that none of the guidelines had full agreement with recommendations on serial assessment of fetal growth by ultrasound for high-risk women, plotting EFW onto a chart, screening for SGA using fundal height measurement (2-3 weekly from 26-28 weeks) in low-risk women, or ongoing audit and report of rates of SGA and its detection. Only 10 of the 19 sites were fully adherent to the NHS-England risk stratification algorithm.[337] These findings suggest that the variation in fidelity to GAP recommendations may be more widespread in the UK and is not limited to DESiGN trial sites.

There is no literature which evidences the training target of >75% staff from each professional group for both types of training as an implementation strategy for this protocol. It is therefore unclear what the effect is likely to have been at sites which did not achieve this target. Furthermore, we do not have data assessing the quality of training received. A Cochrane review of randomised trials comparing e-learning to traditional training methods found that e-learning made little or no difference to either patient outcomes or on healthcare professionals' behaviours, knowledge or skills.[338]

Demonstrating high levels of implementation reach, most women had a GROW chart. Whilst these GROW charts are one of the major ways in which GAP differs from standard care, this alone was not sufficient to make the intervention more clinically effective.

Implementation dose is the measure in which the cluster sites consistently scored the lowest. This affected both women at low- and high-risk of SGA. In the case of high-risk women, this is strongly linked to implementation fidelity and modifications i.e., at some sites women received less than the expected number of ultrasound scans because the local policy recommended less frequent surveillance. Serial fundal height measurements for women at low-risk of SGA have been associated with an improvement in predictive accuracy

for SGA,[339] and serial monitoring of fetal size by ultrasound is a direct measure of fetal growth hence should facilitate earlier diagnosis of SGA.[1] It therefore follows that, if implemented as intended, this intervention component may be effective in improving the clinical effectiveness of the intervention. However, the same methods were used by control sites, with the only difference being the standard by which the sites judged the measurements (population references versus customised standards, or no fundal height charts at some control sites) and so, this may not have the desired effect if similar methods were implemented with the same strength in both arms of the trial. It is important to note that the dose of SGA surveillance for women at high risk of SGA was not measured and is therefore not known at control sites.

Through qualitative interviews, staff identified numerous barriers and facilitators to implementation. A common barrier to implementation was on-site resource with reference to computer hardware, clinical space, training time and sufficient staffing for clinical demand. These resource barriers exist because of the socioeconomic context of implementation and are likely to have impacted on all domains of implementation strength. Overcoming this barrier is likely to be integral to improved implementation and sustainment of the GAP intervention. Other barriers, such as the inability to find the GROW chart in the notes, require simple modifications to practice, such as a policy of placing the chart in a particular section, or printing it on coloured paper. Similar modifications were already implemented by resourceful staff at some cluster sites in response to early barriers to implementation, described in this chapter as materials which facilitated implementation.

3.4.3   Strengths and limitations

The strengths in this evaluation lie in the mixed methods approach to examine both the quality of implementation and the associated factors. Whilst process evaluation frameworks are not new, guidance and frameworks for hybrid evaluations of implementation process and clinical effectiveness of interventions are still in their infancy.[265,340] Such hybrid approaches are absolutely essential to understand both whether an intervention is effective and what mechanisms lie behind this effectiveness, or in the case of the DESiGN trial – non-superiority. The process evaluation nested within the DESiGN trial, has enabled a detailed study of the implementation of the complex GAP intervention, comparing this implementation to that which was intended, and allowing the generation of hypotheses as to whether the intervention was ineffective because of a failure in the intervention itself, or because the strength of implementation was low.

The characteristics of the subgroup of women included in the notes review are comparable to the wider group of women included in the whole trial comparison period for

assessment of the primary outcome, including similar birthweight and gestational age of the baby. The main difference is that women in the notes review were more often of African origin and less often of British European origin. The notes review was conducted on the notes of women who gave birth during the trial outcome phase and since these were randomly chosen from postnatal stores, they are expected to be mostly reflective of the actual practice. If there were a bias in these notes collected, it would be that notes were possibly not available for births in which the outcome was poor and the care required investigation; these outcomes are rare. If there was evidence of poor quality of care associated with an adverse outcome, I would expect the inclusion of these notes to lower the overall strength of implementation determined from the notes review. The extent to which the review was limited by missing data is unknown, for example, if a high risk woman did not have any risk factors recorded in the notes, I assumed her to have been low risk – this may have caused a small bias in the assessment of appropriate risk stratification, but given that the dose domains were then conducted according to risk status as determined by the clinician, this is unlikely to have affected other assessments. The conclusions are also limited by the absence of notes reviews carried out at control sites, knowledge of the dose of surveillance for common methods, e.g., serial ultrasound scans for women at high-risk of SGA, would inform whether these implementation problems were restricted to GAP.

The analysis is limited by the absence of semi-structured interviews conducted with pregnant women exposed to the intervention. These were planned in the trial protocol to assess the acceptability of the intervention, but the trial team encountered difficulties in recruiting pregnant participants, only one woman was identified from the four cluster sites approached over a 4-month period. A previous systematic review on women's views of pregnancy ultrasound identified that women generally consider this intervention to be acceptable,[341] but the whole GAP intervention including the use of customised fetal growth charts has not previously been explored qualitatively with women. Also absent is the opinion of frontline obstetricians, who play a significant role in providing antenatal care to women at high-risk of complications, including women at high-risk of SGA. I do not think that interviews with obstetricians would have identified many additional barriers or facilitators to implementation, since these are barriers which are also likely to have faced midwives or sonographers, although it would be necessary to interview a small sample of obstetricians to test this hypothesis and ensure the data were truly saturated.

The application of the CICI framework to this trial-based process evaluation was novel, but we found it to be generally well-suited to this analysis. It enabled us to consider the interplay between context and implementation processes and strength in detail, however, the framework was not perfectly aligned with the needs of this evaluation and additional

codes were required to enable study of feasibility and to easily identify implementation barriers and facilitators from the coded data; these are all widely acknowledged to be important questions of process evaluation.[304,306]

Another limitation is that I have not been able to statistically correlate implementation strength with intervention effectiveness because of the low number of clusters available for a sensitivity analysis. The formal comparison of clinical outcomes at cluster sites that achieved high scores for each domain of implementation strength to those with lower scores would be a useful analysis to generate hypotheses about which elements of the complex intervention have most weight in affecting the outcome. A simple way to do this might be to repeat the primary outcome analysis following exclusion of site 7, which achieved lowest scores for all domains of implementation strength, but this would also likely lose statistical power to identify a difference in clinical effect. This is not the first study in which it was concluded inappropriate to study this using anything other than descriptive methodology.[329] Hargreaves et al (2016) also expressed caution about the application of arbitrarily determined weights to component measures of implementation strength.[327] Furthermore, as noted in the MRC framework on process evaluation for complex interventions, such data integration is expected to be challenging, with significant limitations of statistical power, where assessment of implementation strength is based upon measures collected at only a few sites.[265]

The degree to which the data can be candidly presented to support the analytical findings is somewhat limited by the risk of revealing the identity of the participant or losing anonymity for the site. Participants were assured through a formal consenting process that we will endeavour to prevent their identification. Whilst I have taken significant steps to ensure anonymity for all interview participants, including assigning a pseudonym for both person and site, and removing identifiable information from quotations, a small risk to confidentiality remains because of the small number of participants and sites.

With regards to the generalisability of these findings, Bonell et al (2006) recommend assessment of generalisability to include evaluation of whether the intervention can be delivered elsewhere (to include assessment of feasibility, coverage within a health system and acceptability to the target population) and assessment of whether there is potential for recipients to benefit from the intervention.[342] Furthermore, assessment of generalisability must also consider the context of implementation in the study and the diversity of the sample population, compared to the context and population characteristics in the settings in which generalisability may apply. Implementation of the GAP-GROW programme, or development of GROW charts has previously been reported across the UK and in high-income countries in Europe, Australasia and North America.[78-80,82,83,85,86,270] At the time of

writing (June 2020), GAP-GROW has already been taken up by 78% of Trusts or health Boards within the UK.[270] The diversity of the implementation population was described in section 2.6.2. Many of the barriers and facilitators identified exist nationally,[343] and some may also be relevant to other contexts, particularly in publicly funded health systems. Examples of national barriers include the national shortage of sonographers and restricted funds across the NHS. National facilitators include a belief that GAP was needed because of a national concern that stillbirth rates are higher in the UK than comparable Western countries, and external pressures to implement following a national requirement to be compliant with the Saving Babies' Lives care bundle.[176,217] Consideration of both these and local barriers or facilitators e.g. nomination of a full-time GAP implementation lead, are relevant to both a national and an international audience when making decisions to implement GAP-GROW, or exploring strategies to support its implementation.[344,345]

3.4.4 Implication of the findings

The lack of clinical superiority of GAP over standard care identified in the DESiGN trial was either because of the simultaneous implementation of a national care bundle targeting stillbirth, including protocols aimed at improving the detection of SGA in both trial arms, because GAP is truly not more clinically effective than standard care even when implemented as intended, or because GAP was not implemented entirely as intended. It is quite possible that the truth lies in a combination of these reasons. If it is the former, then a cost-effectiveness analysis is essential to guide clinical leadership teams in their choice of clinically equivalent interventions to increase the detection of SGA.

If incomplete implementation was the cause, this has implications for the sustainability and spread of GAP – namely that the intervention must be properly resourced in terms of time away from clinical duties for staff to be trained, increased availability of ultrasound scan appointments so that women at high-risk of SGA can be offered the full range of scans, and removal of barriers to computers or printer access. Furthermore, integration of the intervention into current practice could be used to remind staff to engage with it (e.g. through computer prompts in the maternity records systems), prevent the wastage of time when looking for the loose paper GROW chart (either through availability of electronic customised charts, or a standardised place for storage in the maternity notes) and ease the flow of clinical care (e.g. by incorporating the customised centiles into electronic ultrasound-derived fetal biometry assessments).

It is unlikely that a further trial will be possible in the UK, to study the clinical effectiveness of GAP following complete implementation and to consider the weight or independent effect of each component of the GAP intervention. This is because GAP has

already been rolled out to approximately 80% of hospital trusts and the DESiGN trial was seen as the last opportunity to test this intervention in a trial setting. However, full implementation of GAP could be assessed internationally, where the intervention has not already been widely spread. In this case, a quantitative process evaluation of implementation is key to understanding the mechanisms behind any clinical effect and ideally, this would be studied in a large number of sites to facilitate sensitivity analyses examining the effect of low or high implementation strength in each of the implementation domains (fidelity, dose and reach), and to include a quantitative assessment of compliance with protocols in both control and intervention sites.

With regards to the relatively novel method of studying implementation strength both qualitatively and quantitatively, this also has implications for future research. This analysis was integral to the DESiGN trial and essential to the generation of hypotheses on the mechanisms behind the clinical finding of equivalent effect for both GAP and standard care. The methods used add to the currently scarce literature on the assessment of implementation strength.[346] Formal guidance on how to assess weight of individual components of implementation strength, particularly in complex interventions, and how to combine qualitative and quantitative scores would aid global assessments in hybrid implementation-effectiveness trials and ensure that analytical practices were standardised.

Hybrid-effectiveness trials and associated frameworks and guidance have been developed, driven by the need to understand if lack of effect is attributable to the intervention itself, rather than to poor implementation, and to replicate and scale up the intervention and implementation strategies in different health system contexts if promising.[285] The sample size of the DESiGN trial was determined following a statistical power calculation for the primary clinical outcome. The implication of a small number of cluster sites on the potential conclusions to be drawn from the process evaluation was not considered in the early planning of the trial. These conclusions have been limited by the variation in implementation, in combination with a small number of cluster sites. This is an important learning point for the future of hybrid clinical-effectiveness process-evaluation randomised trials.

3.4.5   Conclusion

The findings of equivalence of clinical effectiveness between GAP and standard care in the DESiGN trial may be explained by a combination of the simultaneous implementation of a national care bundle targeting fetal growth surveillance, and by variable strength of implementation of the intervention that sometimes had low concordance with guidelines from the Perinatal Institute. This was affected by intended adaptations, which were largely

driven by the availability of key resources, and by unintended deviations from protocol, driven both by resource availability and by the intervention not being fully integrated into the course of usual care.

Clinical leadership teams will require information on the barriers to implementation as intended by the provider, and the results of the cost-effectiveness evaluation, to aid decision making regarding the sustainment and spread of GAP which has been found to be equally effective to standard care in the antenatal detection of SGA.

4 COSTING MATERNITY CARE IN ECONOMIC EVALUATIONS OF INTERVENTIONS: A SYSTEMATIC REVIEW

I have previously published a version of this chapter (full manuscript is free to access through the link provided):

4.1 Introduction

Healthcare economic evaluations are pertinent components of healthcare research and quality improvement, informing policymakers on the cost effectiveness of new interventions and thereby assisting in decisions regarding their uptake. The UK NICE bases its recommendations on the implementation of new interventions according to both their clinical efficacy and cost-effectiveness.[309] The number of cost-effectiveness evaluations published annually in obstetrics and gynaecology has increased since 2000, with the majority being conducted alongside a clinical trial.[347]

Internationally, maternity care is funded using different payment models, including itemised bills and payment using composite 'bundled' costs.[310,348,349] Bundled pricing describes a model where a single price is used to cover a full package of care for a specific indication. Bundled costs can be uplifted locally using factors which account for geographical variation in the cost of providing care and by the level of comorbidity or complexity of the woman and her pregnancy, but are not explicitly changed by differences in utilisation of maternity services.[310,350] Such models are often used because they are easier for hospitals to manage, allow flexibility within the pathway and are intended to encourage improvements in care including standardisation of evidence-based care.[348] Bundled payments were introduced by the Medicaid initiative (United States of America) to also reduce interventions which are not medically indicated and potentially harmful.[349] Bundled payment models are also used for non-maternity indications.[351]

Whilst bundled costs represent the cost of a woman's care to a commissioner or insurer, to the hospital they only reflect the average cost of women who experience the same level of complexity in pregnancy. Bundled payments present difficulties in estimating small changes to the overall cost of a woman's maternity care because of a new intervention, because the cost of her care is estimated as a composite which is often not affected by small changes in resource use (e.g., an additional antenatal appointment or ultrasound scan). This represents a significant limitation for using these tariffs in estimating costs within economic evaluations, which seek to identify the true clinical resource impact of initiatives aimed at improving quality of maternal and perinatal care, as distinct from potential financial impacts to commissioners or insurers.

In England, maternity services are currently funded by regional commissioners, who determine payment using reference costs from a national tariff and uplift these locally using Market Forces Factors that account for geographical variation.[350] The tariff was determined following costing exercises conducted by NHS Trusts across the country. Using the most recent tariff (2018-19), antenatal, intrapartum and postnatal care attract a composite

payment that varies according to the women's comorbidities or complexities of her pregnancy, but not by specific utilisation of maternity services.[310] Whilst the 2018-19 national tariff was designed to simplify payments by commissioners to hospitals for maternity care, it presents difficulty in estimating small changes to the overall cost of a woman's maternity care as a result of a new intervention, because the costs of her care are estimated as an average composite. The 2015-16 tariff was more suited to estimating costs for individual activities but has now been superseded (Table 4.1).

Table 4.1 - Summary of UK Department of Health reference costs 2015-16 and 2018-19

Activity costed	Department of Health Reference costs 2015-16 [284]	Department of Health Reference costs 2018-19 [310]
Estimated cost for antenatal appointment	Non-consultant-led outpatient attendance for obstetrics - £117.16 Consultant-led outpatient attendance for obstetrics - £141.95	Tariff for all care within an average maternity pathway is costed as a composite, with higher prices paid for higher levels of obstetric complexity or presence of medical co-morbidities: £1,019-£2,713.00.
Estimated cost for attendance to maternity day unit/triage	£259.48 (NZ16Z - Day case)	
Estimated cost for an antenatal inpatient admission	£1,133.76 for first day and £184.67 for every day after (NZ16Z - Antenatal Routine Observation)	
Estimated cost for an ultrasound scan	Sonography-led: £109.07 / scan (NZ21Z) Specialist-led: £151.15 / scan (NZ22Z)	
Induction of labour	£537.90-£850.96 in addition to cost of normal vaginal delivery	Upgrades the price for vaginal birth to a higher level of complexity.
Epidural	£537.90-£850.96 in addition to cost of normal vaginal delivery	
Unassisted vaginal birth	£1,832.10-£2,259.20	£1,957-£3,357, depends on level of complexity and additional activity
Assisted vaginal birth	£2,441.01-£3,204.83	£1,957-£3,357, depends on level of complexity and additional activity
Elective Caesarean section	£3,383.41-£5,042.31	£3,357
Emergency Caesarean section	£4,594.77-£6,867.95	
Repair 3/4th degree tear	£242.06-£762.75 in addition to cost for normal vaginal birth	£0.00 - £1,400.00 (upgrades vaginal birth price to £3,357, if not already being paid)

4.1.1 Objective

The objective of this systematic review was to summarise the current evidence on the costing of resource use within UK maternity care, to facilitate the estimation of incremental resource and cost impacts potentially attributable to maternity care interventions including the cost of GAP when studied in the DESiGN trial.

4.2 METHODS

This review was registered on PROSPERO during the data collection stage (registration number CRD42019145309) and has been reported with reference to the PRISMA statement and checklist (Appendix section 10.12).[315]

4.2.1 Search Strategy

A systematic review was conducted in August 2019 of the Medline, Health Management Information Consortium (HMIC), the NHS Economic Evaluations Database, the Cumulative Index to Nursing and Allied Health Literature (CINAHL) database, and NICE guidelines for economic evaluations of maternity care and the economic impact of research interventions. Search terms included free text and expanded synonyms for terms relevant to economic evaluations (e.g., cost-effectiveness, price tariffs) and to pregnancy healthcare (e.g. midwife, maternity, pregnancy). An exemplar search strategy for the Medline database is included within Box 4.1. The search was limited to papers written in the English language (because the perspective was that of UK maternity care providers) and published since 2010 (to ensure the cost estimates were recent and more reliable when inflated to current prices).

Box 4.1- Search strategy for Medline database

Database: Medline
((exp PREGNANCY/ OR exp "PREGNANT WOMEN"/ OR exp "HOSPITALS, MATERNITY"/ OR exp "PRENATAL CARE"/ OR exp "MATERNAL HEALTH SERVICES"/ OR exp "BIRTHING CENTERS"/ OR exp PARTURITION/ OR exp PREGNANCY/ OR exp "HOME CHILDBIRTH"/ OR exp "NATURAL CHILDBIRTH"/ OR exp "TERM BIRTH"/ OR exp "PERINATAL CARE"/ OR exp "POSTNATAL CARE"/ OR (pregnan* OR matern* OR antenat* OR prenat* OR intrapartum OR postnat* OR birth OR obstetric* OR midwi*).ti) AND (exp "COSTS AND COST ANALYSIS"/ OR exp "COST-BENEFIT ANALYSIS"/ OR (health economic*).ti OR (economic evaluation).ti OR (cost effectiv*).ti OR (cost benefi*).ti)) [DT 2010-2018] [Languages English]

For inclusion in the review, it was predetermined that manuscripts must be full reports of primary research studies or systematic reviews (including those in NICE guidelines), where an economic evaluation of an antenatal or intrapartum intervention was performed and assessed within the UK context only. A UK context was chosen because it is well-established internationally that different countries vary in their approach to providing maternity care, the type of clinical resource inputs used to deliver specific types of clinical activity and in terms of the efficiency with which this is delivered. A review on the international economics of childbirth identified no accepted cost which was translatable across international settings due to differences in national clinical practices, outcome

definitions and healthcare funding mechanisms.[352] Given that this review was undertaken to inform an economic evaluation of an intervention in the UK, it was therefore taken as reasonable to assume that the costs of maternity resource-use estimated from other countries would not be appropriate in the UK.[353]

It was also necessary that the papers reported unit costs for any of the items from a list of *a-priori* key activities within the maternity pathway (see Table 4.2) because these were common and presumed high-cost maternity activities that may commonly vary following introduction of interventions in maternity services. There were no specific exclusion criteria.

I screened the titles and abstracts; the remaining full texts were reviewed in full against the inclusion criteria by both a research assistant (Dr Louisa Delaney) and I.

Table 4.2 - Key activities costed within the maternity pathway

Antenatal Activity	Intrapartum Activity	Postnatal/Neonatal Activity
• Midwife-led antenatal appointment • Obstetrician-led antenatal appointment • Glucose tolerance test • Attendance to day assessment unit/triage • Antenatal inpatient admission • Sonography-led ultrasound scan • Consultant-led ultrasound scan	• Induction of labour • Augmentation of labour • Epidural • Normal vaginal birth • Instrumental vaginal birth • Elective Caesarean section • Emergency Caesarean section • Repair 3rd/4th degree tear • Manual removal of placenta • Treatment of postpartum haemorrhage (500mL-1500mL) • Treatment of major obstetric haemorrhage (>1500mL) • Examination under anaesthesia for haemorrhage	• Maternal stay in postnatal ward (with/without baby) • Maternal stay in high dependency unit • Maternal stay in intensive care unit

4.2.2 Data Extraction

Data were extracted from each paper on the cost perspective taken by the study, year and methodology used for costing the resource use, by the same research assistant (LD) and I onto a pre-specified study spreadsheet. Unit costs quoted for any of the key activities listed in Table 4.2 were collected. Costs were inflated to the 2018/19 financial year using the Department of Health's Pay & Price Series for financial years 2008/09 - 2015/16 and the NHS Improvement Economic Assumptions for years 2016/17 to 2018/19.[312,313]

4.2.3 Assessment of study quality

Assessment of study quality was performed using the Quality of Health Economic Analyses (QHES) checklist.[354] This was designed by health economists and validated by both

clinicians and economists with the aim of providing a tool suitable to evaluate all common types of health economic analyses, by reviewers of either profession. As per the case study in the original QHES paper, papers have been assessed as high (≥75/100 points), medium (50-74 points), or low (<50 points) quality.

4.2.4 Data Analysis

For the cost of each activity within the maternity pathway, the range, mean, standard deviation and relative difference between the minimum and maximum estimates were reported. The distribution of costs was represented graphically on a scatter plot, with data divided by cost source for comparison (national guideline, review article and primary research study).

Simulated low and high-risk patients were agreed through consensus of the DESiGN trial team with reference to risk stratification guidance produced by NICE, to demonstrate the difference in cost estimates for common exemplar clinical scenarios when applied across the whole maternity pathway.[355] The planned low risk pregnant woman was 35 years old and multiparous, having had two previous vaginal births with no medical or obstetric complicating factors. She had an uncomplicated pregnancy and spontaneous vaginal birth, followed by a 6-hour postnatal discharge. The planned high-risk woman was 42 years old, nulliparous having conceived with in-vitro fertilisation. She develops pre-eclampsia in the 35th gestational week and is induced at 37 weeks' gestation. She labours with an epidural but requires an emergency caesarean section for presumed fetal compromise.

4.2.5 Sub-group analyses

Two post-hoc sensitivity analyses were conducted to determine (i) the extent to which removal of low-quality primary research papers reduced the variability in cost estimates and (ii) whether the variation in cost estimates for each named activity was caused by the range of different cost perspectives used. For the first analysis, costs derived from published papers deemed to be of low quality were removed and the effect on the mean and range costs per activity described. For the second analysis, cost estimates for each activity included in the primary research articles were presented graphically, stratified by the perspective.

4.3 RESULTS

Of 5,081 papers identified in the electronic database search, 3 economic evaluations in relevant NICE guidelines and 10 publications identified from handsearching systematic reviews, 848 were duplicates and 4,080 were excluded through screening of the titles and abstracts, leaving 140 full texts for screening. Following exclusion of papers which didn't meet the inclusion criteria, 37 papers were included in the final review, including 27 primary research articles, seven review articles, and three economic evaluations from NICE guidelines. This is represented diagrammatically in Figure 4.1. The characteristics of the included studies are detailed in Table 4.3.

Figure 4.1 - Study Selection Process
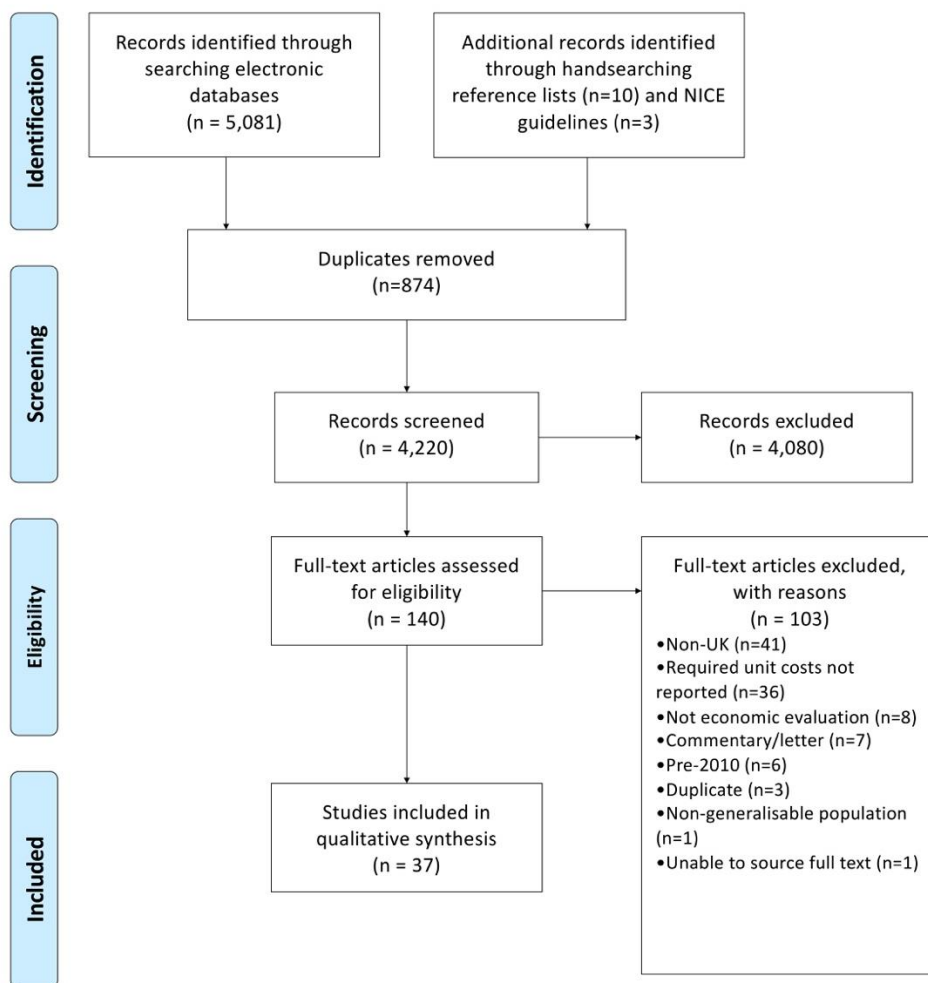
Table 4.3 - Characteristics of included studies

	Article	Perspective	Original cost year	Methodology
Economic evaluations for national guidelines	NICE - Hypertension in Pregnancy [356]	NHS	2010	Used NHS reference costs and costs published as parts of other trials.
	NICE - Intrapartum Care [355]	NHS	2014	Uses the bottom-up costing calculations from a primary research study (Schroeder E, 2012) and revised the cost calculations by consensus with an expert committee.
	NICE - Diabetes in Pregnancy [357]	NHS	2015	Used a combination of NHS reference costs and a bottom-up costing exercise
Review articles	Mistry, H (2013) [358]	NHS	2009-10	Health Technology Assessment Uses NHS costs and costs from a literature review of primary research studies.
	Deshpande, SN (2013) [359]	NHS	2011	Weighted averages from NHS reference costs.
	Thomas, CM (2013) [360]	NHS	2011	NHS reference costs applied to a cost-effectiveness model derived through published data sources on resource use.
	O'Donnell, A (2016) [361]	NHS	2012-13	Health Technology Assessment uses NHS reference costs and calculations from the Personal Social Services Research Unit.
	Alfirevic, Z (2016) [362]	NHS	2012-13	Health Technology Assessment using a de novo decision model and NHS reference/manufacturer costs inputs
	Farrar, D (2016) [363]	NHS	2013-14	Review article for Health Technology Assessment. Uses cost from NHS reference costs, NICE guidelines and the Personal Social Services Research Unit.
	Gallos, I (2019) [364]	NHS	2016	Health Technology Assessment using NHS reference costs and drug costs from the British National Formulary
Primary research studies	Petrou, S (2011) [365]	NHS	2008	Primary bottom-up methodology
	Eddama, O (2010) [366]	Hospital	2008	Bottom-up costing exercise
	Jit, M (2010) [367]	NHS	2008	NHS reference costs applied to a decision tree model developed through consultation of national hospital admission data
	Round, JA (2011) [368]	NHS	2009	Bottom-up methodology
	Schroeder, E (2012) [369]	Hospital	2009-10	Costing exercise involved interviews with staff to describe resource use, costs from hospital finance departments and reference lists
	Essex, HN (2015) [370]	NHS	2009-10	NHS reference costs applied to resource use determined through a randomised controlled trial.

Coomarasamy, A (2016) [371]	NHS	2011-12	Health Technology Assessment. Uses NHS reference costs to attribute cost estimates to calculated resource use from a randomised controlled trial.
Carolan-Rees, G (2015) [372]	NHS	2011-12	Applied NHS reference costs to resource use.
Lain, SJ (2017) [373]	NHS	2012	NHS reference costs applied to resource use in a randomised controlled trial.
Parisaei, M (2016) [374]	Central London Hospital	2012	Bottom-up methodology
Ussher, M (2015) [375]	NHS	2012-13	National Reference costs applied to resource use in a randomised controlled trial
Walker, KF (2017) [376]	NHS and personal social services	2012-13	Cost-utility analysis of a randomised controlled trial using National reference and manufacturer costs
van der Nelson, H (2017) [377]	NHS	2012-13	Costs taken from NHS Reference costs and the British National Formulary
Bick, D (2017) [378]	NHS	2013-14	Health Technology Assessment of a randomised controlled trial cost-effectiveness study using National reference costs and those from other primary research studies
Campbell, HE (2018) [379]	Society	2013-14	Assessed the prevalence in a UK cohort and applied costs assessed from secondary sources.
Duckworth, S (2016) [380]	Commissioner	2013-14	Decision analytic model developed using data from an observational cohort study and National reference costs were applied
Orlovic, M (2017) [381]	NHS	2013-14	Applied NHS reference costs to resource use derived from a population study using national Hospital Episode Statistics data
Vatish, M (2016) [382]	NHS	2013-14	Applied NHS reference costs to an economic model derived from an observational cohort study.
Bowers, J (2016) [383]	NHS	2014	Used data from the Scottish Nursing and Midwifery Workload and Workforce planning project to develop a financial model
Luni, Y (2017) [384]	South West England Hospital	2014-15	Bottom-up costing exercise
Khan, KS (2018) [385]	NHS	2014-15	Bottom-up costing attached to data on resource use.
Waugh, J (2017) [386]	NHS	2014-15	Cost-effectiveness analysis conducted by NHS reference and manufacturer costs to estimated resource use taken from the NICE hypertension guideline.
Jones, M (2019) [387]	NHS	2014-15	Costs derived from NHS reference costs and the expert opinion of an NHS midwife.
Jacklin, PB (2017) [388]	NHS	2015	Applied costs taken from published UK sources to UK and Australian cohorts

Xydopoulos, G (2019)[389]	NHS	2015	Cost inputs derived from a series of costing templates based on NICE guidelines and NHS practice reports as well as other relevant scientific literature. NHS hospital tariffs could not be extrapolated to these costs.
Wastlund, D (2019 - BJOG)[390]	NHS	2016-17	Values were identified from relevant literature by two authors, systematic reviews using UK data were prioritised where possible. Where multiple sources where available, those which provided ranges were preferred and if not, a decision was made by consensus or arbitration by the senior author.
Wastlund, D (2019 - PLOS Med)[391]	NHS	2017	Costs were determined using a combination of expert opinion, relevant scientific literature and NHS reference costs.

Of the 27 included primary research studies, the cost perspective was as follows:

- local hospital (i.e., direct costs of procuring items and paying staff salaries at those hospitals) in four cases,
- commissioner (i.e., direct costs of paying the hospital for providing a service) in one study,
- indirect societal perspective (i.e., the wider costs to society, including workdays lost; in one case this included the NHS perspective) in two studies
- the NHS perspective only (i.e., directly attributed, nationally agreed costs for procurement, staff salaries, etc) in the remaining studies (n=20) (including guidelines and review articles).

Through quality assessment of the 27 primary research articles, 21 were scored as high quality (≥75/100 points)[365-371,373,375-378,380-382,385-388,390,391], 3 as medium quality (50-74/100 points),[372,379,389] and 3 were scored as low quality (<50/100 points)[374,383,384]. The detailed QHES evaluations are provided in Table 4.4 and Table 4.5 (separated for ease of reading).

Table 4.4 - Results of the quality assessments on primary research articles

	Petrou, S (2011)	Eddama, O (2010)	Jit, M (2010)	Round, JA (2011)	Schroeder, E (2012)	Essex, HN (2015)	Coomarasamy, A (2016)	Carolan-Rees, G (2015)	Lain, SJ (2017)	Parisaei, M (2016)	Ussher, M (2015)	Walker, KF (2017)	van der Nelson, H (2017)	Bick, D (2017)
1. Was the study objective presented in a clear, specific and measurable manner? (Y/N)	7	7	0	7	7	7	7	7	7	0	7	7	7	7
2. Were the perspective of the analysis (societal, third-party payer, etc) and reasons for its selection stated? (Y/N)	4	0	0	0	0	4	0	0	0	0	4	0	4	0
3. Were variable estimates used in the analysis from the best available source (i.e. randomized control trial - best, expert opinion - worst)? (Y/N)	8	8	8	8	0	8	8	0	8	0	8	8	8	8
4. If estimates came from a subgroup analysis, were the groups prespecified in the beginning of the study? (Y/N)	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5. Was uncertainty handled by (1) statistical analysis to address random events, (2) sensitivity analysis to cover a range of assumptions? (Y/N)	9	9	9	9	9	0	9	9	9	0	9	9	9	9
6. Was incremental analysis performed between alternatives for resources and costs? (Y/N)	6	6	6	6	6	6	6	0	6	6	6	6	6	6
7. Was the methodology for data abstraction (including the value of health states and other benefits) stated? (Y/N)	5	5	5	5	5	0	5	0	0	0	0	5	5	5
8. Did the analytic horizon allow time for all relevant and important outcomes? Were benefits and costs that went beyond 1 year discounted (3%-5%) and justification given for the discount rate? (Y/N)	7	7	7	7	7	7	7	7	7	7	7	7	7	7

	Petrou, S (2011)	Eddama, O (2010)	Jit, M (2010)	Round, JA (2011)	Schroeder, E (2012)	Essex, HN (2015)	Coomarasamy, A (2016)	Carolan-Rees, G (2015)	Lain, SJ (2017)	Parisaei, M (2016)	Ussher, M (2015)	Walker, KF (2017)	van der Nelson, H (2017)	Bick, D (2017)
9. Was the measurement of costs appropriate and the methodology for the estimation of quantities and unit costs clearly described? (Y/N)	8	8	8	8	8	8	8	8	8	8	8	8	8	8
10. Were the primary outcome measure (s) for the economic evaluation clearly stated and did they include the major short-term. Was justification given for the measures/scales used? (Y/N)	6	6	6	6	6	6	6	6	6	0	6	6	6	6
11. Were the health outcomes measures/scales valid and reliable? If previously tested valid and reliable measures were not available, was justification given for measures and scales used? (Y/N)	7	7	7	7	7	7	7	0	7	7	7	7	7	7
12. Were the economic model (including structure), study methods and analysis, and the components of numerator and denominator displayed in a clear, transparent manner? (Y/N)	8	8	8	8	8	8	8	8	8	8	8	8	8	8
13. Were the choice of economic model, main assumptions, and limitations if the study stated and	7	0	0	7	0	7	7	0	0	0	7	7	7	7
14. Did the author (s) explicitly discuss direction and magnitude of potential biases? (Y/N)	6	0	0	0	0	0	6	0	0	0	6	6	6	6
15. Were the conclusions/recommendations of the study justified and based on the study results? (Y/N)	8	8	8	8	8	8	8	8	8	8	8	8	8	8
16. Was there a statement disclosing the source of funding for the study? (Y/N)	3	0	3	3	3	3	3	3	3	3	3	3	3	3
Total (Max 100)	99	79	75	89	75	79	95	56	77	47	94	95	99	95
Overall assessment	High	High	High	High	High	High	High	Medium	High	Low	High	High	High	High

Table 4.5 - Results of the quality assessments on primary research articles

Question	Campbell, HE (2018)	Duckworth, S (2016)	Orlovic, M (2017)	Vatish, M (2016)	Bowers, J (2016)	Luni, Y (2017)	Khan, KS (2018)	Waugh, J (2017)	Jones, M (2019)	Jacklin, PB (2017)	Xydopoulos, G (2019)	Wastlund, D (2019 - BJOG)	Wastlund, D (2019 - PLOS Med)
1. Was the study objective presented in a clear, specific and measurable manner? (Y/N)	7	7	7	7	0	7	7	7	7	7	7	7	7
2. Were the perspective of the analysis (societal, third-party payer, etc) and reasons for its selection stated? (Y/N)	4	4	4	4	0	4	4	0	0	0	0	0	0
3. Were variable estimates used in the analysis from the best available source (i.e. randomized control trial - best, expert opinion - worst)? (Y/N)	0	0	8	0	0	0	8	8	8	8	0	0	0
4. If estimates came from a subgroup analysis, were the groups prespecified in the beginning of the study? (Y/N)	0	1	0	0	0	0	0	0	0	1	0	0	0
5. Was uncertainty handled by (1) statistical analysis to address random events, (2) sensitivity analysis to cover a range of assumptions? (Y/N)	9	9	9	9	0	0	9	9	9	9	0	9	9
6. Was incremental analysis performed between alternatives for resources and costs? (Y/N)	0	6	0	6	6	0	6	6	6	6	6	6	6
7. Was the methodology for data abstraction (including the value of health states and other benefits) stated? (Y/N)	0	0	0	0	0	0	0	5	5	5	0	5	5
8. Did the analytic horizon allow time for all relevant and important outcomes? Were benefits and costs that went beyond 1 year discounted (3%-5%) and justification given for the discount rate? (Y/N)	0	7	7	7	7	7	7	7	7	7	7	7	7

210

	Campbell, HE (2018)	Duckworth, S (2016)	Orlovic, M (2017)	Vatish, M (2016)	Bowers, J (2016)	Luni, Y (2017)	Khan, KS (2018)	Waugh, J (2017)	Jones, M (2019)	Jacklin, PB (2017)	Xydopoulos, G (2019)	Wastlund, D (2019 - BJOG)	Wastlund, D (2019 - PLOS Med)
9. Was the measurement of costs appropriate and the methodology for the estimation of quantities and unit costs clearly described? (Y/N)	8	8	8	8	8	0	8	8	8	8	8	8	8
10. Were the primary outcome measure (s) for the economic evaluation clearly stated and did they include the major short-term. Was justification given for the measures/scales used? (Y/N)	6	6	6	6	0	6	6	6	6	6	6	6	6
11. Were the health outcomes measures/scales valid and reliable? If previously tested valid and reliable measures were not available, was justification given for measures and scales used? (Y/N)	0	7	7	7	7	7	7	7	7	7	7	7	7
12. Were the economic model (including structure), study methods and analysis, and the components of numerator and denominator displayed in a clear, transparent manner? (Y/N)	8	8	8	8	0	0	8	8	8	8	8	8	8
13. Were the choice of economic model, main assumptions, and limitations if the study stated and justified? (Y/N)	7	7	7	7	7	0	7	7	7	7	7	7	7
14. Did the author (s) explicitly discuss direction and magnitude of potential biases? (Y/N)	6	6	6	0	0	0	6	0	6	6	6	6	6
15. Were the conclusions/recommendations of the study justified and based on the study results? (Y/N)	8	8	8	8	8	0	8	8	8	8	8	8	8
16. Was there a statement disclosing the source of funding for the study? (Y/N)	3	3	3	0	0	3	3	3	3	3	3	3	3
Total (Max 100)	66	87	84	77	43	30	94	89	95	96	73	81	87
Overall assessment	Medium	High	High	High	Low	Low	High	High	High	High	Medium	High	High

211

Of the primary research articles (n=27), seven estimated activity costs using bottom-up methodologies (i.e., individually micro-costed each item e.g., drug costs, equipment costs, cost per staff-hour worked), ten costed activities using national reference costs for NHS diagnosis or procedure codes and staff, three used other available literature to cost activities, and the remaining seven studies used a combination of costing methodologies.

The cost estimates from the NICE guidelines, review articles and primary research papers (inflated to 2018/19 prices) are presented separately for activity items within antenatal, intrapartum and, postnatal care within Figure 4.2 and in more detailed and referenced format in Table 4.6 - Table 4.8 respectively.

Figure 4.2 - Variation in extracted unit costs for activities within the maternity care pathway
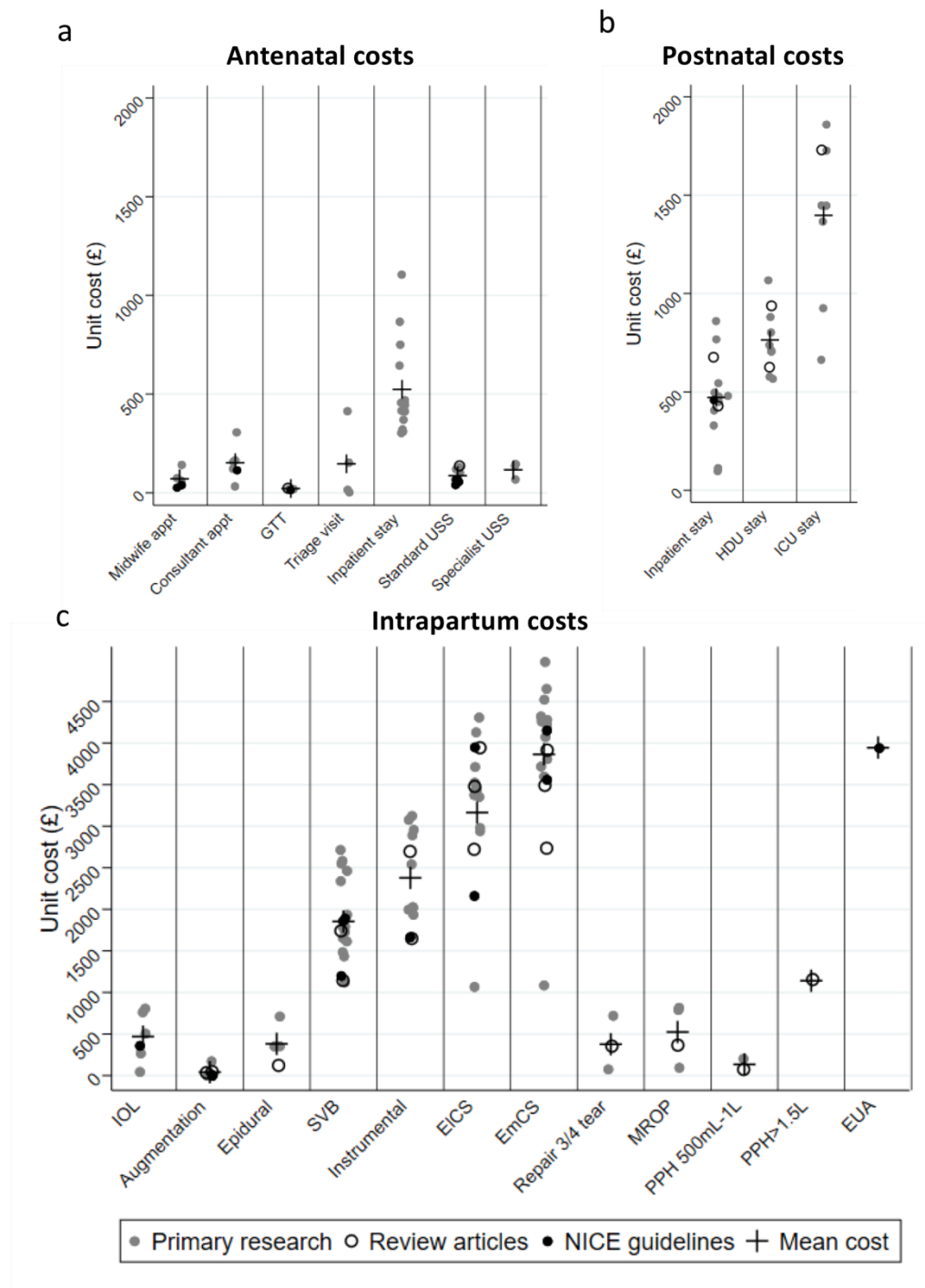
Table 4.6- Extracted costs for antenatal care

Activity costed	Economic evaluations for National guidelines	Review articles	Primary research studies Unit costs reported (ascending order)	Range	Mean (SD)	Comments
Estimated cost for standard midwife antenatal appointment	Not costed	£27.34/20 minute appointment,[358] £224.99/pregnancy for community care,[360] £264.56/pregnancy for midwife-led hospital care,[360] £131.94/hour.[363]	£76.03/appointment,[379] £146.25/appointment,[382] £60.49/appointment.[387]	£27.34 - £146.25	£70.82 (£45.92)	Only 'per appointment' estimates used
Estimated cost for a standard obstetric antenatal appointment	Not costed	£114.1/follow-up appointment,[358] £515.55/pregnancy.[360]	£43.36,[376] £166.68/appointment,[379] £312.29/appointment,[380] £153.94/appointment,[382] £124.12/appointment.[387]	£43.36 - £312.29	£152.42 (£89.36)	Only 'per appointment' estimates used
Estimated cost for any antenatal appointment	£95.75/ appointment.[357]	Not costed	£115.62/appointment,[373] £103.28/appointment.[388]	£95.75 - £115.62	£104.88 (£10.03)	Only 'per appointment' estimates used
Estimated cost for a glucose tolerance test	£23.73.[357]	£24.26.[363]	£26.16,[368] £13.03.[388]	£13.03 - £26.16	£21.80 (£5.94)	
Estimated cost for attendance to maternity day unit/triage	Not costed	Not costed	£150.37/visit,[371] £415.65/visit,[380] £6.56 for nurse-led or £15.49 for doctor-led triage review.[389]	£6.56 - £415.65	£147.02 (£190.79)	

214

Activity costed	Economic evaluations for National guidelines	Review articles	Primary research studies — Unit costs reported (ascending order)	Range	Mean (SD)	Comments
Estimated cost for an antenatal inpatient admission	Not costed	Not costed	£330.84 on antenatal ward or £1115.87 on labour ward,[365] £366.00/day,[366] £1403.55/admission,[370] £298.47/day,[371] £758.18/day for first three days and £448.01/day after this,[373] £466.91/day,[374] £457.75/day,[379] £867.59/day for first five days and £414.55/day after this,[380] £414.55/day,[382] £447.12/day,[386] £1658.89 for standard antenatal stay and £304.45 for additional bed days.[389]	£298.47 - £1,115.87	£524.11 (£239.07)	Only 'per day' estimates used
Estimated cost for an ultrasound scan (sonographer)	£139.85,[357]	£74.90/first scan and £61.83/subsequent scans,[358] £56.07/scan,[359] £42.24/scan,[361] £142.95/3 scans.[363]	£57.66/scan,[372] £121.87/scan,[379] £120.49/scan,[388] £120.72/scan,[387] £112.21/scan.[390]	£42.24 - £139.85	£86.86 (£36.13)	Only 'per scan' estimates used
Estimated cost for an ultrasound scan (specialist)	Not costed	Not costed	£77.82/ scan,[374] £127.55/scan,[380] £143.65/scan.[387]	£77.82 - £143.65	£116.34 (£34.32)	Only 'per scan' estimates used

215

Table 4.7 - Extracted costs for intrapartum care

Activity costed	Economic evaluations for National guidelines	Review articles	Primary research studies — Unit costs reported (ascending order)	Range	Mean (SD)
Induction of labour	£31.17 for drugs only.[356]	£33.35 for Propess or £39.53 for vaginal prostaglandin gel/tablet (inpatient days costed separately),[362] £361.77 for induction.[363]	£47.56,[368] £518.46-£805.42 in addition to cost of vaginal birth,[371] £791.53 in addition to the cost of a vaginal birth,[373] £33.35 for Propess pessary or £30.02 for 2 doses of Prostin,[376] £28.94 for 2 doses of 3mg dinoprostone tablet,[386] £290.60.[390]	£31.17 - £805.42	£469.22 (£323.86) Only fully-costed induction included in the
Augmentation of labour	£34.98,[356] £56.96.[355]	£1.90 for oxytocin and £2.97 for amniotomy and oxytocin.[362]	£189.16,[369] £1.01 for oxytocin and £0.95 for Amnihook,[376] £1.10 for oxytocin.[378]	£1.10 - £189.16	£41.29 (£68.72)
Epidural	£118.08.[355]	Not costed	£369.89,[369] £693.70 in addition to the cost of a vaginal birth,[373] £345.73.[376]	£118.08 - £693.70	£381.85 (£266.89)
Spontaneous vaginal birth	£1,170.50,[356] £1,762.19.[355]	£1,888.08,[358] £1,222.87 if within 24 hours of commencing induction of labour;[362] £1,905.40.[364]	£1,125.95 for birth without complication or £2,474.24 for birth with complication,[369] £1,729.09,[370] £1,460.77-£1,812.38 (dependent upon complexity),[371] £1,782.05,[373] £1,473.00,[376] £2,572.02,[375] £1,943.23,[379] £1,607.61,[378] £1,648.92,[386] £2,721.27,[387] £2,586.11,[390] £2,343.25.[391]	£1,125.95 - £2,721.27	£1,854.15 (£486.97)
Instrumental birth	£1,662.25,[356] £1194.17 £2,663.45.[355]	£1,633.64,[358] £1194.17 more than a spontaneous vaginal birth.[363]	£52.64 in addition to cost of labour;[366] £510.31 - 677.59 for birth plus staffing and overhead costs for labour,[369] £2,961.67,[370] £1,922.20 - £2,524.68 (dependent on complexity),[371] £3,100.37,[373] £2,003.28,[376] £3,100.37,[375] £2,866.18[379] £3,100.37,[373] £2,045.26,[378] £2,056.62.[386]	£1,633.64 - £3,866.18	£2,378.33 (£552.20) Only estimates of labour and birth included.

Activity costed	Economic evaluations for National guidelines	Review articles	Primary research studies — Unit costs reported (ascending order)	Range	Mean (SD)
Elective Caesarean section	£2,724.24 - £3,494.18,[356] £3,923.25.[355]	£2,166.30,[358] £972.05 more than a spontaneous vaginal birth,[363]	£1,251.51 plus staffing and overhead costs,[369] £4,281.45,[370] £2,935.27 - £3,438.43 (depends on complexity),[371] £3,380.67,[373] £2,983.79,[376] £3,402.19,[379] £3,385.25,[386] £4,120.81,[387] £1,056.44,[388] £3710.33,[390] £3,506.86.[391]	£1,056.44 - £4,281.45	£3,164.49 (£801.63) Only full estimates of elective Caesarean birth included.
Emergency Caesarean section	£2,724.24 - £3,494.18,[356] £3,923.25.[355]	£3,541.94,[358] £4,143.29,[362] £972.05 more than a spontaneous vaginal birth.[363]	£318.78 - in addition to cost of labour,[366] £1,251.51 for birth plus staffing and overhead costs for labour,[369] £1,432.71 (in addition to cost of vaginal birth),[368] £3,600.98,[370] £3,717.67 - £4,284.10 (dependent on complexity),[371] £4,212.22,[373] £3,795.33,[376] £4,278.73,[375] £4,325.86,[379] £4,039.93,[378] £4,244.57,[386] £4,555.92,[387] £1,056.44,[388] £4,982.21,[390] £4,644.47,[391]	£2,724.24 - £4,982.21	£3864.74 (£867.58) Only estimates of labour and birth included.
Repair 3/4th degree tear	£351.95.[355]	Not costed	£707.79,[369] £70.37.[378]	£70.37 - £707.79	£376.70 (£321.71)
Manual removal of placenta	£152.35 for staff and £236.49 for consumables.[355]	Not costed	£819.58,[369] £807.09 (in addition to the cost of birth,[373] £81.37.[378]	£81.37 - £819.58	£524.22 (£387.80)
Postpartum haemorrhage (500-1500mL without shock)	£60.40 for staff and £40.32 for consumables.[355]	Not costed	£169.34.[378]	£100.72 - £169.34	£135.03 (£85.18)
Major obstetric haemorrhage (>1500mL)	£950.28 for staff, £150.39 for one unit blood transfusion and £40.32 for consumables.[355]	Not costed	Not costed	Only one cost provided	
Examination under anaesthesia for postpartum haemorrhage	Not costed	£3,944.70 [364]	Not costed	Only one cost provided	

Table 4.8 - Extracted costs for postnatal care

Activity costed	Economic evaluations for National guidelines	Review articles	Primary research studies		
			Unit costs reported (ascending order)	Range	Mean (SD)
Maternal inpatient stay on postnatal ward	£676.44 / day,[356] £431.56 / day.[355]	£74.62 for average postnatal stay,[360] £459.63/day.[364]	£330.84 / day,[365] £549.09 / day,[366] £112.95 / day,[369] £870.10 / day,[375] £1,289.57 / admission,[376] £488.42/day,[377] £103.00 - £757.00/day,[378] £407.83 / day,[379] £1,406.39 for average length of stay (<3 days),[381] £754.14 / standard 36-hour postnatal stay,[383] £470.19 / day,[385] £20.47 / hour (staff costs only),[384] £447.12 / day.[386]	£103.00 – £870.10	£471.12 (£211.26) Full costs only included.
Maternal inpatient stay in HDU	£936.17 / day,[356] £636.44/day.[355]	Not costed	£804.38 / day,[365] £95.12 for 4 hours,[369] £712.17 / day,[371] £700.37/day,[377] £890.00 / day,[378] £587.40 / day (level 1) or £734.52 / day (level 2),[385] £2,150.63 for first 2 days.[386]	£570.72 - £1075.32	£764.75 (£277.20)
Maternal inpatient stay in ITU	£1,737.28 / day.[356]	Not costed	£1,729.56 / day,[367] £1,444.56 / day,[366] £665.82 / day,[369] £1,367.21 / day,[371] £1,449.00 / day,[378] £924.14 / day,[385] £3,722.01 for first 2 days.[386]	£665.82 - £1,861.01	£1,397.32 (£606.35)

With respect to antenatal care, estimates for 20-minute long antenatal clinic appointments were provided for midwifery-led clinics (range £27.34 - £146.25, mean £74.70, 5.3-fold difference)[358,363,379,382,387,392] or consultant obstetrician-led clinics (range £43.36 - £312.29, mean £144.15, 7.2-fold difference).[358,360,376,379,380,382,387,392] There were only two cost estimates identified for glucose tolerance tests (range £13.03 - £26.16, mean £21.80, 2.0-fold difference).[368,388] A larger absolute unit cost range was found when estimating the cost for one day of an antenatal inpatient admission (range £298.47 - £1,115.87, mean £546.08, 3.7-fold difference).[359-361,365,366,370,371,373-375,379,380,382,386,389] Similarly, cost estimates for antenatal scans were variable for both general scans conducted by a sonographer (range £40.67 - £139.85, mean £80.86, 3.4-fold difference)[357-359,361,363,372,379,387,388,390,393] and 'specialist' scans, usually conducted by a fetal medicine consultant (range £77.82 - £143.65, mean £116.34, 1.8-fold difference).[374,380,387]

When estimating cost for intrapartum activities, there was wide variation and the costs for each activity item were generally higher than for antenatal or postnatal care. For example, the estimated cost of induction of labour varied between £47.56 - £805.42 (mean £450.08, 16.9-fold difference).[356,362,363,368,371,373,376,386,390,392,393] In some studies, it was clear that this variation followed decisions to cost the induction with or without the cost of staffing and antenatal admission, but this was not always the case. The estimated cost of an emergency caesarean section varied between £1,056.44 - £4,982.21 (mean £3,508.93, 4.7-fold difference),[355,356,358,362,370,371,373,375,376,378,379,386-388,390-393] this included the staffing and bed space required for the intrapartum admission. There were lower estimates which cost the surgery only (£318.78-£1,432.71, mean £993.76, 4.5-fold difference).[363,366,368,369]

With regards to postnatal care, the inpatient postnatal stay for a healthy woman and baby on a postnatal ward varied between £103.00 - £870.10 per day (mean £469.55, 8.4-fold difference).[355,356,360,364-366,369,375,377-379,383-386]

Through application of cost estimates to the exemplar activities within the care pathway for a low-risk multiparous woman and high-risk nulliparous woman and applying the lowest and highest cost estimates, the significant effects that these cost variations can have on the estimated cost of care provided to a single woman are demonstrated (Table 4.9 and Table 4.10).

Table 4.9 - Estimating costs for a low-risk pregnant woman.

Activity within care	Lowest cost estimate	Highest cost estimate
Antenatal booking appointment	£27.34[358]	£146.25[382]
2 sonography-led ultrasound scans	2x£42.24[361]	2x£139.85[357]
	£84.48	£279.70
5 midwifery-led antenatal appointments	5x£27.34[358]	5x£146.25[382]
	£136.70	£731.25
1 attendance to maternity triage	£6.56[389]	£415.65[380]
Uncomplicated spontaneous vaginal birth	£1125.95[369]	£2572.02[375]
6 hour discharge	£0	£0
TOTAL:	**£1,381.03**	**£4,144.87**

Table 4.10 - Estimating costs for a higher-risk pregnant woman

Activity within care	Lowest cost estimate	Highest cost estimate
Antenatal booking appointment	£27.34[358]	£146.25[382]
2 sonography-led ultrasound scans	2x£42.24[361]	2x£139.85[357]
	£84.48	£279.70
7 midwifery-led antenatal appointments	7x£27.34[358]	7x£146.25[382]
	£191.38	£1,023.75
2 consultant-led appointments	2x£43.36[376]	2x£312.29[380]
	£86.72	£624.58
3 attendances to maternity triage with pre-eclampsia	3x£15.49[389]	3x£415.65[380]
	£46.47	£1,246.95
2 Specialist growth scans	2x£77.82[374]	2x£127.55[380]
	£155.64	£255.10
3 day antenatal admission	3x£298.47[371]	3x£1115.87[365]
	£895.41	£3,347.61
Induction of labour with 2 day antenatal admission	£361.77[363]	£805.42[371]
Epidural	£118.08[355]	£693.70[373]
Labour augmentation	£1.10[378]	£189.16[369]
Emergency caesarean section	£1,056.44[388]	£4,982.21[390]
3 day postnatal inpatient stay	3x£103.00[378]	3x£870.10[375]
	£309.00	£2,610.30
TOTAL:	**£3,333.83**	**£16,204.73**

4.3.1 Sub-group analyses

Sensitivity analysis with exclusion of the three primary research papers which were assessed as low quality using the QHES instrument resulted in only two cost estimates being removed from the overall results. The mean cost per antenatal admission day changed from £524.11 (SD: £239.07) to £528.51 (SD: £248.24) and the mean cost per day on postnatal ward changed from £471.92 (SD: £211.26) to £469.55 (SD: 219.69). There were no changes to intrapartum costs.

Sensitivity analysis to determine the effect of including a range of cost perspectives on the variation identified between cost estimates for each activity was conducted graphically and the results are presented in Figure 4.3. Most costs were derived from economic evaluations which use the perspective of the health service. Despite separation by perspective, variation still exists, even for costs derived from studies conducted from the perspective of the health service. There were too few data points to examine variation across studies conducted from other perspectives.

Figure 4.3 - Presentation of unit costs by cost perspective quoted in source paper
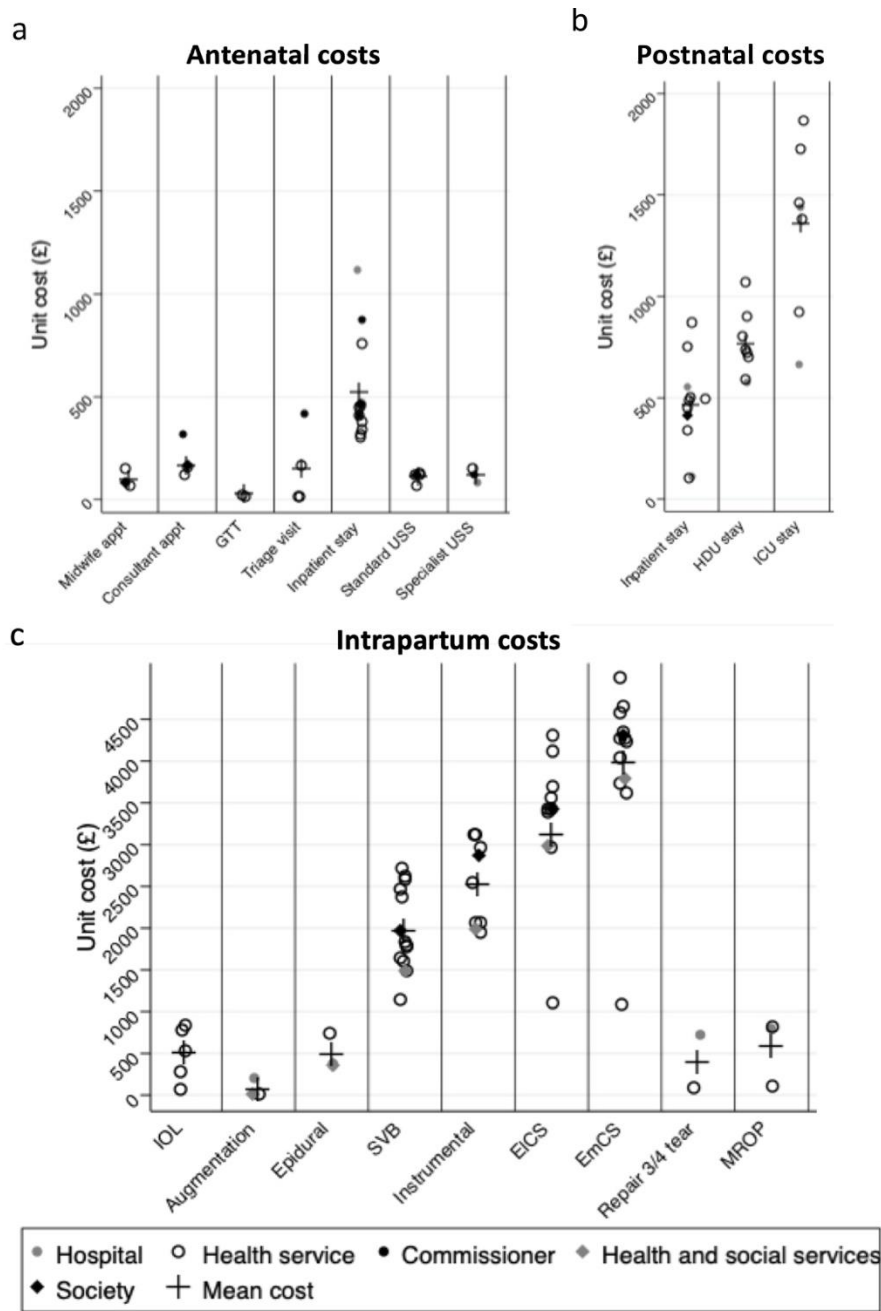
4.4 Discussion

4.4.1 Summary of the key findings

The aim of this analysis was to document the current evidence on the costing of resource use within maternity care, to inform the economic evaluation nested within the DESiGN trial and reported in Chapter 5. I have reviewed seven economic evaluations with UK costs applied following a systematic literature review, three economic evaluations from UK NICE guidelines and 27 primary research articles that have attributed unit costs to activities within the antenatal, intrapartum and postnatal pathways, specific to the UK context. I have noted wide ranges in published cost estimates, including a 16.9-fold difference between the minimum and maximum cost estimates for induction of labour, despite limiting the search to studies within the last 10 years and inflating costs to 2018/19 prices.

For intrapartum costs in particular, the absolute difference between the minimum and maximum costs are greater, because these are usually higher cost interventions. This is likely to have more of an impact on the results of cost-effectiveness evaluations. Even where absolute cost differences are small because the activity itself is relatively inexpensive e.g., cost estimates for the glucose tolerance test, the relative difference shows that the maximum estimate is twice as high (or more) than the minimum estimate, although a low absolute difference is less likely to impact when estimating the financial impact of new interventions.

Whilst the estimate ranges are tighter for the unit costs supplied in the seven review articles, with the overall estimate tending towards the middle of the range of the primary research articles, this may be accounted for by both the smaller number of studies and that these studies are often based on estimates from national guidelines and primary research articles. Wide variation also exists within cost estimates supplied by some NICE guidelines, where the same activity (e.g., emergency Caesarean section) is priced differently by economic evaluations featured in different guidelines.[355,356,392,393]

4.4.2 Interpretation of the findings

There are several potential explanations for some of this variation. Variation can reflect different methodology of cost calculation, and differing definitions of each activity, for example, the average cost for an inpatient admission varied from cost per day/night to costs estimated for a time-defined (e.g., 3 nights) admission (although only costs for a single night are presented in the results). Costs may also vary with changing geographical perspective

and varying approaches to clinical practice resource use between localities; it is well-established that costs are higher in Southern than Northern England, particularly in inner-city London.[394]

Methodological quality is another explanation for the variation found in this review, although the estimates changed little after exclusion of papers determined to be of low methodological quality. Poorly applied methods and incomplete reporting make the results of economic evaluations less reliable, less comparable on a consistent like-with-like basis and difficult to interpret. It also introduces additional uncertainty when seeking to transfer evidence on costs to other study contexts.

4.4.3   Strengths and limitations

The strength of this study is in the extensive literature search of four relevant databases and the wide, clinically-generalisable (within UK maternity care) inclusion criteria. Unit costs extracted from published research articles were compared to one-another and to summary costs published as part of national Health Technology Assessments and NICE guidelines. Unit costs extracted from published reports of economic evaluations were inflated to 2018/19 prices and stratified by quoted cost perspective to ensure comparability.

Due to the lack of comparability in international health systems and maternity reimbursement policies, it was not appropriate to extend the search outside of economic evaluations conducted within the UK. Whilst the specific cost findings are only generalisable to UK maternity care, the overall findings regarding the challenges of estimating the financial impact of interventions using bundled prices, and the risk of cost variation where nationally agreed costs are not available, are relevant to maternity care providers internationally and potentially also to other medical specialities where bundled costs are commonplace.

4.4.4   Implication of the findings

I have shown how variation in reported costs can introduce uncertainty into estimates of the overall cost of pregnancy management at different levels of pregnancy risk. This is likely to have important implications where "bottom-up" costing methodologies are required to support the evaluation of interventions that are expected to change the type and volume of clinical activity that patients are exposed to along the pregnancy care pathway. This will be further magnified in cases where an intervention impacts on comparatively expensive areas of clinical activity e.g., antenatal admissions, mode of birth, infant admission to neonatal units.

Estimating the financial impact of introducing GAP into the antenatal pathway for the DESiGN trial was expected to be challenging because of the bundled nature of national reference costs in England (Table 4.2).[310] Our hypothesis specified that the intervention was expected to increase antenatal activities such as clinic appointments or scans and intrapartum activities such as induction of labour (through increases in the diagnosis of SGA). Whilst these changes were expected to incur cost to the hospital, this would not be reflected in the bundled price charged to the commissioner. Itemised costs were therefore required, and the results of this review will be used to inform the costing framework of Chapter 5.

The variation in quoted costs suggests uncertainty around methods to calculate costs. Whilst the included studies have mostly been appraised as having medium or high reporting quality (according to the QHES checklist), how costs were calculated and exactly what was included in each estimate e.g., length of appointment, salary of healthcare professional used, inclusion of indirect costs was not always explicitly described. Guidelines on what should be included when calculating the cost of common activities, including how to account for variable staff salaries and indirect costs, on the appropriate cost perspective to choose (and report) and how to translate costs geographically would be invaluable in achieving lower variation in published estimates.

An alternative strategy would be publication of a list of nationally agreed itemised costs for use in the economic evaluation of interventions, with guidance on which costs within a range of estimates are more likely to be applicable to specific circumstances. This would facilitate greater consistency in the application of cost data across different evaluations. Such a list was previously available in England (although not published for this purpose), but has since been replaced by a national bundled tariff.[284]

4.4.5 Conclusions

Through this systematic review of economic evaluations within maternity care I have described significant variation in costs applied to maternity care activities, even after controlling for study reporting quality and cost perspective. I have outlined the challenges in attributing cost to maternity activities, due to non-standardised activity descriptions and provision of composite 'bundled' cost estimates.

Overall, the level of variation in cost calculations is likely to reflect the uncertainty within the system and must be dealt with by accounting for this uncertainty during economic evaluations. The development of nationally agreed unit costs for key areas of clinical activity within the pregnancy care pathway would serve to standardise cost-

effectiveness analyses of new interventions within maternity care, to be used either for research purposes or national decisions regarding intervention uptake.

5 EVALUATING THE COST-EFFECTIVENESS OF THE GROWTH ASSESSMENT PROTOCOL

5.1 INTRODUCTION

Inclusion of economic evaluations in trials of healthcare interventions is recommended to assist decision making on the adoption or spread of intervention implementation.[265] In the UK, NICE makes recommendations about implementation of new interventions based partly on cost-effectiveness. This considers whether the intervention is clinically effective, the costs associated with its use and whether the clinical resource could be better spent to achieve improved health outcomes differently. In healthcare settings where resources are limited, resources must be prioritised to care strategies with greatest patient benefit.[395]

A trial-based economic evaluation of GAP has not previously been conducted although a cost-benefit analysis to determine the costs and benefits of increasing serial fetal ultrasound provision for women at high risk of SGA has been published on the website of the Perinatal Institute.[87] This estimated that such a policy would cost an additional £10 per pregnancy (cost is spread across all low- and high-risk pregnancies). This cost of serial scans was compared to an estimated saving of £120 per pregnancy by reducing neonatal admissions, perinatal morbidity and mortality, cerebral palsy and litigation through an increase in the detection of SGA. A Cochrane review which compared customised to population-based perinatal growth charts for antenatal screening of SGA concluded both that there was insufficient evidence to prioritise either approach but also stated that the potential financial costs associated with implementation of customised growth charts, including the costs of additional scans, iatrogenic birth, and further investigation must be considered when recommending such an approach.[136]

During the DESiGN trial, it was evident that individual clinicians and NHS Trusts were concerned about the cost-effectiveness of GAP, with two trusts withdrawing from GAP implementation as randomised early in the trial because of local financial concerns, and members of staff interviewed during the process evaluation, citing concerns about the inability to cope with the increase in ultrasound resource use.[396]

5.1.1 Objectives

The objective of this economic evaluation was to determine whether the Growth Assessment Protocol is cost-effective in the antenatal detection of the SGA fetus, compared to standard practice.

5.2 METHODS

5.2.1 Study design

The economic evaluation of this trial was planned as a cost-effectiveness analysis.

This chapter has been reported in accordance with the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement for reporting economic evaluations (Appendix section 10.13).

5.2.2 Economic perspective and time horizon

The cost perspective was that of an NHS maternity care provider and the time horizon was from 24^{+0} weeks of gestation until hospital discharge of the mother and baby following the birth admission. The economic perspective decision was explained in detail in section 2.4.3.1.

Antenatal costs have been estimated for care at or after 24 weeks' gestation, because this is the time at which fetal surveillance using the intervention commences during pregnancy and therefore, costs incurred before this gestational age were not expected to vary by the intervention. After this threshold, we planned to include in this analysis all major antenatal, intrapartum, neonatal, and postnatal costs until the mother or infant are discharged from their inpatient episode for birth at the cluster site in which birth occurred.

Maternal and neonatal readmissions were not included since these were not expected to be related to the intervention, nor were costs incurred in the community or social services. The long-term health and societal costs of stillbirth and preterm birth have previously been estimated elsewhere.[379,397,398] Costs were therefore not discounted because all were expected to occur within a single year.

5.2.3 Clinical outcomes

An analysis was undertaken to evaluate the cost-effectiveness of the GAP intervention, compared to standard care, in the antenatal detection of SGA_{both} neonates (using the trial primary outcome definition of SGA, section 2.2.1.1). The incremental cost and incremental clinical effects attributable to GAP compared to standard care were estimated and expressed as an incremental cost-effectiveness ratio (ICER), i.e., the cost of the intervention for every additional case of SGA diagnosed.

5.2.4 Measurement of resource use

Data were collected on all significant hospital or community midwifery activities and clinical outcomes for women and babies, during the time horizon of the trial (section 2.4.3.2 and Table 2.15), using data collection methods and data management strategies outlined in Chapter 2 and previously published.[399] These activities and outcomes include pre-specified secondary outcomes of the trial relevant to estimating the impact of GAP on service provision:

- rate of attendance to antenatal clinics or day units,
- number of ultrasound scans,
- length of antenatal or maternal postnatal stays in hospital,
- rate of induction of labour,
- rate of caesarean births,
- rate of admission including length of stay at each level of neonatal care.

For some maternity activities, we were aware that not all women would have needed to engage with them during pregnancy, but that this was not recorded in the dataset i.e., the record appeared to be missing when in fact it had just not happened. Important examples include fetal growth scans in the third trimester, antenatal admissions and epidural anaesthesia. For some activity items e.g., epidural, repair of third degree tear, or neonatal care we used the assumptions previously described in section 2.2.4. For fetal growth ultrasound and antenatal admissions we made the following assumptions:

- That women who had any record of an ultrasound scan prior to 24 weeks' gestation but not after, had not had a third trimester ultrasound scan.
- That women with no record of an antenatal admission to hospital, but who had evidence of antenatal care commencing at the cluster site prior to 32 weeks' gestation, through evidence of antenatal clinic contacts, had really not been admitted to hospital in the antenatal period.

Data were also collected on the most resource intense activities relevant to GAP implementation (section 2.4.3.2) – this did not include planning meetings, time taken to re-write guidelines or perform engagement activities. The number of staff employed in maternity departments was collected from site clinical leads. The number of these staff members from each professional group (doctors, midwifes and sonographers) who attended the site-wide training launch was collected from training records supplied by the Perinatal Institute. The time taken to complete face-to-face and e-learning training not provided as part of the training day was estimated by participants in semi-structured interviews conducted with participants as part of the trial process evaluation. The time

taken to generate the GROW chart and to plot the fundal height or EFW measurements onto the chart has not been costed because we found no evidence in qualitative interviews that these interventions were prolonging the standard appointment time, despite interviewed staff estimating that the median time to generate each GROW chart was 4 minutes. For each time estimate, it was assumed that this did not differ between sites and the median response across all sites was used.

5.2.5   Valuation of resource use

Unit costs for each hospital activity or maternal outcome were estimated using the strategy described in Chapter 2, section 2.4.3.3. The cost estimate applied to each resource or activity is detailed in Table 5.1.

Table 5.1 - Clinical care activities and maternal or neonatal outcomes and costs estimated for the cost-effectiveness analysis

Cost item	Cost applied (inflated to 2018/19)	Cost source
Antenatal costs		
Standard midwifery outpatient appointment	£117.16 / appt	NHS reference costs 2015-16[284]
Standard obstetric outpatient appointment	£141.95 / appt	
Standard outpatient appointment (lead professional not known)	£129.56 / appt	Midpoint between midwife and consultant costs (above).
Attendance to maternity day assessment unit or maternity triage centre.	£259.48 / attendance	NHS reference costs 2015-16 (NZ16Z - Day case)[284]
Admission on antenatal ward	£1,133.76 for first day and £184.67 for every day which follows	NHS reference costs 2015-16 (NZ16Z - Antenatal Routine Observation)[284]
Fetal growth ultrasound scan*	£109.07 / scan	NHS reference costs 2015-16 (NZ21Z)[284]
Intrapartum costs		
Induction of labour (includes admission)	£394.71	NICE – Inducing Labour (2008)[392]
Epidural	£118.08	NICE – Intrapartum Care for healthy women and babies (2014)[355]
Spontaneous vaginal birth	£1,762.19	
Instrumental vaginal birth	£2,663.45	
Elective Caesarean Section	£3,923,25	
Emergency Caesarean Section	£3,923.25	
Repair 3/4th degree tear	£351.95	
Postpartum haemorrhage (500-1500mL)	£100.32	
Postpartum haemorrhage > 1500mL	£1,140.99	
Postnatal costs		
Inpatient stay on postnatal ward	£431.56 / day	NICE – Intrapartum Care for healthy women and babies (2014)[355]
Neonatal costs		
Admission to Neonatal Intensive Care Unit (level 3)	£1,157.70 / day	NHS reference costs 2017-18 XA01Z[310]
Admission to Local Neonatal Unit (level 2)	£780.30 / day	NHS reference costs 2017-18 XA02Z[310]
Admission to Special Care Baby Unit (level 1)	£542.64 / day	NHS reference costs 2017-18 XA03Z[310]

*Data which indicated which healthcare professional conducted the scan were not available at most sites and so all scans are assumed to have been sonographer-led.

Hourly costs were estimated for each staff group from the Unit Costs of Health and Social Care 2018 data published by the Personal Social Services Research Unit (PSSRU).[400] These estimate the mean full-time equivalent salary of each healthcare professional and direct or indirect costs associated with their employment, and are presented in Table 5.2. All midwifery staff were assumed to be band 6 on the pay scale, all sonographers were assumed to be band 7, the published mean costs of junior doctors and surgical consultants (as per the PSSRU document) were used.

Table 5.2 - Salary costs used in economic evaluation

Cost type	Midwife (band 6)	Sonographer (band 7)	Consultant obstetrician	Junior doctor (Registrar)
Annual salary	£32,563	£39,181	£90,535	£41,583
Annual on-costs	£8,050	£9,912	£24,386	£10,591
Other overheads	£30,794	£42,638	£83,457	£39,823
Total annual cost	**£71,407**	**£91,731**	**£198,378**	**£91,897**
Annual hours of work (full time)	1,573	1,599	1,842	2,138
Estimated hourly cost	**£45**	**£57**	**£108**	**£43**

Finally, implementation of GAP incurs an annual license fee to the Perinatal Institute which pays for the online GAP resources (generation of customised birthweight centile charts, birthweight customised centile calculator, audit tools for missed cases). It includes a one-off set-up cost of £500, plus an annual cost titrated by the annual birth rate at the purchasing trust (Table 5.3).

Table 5.3 - Annual cost of implementing GAP (2018/19)

Annual birth rate	Set up cost (includes whole day of training)	Annual cost
<3000 babies	£500	£1500
3000-5000 babies	£500	£2000
5000 – 7000 babies	£500	£3000
>7000 babies	£500	£4000

5.2.6 Evaluation and management of missing data

Following application of the assumptions described in section 5.2.4 (to determine which women had most likely not needed to engage with some of the antenatal activities measured), we considered the following to be incomplete records of care, indicating that it

was most likely that the woman had received earlier antenatal care elsewhere, and converted the values missing:

- Ultrasound records in which there was no evidence of any ultrasound prior to 24 weeks' gestation.
- Antenatal clinic records or antenatal admission records in which the woman had no evidence of any clinic appointment prior to 32 weeks.
- Postnatal length of stay for the hospital admission for birth in which there was no documented admission for birth (recognising the limitation that a small proportion of women had a home birth).

The number and proportion of pregnancies with missing data for each activity item was assessed for all clusters. We managed missing data for the following activities through multiple imputation using chained equations (additional to the clinical data imputed for the main clinical analysis, described in section 2.2.4):

- Number of ultrasound scans after 24 weeks' gestation,
- Number of antenatal clinic appointments after 24 weeks' gestation – this was treated as missing if the woman had no record of antenatal appointments prior to 32 weeks',
- Number and length of antenatal stay after 24 weeks' gestation,
- Length of postnatal stay during the admission in which the woman gave birth.

Data were only imputed in cluster sites which contributed at least some data on the activity item because imputation was conducted within cluster (the non-missing data in each cluster was used to predict missing data for that cluster only). Where a site did not contribute any data for an entire resource item, it was excluded from the analysis.

Data were not available to define false positive and true negative cases at one cluster site in the intervention arm of the trial. These screening outcomes are essential to the economic model and so the site was excluded from any analysis relevant to these two screening outcomes only (n=4,950 pregnancies in pre-randomisation period and n=2,214 in outcome period).

All analyses are presented using imputed data.

5.2.7   Analytical methods

5.2.7.1   *Describing and comparing resource use*

Resource use (as listed in section 5.2.4) was described using summary statistics (section 2.2.6.1) for both trial arms during the pre-randomisation and trial outcome phases.

Resource use was compared between trial arms to determine the mean unadjusted and adjusted differences. Adjustments were made using individual-level data on maternal age, parity and ethnicity, and cluster-level data on randomisation strata and the cluster summary value from the pre-randomisation (baseline) period using the cluster summary method described in section 2.2.6.3.1.

5.2.7.2 *Describing costs*

Costs were applied to individual pregnancy records for each activity of resource item included. Costs were also described using mean and standard deviation for all subtotals of antenatal, intrapartum, postnatal, and neonatal care, and an overall total for maternity care for both trial arms across the same two trial phases. For the GAP implementers, the total also included the additional cost of GAP implementation.

5.2.7.3 *Developing an economic model*

An economic model (Figure 5.1) was developed to consider the cost implications of GAP on all screening outcomes of antenatal detection of SGA_{both} (true positive, false positive, true negative and false negative). Non-SGA included all pregnancies which were not defined as SGA by both the GROW chart and population reference (including some pregnancies defined as SGA_{pop} but not SGA_{cust}, or vice versa). Given that GAP is an antenatal intervention, and SGA status is not confirmed until birth, we hypothesised that any incremental cost difference of GAP would be mediated through changes in the probability of each screening outcome and through changing the volume and composition of antenatal resource use for cases detected antenatally as SGA, for example GAP could increase the volume of ultrasound scans undertaken compared to standard care. The analysis assumed that GAP did not impact on intrapartum, postnatal, or neonatal costs within screening outcomes given that the protocol does not recommend different management strategies to standard care once SGA is antenatally detected.

Figure 5.1 - Economic model to explain the expected pathways through which GAP would impact on cost-effectiveness
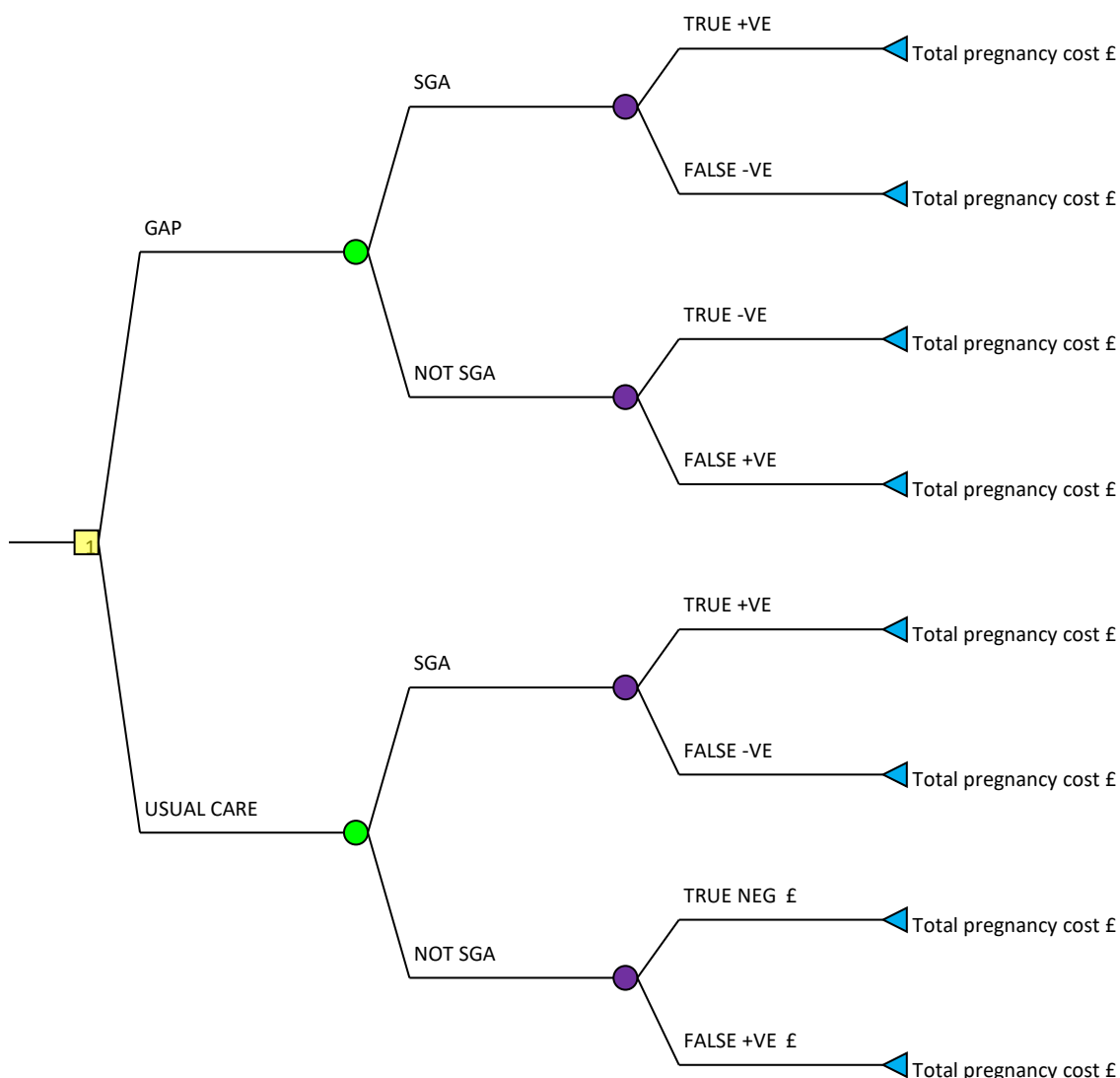
Green nodes represent the probability that a pregnancy was SGA_{both}, which was not expected to be affected by GAP. Purple nodes represent the probability of each screening programme and blue notes represent the individual pregnancy cost incurred by resource use for each screening outcome; both purple and blue nodes were hypothesised to be affected by GAP.

5.2.7.4 *Estimating the effect of GAP on SGA screening outcomes*

Cluster summary values for the proportion of pregnancies in which the infant was confirmed as either SGA or not SGA at birth were obtained from the pregnancy level data, based on the observed means for each cluster site. The clinical effect was estimated using a similar approach to the primary trial analysis, using cluster summary statistics (section 2.2.6.3.1):

1. For SGA births, adjusted cluster summary values for the true positive (TP) and true negative (TN) rates during the pre-randomisation and trial outcome phases were estimated from a multivariable logistic regression model that included adjustment for hospital site fixed effects and maternal ethnicity, age and parity.

The TP and TN estimates were derived as the cluster-specific mean predicted probabilities of observing the outcome using separate models including only SGA births for the TP value and only non-SGA births for the TN values.

2. The false positive (FP) and false negative (FN) values were derived as from the TP and TN values (FP=1-TN; FN=1-TP).

3. Cluster summary expected values for the trial outcome phase were then included into a further linear regression model. This involved regressing the logit transformation of the cluster-level outcome proportions on the expected pre-randomisation cluster summary value, adjusting for the randomisation strata (as per section 2.2.6.3.1), to identify the effect of the intervention on the proportion of births observed with each of the screening outcomes.

4. At a later stage, predictions from this model were re-transformed back to the original proportions scale, ensuring that the predicted proportions by trial arm lie between 0 and 1.

5.2.7.5 *Calculating the incremental antenatal cost attributable to GAP*

We expected that GAP would affect both antenatal costs applied to each screening outcome as well as the rate of each outcome. Because we intended to identify any treatment effect of GAP on resource use, particularly for pregnancies that screened positive, we restricted the estimation of antenatal cost for each of four screening outcomes (TP/FP/FN/TN) to clusters in the modified intention to treat sample (excluded cluster sites 12 and 13 who but did not contact the Perinatal Institute to implement GAP as randomised).

Cluster summary values were predicted by fitting a generalised linear model (GLM) to pregnancy level cost data, adjusting for age, ethnicity, parity and BMI. The model included screening outcome as an explanatory variable and fixed effects for cluster site. The predictive models were estimated separately for the pre-randomisation and trial outcome phases, the latter providing the cluster summaries and the former used to obtain a baseline prediction for each cluster. The baseline predictions were required to make adjustments for pre-existing antenatal cost differences when comparing costs by trial arm at the second stage (cluster-level) analysis. For the model fitted to the comparison period data, screening outcome was also interacted with the cluster identifiers because we expected that GAP would affect antenatal resource use differently for alternative screening outcomes at each cluster (i.e. we expected that resource use would be more affected for pregnancies screening positive for SGA, regardless of whether they were FP or TP).

5.2.7.6 *Calculating the incremental intrapartum, postnatal and neonatal costs attributable to screening outcomes*

Protocols of SGA management did not differ between GAP and standard care and we therefore expected that GAP would only affect costs that followed screening by altering the rate of each screening outcome, but not by the resource use related to those screening outcomes. Cluster values for costs by screening outcome were therefore obtained for intrapartum, post-natal and neonatal care, regardless of GAP implementation. Again, these were obtained by fitting GLM models to pregnancy-level cost data, using a model that included all trial clusters where relevant data were available and all pregnancies occurring during either trial phase (to maximise information on cost yielded by the trial data). As before, the models included adjustment for pregnancy-level characteristics, the screening outcomes and a fixed effect for cluster site. A further adjustment was also made to births that occurred during the trial outcome phases, using the cluster summary values from the pre-randomisation phase.

5.2.7.7 *Cost-effectiveness analysis*

Statistical parameters required to simulate probability distributions for the value of all parameters required for the cost-effectiveness analysis were extracted from the cluster summary data: proportion of births SGA or not SGA, proportion of SGA births with a TP/FN screening outcome, proportion of not-SGA births with a TN/FP screening outcome, and costs associated with each screening outcome (separately for antenatal, intrapartum, postnatal and neonatal care, and total pregnancy cost).

Probabilities distributions were simulated in Excel using Monte Carlo simulation. The simulated probability distributions for all parameters were combined to: identify an ICER (incremental cost per additional SGA detected antenatally) based on the ratio of the difference in the expected total pregnancy cost per 1000 births between GAP and standard care and the difference in the expected number of pregnancies that result in an SGA birth being detected antenatally, and to quantify uncertainty around the central (expected) cost-effectiveness estimate.

5.2.8 Sensitivity analyses

We pre-planned to test the effect of the clinical outcome definition (SGA_{both}) by repeating the analysis using the secondary outcome definitions of SGA detection – SGA was defined by GROW charts in the intervention arm and by population references in the standard care arm.

For the primary economic analysis, we were unable to include resource-use data on antenatal clinic appointments and unplanned antenatal day attendances to the maternity unit (e.g. maternity triage or day assessment unit) because of variability in the availability of data on antenatal appointments (missing data from some sites on midwifery appointments, or on community appointments, or because appointments were frequently offered by general practitioners at one cluster) and difficulty in obtaining complete records of unplanned hospital attendances. We therefore also conducted post-hoc sensitivity analyses to:

1. Include antenatal clinic appointment resource use and cost in the antenatal cost subtotal, but only for maternity clusters in which we were not aware of systematically missing data (i.e., we excluded units, amongst others, who provided no data on midwife-led antenatal appointments).

2. Include unplanned antenatal hospital attendance resource use and cost in the antenatal cost subtotal. We defined this as recorded attendance to a maternity triage unit, an antenatal day unit, or an antenatal day admission.

5.3 RESULTS

The study population was the same as that described for the main DESiGN trial analysis (section 2.6.2 with consort Figure 2.10). The population included data from 24,906 women and babies during the outcome phase and adjusted using data from 55,950 women and babies with births during the pre-randomisation (baseline) phase. Characteristics were previously described in Table 2.16 and Table 2.17. Data availability during the trial outcome phase for each activity item studied is summarised in Table 5.4, presented by trial arm. Neonatal care is not included in the table because following application of the assumption that all babies without a neonatal care record had not required additional care, the data were 100% complete.

Table 5.4 - Summary of data availability and completeness for maternal and neonatal care activity studied

Activity item	Number of sites contributing any data		Completeness at sites where data were available	
Antenatal care	**GAP**	**Standard care**	**GAP**	**Standard care**
Women with a record of ultrasound before 24 weeks'	5	6	80.9%	82.4%
Women with a record of antenatal appointments before 32 weeks	5	4§	49.0%	70.9%
Women with a recorded antenatal admission*	5	4	25.7%	31.2%
Women with a recorded unscheduled antenatal attendance*	4	4	57.5%	70.9%
Intrapartum care				
Women with recorded onset of labour	5	5	97.7%	99.4%
Women with recorded mode of birth	5	5	97.4%	99.8%
Women with a record of labour analgesia (including none)	5	5	99.8%	82.6%
Women with a record of perineal trauma (including none)	5	5	85.0%	74.5%
Women with a record of blood loss at birth	5	5	97.4%	99.4%
Postnatal care				
Women with a record of postnatal length of stay	3±	5	85.0%	99.0%

§ Three sites excluded later because they provided data only on consultant-led but provided either no data on midwife-led appointments, or only data on midwife-led appointments held in hospital settings.
± Also excludes one site which provided some data but for only 15% of births.

For estimation of resource use, availability and quality of data varied by cluster and activity. Sites were excluded if comprehensive data were not provided on all components of resource items (i.e., sites that did not provide data on midwife-led antenatal appointments were excluded from contributing resource data on any type of antenatal appointment). The number of cluster sites and births which contributed data on each resource type or cost within both the standard care and intervention arms are detailed in Table 5.5. There were no unadjusted or adjusted differences in the frequency or intensity of the resource items studied at the cluster level, nor were there any unadjusted differences in cost including for subtotals of maternal and neonatal care (Table 5.6). Adjusted cost differences are presented with the results of the economic model further below.

Table 5.5 - Summary of unit resource use and costs using imputed data (where available)

Hospital activity	GAP		Standard care		Difference in activity		
	Number of sites (births) included N=5 (11,096)	Value	Number of sites (births) included N=6 (13,810)	Value	Unadjusted, mean (95% CI)	Adjusted, mean (95% CI)	P-value (adjusted)
Antenatal activity							
Number of scans after 24 weeks, mean(sd)	5 (11,096)	1.72 (1.67)	6 (13,810)	1.95 (1.88)	-0.09 (-0.99, 0.82)	-0.05 (-0.53, 0.43)	0.83
Antenatal length of stay, mean nights (sd)	5 (11,096)	0.97 (2.51)	4 (10,326)	0.80 (1.84)	0.14 (-0.19, 0.46)	-0.03 (-0.15, 0.09)	0.61
Number of antenatal appointments, mean(sd)	5 (11,096)	5.42 (3.11)	1 (2,361)	2.43 (2.86)	1.21 (-1.51, 3.93)	1.12 (-0.11, 2.35)	0.08
Number of unscheduled antenatal day attendances, mean (sd)	4 (8,882)	1.45 (1.89)	4 (10,326)	1.46 (1.80)	0.02 (-1.23, 1.26)	-0.08 (-0.89, 0.72)	0.84
Intrapartum activity							
Percentage of women with induced labour onset (%)	5 (11,096)	29.5%	6 (13,810)	26.9%	2.8% (-4.2, 9.8)	1.7% (-0.4, 3.7)	0.11
Percentage of women who had an epidural (%)*	5 (11,096)	23.7%	6 (13,810)	28.6%	-4.6% (-14.3, 5.1)	-0.4% (-4.2, 3.5)	0.86
Percentage of women with unassisted vaginal birth (%)	5 (11,096)	54.0%	6 (13,810)	54.5%	1.5% (-4.5, 7.5)	-0.1% (-2.6, 2.4)	0.94
Percentage of women with instrumental birth (%)	5 (11,096)	14.4%	6 (13,810)	14.1%	0.3% (-3.1, 3.6)	-0.1% (-1.6, 1.4)	0.87
Percentage of women with emergency caesarean (%)	5 (11,096)	16.7%	6 (13,810)	17.2%	-0.8% (-4.4, 2.8)	0.6% (-1.6, 2.8)	0.59
Percentage of women with elective caesarean (%)	5 (11,096)	14.6%	6 (13,810)	13.9%	-0.9% (-5.7, 3.8)	-0.6% (-1.5, 0.4)	0.24

Hospital activity	GAP		Standard care		Difference in activity		
	Number of sites (births) included N=5 (11,096)	Value	Number of sites (births) included N=6 (13,810)	Value	Unadjusted, mean (95% CI)	Adjusted, mean (95% CI)	P-value (adjusted)
Intrapartum activity (continued)							
Percentage of women requiring repair of 3rd/4th degree perineal tear (n/%)*	5 (11,096)	1.8%	6 (13,810)	1.9%	0.0% (-0.8, 0.7)	-0.1% (-0.6, 0.4)	0.78
Percentage of women with postpartum haemorrhage: 500-1500mls (%)	5 (11,096)	40.9%	6 (13,810)	36.6%	1.5% (-7.9, 10.9)	0.3% (-1.4, 2.0)	0.73
Percentage of women with postpartum haemorrhage: over 1500mls (%)	5 (11,096)	3.2%	6 (13,810)	3.5%	-0.6% (-1.6, 0.4)	-0.2% (-0.6, 0.2)	0.37
Postnatal activity							
Postnatal length of stay, mean nights (sd)	5 (11,096)	0.90 (1.81)	3 (7,944)	0.75 (1.65)	-0.06 (-0.56, 0.45)	0.04 (-0.08, 0.16)	0.51
Neonatal activity							
Number of babies admitted to neonatal intensive care (n/%)*	5 (11,096)	2.3%	6 (13,810)	2.9%	-0.2% (-2.2, 1.8)	-0.2% (-1.5, 1.1)	0.77
Length of stay in neonatal intensive care, mean days (sd)*	5 (11,096)	0.10 (1.05)	6 (13,810)	0.19 (2.28)	-0.07 (-0.21, 0.8)	-0.07 (-0.15, 0.01)	0.09
Number of babies admitted to neonatal high dependency unit (n/%)*	5 (11,096)	2.7%	6 (13,810)	2.4%	0.2% (-1.3, 1.6)	-0.2% (-1.1, 0.7)	0.64
Length of stay in neonatal high dependency unit, mean days (sd)*	5 (11,096)	0.18 (2.04)	6 (13,810)	0.19 (2.60)	0.00 (-0.14, 0.13)	-0.02 (-0.09, 0.05)	0.54
Number of babies admitted to special care baby unit (n/%)*	5 (11,096)	6.3%	4 (10,099)	5.3%	1.4% (-0.3, 3.0)	1.0% (0.4, 1.6)	0.002
Length of stay in special care baby unit, mean days (sd)	5 (11,096)	0.41 (2.68)	4 (10,099)	0.27 (2.22)	0.09 (-0.2, 0.38)	0.12 (0.02, 0.21)	0.01

⁺Sites were excluded if they did not provide data across the whole site for the measure, or some of its component parts; *These variables were not imputed.

Table 5.6 - Summary of cost data for each maternity period (antenatal, intrapartum, postnatal and neonatal) and for GAP implementation

Hospital activity	Intervention		Standard care		Unadjusted cost difference, mean (95% CI)
	Number of sites (births) included N=5 (11,096)	Cost, mean(sd)	Number of sites (births) included N=6 (13,810)	Cost, mean(sd)	
Total antenatal cost 1* (main analysis definition)	5	£828.73 (£829.23)	3	£730.31 (£748.23)	£108.47 (-£62.31, £279.26)
Total antenatal cost 2*	5	£1,530.92 (£949.04)	1	£1,044.88 (£870.20)	£196.73 (-£379.73, £773.19)
Total antenatal cost 3*	5	£1,216.17 (£1,030.85)	3	£1,109.69 (£971.86)	£121.91 (-£260.38, £504.21)
Total cost of GAP implementation	5	£10.97 (£0.63)	6	£0.00 (£0.00)	£10.79 (£10.23, £11.36)
Total intrapartum cost	5	£2,786.76 (£1,039.94)	6	£2,799.79 (£1,039.27)	-£33.10 (-£150.22, £84.03)
Total postnatal cost	5	£388.82 (£781.79)	3	£324.08 (£713.75)	-£24.03 (-£240.35, £192.28)
Total neonatal cost	5	£482.87 (£3,389.01)	4	£516.93 (£4,546.84)	-£46.40 (-£350.29, £257.48)

*Total antenatal cost 1 = ultrasound scans + antenatal admissions; Total antenatal cost 2 = ultrasound scans + antenatal admissions + antenatal appointments; Total antenatal cost 3 = ultrasound scans + antenatal admissions + unscheduled day attendances

The expected number of births with a true positive screening outcome was 22.6/1,000 (95% CI: 10.0-41.5) for GAP and 20.7/1,000 (95% CI: 9.6-37.0) for standard care. The expected number of births with all other screening outcomes, expressed as the number per 1,000 births, is detailed for GAP and standard care in Table 5.7. GAP was associated with a slightly higher antenatal cost (£1,508, 95% CI: £726-£2,321) than standard care (£1,276, 95% CI: £835-£1,825) for births with the true positive screening outcome. The costs for all four screening outcomes are presented for antenatal (separated by trial arm), intrapartum, postnatal, and neonatal care phases in Table 5.8.

Table 5.7 - Expected number of screening outcomes per 1,000 births in GAP and standard care trial arms

Screening outcome	GAP, n/1,000 (95% CI)	Standard care, n/1,000 (95% CI)
True positive	22.6 (10.0-41.5)	20.7 (9.6-37.0)
False negative	51.1 (29.6-78.2)	53.0 (31.9-79.1)
True negative	904.9 (869.4-934.8)	912.6 (879.9-939.2)
False positive	21.4 (9.3-42.1)	13.7 (9.2-19.4)

Table 5.8 - Expected costs for each phase of maternity or neonatal care, presented by SGA screening outcome

Screening outcome	Cost of antenatal care, £ (95% CI)		Cost of intrapartum care, £ (95% CI)	Cost of postnatal care, £ (95% CI)	Cost of neonatal care, £ (95% CI)
	GAP	Standard care			
True positive	£1,508 (£726 - £2,321)	£1,276 (£835 - £1,825)	£3,022 (£2,858 - £3,191)	£729 (£329 - £1,277)	£2,803 (£906 - £5,908)
False negative	£894 (£682 - £1,143)	£848 (£639 - £1,093)	£2,724 (£2,573 - £2,881)	£467 (£209 - £824)	£1,010 (£315 - £2,115)
True negative	£689 (£461 - £938)	£690 (£494 - £918)	£2,708 (£2,583 - £2,836)	£364 (£197 - £582)	£416 (£173 - £762)
False positive	£841 (£181 - £1,553)	£1074 (£653 - £1,616)	£2,801 (£2,673 - £2,934)	£561 (£303 - £900)	£2,351 (£984 - £4,336)

Overall, GAP is expected to cost an additional £34,450 per 1,000 live births (95% CI: -£111,298 to £192,610). The estimated incremental cost for each phase of maternity and neonatal care is detailed in Table 5.9.

.

Table 5.9 - Incremental cost of GAP compared to standard care

Incremental cost of GAP (per 1,000 births)	Expected cost, mean (95% CI)
GAP implementation cost	£10,796 (£9,600 to £12,033)
Antenatal	£4645 (-£94,845 to £107,000)
Labour	£1,122 (-£7972 to £10,985)
Postnatal	£1721 (-£10,742 to £17,269)
Neonatal	£16,165 (-£63,679 to £118,336)
Total pregnancy cost	£34,450 (-£111,298 to £192,610)

Combining the estimated additional overall cost with the estimated change in the rate of detection of SGA results in an expected mean ICER from implementing GAP of £19,463 per additional case of SGA detected. This estimate is associated with uncertainty: the probability of GAP being dominant over standard care (better clinical effect, lower cost) was 11.3% and the probability of GAP being more clinically effective but with higher cost is 44.1%. Conversely, the probability of usual care being dominant was 23.4% or being more clinically effective with higher cost was 21.1% (Figure 5.2).

Figure 5.2 - Cost-effectiveness plane demonstrating the incremental cost-effectiveness ratio of GAP implementation, with associated uncertainty

245
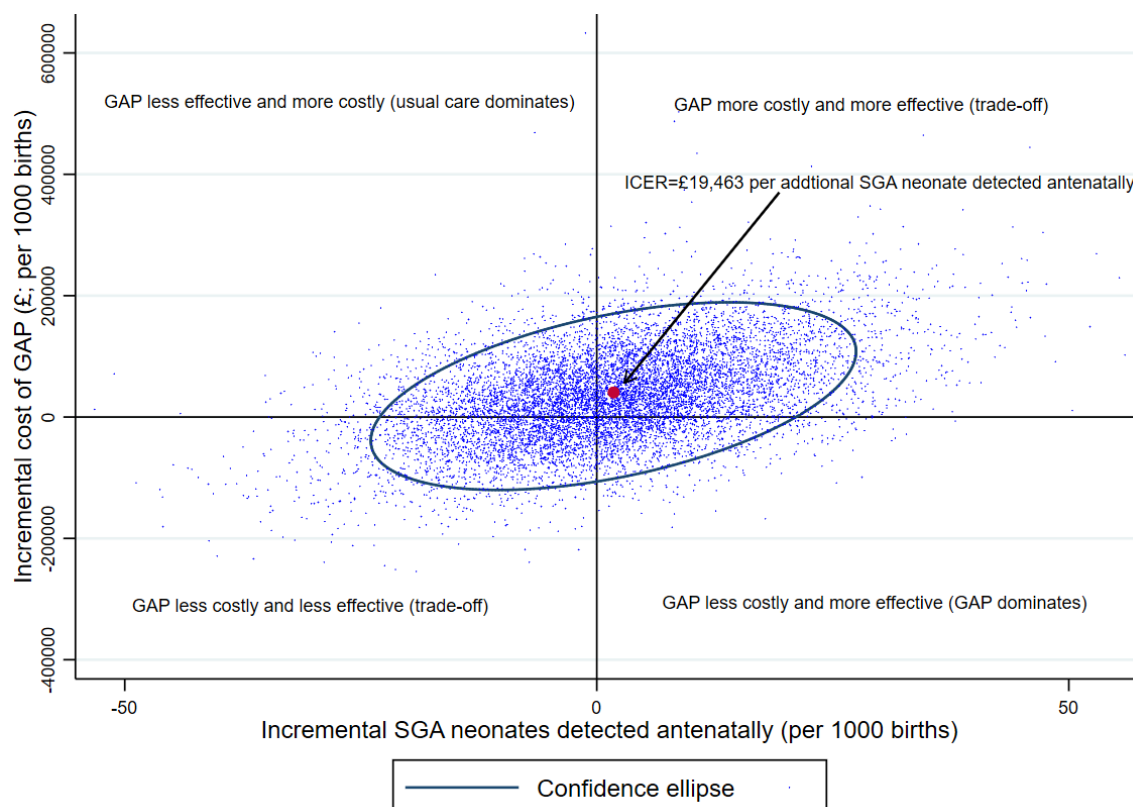
5.3.1   Sensitivity analyses

The incremental costs of GAP identified from all three sensitivity analyses are included in Table 5.10. When the primary clinical outcome definition was varied to be SGA_{cust} in the GAP-implementing clusters and SGA_{pop} in the standard care arm of the trial (sensitivity analysis 1), the incremental cost of GAP per 1,000 births fell for antenatal care to -£378 (95% CI: -£98,681 to £101.522), this was the primary contributing factor to the reduced total incremental cost per pregnancy (£30,504, 95% CI: -£123,625 to £197,527). Overall, the ICER was expected to an additional cost of £12,110 per SGA neonate detected, although this was associated with uncertainty – GAP being of higher cost and more clinically effective was still the most likely outcome (44.9% probability) on the cost-effectiveness plane (Table 5.11).

Table 5.10 – Incremental costs of GAP (sensitivity analyses)

Incremental cost of GAP (per 1,000 births)	Expected cost, mean (95% CI)		
	Sensitivity analysis 1 (varying clinical outcome definition)	Sensitivity analysis 2 (including cost of antenatal clinic attendances)	Sensitivity analysis 3 (including cost of unscheduled antenatal attendances)
GAP implementation cost	£10,796 (£9,600 to £12,033)	Unchanged	Unchanged
Antenatal	-£378 (-£98,681 to £101.522)	£5,244 (-£97,189 to £110,719)	£5,895 (-£98,121 to £113,604)
Labour	£1,308 (-£10,146 to £13,910)	Unchanged	Unchanged
Postnatal	£1,966 (-£11,850 to £19,316)	Unchanged	Unchanged
Neonatal	£16,812 (-£77,993 to £123,631)	Unchanged	Unchanged
Total pregnancy cost	£30,504 (-£123,625 to £197,527)	£35,050 (-£115,115 to £201,627)	£35,704 (-£116,321 to £206,027)

For the second and third sensitivity analyses, in which the removal of antenatal clinic and unscheduled attendance costs from the main analysis was tested by including these costs only for the clusters that provided data required to estimate them, the results differed very little compared to those from the primary analysis (Table 5.10). In both cases, the best estimate of the incremental cost of GAP increased slightly for the antenatal period, therefore also causing a slight increase in the overall pregnancy cost. The incremental costs from all other aspects of the maternity and neonatal pathways were unchanged.

Similarly, these inclusions had little impact on the expected (mean) ICER. Inclusion of antenatal clinic costs resulted in an ICER of £19,802 and inclusion of antenatal clinic costs resulted in an ICER of £20,173. In both cases, the most likely cost-effectiveness outcome remained that GAP would be more clinically effective and more costly, with a 44% probability of this outcome occurring.

Table 5.11 - Cost-effectiveness of GAP versus standard care (sensitivity analyses)

		Sensitivity analysis 1 (varying clinical outcome definition)	Sensitivity analysis 2 (including cost of antenatal clinic attendances)	Sensitivity analysis 3 (including cost of unscheduled antenatal attendances)
Expected (mean) incremental cost per SGA neonate detected antenatally (ICER)		£12,110	£19,802	£20,173
Probability (%) of cost-effectiveness outcomes:	**Usual care is dominant**	19.87	23.15	22.95
	GAP is dominant	10.33	11.43	10.95
	GAP higher cost, more SGA detected antenatally	44.91	43.97	44.45
	GAP lower cost, fewer SGA detected antenatally	24.89	21.45	21.65

5.4 DISCUSSION

5.4.1 Summary of the key findings

This trial-based evaluation of the cost-effectiveness of GAP in improving the rate of antenatal detection of SGA found that GAP was associated with an incremental cost increase of £19,463 per additional case of SGA detected. This estimate is associated with significant uncertainty, as demonstrated by the cost-effectiveness plane and probability of the outcome being consistent with each of the possible four: GAP was most likely (44.1% chance) to be more clinically effective but associated with a higher cost than standard care but there was a 44.5% chance that standard care was more clinically effective than GAP (regardless of associated cost), and an 11.3% chance that GAP was more clinically effective and less costly. The higher cost is attributable to the introduction of costs to implement GAP (£10,796 per 1,000 pregnancies) that were mostly mediated through staff training costs, and greater antenatal costs for true positive SGA diagnoses.

5.4.2 Interpretation of the findings

The economic impact of GAP has only previously been studied using a cost-benefit analysis that was limited to the costs associated with introduction of serial fetal growth scans for women at high risk of SGA, with frequency and indication depicted by RCOG guidelines. Using these methods, the Perinatal Institute team estimated that increasing the frequency of fetal growth scans for this group of women would cost an additional £10,000 per 1,000 pregnancies.[284] The expected benefit was a cost-saving of £120,000 per 1,000 pregnancies (net benefit of £110,000 per 1,000 pregnancies) through presumed reductions in neonatal admissions, perinatal morbidity and mortality, cerebral palsy and litigation that were expected to follow the higher rate of antenatal detection of SGA.[87] This analysis differs to that presented here for several reasons. Firstly, it did not consider the training costs associated with the implementation of GAP, including training on a standardised technique for fundal height measurement, plotting onto, and interpretation of the completed GROW chart. Secondly, the source of model estimates was not always clear but where it was, these were often derived from retrospectively conducted observational studies which are subject to bias. Thirdly, it was assumed that each additional ultrasound scan cost only £15 per scan in a service that was already functional, but this estimate is much lower than the NHS reference cost at the same time of £103.84 for a standard antenatal ultrasound scan.[284] Finally, the cost-benefit model did include expected cost savings related to litigation avoided, which our model did not include due to the time horizon of the study, but costs associated with iatrogenic birth interventions including iatrogenic preterm birth were not considered. These factors may explain the large difference in the direction and magnitude

of the conclusions, although it is important to note that the cost estimation derived from this cost-benefit analysis was at the lowermost limit of the plausible range of cost associated with GAP implementation calculated for the cost-effectiveness analysis presented in this chapter.

Whilst the findings of this analysis have suggested that there is a 55.4% chance that GAP was more clinically effective than standard care (44.1% more costly and effective, 11.3% more effective but less costly), this does not contradict the primary outcome analysis of the DESiGN trial that identified no difference in the rate of detection of SGA from either screening policy (25·9% vs 27·7%, adjusted difference 2·4%, 95% CI: -6·1% to 10·8%, p=0·58). It would be necessary for the economic evaluation to find that GAP had more than a 95% chance of being clinically effective (and more or less costly) for the clinical analysis to find a significant difference at the 95% confidence level. However, policy decisions regarding implementation of new interventions that consider economic evaluations are not made in the same manner as decisions which are based only on studies of clinical effectiveness.[401] From an economic perspective, rejection of GAP on the basis of the findings of the DESiGN trial would have the consequence of continuing standard care, a programme that is less likely to be either clinically- or cost-effective.

We found that GAP was expected to cost £19,463 per additional case of SGA detected, but NICE measure patient benefit in terms of quality adjusted life years (QALY), which are comparable across the healthcare spectrum. QALYs are a measure of patient/population outcome that combine survival and the quality of years survived into a single measure. NICE use a maximum threshold of £20,000-£30,000 per QALY gained to recommend whether an intervention should be introduced into generalised use in the NHS, although QALYs are not the only measures taken into account.[402] The time horizon of this RCT prevented calculation of QALYs gained or lost, expected to occur either through additional QALYs following prevention of stillbirth, or QALYs lost following iatrogenic preterm birth and associated infant, child or adult morbidity or mortality beyond that measured within the time to first postnatal/neonatal hospital discharge studied in this trial, particularly for those neonates who would not otherwise have been stillborn.

However, an estimate can be made of the potential cost per QALY gained through implementation of GAP by adopting point estimates published in the literature, and knowledge of rates of stillbirth in detected and missed cases of SGA in the DESiGN trial. It must be stressed that what follows is a rough calculation, with significant uncertainty arising from the confidence intervals around the point estimates, and probabilities suggesting that the outcome may have fallen within any of the four quadrants of the cost-

effectiveness plane. NICE have previously provided estimates that each prevented stillbirth results in 23.73 QALYs gained, although this is based on the assumption that babies in whom stillbirth is presented have a normal quality of life and life expectancy.[403] Using this QALY estimate and evidence from the DESiGN trial that one out of every 72 undetected SGA babies were stillborn, and that the best estimate would be that detection and intervention could halve the rate of SGA-associated stillbirth (i.e. 144 SGA babies must be detected to avoid one stillbirth),[404,405] I estimate that the cost of implementing GAP to avoid one stillbirth is £2,802,672 (£19,463 multiplied by 144). This equates to a cost per QALY gained of £118,106 (£2,802,672 divided by 23.73 QALYs gained per stillbirth prevented), which is above the NICE willingness to pay threshold of £20,000-£30,000, therefore suggesting that implementation of GAP is not likely to be considered cost-effective in the context of the UK NHS.[402] This is an oversimplification of the calculation because the gain in QALYs though stillbirth prevention is offset by a likely QALY loss elsewhere through lifelong morbidity caused by preterm birth of SGA infants who might not otherwise have been stillborn,[196,406] but this would mean that the cost per QALY gained overall is likely to be much higher than that presented. Conversely, this estimate must be interpreted in the context of the estimated societal and long-term indirect healthcare costs associated with stillbirth which have not been accounted for, these include parental time off work including for future mental health, staff wellbeing, bereavement counselling or mental health care and costs of litigation or investigation.[379]

Finally, these findings must be interpreted in the context of the implementation strength seen in this trial, and already reported in Chapter 3, section 3.3.1. This evidenced that women at high-risk of SGA were not offered fetal growth scans as frequently as recommended by GAP, and that women at low risk of SGA were not all referred for a growth scan when indicated by the fundal height measurement plotted on the GROW chart. Practice that is concordant with GAP is therefore expected to cost more than that which was observed. However, the effect that this would have had on the rate of detection of SGA, and therefore on the additional cost of GAP per case of SGA detected (or stillbirth prevented) is not known.

5.4.3 Strengths and limitations

The main strength of this economic evaluation is that it was conducted using data on resource use recorded routinely during clinical practice and is therefore expected to reflect the true level of resource required to implement GAP in this setting, as compared to standard care. GAP was compared to standard contemporaneous practice, rather than to no care, and so the findings reflect the expected cost of implementing GAP over and above those of current practice, as is recommended by NICE.[401] The primary analysis did not include

economic assumptions drawn from other research or models, meaning that the findings reflect the reality of what was observed when GAP was pragmatically implemented into clinical practice.

One limitation includes the cost perspective and time horizon. The costs associated with GAP for the detection of SGA and potential stillbirth prevention were restricted to those incurred by the NHS trust up until the time of hospital discharge following birth. As previously discussed, costs of additional infant, child or adult care which occur following preterm birth, particularly in the early third trimester, have not been accounted for, nor have societal costs associated with parental time off work (in addition to maternity leave), long-term costs of providing emotional or mental health care to parents or hospital costs associated with either clinical governance procedures or litigation. However, GAP is expected to mediate its effect through screening and detection of SGA and costs subsequent to SGA detection, or as a result of missed SGA, are not expected to differ between GAP and standard care which both employ guidelines on SGA management from the RCOG.[1] The main exception to this is that GAP may have an effect on gestational age at the time of SGA detection, and therefore on gestation at timing of birth, which may affect costs of mode of birth and neonatal care.

Another limitation includes the availability and quality of data from some clusters. Data were missing on true negative and false positive SGA diagnoses at one cluster site allocated to GAP implementation. This site was therefore excluded from contributing to cost calculations for these screening outcomes. Hospital administrative data were entirely missing from one cluster site allocated to standard care and were not usable for another site allocated to standard care (only first antenatal appointment provided, no data on length of stay during antenatal/postnatal admissions). There were also problems caused by lack of electronic recording of midwifery antenatal appointments at the remaining three clusters allocated to standard care –data on these were either not provided at all, or only provided for appointments which occurred in the hospital setting. In all clusters, we were unable to distinguish between women who had absence of an activity recorded because it had occurred at a different maternity unit, because the woman had not received care at all, or because it had occurred at the cluster site but had not been recorded, we introduced assumptions in which plausible limits were applied to deal with this. Finally, some cluster sites only kept paper records of unscheduled hospital attendances to maternity day units or triage. The quality of antenatal data disproportionately affected cluster sites allocated to standard care meaning that antenatal appointments and unscheduled attendances were completely excluded from the primary cost-effectiveness analysis. It is feasible that antenatal detection of SGA leads to an increase in either of these maternity activities, but

these follow SGA diagnosis and so intensity of resource use is not expected to differ between GAP and standard care – this was tested through sensitivity analysis in which I found very little difference was made to the incremental cost-effectiveness ratios or likely outcomes following inclusion of either activity in the economic model.

Another limitation of this analysis is the choice of primary clinical outcome definition – detection of SGA was defined only amongst babies whose birthweight was below the 10th centile for gestational age both by the population reference and customised standard charts. The reasons for this choice have previously been described (section 2.2.1.1). This meant that false positives were defined in the group of babies who did not meet the SGA definition by both birthweight charts, this includes approximately 5% of babies who are defined as SGA by one but not both birthweight charts. The effect that this choice had on the ICER was studied through sensitivity analysis which also concluded that the most likely cost-effectiveness outcome was that GAP was more clinically effective and more expensive, but with a slightly lower ICER (£12,110), largely driven by negative incremental costs of GAP on antenatal care.

5.4.4 Implication of the findings

The DESiGN trial found that GAP was no more clinically effective than standard care (section 2.6.3) and this economic evaluation of GAP implementation in the trial has found that whilst GAP (as implemented) was most likely to be more effective and more costly than standard care, it cannot be recommended for further implementation in the UK at the current willingness to pay threshold given that the additional cost per SGA detected was £19,463, which translates into an estimated cost per QALY gained by preventing stillbirth in excess of £100,000. However, this finding must be interpreted with caution given the uncertainty presented in the economic model and that GAP was implemented with lower fidelity and dose (section 3.3.1) than recommended by the Perinatal Institute. The effect of this on the rate of SGA detected and the use of healthcare resource has not been quantified.

The findings and implications of this economic evaluation are generalisable only to healthcare systems that have similar resource constraints as the clusters included within the DESiGN trial and with willingness to pay thresholds of less than £100,000 per QALY gained. In the context of this trial, these resource restrictions were shown qualitatively (section 3.3.2.3) to have likely had implications on strength of implementation and therefore possibly on clinical effectiveness in this setting.

Further evidence is therefore required to determine the clinical- and cost-effectiveness of GAP when implemented as recommended, in comparison to standard care. Unfortunately,

the DESiGN trial is believed to have been the last opportunity to conduct this sort of trial in the UK setting,[276] given that 78% of UK maternity units have now implemented GAP.[270] Whilst a randomised control trial of GAP could be repeated elsewhere, it would only be informative of UK practice if conducted in a country where standard care was comparable to that in the UK, particularly where this includes a similar schedule of antenatal care as that recommended by NICE,[8] and selective offer of serial fetal growth scans to women at highest risk of SGA. Furthermore, cost-effectiveness findings of an internationally conducted trial may have limited translation to UK settings, given the previously documented variation in international costs of maternity care.[352] For these reasons, it may not be possible to repeat this economic evaluation of GAP in a setting which is generalisable to UK practice and this may be the last example to inform UK policy on this research question.

5.4.5 Conclusion

This economic evaluation comparing the cost-effectiveness of GAP to standard care in terms of the rate of detection of SGA has identified that whilst GAP is most likely to be more clinically effective and costly than standard care, its likely cost per additional case of SGA detected is £19,463 (albeit with significant uncertainty surrounding this point estimate). This point estimate is most likely to translate into a cost per QALY gained by preventing stillbirth in excess of £100,000. Since this cost per QALY gained is above the national willingness to pay threshold of £20,000-£30,000 as determined by NICE, GAP cannot be recommended for prioritisation over and above other healthcare interventions which offer greater patient benefit for less cost. This finding is associated with both uncertainty in the economic model and a lower than recommended implementation strength of GAP in the DESiGN trial. Both these factors must be considered when using these findings to make national policy decisions.

6 CHARACTERISTICS ASSOCIATED WITH MISSED ANTENATAL DIAGNOSES OF SGA: A CASE-CONTROL STUDY

6.1 INTRODUCTION

The reduction of stillbirth and perinatal death is an international priority, with the WHO Global Strategy-supported Sustainable Development Goal (SDG 3) aiming to end preventable newborn death and stillbirth by 2030.[407] The UK Department of Health and Social Care also committed to halve the rate of stillbirth and neonatal deaths between 2015 and 2025.[332] It is accepted that up to half of stillborn babies are growth restricted in utero,[408,409] and therefore that improvements in antenatal detection of SGA and appropriate perinatal care should be targeted to reduce stillbirth.[246]

Routine antenatal care in the UK involves screening for fetal growth anomalies through fundal height measurement and targeted ultrasound for low-risk women or serial ultrasound assessment for women at high risk of SGA. This is associated with a rate of detection of SGA under 50%.[4,210-216] Ultrasound screening is also offered to women with diabetes - who are at risk of disproportionate macrosomia, or because other screening methods (fundal height measurements) are unreliable in some groups of women (section 1.4.1.1).[1,217,357] Alternatively, universal serial ultrasound screening has been shown to detect a higher proportion of SGA when studied under trial conditions but it does not have such success when used in routine care.[211,212] Such universally-applied serial ultrasound is associated with a high false-positive rate,[211] which may lead to a cascade of avoidable intervention and adverse outcomes. There is no UK consensus on how frequently serial ultrasound scans should be offered. GAP recommends that scans are conducted 3-weekly starting between 26-28 weeks' gestation,[74] but RCOG guidelines do not define an expected frequency.[1] Clusters recruited to the DESiGN trial offered screening which was heterogeneous both between, and within clusters (i.e. different frequency for each indication), see Table 3.4.

The improvement of strategies that target antenatal detection of SGA by increasing the sensitivity without major detriment to specificity, requires an understanding of the characteristics of women and babies in whom SGA is not detected. In a Danish study, Andreasen et al (2020) compared demographic and service characteristics for women in whom FGR (defined as birthweight<2.3rd centile on population weight charts) was or wasn't detected antenatally (detection rate of 31% in 3,069 pregnancies with FGR).[410] As in the UK, Danish women are selectively offered fetal growth scans by indication. Detection of FGR was more likely for multiparous women (especially those who previously had an FGR baby), of lower BMI, who had assisted conception, and/or those who saw a midwife with over 10 years of experience at the first antenatal appointment. In a single centre Australian study, Diksha et al (2018) compared characteristics for FGR babies (birthweight <3rd centile) born

with birth planned for FGR at 37[+0]-39[+6] weeks' (n=187) to those who were undelivered by 40[+0] weeks (n=233).[411] They found that there was no difference in the prevalence of risk factors for FGR, but that babies born after 40[+0] weeks' were less likely to have had a third-trimester fetal growth scan, more likely to have had a reassuring fetal growth scan (EFW or AC>10th centile) and more likely to have been cared for in low-risk midwifery-led settings than in collaborative care models, the former two associations are unsurprising when a non-reassuring third trimester scan is necessary to diagnose FGR. Similar studies have not been conducted in the UK. Given that detection of SGA (not FGR) is the target of most UK antenatal care,[1,217] and that risk stratification and screening strategies may differ between countries, these findings may not be translatable to UK practice.

This analysis aimed to identify the clinical and ultrasound utilisation characteristics of pregnancies in which an antenatal diagnosis of SGA is missed, compared to those in which it is made, to understand how we can better target interventions intended to improve detection.

6.1.1 Objectives

The objectives of this analysis were to:

1. Compare the characteristics of women and babies in whom SGA was missed antenatally, to those in whom SGA was detected.
2. Determine what characteristics are associated with missed SGA.
3. Describe the patterns of ultrasound usage when screening for SGA in each group and determine whether these differ by the presence of risk factors for SGA.

6.2 METHODS

6.2.1 Study design

A case-control study design was planned to compare pregnancies in which SGA was missed antenatally (cases) to pregnancies in which it was detected (controls). The objectives were addressed using the data collected on pregnancies included in the DESiGN trial. Given that the DESiGN trial did not find a difference in the rate of the primary outcome (SGA detection) between trial arms, it is suitable for use in this case-control study.

6.2.2 Reporting checklist

This study has been reported according to the recommendations of the STROBE statement and checklist for case-control studies.[314] The completed checklist is included in Appendix section 10.14.

6.2.3 Study population

This analysis has been restricted to women and babies who met the criteria for inclusion in the DESiGN trial (section 2.1.3). The sample was limited to births occurring at sites that provided data on co-morbidities and antenatal complications, which were among the key characteristics studied. The complete case analysis was also limited to pregnancies for which we had complete data on fetal growth status (birthweight, gestational age and maternal characteristics required to calculate the customised centile, section 1.1.7). Women and babies in whom SGA detection status could not be determined because data on ultrasound were missing during an entire trial phase at the cluster site (pre-randomisation and washout phase for site 8, washout phase for site 11) were also excluded.

6.2.4 Defining cases and controls

The cases ('missed SGA') were defined as pregnancies in which the baby was diagnosed as being SGA at birth, but for whom there was no evidence that an antenatal ultrasound diagnosis had been made i.e., the woman had not received growth scans, or the EFW at the last fetal growth scan was above the 10th centile for gestational age (not SGA). The controls were pregnancies in which an antenatal diagnosis of SGA had been made i.e., the EFW at the last fetal growth ultrasound recorded in the electronic patient ultrasound record was below the 10th centile for gestational age. SGA at birth was defined as a birthweight that was below the 10th centile for gestational age on both population and customised centile charts (SGA_{both}) – rationale described in section 6.2.4.1 below. For both cases and controls, the EFW was judged against the fetal weight charts used in the cluster at that time (GROW charts

when GAP had been implemented in sites allocated to GAP in the trial, Hadlock population charts for non-GAP clusters and pre-GAP trial phases).

6.2.4.1 *Considerations made when determining the SGA definition used*

In the DESiGN trial and the cluster sites recruited to it, SGA was defined in two ways: by population or customised centiles. For this analysis, where pregnancies in both arms of the trial were treated together in the sample, a unified definition was required. The options were:

- Define SGA by population centiles (approximately 10% of all babies for whom this definition was applicable)
    - Population centiles were used to define and detect SGA amongst all babies born during the pre-randomisation and early washout phase of the trial arm allocated to GAP and all trial phases of the control arm.
    - However, approximately 80% of NHS Trusts in the UK now use customised centiles. Application of customised centiles defines SGA in different babies than when using population centiles, and so an analysis of predictors of SGA detection using population centiles only is, therefore, less relevant to current UK practice.
- Define SGA by customised centiles (approximately 10% of all babies for whom this definition was applicable)
    - Customised centiles were used to define and detect SGA only amongst babies born during the outcome and late washout phases of the trial arm allocated to the intervention. This would limit the size of the population and the power to conduct an analysis.
- Define SGA in babies for whom the criteria are met by both population and customised centiles (approximately 7.5% of all babies, regardless of trial arm)
    - These are the SGA babies most at risk of stillbirth (section 1.1.9). The definition is relevant to all babies born in the trial because both trial arms were intended to detect SGA in these babies, regardless of whether the cluster sites were using population or customised centiles. However, this definition alone is not as clinically useful as the other two definitions, because it is not used to define SGA in practice.

In this analysis, SGA at birth has been defined as those babies who are SGA by both population and customised centiles (SGA_{both}). This maximised the sample size and studied the babies who are at the highest risk of stillbirth. Sensitivity analysis was conducted to

repeat all analyses using the SGA$_{pop}$ sample. The SGA$_{cust}$ sample was small and therefore not studied.

6.2.5   Exposures

6.2.5.1   *Maternal and fetal factors associated with missed antenatal diagnosis of SGA$_{both}$*

The characteristics studied to determine maternal or fetal factors which are associated with missed SGA are listed in Table 6.1. The maternal co-morbidities and antenatal complications were chosen because each is an indication for serial fetal growth scans in pregnancy and we had access to data on these items. Drug use, history of a previous stillbirth of SGA baby, renal impairment, large fibroids, fetal echogenic bowel and significant antepartum haemorrhage are also indications for serial fetal growth scans,[217] but data on these were not collected for the DESiGN trial, and are often not collected in electronic maternity records. PAPP-A was categorised as low using two thresholds (0.3 MoM and 0.415 MoM) because both were used to indicate the risk of SGA in DESiGN trial clusters. Birthweight centile was defined by the chart used at the time and cluster site of the birth (allocated centile).

Table 6.1- Planned maternal and fetal factors to be studied for association with SGA$_{both}$ that was missed antenatally

	Factors studied
Maternal demographics	• Age (continuous and binary – above/below 40 years at the time of the baby's birth) • IMD quintile • Ethnicity (white, Black, Asian, mixed, other)
Maternal clinical characteristics	• BMI (categorical using WHO categories) • Parity (0,1,2,3,≥4) • Smoking status in pregnancy (smoker, non-smoker)
Maternal co-morbidities	• Pre-existing diabetes • Pre-existing hypertension • Anti-phospholipid syndrome
Antenatal complications	• Pre-eclampsia • Pregnancy-induced hypertension • Gestational diabetes
Other obstetric factors	• Low placenta-associated plasma protein A (<0.3 MoM, 0.3-0.415 MoM, ≥0.415 MoM, missing) • Non-cephalic fetal presentation at birth • Birthweight centile (continuous or <3rd centile, 3rd-5th centile, 5-10th centile)
Composite characteristic	• Any indication for serial fetal growth scans – age>40years, BMI≥35 kg/m², smoking, any of the above maternal co-morbidities or antenatal complications, PAPP-A<0.415 MoM.

Given that the sample was drawn from a randomised control trial dataset, I also planned to document how the rate of SGA detection varied by trial phase (pre-

randomisation, washout, or outcome phase) or intervention group (standard care, GAP implementers, clusters allocated to implement GAP that withdrew early).

6.2.5.2 *Patterns of ultrasound usage associated with detected or missed SGA$_{both}$*

Growth scans were defined as in the DESiGN trial (any scan occurring after 24^{+0} weeks', in which EFW was documented or could be calculated from the available biometry). Scans that only assessed measures of amniotic fluid or fetal Doppler studies were excluded. Fetal growth scans were categorised into screening or surveillance scans based on when the EFW was first calculated below the 10th centile. This is detailed below.

- For pregnancies in which SGA was detected:
    - All scans before and including the first scan with EFW<10th centile were categorised as screening scans;
    - All scans after the scan at which the EFW was first below the 10th centile were categorised as surveillance scans.
- For pregnancies in which SGA was missed, all scans were considered to be screening scans.

The patterns of ultrasound scans studied were:

- Presence of a record for a presumed fetal anomaly scan at the same cluster as that at which the birth occurred (anomaly defined as any scan conducted between 18^{+0} to 24^{+0} weeks').
- Gestational age at the time of first fetal growth scan
- Mean frequency of serial screening scans:
    - Continuously reported i.e., one scan every *n* weeks
    - Categorical: 3-weekly or more, 4-weekly, less than 4-weekly)
- Duration from the last (screening or surveillance) scan until birth.
- Difference between the EFW at the last scan and the birthweight expressed in terms of:
    - Absolute number of centiles
    - Weight difference as a percentage of the birthweight

The calculation of mean screening frequency accounted for preterm birth and indications which arise later in pregnancy (e.g., pre-eclampsia), by dividing the period from the gestation of the first scan until the gestation of the last screening scan, by one less than the number of screening scans performed during the period. For this, only pregnancies that had at least two screening scans could be included.

6.2.6 Management of missing data

Patterns of missing data were summarised for each characteristic and outcome using descriptive statistics. Missing data were multiply imputed as described in section 2.2.4.2. The primary analysis of factors associated with SGA_{both} detection status used imputed data on demographics and growth status, and only included records of pregnancies with complete data on the non-imputed co-morbidities and antenatal complications. For evaluation of fetal presentation, it was also necessary that the record had complete data on this item. Since PAPP-A is an important characteristic (it is a major indicator for serial fetal growth ultrasound), but there was wide variation in its availability, missing data on PAPP-A was included as an exposure category, and in addition to the principles above, PAPP-A was only studied in sites which provided data on it.

For the study of ultrasound patterns, it was assumed that pregnancies without a record of a fetal growth scan had not had a scan. This assumption is tested in the third sensitivity analysis (section 6.2.7.1 below). For the analysis stratified by presence or absence of an indication for serial fetal scans, only women who had complete data on co-morbidities and antenatal complications, and who gave birth to their baby at a site that provided any data on PAPP-A were included. In this analysis, missing PAPP-A was not treated as an indication for serial fetal growth scans.

6.2.7 Statistical analysis

The number and proportion of pregnancies in which the baby was SGA at birth, as defined by SGA_{pop}, SGA_{cust} or SGA_{both} definitions, was calculated. The rate of SGA detection was presented over time (split into 4-month periods), by the trial intervention group.

Characteristics of pregnancies in which SGA was missed were summarised using descriptive statistics (section 2.2.6.1). Characteristics of pregnancies in which SGA was missed were compared to those of pregnancies in which SGA was detected using univariate logistic or linear regression and presented as odds ratios or mean differences. Multivariate comparisons were conducted by adjusting the relationship between each exposure and the outcome (missed SGA) by all other demographic and clinical characteristics, the allocated birthweight centile of the neonate, and the maternal co-morbidities and antenatal complications, but not by PAPP-A or fetal presentation. Because the data were collected from a cluster trial population, all models were adjusted by the cluster site and an interaction parameter between the trial phase and the intervention group.

Patterns of screening ultrasound utilisation were also summarised using descriptive statistics and univariate comparisons as described above. Comparisons were adjusted using

trial factors only (site and interaction parameter between the trial phase and intervention group). To determine the impact of ultrasound patterns on the rate of detection of SGA amongst women with and without indications for serial fetal ultrasound scans, the comparisons were stratified by the presence or absence of an indication.

6.2.7.1 *Sensitivity analyses*

The analyses were repeated to determine whether any of the methodological choices had influenced the findings.

1) The analysis was repeated using non-imputed data.
2) An analysis was conducted using the SGA_{pop} pregnancies, regardless of whether they were also defined as SGA_{cust}. In this case, pregnancies were excluded if it was expected that fetal growth had been judged clinically using customised centiles because the site had implemented GAP.
3) The analysis was repeated but restricted to pregnancies in which there was evidence of a *presumed* anomaly scan in the cluster to determine the effect of having continuous third-trimester care at the same cluster site, in the context of the assumption that the woman had not received a fetal growth scan if it were not conducted at the site in which she gave birth.

6.3 RESULTS

6.3.1 Description of data quality

Four sites did not provide data on the studied maternal co-morbidities and antenatal complications, these were excluded leaving 45.9% (n=4,596) of the SGA$_{both}$ individual pregnancy records with complete data on demographic and clinical characteristics, co-morbidities and antenatal complications, this rose to 78.4% following imputation of demographic data. In this imputed dataset, only 44.3% of women had data on PAPP-A and 47.7% had data on fetal presentation (Figure 6.1). A summary of the missing data in the included SGA$_{both}$ sample, stratified by SGA detection status, is included in Table 6.2. Missing data were distributed similarly between cases and controls.

Table 6.2 - Description of missing data for each characteristic in the available case SGA population, stratified by detection status

Missing characteristic	Proportion of data which were missing			
	Missed SGA (n=7,753)		Detected SGA (n=2,269)	
	n	%	n	%
Age*	751	9.7%	153	6.7%
Index of multiple deprivation*	90	1.2%	28	1.2%
Ethnicity*	501	6.5%	113	5.0%
Body mass index*	1,206	15.6%	321	14.1%
Parity*	915	11.8%	207	9.1%
Smoking status	515	6.6%	148	6.5%
Pre-existing hypertension	45	0.6%	20	0.9%
Pre-existing diabetes	156	2.0%	22	1.0%
Antiphospholipid syndrome	2,045	26.4%	528	23.3%
Pre-eclampsia	1,049	13.5%	256	11.3%
Pregnancy-induced hypertension	445	5.7%	40	1.8%
Gestational diabetes	28	0.4%	5	0.2%
PAPP-A+	3,693	69.6%	928	50.2%
Fetal presentation at birth+	118	1.8%	57	3.0%

These data were imputed for the primary analysis.

+Excludes sites that did not provide any data on these characteristics.

NB. Data on gestational age and birthweight were not missing in this population because the completeness of these characteristics was required to define SGA$_{both}$.

Figure 6.1 - Consort diagram detailing the construction of the study population (imputed data)
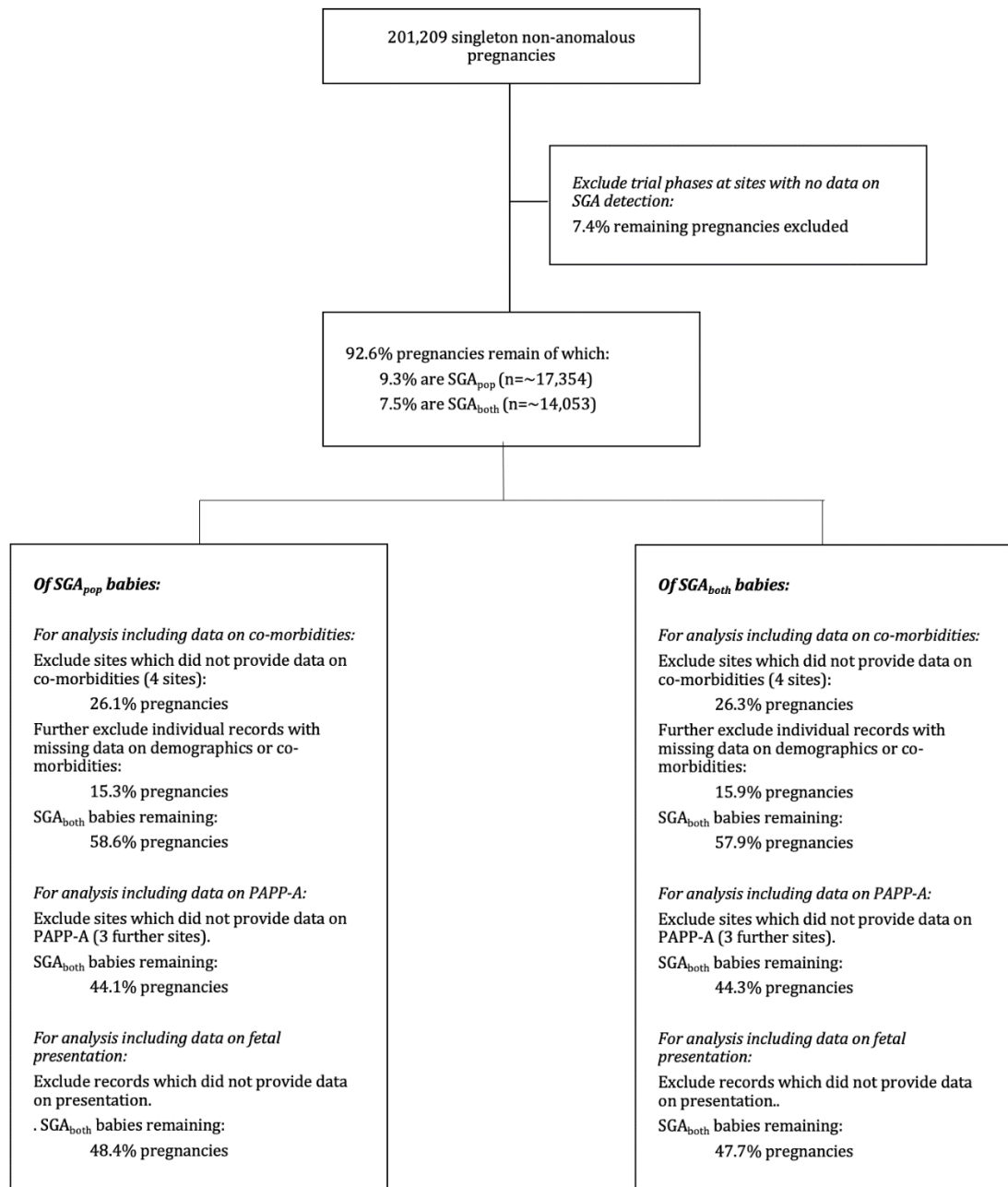
6.3.2 Description of the study population

Of the 201,209 singleton non-anomalous pregnancies included in the whole DESiGN trial, 7.4% were excluded because they occurred during trial phases for which the two clusters did not provide data on SGA detection. Of the remaining pregnancies, 7.5% were SGA_both and 9.3% were SGA_pop (Figure 6.1). This was the sample used to describe patterns of ultrasound utilisation.

For the study of maternal and fetal characteristics associated with detection of SGA, four sites were excluded from the analyses because they did not provide data on the co-

morbidities or antenatal complications studied (26.2% of pregnancies). Individual records were also excluded if they did not provide complete or imputed data on all the studied maternal characteristics (15.9% pregnancies) leaving 57.9% of SGA$_{both}$ pregnancies for analysis (Figure 6.1).

The rate of detection in the whole study population (estimated n=14,053 in imputed sample) was 24.1%. The change in this rate over time is illustrated in Figure 6.2, stratified by the intervention group. Whilst it appears on the graph that the rate of detection improved over time in the control arm of the trial, changed little in the GAP implementing clusters, and decreased in the non-implementing GAP sites, it is important to note that the statistical analysis of the DESiGN trial did not find a significant difference between sites randomised to standard care or GAP following adjustments.

Figure 6.2 - Rate of detection of SGA over time, for each intervention group in the DESiGN trial
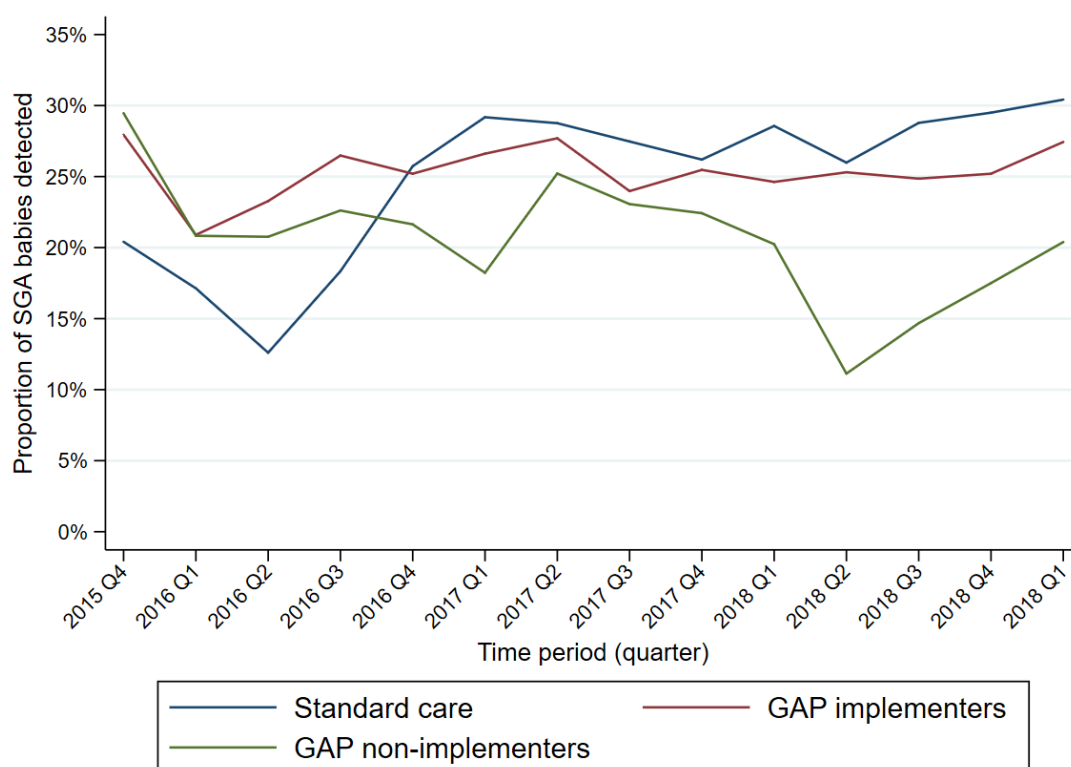
6.3.3   Characteristics of the study population

The characteristics of the included women are summarised in Table 6.3, these are compared statistically in section 6.3.4. A lower proportion of women with missed SGA were over the age of 40 years, living in the least deprived areas, had an underweight BMI (<18.5 kg/m^2) or were smokers. A higher proportion of women with missed SGA were white and a lower proportion were of Asian ethnicity. Maternal parity was similarly distributed across women with missed and detected SGA.

Table 6.3 - Characteristics of the included SGA$_{both}$ pregnancies, presented for all pregnancies, and stratified by detection status (imputed data)

			All SGA$_{both}$ babies	Missed SGA$_{both}$	Detected SGA$_{both}$
Age (years),		**Mean (SD)**	30.9 (5.6)	30.9 (5.6)	31.1 (5.8)
	%	**Age over 40y**	4.5%	4.2%	5.5%
IMD,	**%**	**1=least deprived**	9.5%	9.1%	10.6%
		2	11.7%	11.5%	12.3%
		3	24.0%	24.3%	23.0%
		4	35.0%	35.4%	33.5%
		5=most deprived	19.9%	19.7%	20.6%
Ethnicity,	**%**	**White**	43.6%	44.7%	39.9%
		Black	16.6%	16.4%	17.3%
		Asian	25.2%	24.0%	29.4%
		Mixed	2.0%	1.9%	2.3%
		Other	12.6%	13.1%	11.1%
BMI (kg/m²),		**Mean (SD)**	25.5 (5.5)	25.6 (5.4)	25.2 (5.7)
	%	**Under 18.5**	4.4%	3.8%	6.3%
		18.5-24.9	49.7%	49.5%	50.4%
		25.0-29.9	28.4%	29.0%	26.4%
		30.0-34.9	11.5%	11.7%	10.9%
		35.0-39.9	4.1%	4.1%	4.0%
		40.0 or above	1.9%	1.9%	2.1%
Parity,	**%**	**0**	54.8%	54.7%	55.0%
		1	28.1%	27.9%	28.6%
		2	10.1%	10.4%	9.3%
		3	4.1%	4.0%	4.4%
		4 or above	2.9%	3.0%	2.7%
Smoking, %		**Smoker**	10.4%	10.1%	11.2%
Co-morbidities, %		**Hypertension**	2.3%	2.1%	3.1%
		Diabetes	1.5%	1.3%	2.1%
		APLS	0.1%	0.05%	0.1%
Antenatal complications, %		**Pre-eclampsia**	4.4%	3.4%	7.8%
		PIH	2.6%	2.3%	4.0%
		GDM	5.3%	4.8%	7.0%
PAPP-A, %		**<0.300MoM**	2.4%	1.7%	4.9%
		0.3-0.415MoM	3.9%	3.1%	6.5%
		>0.415MoM	42.1%	41.0%	45.9%
		Missing data	51.6%	54.3%	42.6%
Indication for serial fetal scans,+ %		**Any indication**	34.8%	31.7%	44.3%
		No indication*	65.2%	68.3%	55.7%

*No recorded indication, complete case data except that information on PAPP-A may be missing.

Only 34.8% of women with an SGA$_{both}$ baby had any recorded indication for serial fetal growth ultrasound scans; the rate was higher amongst women with detected SGA$_{both}$ than missed SGA$_{both}$ (44.3% vs 31.7%). Given that the rate of recorded APLS was very low (0.1%) and there was a high proportion of missing data on this co-morbidity (29.6% included records), a post-hoc decision was made to exclude it from the composite 'any indication for serial growth scans' because it was unlikely to be informative and was a major contributor to missing data in this composite. Furthermore, 57.8% of women with no known scan

indication (including those with missing data) were nulliparous and therefore also could not have had a previous history of SGA or stillbirth; these are two other important indications for serial fetal ultrasounds that we did not have data on.

Characteristics of the SGA$_{both}$ babies are summarised in Table 6.4, categorised by whether SGA$_{both}$ was missed or detected. Overall, 12.3% of SGA$_{both}$ babies were born preterm. A higher proportion of babies with detected SGA$_{both}$ were born during each gestational age category before 39 weeks' gestation; following 39 weeks', a higher proportion of babies with missed SGA$_{both}$ were born (Figure 6.3). Of missed SGA$_{both}$ babies, 56.8% were born after their expected due date. Regardless of the centile chart used, a higher proportion of detected SGA$_{both}$ babies were born with birthweight below the 3rd centile than babies in whom the diagnosis was missed antenatally.

Table 6.4 - Characteristics of the included SGA$_{both}$ babies, presented by all babies, and stratified by detection status (imputed data)

		All SGA$_{both}$ babies	Missed SGA$_{both}$	Detected SGA$_{both}$
Neonatal presentation at birth, %	**Non-cephalic**	6.3%	5.5%	9.4%
Gestational age at birth (weeks),	**Mean (SD)**	39.1 (2.9)	39.6 (2.6)	37.4 (3.2)
%	**<28^{+0}**	1.3%	1.1%	2.2%
	28^{+0} – 33^{+6}	4.0%	2.4%	9.4%
	34^{+0} – 36^{+6}	7.0%	4.3%	16.2%
	37^{+0} – 37^{+6}	7.7%	5.2%	16.3%
	38^{+0} – 38^{+6}	12.6%	10.6%	19.3%
	39^{+0} – 39^{+6}	18.8%	19.6%	16.2%
	40^{+0} or above	48.6%	56.8%	20.3%
Birthweight customised centile,	**Mean (SD)**	3.7 (2.8)	4.0 (2.8)	2.5 (2.5)
%	**<3rd centile**	46.8%	41.3%	65.5%
	3rd – 5th centile	22.2%	23.6%	17.2%
	5th-10th centile	31.0%	35.0%	17.3%
Birthweight population centile,	**Mean (SD)**	4.5 (2.8)	4.8 (2.8)	3.7 (2.7)
%	**<3rd centile**	35.0%	31.4%	47.3%
	3rd – 5th centile	21.6%	21.5%	22.0%
	5th-10th centile	43.4%	47.1%	30.6%
Birthweight allocated centile,	**Mean (SD)**	4.4 (2.8)	4.7 (2.8)	3.5 (2.7)
%	**<3rd centile**	36.4%	32.4%	50.0%
	3rd – 5th centile	21.4%	21.5%	21.4%
	5th-10th centile	42.2%	46.2%	28.6%

Figure 6.3 - Proportion of SGA detected and SGA missed babies born during each gestational week
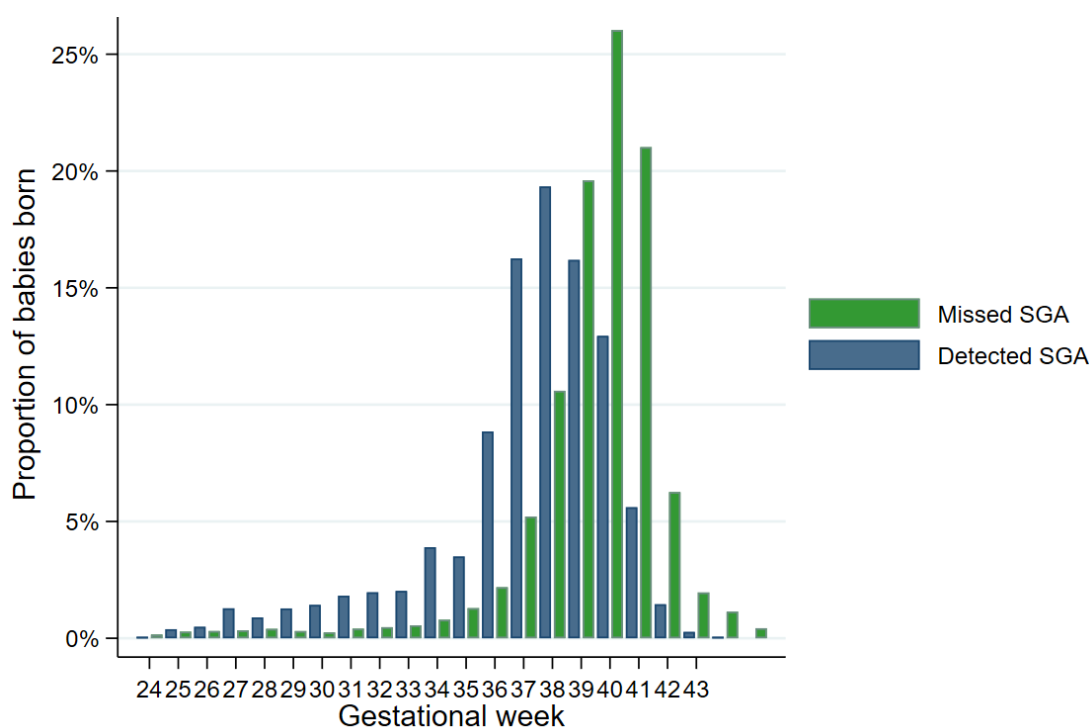
6.3.4 Comparing characteristics between cases and controls

Unadjusted and adjusted statistical comparisons between the odds of detected and missed SGA_{both} for each characteristic are reported in Table 6.5 and Table 6.6. SGA_{both} was less likely to be missed (more likely to be detected) amongst women with many of the indications for serial fetal ultrasound: smoking (aOR 0.79, CI: 0.66-0.96, p=0.02), pre-existing (aOR 0.55, CI: 0.36-0.84, p=0.01) or gestational diabetes (aOR 0.67, CI: 0.53-0.85, p=0.001), pregnancy-induced hypertension (aOR 0.61, CI: 0.44-0.83, p=0.002), pre-eclampsia (aOR 0.44, CI: 0.34-0.56, p<0.001) or low PAPP-A (aOR 0.44, CI: 0.31-0.63, p<0.001 for <0.3 MoM) when compared to absence of the indication. The same was true for women with any indication for serial fetal growth scans (composite aOR: 0.60, CI: 0.51-0.70, p<0.001).

Age over 40 years and pre-existing hypertension were associated with missed SGA_{both} in univariate comparisons but not in adjusted analyses, following the inclusion of hypertension and pre-eclampsia in the model respectively. Asian women were less likely to have missed SGA_{both}, even after adjusting for all other characteristics and co-morbidities. Women with BMI < 18.5 kg/m² were less likely to have missed SGA_{both} (aOR 0.58, CI: 0.44-0.77, p<0.001) and women of BMI 25.0-29.9 kg/m² were more likely to have missed SGA_{both} (aOR 1.15, CI: 1.01-1.32, p=0.04) when compared to women of healthy weight (BMI 18.5-24.9 kg/m²).

Table 6.5 - Unadjusted and adjusted odds ratios comparing demographic or clinical characteristics of women with missed SGA$_{both}$ to women in whom SGA$_{both}$ was antenatally detected.

		Missed SGA$_{both}$ (77.4%)	Detected SGA$_{both}$ (22.6%)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)	Adjusted p value
Age (years), %	**≤40y**	77.7%	22.3%	Ref	Ref	Ref
	>40y	72.5%	27.5%	0.76 (0.58-0.99)	0.77 (0.58-1.02)	0.07
IMD, %	**1=least deprived**	74.7%	25.3%	Ref	Ref	Ref
	2	76.3%	23.7%	1.08 (0.86-1.35)	0.96 (0.76-1.22)	0.76
	3	78.4%	21.6%	1.23 (1.01-1.5)	1.11 (0.89-1.38)	0.34
	4	78.4%	21.6%	1.21 (1.00-1.46)	1.06 (0.86-1.31)	0.59
	5=most deprived	76.7%	23.3%	1.08 (0.88-1.32)	1.02 (0.81-1.30)	0.86
Ethnicity, %	**White**	79.4%	20.6%	Ref	Ref	Ref
	Black	76.5%	23.5%	0.83 (0.71-0.97)	0.86 (0.71-1.02)	0.09
	Asian	73.7%	26.3%	0.73 (0.64-0.83)	0.70 (0.60-0.82)	<0.001
	Mixed	73.5%	26.5%	0.72 (0.48-1.07)	0.76 (0.50-1.15)	0.19
	Other	80.2%	19.8%	1.02 (0.85-1.23)	0.81 (0.66-1.00)	0.05
BMI (kg/m²), %	**<18.5**	67.3%	32.7%	0.57 (0.44-0.74)	0.58 (0.44-0.77)	<0.001
	18.5-24.9	77.2%	22.8%	Ref	Ref	Ref
	25.0-29.9	79.1%	20.9%	1.11 (0.98-1.27)	1.15 (1.01-1.32)	0.04
	30.0-34.9	78.6%	21.4%	1.07 (0.89-1.30)	1.16 (0.94-1.42)	0.16
	35.0-39.9	77.9%	22.1%	1.02 (0.77-1.36)	1.08 (0.80-1.46)	0.60
	≥40.0	75.6%	24.4%	0.90 (0.59-1.38)	1.04 (0.66-1.62)	0.87
Parity, %	**0**	77.3%	22.7%	Ref	Ref	
	1	77.1%	22.9%	1.01 (0.89-1.15)	0.93 (0.82-1.07)	0.31
	2	79.2%	20.8%	1.15 (0.95-1.40)	1.11 (0.90-1.36)	0.33
	3	75.7%	24.3%	0.93 (0.71-1.22)	0.89 (0.66-1.18)	0.41
	4 or above	79.3%	20.7%	1.12 (0.80-1.56)	1.07 (0.74-1.55)	0.71
Smoking, %	**Non-smoker**	77.7%	22.3%	Ref	Ref	
	Smoker	75.6%	24.4%	0.89 (0.75-1.06)	0.79 (0.66-0.96)	0.02

Babies with non-cephalic presentations at the time of birth were less likely to have missed SGA$_{both}$ (aOR 0.61, CI: 0.48-0.77, p<0.001). SGA$_{both}$ was more likely to be missed with higher birthweight centiles within the SGA$_{both}$ range, see Table 6.6.

Table 6.6 - Unadjusted and adjusted odds ratios or mean differences comparing co-morbidities or obstetric factors of women with missed SGA$_{both}$ to women in whom SGA$_{both}$ was antenatally detected.

		Missed SGA$_{both}$ (77.4%)	Detected SGA$_{both}$ (22.6%)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/ mean diff (95% CI)	Adjusted p value
Co-morbidities, %	**No hypertension**	77.6%	22.4%	Ref	Ref	
	Hypertension	70.0%	30.0%	0.69 (0.50-0.95)	0.84 (0.59-1.19)	0.33
	No diabetes	77.6%	22.4%	Ref	Ref	
	Diabetes	67.6%	32.4%	0.56 (0.37-0.84)	0.55 (0.36-0.84)	0.01
Antenatal complications, %	**No pre-eclampsia**	78.3%	21.7%	Ref	Ref	
	Pre-eclampsia	59.5%	40.5%	0.39 (0.31-0.49)	0.44 (0.34-0.56)	<0.001
	No PIH	77.8%	22.2%	Ref	Ref	
	PIH	65.9%	34.1%	0.53 (0.40-0.71)	0.61 (0.44-0.83)	0.002
	No GDM	77.9%	22.1%	Ref	Ref	
	GDM	70.1%	29.9%	0.66 (0.53-0.82)	0.67 (0.53-0.85)	0.001
PAPP-A, %	**<0.3 MoM**	53.2%	46.8%	0.39 (0.28-0.54)	0.44 (0.31-0.63)	<0.001
	0.3-0.415 MoM	61.0%	39.0%	0.53 (0.40-0.70)	0.58 (0.43-0.78)	<0.001
	>0.415MoM	74.8%	25.2%	Ref	Ref	
	Missing data	80.9%	19.1%	1.28 (1.13-1.45)	1.35 (1.15-1.60)	<0.001
Any indication for serial growth scans[+], %	**No indication**	78.7%	21.3%	Ref	Ref	Ref
	Any indication	68.3%	31.7%	0.58 (0.50-0.67)	0.59 (0.51-0.67)	<0.001
Fetal presentation at birth, %	**Cephalic**	78.8%	21.2%	Ref	Ref	Ref
	Non-cephalic	67.3%	32.7%	0.52 (0.42-0.65)	0.61 (0.48-0.77)	<0.001
Allocated centile at birth, mean (SD)		4.7 (2.8)	3.5 (2.7)	1.27* (1.24-1.30)	1.19* (1.17-1.22)	<0.001

*Change in OR with a one centile increase (<10th centile).

+Adjusted only for IMD, parity, ethnicity, and allocated birthweight centile (not for other adjustment characteristics which are included in this composite).

6.3.5   Comparing measures of ultrasound utilisation between cases and controls

Unstratified measures of ultrasound utilisation were studied amongst the entire sample of SGA$_{both}$ pregnancies (n≈14,053 across imputed datasets). Patterns were also stratified by the presence or absence of an indication for serial fetal ultrasound scans; this required a complete case analysis and therefore included 47.7% of all SGA$_{both}$ babies in the

imputed datasets. Ultrasound patterns are described in Table 6.7 for the whole SGA$_{both}$ sample and for the stratified sample.

Almost half of the pregnancies with missed SGA$_{both}$ (46.2%) had no record of a fetal growth scan conducted at the site at which they gave birth. Over half (55.7%) of women who had SGA$_{both}$ diagnosed antenatally required only one screening scan, meaning that the EFW was below the 10th centile at the time of the first scan. Regardless of the presence of indications for serial scans, a lower proportion of women with missed SGA$_{both}$ received 3-weekly or 4-weekly scans than women with detected SGA$_{both}$. Screening scans were generally commenced slightly later for women with missed SGA$_{both}$ compared to those with detected SGA$_{both}$, with a lower proportion commencing scans before 31 weeks' (48.9% vs 58.4%). The patterns for women with a documented indication for serial scans were similar, although a lower proportion of women with a scan indication were offered no scans, and conversely, the proportion with no scans was higher for women with no documented scan indication. More women with a scan indication commenced their scans before 31 weeks (71.5% if SGA was detected, 61.5% if SGA was missed).

Table 6.7 - Patterns of ultrasound utilisation for all SGAboth pregnancies, and stratified by presence or absence of a recorded indication for serial fetal growth scans

		All SGA$_{both}$		SGA$_{both}$ with serial scan indication		SGA$_{both}$ with no recorded serial scan indication*	
		Missed SGA$_{both}$ (75.9%)	Detected SGA$_{both}$ (24.1%)	Missed SGA$_{both}$ (70.3%)	Detected SGA$_{both}$ (29.7%)	Missed SGA$_{both}$ (80.4%)	Detected SGA$_{both}$ (19.6%)
Number of screening scans received, %	0	46.2%	-	37.1%	-	54.5%	-
	1	20.8%	55.7%	16.6%	54.2%	20.5%	58.2%
	2	14.7%	26.1%	19.6%	25.0%	11.8%	25.6%
	3	11.6%	12.5%	18.1%	14.7%	8.5%	10.7%
	4	4.6%	4.5%	5.6%	4.6%	3.3%	4.7%
	≥5	2.2%	1.1%	2.9%	1.5%	1.3%	0.7%
Screening scan frequency for pregnancies with at least two scans:	≤3-weekly	16.5%	44.8%	14.6%	46.4%	17.7%	43.9%
	4-weekly	14.1%	29.6%	12.6%	26.8%	13.3%	27.3%
	>4-weekly	69.4%	25.5%	72.8%	26.7%	69.0%	28.9%
Gestation at the time of the first scan, if scans conducted, %	<31^{+0}	48.9%	58.4%	61.5%	71.5%	44.7%	48.2%
	31^{+0}-33^{+6}	15.1%	13.6%	14.8%	12.0%	14.9%	16.4%
	34^{+0}-36^{+6}	25.6%	18.4%	17.7%	11.8%	25.1%	22.1%
	≥37^{+0}	10.3%	9.6%	6.1%	4.8%	15.3%	13.3%

Includes records for which PAPP-A was not documented.

For pregnancies in which screening for SGA remained relevant (pregnancy ongoing and SGA had not yet been detected), the proportion of women receiving any ultrasound scan during each gestational week, starting from 26 weeks', is presented in Figure 6.4. Screening

ultrasound scans remained applicable to over 90% of women with missed SGA until 37 weeks', after which the proportion of women for whom it remained applicable decreased as the babies had been born. Amongst women in whom SGA was detected antenatally the proportion for whom screening scans remained relevant decreased in a linear fashion throughout the pregnancy. Amongst pregnancies in which SGA was missed, screening scans were less common at all gestations than amongst women with detected SGA. Despite screening scans remaining relevant to a larger proportion of pregnancies at term amongst women with missed SGA than detected SGA, less than 10% of remaining women received a scan during any week of gestation at term.

Figure 6.4 - Proportion of women receiving a screening ultrasound for fetal growth, amongst the proportion in whom screening for SGA remains relevant, presented by SGA$_{both}$ detection status

272
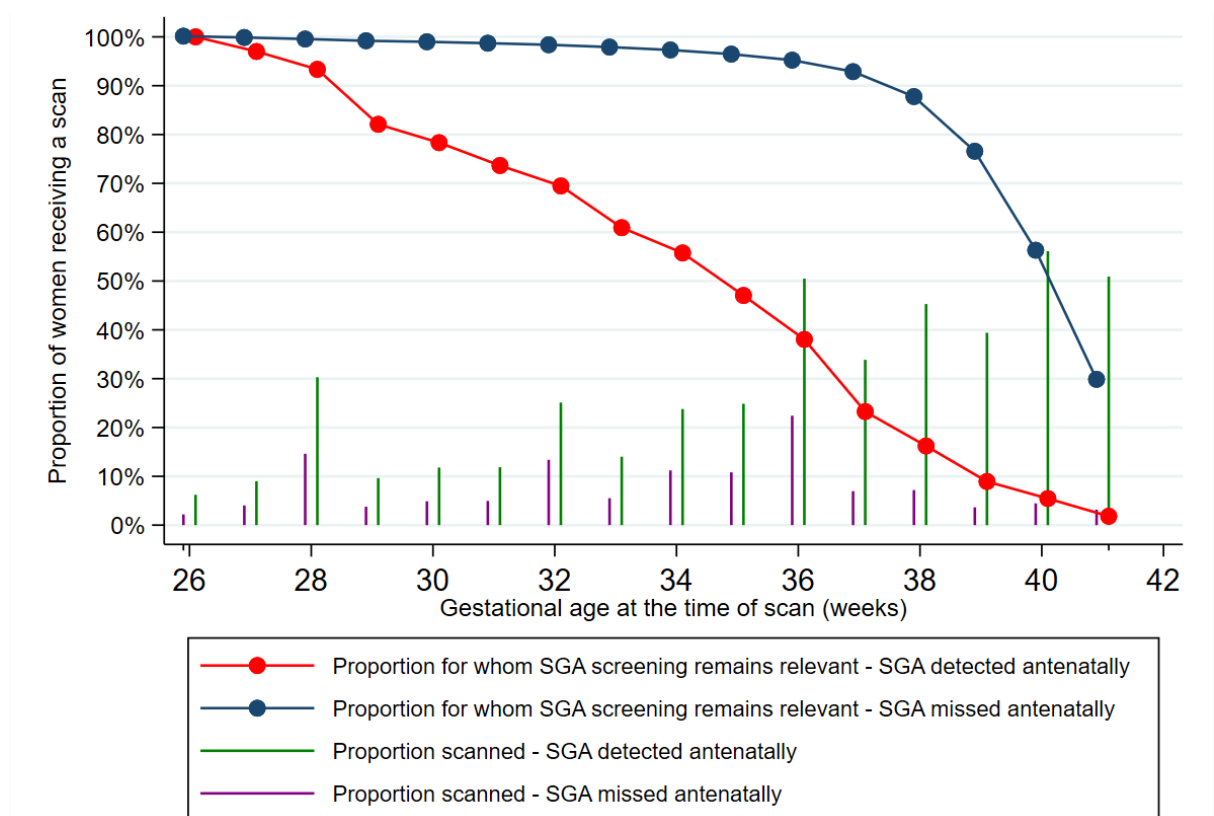
For women in whom SGA_{both} was detected, the gestation at which SGA was first detected was distributed throughout the gestational period (Figure 6.5).

Figure 6.5 - Bar chart showing the gestational age at the time of the first scan at which the EFW was below the 10th centile for pregnancies in which SGA_{both} was antenatally detected.
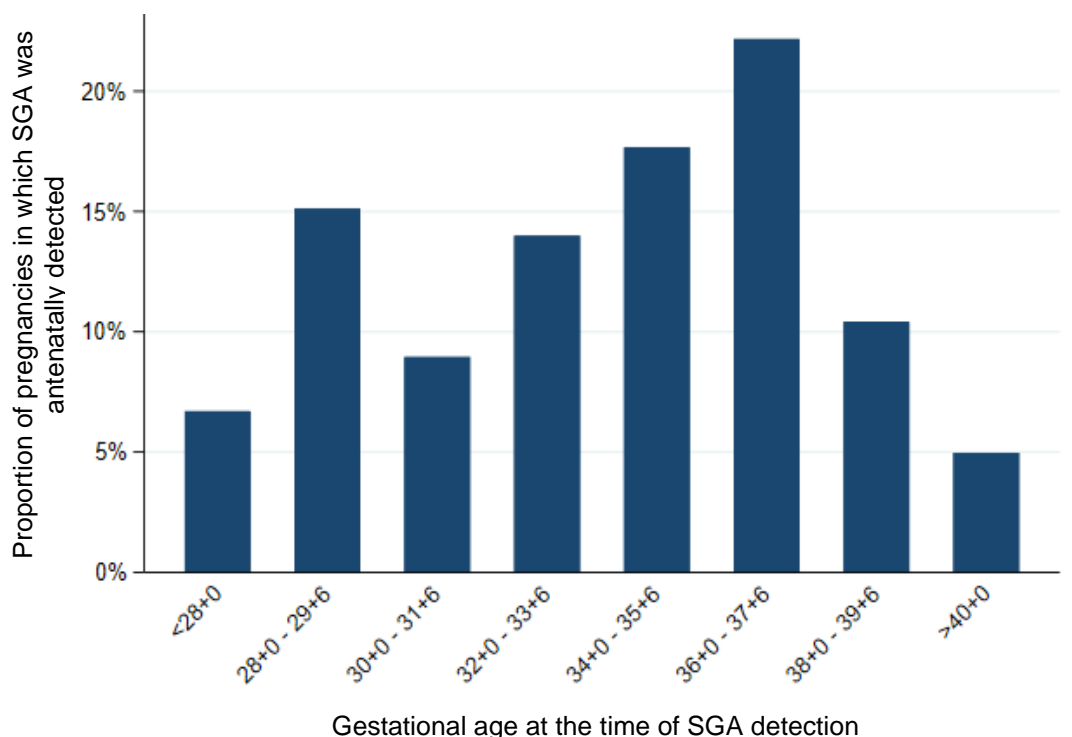
The frequency and timing of screening scans conducted for each pregnancy are compared in Table 6.8. Of the 20.2% of women with missed SGA_{both} who had at least two scans during the pregnancy, the mean frequency of screening scans was lower than for women with detected SGA and at least two screening scans (6.5-weekly vs 3.9-weekly, adjusted difference 2.6 weeks, CI: 2.4-2.8, p=0.001). SGA was less likely to be missed in women who had 3-weekly scans (aOR 0.75, CI: 0.63-0.89, p=0.001) and more likely to be missed in pregnancies who had scans conducted less often than 4-weekly (aOR 5.89, CI: 4.96-7.01, p<0.001) when compared to women with 4-weekly scans. Pregnancies in which SGA_{both} was missed had a slightly later onset of screening scans than those in which it was detected (31.7 weeks' vs 30.8 weeks', adjusted difference 0.9 weeks, CI: 0.7-1.1, p<0.001). Very similar findings were seen for women with or without documented indications for serial fetal growth scans (Table 6.9).

Women with missed SGA also had an adjusted mean of 17.6 additional days between their last scan and their birth compared to women with detected SGA (27.4 days vs. 9.8 days, adjusted difference 17.6 days, CI: 16.9-18.4, p<0.001); this is partly because many of the women with detected SGA_{both} were receiving surveillance of diagnosed SGA. The duration between the last scan and the birth increased with increasing gestational age at birth;

pregnancies in which SGA was missed had the last scan conducted 30.3 days before the birth if it occurred after 39 weeks' gestation, compared to 15.5 days for pregnancies with detected SGA (adjusted difference 14.8 days, CI: 13.4-16.0, p<0.001). The periods for each gestational age category are detailed in Table 6.8. This finding was also very similar for women when stratified by the presence of a recorded indication for serial fetal growth scans (Table 6.9).

Table 6.8- Patterns of ultrasound utilisation for pregnant women and their SGA$_{both}$ babies, by detection status of SGA$_{both}$

		Missed SGA$_{both}$ (75.9%)	Detected SGA$_{both}$ (24.1%)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/mean diff (95% CI)	p value
Frequency of screening scans - one scan every *n* weeks, mean *n* (SD)		6.5 (1.5)	3.9 (1.1)	2.6 (2.4-2.8)	2.6 (2.4-2.8)	<0.001
Scan frequency for pregnancies with at least two scans:	**≤3-weekly**	46.4%	53.6%	0.77 (0.65-0.92)	0.75 (0.63-0.89)	0.001
	4-weekly	52.9%	47.1%	Ref	Ref	Ref
	≥5-weekly	86.5%	13.5%	5.70 (4.80-6.76)	5.89 (4.96-7.01)	<0.001
If screening scan performed, gestation at the time of the first in weeks, mean (SD)		31.7 (4.2)	30.8 (4.3)	0.9 (0.7-1.1)	0.9 (0.7-1.1)	<0.001
Duration from the last scan until birth in days	**Mean (SD)**	27.4 (20.9)	9.8 (12.1)	17.7 (17.0-18.5)	17.7 (16.9-18.4)	<0.001
Duration by gestational age at birth, mean (SD):	**<28^{+0}**	0.9 (24.1)	3.7 (3.7)	-2.5 (-0.1 – 5.0)	-2.5 (-0.1-5.0)	0.06
	28^{+0} – 30^{+6}	12.3 (13.4)	4.6 (5.3)	11.4 (7.3 – 15.4)	11.4 (7.3-15.4)	<0.001
	31^{+0} – 33^{+6}	15.3 (15.2)	5.3 (8.5)	11.1 (7.8 – 14.4)	11.1 (7.8-14.4)	<0.001
	34^{+0} – 36^{+6}	14.6 (15.2)	5.8 (7.2)	8.9 (7.3 – 10.5)	8.9 (7.3-10.5)	<0.001
	37^{+0} – 38^{+6}	18.2 (16.1)	7.9 (7.8)	10.3 (9.3 – 11.3)	10.3 (9.3-11.3)	<0.001
	≥39^{+0}	30.3 (21.2)	15.5 (16.3)	14.9 (13.6 – 16.2)	14.9 (13.6-16.2)	<0.001

Table 6.9 - Patterns of ultrasound utilisation for pregnant women and their SGA_both babies, by the presence of a recorded indication for serial fetal growth scans and detection status of SGA

	Pregnancies with any indication for serial growth scans					Pregnancies with no known indication for serial growth scans*				
	Missed SGA_both (71.0%)	Detected SGA_both (29.0%)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/mean diff (95% CI)	Adjusted p value	Missed SGA_both (80.4%)	Detected SGA_both (19.6%)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/mean diff (95% CI)	Adjusted p value
Frequency of screening scans – one scan every n weeks, mean n (SD)*	6.7 (3.4)	3.7 (1.9)	2.8 (2.3-3.3)	2.8 (2.3-3.3)	<0.001	6.7 (1.7)	3.9 (1.3)	2.8 (2.4-3.3)	2.9 (2.5-3.3)	<0.001
Scan frequency for pregnancies with at least two scans: ≤3-weekly	44.0%	56.0%	0.72 (0.50-1.03)	0.71 (0.49-1.03)	0.07	49.6%	50.4%	0.80 (0.56-1.15)	0.76 (0.52-1.11)	0.16
4-weekly	51.1%	48.9%	Ref	Ref	Ref	54.5%	45.5%	Ref	Ref	Ref
≥5-weekly	87.6%	12.4%	6.62 (4.63-9.45)	7.03 (4.86-10.16)	<0.001	85.4%	14.6%	4.93 (3.47-7.00)	5.39 (3.76-7.75)	<0.001
If scan performed, gestation at the time of the first in weeks, mean (SD)	30.6 (3.9)	29.6 (3.9)	1.1 (0.7-1.4)	1.1 (0.8-1.5)	<0.001	32.5 (4.5)	31.8 (4.4)	0.7 (0.2-1.2)	0.7 (0.2-1.2)	<0.001
Duration from the last scan until birth in days Mean (SD)	26.2 (19.2)	9.0 (10.0)	17.4 (15.9-18.8)	16.9 (15.5-18.4)	<0.001	28.7 (23.4)	10.9 (12.6)	17.9 (15.7-20.3)	17.7 (15.5-20.0)	<0.001
Duration by gestational age at birth: <28+0	4.3 (4.1)	4.1 (4.1)	0.6 (-4.4-5.64)	-2.9* (-11.4-5.5)	0.48	-39.5* (56.7)	5.3 (4.1)	-2.3* (-14.7-10.1)	-7.0* (-38.0-24.0)	0.43
28+0 – 30+6	11.9 (9.8)	5.3 (5.9)	#	#	#	16.8 (33.1)	4.4 (4.9)	32.6 (17.9-47.2)	34.3 (13.7-54.9)	0.01
31+0 – 33+6	17.8 (16.5)	5.6 (5.8)	14.3 (9.5-19.9)	12.7 (7.6-17.9)	<0.001	23.5 (24.8)	6.7 (4.1)	24.2 (18.0-30.4)	25.1 (20.2-30.1)	<0.001
34+0 – 36+6	15.1 (16.5)	5.0 (6.1)	10.5 (7.5-13.5)	10.4 (7.3-13.4)	<0.001	20.7 (19.5)	5.4 (6.7)	15.2 (8.6-21.7)	14.7 (7.9-21.4)	<0.001
37+0 – 38+6	17.1 (14.5)	8.4 (7.0)	8.8 (6.8-10.7)	8.5 (6.5-10.5)	<0.001	17.3 (15.1)	7.4 (7.3)	9.9 (6.9-12.8)	10.0 (7.0-13.0)	<0.001
≥39+0	30.5 (19.3)	16.8 (14.2)	13.6 (10.6-16.5)	13.6 (10.7-16.4)	<0.001	30.8 (23.6)	14.8 (15.1)	15.9 (12.7-19.2)	16.0 (12.8-19.1)	<0.001

*Includes records in which PAPP-A was either ≥0.415 MoM, or missing.; # - Too few observations in either group, unable to generate mean difference.

* Negative values suggest that fetal scan occurred after birth, these are not valid and caused by imputation estimates of gestational age at birth in a small sample.

Of all SGA$_{both}$ babies, 87.7% were born at term (Table 6.4). The results of an analysis limited to these babies describing the estimated fetal weights and their centiles at the last ultrasound scan before birth, compared to the birthweights and their centiles, are reported in Table 6.10. SGA$_{both}$ babies born within a week of the last growth scan (41.8% of detected and 13.1% of missed SGA pregnancies) had a mean EFW centile of 4.6 (SD: 2.9) for detected babies and 24.9 (SD: 13.7) for babies in whom SGA was missed antenatally, this equated to an EFW:birthweight centile difference of +0.7 (SD: 3.3) for detected SGA babies and +19.5 (SD: 13.6) for missed SGA babies (adjusted difference 18.6, CI: 17.5-19.7, p<0.001) and an EFW:birthweight difference in grams expressed as a percentage of the birthweight of +3.1% (SD: 10.7%) for detected SGA and +14.0% (SD: 7.5%) for missed SGA (adjusted difference 10.6%, CI: 9.7-11.5%, p<0.001). As the duration between the last growth ultrasound and the birth increased, the centile difference for SGA detected babies remained similar, although the difference between the EFW at the time of scan and the birthweight a few weeks later increased. For pregnancies in whom SGA was missed antenatally, a different relationship was seen. For these pregnancies, as the duration between the last scan and the birthweight increased, the difference between centiles increased, but the percentage difference between EFW and birthweight decreased so that EFW measurements taken 4 weeks before birth were closer to the actual birthweight than EFW measurements taken within one week of birth (-1.4%, SD: 28.4% difference for scans 4 weeks before birth, 14.0%, SD: 7.5% for scans within 1 week of birth).

Table 6.10 - Comparison of estimated fetal weight at the last ultrasound scan and the birthweight, including their centiles, for SGA$_{both}$ babies born at term

	All SGA$_{both}$	Missed SGA$_{both}$ (75.9%)	Detected SGA$_{both}$ (24.1%)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/mean diff (95% CI)	p value
If scan within 1 week*:						
EFW centile at last scan, mean (SD)	14.9 (14.3)	24.9 (13.7)	4.6 (2.9)	20.4 (19.3-21.5)	20.1 (19.0-21.3)	<0.001
EFW <10th centile,%	49.0%	-	100.0%			
EFW 10-20th centile,%	23.7%	46.6%	-	N/A	N/A	N/A
EFW >20th centile,%	27.2%	53.4%	-			
Difference between EFW and birthweight centiles, mean (SD)	10.3 (13.7)	+19.5 (13.6)	+0.7 (3.3)	18.8 (17.7 – 19.9)	18.6 (17.5-19.7)	<0.001
Percentage difference between EFW and birthweight, mean (SD)	8.7% (10.7%)	+14.0% (7.5%)	+3.1% (10.7%)	10.9% (9.9-12.0%)	10.6% (9.7-11.5%)	<0.001
If scan within 2 weeks*:						
EFW centile at last scan, mean (SD)	18.8 (15.2)	26.1 (14.1)	5.2 (2.8)	20.9 (19.3-22.6)	20.8 (19.1-22.5)	<0.001
EFW <10th centile,%	35.0%	-	100.0%			
EFW 10-20th centile,%	27.7%	42.7%	-	N/A	N/A	N/A
EFW >20th centile,%	37.2%	57.3%	-			
Difference between EFW and birthweight centiles, mean (SD)	14.0 (14.8)	+20.9 (14.0)	+1.12 (3.2)	19.6 (18.0 – 21.3)	19.5 (17.9-21.2)	<0.001
Percentage difference between EFW and birthweight, mean (SD)	6.6% (33.1%)	+11.2% (39.9%)	-1.8% (8.3%)	12.8% (8.2-17.4%)	12.9% (8.2-17.6%)	<0.001
If scan within 3 weeks*:						
EFW centile at last scan, mean (SD)	22.1 (15.6)	27.2 (14.3)	5.3 (2.7)	22.0 (20.1-23.9)	21.7 (19.8-23.7)	<0.001
EFW <10th centile,%	23.5%	-	100.0%			
EFW 10-20th centile,%	29.0%	38.0%	-	N/A	N/A	N/A
EFW >20th centile,%	47.4%	62.0%	-			
Difference between EFW and birthweight centiles, mean (SD)	5.0 (2.8)	21.7 (14.1)	+1.8 (3.4)	20.1 (18.2-21.9)	19.8 (17.9-21.7)	<0.001
Percentage difference between EFW and birthweight, mean (SD)	1.4% (26.1%)	+3.9% (28.7%)	-6.7% (11.3%)	10.7% (6.8-14.5%)	9.8% (5.9-13.7%)	<0.001
If scan within 4 weeks*:						
EFW centile at last scan, mean (SD)	26.5 (16.4)	30.1 (15.1)	5.7 (3.2)	24.3 (21.8-26.9)	24.4 (21.7-27.0)	<0.001
EFW <10th centile,%	14.7%	-	100.0%			
EFW 10-20th centile,%	25.5%	29.9%	-	N/A	N/A	N/A
EFW >20th centile,%	59.8%	70.1%	-			
Difference between EFW and birthweight centiles, mean (SD)	21.5 (16.1)	+24.9 (14.9)	+2.0 (4.1)	22.8 (20.3-25.3)	22.7 (20.1-25.3)	<0.001
Percentage difference between EFW and birthweight, mean (SD)	-2.9% (28.5%)	-1.4% (28.4%)	-11.8% (27.2%)	10.3% (5.2-15.4%)	8.5% (3.2-13.7%)	0.002

6.3.6 Sensitivity analyses

6.3.6.1 *Available case sensitivity analysis*

The composition of the sample and results tables for the available case sensitivity analysis are presented in Appendix section 10.16 (Figure 10.1 and Table 10.1 - Table 10.5). The characteristics for the included women and babies with SGA$_{both}$ pregnancies were broadly similar, except that there was a lower proportion of nulliparous women in the available case analysis compared to the imputed analysis (50.9% vs 54.8%). This is the opposite of the finding in the entire DESiGN dataset, where the proportion of nulliparous women decreased after imputation of missing data (section 2.2.4.2).

In univariate and multivariate comparisons of factors associated with detection status, the findings were also broadly similar. There were three differences between the analyses using the available case or imputed data: pre-existing hypertension was associated with a lower odds of having missed SGA$_{both}$ (aOR 0.64, CI: 0.40-0.999, p=0.049). Neither pre-existing diabetes (aOR 1.01, CI: 0.56-1.83, p=0.96) nor pregnancy-induced hypertension (aOR 0.75, CI: 0.48-1.18, p=0.22) were found to be associated with SGA$_{both}$ detection.

Only the stratified analysis is presented for comparison of ultrasound utilisation using the available case data, the findings of this were very similar to that using imputed data. The results of the non-stratified analysis are expected to be very similar to those using imputed data because in both cases, the whole SGA$_{both}$ population was used.

6.3.6.2 *Repeating the analysis for pregnancies in which the baby was SGA$_{pop}$*

The construction of the sample for this sensitivity analysis and the results tables are reported in Appendix section 10.16 (Table 10.6 - Table 10.8). The rate of detection of SGA was higher amongst the SGA$_{pop}$ sample than the SGA$_{both}$ sample (33.7% vs. 22.6%). The babies in the SGA$_{pop}$ sample (detected SGA population centile 4.0 vs missed SGA 5.7, adjusted difference 1.2, CI: 1.2-1.3, p<0.001) were overall slightly larger than those in the SGA$_{both}$ sample. As for the main analysis, Asian ethnicity, BMI<18.5 kg/m^2, pre-eclampsia, pregnancy-induced hypertension, any indication for serial fetal growth scans and non-cephalic fetal presentation at birth were associated with a lower chance of missed SGA$_{pop}$. However, the following factors were no longer found to be associated with the rate of detection of SGA$_{pop}$: BMI 25.0-29.9 kg/m^2 (aOR 1.12, CI: 0.97-1.30, p=0.12), smoking (aOR 0.84, CI:0.69-1.03, p=0.09), and pre-existing (aOR 0.77, CI: 0.51-1.17, p=0.22) or gestational diabetes (aOR 0.99, CI: 0.79-1.23, p=0.92).

6.3.6.3 *Restricting the sample to pregnancies with a record of an anomaly scan at the site of birth*

This analysis was conducted in the same samples as the main SGA$_{both}$ analysis (Figure 6.1), but further restricted to pregnancies with a record of a fetal anomaly scan at the site at which the woman later gave birth. Report tables are in Appendix section 10.16 (Table 10.9 - Table 10.12). Of women with missed SGA$_{both}$, 75.5% had received a presumed anomaly scan at the cluster site in which they later gave birth. This was lower than the 90.4% of women with detected SGA$_{both}$ who had received an anomaly scan.

Compared to women with an anomaly scan, women without a recorded anomaly scan had higher rates of Black ethnicity (18.1% vs 16.2%), BMI in all categories above the normal range (18.5-24.9 kg/m^2), nulliparity (57.1% vs 54.1%), smoking (11.6% vs. 10.0%), pre-existing hypertension (3.3% vs 2.1%) and pre-eclampsia (5.6% vs. 4.0%). They had a lower rate of Asian ethnicity (22.8% vs 25.9%) and GDM (3.6% vs 5.7%). Of those without a recorded fetal anomaly scan, 94.6% had missing PAPP-A, compared to 41.0% of those with a scan recorded. The babies born from these pregnancies also had higher rates of severe SGA (birthweight <3rd allocated centile: 42.0% vs 34.8%).

The rate of detection of SGA in this restricted sample was slightly higher than that in the main analysis – 27.6% of all pregnancies had an EFW<10th centile at the last fetal growth scan (or 25.9% of the sample when restricted to those with complete demographic and co-morbidity data). Compared to the whole SGA$_{both}$ sample, the sample restricted to pregnancies with an anomaly scan showed very similar findings except Black ethnicity (aOR 0.81, CI: 0.66-0.98, p=0.03) and pre-existing hypertension (aOR 0.62, CI: 0.42-0.91, p=0.02) were now also associated with missed SGA. The association with missing PAPP-A was lost in this sensitivity analysis (aOR 0.88, CI: 0.72-1.07, p=0.19).

With regards to the patterns of ultrasound utilisation, restricted to women with a record of a fetal anomaly scan, a lower proportion of women with missed SGA$_{both}$ did not receive any scan in pregnancy (34.8% of all women, 22.6% of women with an indication for serial scans). The gestation at which scans were commenced and the gestation at which they were stopped (measured by the mean duration between the last scan and the birth) were broadly like those found in the main analysis. The frequency of scans for both pregnancies with missed and detected SGA$_{both}$, however, was higher. Of women with detected SGA and at least two screening scans, the mean frequency was 3-weekly or more often for 88.5% of women. For women with missed SGA, 41.1% of women with at least two scans had them conducted this frequently.

Discussion

6.3.7   Summary of the key findings

Overall SGA$_{both}$ was missed antenatally in 75.9% of pregnancies. As expected, SGA was less likely to be missed amongst women with indications for serial fetal ultrasound in pregnancy, but 55.7% of all pregnancies in which SGA was detected antenatally and 68.3% of pregnancies in which SGA was missed antenatally had no recorded indication for serial fetal ultrasound in this dataset. Whilst we did not have data on some indications, 57.8% of all pregnancies without a recorded indication for serial ultrasound were nulliparous and therefore also did not have an obstetric history of stillbirth or previous birth of an SGA baby. The SGA detection rate was also higher for babies with a non-cephalic presentation at birth.

Only two factors were identified that increased the odds of missing SGA: having a BMI in the range 25.0-29.9 kg/m^2 and having a birthweight centile that was higher within the SGA range (i.e., less severe SGA). Having a missing record of PAPP-A was associated with greater odds of missed SGA in the primary analysis, but this association was lost in the sensitivity analysis restricted only to women with a record of a fetal anomaly scan at the same site as the birth. This may have been related to the power to detect a difference in a smaller sample or related either to transfers of care (in which case the PAPP-A is likely to have been available in the hard copy of the maternity records), or the assumption that women with no scan record had received no scans in pregnancy.

I identified that, for pregnancies in which SGA was detected, the gestation at the time of diagnosis is distributed throughout the pregnancy. Of the babies in whom an antenatal diagnosis of SGA$_{both}$ was made, 27.8% were born preterm. Of babies in whom SGA was missed antenatally, 8.8% were born preterm and 56.8% were born after their estimated due date.

With regards to findings from the study of ultrasound utilisation, almost half of the pregnancies with missed SGA$_{both}$ had no record of a fetal growth scan conducted at the site at which they gave birth (one third in a sensitivity analysis restricted to women with a record of a presumed anomaly scan) and only one third received two scans or more. For women with a recorded indication for serial scans, over a third receive no fetal growth scans (one in five women in anomaly scan sensitivity analysis) and under half received two or more scans. In both the main and the sensitivity analysis, having more frequent screening scans was associated with higher odds of having SGA detected.

Finally, a study of the timing and findings of the last scans conducted before birth revealed interesting findings, particularly for SGA$_{both}$ babies born at term. Whilst scans

continued up until birth for pregnancies in which SGA was detected (as would be expected following a diagnosis of SGA), the duration between the last scan widened for pregnancies in which SGA_{both} was missed as the gestational age at birth increased, demonstrating the practice documented in many clinical guidelines from the included clusters (Table 3.4) that ultrasound scans were stopped at 36 weeks' gestation. Of pregnancies in which SGA_{both} was missed, 76.4% were born at 39^{+0} weeks' gestation or later. The mean duration between the last scan (if any scan had been conducted) and the birth of the baby from these pregnancies was 30 days. Furthermore, when a scan was conducted within a week of the birth for a baby born at term, the EFW from scans conducted on all SGA_{both} babies was on average, 10.3 centiles (or 8.7% of birthweight) overestimated, and for pregnancies in which SGA_{both} was missed, the EFW centile was overestimated by 19.5 centiles (or 14.0% of birthweight). The EFW calculated for missed SGA_{both} babies born at term was actually found to be closer to the birthweight if the scan was conducted four weeks before the birth, than when it was conducted one, two or three weeks before the birth.

6.3.8 Interpretation of the findings

The rate of detection of SGA in this observational study conducted using data from over 200,000 births from thirteen maternity units in the UK over three years falls within the range of detection rates previously published by other teams in international settings.[4,210-216,282,283] The rate is likely to be slightly underestimated, given that our data collection methods were unable to account for ultrasound scans that were conducted at other sites. However, this will have only affected a small proportion of women for whom antenatal care is transferred late in the pregnancy after a diagnosis of SGA elsewhere, and for whom a further scan is not conducted at the maternity unit in which they give birth. The rate of detection only increased by 3% when restricted to women who had received their fetal anomaly screening scan at the same site as they gave birth.

As was expected, the presence of a known risk factor for SGA increased the rate of its detection; this has previously been shown in other studies.[184] SGA was also more likely to be detected by ultrasound amongst pregnancies in which the baby is born with a non-cephalic presentation. This may be related to the NICE recommended practice to offer a late third-trimester fetal ultrasound scan, at which fetal growth is also commonly assessed, to women with suspected non-cephalic presentations.[8] A recent UK Health Technology Assessment report recommended universal late pregnancy ultrasound screening for fetal presentation but without evidence for universal fetal growth assessment by ultrasound.[230] If this policy is applied nationally, it is pertinent that the risk of reducing the rate of SGA detection is considered by removing the possibility of incidental diagnoses previously made

at the time of ultrasound scans for other indications (e.g. placenta localisation or fetal presentation).

SGA was less likely to be detected amongst women with BMI in the overweight range (BMI 25.0-29.9 kg/m^2). This finding was only identified amongst SGA$_{both}$ pregnancies, but not amongst SGA$_{pop}$ pregnancies, suggesting that customisation of growth status either has a role by changing the threshold of detection for SGA using fundal height measurements or the EFW. Avci et al (2015) have previously shown that perinatal mortality was correlated with SGA for women with BMI above 25 kg/m^2 when using customised, but not population standards (i.e. customised standards were more likely to classify pregnancies with perinatal mortality as SGA).[412] Gardosi has previously defended his methods for customising fundal height measurements by demonstrating that maternal weight was the second most influential factor on fundal height after gestational age.[413] Preyer et al (2019) identified that ultrasound estimation of fetal weight was more accurate than estimation via abdominal palpation for pregnant women with a BMI above 25 kg/m^2, although equally accurate for women with BMI below 25 kg/m^2.[414] It has also been previously demonstrated that accuracy of fetal weight estimation by ultrasound is lower for women with higher BMI.[415,416] However, studies on the sensitivity of customised fundal height measurements for women according to BMI category have not been conducted and it remains unclear why this finding applied to women with a BMI of 25.0-29.9 kg/m^2, but not to women with a BMI of 30-34.9 kg/m^2, a group who are also not routinely offered serial fetal growth scans.

Whilst it was expected that the presence of an indication for serial fetal growth assessment by ultrasound was associated with higher rates of SGA detection, it is less well established that a large proportion of women who give birth to an SGA baby have no known risk factors for SGA. Two-thirds of women giving birth to an SGA baby (SGA$_{both}$ or SGA$_{pop}$) had no indication for serial fetal growth assessment by ultrasound, and almost half of pregnancies in which SGA was missed did not receive any growth scan during the pregnancy. Whilst this finding should be interpreted cautiously because data were missing on some indications for serial fetal growth ultrasound in pregnancy (see section 6.3.9 below), most of these indications are uncommon or rare and so, had they been available, are unlikely to have made a major change to the finding. Conversely, approximately a quarter of women with a clear indication for serial fetal ultrasound assessment, had no documented scans in pregnancy. This may in part be related to local policies which were not entirely adherent to national guidelines recommending serial scans for all indications (Table 3.4). Whilst SGA cannot be diagnosed antenatally without a fetal growth scan, policies to offer fetal growth ultrasound assessment universally have demonstrated relatively low sensitivity and specificity when implemented in clinical practice (section 1.4.1.3).[211,212]

Almost half of all SGA_{both} babies were born after their estimated due date, despite iatrogenic early delivery being indicated for pregnancies in which SGA is detected.[1] This is unlikely to be because pathological factors associated with SGA also cause post-dates pregnancy, given that FGR is associated with preterm birth,[53] and many of the risk factors for SGA (e.g. hypertension, diabetes, maternal age) indicate iatrogenic earlier birth.[248,357] A more likely explanation is that the babies born from pregnancies in which SGA was missed did not become SGA until late in the pregnancy. This is particularly likely to be the case for the pregnancies in which SGA was missed, given that these babies were shown to have less severe SGA than the babies in whom SGA was detected (as assessed by birthweight centile) and that the gestation at the time of SGA first being detected was distributed throughout the third trimester for pregnancies in which SGA was antenatally detected.

Whilst over half of the women with missed SGA received at least one scan during the pregnancy, the EFW at the time of the scan was not calculated to be below the 10th centile. There are many possible reasons for this: that the measurement of fetal biometry was overestimated, that the process of growth restriction commenced later in pregnancy than the scan was conducted, or that the fetal growth was restricted but the weight had not yet dropped below the 10th centile. In a cohort study of nulliparous pregnancies, MacDonald et al (2017) documented that even babies who are born AGA can have fetal growth restriction during pregnancy when assessed longitudinally through serial antenatal ultrasound.[417] In the analysis reported in this chapter, I have documented the extent to which fetal weight was overestimated during a scan performed during the week before the birth of babies in whom SGA was missed antenatally, this has previously been noted including in other babies born with low birthweight at term.[21,418] A meta-analysis has also demonstrated that estimation of fetal weight has low sensitivity (35%) for predicting an SGA birthweight, but that it performed better for prediction of FGR (sensitivity of 70% for birthweight<3rd centile or <10th centile with ultrasound Doppler changes).[229] I have also documented the widening duration between the last scan and the birth for babies born at later gestations, likely caused by local guidelines from clusters in the trial that recommend serial fetal ultrasound scans until 36 weeks' gestation but not after (Table 3.4). Stopping routine offer of fetal growth assessment before term may prevent the diagnosis of SGA that only becomes evident at term.

6.3.9   Strengths and limitations

The strengths of this study lie in its sample size and the richness of the available ultrasound data. This is the largest and most comprehensive study conducted to date and only UK study, comparing the characteristics of women in whom SGA is missed antenatally

to those receiving an antenatal diagnosis.[184,410,411] Given the cluster trial design, the analysis included all births occurring at 13 clusters during the trial period, reducing the risk of recruitment bias at the individual level. There are no other published studies that evaluated the patterns of ultrasound utilisation between these two groups, an analysis that was only possible because of the decision to use EPR data for the DESiGN trial.

Whilst EPR allows the study of a large sample, it is limited by the quality and availability of data.[399] Following exclusion of records that did not have complete data on maternal co-morbidities and antenatal complications, and with the imputation of missing data on maternal demographics, the final dataset included 57.8% of SGA pregnancies. The robustness of the imputed results was previously considered in section 2.2.4.2. Data were not available to distinguish between severe and mild-to-moderate pregnancy-induced hypertension, and so both were considered as an indication for serial fetal growth scans, contrary to current guidelines.[217,248] Data were also not available on some indications for serial fetal growth assessment in pregnancy, although the missing indicators are mostly uncommon or rare: illicit drug use, history of stillbirth or SGA baby, renal impairment, fetal echogenic bowel or significant antepartum haemorrhage. Previous stillbirth or SGA pregnancy could only have been a factor amongst 42.2% of the women with no other known risk factor who were multiparous. A stillbirth occurs in approximately 0.4% of all UK births and given its rarity, including as a recurrent outcome,[419] it is unlikely to have been much more prevalent amongst the otherwise low-risk multiparous women having an SGA baby.[420] In one large cohort study, SGA was diagnosed in 8.2% of babies born to women who had an SGA baby in the first pregnancy.[421] History of SGA is therefore likely to have been the most important risk factor on which information was missing.[1] Illicit drug use, renal impairment, fetal echogenic bowel or significant antepartum haemorrhage are all uncommon or rare conditions.[422-425] As previously described in section 5.2.8, data on other aspects of service use (i.e. attendance at antenatal appointments or unscheduled attendances to antenatal day units) were unreliable and could not be included. Both of the previous international studies on detection of SGA identified associations with aspects of service use (years of midwifery experience of the care provider, care in low-risk midwifery settings) that could not be studied here. [410,411]

A key assumption for this analysis was that women with no ultrasound record had no scans at the site at which they gave birth and did not receive an SGA diagnosis elsewhere. I acknowledge that there will be instances where this assumption is not true, particularly for women who transfer their antenatal care late, either by choice or because they require transfer to a maternity unit with services appropriate for the management of their pregnancy (e.g., higher-level neonatal care if early preterm birth of an SGA fetus is

indicated). However, sensitivity analysis tested this assumption by restricting the sample to women who are expected to have had a near-complete third-trimester ultrasound record as measured by evidence of care both before the third trimester, and at birth using the record of a fetal anomaly scan between 18-24 weeks' gestation as the marker of care commencing at the site before the third trimester. This did not lead to major differences in the findings.

Finally, this analysis was conducted in a sample of babies who were SGA by both the population and customised centile definitions. The reasons for this were described in section 6.2.4.1. Whilst this is not a definition of SGA in common practice in the UK, it was decided it was more relevant that the definition included in the sensitivity analysis (SGA by population centiles) given the high rate of national uptake of GAP and therefore use of customised weight centiles. Babies born SGA_{both} tend to be at highest risk of stillbirth (section 1.1.9) and there were no major differences in the findings of the sensitivity analysis – most differences identified are likely to have been caused by the effect of customisation. A study limited only to babies who were born SGA by customised definitions was not conducted because of the small sample size (n~900), limiting such a detailed study of associated characteristics and ultrasound patterns.

6.3.10 Implication of the findings

I have demonstrated the importance of offering high-quality fetal growth screening for women at low risk of SGA and practice which is concordant with existing national policies recommending an offer of serial scans to women with an indication. In chapter 3, I demonstrated the variation with which both strategies were implemented in practice. Gardosi et al (2020) demonstrated that maternity units that implement GAP with high concordance to the SGA detection reporting and missed case audit components of the intervention achieve higher rates of SGA detection.[281] Reporting and audit are therefore likely to play a major role in quality improvement relevant to this problem. Furthermore, maternity units that were not compliant with offering serial fetal growth scans to all women with an indication (as recommended by national policy)[217] cited resource availability, including sonographer shortages and cost as reasons for not being able to do this (section 3.3.2.3). Common deviations were for lighter smokers (<10 per day) or women with BMI 35-39.9 kg/m[2], or PAPP-A 0.3-0.415 MoM. Since these indications have already been established as risk factors for SGA,[1,217] economic evaluations that assess specific indications for serial fetal growth scans may be required to convince maternity units that offering scans for such indications is cost-effective.

If, as recommended in a recent Health Technology Assessment report (Smith, 2021),[230] the UK does move towards a universal offer of a late pregnancy ultrasound to assess fetal

presentation without concomitant ultrasound assessment of fetal growth, maternity units must monitor their rate of detection of SGA to determine whether the removal of these incidental diagnoses has any clinically meaningful effect on the rate of SGA detection, including the false positive rate.

Further research is required to understand how the rate of SGA detection differs with BMI, including with the use of different screening strategies and centile charts. Overweight BMI was found in 26.4% of pregnancies with missed SGA, and therefore may be a useful target in further research to increase the rate of detection of SGA, but given that approximately a quarter of all women have this characteristic,[426] its use as a target must be studied for both clinical and cost-effectiveness.

Further research is also required to assess alternative strategies of fetal growth screening for women without an indication for serial scans in pregnancy. Whilst policies of universal ultrasound screening have not so far been shown to be effective, many questions remain unanswered. These include the optimal timing of a single growth scan offered to all women at low risk of SGA, or the effect of instead offering two scans to measure the change in the EFW centile as opposed to a single estimate of fetal weight for women who have a one-off indication to assess fetal wellbeing (e.g. because of suspected SGA on measurement of the fundal height, or reduced fetal movements). Furthermore, policies to continue serial fetal growth assessment with scans until birth (including at term) have been introduced into common UK practice through GAP,[87] but were not widely implemented in DESiGN trial clusters (Table 3.4). There is currently no published research studying the benefit of these resource-intense policies, except for when studied through observation of implementation as a component of an otherwise complex intervention.[281-283] Studies of the accuracy of ultrasound assessment of EFW at term vary in their findings,[21,418,427,428] and a policy to introduce growth scans at term must be supported by evidence demonstrating a beneficial effect on maternal and perinatal outcomes, including its scope to prevent stillbirth.[429] Accuracy of fetal weight estimation is problematic and therefore, further research is also required to improve this. Possible areas of research may include adjusted fetal weight formulas, or initiatives to reduce operator bias, such as blinding the operator to the EFW centile until after the scan is finished. MRI is one other possible strategy that has proved accurate in the estimation of fetal weight for LGA babies, however, the cost is currently prohibitive, and it is not yet established whether women would find this acceptable.

6.3.11 Conclusion

SGA was more likely to be missed amongst pregnancies without an indication for serial fetal growth scans, amongst women with a BMI in the range 25.0-29.9 kg/m² and for babies

with less severe SGA. It was less likely to be missed amongst pregnancies where the fetus had a non-cephalic presentation at birth. Nevertheless, two-thirds of pregnancies in the studied SGA sample had no indication for serial fetal growth scans, emphasising the importance of improving the accuracy of techniques that screen for SGA in low-risk populations. For pregnancies in which SGA was missed, over half were born after their estimated due date, and just under half received no growth scan in pregnancy. Amongst those who did receive a scan, the EFW was generally over-estimated, precluding SGA diagnosis, and the duration between the last scan and the birth increased with advancing gestation so that babies born after 39 weeks' had not received a scan for an average of 30 days before the birth.

Reporting of SGA detection rates and local audit are both likely to play significant roles in improving the rate of detection of SGA. There is insufficient evidence at present to offer fetal growth scans to all pregnant women, regardless of the risk of SGA. Further research is needed to determine how the rate of detection of SGA differs by BMI and how this can be improved for women who are overweight, and also into the optimal strategy of fetal growth screening for women who are at low-risk of SGA, including whether altering the timing or number of fetal growth scans can improve the rate of detection if such scans are offered to low-risk populations.

7 THE EFFECT OF THE GROWTH ASSESSMENT PROTOCOL ON THE DETECTION OF THE LARGE FOR GESTATIONAL AGE FETUS: SECONDARY ANALYSIS OF A RANDOMISED CONTROL TRIAL

7.1 INTRODUCTION

Large for gestational age refers to fetal or neonatal size above the 90th centile for gestational age. It is more likely amongst women with obesity or gestational diabetes, and is associated with adverse maternal and perinatal outcomes such as emergency caesarean birth, severe perineal trauma and shoulder dystocia.[2]

Unlike the screening program which exists for the detection of SGA,[217] guidelines on antenatal care in the UK do not recommend screening for LGA amongst all women.[8] There are some situations in which selective fetal growth screening is indicated for women at risk of having an LGA baby, although not necessarily intended to screen for LGA alone; women with diabetes in pregnancy,[357] and women with a BMI of 35 kg/m[2] or higher,[1,93] are both recommended to have serial ultrasound assessment of fetal growth. This policy is in part due to the limited accuracy for LGA screening and the evidence against cost-effectiveness. In a meta-analysis of 29 studies, antenatal ultrasound diagnosis of LGA (EFW>90th centile or >4,000g) amongst nulliparous women had only moderate sensitivity (53.5%), although good specificity (93.9%) of identifying babies who were born LGA (birthweight >90th centile or >4,000g), its performance was better (sensitivity 70.2%, specificity 89.2%) for identifying babies with birthweight >95th centile or >4,500g.[230] Other studies have shown that sensitivity worsens with increasing fetal weight.[430,431] A cost-effectiveness analysis of universal compared to selective ultrasound screening for fetal macrosomia identified that the health benefits were too small to justify a policy of universal screening.[390]

Whilst a universal screening program to detect LGA does not exist in the UK, any program intended to screen for SGA fetuses has the unintended consequence of also identifying LGA fetuses. This can cause maternal or clinician anxiety but without clear strategies for further management.[2,432] GAP is one such example that, while not intended to screen for LGA, does recommend that an accelerative trajectory of the fundal height plotted on the GROW chart should initiate referral for a fetal growth ultrasound assessment. Qualitative evaluation of the acceptability of GAP during the DESiGN trial identified concerns amongst healthcare staff that GAP was inadvertently leading to identification of LGA babies, causing anxiety amongst women about giving birth to a 'big baby' and uncertainty amongst clinicians about which management strategies to offer.[396]

> *"I have a personal concern, […] that the charts are being used to identify large for dates babies as well as small for dates. And that is not what they're designed to do. And I believe that the criteria's saying unless there's been one significant jump and you think actually maybe it's actually maybe it's amniotic fluid, that kind of thing, we shouldn't be, investigating or offering induction for large for dates babies and I think that yeah, I'm worried that that might be a side effect of the personalised growth chart. […] Its some people*

unnecessarily having caesareans, lots of intervention that may not be warranted, it I mean it happens and also with the margin of error it's not always accurate, so". (HP3, site 7)

Following an ultrasound scan that raises the suspicion of LGA, there is limited UK guidance on what management options the pregnant women should be offered. Until March 2019, UK national guidance on the management of suspected macrosomia was only available for pregnant women with diabetes, recommending that this group of women are informed of the risks and benefits of induced labour, vaginal and caesarean birth.[357] More recent guidelines from both NICE and the American College of Obstetricians and Gynaecologists (ACOG) have been published for non-diabetic women.[2,403] The NICE guideline 'Intrapartum care for women with existing medical conditions or obstetric complications and their babies' recommends that women in whom the baby is suspected to be LGA should be counselled that there is uncertainty regarding the diagnosis, which should be taken into account when making a decision about mode of birth. Whilst there is no guidance on making an antenatal decision about mode of birth, the guideline does recommend that a discussion should be had with a woman in labour whose baby is suspected to be LGA, which includes a description of the risks and benefits of the different available modes of birth.[403] The ACOG guideline recommends that women are not offered induction of labour prior to 39 weeks' gestation for fetal macrosomia, but remains uncertain about the benefits of inducing labour for suspected macrosomia after this gestation. The same ACOG guideline does recommend offering caesarean birth for EFW over 4,500g in diabetic pregnancies and over 5,000g in non-diabetic pregnancies, given the high risk of shoulder dystocia for babies born with weights above these thresholds.[2]

7.1.1 Aim and objectives

The aim of this analysis was to assess for the presence of an unintended impact from implementing GAP on the detection and management of pregnancies in which the baby is LGA.

The objectives were to:

1. Compare the characteristics of women and their LGA babies by category of LGA (as defined by the available centile charts).
2. Determine the effect of GAP on the rate of antenatal detection of the LGA fetus, compared to standard practice.
3. Compare maternal and perinatal outcomes for LGA fetuses, between sites implementing GAP and sites offering standard care.

7.2 Methods

7.2.1 Study design

Secondary analysis of a randomised cluster control trial comparing the effect of GAP to standard care. The intention to conduct this secondary analysis was pre-specified in the DESiGN trial protocol.

This study has been reported according to the recommendations of the CONSORT checklist with cluster extension for reporting the results of randomised control trials.[319] The completed checklist is included in Appendix section 10.15.

7.2.2 Study population

The study population was the same as that for the modified intention to treat analysis of the DESiGN trial (section 2.1.3). Births at the two clusters randomly allocated to the intervention that withdrew from implementation of GAP prior to contacting its provider (sites 12 and 13) were excluded. For study of ultrasound utilisation and maternal or perinatal clinical outcomes, only babies who were born LGA as defined by both customised and population centiles (LGA_{both}), at 36^{+0} weeks of gestation or later were included.

7.2.3 Outcomes and exposures

The effect of GAP was compared to standard care. The primary outcome of this analysis was the rate of antenatal ultrasound detection of LGA at or after 34^{+0} weeks of gestation in infants who were confirmed to be LGA_{both} when born at 36^{+0} weeks' gestation or later. Antenatal ultrasound detection of LGA was defined as evidence of an EFW at the time of the last fetal growth scan (conducted at or after 34^{+0} weeks gestation and recorded in the ultrasound EPR) that was above the 90th centile on the fetal weight chart applicable to the trial arm at the time:

- EFW above the 90th centile for population (Hadlock) fetal weight charts for:
    - births in both trial arms during the pre-randomisation phase
    - births in the standard care arm during the trial comparison phase
- EFW above the 90th centile on customised (GROW) fetal weight charts for babies born in intervention arm cluster sites during the trial comparison phase.

Babies born prior to 36^{+0} weeks were excluded from the analysis for the primary outcome because preterm birth does not usually occur with, nor is it indicated by, pathologically excess growth as is often seen in LGA. Babies born between 36^{+0} and 36^{+6}

weeks of gestation were included because babies born to mothers with diabetes may be iatrogenically delivered during this period in response to excessive fetal growth, or uncontrollable maternal hyperglycaemia. Ultrasound scans conducted prior to 34^{+0} weeks of gestation were less likely to be considered informative of LGA confirmed at birth at or after 36^{+0} weeks', or expected to affect maternal or perinatal outcomes through clinical actions.

Secondary outcomes of the study included:

- the rate of antenatal ultrasound detection of LGA at or after 34^{+0} weeks of gestation for babies confirmed to be LGA by customised centiles (LGA_{cust}) at birth (at or after 36^{+0} weeks of gestation), and the same but using population centiles (LGA_{pop}). Detection was defined as per allocated centile chart in the relevant trial arm and phase.
- Screening outcomes (TP, FP) for each definition of LGA.
- Measures of ultrasound utilisation for women giving birth to an LGA_{both} baby at or after 36^{+0} weeks of gestation – percentage of women receiving any ultrasound in pregnancy, number of scans in each pregnancy, percentage of women receiving an ultrasound scan at or after 34^{+0} weeks (with or without EFW), number of scans received at or after 34^{+0} weeks.
- Maternal and perinatal outcomes known to be associated with LGA (Table 7.1).

Table 7.1 - Secondary maternal and perinatal outcomes to be studied in women and their babies who are LGA at birth

Maternal outcomes	Perinatal outcomes
Induction of labour	Gestational age at birth
Mode of birth	Early term birth (<39^{+0} gestational weeks)
Post-partum haemorrhage	Birthweight
Severe perineal trauma (3rd/4th degree tear)	Apgar score <7 at 5 minutes
Episiotomy	Umbilical arterial cord pH <7.10
Epidural use	Admission to neonatal unit
	Hypoxic ischaemic encephalopathy
	Neonatal hypoglycemia
	Naso-gastric tube feeding
	Stillbirth, neonatal death, perinatal death

7.2.4 Management of missing data

The proportion of pregnancy records in which data on maternal or perinatal characteristics and outcomes used for this analysis were missing was calculated using number and percentage. Where individual patient data were missing, these were multiply imputed as per the description in section 2.2.4.2. All results are primarily presented using multiply imputed data, where available.

When there were no data on ultrasound utilisation available for an individual record, it was assumed that the woman had not received an ultrasound at that cluster site unless ultrasound data were missing for an entire trial phase and site (individual ultrasound records were not imputed). Ultrasound-derived biometric measurements were not available for LGA babies born at site 11 during either the pre-randomisation or trial comparison phases, or at site 8 during the pre-randomisation period. Since these data were not available cluster-wide at site 11, they were not imputed and site 11 was excluded from the relevant analyses. Since data were available at site 8 to estimate the rate of detection of LGA during the trial comparison phase, this site was included but the rate of detection of LGA (for the three definitions) was estimated at the cluster level for the pre-randomisation trial phase using a value predicted by a model fitted to values for detection rate from the other clusters and number of infants with an ultrasound scan after 34^{+0} weeks' from all clusters. This predictor was found to be well correlated (r=0.90) with the rate of LGA detection (Figure 7.1), except for at site 10, which was excluded from the model for being an outlier (r=0.98 after exclusion).

Figure 7.1 - Performance of 'number of women with an ultrasound scan after 34 weeks of gestation' as a predictor for 'number of women in whom an antenatal diagnosis of LGA was made at or after 34 weeks' at each cluster site amongst women giving birth to an LGA$_{both}$ baby
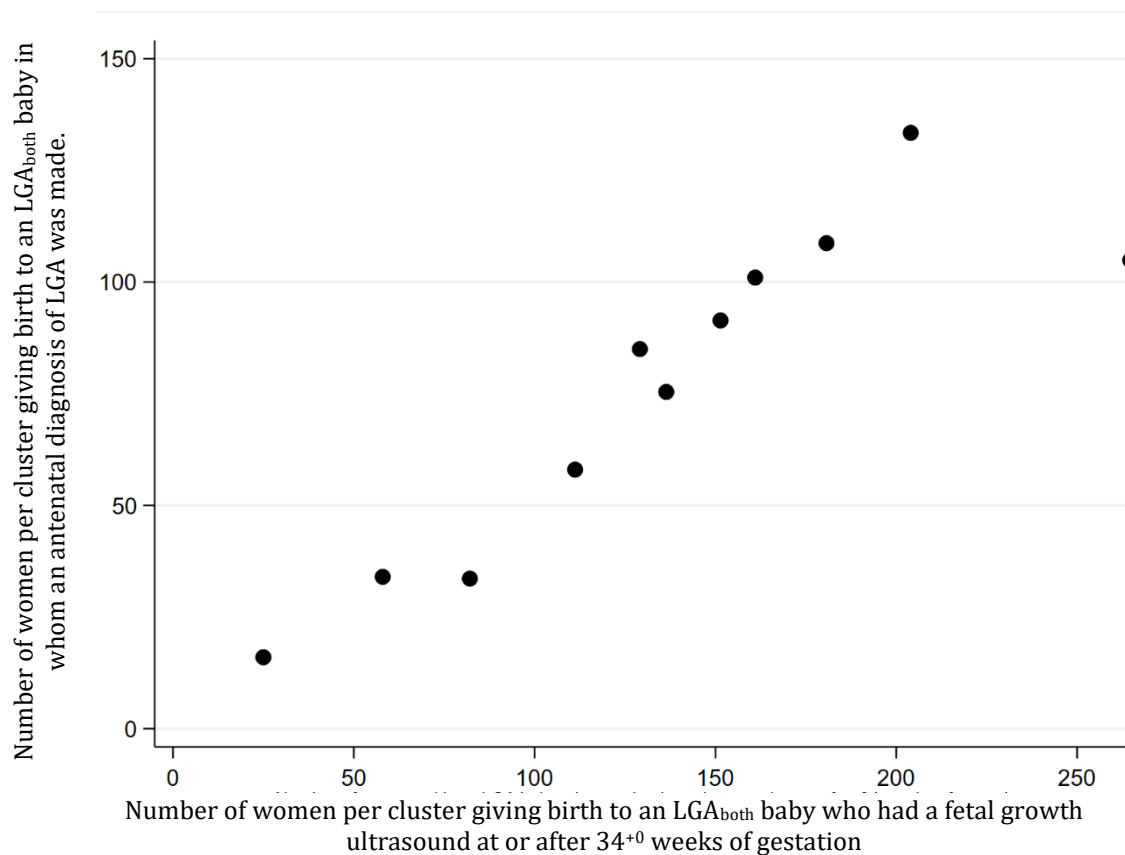
7.2.5 Statistical analysis

Maternal and neonatal characteristics were compared between trial arms and phases for all births, and for those in which the baby was born LGA (using three definitions of LGA: LGA_{both}, LGA_{cust} but not LGA_{pop}, and LGA_{pop} but not LGA_{cust}, Figure 7.2) using summary statistics as described in section 2.2.6.1.

Figure 7.2 - Illustration demonstrating the overlap between LGA_{pop} and LGA_{cust} babies, amongst all babies
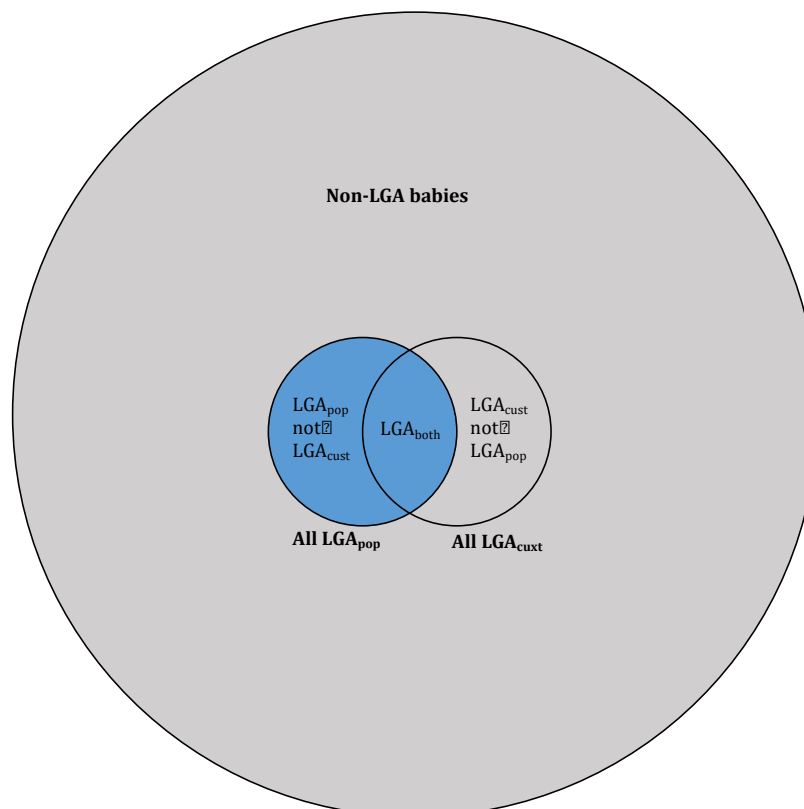
The number and proportion of babies who were LGA_{both}, LGA_{cust} not LGA_{pop}, and LGA_{pop} not LGA_{cust} at birth were calculated. The number and proportion of babies born LGA (LGA_{pop} in the standard care arm and LGA_{cust} in the intervention arm) was also stratified by gestational age at birth categories: $24^{+1}-27^{+6}$ weeks', $28^{+0}-31^{+6}$ weeks', $32^{+0}-35^{+6}$ weeks', $36^{+0}-37^{+6}$, $38^{+0}-39^{+6}$ weeks' and $40^{+0}-42^{+6}$ weeks of gestation. Further analyses are only conducted using data from pregnancies in which the baby was born LGA at or after 36^{+0}.

The number and percentage of LGA_{both}, LGA_{cust} not LGA_{pop}, and LGA_{pop} not LGA_{cust} babies who were antenatally detected by ultrasound at or after 34^{+0} weeks' gestation was calculated. The proportion denominator was all babies born at 36^{+0} weeks' or later who meet the studied definition for LGA. In all cases, the numerator of the proportion was those babies in the denominator for whom the EFW from the last recorded fetal growth ultrasound scan was greater than that for the 90th centile (using Hadlock fetal charts for the

population reference definition and GROW charts for the customised standard definition). Where there was no record of a fetal growth ultrasound scan conducted at or after 34^{+0} weeks', it was assumed that one had not been done and that LGA had not been antenatally detected. The unadjusted difference between the mean cluster proportion of detection in the standard care and intervention arms of the trial was calculated, and the adjusted difference was also calculated using the cluster summary methods described in section 2.2.6.3.1. Unadjusted and adjusted differences in means or proportions were also calculated for the remaining secondary outcomes: FP rate, ultrasound utilisation, maternal and perinatal outcomes.

A sensitivity analysis was conducted using available case data only.

7.3 Results

The characteristics of mothers and babies included in the whole trial have already been presented in Table 2.16 and Table 2.17. Of the 80,856 women and babies included across the pre-randomisation and outcome comparison phases of both arms of the trial, 5.4% were LGA$_{both}$, 1.4% were LGA$_{pop}$ not LGA$_{cust}$ and 3.0% were LGA$_{cust}$ not LGA$_{pop}$. The majority of LGA babies (95.4%) were born at or after 36^{+0} weeks' when LGA was defined by trial arm allocation. The number and proportion of babies who were LGA$_{both}$, LGA$_{cust}$ not LGA$_{pop}$, or LGA$_{pop}$ not LGA$_{cust}$, including stratification by gestational age, are presented by trial arm and time period in Table 7.2 (imputed data) and Appendix section 10.17 (available case data, including raw numbers).

Table 7.2 – Number and proportion of babies who were LGA$_{both}$, LGA$_{cust}$ not LGA$_{pop}$, or LGA$_{pop}$ not LGA$_{cust}$ at birth, presented by trial arm and phase (imputed data)

	Pre-randomisation phase		Outcome comparison phase	
	Standard Care (n=29,404)	**Intervention (n=26,546)**	**Standard Care (n=13,810)**	**Intervention (n=11,096)**
LGA$_{both}$, %	5.7%	5.5%	4.8%	5.0%
LGA$_{pop}$ not LGA$_{cust}$, %	1.4%	1.3%	1.6%	1.4%
LGA$_{cust}$ not LGA$_{pop}$, %	3.0%	3.2%	2.7%	3.1%
Subtotals:				
All LGA$_{pop}$, %	7.1%	6.8%	6.4%	6.4%
All LGA$_{cust}$, %	8.7%	8.7%	7.5%	8.1%
LGA (defined by intervention allocation)* presented by gestational age at birth categories, %				
24^{+1}-27^{+6}	0.1%	0.6%	0.1%	0.7%
28^{+0}-31^{+6}	0.6%	0.8%	0.8%	1.6%
32^{+0}-35^{+6}	2.9%	3.6%	3.7%	3.6%
36^{+0}-37^{+6}	13.2%	13.0%	15.8%	13.1%
38^{+0}-39^{+6}	50.3%	47.0%	53.6%	48.2%
40^{+0}-42^{+6}	33.0%	35.0%	26.0%	32.8%

* All LGA$_{pop}$ in standard care arm LGA$_{cust}$ in intervention arm

The characteristics of women and their LGA babies born during the outcome comparison trial phase are summarised in Table 7.3 (imputed data where available, available case data if values not imputed), stratified by women who had an LGA$_{both}$, LGA$_{pop}$ not LGA$_{cust}$, or LGA$_{cust}$ not LGA$_{pop}$ baby. Compared to women giving birth to LGA$_{both}$ babies in the standard care trial arm, women giving birth to LGA$_{both}$ babies in the trial intervention arm were of a similar age (32.6y, IQR: 28.8-36.5 vs 33.0y, IQR: 29.3-36.2), more likely to be Asian (15.9% vs. 9.0%) and less likely to be white (62.6% vs 66.6%) or black (13.2% vs. 16.5%), less likely to live in the least deprived areas (8.8% vs. 22.3%), had a similar BMI (26.8 kg/m^2, IQR: 23.4-31.6 vs. 26.6 kg/m^2, IQR: 23.4-31.5) and were more likely to be

nulliparous (36.8% vs. 29.9%). Rates of smoking were similar between trial arms. Rates of maternal co-morbidities and antenatal complications are less reliable because of a high proportion of missing values in both arms of the trial.

The characteristics of women and their LGA babies born during the pre-randomisation trial phase, used for adjustments, are presented in Appendix section 10.17. The characteristics of women giving birth to LGA_{both} babies were similar across trial phases for each trial arm, except that there was a lower rate of women with BMI in the range 18.5-24.9 kg/m^2 and higher rate of women with BMI of 25.0 kg/m^2 or greater during the outcome comparison phase for both arms of the trial, and a decrease in nulliparous women during the outcome comparison phase for the intervention arm in both available case and imputed data (pre-randomisation imputed value: 45.5%, outcome comparison imputed value: 36.8%). The change in rate of nulliparity may have been an artefact of the reduction in missing values for parity in this trial arm (21.3% missing during pre-randomisation phase, 2.6% missing during outcome comparison phase).

During both trial phases and across both trial arms (values given for standard care arm during the outcome comparison phase), compared to women giving birth to LGA_{cust} not LGA_{pop} babies, women giving birth to LGA_{pop} but not LGA_{cust} babies were more likely to be white (77.8% vs. 53.8%) and less likely to be black (8.2% vs. 15.9%) or Asian (1.1% vs. 32.2%), and have a higher BMI (29.1 kg/m^2, IQR: 25.7-33.7, vs. 23.7 kg/m^2, IQR: 21.1-26.8). Babies born LGA_{pop} not LGA_{cust} were less likely to be male than babies born LGA_{cust} not LGA_{pop} (19.9% vs. 86.9%). There was no obvious difference in the IMD quintiles or parity of women giving birth to LGA_{pop} not LGA_{cust}, compared to LGA_{cust} not LGA_{pop} babies.

Table 7.3- *Characteristics of pregnancies in which the baby was born LGA during the outcome comparison trial phase, presented by trial arm*

		Standard care (n=13,810)			Intervention (n=11,096)		
		LGA$_{pop}$ not LGA$_{cust}$ (n=221)	LGA$_{cust}$ not LGA$_{pop}$ (n=367)	LGA$_{both}$ (n=618)	LGA$_{pop}$ not LGA$_{cust}$ (n=153)	LGA$_{cust}$ not LGA$_{pop}$ (n=340)	LGA$_{both}$ (n=506)
Imputed data							
Age at estimated conception, years, median [IQR]		33.24 (29.70-36.19)	32.12 (28.35-35.27)	33.03 (29.29-36.15)	32.69 (29.03-36.01)	31.68 (28.45-34.78)	32.63 (28.82-36.45)
Ethnicity, n (%)	White	77.8%	53.8%	66.6%	85.1%	37.7%	62.6%
	Black	8.2%	15.9%	16.5%	4.2%	15.8%	13.2%
	Asian	1.1%	32.2%	9.0%	3.7%	40.0%	15.9%
	Mixed	4.0%	1.4%	1.4%	0.7%	0.6%	0.9%
	Other	8.9%	5.6%	6.5%	6.3%	5.9%	7.5%
Index of Multiple Deprivation Quintiles, n (%)	1 (Least deprived)	17.9%	20.6%	22.3%	7.6%	9.5%	8.8%
	2	13.8%	15.8%	13.9%	11.6%	9.6%	12.2%
	3	15.0%	15.3%	14.6%	27.1%	27.2%	24.6%
	4	23.4%	27.4%	26.0%	31.8%	36.3%	31.7%
	5 (Most deprived)	30.0%	23.5%	23.2%	22.0%	17.5%	22.7%
Body Mass Index, median (IQR)		29.07 (25.65-33.67)	23.71 (21.06-26.83)	26.64 (23.39-31.51)	28.72 (25.04-33.86)	23.14 (20.59-25.15)	26.84 (23.44-31.59)
n (%)	<18.5	0.3%	5.6%	1.4%	0.1%	5.00%	2.1%
	18.5-25	22.2%	55.9%	35.0%	24.5%	68.4%	34.0%
	25-30	35.3%	24.8%	32.9%	31.6%	19.9%	31.3%
	30-35	21.7%	9.6%	19.1%	24.1%	4.4%	18.5%
	35-40	13.7%	2.4%	6.6%	11.7%	1.2%	9.8%
	≥40	6.8%	1.8%	5.0%	8.0%	1.1%	4.3%
Parity, n (%)	Nulliparous	37.6%	35.5%	29.9%	35.4%	42.3%	36.8%
	1	37.3%	42.8%	44.2%	40.0%	36.5%	39.2%
	2	14.6%	12.4%	14.4%	16.9%	13.9%	14.6%
	3	5.1%	5.6%	6.9%	3.3%	5.9%	5.3%
	≥4	5.5%	3.8%	4.6%	4.4%	1.5%	4.0%

	Standard care (n=13,810)			Intervention (n=11,096)		
	LGA$_{pop}$ not LGA$_{cust}$ (n=221)	LGA$_{cust}$ not LGA$_{pop}$ (n=367)	LGA$_{both}$ (n=618)	LGA$_{pop}$ not LGA$_{cust}$ (n=153)	LGA$_{cust}$ not LGA$_{pop}$ (n=340)	LGA$_{both}$ (n=506)
Non-imputed data						
Smoking in pregnancy, n (%)*	7 (3.7%)	10 (2.9%)	20 (3.2%)	8 (6.0%)	5 (1.8%)	13 (2.9%)
Missing smoking	5 (2.6%)	8 (2.3%)	15 (2.4%)	8 (5.7%)	26 (8.7%)	43 (8.7%)
Pre-existing comorbidities, n (%)* **Diabetes**	3 (2.1%)	4 (1.9%)	25 (6.1%)	4 (3.7%)	8 (3.2%)	32 (7.7%)
Missing diabetes	50 (25.5%)	143 (40.7%)	225 (35.6%)	32 (22.7%)	50 (16.8%)	82 (16.5%)
Hypertension	3 (2.1%)	2 (1.0%)	8 (1.9%)	1 (0.9%)	5 (2.0%)	9 (2.1%)
Missing hypertension	50 (25.5%)	140 (39.9%)	219 (34.6%)	26 (18.4%)	45 (15.1%)	62 (12.5%)
Antenatal complications, n (%)* **Gestational diabetes (GDM)**	12 (7.1%)	22 (7.8%)	57 (11.5%)	12 (11.3%)	21 (8.8%)	55 (13.6%)
Missing GDM	27 (13.8%)	68 (19.4%)	137 (21.6%)	35 (24.8%)	59 (19.8%)	92 (18.5%)
Gestational hypertension	*2 (1.8%)*	*4 (1.6%)*	*12 (3.2%)*	*1 (2.6%)*	*8 (10.8%)*	*25 (14.5%)*
Missing gestational hypertension	83 (42.4%)	107 (30.5%)	256 (40.4%)	102 (72.3%)	224 (75.2%)	325 (65.4%)
Infant sex, male*	39 (19.9%)	305 (86.9%)	370 (58.5%)	26 (18.4%)	242 (81.2%)	271 (54.5%)
Missing Infant sex	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

*Mothers and babies with missing data have been excluded from the denominator

299

The percentage of women who had received at least one scan during pregnancy at the cluster site in which they gave birth to an LGA$_{both}$ baby was similar in trial arms (94.5% in the standard care arm vs. 94.8% in the intervention arm, p=0.23) however, there was strong evidence to suggest that babies born LGA$_{both}$ in the intervention arm of the trial had a lower total number of scans than babies born LGA$_{both}$ in the standard care arm (adjusted effect size -0.9, CI: -1.4 to -0.4, p=0.002). There were no differences between trial arms in the proportion of women who received an ultrasound scan after 34^{+0} weeks' with (adjusted effect size -12.0%, CI: -39.7 to 15.6, p=0.29) or without (adjusted effect size -14.2%, CI: -34.7 to 6.4, p=0.14) a measured EFW. Utilisation of ultrasound scans for women giving birth to an LGA$_{both}$ baby in both trial arms and phases is detailed in Table 7.4.

There was no difference between trial arms in the rate of detection of LGA$_{both}$ after 34^{+0} weeks' for babies born at 36^{+0} weeks or later (48.0% vs. 38.0%, adjusted effect size -4.9%, CI: -20.5 to 10.7, p=0.54). There were also no differences in the test positive rate, the rate of detection using other definitions of LGA or any of the other screening test statistics studied. The screening outcomes for mothers and their LGA babies are available in Table 7.5.

There were no differences in secondary outcomes for mothers giving birth to LGA$_{both}$ babies at 36^{+0} weeks' or later between the standard care and intervention arms of the DESiGN trial. There were also no differences between trial arms for any of the neonatal outcomes. There were two few events in either one or both arms to estimate an adjusted effect size for stillbirth and perinatal death; there were no differences in the unadjusted estimates. The secondary outcomes for mothers and their LGA$_{both}$ babies are available in Table 7.6 (maternal outcomes) and Table 7.7 (perinatal outcomes).

Table 7.4 - Ultrasound utilisation amongst pregnancies in which the baby was born LGA$_{both}$ at or after 36^{+0} weeks of gestation, presented by trial arm and phase (data imputed for LGA definitions, ultrasound utilisation not imputed)

	Pre-randomisation phase		Comparison phase		Intervention effect size - unadjusted (95%CI)	Intervention effect size - adjusted (95%CI)	p-value
	Standard	Intervention	Standard Care	Intervention			
Proportion of pregnancies with at least one scan, n (%)	72.4%	88.1%	94.0%	91.0%	-1.3% (-10.4, 7.8)	-0.7% (-9.7, 8.4)	0.86
Number of all scans, mean (SD)	3.2 (3.0)	3.5 (2.3)	4.7 (2.9)	3.8 (2.2)	-0.9 (-2.4, 0.7)	-0.9 (-1.4, -0.4)	0.002
Proportion of pregnancies with scans $\geq34^{+0}$ weeks, n (%)	39.9%	56.7%	73.0%	62.2%	-8.5% (-26.4, 9.5)	-9.8% (-25.8, 6.2)	0.19
Number of scans $\geq34^{+0}$ weeks, mean (SD)	0.9 (1.2)	0.9 (0.8)	1.2 (1.2)	1.0 (1.0)	-0.2 (-0.9, 0.6)	-0.2 (-0.6, 0.3)	0.38
Proportion of pregnancies with scans $\geq34^{+0}$, including EFW, n (%)*	38.3%	57.3%	71.1%	59.1%	-11.8% (-33.2, 9.7)	-12.9% (-31.6, 5.9)	0.13
Number of scans $\geq34^{+0}$ with EFW, mean (SD)*	0.5 (0.7)	0.7 (0.7)	0.9 (0.7)	0.7 (0.7)	-0.2 (-0.5, 0.1)	-0.2 (-0.4, 0.02)	0.07
Of women with EFW recorded $\geq34^{+0}$ weeks, median duration in days between the last growth scan and the birth	18.3 (12.9)	17.7 (10.8)	17.7 (11.5)	18.2 (10.4)	1.0 (-4.2, 6.1)	0.0 (-5.3, 5.4)	0.99

Values for the intervention arm exclude two cluster sites at which data on estimated fetal weight was not available during both trial phases for non-SGA babies.

Table 7.5 - Rate of detection of LGA by each definition, presented by trial arm and phase (imputed data)

	Pre-randomisation phase		Comparison phase		Intervention effect size - unadjusted (95%CI)	Intervention effect size - adjusted (95%CI)	p-value
	Standard Care	GAP	Standard Care	GAP			
Primary outcome							
LGA$_{both}$ at birth, %	5.7%	5.4%	4.7%	4.8%	0.4% (-1.0, 1.8)	0.02% (-0.5, 0.5)	0.93
Antenatal detection, %	24.1%	38.0%	48.0%	38.1%	-6.2% (-21.1, 8.7)	-4.9% (-20.5, 10.6)	0.53
Test positive rate*, %	4.7%	4.2%	3.7%	3.9%			
Secondary outcomes							
All LGA$_{cust}$ at birth, %	8.7%	8.5%	7.5%	7.6%	0.2% (-1.4, 1.8)	0.3% (-0.2, 0.9)	0.20
Antenatal detection, %	19.0%	29.8%	38.2%	36.1%	0.8% (-13.6, 15.2)	0.9% (-13.3, 15.1)	0.90
False positive rate*, %	3.1%	2.4%	6.7%	3.9%	-3.2% (-6.4, 0.1)	-2.0% (-4.4, 0.5)	0.12
All LGA$_{pop}$ at birth, %	7.0%	6.7%	6.3%	6.2%	0.6% (-1.5, 2.7)	-0.1% (-0.8, 0.7)	0.89
Antenatal detection, %	23.2%	36.9%	45.2%	33.1%	-10.3% (-21.5, 0.9)	-7.4% (-19.8, 5.1)	0.25
False positive rate*, %	3.1%	2.4%	6.6%	4.7%	-2.3% (-5.5, 0.8)	-1.5% (-3.7, 0.7)	0.18

One site did not contribute data on detection of LGA during the pre-randomisation phase. Pre-randomisation estimate imputed at cluster level for rate of LGA detection (any definition) at this site to enable calculation of adjusted effect size; cluster excluded from results for other screening outcomes.

Table 7.6 - Secondary outcomes for mothers who gave birth to LGA$_{both}$ babies at or after 36^{+0} weeks of gestation, presented by trial arm and phase (imputed data where available).

	Pre-randomisation phase*		Comparison phase*		Intervention effect size - unadjusted (95%CI)	Intervention effect size - adjusted (95%CI)	p-value
	Standard care	GAP	Standard Care	GAP			
Induction of Labour, %	24.4%	29.9%	24.8%	31.1%	8.1% (-3.0,19.2)	1.6% (-2.4,5.6)	0.42
Mode of birth, %							
Spontaneous vaginal delivery	45.4%	49.1%	43.1%	42.7%	0.9% (-9.5,11.4)	-2.0% (-5.1,1.1)	0.21
Instrumental delivery	9.6%	12.0%	9.7%	10.3%	1.6% (-4.1,7.3)	-0.1% (-3.5,3.3)	0.95
Elective caesarean section	25.5%	23.7%	29.8%	28.9%	-3.3% (-16.2,9.6)	-1.2% (-6.8, 4.3)	0.67
Emergency caesarean section	19.5%	15.0%	17.4%	18.2%	0.7% (-3.3,4.7)	-0.1% (-2.7, 2.5)	0.92
Estimated blood loss, mls mean (SD)	625.7 (482.1)	638.0 (481.7)	652.6 (550.6)	642.6 (454.1)	-31.9 (-117.1, 53.4)	-12.7 (-64.7, 39.3)	0.63
Post-partum haemorrhage (>1500mls), %	5.5%	4.4%	6.1%	4.2%	-2.5% (-5.5, 0.4)	-1.5% (-3.8, 0.8)	0.21
3rd/4th degree tears, %[+]	2.2%	3.0%	1.4%	2.2%	1.4% (-1.1, 3.9)	1.0% (-1.0, 2.9)	0.33
Epidural, %[+]	31.6%	29.1%	31.6%	29.1%	-5.3% (-23.5, 12.9)	2.4% (-7.8, 12.5)	0.65
Episiotomy, %[+]	12.3%	18.5%	13.4%	14.7%	14.5% (-7.8, 36.8)	-4.4% (-9.2, 0.4)	0.07

*Raw numbers cannot be provided for imputed datasets;

[+] LGA definitions imputed but not the marked outcomes.

Table 7.7 - Secondary outcomes for LGAboth babies born at or after 36+0 weeks of gestation, presented by trial arm and phase (imputed data).

	Pre-randomisation phase		Comparison phase		Intervention effect size - unadjusted (95%CI)	Intervention effect size - adjusted (95%CI)	p-value
	Standard Care (n=1607)*	Intervention (n=1358)*	Standard Care (n=627)*	Intervention (n=513)*			
Gestational age at birth, weeks mean (SD)	39.3 (1.3)	39.2 (1.3)	39.2 (1.3)	39.3 (1.3)	0.1 (-0.1, 0.2)	0.1 (-0.1, 0.2)	0.30
Babies born before 39+0 weeks, n/N (%)	36.6%	38.9%	39.5%	37.5%	-2.1% (-9.3, 5.1)	-2.8% (-8.2, 2.7)	0.32
Birthweight, g mean (SD)	4208.0 (352.6)	4179.4 (372.3)	4184.3 (338.9)	4196.2 (333.1)	18.5 (-19.5, 56.6)	24.6 (-2.4, 51.6)	0.07
Apgar score < 7 at 5 minutes, %+	1.8%	1.7%	2.4%	1.3%	-1.0% (-2.4, 0.4)	0.4% (-1.8, 0.9)	0.53
Arterial cord pH < 7.1, %+	2.8%	3.1%	2.8%	3.4%	0.2% (-1.9, 2.4)	0.2% (-1.2, 1.5)	0.81
Neonatal unit admission, %+	16.1%	10.8%	19.9%	9.3%	-10.0% (-27.9, 7.9)	-1.1% (-4.7, 2.5)	0.54
Hypoxic-Ischaemic encephalopathy, %+	0.1%	0.2%	0.3%	0.6%	0.2% (-0.4, 0.8)	0.5% (-0.1, 1.1)	0.12
Hypoglycaemia, %+	2.3%	2.9%	2.9%	2.0%	-0.4% (-2.6, 1.7)	0.4% (-1.9, 2.6)	0.75
Nasogastric tube feeding, %+	1.2%	2.8%	1.6%	2.8%	0.9% (-1.1, 2.9)	0.4% (-1.3, 2.1)	0.64
Stillbirth, %+	0.0%	0.2%	0.2%	0.0%	-0.2% (-0.5, 0.2)	#	#
Neonatal death, %+	0.0%	0.0%	0.0%	0.0%	0.0% (0.0%, 0.0%)	0.0% (0.0%, 0.0%)	1.00
Perinatal mortality, %+	0.0%	0.2%	0.2%	0.0%	-0.2% (-0.5, 0.2)	#	#

*Estimated for imputed dataset
Cannot be calculated due to too few events
+ These outcomes were not imputed

304

7.3.1 Sensitivity analyses

The results of a sensitivity analysis including only available case data are included in Appendix section 10.17. There remained no difference in the rate of detection of LGA between trial arms when LGA was defined by any definition. There were no differences in the findings on utilisation of ultrasound when examined for LGA babies as defined by available case data only (ultrasound data was not otherwise imputed). There were differences in the secondary clinical outcomes when studied in available case data only – there was a trend towards a lower rate of assisted vaginal birth (adjusted effect size -3.75%, 95% CI: -7.90 to 0.40, p=0.07) and evidence of a lower rate of major obstetric haemorrhage (postpartum bleeding of >1500mL, adjusted effect size -2.40%, 95% CI: -4.77 to -0.03, p=0.048) in the intervention arm.

7.4 DISCUSSION

7.4.1 Summary of the key findings

In this secondary analysis of the DESiGN trial, there was no difference in the antenatal ultrasound rate of detection of LGA fetuses born at 36^{+0} weeks' or later, when LGA was defined using any centile definition. This occurred despite women in the intervention arm having almost one fewer scan during the pregnancy. Sensitivity analysis using available case data only identified a lower rate of postpartum haemorrhage above 1,500 mls for women giving birth in the intervention arm which was possibly accompanied by a trend towards lower rates of assisted vaginal birth and episiotomy.

7.4.2 Interpretation of the findings

In the process evaluation of the DESiGN trial, frontline members of clinical staff working in maternity units that had been randomised to implement GAP expressed concerns that GAP was leading to an increase in detection of LGA babies (Chapter 3), without providing guidance on how to manage the pregnancies following an LGA diagnosis. UK guidance on this topic was also not available from NICE or the RCOG at the time of trial, although brief guidance has since been published by NICE (2019), recommending that pregnant women whose babies are suspected to be LGA are counselled regarding the uncertainty of the diagnosis, and are offered an informed discussion about the risks and benefits of the available modes of birth.[403] This perceived increase in the rate of detection of LGA without concurrent guidance on how to manage LGA pregnancies caused anxiety amongst staff and was also believed to have caused anxiety amongst women who were told that they were expecting a 'big baby'. This analysis has shown that there was no difference in the rate of LGA detection for the majority of LGA babies (those born at or near term) when babies were defined as LGA_{both}.

Women giving birth to LGA_{both} babies in the control arm of the DESiGN trial had a higher number of all ultrasound scans during pregnancy, which may have been caused in part by a higher number of ultrasound scans at or after 34 weeks'. Whilst GAP recommends that low-risk women with a fundal height plot above the 90th centile should not be referred for a fetal growth scan unless the growth trajectory is accelerative,[74] half of the guidelines received from maternity clusters in the control arm of the trial did recommend that women suspected to have an LGA baby be referred for a fetal growth scan, albeit heterogeneously (indications in three sites: refer if fundal height>90th centile on population fundal height chart, refer if fundal height >95th centile on fundal height chart or refer if fundal height is more than 3cm greater than the expected number for gestational weeks - using McDonald's rule). This is likely to explain the higher number of ultrasound scans received by women

with LGA^{both} babies in the control arm of the trial. Nevertheless, this difference in number of ultrasound scans did not translate into a difference in the rate of antenatal detection of LGA$_{both}$.

It is possible that the higher number of fetal ultrasound scans conducted in the control arm of the trial did not translate into a difference in the rate of detection of LGA$_{both}$ because of the established inaccuracy of estimating fetal weight, which is magnified for babies with the highest weights causing missed diagnoses. Scioscia et al (2008) found that all 29 studied algorithms for the estimation of fetal weight tended to underestimate the weight of larger babies. Whilst algorithms which estimated fetal weight using only abdominal circumference and femur length performed best (over 60% of EFW were within 10% of the birthweight when calculated a maximum of 48 hours pre-birth), algorithms which also use head circumference (or biparietal diameter) are more commonly used and these tended to underestimate fetal weight by approximately 400g when birthweight was over 4,000g (only 40% of EFWs were within 10% of the measured birthweight for babies of this size).[430] In a different cohort study, 33% of babies born above 4,000g birthweight had a scan which estimated them to be smaller than 4,000g within 3 days prior to birth.[433] Similarly, Malin et al (2015) conducted a systematic review of cohort or diagnostic accuracy studies conducted in women who had third trimester scans to predict fetal macrosomia. The summary sensitivity across 29 studies for the prediction of LGA or macrosomia (birthweight above 4,000g) was 0.56 (95% CI 0.49–0.61) for Hadlock EFW calculated at ultrasound, but this increased to 0.80 (95% confidence interval [95% CI] 0.69–0.87, 4 studies) when ultrasound measured abdominal circumference (AC) >35 cm was used to predict the outcome and 0.93 (95% CI 0.76–0.98, 3 studies) when EFW was calculated following 3D measurement at MRI.[434]

In sensitivity analysis using only available case data, the rate of major postpartum haemorrhage (>1,500 mls) was lower in the intervention arm and with a trend towards a lower rate of assisted vaginal birth and episiotomy. Whilst these findings make sense when presented together, because assisted vaginal birth is an indication for episiotomy and both are associated with greater postpartum blood loss,[435] this finding was not replicated in the primary analysis using imputed data, except for a possible lower rate of episiotomy. All these findings should be interpreted with caution given the number of statistical tests performed. Furthermore, the finding was not clearly accompanied by an increase in the rate of induced labour or caesarean birth, which would explain how this difference was mediated through LGA diagnoses. Whilst mode of birth and estimated blood loss were imputed, episiotomy was not and only 0.24% of records in the control arm and 2.61% in the intervention arm had missing mode of birth during the trial outcome period. The difference

in the sensitivity analysis may have also been caused by differences in the available case and imputed data for the rate of LGA$_{both}$ – 0.2% of all babies were re-categorised as LGA$_{both}$ for the intervention arm when maternal characteristic and neonatal outcome data were imputed. Since all these differences are small, the findings in the sensitivity analysis are likely to have been caused by a combination of these factors.

7.4.3   Strengths and limitations

This is a secondary analysis of a pragmatic RCT that was designed to explore the effect of GAP under real-world conditions, I therefore expect it to have captured, as close as possible, the real effects of GAP when implemented outside of trial conditions. Another strength is in the choice of definition of LGA as LGA$_{both}$. Standard care and GAP both diagnose LGA using different centile charts, thereby LGA$_{pop}$ is not wholly relevant to maternity clusters implementing GAP, or vice versa for LGA$_{cust}$. Both population and customised charts can identify LGA$_{both}$ babies.

This analysis does however have limitations. The primary outcome of the DESiGN trial was the rate of detection of SGA, data collection was prioritised towards enabling analysis for this primary outcome, including manual collection of data on the EFW at the time of the last growth scan in one GAP-implementing cluster that could not provide these data via EPR download. Manual data collection in this cluster was only conducted for babies born SGA; the cluster therefore had to be excluded from the analysis examining rate of detection of LGA. A similar problem also occurred for data on the last EFW of babies born during the pre-randomisation phase at another GAP-implementing site, however, as for the primary DESiGN trial analysis, the rate of detection was predicted for this site using the rate of detection at all other sites, and the proportion of women who had an ultrasound scan at 34^{+0} weeks' or later at all sites (including the site with missing data). This predicted rate was only used to adjust the observed rate for that cluster during the outcome phase, and so its impact is expected to be small.

The DESiGN trial was statistically powered to find a doubling in the odds of the rate of detection of SGA (from 20% to 30%), not detection of LGA. The ability for this secondary analysis to detect a difference in the rate of detection of LGA was further reduced by the loss of one cluster site from the intervention arm. Given the intra-cluster correlation coefficient of 0.028 for the rate of detection of LGA$_{both}$, this secondary analysis had 80% power to find a 2.2-fold change in the odds (from the observed 38% in the control arm to either 21% or 57% in the intervention arm). To find a statistical difference in the observed rate of LGA detection with 80% power, it would be necessary to recruit 16 clusters per trial arm; this is unlikely to be feasible in the UK where most maternity units have already implemented GAP.

Another limitation was the availability or quality of data obtained from EPR systems. We did not originally request data on shoulder dystocia, since this is not associated with being SGA, nor did we collect data on the specific ramifications of shoulder dystocia on the baby (e.g., fetal fractures or brachial plexus injuries). Nevertheless, we did not find a difference in the rate of admission to the neonatal unit, hypoxic ischaemic encephalopathy, or low Apgar scores, all of which I would expect to be affected if there were a clinically important different in shoulder dystocia. We also did not collect data on clinical suspicion of LGA (i.e., where a clinician plots a fundal height above the 90th centile on a growth chart but does not arrange an ultrasound scan). It is possible that anxiety regarding birth of a 'big baby' is also mediated through this route, which we did not test. Whilst we did request and collect data on both pre-existing and gestational diabetes, maternal conditions commonly associated with an LGA fetus, data quality was poor, and these fields were often missing. I therefore cannot assess whether this characteristic was balanced across both trial arms – this may have mediated the observed difference in number of ultrasound scans received for LGA babies between the control and intervention arms. Finally, we assumed that women without a record of an ultrasound scan after 34^{+0} weeks' in the cluster in which they gave birth had not had a scan. It is possible that some women accessed a late third trimester ultrasound scan privately, or others were offered a scan in another maternity unit prior to a late transfer of care. This may have led to an underestimate of the rate of detection of LGA, but there is no reason to think that this would have been unbalanced between trial arms, which were randomly allocated to trial arms according to size and type of maternity unit.

I expect the findings of this study to be generalisable to maternity units with similar fidelity of GAP implementation and similar availability of resource. The DESiGN trial also didn't find a difference in the rate of detection of SGA, and the associated process evaluation noted that this may have been caused by variation in concordance with the GAP protocols at implementing clusters. Maternity staff frequently cited limited resources, specifically limited availability of ultrasound scan appointments, as one reason for low fidelity and dose of GAP implementation (Chapter 3). Just as for the DESiGN trial, I do not know whether the findings of this study would be replicated in maternity units with greater availability of ultrasound appointments, despite GAP recommending that single fundal height plots above the 90th centile on the GROW chart do not indicate referral for a fetal growth ultrasound assessment.

7.4.4 Implication of the findings

It is reassuring to find that, in the context of the DESiGN trial, GAP did not incidentally cause a higher rate of detection of LGA than standard care, nor did it lead to any important difference in secondary outcomes for LGAboth babies in the primary analysis. It is expected

that clinicians implementing GAP will also consider these findings reassuring, given the anxieties that motivated this study.

Nevertheless, the management of pregnancies in which the baby is suspected to be LGA remains uncertain, but with building evidence from a relatively recent Cochrane systematic review which found that induction of labour at or near term (prior to 40 weeks' gestation) resulted in better outcomes for the baby (lower rates of shoulder dystocia and fetal fractures) but not the mother (no difference in caesarean or assisted vaginal birth, higher rates of severe perineal trauma, but only reported from one study).[9] Given that this systematic review was dominated by a single RCT,[241] and still presents uncertainty regarding other perinatal (e.g. Apgar scores at birth, fetal acidosis, brachial plexus injury) and maternal outcomes (severe perineal trauma), further research in this area is needed. The 'Big Baby Trial' is currently underway (expected completion in 2022),[436] being led by a partnership between the University of Warwick and the Perinatal Institute, to determine whether induction of labour at 38 weeks' for babies suspected to be LGA as defined by customised fetal GROW charts, compared to expectant management, reduces the incidence of shoulder dystocia. If this finds that intervention is indicated, it will then be necessary to explore whether selective or universal screening for LGA at term also contributes to an improvement of outcomes.

7.4.5   Conclusion

GAP was not found to inadvertently increase the rate of antenatal ultrasound detection of LGA after 34^{+0} weeks' in LGA_{both} babies born at 36^{+0} weeks of gestation or later, when compared to standard care in the DESiGN trial. Women receiving care in GAP implementing clusters who gave birth to LGA_{both} babies received fewer fetal growth ultrasound scans than similar women receiving care in clusters continuing standard care. This difference is likely to have been caused by variation in local guidelines on referral for suspected LGA. Further research is needed from RCTs to inform clinicians on the safest and most cost-effective methods to manage pregnancies with suspected LGA, before research can be conducted on the clinical usefulness of routine screening for LGA.

8 GENERAL DISCUSSION

The aim of this thesis was to conduct a detailed evaluation of the implementation and cost-effectiveness of the Growth Assessment Protocol in the context of the DESiGN trial, including an assessment of its impact on both large and small-for-gestational-age babies. The DESiGN trial was the first randomised control trial which evaluated the mechanisms and effects of this complex antenatal intervention. Given that the DESiGN trial is the only randomised trial published on this subject to date and that antenatal detection of fetal growth abnormalities is presently highly topical in high-income countries as a major target in the international drive to reduce stillbirth rates, the results presented in this thesis are very likely to influence further policy and research on this subject area.

In this final chapter, I present an overall discussion of the main findings of the thesis, consider the general methodological strengths and weaknesses of the work, and consider the ways in which the findings may influence future policy and research agendas.

8.1 SUMMARY OF FINDINGS AND INTERPRETATION IN THE CONTEXT OF EXISTING LITERATURE

The findings of this thesis can be categorised under two headings. Section 8.1.1 will summarise the findings which are specifically relevant to the Growth Assessment Protocol, as implemented during the DESiGN trial. Section 8.1.2 will summarise the findings relevant to the international drive to improve the rate of stillbirth through increasing the rate of antenatal detection of SGA. During both sections I will interpret the findings in the context of existing literature on this topic.

8.1.1 The Growth Assessment Protocol

The Growth Assessment Protocol is a complex antenatal intervention which aims to prevent many adverse outcomes in maternal and perinatal care through improving the detection of fetal growth problems.[273] It aims to do this through staff training, standardised protocols, implementation of customised fetal growth charts, benchmarking and missed case audit.[273] At the time of writing (June 2021), GAP had already been implemented in 78% of UK maternity units, as well as in New Zealand and Australia.[270] Observational studies have previously demonstrated that GAP is associated with an increase in the rate of antenatal detection of SGA and a decrease in the rate of stillbirth.[277,281-283] The cost-effectiveness of GAP has not previously been studied, nor has its implementation using standardised implementation outcomes. The DESiGN trial was the first example of an RCT that compared the implementation of GAP and standard care on the rate of antenatal detection of SGA,[276] finding no difference in the thirteen cluster sites recruited.[320]

In Chapter 3 of this thesis, I summarised the context, process and strength of GAP implementation, including any barriers and facilitators affecting it. In work drafted for publication elsewhere, I have also reported the acceptability and feasibility of GAP implementation from the perspectives of clinical leaders, midwives and sonographers involved in its use.[396] GAP was implemented at a time when national policy, strategy and guidelines supported the uptake of interventions aiming to reduce the rate of stillbirth, and its role was clear to staff who were aware that it was intended to reduce the rate of stillbirth by increasing the rate of detection of SGA, a risk factor. However, despite the national support, clinical leaders spoke about financial pressures and staff shortages which precluded tight concordance to GAP implementation as recommended. Overall, members of clinical staff generally found GAP to be an acceptable package, but with issues of feasibility. Staff spoke about variable access to the necessary computer hardware, pressures on time caused by staff shortages, conflicting priorities on ultrasound appointments, confusion or partial understanding of the protocol and the inconvenience of not having the necessary information or paperwork to hand. Nevertheless, staff spoke about working in a collaborative environment including with a hands-on clinical leader, and they innovated using a range of materials and methods, both of which enabled them to support each other with GAP implementation.

GAP was implemented with variable strength, as measured by implementation fidelity, reach and dose. Whilst the face-to-face training target was achieved at all sites because it was a pre-requisite to further implementation, only one site achieved the e-learning training target. Local guidelines had variable concordance to the recommendations made by the Perinatal Institute with two sites issuing highly concordant guidance, but one site writing a guideline with low concordance. Clinical leads attributed lower fidelity of both training and guidelines to pressures on staff time and resources. Nevertheless, women were generally assessed for risk of SGA with high fidelity and a high proportion of women had a GROW chart issued at four of the five sites. The findings of dose implemented were variable across maternity units, but dose of implementation was generally low, with few low-risk women receiving the recommended number of fundal height plots on their GROW chart, and few high-risk women receiving the recommended frequency of fetal growth scans. The latter finding was likely caused by local guidelines which recommended a lower frequency of scans that that recommended by GAP.

Process evaluation is key to the understanding of implementation effectiveness including the generation of hypotheses on whether a lack of clinical effect was caused by an ineffective intervention or by inadequate implementation.[298] We do not know if GAP would be more clinically effective if it had been implemented exactly as recommended. The

process evaluation of GAP reported in this thesis is the most detailed study of GAP implementation to date and the first that adopted standard implementation outcomes, but previous observational studies have similarly identified variation in the strength of GAP implementation, demonstrating that this is not a problem confined to this trial. Hugh et al (2020) recently presented a retrospective study in which concordance with the GAP benchmarking component was measured, finding implementing sites who reported the birthweight centile and SGA detection rate of over 75% of their births (n=65/94, 69.2%) had lower stillbirth rates than sites with lower levels of reporting concordance (3.99/1000 vs 4.27/1000).[281] However an observational study such as this cannot prove causation and it is quite likely that sites who performed well in this measure also implement other initiatives intended to reduce stillbirth with high strength, meaning that it is not possible to attribute low stillbirth rates to a single intervention. Variation in GAP implementation was also demonstrated by Lau et al (2020) who studied concordance with components of the national Saving Babies' Lives care bundle, the fetal growth element of which overlaps with GAP guidelines. Of 15 included maternity units who had implemented GAP (none of which were DESiGN sites), the majority were only partially compliant with four out of five components of the element that also feature in GAP guidelines.[437]

The benefit of identifying the target disease during a clinical investigation must be offset against the risk, amongst others, of other incidental and sometimes unimportant diagnoses. In the case of screening for SGA, incidental identification of LGA is the risk. A universal screening programme for LGA has previously been shown to be cost-ineffective,[390] with poor sensitivity,[434] and with uncertainty regarding the most clinically and cost-effective birth plan to offer following diagnosis (section 1.5.4). During assessment of GAP acceptability to staff, I identified that staff members were concerned about incidental diagnosis of LGA, the maternal anxiety it was believed to cause and the subsequent uncertainty around management (section 7.1). In Chapter 7, I demonstrated that GAP was not associated with an increase in the rate of detection of LGA during the DESiGN trial, nor did the women or babies affected by LGA have different outcomes when compared to standard care. However, unlike the analysis studying detection of SGA, the LGA analysis was unfortunately not powered to demonstrate a difference. Furthermore, as for the study of the detection of SGA, the lack of difference between intervention groups may have been caused by variation in implementation. I do not know whether there would have been a difference in LGA diagnoses or outcomes had GAP been implemented as recommended by the provider.

During interviews with clinical leads, cost was frequently cited as a reason for lower implementation fidelity (Chapter 3), cost was also cited as the reason for two cluster sites

allocated to implement GAP at the start of the trial to withdraw from GAP implementation prior to contacting the Perinatal Institute. While the team at the Perinatal Institute have previously estimated that its implementation is associated with an overall cost saving of £110 per pregnancy, this analysis was not conducted using formal economic evaluation methods and its costing framework was flawed (section 5.4.2).[87] In Chapter 5, I studied whether GAP was cost-effective in its aim to decrease the rate of stillbirth through increasing the rate of detection of SGA, when compared to standard clinical care. We calculated that each additional SGA baby detected by GAP would cost £19,463, although this value is associated with significant uncertainty and the chance that GAP is associated with an increase in cost and effectiveness is less than 50% (44.1%), with the chance of other outcomes occurring spread fairly equally throughout the cost-effectiveness plane. As described in Chapter 5, this estimate is most likely to be consistent with a cost per QALY gained of over £100,000, which is above the NICE willingness to pay threshold of £20,000-£30,000.[402] This is an oversimplification of the calculation because the estimates are associated with significant uncertainty and the gain in QALYs though stillbirth prevention is offset by a likely QALY loss elsewhere through lifelong morbidity caused by preterm birth of SGA infants who might not otherwise have been stillborn,[196,406] meaning that the cost per QALY gained overall is likely to be much higher than that presented.

8.1.2 Improving the detection of small for gestational age

Given that SGA is associated with morbidity and mortality (section 1.3.1), that less than half of SGA babies are detected antenatally (section 1.4), that previous studies have demonstrated that improving the rate of antenatal detection of SGA can improve outcomes,[4,196] and that the DESiGN trial did not find that GAP was more clinically effective than standard care at improving this rate,[320] alternative methods are needed.

In Chapter 5, I identified that 65.2% of women who gave birth to an SGA baby did not have a risk factor for SGA indicating serial fetal growth ultrasound scans recorded in the trial dataset. Whilst the dataset did not have complete data on all indications, those with missing data are uncommon or rare and, even if recorded, are unlikely to make a clinically-important change on this estimate. Not surprisingly, absence of any risk factor was associated with an increase in the odds of missing the diagnosis of SGA antenatally. Current UK screening protocols recommend that women without risk factors are monitored for SGA using serial fundal height measurements plotted onto a growth chart, a method which has previously been shown to have a sensitivity of approximately 50%.[226] Unfortunately this sensitivity was not replicated despite augmenting screening with the addition of risk-factor indicated serial fetal growth scans in the DESiGN trial.[320] As the only characteristic identifiable in early pregnancy and associated with a quarter of missed SGA, BMI 25.0-29.9

kg/m² may be a useful target to increase the rate of its detection but given that over a quarter of all pregnant women are overweight, its use as a target must be studied using both clinical- and cost-effectiveness methods.

Also in Chapter 5, I identified that almost half of women with missed SGA in pregnancy did not have a fetal growth scan conducted at the site at which they gave birth. SGA can only be suspected and not diagnosed without a fetal growth scan, but universal fetal growth ultrasound has not previously been shown outside of a trial setting to significantly increase the rate of detection of SGA,[211,212] its use has not been shown to improve antenatal, obstetric or neonatal morbidity, or intervention rates,[231] it can be associated with low specificity,[211] and it is of borderline cost-effectiveness.[230] I hypothesised that even where a woman did receive a scan during the pregnancy, SGA was missed for two reasons. The first is that the EFW at scan was consistently over-estimated for babies born SGA. The second is that recognisable FGR (i.e. restriction which is severe enough to drop a previously normally growing fetus below the 10th centile for gestational age) often does not arise until later in the pregnancy, and most likely at a point after the last fetal growth scan is performed, this is likely to have been further compounded by site-wide policies which did not routinely recommend continuation of serial fetal growth scans beyond 36 weeks of gestation, even for women with risk factors. A meta-analysis previously identified that fetal growth scans conducted later in pregnancy had higher sensitivity for detection of SGA than scans conducted earlier.[229]

8.2 METHODOLOGICAL STRENGTHS AND LIMITATIONS

This thesis was conducted as part of the DESiGN trial. Three out of five of the analyses reported here were pre-planned in the original DESiGN trial protocol. I have considered the methodological strengths and limitations of the trial itself (as the context in which the analyses were conducted), the data collection and management strategies, and the analytical methods for each type of study in the sections that follow.

8.2.1 Study design

The DESiGN trial was planned as a randomised cluster control trial because GAP implementation mandates site-wide training and protocols. Individual recruitment would have meant a serious risk of intervention contamination between recruits. Such a trial design required a moderate number of clusters (and therefore a high number of pregnant women included) to ensure that the analysis was sufficiently powered to identify a difference, if present, in the primary outcome. The early withdrawal from GAP implementation in two clusters risked the trial statistical power but a post-hoc calculation demonstrated that it was retained despite this. A step-wedged cluster trial was an

alternative trial design, but would have meant a longer trial duration, further complicated by uncertainty caused by the variation in time from randomisation until women exposed to the intervention were giving birth (mean 17.4 months). This occurred in the context of national pressure to quickly reduce stillbirth rates and to implement new strategies to detect SGA (such as GAP).[217,332]

Furthermore, most trial clusters were based in London, UK. Evidence from other studies (described in section 8.1.1) demonstrated that maternity units outside London also had difficulty in implementing GAP as recommended and so the trial findings, including those presented in this thesis are likely to be generalisable across the UK. However, the UK has a publicly funded health system with universal access to healthcare without payment at the point of use, but in which resources are limited because of budget constraints and national guidelines on willingness to pay thresholds. Furthermore, the context of implementation (as described in Chapter 3) arises from political policy and is therefore specific to the UK. The generalisability of the findings to other countries, particularly to those with lower economic standing and non-public healthcare, is therefore limited.

8.2.2 Data collection and management

The DESiGN trial was only possible because of our adoption of routinely collected data, obtained from EPR. This method was both feasible and cost-effective, when compared to the workload expected by bespoke data entry of individual patient records into a research database. However, the use of EPR was limited by the availability or completeness of data items. Where data were available on common co-morbidities, the rates were often lower than would be expected for these maternity units. Unfortunately, data were not available for rarer co-morbidities such as renal or autoimmune diseases. There was heterogeneity in how the absence of a condition or intervention was recorded (either required as 'no' or as an absence of a value), requiring rules to be developed to manage this. Data on ethnicity were not as granular as required by the GROW calculator, and so a series of assumptions were also developed to manage this. Furthermore, two sites used ultrasound reporting systems during at least one trial phase from which we were unable to generate a database of fetal biometric measurements. At one of these sites, data were collected manually, but only for the last fetal growth scan for babies in whom SGA was confirmed at birth during the pre-randomisation and trial outcome periods – we therefore only have data on true positive and false negative SGA diagnoses at this site (no data on false positive diagnoses or for LGA babies). At the second site, an ultrasound reporting system with this capability was introduced and used for scans on women who gave birth during the trial outcome period. Manual data collection for the baseline period was planned at this site but was prevented by infection prevention and control policies introduced nationally during the COVID-19

pandemic. Estimates for the screening outcomes in both SGA and LGA babies were therefore predicted at this site using observed data from all other sites and characteristics of women at all sites including the one with missing data.

Missing data for the major characteristics and outcomes were multiply imputed. This was an essential strategy given the quantity and distribution of missing values. It was reassuring that imputation had no major consequence on the value of summary statistics for the imputed values, except for where a change was expected because of a known bias in the data. However, multiple imputation is not a strategy without limitations, the assumption that data is missing at random is not testable and the method is not guaranteed to reduce bias.

8.2.3 Methods of process evaluation

The strengths of the process evaluation reported for this trial lie in the comprehensive and mixed-methods assessment of a wide range of implementation outcomes. Through an innovative development of novel methodology, including case note reviews to investigate implementation strength, we have developed hypotheses to explain the non-superiority of GAP over standard care in the DESiGN trial. Strengths of the qualitative process evaluation included good recruitment that overall led to collection of rich and detailed data.

This process evaluation was limited by the lack of guidance on summarising implementation strength into a composite score, and the low number of sites included in the cluster randomisation, preventing conduct of a mediation analysis to examine the relationship between the site-specific composite or outcome-level implementation strength and the clinical effectiveness of the GAP intervention. We were also limited by an inability to distinguish between the effects of the studied intervention, and those of the Saving Babies' Lives care bundle that had similar aims and was implemented simultaneously, including in clusters allocated to standard care.

The qualitative inquiry was limited by difficulty in recruiting sonographers at the implementing site with lowest overall implementation strength, and at those sites randomised to implement GAP but that did not implement, so we lack data on sonographer perspectives. Frontline obstetricians were not targeted for recruitment, except for where they acted as GAP leads, and so the staff perspectives are drawn more from sonographers or midwives providing routine care.

8.2.4 Methods of economic evaluation

A comprehensive economic evaluation was planned from the outset of the trial to include all major maternity care activities that were hypothesised to potentially be influenced by implementation of GAP or by any subsequent effect on SGA screening

outcomes. This evaluation was limited by low availability of routine electronic recording of midwifery-led antenatal appointments, particularly those that occurred in community settings, and of recording of unplanned antenatal attendances to maternity assessment or day units. It was also limited by a difficulty in accessing hospital administrative data in its entirety at one site.

During the planning stages of the economic evaluation, I identified wide variation in the published costs for each item of maternity activity, these persisted despite inflating to the same year or when costs were compared only within groups that had been derived using the same economic perspective. The reported economic evaluation has been strengthened by the conduct of a detailed systematic review to understand the variation in published costs, highlighting to us the uncertainty in the economic model.

8.2.5   Methods studying clinical outcomes

The analysis presented in Chapter 5 is the largest known and most comprehensive study of factors associated with missed SGA conducted to date, and the only analysis reported from the UK context. The use of data from a cluster trial which included all women giving birth in the site over three years reduces any bias that could be present from individual recruitment and allows the study of less common co-morbidities or characteristics such as hypertension or pre-existing diabetes. Given that there were no differences in the rates of the primary or most of the secondary trial outcomes, the trial population was adopted as a single sample, although adjustments were still made for trial factors to attempt to eliminate potential bias caused by implementation of different interventions.

The analysis assumption that an absent record of a fetal growth scan meant that a scan was not done and therefore SGA was not detected antenatally may have introduced a systematic bias. Although this was tested by excluding women who also had no record of a fetal anomaly scan at the site, making little difference to the results. This is likely to be a more reasonable assumption, given that having both an anomaly scan and a birth at the same site demonstrates continuity of care and presence of an ultrasound record.

The choice of definition for SGA was also a limitation. As described in section 6.2.4.1, I chose to define SGA as birthweight below the 10th centile on both population and customised charts because this was a definition applicable to all babies in all sites throughout the trial. The definition is however limiting because it is not routinely used in clinical practice in the UK, nevertheless, it does represent the group of SGA babies who are most at risk of stillbirth (section 1.1.9). Alternative definitions that are in more common use (either SGA as defined by customised centiles or SGA as defined by population centiles,

rather than by both centile types), were not applicable to both trial arms and therefore unsuitable for application in these analyses.

Finally, the analysis presented in Chapter 5 was also limited by the lack of available data on some risk factors for SGA that indicate serial fetal growth scans. However, most of these risk factors are rare, or uncommon and only have potential to affect less than half of women who were multiparous.

The analysis presented in Chapter 7 is the first study of the effect of GAP, when compared to standard care, on the rate of incidental detection of LGA. This is a secondary analysis of an RCT and therefore likely to lead to a more reliable estimate than a retrospective observational study which is less efficient at controlling for external factors. The analysis was however limited by statistical power, with the sample size chosen for the trial primary outcome, which was seen more commonly in the dataset. The analysis was also limited by the lack of data on shoulder dystocia and poor-quality data on gestational diabetes; these are both important associations of LGA.

8.3 FUTURE POLICY AND RESEARCH

In this thesis, I have explored the implementation and cost-effectiveness of GAP, a complex antenatal intervention aiming to reduce the national rate of stillbirth through increasing the detection of SGA. GAP has already been adopted by 78% of maternity units in the UK,[270] without high quality evidence supporting its implementation and now with trial evidence demonstrating it to be neither clinically- nor cost-effective. The rate of stillbirth in England and Wales is falling (Figure 8.1), but the mechanism behind this decrease is less clear. Whilst implementation of the entire Saving Babies Lives' care bundle was associated with a fall in the rate of stillbirth,[269] the trend was already downwards prior to its implementation. Aside from the fetal growth restriction element of the Saving Babies' Lives care bundle, it also includes an element to improve recognition and management of women with reduced fetal movements, an element of smoking cessation support and one to improve intrapartum fetal monitoring. The AFFIRM trial did not find a fall in the stillbirth rate when the fetal movement bundle was implemented in Scotland.[438] The RCOG Each Baby Counts report found a slight increase in the rate of babies born with intrapartum stillbirth, early neonatal death or severe brain injury.[439] Only smoking cessation has clearly been shown to reduce the rate of stillbirth.[440] It is quite possible that it is a combination of these policies, and of national smoking policies (the UK 'Smoke-free Regulations 2006, enforced 2007),[441] an improvement in surveillance and management of multiple pregnancies,[442] and

the 2013 major update to RCOG guidance on the Investigation and Management of the SGA fetus,[271] that has led to this fall in stillbirth.

Figure 8.1 - Change in rate of stillbirth over time in England and Wales (Office for National Statistics data)
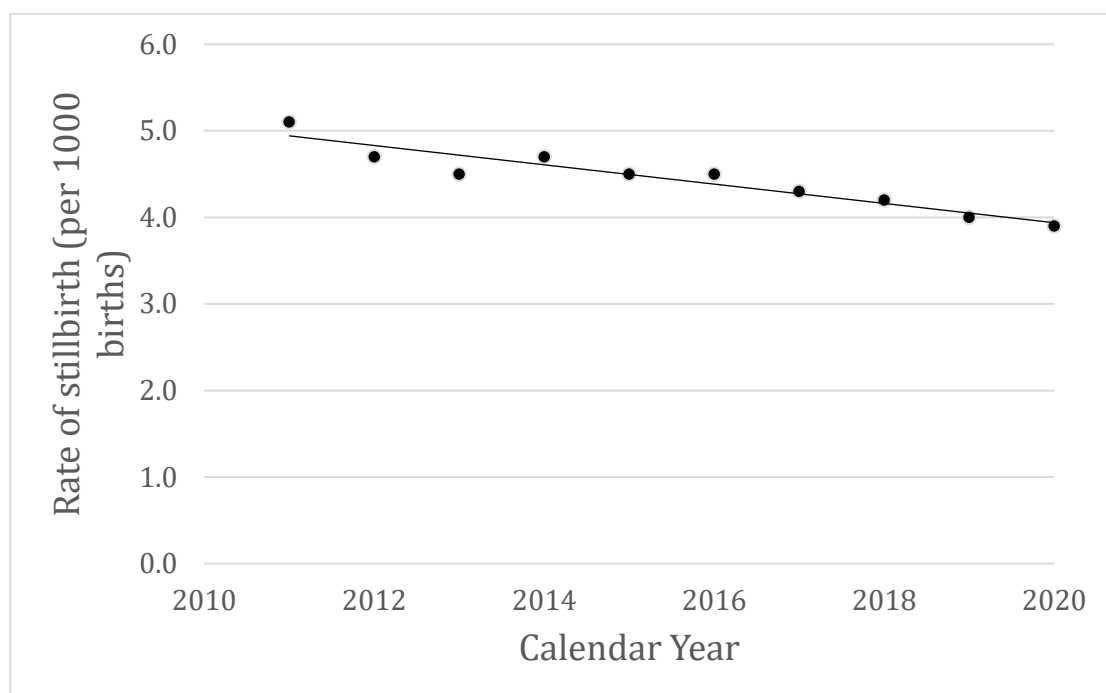
The results of the DESiGN trial do not mean that when implemented with high strength, GAP is no more effective than standard care. This has yet to be proven. At the time of planning the trial, given the extent of national GAP uptake already present, DESiGN was seen as the last opportunity to study GAP implementation in the UK.[276] It is therefore imperative that UK maternity units reflect on the level of GAP implementation achieved locally, ideally by using the implementation outcomes described in Chapter 3, and consider whether the results of the DESiGN trial are likely to be applicable to them. If so, it is likely that standard care is an equally effective and less costly method of SGA screening. For sites who decide to continue with use of GAP, the research presented in Chapter 3 should be used to aid implementation improvement. I identified that the primary barrier was resource availability, in terms of staff time, access to computer hardware and ultrasound appointments. Any site who does continue this path should be aware that GAP was not likely to be cost-effective (in terms of QALYs gained) when implemented in the DESiGN trial, and that any drive to improve implementation is likely to be associated with an increase in cost but with an uncertain impact on SGA detection.

Whilst the work presented in Chapter 7 did not suggest that GAP leads to an increase in the incidental diagnosis of LGA, LGA was diagnosed antenatally with the use of either method. At present, clinical guidelines are vague in terms of recommendations for birth

321

planning when LGA has been diagnosed antenatally. The 'Big Baby Trial' is currently underway (expected completion in 2022) and aims to guide management in this scenario.[436] The research team will determine whether inducing labour at 38 weeks' for LGA babies (suspected after plotting on customised fetal GROW charts) reduces the incidence of shoulder dystocia when compared to expectant management. Unfortunately, the study does not include a qualitative component assessing the acceptability of the strategy to women, this is likely to be essential if the findings are favourable towards an offer of induced labour.

If we are to further pursue an improvement in the detection of SGA to prevent stillbirth, it is essential that we identify which women are at highest risk of missed SGA and that we test alternative strategies either targeted at this group, or universally, which aim to improve the rate of detection. The finding that most women with SGA did not have a risk factor suggests that universal screening is indicated. With current practice, serial fundal height measurement has been ineffective at identifying these women. New strategies which warrant testing under trial conditions include either those that target the effectiveness of fundal height measurement, the dose of which was demonstrated to be universally low in GAP implementing sites of the DESiGN trial, or adopt a more sensitive screening method. For the latter, both single and serial universal ultrasound screening have not been shown to be effective but are both often stopped at 36 weeks' gestation. I hypothesised that this may be too early to detect FGR using the SGA<10th centile definition, particularly for over half of pregnancies with missed SGA where the baby is born after the due date. One strategy that I believe to be worthy of study is a universal offer of a fetal growth ultrasound scan at 38-39 weeks' gestation for pregnancies in which the baby's birth is not expected prior to 39 weeks'. A meta-analysis of studies in which a single screening ultrasound for fetal growth was offered after 32 weeks' gestation found that ultrasound scans had greater sensitivity for detection of AC/EFW<10th centile if conducted at a later gestation, including 4 of the 21 studies that offered ultrasound at term.[229] Such a strategy may be limited by the accuracy of estimating fetal weight at term including a tendency to overestimate (section 6.3.5) and is yet to be proven to be either clinically- or cost-effective (particularly in terms of its potential to impact on perinatal outcomes), both of which must be demonstrated prior to implementation outside of trial settings. Other authors have previously proposed a policy to induce labour at 39 weeks' gestation for all women, but it also remains unknown whether this is a cost-effective strategy to reduce stillbirth, or whether it would be acceptable to women.[443,444] With regards to the other findings that women with BMI in the overweight range are most at risk of missed SGA, a study of how GAP performs for women according to BMI category is planned within my research group and should inform hypothesis on the reasons for this finding.

Finally, I wish to make generic comment on the importance of detailed process evaluation of complex antenatal interventions such as GAP. The process evaluation reported here was fundamental in developing hypotheses about why the DESiGN trial found no difference of effect between the two strategies. The approach to the assessment of implementation strength presented was novel and is expected to influence other process evaluation research, including in non-obstetric specialties. The approach was limited by a lack of guidance on developing a composite measure of implementation strength that could be used to conduct mediation analysis. Further methodological research in this area of implementation science is important to aid future understanding regarding mechanisms of effect.

8.4 Conclusion

The Growth Assessment Protocol has been implemented across much of the UK in response to the national effort to reduce the stillbirth rate through increasing the rate of detection of SGA. In the context of the DESiGN trial, GAP was found to be neither clinically- nor cost-effective when compared to standard care at increasing the rate of detection of SGA, neither did it result in an incidental change in the rate of detection of LGA. The implementation of GAP was challenged by resource availability, this may have been contributory to its lack of effect. The search to find an intervention which improves SGA detection continues. Interventions worthy of further research include those which seek to improve the dose of fundal height measurement for women at low-risk of SGA, a universal offer of a fetal growth scan at term, or strategies targeted specifically to women with BMI in the overweight range. The research presented in this thesis has demonstrated that process evaluation of complex interventions is imperative to develop hypotheses about why or how an intervention is (or is not) effective. Further research is also needed to guide development of composite implementation strength measures, or guide mediation analysis to compare effectiveness between groups of high or low implementation strength.

9 REFERENCES

1. Royal College of Obstetricians & Gynaecologists. Small for Gestational Age Fetus: Investigation & Management. Green-top Guideline No. 31. 2013.
2. Macrosomia: ACOG Practice Bulletin, Number 216. *Obstetrics & Gynecology* 2020; **135**(1).
3. Froen JF, Gardosi JO, Thurmann A, Francis A, Stray-Pedersen B. Restricted fetal growth in sudden intrauterine unexplained death. *Acta Obstet Gynecol Scand* 2004; **83**(9): 801-7.
4. Lindqvist PG, Molin J. Does antenatal identification of small-for-gestational age fetuses significantly improve their outcome? *Ultrasound Obstet Gynecol* 2005; **25**(3): 258-64.
5. Weaver LT. In the balance: weighing babies and the birth of the infant welfare clinic. *Bull Hist Med* 2010; **84**(1): 30-57.
6. Campbell S. A short history of sonography in obstetrics and gynaecology. *Facts Views Vis Obgyn* 2013; **5**(3): 213-29.
7. Nicolson M, Fleming JEE. Imaging and Imagining the Fetus. Baltimore, USA: John Hopkins University Press; 2013.
8. National Institute for Health and Care Excellence. Antenatal care for uncomplicated pregnancies. 2008. https://www.nice.org.uk/Guidance/CG62 (accessed 02 July 2020).
9. Boulvain M, Irion O, Dowswell T, Thornton JG. Induction of labour at or near term for suspected fetal macrosomia. *Cochrane Database Syst Rev* 2016; **2016**(5): CD000938.
10. McLaughlin EJ, Hiscock RJ, Robinson AJ, et al. Appropriate-for-gestational-age infants who exhibit reduced antenatal growth velocity display postnatal catch-up growth. *PLoS One* 2020; **15**(9): e0238700.
11. World Health Organisation. International statistical classification of diseases and related health problems. 2nd ed; 2004.
12. National Institute for Health and Care Excellence. Postnatal care up to 8 weeks after birth. Clinical guideline [CG37]. 2015. https://www.nice.org.uk/guidance/cg37 (accessed 30 November 2020).
13. Royal College of Paediatrics and Child Health. UK-WHO Growth Charts - Fact Sheet 6. Plotting and assessing infants and toddlers up to age 4 years. 2009. https://www.rcpch.ac.uk/sites/default/files/Plotting_toddlers.pdf (accessed 30 November 2020).
14. Haggarty P, Campbell DM, Bendomir A, Gray ES, Abramovich DR. Ponderal index is a poor predictor of in utero growth retardation. *BJOG* 2004; **111**(2): 113-9.
15. Elizabeth NL, Christopher OG, Patrick K. Determining an anthropometric surrogate measure for identifying low birth weight babies in Uganda: a hospital-based cross sectional study. *BMC Pediatr* 2013; **13**: 54.
16. Ioannou C, Talbot K, Ohuma E. Systematic review of methodology used in ultrasound studies aimed at creating charts of fetal size. *BJOG Int J Obstet Gynaecol* 2012; **119**(12): 1425-39.
17. Giuliani F, Ohuma E, Spada E, et al. Systematic review of the methodological quality of studies designed to create neonatal anthropometric charts. *Acta Paediatr* 2015; **104**(10): 987-96.
18. Willocks J, Donald I, Duggan TC, Day N. Foetal Cephalometry by Ultrasound. *J Obstet Gynaecol Br Commonw* 1964; **71**(1): 11-20.
19. Campbell S, Wilkin D. Ultrasonic measurement of fetal abdomen circumference in the estimation of fetal weight. *Br J Obstet Gynaecol* 1975; **82**(9): 689-97.
20. Hadlock FP, Harrist RB, Carpenter RJ, Deter RL, Park SK. Sonographic estimation of fetal weight. The value of femur length in addition to head and abdomen measurements. *Radiology* 1984; **150**(2): 535-40.
21. Milner J, Arezina J. The accuracy of ultrasound estimation of fetal weight in comparison to birth weight: A systematic review. *Ultrasound* 2018; **26**(1): 32-41.
22. Salomon LJ, Alfirevic Z, Da Silva Costa F, et al. ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. *Ultrasound Obstet Gynecol* 2019; **53**(6): 715-23.
23. Calvert JP, Crean EE, Newcombe RG, Pearson JF. Antenatal screening by measurement of symphysis-fundus height. *Br Med J (Clin Res Ed)* 1982; **285**(6345): 846-9.
24. Persson B, Stangenberg M, Lunell NO, Brodin U, Holmberg NG, Vaclavinkova V. Prediction of size of infants at birth by measurement of symphysis fundus height. *Br J Obstet Gynaecol* 1986; **93**(3): 206-11.
25. Rogers MS, Needham PG. Evaluation of fundal height measurement in antenatal care. *Aust N Z J Obstet Gynaecol* 1985; **25**(2): 87-90.
26. Pattinson RC. Antenatal detection of small-for-gestational-age babies. Choice of a symphysis-fundus growth curve. *S Afr Med J* 1988; **74**(6): 282-3.
27. Stuart JM, Healy TJ, Sutton M, Swingler GR. Symphysis-fundus measurements in screening for small-for-dates infants: a community based study in Gloucestershire. *J R Coll Gen Pract* 1989; **39**(319): 45-8.
28. Quaranta P, Currell R, Redman CW, Robinson JS. Prediction of small-for-dates infants by measurement of symphysial-fundal-height. *Br J Obstet Gynaecol* 1981; **88**(2): 115-9.
29. Cnattingius S, Axelsson O, Lindmark G. Symphysis-fundus measurements and intrauterine growth retardation. *Acta Obstet Gynecol Scand* 1984; **63**(4): 335-40.
30. Wallin A, Gyllensward A, Westin B. Symphysis-fundus measurement in prediction of fetal growth disturbances. *Acta Obstet Gynecol Scand* 1981; **60**(3): 317-23.
31. Mathai M. Prediction of small-for-gestational-age infants using a specially calibrated tape measure. *Br J Obstet Gynaecol* 1988; **95**(3): 313-4.
32. American College of Obstetricians and Gynecologists. Fetal growth restriction. ACOG Practice bulletin no. 134. *Obstet Gynecol* 2013; **121**(5): 1122-33.

33. Lausman A, Kingdom J, Maternal Fetal Medicine C. Intrauterine growth restriction: screening, diagnosis, and management. *J Obstet Gynaecol Can* 2013; **35**(8): 741-8.
34. New Zealand Maternal Fetal Medicine Network. Guideline for the management of suspected small for gestational age singleton pregnancies and infants after 34 weeks' gestation. 2014.
35. Institute of Obstetricians and Gynecologists Royal College of Physicians of Ireland. Fetal growth restriction recognition, diagnosis management. Clinical practice guideline no. 28. 2017. http://www.hse.ie/eng/services/publications/Clinical-Strategyand-Programmes/Fetal-Growth-Restriction.pdf. (accessed September 2020).
36. Vayssiere C, Sentilhes L, Ego A, et al. Fetal growth restriction and intra-uterine growth restriction: guidelines for clinical practice from the French College of Gynaecologists and Obstetricians. *Eur J Obstet Gynecol Reprod Biol* 2015; **193**: 10-8.
37. Lees CC, Stampalija T, Baschat A, et al. ISUOG Practice Guidelines: diagnosis and management of small-for-gestational-age fetus and fetal growth restriction. *Ultrasound Obstet Gynecol* 2020; **56**(2): 298-312.
38. Gordijn SJ, Beune IM, Thilaganathan B, et al. Consensus definition of fetal growth restriction: a Delphi procedure. *Ultrasound Obstet Gynecol* 2016; **48**(3): 333-9.
39. Selvaratnam RJ, Davey MA, Mol BW, Wallace EM. Increasing obstetric intervention for fetal growth restriction is shifting birthweight centiles: a retrospective cohort study. *BJOG* 2020; **127**(9): 1074-80.
40. Divon MY, Haglund B, Nisell H, Otterblad PO, Westgren M. Fetal and neonatal mortality in the postterm pregnancy: the impact of gestational age and fetal growth restriction. *Am J Obstet Gynecol* 1998; **178**(4): 726-31.
41. Piper J, Xenakis E, McFarland M, Elliott B, Berkus M, Langer O. Do growth-retarded premature infants have different rates of perinatal morbidity and mortality than appropriately grown premature infants? *Obstetrics & Gynecology* 1996; **87**(2): 169-74.
42. Kramer MS, Olivier M, McLean FH, Willis DM, Usher RH. Impact of intrauterine growth retardation and body proportionality on fetal and neonatal outcome. *Pediatrics* 1990; **86**(5): 707-13.
43. Vieira MC, Relph S, Persson M, Seed PT, Pasupathy D. Determination of birth-weight centile thresholds associated with adverse perinatal outcomes using population, customised, and Intergrowth charts: A Swedish population-based cohort study. *PLoS Med* 2019; **16**(9): e1002902.
44. Choi SKY, Gordon A, Hilder L, et al. Performance of six birthweight and estimated fetal weight standards for predicting adverse perinatal outcomes: a 10-year nationwide population-based study. *Ultrasound Obstet Gynecol* 2020.
45. McCowan LM, Figueras F, Anderson NH. Evidence-based national guidelines for the management of suspected fetal growth restriction: comparison, consensus, and controversy. *Am J Obstet Gynecol* 2018; **218**(2S): S855-S68.
46. Frick AP, Syngelaki A, Zheng M, Poon LC, Nicolaides KH. Prediction of large-for-gestational-age neonates: screening by maternal factors and biomarkers in the three trimesters of pregnancy. *Ultrasound Obstet Gynecol* 2016; **47**(3): 332-9.
47. Norris T, Johnson W, Farrar D, Tuffnell D, Wright J, Cameron N. Small-for-gestational age and large-for-gestational age thresholds to predict infants at risk of adverse delivery and neonatal outcomes: are current charts adequate? An observational study from the Born in Bradford cohort. *Bmj Open* 2015; **5**(3): e006743.
48. Ohuma EO, Altman DG, International F, Newborn Growth Consortium for the 21st C. Design and other methodological considerations for the construction of human fetal and neonatal size and growth charts. *Stat Med* 2019; **38**(19): 3527-39.
49. Wilcox AJ. Birth weight, gestation, and the fetal growth curve. *Am J Obstet Gynecol* 1981; **139**(8): 863-7.
50. Ott WJ. Intrauterine growth retardation and preterm delivery. *Am J Obstet Gynecol* 1993; **168**(6 Pt 1): 1710-5; discussion 5-7.
51. Secher NJ, Kern Hansen P, Thomsen BL, Keiding N. Growth retardation in preterm infants. *Br J Obstet Gynaecol* 1987; **94**(2): 115-20.
52. Doubilet PM, Benson CB, Wilkins-Haug L, Ringer S. Fetuses subsequently born premature are smaller than gestational age-matched fetuses not born premature. *J Ultrasound Med* 2003; **22**(4): 359-63.
53. Tamura RK, Sabbagha RE, Depp R, Vaisrub N, Dooley SL, Socol ML. Diminished growth in fetuses born preterm after spontaneous labor or rupture of membranes. *Am J Obstet Gynecol* 1984; **148**(8): 1105-10.
54. Joseph FA, Hyett JA, Schluter PJ, et al. New Australian birthweight centiles. *Med J Aust* 2020.
55. Hutcheon JA, Platt RW. The missing data problem in birth weight percentiles and thresholds for "small-for-gestational-age". *Am J Epidemiol* 2008; **167**(7): 786-92.
56. Bukowski R, Gahn D, Denning J, Saade G. Impairment of growth in fetuses destined to deliver preterm. *Am J Obstet Gynecol* 2001; **185**(2): 463-7.
57. Groom KM, Poppe KK, North RA, McCowan LM. Small-for-gestational-age infants classified by customized or population birthweight centiles: impact of gestational age at delivery. *Am J Obstet Gynecol* 2007; **197**(3): 239 e1-5.
58. Pritchard NL, Hiscock RJ, Lockie E, et al. Identification of the optimal growth charts for use in a preterm population: An Australian state-wide retrospective cohort study. *PLoS Med* 2019; **16**(10): e1002923.
59. Lackman F, Capewell V, Richardson B, daSilva O, Gagnon R. The risks of spontaneous preterm delivery and perinatal mortality in relation to size at birth according to fetal versus neonatal growth standards. *Am J Obstet Gynecol* 2001; **184**(5): 946-53.
60. Ding G, Tian Y, Zhang Y, Pang Y, Zhang JS, Zhang J. Application of a global reference for fetal-weight and birthweight percentiles in predicting infant mortality. *BJOG* 2013; **120**(13): 1613-21.

61. Ananth CV, Brandt JS, Vintzileos AM. Standard vs population reference curves in obstetrics: which one should we use? *Am J Obstet Gynecol* 2019; **220**(4): 293-6.
62. Hadlock FP, Harrist RB, Martinez-Poyer J. In utero analysis of fetal growth: a sonographic weight standard. *Radiology* 1991; **181**(1): 129-33.
63. Cole TJ, Freeman JV, Preece MA. British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med* 1998; **17**(4): 407-29.
64. Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. *BMC Pediatr* 2013; **13**: 59.
65. Cole TJ, Williams AF, Wright CM, Group RGCE. Revised birth centiles for weight, length and head circumference in the UK-WHO growth charts. *Ann Hum Biol* 2011; **38**(1): 7-11.
66. Freeman JV, Cole TJ, Chinn S, Jones PR, White EM, Preece MA. Cross sectional stature and weight reference curves for the UK, 1990. *Arch Dis Child* 1995; **73**(1): 17-24.
67. Norris T, Seaton SE, Manktelow BN, et al. Updated birth weight centiles for England and Wales. *Arch Dis Child Fetal Neonatal Ed* 2018; **103**(6): F577-F82.
68. Thomson AM, Billewicz WZ, Hytten FE. The assessment of fetal growth. *J Obstet Gynaecol Br Commonw* 1968; **75**(9): 903-16.
69. Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. *Lancet* 1992; **339**(8788): 283-7.
70. Mikolajczyk RT, Zhang J, Betran AP, et al. A global reference for fetal-weight and birthweight percentiles. *Lancet* 2011; **377**(9780): 1855-61.
71. Sahota DS, Kagan KO, Lau TK, Leung TY, Nicolaides KH. Customized birth weight: coefficients and validation of models in a UK population. *Ultrasound Obstet Gynecol* 2008; **32**(7): 884-9.
72. Sanderson DA, Wilcox MA, Johnson IR. The individualised birthweight ratio: a new method of identifying intrauterine growth retardation. *Br J Obstet Gynaecol* 1994; **101**(4): 310-4.
73. Perinatal Institute. www.perinatal.org.uk (accessed 05 January 2021).
74. Clifford S, Giddings S, South M, Williams M, Gardosi J. The Growth Assessment Protocol: a national programme to improve patient safety in maternity care. *MIDIRS Midwifery Digest* 2013; **23**(4): 516-23.
75. Deter RL, Harrist RB, Hadlock FP, Carpenter RJ. Fetal head and abdominal circumferences: II. A critical re-evaluation of the relationship to menstrual age. *J Clin Ultrasound* 1982; **10**(8): 365-72.
76. Gardosi J, Mongelli M, Wilcox M, Chang A. An adjustable fetal weight standard. *Ultrasound Obstet Gynecol* 1995; **6**(3): 168-74.
77. Wilcox M, Gardosi J, Mongelli M, Ray C, Johnson I. Birth weight from pregnancies dated by ultrasonography in a multicultural British population. *BMJ* 1993; **307**(6904): 588-91.
78. Unterscheider J, Geary MP, Daly S, et al. The customized fetal growth potential: a standard for Ireland. *Eur J Obstet Gynecol Reprod Biol* 2013; **166**(1): 14-7.
79. McCowan L, Stewart AW, Francis A, Gardosi J. A customised birthweight centile calculator developed for a New Zealand population. *Aust N Z J Obstet Gynaecol* 2004; **44**(5): 428-31.
80. Mongelli M, Figueras F, Francis A, Gardosi J. A customized birthweight centile calculator developed for an Australian population. *Aust N Z J Obstet Gynaecol* 2007; **47**(2): 128-31.
81. Pain S, Chang AM, Flenady V, Chan FY. Customised birthweight: coefficients for an Australian population and validation of the model. *Aust N Z J Obstet Gynaecol* 2006; **46**(5): 388-94.
82. Figueras F, Meler E, Iraola A, et al. Customized birthweight standards for a Spanish population. *Eur J Obstet Gynecol Reprod Biol* 2008; **136**(1): 20-4.
83. Gardosi J, Francis A. A customized standard to assess fetal growth in a US population. *Am J Obstet Gynecol* 2009; **201**(1): 25 e1-7.
84. Premru-Srsen T, Verdenik I, Mihevc Ponikvar B, Hugh O, Francis A, Gardosi J. Customised birthweight standard for a Slovenian population. *J Perinat Med* 2019; **47**(3): 270-5.
85. Ego A, Subtil D, Grange G, et al. Customized versus population-based birth weight standards for identifying growth restricted infants: a French multicenter study. *Am J Obstet Gynecol* 2006; **194**(4): 1042-9.
86. Gardosi J, Clausson B, Francis A. The value of customised centiles in assessing perinatal mortality risk associated with parity and maternal size. *BJOG* 2009; **116**(10): 1356-63.
87. Williams M, Turner S, Butler E, Gardosi J. Fetal growth surveillance - Current guidelines, practices and challenges. *Ultrasound* 2018; **26**(2): 69-79.
88. Figueras F, Gardosi J. Should we customize fetal growth standards? *Fetal Diagn Ther* 2009; **25**(3): 297-303.
89. de Jong CL, Gardosi J, Dekker GA, Colenbrander GJ, van Geijn HP. Application of a customised birthweight standard in the assessment of perinatal outcome in a high risk population. *Br J Obstet Gynaecol* 1998; **105**(5): 531-5.
90. Bann D, Johnson W, Li L, Kuh D, Hardy R. Socioeconomic inequalities in childhood and adolescent body-mass index, weight, and height from 1953 to 2015: an analysis of four longitudinal, observational, British birth cohort studies. *Lancet Public Health* 2018; **3**(4): e194-e203.
91. Marshall NE, Biel FM, Boone-Heinonen J, Dukhovny D, Caughey AB, Snowden JM. The Association between Maternal Height, Body Mass Index, and Perinatal Outcomes. *Am J Perinatol* 2019; **36**(6): 632-40.
92. Persson PH, Grennert L, Gennser G. Impact of fetal and maternal factors on the normal growth of the biparietal diameter. *Acta Obstet Gynecol Scand Suppl* 1978; **78**: 21-7.
93. Denison FC, Aedla NR, Keag O, et al. Care of Women with Obesity in Pregnancy: Green-top Guideline No. 72. *BJOG* 2019; **126**(3): e62-e106.

94. Zhang X, Platt RW, Cnattingius S, Joseph KS, Kramer MS. The use of customised versus population-based birthweight standards in predicting perinatal mortality. *BJOG* 2007; **114**(4): 474-7.
95. Sjaarda LA, Albert PS, Mumford SL, Hinkle SN, Mendola P, Laughon SK. Customized large-for-gestational-age birthweight at term and the association with adverse perinatal outcomes. *Am J Obstet Gynecol* 2014; **210**(1): 63 e1- e11.
96. Rossen LM. Neighbourhood economic deprivation explains racial/ethnic disparities in overweight and obesity among children and adolescents in the U.S.A. *J Epidemiol Community Health* 2014; **68**(2): 123-9.
97. Anderson NH, Sadler LC, Stewart AW, McCowan LM. Maternal and pathological pregnancy characteristics in customised birthweight centiles and identification of at-risk small-for-gestational-age infants: a retrospective cohort study. *BJOG* 2012; **119**(7): 848-56.
98. Wood AM, Pasupathy D, Pell JP, Fleming M, Smith GC. Trends in socioeconomic inequalities in risk of sudden infant death syndrome, other causes of infant mortality, and stillbirth in Scotland: population based study. *BMJ* 2012; **344**: e1552.
99. Vos AA, Posthumus AG, Bonsel GJ, Steegers EA, Denktas S. Deprived neighborhoods and adverse perinatal outcome: a systematic review and meta-analysis. *Acta Obstet Gynecol Scand* 2014; **93**(8): 727-40.
100. Papageorghiou AT, Ohuma EO, Altman DG, et al. International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the INTERGROWTH-21st Project. *Lancet* 2014; **384**(9946): 869-79.
101. Alberman E. Are our babies becoming bigger? *J R Soc Med* 1991; **84**(5): 257-60.
102. Smith R, Mohapatra L, Hunter M, et al. A case for not adjusting birthweight customized standards for ethnicity: observations from a unique Australian cohort. *Am J Obstet Gynecol* 2019; **220**(3): 277 e1- e10.
103. Knight M, Kurinczuk JJ, Spark P, Brocklehurst P, Ukoss. Inequalities in maternal health: national cohort study of ethnic variation in severe maternal morbidities. *BMJ* 2009; **338**: b542.
104. Nishikawa E, Oakley L, Seed PT, Doyle P, Oteng-Ntim E. Maternal BMI and diabetes in pregnancy: Investigating variations between ethnic groups using routine maternity data from London, UK. *PLoS One* 2017; **12**(6): e0179332.
105. Ganzevoort W, Thilaganathan B, Baschat A, Gordijn SJ. Point. *Am J Obstet Gynecol* 2019; **220**(1): 74-82.
106. Tuck SM, Cardozo LD, Studd JW, Gibb DM, Cooper DJ. Obstetric characteristics in different racial groups. *Br J Obstet Gynaecol* 1983; **90**(10): 892-7.
107. Brooke OG, Butters F, Wood C, Bailey P, Tukmachi F. Size at birth from 37 - 41 weeks gestation: ethnic standards for British infants of both sexes. *J Hum Nutr* 1981; **35**(6): 415-30.
108. Shiono PH, Klebanoff MA, Graubard BI, Berendes HW, Rhoads GG. Birth weight among women of different ethnic groups. *JAMA* 1986; **255**(1): 48-52.
109. Wang X, Guyer B, Paige DM. Differences in gestational age-specific birthweight among Chinese, Japanese and white Americans. *Int J Epidemiol* 1994; **23**(1): 119-28.
110. Freni-Sterrantino A, Afoakwah P, Smith RB, Ghosh RE, Hansell A. Birth weight centiles and small for gestational age by sex and ethnicity for England and Wales. *Arch Dis Child* 2019; **104**(12): 1188-92.
111. Graafmans WC, Richardus JH, Borsboom GJ, et al. Birth weight and perinatal mortality: a comparison of "optimal" birth weight in seven Western European countries. *Epidemiology* 2002; **13**(5): 569-74.
112. Boshari T, Urquia ML, Sgro M, De Souza LR, Ray JG. Differences in birthweight curves between newborns of immigrant mothers vs. infants born in their corresponding native countries: systematic overview. *Paediatr Perinat Epidemiol* 2013; **27**(2): 118-30.
113. North A, MacDonald H. Why are neonatal mortality rates lower in small black infants than in white infants of similar birthweight? *J Pediatr* 1977; **90**: 809-10.
114. Dawson I, Golder RY, Jonas EG. Birthweight by gestational age and its effect on perinatal mortality in white and in Punjabi births: experience at a district general hospital in West London 1967-1975. *Br J Obstet Gynaecol* 1982; **89**(11): 896-9.
115. Alshimmiri MM, Hammoud MS, Al-Saleh EA, Alsaeid KM. Ethnic variations in birthweight percentiles in Kuwait. *Paediatr Perinat Epidemiol* 2003; **17**(4): 355-62.
116. Drooger JC, Troe JW, Borsboom GJ, et al. Ethnic differences in prenatal growth and the association with maternal and fetal characteristics. *Ultrasound Obstet Gynecol* 2005; **26**(2): 115-22.
117. Kierans WJ, Joseph KS, Luo ZC, Platt R, Wilkins R, Kramer MS. Does one size fit all? The case for ethnic-specific standards of fetal growth. *BMC Pregnancy Childbirth* 2008; **8**: 1.
118. Hanley GE, Janssen PA. Ethnicity-specific birthweight distributions improve identification of term newborns at risk for short-term morbidity. *Am J Obstet Gynecol* 2013; **209**(5): 428 e1-6.
119. Hiersch L, Shinar S, Melamed N, et al. Birthweight and large for gestational age trends in non-diabetic women with three consecutive term deliveries. *Arch Gynecol Obstet* 2018; **298**(4): 725-30.
120. Ego A, Subtil D, Grange G, et al. Should parity be included in customised fetal weight standards for identifying small-for-gestational-age babies? Results from a French multicentre study. *BJOG* 2008; **115**(10): 1256-64.
121. Bai J, Wong FW, Bauman A, Mohsin M. Parity and pregnancy outcomes. *Am J Obstet Gynecol* 2002; **186**(2): 274-8.
122. Ego A, Zeitlin J. Parity and growth restriction: on whom the burden of proof? *BJOG: An International Journal of Obstetrics & Gynaecology* 2009; **116**(8): 1136-7.
123. Morrison J, Williams GM, Najman JM, Andersen MJ. The influence of paternal height and weight on birth-weight. *Aust N Z J Obstet Gynaecol* 1991; **31**(2): 114-6.
124. Derraik JGB, Pasupathy D, McCowan LME, et al. Paternal contributions to large-for-gestational-age term babies: findings from a multicenter prospective cohort study. *J Dev Orig Health Dis* 2019; **10**(5): 529-35.

125. Mattsson K, Rylander L. Influence of maternal and paternal birthweight on offspring birthweight - a population-based intergenerational study. *Paediatr Perinat Epidemiol* 2013; **27**(2): 138-44.
126. Owen P, Farrell T, Hardwick JC, Khan KS. Relationship between customised birthweight centiles and neonatal anthropometric features of growth restriction. *BJOG* 2002; **109**(6): 658-62.
127. Hutcheon JA, Zhang X, Cnattingius S, Kramer MS, Platt RW. Customised birthweight percentiles: does adjusting for maternal characteristics matter? *BJOG* 2008; **115**(11): 1397-404.
128. Smith NA, Bukowski R, Thomas AM, Cantonwine D, Zera C, Robinson JN. Identification of pathologically small fetuses using customized, ultrasound and population-based growth norms. *Ultrasound Obstet Gynecol* 2014; **44**(5): 595-9.
129. Zhang J, Mikolajczyk R, Grewal J, Neta G, Klebanoff M. Prenatal application of the individualized fetal growth reference. *Am J Epidemiol* 2011; **173**(5): 539-43.
130. Gaillard R, de Ridder MA, Verburg BO, et al. Individually customised fetal weight charts derived from ultrasound measurements: the Generation R Study. *Eur J Epidemiol* 2011; **26**(12): 919-26.
131. Villar J, Papageorghiou AT, Pang R, et al. The likeness of fetal growth and newborn size across non-isolated populations in the INTERGROWTH-21st Project: the Fetal Growth Longitudinal Study and Newborn Cross-Sectional Study. *Lancet Diabetes Endocrinol* 2014; **2**(10): 781-92.
132. Gardosi J. Fetal growth and ethnic variation. *Lancet Diabetes Endocrinol* 2014; **2**(10): 773-4.
133. Stirnemann J, Villar J, Salomon LJ, et al. International estimated fetal weight standards of the INTERGROWTH-21(st) Project. *Ultrasound Obstet Gynecol* 2017; **49**(4): 478-86.
134. Stampalija T, Ghi T, Rosolen V, et al. Current use and performance of the different fetal growth charts in the Italian population. *Eur J Obstet Gynecol Reprod Biol* 2020; **252**: 323-9.
135. Kiserud T, Piaggio G, Carroli G, et al. The World Health Organization Fetal Growth Charts: A Multinational Longitudinal Study of Ultrasound Biometric Measurements and Estimated Fetal Weight. *PLoS Med* 2017; **14**(1): e1002220.
136. Carberry AE, Gordon A, Bond DM, Hyett J, Raynes-Greenow CH, Jeffery HE. Customised versus population-based growth charts as a screening tool for detecting small for gestational age infants in low-risk pregnant women. *Cochrane Database Syst Rev* 2014; **2014**(5): CD008549.
137. Clausson B, Gardosi J, Francis A, Cnattingius S. Perinatal outcome in SGA births defined by customised versus population-based birthweight standards. *BJOG* 2001; **108**(8): 830-4.
138. Hemming K, Hutton JL, Bonellie S. A comparison of customized and population-based birth-weight standards: the influence of gestational age. *Eur J Obstet Gynecol Reprod Biol* 2009; **146**(1): 41-5.
139. Rowan JA, Luen S, Hughes RC, Sadler LC, McCowan LM. Customised birthweight centiles are useful for identifying small-for-gestational-age babies in women with type 2 diabetes. *Aust N Z J Obstet Gynaecol* 2009; **49**(2): 180-4.
140. Sciscione AC, Gorman R, Callan NA. Adjustment of birth weight standards for maternal and infant characteristics improves the prediction of outcome in the small-for-gestational-age infant. *Am J Obstet Gynecol* 1996; **175**(3 Pt 1): 544-7.
141. McCowan LM, Harding JE, Stewart AW. Customized birthweight centiles predict SGA pregnancies with perinatal morbidity. *BJOG* 2005; **112**(8): 1026-33.
142. Larkin JC, Hill LM, Speer PD, Simhan HN. Risk of morbid perinatal outcomes in small-for-gestational-age pregnancies: customized compared with conventional standards of fetal growth. *Obstet Gynecol* 2012; **119**(1): 21-7.
143. Figueras F, Figueras J, Meler E, et al. Customised birthweight standards accurately predict perinatal morbidity. *Arch Dis Child Fetal Neonatal Ed* 2007; **92**(4): F277-80.
144. Cha HH, Lee SH, Park JS, et al. Comparison of perinatal outcomes in small-for-gestational-age infants classified by population-based versus customised birth weight standards. *Aust N Z J Obstet Gynaecol* 2012; **52**(4): 348-55.
145. Gonzalez-Gonzalez NL, Gonzalez-Davila E, Cabrera F, et al. Application of customized birth weight curves in the assessment of perinatal outcomes in infants of diabetic mothers. *Fetal Diagn Ther* 2015; **37**(2): 117-22.
146. Francis A, Hugh O, Gardosi J. Customized vs INTERGROWTH-21(st) standards for the assessment of birthweight and stillbirth risk at term. *Am J Obstet Gynecol* 2018; **218**(2S): S692-S9.
147. Gibbons K, Chang A, Flenady V, et al. Customised birthweight models: do they increase identification of at-risk infants? *J Paediatr Child Health* 2013; **49**(5): 380-7.
148. Pasupathy D, McCowan LM, Poston L, et al. Perinatal outcomes in large infants using customised birthweight centiles and conventional measures of high birthweight. *Paediatr Perinat Epidemiol* 2012; **26**(6): 543-52.
149. Costantine MM, Mele L, Landon MB, et al. Customized versus population approach for evaluation of fetal overgrowth. *Am J Perinatol* 2013; **30**(7): 565-72.
150. Cha HH, Kim JY, Choi SJ, Oh SY, Roh CR, Kim JH. Can a customized standard for large for gestational age identify women at risk of operative delivery and shoulder dystocia? *J Perinat Med* 2012; **40**(5): 483-8.
151. Todros T, Plazzotta C, Pastorin L. Body proportionality of the small-for-date fetus: is it related to aetiological factors? *Early Hum Dev* 1996; **45**(1-2): 1-9.
152. Blackwell SC, Moldenhauer J, Redman M, Hassan SS, Wolfe HM, Berry SM. Relationship between the sonographic pattern of intrauterine growth restriction and acid-base status at the time of cordocentensis. *Arch Gynecol Obstet* 2001; **264**(4): 191-3.
153. Colley NV, Tremble JM, Henson GL, Cole TJ. Head circumference/abdominal circumference ratio, ponderal index and fetal malnutrition. Should head circumference/abdominal circumference ratio be abandoned? *Br J Obstet Gynaecol* 1991; **98**(6): 524-7.

154. Lin CC, Su SJ, River LP. Comparison of associated high-risk factors and perinatal outcome between symmetric and asymmetric fetal intrauterine growth retardation. *Am J Obstet Gynecol* 1991; **164**(6 Pt 1): 1535-41; discussion 41-2.

155. Redline RW. Classification of placental lesions. *Am J Obstet Gynecol* 2015; **213**(4 Suppl): S21-8.

156. Khong TY, Mooney EE, Ariel I, et al. Sampling and Definitions of Placental Lesions: Amsterdam Placental Workshop Group Consensus Statement. *Arch Pathol Lab Med* 2016; **140**(7): 698-713.

157. Pandya P, Oepkes D, Sebire N, Wapner R. Fetal Medicine. Basic Science and Clinical Practice. 3rd ed: Elsevier Health Sciences; 2020.

158. Burton GJ, Jauniaux E. Pathophysiology of placental-derived fetal growth restriction. *Am J Obstet Gynecol* 2018; **218**(2S): S745-S61.

159. Kulkarni VG, Sunilkumar KB, Nagaraj TS, et al. Maternal and fetal vascular lesions of malperfusion in the placentas associated with fetal and neonatal death: results of a prospective observational study. *Am J Obstet Gynecol* 2021; **225**(6): 660.e1-.e12.

160. Bustamante Helfrich B, Chilukuri N, He H, et al. Maternal vascular malperfusion of the placental bed associated with hypertensive disorders in the Boston Birth Cohort. *Placenta* 2017; **52**: 106-13.

161. Derricott H, Jones RL, Heazell AE. Investigating the association of villitis of unknown etiology with stillbirth and fetal growth restriction - a systematic review. *Placenta* 2013; **34**(10): 856-62.

162. Redline RW. Villitis of unknown etiology: noninfectious chronic villitis in the placenta. *Human Pathology* 2007; **38**(10): 1439-46.

163. Brady CA, Williams C, Sharps MC, et al. Chronic histiocytic intervillositis: A breakdown in immune tolerance comparable to allograft rejection? *Am J Reprod Immunol* 2021; **85**(3): e13373.

164. Redline RW. Extending the Spectrum of Massive Perivillous Fibrin Deposition (Maternal Floor Infarction). *Pediatric and Developmental Pathology* 2020; **24**(1): 10-1.

165. Spinillo A, Gardella B, Muscettola G, Cesari S, Fiandrino G, Tzialla C. The impact of placental massive perivillous fibrin deposition on neonatal outcome in pregnancies complicated by fetal growth restriction. *Placenta* 2019; **87**: 46-52.

166. Umbers AJ, Aitken EH, Rogerson SJ. Malaria in pregnancy: small babies, big problem. *Trends Parasitol* 2011; **27**(4): 168-75.

167. Freeman K, Oakley L, Pollak A, et al. Association between congenital toxoplasmosis and preterm birth, low birthweight and small for gestational age birth. *Bjog* 2005; **112**(1): 31-7.

168. Hendrix N, Berghella V. Non-placental causes of intrauterine growth restriction. *Semin Perinatol* 2008; **32**(3): 161-5.

169. Yamamoto R, Ishii K, Shimada M, et al. Significance of maternal screening for toxoplasmosis, rubella, cytomegalovirus and herpes simplex virus infection in cases of fetal growth restriction. *J Obstet Gynaecol Res* 2013; **39**(3): 653-7.

170. Borrell A, Grande M, Pauta M, Rodriguez-Revenga L, Figueras F. Chromosomal Microarray Analysis in Fetuses with Growth Restriction and Normal Karyotype: A Systematic Review and Meta-Analysis. *Fetal Diagn Ther* 2018; **44**(1): 1-9.

171. Meler E, Sisterna S, Borrell A. Genetic syndromes associated with isolated fetal growth restriction. *Prenat Diagn* 2020; **40**(4): 432-46.

172. Ballard JL, Rosenn B, Khoury JC, Miodovnik M. Diabetic fetal macrosomia: significance of disproportionate growth. *J Pediatr* 1993; **122**(1): 115-9.

173. Jolly MC, Sebire NJ, Harris JP, Regan L, Robinson S. Risk factors for macrosomia and its clinical consequences: a study of 350,311 pregnancies. *Eur J Obstet Gynecol Reprod Biol* 2003; **111**(1): 9-14.

174. Lawn JE, Blencowe H, Waiswa P, et al. Stillbirths: rates, risk factors, and acceleration towards 2030. *Lancet* 2016; **387**(10018): 587-603.

175. Blencowe H, Cousens S, Jassir FB, et al. National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: a systematic analysis. *Lancet Glob Health* 2016; **4**(2): e98-e108.

176. Flenady V, Wojcieszek AM, Middleton P, et al. Stillbirths: recall to action in high-income countries. *Lancet* 2016; **387**(10019): 691-702.

177. Efkarpidis S, Alexopoulos E, Kean L, Liu D, Fay T. Case-control study of factors associated with intrauterine fetal deaths. *MedGenMed* 2004; **6**(2): 53.

178. Gardosi J. Customized fetal growth standards: rationale and clinical application. *Semin Perinatol* 2004; **28**(1): 33-40.

179. Halimeh R, Melchiorre K, Thilaganathan B. Preventing term stillbirth: benefits and limitations of using fetal growth reference charts. *Curr Opin Obstet Gynecol* 2019; **31**(6): 365-74.

180. Flenady V, Koopmans L, Middleton P, et al. Major risk factors for stillbirth in high-income countries: a systematic review and meta-analysis. *Lancet* 2011; **377**(9774): 1331-40.

181. Aminu M, Bar-Zeev S, van den Broek N. Cause of and factors associated with stillbirth: a systematic review of classification systems. *Acta Obstet Gynecol Scand* 2017; **96**(5): 519-28.

182. Reinebrant HE, Leisher SH, Coory M, et al. Making stillbirths visible: a systematic review of globally reported causes of stillbirth. *BJOG* 2018; **125**(2): 212-24.

183. Trudell AS, Cahill AG, Tuuli MG, Macones GA, Odibo AO. Risk of stillbirth after 37 weeks in pregnancies complicated by small-for-gestational-age fetuses. *Am J Obstet Gynecol* 2013; **208**(5): 376 e1-7.

184. Kajdy A, Modzelewski J, Jakubiak M, Pokropek A, Rabijewski M. Effect of antenatal detection of small-for-gestational-age newborns in a risk stratified retrospective cohort. *PLoS One* 2019; **14**(10): e0224553.

185. Pallotto EK, Kilbride HW. Perinatal outcome and later implications of intrauterine growth restriction. *Clin Obstet Gynecol* 2006; **49**(2): 257-69.

186. de Jong CL, Francis A, van Geijn HP, Gardosi J. Fetal growth rate and adverse perinatal events. *Ultrasound Obstet Gynecol* 1999; **13**(2): 86-9.
187. Dowdall D, Flatley C, Kumar S. Birth weight centiles, risk of intrapartum compromise, and adverse perinatal outcomes in term infants. *J Matern Fetal Neonatal Med* 2017; **30**(17): 2126-32.
188. Doctor BA, O'Riordan MA, Kirchner HL, Shah D, Hack M. Perinatal correlates and neonatal outcomes of small for gestational age infants born at term gestation. *Am J Obstet Gynecol* 2001; **185**(3): 652-9.
189. Boghossian NS, Geraci M, Edwards EM, Horbar JD. Morbidity and Mortality in Small for Gestational Age Infants at 22 to 29 Weeks' Gestation. *Pediatrics* 2018; **141**(2).
190. Lees C, Marlow N, Arabin B, et al. Perinatal morbidity and mortality in early-onset fetal growth restriction: cohort outcomes of the trial of randomized umbilical and fetal flow in Europe (TRUFFLE). *Ultrasound Obstet Gynecol* 2013; **42**(4): 400-8.
191. McIntire DD, Bloom SL, Casey BM, Leveno KJ. Birth weight in relation to morbidity and mortality among newborn infants. *N Engl J Med* 1999; **340**(16): 1234-8.
192. Malloy MH. Size for gestational age at birth: impact on risk for sudden infant death and other causes of death, USA 2002. *Arch Dis Child Fetal Neonatal Ed* 2007; **92**(6): F473-8.
193. Jacobsson B, Ahlin K, Francis A, Hagberg G, Hagberg H, Gardosi J. Cerebral palsy and restricted growth status at birth: population-based case-control study. *BJOG* 2008; **115**(10): 1250-5.
194. Jarvis S, Glinianaia SV, Torrioli MG, et al. Cerebral palsy and intrauterine growth in single births: European collaborative study. *Lancet* 2003; **362**(9390): 1106-11.
195. Clayton PE, Cianfarani S, Czernichow P, Johannsson G, Rapaport R, Rogol A. Management of the child born small for gestational age through to adulthood: a consensus statement of the International Societies of Pediatric Endocrinology and the Growth Hormone Research Society. *J Clin Endocrinol Metab* 2007; **92**(3): 804-10.
196. Andreasen LA, Tabor A, Norgaard LN, Rode L, Gerds TA, Tolsgaard MG. Detection of growth-restricted fetuses during pregnancy is associated with fewer intrauterine deaths but increased adverse childhood outcomes: an observational study. *BJOG* 2021; **128**(1): 77-85.
197. Hediger ML, Overpeck MD, McGlynn A, Kuczmarski RJ, Maurer KR, Davis WW. Growth and fatness at three to six years of age of children born small- or large-for-gestational age. *Pediatrics* 1999; **104**(3): e33.
198. Kramer MS. Invited commentary: association between restricted fetal growth and adult chronic disease: is it causal? Is it important? *Am J Epidemiol* 2000; **152**(7): 605-8.
199. Barker DJ. The developmental origins of adult disease. *J Am Coll Nutr* 2004; **23**(6 Suppl): 588S-95S.
200. Geelhoed JJ, Jaddoe VW. Early influences on cardiovascular and renal development. *Eur J Epidemiol* 2010; **25**(10): 677-92.
201. Boulet SL, Alexander GR, Salihu HM, Pass M. Macrosomic births in the united states: determinants, outcomes, and proposed grades of risk. *Am J Obstet Gynecol* 2003; **188**(5): 1372-8.
202. Zhang X, Decker A, Platt RW, Kramer MS. How big is too big? The perinatal consequences of fetal macrosomia. *Am J Obstet Gynecol* 2008; **198**(5): 517 e1-6.
203. Mondestin MA, Ananth CV, Smulian JC, Vintzileos AM. Birth weight and fetal death in the United States: the effect of maternal diabetes during pregnancy. *Am J Obstet Gynecol* 2002; **187**(4): 922-6.
204. Iliodromiti S, Mackay DF, Smith GC, et al. Customised and Noncustomised Birth Weight Centiles and Prediction of Stillbirth and Infant Mortality and Morbidity: A Cohort Study of 979,912 Term Singleton Pregnancies in Scotland. *PLoS Med* 2017; **14**(1): e1002228.
205. American College of Obstetricians and Gynecologists. Fetal macrosomia. ACOG Practice Bulletin No. 22. *ACOG* 2000.
206. Persson M, Pasupathy D, Hanson U, Norman M. Disproportionate body composition and perinatal outcome in large-for-gestational-age infants to mothers with type 1 diabetes. *BJOG* 2012; **119**(5): 565-72.
207. Vieira MC, McCowan LME, North RA, et al. Antenatal risk factors associated with neonatal morbidity in large-for-gestational-age infants: an international prospective cohort study. *Acta Obstet Gynecol Scand* 2018; **97**(8): 1015-24.
208. Boney CM, Verma A, Tucker R, Vohr BR. Metabolic syndrome in childhood: association with birth weight, maternal obesity, and gestational diabetes mellitus. *Pediatrics* 2005; **115**(3): e290-6.
209. Rogers I, Group E-BS. The influence of birthweight and intrauterine environment on adiposity and fat distribution in later life. *Int J Obes Relat Metab Disord* 2003; **27**(7): 755-77.
210. Backe B, Nakling J. Effectiveness of antenatal care: a population based study. *Br J Obstet Gynaecol* 1993; **100**(8): 727-32.
211. Monier I, Blondel B, Ego A, Kaminiski M, Goffinet F, Zeitlin J. Poor effectiveness of antenatal detection of fetal growth restriction and consequences for obstetric management and neonatal outcomes: a French national study. *BJOG* 2015; **122**(4): 518-27.
212. Jahn A, Razum O, Berle P. Routine screening for intrauterine growth retardation in Germany: low sensitivity and questionable benefit for diagnosed cases. *Acta Obstet Gynecol Scand* 1998; **77**(6): 643-8.
213. Mattioli KP, Sanderson M, Chauhan SP. Inadequate identification of small-for-gestational-age fetuses at an urban teaching hospital. *Int J Gynaecol Obstet* 2010; **109**(2): 140-3.
214. Kean L, Liu D. Antenatal care as a screening tool for the detection of small for gestational age babies in the low risk population. *Journal of Obstetrics & Gynaecology* 1996; (16): 77-82.
215. Chauhan SP, Beydoun H, Chang E, et al. Prenatal detection of fetal growth restriction in newborns classified as small for gestational age: correlates and risk of neonatal morbidity. *Am J Perinatol* 2014; **31**(3): 187-94.

216. Fratelli N, Valcamonico A, Prefumo F, Pagani G, Guarneri T, Frusca T. Effects of antenatal recognition and follow-up on perinatal outcomes in small-for-gestational age infants delivered after 36 weeks. *Acta Obstet Gynecol Scand* 2013; **92**(2): 223-9.
217. O'Conner D. Saving Babies' Lives: A care bundle for reducing stillbirth: NHS England, 2016.
218. Morse K, Williams A, Gardosi J. Fetal growth screening by fundal height measurement. *Best Pract Res Clin Obstet Gynaecol* 2009; **23**(6): 809-18.
219. Bailey SM, Sarmandal P, Grant JM. A comparison of three methods of assessing inter-observer variation applied to measurement of the symphysis-fundal height. *Br J Obstet Gynaecol* 1989; **96**(11): 1266-71.
220. Jelks A, Cifuentes R, Ross MG. Clinician bias in fundal height measurement. *Obstet Gynecol* 2007; **110**(4): 892-9.
221. Goto E. Prediction of low birthweight and small for gestational age from symphysis-fundal height mainly in developing countries: a meta-analysis. *J Epidemiol Community Health* 2013; **67**(12): 999-1005.
222. Lindhard A, Nielsen PV, Mouritsen LA, Zachariassen A, Sorensen HU, Roseno H. The implications of introducing the symphyseal-fundal height-measurement. A prospective randomized controlled trial. *Br J Obstet Gynaecol* 1990; **97**(8): 675-80.
223. Robert Peter J, Ho JJ, Valliapan J, Sivasangari S. Symphysial fundal height (SFH) measurement in pregnancy for detecting abnormal fetal growth. *Cochrane Database Syst Rev* 2015; **2015**(9): CD008136.
224. Westin B. Gravidogram and fetal growth. Comparison with biochemical supervision. *Acta Obstet Gynecol Scand* 1977; **56**(4): 273-82.
225. Gardosi J, Francis A. Controlled trial of fundal height measurement plotted on customised antenatal growth charts. *Br J Obstet Gynaecol* 1999; **106**(4): 309-17.
226. Roex A, Nikpoor P, van Eerd E, Hodyl N, Dekker G. Serial plotting on customised fundal height charts results in doubling of the antenatal detection of small for gestational age fetuses in nulliparous women. *Aust N Z J Obstet Gynaecol* 2012; **52**(1): 78-82.
227. Hadlock FP, Harrist RB, Sharman RS, Deter RL, Park SK. Estimation of fetal weight with the use of head, body, and femur measurements--a prospective study. *Am J Obstet Gynecol* 1985; **151**(3): 333-7.
228. Heazell AE, Hayes DJ, Whitworth M, Takwoingi Y, Bayliss SE, Davenport C. Biochemical tests of placental function versus ultrasound assessment of fetal size for stillbirth and small-for-gestational-age infants. *Cochrane Database Syst Rev* 2019; **5**: CD012245.
229. Caradeux J, Martinez-Portilla RJ, Peguero A, Sotiriadis A, Figueras F. Diagnostic performance of third-trimester ultrasound for the prediction of late-onset fetal growth restriction: a systematic review and meta-analysis. *Am J Obstet Gynecol* 2019; **220**(5): 449-59 e19.
230. Smith GC, Moraitis AA, Wastlund D, et al. Universal late pregnancy ultrasound screening to predict adverse outcomes in nulliparous women: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2021; **25**(15): 1-190.
231. Bricker L, Neilson JP, Dowswell T. Routine ultrasound in late pregnancy (after 24 weeks' gestation). *Cochrane Database Syst Rev* 2008; (4): CD001451.
232. Gagnon A, Wilson RD, Society Of O, Gynaecologists Of Canada Genetics C. Obstetrical complications associated with abnormal maternal serum markers analytes. *J Obstet Gynaecol Can* 2008; **30**(10): 918-32.
233. Morris RK, Cnossen JS, Langejans M, et al. Serum screening with Down's syndrome markers to predict pre-eclampsia and small for gestational age: systematic review and meta-analysis. *BMC Pregnancy Childbirth* 2008; **8**: 33.
234. Benton SJ, McCowan LM, Heazell AE, et al. Placental growth factor as a marker of fetal growth restriction caused by placental dysfunction. *Placenta* 2016; **42**: 1-8.
235. Dugoff L, Hobbins JC, Malone FD, et al. First-trimester maternal serum PAPP-A and free-beta subunit human chorionic gonadotropin concentrations and nuchal translucency are associated with obstetric complications: a population-based screening study (the FASTER Trial). *Am J Obstet Gynecol* 2004; **191**(4): 1446-51.
236. Spencer K, Cowans NJ, Avgidou K, Molina F, Nicolaides KH. First-trimester biochemical markers of aneuploidy and the prediction of small-for-gestational age fetuses. *Ultrasound Obstet Gynecol* 2008; **31**(1): 15-9.
237. Lin S, Shimizu I, Suehara N, Nakayama M, Aono T. Uterine artery Doppler velocimetry in relation to trophoblast migration into the myometrium of the placental bed. *Obstet Gynecol* 1995; **85**(5 Pt 1): 760-5.
238. Sotiriadis A, Hernandez-Andrade E, da Silva Costa F, et al. ISUOG Practice Guidelines: role of ultrasound in screening for and follow-up of pre-eclampsia. *Ultrasound Obstet Gynecol* 2019; **53**(1): 7-22.
239. Cnossen J, Morris R, ter Riet G. Use of uterine artery Doppler ultrasonography to predict pre-eclampsia and intrauterine growth restriction: a systematic review and bivariable meta-analysis. . *CMAJ* 2008; **178**(6): 701-11.
240. Garcia B, Llurba E, Valle L, et al. Do knowledge of uterine artery resistance in the second trimester and targeted surveillance improve maternal and perinatal outcome? UTOPIA study: a randomized controlled trial. *Ultrasound Obstet Gynecol* 2016; **47**(6): 680-9.
241. Boulvain M, Senat MV, Perrotin F, et al. Induction of labour versus expectant management for large-for-date fetuses: a randomised controlled trial. *Lancet* 2015; **385**(9987): 2600-5.
242. Ben-Haroush A, Yogev Y, Hod M, Bar J. Predictive value of a single early fetal weight estimate in normal pregnancies. *Eur J Obstet Gynecol Reprod Biol* 2007; **130**(2): 187-92.
243. Coomarasamy A, Connock M, Thornton J, Khan KS. Accuracy of ultrasound biometry in the prediction of macrosomia: a systematic quantitative review. *BJOG* 2005; **112**(11): 1461-6.
244. Hehir MP, Burke N, Burke G, et al. Sonographic markers of fetal adiposity and risk of Cesarean delivery. *Ultrasound Obstet Gynecol* 2019; **54**(3): 338-43.

245. Mazzone E, Dall'Asta A, Kiener AJO, et al. Prediction of fetal macrosomia using two-dimensional and three-dimensional ultrasound. *Eur J Obstet Gynecol Reprod Biol* 2019; **243**: 26-31.

246. Kadji C, Cannie MM, Resta S, et al. Magnetic resonance imaging for prenatal estimation of birthweight in pregnancy: review of available data, techniques, and future perspectives. *Am J Obstet Gynecol* 2019; **220**(5): 428-39.

247. Kadji C, Cannie MM, De Angelis R, et al. Prenatal prediction of postnatal large-for-dates neonates using a simplified MRI method: comparison with conventional 2D ultrasound estimates. *Ultrasound Obstet Gynecol* 2018; **52**(2): 250-7.

248. National Institute for Health and Care Excellence. Hypertension in pregnancy: diagnosis and management. 2019. https://www.nice.org.uk/guidance/ng133.

249. Duley L, Henderson-Smart DJ, Meher S, King JF. Antiplatelet agents for preventing pre-eclampsia and its complications. *Cochrane Database Syst Rev* 2007; (2): CD004659.

250. Lumley J, Chamberlain C, Dowswell T, Oliver S, Oakley L, Watson L. Interventions to promote smoking cessation during pregnancy. . *Cochrane Database Syst Rev* 2009; (3): CD001055.

251. Hodgetts VA, Morris RK, Francis A, Gardosi J, Ismail KM. Effectiveness of folic acid supplementation in pregnancy on reducing the risk of small-for-gestational age neonates: a population study, systematic review and meta-analysis. *BJOG* 2015; **122**(4): 478-90.

252. Kramer M, Kakuma R. Energy and protein intake in pregnancy. . *Cochrane Database Syst Rev* 2010; (9): CD000032.

253. Meher S, Duley L. Progesterone for preventing pre-eclampsia and its complications. . *Cochrane Database Syst Rev* 2006; (4): CD006175.

254. Hofmeyr G, Lawrie T, Atallah A, Duley L. Calcium supplementation during pregnancy for preventing hypertensive disorders and related problems. . *Cochrane Database Syst Rev* 2010; (8): CD001059.

255. Sharp A, Cornforth C, Jackson R, et al. Maternal sildenafil for severe fetal growth restriction (STRIDER): a multicentre, randomised, placebo-controlled, double-blind trial. *Lancet Child Adolesc Health* 2018; **2**(2): 93-102.

256. McCowan LM, Harding J, Roberts A, Barker S, Ford C, Stewart A. Administration of low-dose aspirin to mothers with small for gestational age fetuses and abnormal umbilical Doppler studies to increase birthweight: a randomised double-blind controlled trial. *Br J Obstet Gynaecol* 1999; **106**(7): 647-51.

257. Mazarico E, Molinet-Coll C, Martinez-Portilla RJ, Figueras F. Heparin therapy in placental insufficiency: Systematic review and meta-analysis. *Acta Obstet Gynecol Scand* 2020; **99**(2): 167-74.

258. Johansson K, Cnattingius S, Naslund I, et al. Outcomes of pregnancy after bariatric surgery. *N Engl J Med* 2015; **372**(9): 814-24.

259. Davenport MH, Meah VL, Ruchat SM, et al. Impact of prenatal exercise on neonatal and childhood outcomes: a systematic review and meta-analysis. *Br J Sports Med* 2018; **52**(21): 1386-96.

260. Wiebe HW, Boule NG, Chari R, Davenport MH. The effect of supervised prenatal exercise on fetal growth: a meta-analysis. *Obstet Gynecol* 2015; **125**(5): 1185-94.

261. Muktabhant B, Lawrie TA, Lumbiganon P, Laopaiboon M. Diet or exercise, or both, for preventing excessive weight gain in pregnancy. *Cochrane Database Syst Rev* 2015; (6): CD007145.

262. Crowther CA, Hiller JE, Moss JR, et al. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *N Engl J Med* 2005; **352**(24): 2477-86.

263. Landon MB, Spong CY, Thom E, et al. A multicenter, randomized trial of treatment for mild gestational diabetes. *N Engl J Med* 2009; **361**(14): 1339-48.

264. Royal College of Obstetricians & Gynaecologists. Shoulder Dystocia (Green-top Guideline No. 42). 2012. https://www.rcog.org.uk/en/guidelines-research-services/guidelines/gtg42/ (accessed 28 December 2020).

265. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008; **337**: a1655.

266. National Perinatal Epidemiology Unit. Perinatal Mortality Review Tool. 2017. https://www.npeu.ox.ac.uk/pmrt (accessed 28 December 2020).

267. Robertson L, Knight H, Prosser Snelling E, et al. Each baby counts: National quality improvement programme to reduce intrapartum-related deaths and brain injuries in term babies. *Semin Fetal Neonatal Med* 2017; **22**(3): 193-8.

268. NHS England. Saving Babies' Lives Version Two. 2019. https://www.england.nhs.uk/wp-content/uploads/2019/07/saving-babies-lives-care-bundle-version-two-v5.pdf (accessed 25 June 2021).

269. Widdows K, Roberts SA, Camacho EM, Heazell AEP. Stillbirth rates, service outcomes and costs of implementing NHS England's Saving Babies' Lives care bundle in maternity units in England: A cohort study. *PLoS One* 2021; **16**(4): e0250150.

270. Perinatal Institute. National Uptake of GAP. https://perinatal.org.uk/Gap/Uptake (accessed 28 May 2021).

271. RCOG. Small-for-Gestational-Age Fetus, Investigation and Management. Green-top Guideline No. 31 (2nd edn). Royal College of Obstetricians and Gynaecologists (RCOG) Press: London, 2013.; 2013.

272. Perinatal Institute. Growth Assessment Protocol (GAP): Outline Specification - 2016/17. 2016. https://www.perinatal.org.uk/FetalGrowth/PDFs/GAPplus_outline_specification_new_users_2016_17_V3.pdf (accessed 21 January 2020).

273. Perinatal Institute. Growth Assessment Protocol. 2019. https://www.perinatal.org.uk/GAP/Programme (accessed 18 June 2020).

274. Perinatal Institute. GROW-App UK. https://app.growservice.org/uk/ (accessed 14 June 2019).

275. Perinatal Institute. Fetal growth examples. http://www.perinatal.org.uk/FetalGrowth/Examples.aspx (accessed 14 June 2019).
276. Vieira MC, Relph S, Copas A, et al. The DESiGN trial (DEtection of Small for Gestational age Neonate), evaluating the effect of the Growth Assessment Protocol (GAP): study protocol for a randomised controlled trial. *Trials* 2019; **20**(1): 154.
277. Office for National Statistics. Characteristics of birth 1. England & Wales, 2013. 2014. http://www.ons.gov.uk/ons/rel/vsob1/characteristics-of-birth-1--england-and-wales/2013/index.html (accessed 28 January 2017).
278. Manktelow B, Smith L, Evans T, et al. Perinatal Mortality Surveillance Report UK Perinatal Deaths for births from January to December 2013 Leicester: The Infant Mortality and Morbidity Group, University of Leicester., 2015.
279. Iliodromiti S, Smith GCS, Lawlor DA, Pell JP, Nelson SM. UK stillbirth trends in over 11 million births provide no evidence to support effectiveness of Growth Assessment Protocol program. *Ultrasound Obstet Gynecol* 2020; **55**(5): 599-604.
280. Gardosi J, Turner S, Williams M, Buller S, Hugh O, Francis A. The Growth Assessment Protocol: a major cause of declining stillbirth rates in the UK. *Ultrasound Obstet Gynecol* 2020; **56**(1): 117-9.
281. Hugh O, Williams M, Turner S, Gardosi J. Reduction of stillbirths in England from 2008 to 2017 according to uptake of the Growth Assessment Protocol: 10-year population-based cohort study. *Ultrasound Obstet Gynecol* 2021; **57**(3): 401-8.
282. Jayawardena L, Sheehan P. Introduction of a customised growth chart protocol increased detection of small for gestational age pregnancies in a tertiary Melbourne hospital. *Aust N Z J Obstet Gynaecol* 2019; **59**(4): 493-500.
283. Cowan FJ, McKinlay CJD, Taylor RS, et al. Detection of small for gestational age babies and perinatal outcomes following implementation of the Growth Assessment Protocol at a New Zealand tertiary facility: An observational intervention study. *Aust N Z J Obstet Gynaecol* 2021; **61**(3): 339-46.
284. Department of Health. Reference costs 2015-16. 2016. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/577083/Reference_Costs_2015-16.pdf (accessed 20 December 2018).
285. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical care* 2012; **50**(3): 217-26.
286. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *J Clin Epidemiol* 2009; **62**: 499-505.
287. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *Bmj* 2015; **350**: h2147.
288. Gardosi J, Francis A. Adverse pregnancy outcome and association with small for gestational age birthweight by customized and population-based percentiles. *Am J Obstet Gynecol* 2009; **201**(1): 28 e1-8.
289. Hepburn M, Rosenberg K. An audit of the detection and management of small-for-gestational age babies. *Br J Obstet Gynaecol* 1986; **93**(3): 212-6.
290. Laios G, Haddad S, Tedesco R, Passini R, Dias T, Nomura M. Intracluster correlation coefficients for the Brazilian Multicenter Study on Preterm Birth (EMIP): methodological and practical implications. *BMC medical research methodology* 2014; (14): 54.
291. Government Digital Service. National Statistics Postcode Lookup. 2017. https://data.gov.uk/dataset/7709b64e-369f-41f4-96ce-1f05efde9834/national-statistics-postcode-lookup-august-2017 (accessed 17 November 2017).
292. NHS Digital. NHSmail. https://digital.nhs.uk/services/nhsmail (accessed 16 March 2020).
293. National Maternity and Perinatal Audit. Clinical Report 2019: Based on births in NHS maternity services between 1 April 2016 and 31 March 2017. 2019. https://maternityaudit.org.uk/FilesUploaded/NMPA%20Clinical%20Report%202019.pdf (accessed 10 July 2020).
294. Royston P, White I. Multiple Imputation by Chained Equations (MICE): Implementation inStata. *Journal of Statistical Software* 2011; **45**(4): 20.
295. Hayes R, Moulton L. Cluster Randomized Trials. Abingdon, UK: Taylor & Francis; 2009.
296. Kirkwood B, Sterne J. Essential Medical Statistics. 2 ed. Oxford, UK: Blackwell Science Ltd; 2003.
297. Rubin D. Multiple imputation for nonresponse in surveys. New York: Wiley & Sons; 1987.
298. Anderson R. New MRC guidance on evaluating complex interventions. *BMJ* 2008; **337**: a1937.
299. Hawe P, Shiell A, Riley T, Gold L. Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *J Epidemiol Community Health* 2004; **58**(9): 788-93.
300. Basch CE, Sliepcevich EM, Gold RS, Duncan DF, Kolbe LJ. Avoiding type III errors in health education program evaluations: a case study. *Health Educ Q* 1985; **12**(4): 315-31.
301. Steckler A, Linnan L. Process Evaluation for Public Health Interventions and Research. San Francisco: J Wiley; 2002.
302. Bryce J, Amouzou A, Hazel E, et al. Measuring the strength of implementation of community case management of childhood illness within the Catalytic Initiative to Save a Million Lives. 2011. https://www.jhsph.edu/research/centers-and-institutes/institute-for-international-programs/_documents/rapid_scaleup/wp-implementation-strength.pdf (accessed 04 May 2020).

303. Pfadenhauer LM, Gerhardus A, Mozygemba K, et al. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. *Implement Sci* 2017; **12**(1): 21.

304. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci* 2009; **4**(1): 50.

305. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health* 2011; **38**(2): 65-76.

306. Karsh BT. Beyond usability: designing effective technology implementation systems to promote patient safety. *Qual Saf Health Care* 2004; **13**(5): 388-94.

307. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol* 2013; **13**(1): 117.

308. Public Health England. Guidance: Economic evaluation. 07 August 2018. https://www.gov.uk/government/publications/evaluation-in-health-and-well-being-overview/economic-evaluation (accessed 17 July 2020).

309. National Institute for Health and Care Excellence. NICE Technology Appraisal Guidance. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance.

310. NHS Improvement. National tariff payment system 2017/18 and 2018/19. 2018. https://improvement.nhs.uk/resources/national-tariff-1719/ (accessed 05 December 2018.

311. Relph S, Delaney L, Melaugh A, et al. Costing the impact of interventions during pregnancy in the UK: a systematic review of economic evaluations. *Bmj Open* 2020; **10**(10): e040022.

312. Department of Health. Pay & Price Series 2015/16. 2016. www.info.doh.gov.uk (accessed 11 December 2018).

313. NHS Improvement. Economic Assumptions for 2016/17 to 2020/21. 2016. https://www.gov.uk/government/publications/economic-assumptions-201617-to-202021/economic-assumptions-201617-to-202021 (accessed 11 December 2018).

314. Pinnock H, Barwick M, Carpenter CR, et al. Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ* 2017; **356**: i6795.

315. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; **339**: b2535.

316. Husereau D, Drummond M, Petrou S, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS)--explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health* 2013; **16**(2): 231-50.

317. von Elm E, Altman D, Egger M, Pocock S, Gotzsche P, Vandenbroucke J. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *The Lancet* 2007; **370**(9596): 1453-7.

318. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010; **340**: c332.

319. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012; **345**: e5661.

320. Vieira M, Relph S, Muruet-Gutierrez W, et al. Effect of the Growth Assessment Protocol (GAP) on the detection of small for gestational age: the DESiGN cluster randomised trial. Unpublished; 2021.

321. Gardosi J, Giddings S, Clifford S, Wood L, Francis A. Association between reduced stillbirth rates in England and regional uptake of accreditation training in customised fetal growth assessment. *Bmj Open* 2013; **3**(12): e003942.

322. Selvaratnam RJ, Davey MA, Anil S, McDonald SJ, Farrell T, Wallace EM. Does public reporting of the detection of fetal growth restriction improve clinical outcomes: a retrospective cohort study. *BJOG* 2020; **127**(5): 581-9.

323. Hebert R, Veil A. Monitoring the degree of implementation of an integrated delivery system. *Int J Integr Care* 2004; **4**(3): e05.

324. Gold RB, Singh S, Frost J. The Medicaid eligibility expansions for pregnant women: evaluating the strength of state implementation efforts. *Fam Plann Perspect* 1993; **25**(5): 196-207.

325. Ryman TK, Elsayed EA, Mustafa AA, Widoa NM, Omer A, Kamadjeu R. Implementation of the reaching every district (RED) approach: experience from North Sudan. *East Mediterr Health J* 2011; **17**(11): 804-12.

326. Wilson MG, Basta TB, Bynum BH, DeJoy DM, Vandenberg RJ, Dishman RK. Do intervention fidelity and dose influence outcomes? Results from the move to improve worksite physical activity program. *Health education research* 2010; **25**(2): 294-305.

327. Hargreaves JR, Goodman C, Davey C, Willey BA, Avan BI, Schellenberg JR. Measuring implementation strength: lessons from the evaluation of public health strategies in low- and middle-income settings. *Health Policy Plan* 2016; **31**(7): 860-7.

328. Sarmandal P, Bailey SM, Grant JM. A comparison of three methods of assessing inter-observer variation applied to ultrasonic fetal measurement in the third trimester. *Br J Obstet Gynaecol* 1989; **96**(11): 1261-5.

329. Tougher S, Group AC, Ye Y, et al. Effect of the Affordable Medicines Facility--malaria (AMFm) on the availability, price, and market share of quality-assured artemisinin-based combination therapies in seven countries: a before-and-after analysis of outlet survey data. *Lancet* 2012; **380**(9857): 1916-26.

330. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant* 2018; **52**(4): 1893-907.
331. Draper ES, Gallimore ID, Kurinczuk JJ, et al. MBRRACE-UK Perinatal Mortality Surveillance Report, UK Perinatal Deaths for Births from January to December 2016. . Leicester: Departement of Health Sciences, University of Leicester, 2018.
332. Department of Health. New ambition to halve rate of stillbirths and infant deaths. 2015. https://www.gov.uk/government/news/new-ambition-to-halve-rate-of-stillbirths-and-infant-deaths (accessed 09 January 2018).
333. NHS Litigation Authority. Clinical Negligence Scheme for Trusts: Maternity Clinical Risk Management Standards. Version 1. London: NHS Litigation Authority, 2013.
334. Hawe P, Shiell A, Riley T. Theorising Interventions as Events in Systems. *American Journal of Community Psychology* 2009; **43**(3): 267-76.
335. May C. Towards a general theory of implementation. *Implement Sci* 2013; **8**: 18.
336. Widdows K, Reid HE, Roberts SA, Camacho EM, Heazell AEP. Saving babies' lives project impact and results evaluation (SPiRE): a mixed methodology study. *BMC Pregnancy Childbirth* 2018; **18**(1): 43.
337. Lau YZ, Widdows K, Roberts SA, et al. Assessment of the quality, content and perceived utility of local maternity guidelines in hospitals in England implementing the saving babies' lives care bundle to reduce stillbirth. *BMJ Open Qual* 2020; **9**(2): e000756.
338. Vaona A, Banzi R, Kwag KH, et al. E-learning for health professionals. *Cochrane Database of Systematic Reviews* 2018; (1).
339. Pearce JM, Campbell S. A comparison of symphysis-fundal height and ultrasound as screening tests for light-for-gestational age infants. *Br J Obstet Gynaecol* 1987; **94**(2): 100-4.
340. Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. *Trials* 2013; **14**: 15.
341. Garcia J, Bricker L, Henderson J, et al. Women's views of pregnancy ultrasound: a systematic review. *Birth* 2002; **29**(4): 225-50.
342. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *BMJ* 2006; **333**(7563): 346-9.
343. Dixon-Woods M, Baker R, Charles K, et al. Culture and behaviour in the English National Health Service: overview of lessons from a large multimethod study. *BMJ Qual Saf* 2014; **23**(2): 106-15.
344. Centre for Workforce Intelligence. Securing the future workforce supply: Sonography workforce review. 2017. https://www.bmus.org/static/uploads/resources/Sonography_workforce_review.pdf (accessed 14 July 2020).
345. Dunn P, McKenna H, Murray R. Deficits in the NHS 2016, 2016.
346. Schellenberg J, Bobrova N, Avan B. Measuring implementation strength: Literature review draft report 2012., 2012.
347. El Alili M, van Dongen JM, Huirne JAF, van Tulder MW, Bosmans JE. Reporting and Analysis of Trial-Based Cost-Effectiveness Evaluations in Obstetrics and Gynaecology. *Pharmacoeconomics* 2017; **35**(10): 1007-33.
348. Independent Hospital Pricing Authority. Bundled pricing for maternity care: Final Report of IHPA and the Bundled Pricing Advisory Group. 2017. https://www.ihpa.gov.au/sites/default/files/bundled_pricing_for_maternity_care_-_final_report.docx (accessed 27 July 2020).
349. Medicaid and CHIP Payment and Access Commission. Medicaid Payment Initiatives to Improve Maternal and Birth Outcomes. 2019. https://www.macpac.gov/publication/medicaid-payment-initiatives-to-improve-maternal-and-birth-outcomes/ (accessed 27 July 2020).
350. NHS England, NHS Improvement. Guidance on the market forces factor: A supporting document for the 2017 to 2019 National Tariff Payment System, 2016.
351. The Commonwealth Fund. Bundled-payment models around the world: how they work and what their impact has been. 2020. https://www.commonwealthfund.org/publications/2020/apr/bundled-payment-models-around-world-how-they-work-their-impact (accessed 27 July 2020).
352. Fahy M, Doyle O, Denny K, McAuliffe FM, Robson M. Economics of childbirth. *Acta Obstet Gynecol Scand* 2013; **92**(5): 508-16.
353. Commonwealth Fund. Mirror, Mirror 2017: International Comparison Reflects Flaws and Opportunities for Better U.S. Health Care. 2017. https://www.commonwealthfund.org/publications/fund-reports/2017/jul/mirror-mirror-2017-international-comparison-reflects-flaws-and?redirect_source=/ (accessed 06 December 2018).
354. Ofman JJ, Sullivan SD, Neumann PJ, et al. Examining the value and quality of health economic analyses: implications of utilizing the QHES. *J Manag Care Pharm* 2003; **9**(1): 53-61.
355. National Institute for Health and Care Excellence. Intrapartum Care for healthy women and babies. Clinical Guideline CG 190. , 2014.
356. National Institute for Health and Care Excellence. Hypertension in pregnancy: diagnosis and management. 2011: 1-51.
357. National Institute for Health and Care Excellence. Diabetes in pregnancy: management from preconception to the postnatal period, 2015.
358. Mistry H, Heazell AE, Vincent O, Roberts T. A structured review and exploration of the healthcare costs associated with stillbirth and a subsequent pregnancy in England and Wales. *BMC Pregnancy Childbirth* 2013; **13**: 236.

359. Deshpande SN, van Asselt AD, Tomini F, et al. Rapid fetal fibronectin testing to predict preterm birth in women with symptoms of premature labour: a systematic review and cost analysis. *Health Technol Assess* 2013; **17**(40): 1-138.
360. Thomas CM, Cameron S. Can we reduce costs and prevent more unintended pregnancies? A cost of illness and cost-effectiveness study comparing two methods of EHC. *Bmj Open* 2013; **3**(12): e003815.
361. O'Donnell A, McParlin C, Robson SC, et al. Treatments for hyperemesis gravidarum and nausea and vomiting in pregnancy: a systematic review and economic assessment. *Health Technol Assess* 2016; **20**(74): 1-268.
362. Alfirevic Z, Keeney E, Dowswell T, et al. Which method is best for the induction of labour? A systematic review, network meta-analysis and cost-effectiveness analysis. *Health Technol Assess* 2016; **20**(65): 1-584.
363. Farrar D, Simmonds M, Griffin S, et al. The identification and treatment of women with hyperglycaemia in pregnancy: an analysis of individual participant data, systematic reviews, meta-analyses and an economic evaluation. *Health Technol Assess* 2016; **20**(86): 1-348.
364. Gallos I, Williams H, Price M, et al. Uterotonic drugs to prevent postpartum haemorrhage: a network meta-analysis. *Health Technol Assess* 2019; **23**(9): 1-356.
365. Petrou S, Taher S, Abangma G, Eddama O, Bennett P. Cost-effectiveness analysis of prostaglandin E2 gel for the induction of labour at term. *BJOG* 2011; **118**(6): 726-34.
366. Eddama O, Petrou S, Regier D, et al. Study of progesterone for the prevention of preterm birth in twins (STOPPIT): findings from a trial-based cost-effectiveness analysis. *Int J Technol Assess Health Care* 2010; **26**(2): 141-8.
367. Jit M, Cromer D, Baguelin M, Stowe J, Andrews N, Miller E. The cost-effectiveness of vaccinating pregnant women against seasonal influenza in England and Wales. *Vaccine* 2010; **29**(1): 115-22.
368. Round JA, Jacklin P, Fraser RB, Hughes RG, Mugglestone MA, Holt RI. Screening for gestational diabetes mellitus: cost-utility of different screening strategies based on a woman's individual risk of disease. *Diabetologia* 2011; **54**(2): 256-63.
369. Schroeder E, Petrou S, Patel N, et al. Cost effectiveness of alternative planned places of birth in woman at low risk of complications: evidence from the Birthplace in England national prospective cohort study. *BMJ* 2012; **344**: e2292.
370. Essex HN, Parrott S, Wu Q, Li J, Cooper S, Coleman T. Cost-Effectiveness of Nicotine Patches for Smoking Cessation in Pregnancy: A Placebo Randomized Controlled Trial (SNAP). *Nicotine Tob Res* 2015; **17**(6): 636-42.
371. Coomarasamy A, Williams H, Truchanowicz E, et al. PROMISE: first-trimester progesterone therapy in women with a history of unexplained recurrent miscarriages - a randomised, double-blind, placebo-controlled, international multicentre trial and economic evaluation. *Health Technol Assess* 2016; **20**(41): 1-92.
372. Carolan-Rees G, Ray AF. The ScanTrainer obstetrics and gynaecology ultrasound virtual reality training simulator: A cost model to determine the cost viability of replacing clinical training with simulation training. *Ultrasound* 2015; **23**(2): 110-5.
373. Lain SJ, Roberts CL, Bond DM, Smith J, Morris JM. An economic evaluation of planned immediate versus delayed birth for preterm prelabour rupture of membranes: findings from the PPROMT randomised controlled trial. *BJOG* 2017; **124**(4): 623-30.
374. Parisaei M, Currie J, O'Gorman N, Morris S, David AL. Implementation of foetal fibronectin testing: Admissions, maternal interventions and costs at 1 year. *J Obstet Gynaecol* 2016; **36**(7): 888-92.
375. Ussher M, Lewis S, Aveyard P, et al. The London Exercise And Pregnant smokers (LEAP) trial: a randomised controlled trial of physical activity for smoking cessation in pregnancy with an economic evaluation. *Health Technol Assess* 2015; **19**(84): vii-xxiv, 1-135.
376. Walker KF, Dritsaki M, Bugg G, et al. Labour induction near term for women aged 35 or over: an economic evaluation. *BJOG* 2017; **124**(6): 929-34.
377. van der Nelson HA, Draycott T, Siassakos D, Yau CWH, Hatswell AJ. Carbetocin versus oxytocin for prevention of post-partum haemorrhage at caesarean section in the United Kingdom: An economic impact analysis. *Eur J Obstet Gynecol Reprod Biol* 2017; **210**: 286-91.
378. Bick D, Briley A, Brocklehurst P, et al. A multicentre, randomised controlled trial of position during the late stages of labour in nulliparous women with an epidural: clinical effectiveness and an economic evaluation (BUMPES). *Health Technol Assess* 2017; **21**(65): 1-176.
379. Campbell HE, Kurinczuk JJ, Heazell A, Leal J, Rivero-Arias O. Healthcare and wider societal implications of stillbirth: a population-based cost-of-illness study. *BJOG* 2018; **125**(2): 108-17.
380. Duckworth S, Chappell LC, Seed PT, Mackillop L, Shennan AH, Hunter R. Placental Growth Factor (PlGF) in Women with Suspected Pre-Eclampsia Prior to 35 Weeks' Gestation: A Budget Impact Analysis. *PLoS One* 2016; **11**(10): e0164276.
381. Orlovic M, Carter AW, Marti J, Mossialos E. Estimating the incidence and the economic burden of third and fourth-degree obstetric tears in the English NHS: an observational study using propensity score matching. *Bmj Open* 2017; **7**(6): e015463.
382. Vatish M, Strunz-McKendry T, Hund M, Allegranza D, Wolf C, Smare C. sFlt-1/PlGF ratio test for pre-eclampsia: an economic assessment for the UK. *Ultrasound Obstet Gynecol* 2016; **48**(6): 765-71.
383. Bowers J, Cheyne H. Reducing the length of postnatal hospital stay: implications for cost and quality of care. *BMC Health Serv Res* 2016; **16**: 16.

384. Luni Y, Borakati A, Matah A, Skeats K, Eedarapalli P. A prospective cohort study evaluating the cost-effectiveness of carbetocin for prevention of postpartum haemorrhage in caesarean sections. *J Obstet Gynaecol* 2017; **37**(5): 601-4.

385. Khan KS, Moore P, Wilson M, et al. A randomised controlled trial and economic evaluation of intraoperative cell salvage during caesarean section in women at risk of haemorrhage: the SALVO (cell SALVage in Obstetrics) trial. *Health Technol Assess* 2018; **22**(2): 1-88.

386. Waugh J, Hooper R, Lamb E, et al. Spot protein-creatinine ratio and spot albumin-creatinine ratio in the assessment of pre-eclampsia: a diagnostic accuracy study with decision-analytic model-based economic evaluation and acceptability analysis. *Health Technol Assess* 2017; **21**(61): 1-90.

387. Jones M, Smith M, Lewis S, Parrott S, Coleman T. A dynamic, modifiable model for estimating cost-effectiveness of smoking cessation interventions in pregnancy: application to an RCT of self-help delivered by text message. *Addiction* 2019; **114**(2): 353-65.

388. Jacklin PB, Maresh MJ, Patterson CC, et al. A cost-effectiveness comparison of the NICE 2015 and WHO 2013 diagnostic criteria for women with gestational diabetes with and without risk factors. *Bmj Open* 2017; **7**(8): e016621.

389. Xydopoulos G, Perry H, Sheehan E, Thilaganathan B, Fordham R, Khalil A. Home blood-pressure monitoring in a hypertensive pregnant population: cost-minimization study. *Ultrasound Obstet Gynecol* 2019; **53**(4): 496-502.

390. Wastlund D, Moraitis AA, Thornton JG, et al. The cost-effectiveness of universal late-pregnancy screening for macrosomia in nulliparous women: a decision analysis. *BJOG* 2019; **126**(10): 1243-50.

391. Wastlund D, Moraitis AA, Dacey A, Sovio U, Wilson ECF, Smith GCS. Screening for breech presentation using universal late-pregnancy ultrasonography: A prospective cohort study and cost effectiveness analysis. *PLoS Med* 2019; **16**(4): e1002778.

392. National Institute for Health and Care Excellence. Inducing Labour. 2008. https://www.nice.org.uk/guidance/cg70/resources/inducing-labour-pdf-975621704389.

393. National Institute for Health and Clinical Excellence. Antenatal Care for uncomplicated pregnancies. NICE Clinical guidelines, 2008.

394. NHS England, NHS Improvement. Market forces factor review and proposed updates: NHS, 2018.

395. National Institute for Health and Care Excellence. How NICE measures value for money in relation to public health interventions. 1 September 2013 2013. https://www.nice.org.uk/Media/Default/guidance/LGB10-Briefing-20150126.pdf (accessed 28 November 2020).

396. Relph S, Coxon K, Vieira M, et al. Effect of the Growth Assessment Protocol on the DEtection of the Small for GestatioNal Age Fetus: Process evaluation from the DESiGN cluster randomised trial. Unpublished; 2021.

397. Petrou S. Health economic aspects of late preterm and early term birth. *Semin Fetal Neonatal Med* 2019; **24**(1): 18-26.

398. Petrou S. The economic consequences of preterm birth during the first 10 years of life. *BJOG* 2005; **112 Suppl 1**(s1): 10-5.

399. Relph S, Elstad M, Coker B, et al. Using electronic patient records to assess the effect of a complex antenatal intervention in a cluster randomised controlled trial-data management experience from the DESiGN Trial team. *Trials* 2021; **22**(1): 195.

400. The Personal Social Services Research Unit. Hospital-based Health Care Staff. 2018. https://www.pssru.ac.uk/pub/uc/uc2018/hospital-based-health-care-staff.pdf (accessed 27 June 2021).

401. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002; **360**(9334): 711-5.

402. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 4 April 2013 2013. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781 (accessed 28 November 2020).

403. National Institute for Health and Care Excellence. Intrapartum care for women with existing medical conditions or obstetric complications and their babies. Supplement 2: Health economics. 2019. https://www.nice.org.uk/guidance/ng121/evidence/supplement-2-health-economics-pdf-6773941838 (accessed 28 November 2020).

404. Gardosi J, Madurasinghe V, Williams M, Malik A, Francis A. Maternal and fetal risk factors for stillbirth: population based study. *BMJ* 2013; **346**: f108.

405. Ego A, Monier I, Skaare K, Zeitlin J. Antenatal detection of fetal growth restriction and risk of stillbirth: population-based case-control study. *Ultrasound Obstet Gynecol* 2020; **55**(5): 613-20.

406. Selvaratnam RJ, Wallace EM, Treleaven S, Hooper SB, Davis PG, Davey MA. Does detection of fetal growth restriction improve neonatal outcomes? *J Paediatr Child Health* 2021; **57**(5): 677-83.

407. Every Woman Every Child: The Global Strategy for Women's, Children's and Adolescents' Health (2016-2030). 2015. https://www.who.int/life-course/partners/global-strategy/ewec-globalstrategyreport-200915.pdf?ua=1 (accessed 25 February 2021).

408. Gardosi J, Kady SM, McGeown P, Francis A, Tonks A. Classification of stillbirth by relevant condition at death (ReCoDe): population based cohort study. *BMJ* 2005; **331**(7525): 1113-7.

409. Flenady V, Middleton P, Smith GC, et al. Stillbirths: the way forward in high-income countries. *Lancet* 2011; **377**(9778): 1703-17.

410. Andreasen LA, Tabor A, Norgaard LN, et al. Why we succeed and fail in detecting fetal growth restriction: A population-based study. *Acta Obstet Gynecol Scand* 2021; **100**(5): 893-9.

411. Diksha P, Permezel M, Pritchard N. Why we miss fetal growth restriction: Identification of risk factors for severely growth-restricted fetuses remaining undelivered by 40 weeks gestation. *Aust N Z J Obstet Gynaecol* 2018; **58**(6): 674-80.
412. Avci ME, Sanlikan F, Celik M, Avci A, Kocaer M, Gocmen A. Effects of maternal obesity on antenatal, perinatal and neonatal outcomes. *J Matern Fetal Neonatal Med* 2015; **28**(17): 2080-3.
413. Mongelli M, Gardosi J. Symphysis-fundus height and pregnancy characteristics in ultrasound-dated pregnancies. *Obstet Gynecol* 1999; **94**(4): 591-4.
414. Preyer O, Husslein H, Concin N, et al. Fetal weight estimation at term - ultrasound versus clinical examination with Leopold's manoeuvres: a prospective blinded observational study. *BMC Pregnancy Childbirth* 2019; **19**(1): 122.
415. Fox NS, Bhavsar V, Saltzman DH, Rebarber A, Chasen ST. Influence of maternal body mass index on the clinical estimation of fetal weight in term pregnancies. *Obstet Gynecol* 2009; **113**(3): 641-5.
416. Aksoy H, Aksoy U, Karadag OI, Yucel B, Aydin T, Babayigit MA. Influence of maternal body mass index on sonographic fetal weight estimation prior to scheduled delivery. *J Obstet Gynaecol Res* 2015; **41**(10): 1556-61.
417. MacDonald TM, Hui L, Tong S, et al. Reduced growth velocity across the third trimester is associated with placental insufficiency in fetuses born at a normal birthweight: a prospective cohort study. *BMC Med* 2017; **15**(1): 164.
418. Stubert J, Peschel A, Bolz M, Glass A, Gerber B. Accuracy of immediate antepartum ultrasound estimated fetal weight and its impact on mode of delivery and outcome - a cohort analysis. *BMC Pregnancy Childbirth* 2018; **18**(1): 118.
419. Sharma PP, Salihu HM, Kirby RS. Stillbirth recurrence in a population of relatively low-risk mothers. *Paediatr Perinat Epidemiol* 2007; **21 Suppl 1**: 24-30.
420. Office for National Statistics. Birth Characteristics. 2016. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/datasets/birthcharacteristicsinenglandandwales (accessed 14 February 2020).
421. Ananth CV, Peltier MR, Chavez MR, Kirby RS, Getahun D, Vintzileos AM. Recurrence of ischemic placental disease. *Obstet Gynecol* 2007; **110**(1): 128-33.
422. Goetzinger KR, Cahill AG, Macones GA, Odibo AO. Echogenic bowel on second-trimester ultrasonography: evaluating the risk of adverse pregnancy outcome. *Obstet Gynecol* 2011; **117**(6): 1341-8.
423. Harlev A, Levy A, Zaulan Y, et al. Idiopathic bleeding during the second half of pregnancy as a risk factor for adverse perinatal outcome. *J Matern Fetal Neonatal Med* 2008; **21**(5): 331-5.
424. European Monitoring Centre for Drugs and Drug Addiction. Pregnancy, childcare and the family: key issues for Europe's response to drugs. Luxembourg: Publications Office of the European Union; 2012.
425. Wiles K, Chappell L, Clark K, et al. Clinical practice guideline on pregnancy and renal disease. *BMC Nephrol* 2019; **20**(1): 401.
426. Relph S, NMPA Project Team. NHS Maternity Care for Women with a Body Mass Index of 30 kg/m2 or Above: Births between 1 April 2015 and 31 March 2017 in England, Wales and Scotland. London: RCOG, 2021.
427. Francis A, Tonks A, Gardosi J. Accuracy of ultrasound estimation of fetal weight at term. *Archives of Disease in Childhood - Fetal and Neonatal Edition* 2011; **96**(Supplement 1): Fa61-Fa.
428. Castro-Vasquez BA, Taboada C. Accuracy of Estimated Fetal Weight in Third Trimester [33A]. *Obstetrics & Gynecology* 2020; **135**.
429. Thilaganathan B. Ultrasound fetal weight estimation at term may do more harm than good. *Ultrasound Obstet Gynecol* 2018; **52**(1): 5-8.
430. Scioscia M, Vimercati A, Ceci O, Vicino M, Selvaggi LE. Estimation of birth weight by two-dimensional ultrasonography: a critical appraisal of its accuracy. *Obstet Gynecol* 2008; **111**(1): 57-65.
431. Zafman KB, Bergh E, Fox NS. Accuracy of sonographic estimated fetal weight in suspected macrosomia: the likelihood of overestimating and underestimating the true birthweight. *J Matern Fetal Neonatal Med* 2020; **33**(6): 967-72.
432. Reid EW, McNeill JA, Holmes VA, Alderdice FA. Women's perceptions and experiences of fetal macrosomia. *Midwifery* 2014; **30**(4): 456-63.
433. Melamed N, Yogev Y, Meizner I, Mashiach R, Ben-Haroush A. Sonographic prediction of fetal macrosomia: the consequences of false diagnosis. *J Ultrasound Med* 2010; **29**(2): 225-30.
434. Malin GL, Bugg GJ, Takwoingi Y, Thornton JG, Jones NW. Antenatal magnetic resonance imaging versus ultrasound for predicting neonatal macrosomia: a systematic review and meta-analysis. *BJOG* 2016; **123**(1): 77-88.
435. Murphy DJ, Strachan BK, Bahl R, Royal College of O, Gynaecologists. Assisted Vaginal Birth: Green-top Guideline No. 26. *BJOG* 2020; **127**(9): e70-e112.
436. Warwick Clinical Trials Unit. Induction of labour for predicted macrosomia - The 'Big Baby Trial'. https://warwick.ac.uk/fac/sci/med/research/ctu/trials/bigbaby/ (accessed 08 March 2021).
437. Lau YZ, Widdows K, Roberts SA, et al. Assessment of the quality, content and perceived utility of local maternity guidelines in hospitals in England implementing the saving babies' lives care bundle to reduce stillbirth. *BMJ Open Qual* 2020; **9**(2).
438. Norman JE, Heazell AEP, Rodriguez A, et al. Awareness of fetal movements and care package to reduce fetal mortality (AFFIRM): a stepped wedge, cluster-randomised trial. *Lancet* 2018; **392**(10158): 1629-38.
439. Royal College of Obstetricians and Gynaecologists. Each Baby Counts: 2020 Final Progress Report. London: RCOG, 2021.

440. Bjornholt SM, Leite M, Albieri V, Kjaer SK, Jensen A. Maternal smoking during pregnancy and risk of stillbirth: results from a nationwide Danish register-based cohort study. *Acta Obstet Gynecol Scand* 2016; **95**(11): 1305-12.
441. UK Government. The Smoke-free (Premises and Enforcement) Regulations. 2006. https://www.legislation.gov.uk/uksi/2006/3368/contents/made (accessed 15 May 2021).
442. Kilby MD, Gibson JL, Ville Y. Falling perinatal mortality in twins in the UK: organisational success or chance? *BJOG* 2019; **126**(3): 341-7.
443. Po G, Oliver EA, Reddy UM, Silver RM, Berghella V. The impact of induction of labor at 39 weeks in low-risk women on the incidence of stillbirth. *Am J Obstet Gynecol* 2020; **222**(1): 88-90.
444. Grobman WA, Rice MM, Reddy UM, et al. Labor Induction versus Expectant Management in Low-Risk Nulliparous Women. *N Engl J Med* 2018; **379**(6): 513-23.

10   APPENDICES

10.1 DESCRIPTION OF THE INTERVENTION USING THE TIDiER GUIDANCE

	Item	Description
	BRIEF NAME	
1.	Provide the name or a phrase that describes the intervention.	The GAP programme is a complex intervention for improved detection of SGA infants through risk stratification, serial fundal height or scans during second and third trimester and use of customized charts for assessment of fetal growth.
	WHY?	
2.	Describe any rationale, theory, or goal of the elements essential to the intervention.	Designed to improve antenatal detection of the small-for-gestational-age neonate thereby reducing stillbirth related to fetal growth restriction. This intervention has been proposed given the context of the UK having a stillbirth rate that ranks poorly compared to other similar health economies.
	WHAT?	
3.	Materials: Describe any physical or informational materials used in the intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed (e.g. online appendix, URL).	• Staff training materials • E-learning module • Evidence-based risk assessment and management protocols • GAP software for generation of customised growth charts and calculation of birthweight centiles • Tools to audit missed cases • Perinatal Institute support • Materials generated by the research sites in disseminating the intervention to staff and patients.
4.	Procedures: Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities.	• Baseline audit on rates of antenatal detection of SGA neonates • Train the Trainers course for senior staff members • A minimum of 75% staff in each staff group at each site to receive the face-to-face and e-learning training on the intervention. • Risk assessment of each woman at antenatal booking and throughout pregnancy • Generation of customised growth charts • Protocols for referral for additional fetal growth scans • Calculation of customised fetal weight centiles at birth • Missed case audit and review tool
	WHO PROVIDED?	
5.	For each category of intervention provider (e.g. psychologist, nursing assistant), describe their expertise,	• The following staff receive the same e-learning and tailored face-to-face training (tailored to assist in the component of the intervention which they will provide): o Health care assistants

	background and any specific training given.	o Antenatal and intrapartum midwives o Antenatal sonographers o Obstetric consultants, registrars, training-grade SHOs (including those in Foundation Year or GP training), GPs.
	HOW?	
6.	Describe the modes of delivery (e.g. face-to-face or by some other mechanism, such as internet or telephone) of the intervention and whether it was provided individually or in a group.	• Training of staff provided face-to-face in group settings (either by Perinatal Institute or by hospital trainers) or individually (by hospital trainers). • Training of staff includes individual e-learning online training. • Risk assessment of women at booking and ongoing assessment of fetal growth throughout pregnancy on an individual basis in a face-to-face scenario.
	WHERE?	
7.	Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features.	• Hospital antenatal clinics • GP clinics • Community antenatal clinics • Birthing centres (obstetric/midwife-led). Necessary infrastructure: access to the website on a computer, printing facilities.
	WHEN and HOW MUCH?	
8.	Describe the number of times the intervention was delivered and over what period of time including the number of sessions, their schedule, and their duration, intensity or dose.	• Training delivered (both face-to-face and e-learning) once per year to a minimum of 75% of staff from each healthcare group (midwives, doctors, sonographers). • Risk assessment for each woman at antenatal booking and at all later antenatal appointments • For low risk women, symphyseal-fundal height measured and plotted onto customized chart every 2-3 weeks from 26-28 weeks until the end of pregnancy. • For high risk women, fetal growth ultrasound every 3 weeks from 26-28 weeks until the end of pregnancy. • For all women, single assessment of neonatal birthweight centile at birth.
	TAILORING?	
9.	If the intervention was planned to be personalised, titrated or adapted, then describe what, why, when, and how.	The intervention to be tailored for women according to risk assessment (for risk of having a small-for-gestational age neonate) performed at antenatal booking. No adaptations to the intervention are recommended for the trial.

	MODIFICATIONS?	
10.	If the intervention was modified during the course of the study, describe the changes (what, why, when, and how).	No modifications to the intervention were recommended but it is recognised that individual cluster sites may modify the intervention for tailored application at their own sites.
	HOW WELL?	
11.	Planned: If intervention adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them.	Recommended investigation of missed cases of SGA.
12.‡	Actual: If intervention adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned.	Further assessment of intervention adherence or fidelity is the subject of this study.

10.2 SUMMARY OF STATEMENTS FROM LOCAL GUIDELINES IN STANDARD CARE SITES, COMPARED TO STATEMENTS FROM THE KEY COMPONENTS OF THE GAP GUIDELINE ON SCREENING FOR SGA.

	Growth assessment protocol	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Low Risk - definition	No known risk factors.	No guidance	No guidance	No guidance	No guidance	Same	No guidance
Low Risk - management	Standardised serial fundal height measurements plotted on a customised growth chart. Not suitable if large fibroids, BMI>35, multiple pregnancy.	No guidance	No guidance	Fundal height plots start at 24w and are plotted onto a population chart. Low risk women with abnormal uterine artery Dopplers (mean PI>1.25 or sum>2.5) are referred for growth scans at 28w and 36w	Fundal height monitoring to be performed from 25w in nulliparous women and from 28w in multiparous women. If >2-3cm (varies between local guidelines) below gestational age in weeks then should be referred for a growth scan.	Same fundal height frequency. Population not customised charts	At each antenatal visit the fundal height should be measured in cm. If there is a discrepancy of +/-3cm or more for the expected gestational age, the woman should be offered a growth scan (within 1 week).

	Growth assessment protocol	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
High risk - definition	(One or more of the following) 1. Maternal risk factors Maternal age > 40y Smoker (any) Drug misuse 2. Previous Pregnancy History Previous SGA baby (<10th customised centile) Previous Stillbirth 3. Maternal Medical History Chronic Hypertension Diabetes Renal Impairment Antiphospholipid syndrome 4. Unsuitable for monitoring by fundal height Large fibroids BMI > 35 5. Current pregnancy complications (Early pregnancy) PAPP-A < 0.415 MoM Fetal echogenic bowel Multiple pregnancy (Late pregnancy) Severe pregnancy induced hypertension or pre-eclampsia Unexplained antepartum haemorrhage Concerns related to growth measurements, as listed above.	Women with the following are considered increased risk: Chronic hypertension Pre-existing diabetes, Pre-eclampsia, BMI>35 Differences: Women with all PIH are considered high risk, women with PAPP-A<0.3MoM only	Women with pre-existing diabetes considered high risk	Additional risk factors: Age<18y BMI<18 Previous pre-term pre-eclampsia, eclampsia or abruption. Chronic disease: IBD, congenital heart disease, sickle cell disease Alcohol abuse Congenital infection Different risk factors: Considers previous baby with birthweight <2.5kg as SGA, rather than <10th centile Doesn't include: Women with large fibroids PAPP-A Fetal echogenic bowel Severe PIH Unexplained APH	Splits risk factors (major and minor): Different major risk factors: Smoking only if >10/day Cocaine users only Diabetes only if vascular disease present PAPP-A cut off is <0.40MoM Additional factors: Maternal major SGA Also lists minor risk factors: Nulliparity BMI>=30 Maternal substance abuse/smoker Exposure smoker 1-10/day IVF singleton pregnancy Low fruit intake pre-pregnancy Previous pre-eclampsia Pregnancy interval <6 months or >=60 months. Pregnancy induced hypertension.	Also stratify using uterine artery Dopplers. Agrees on high risk for: previous SGA, antiphospholipid and stillbirth but previous SGA defined differently. Additional risk factors: previous severe pre-eclampsia, sickle cell disease, alcohol misuse, epilepsy. Uses different cut off for PAPP-A. Multiple pregnancy .not mentioned	The following cases are considered high risk by both: Previous stillbirth Existing diabetes/hypertension/renal disease/APLS, Multiple pregnancy, Pre-eclampsia BMI>35 Drug misuse Additional women considered high risk: BMI 30-35 Thyroid disease Lupus Sickle cell disease Alcohol abuse 2-vessel cord Nulliparous women Multiparous women with last child>10y ago

345

	Growth assessment protocol	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
High risk - management	Women with risk factors for fetal growth restriction should be referred to a consultant obstetrician or fetal medicine specialist. The consultant-led team will refer for serial assessment (3 weekly until delivery) of fetal weight and umbilical Doppler from 26-28 weeks until delivery. EFWs plotted on customised charts. These women will not require plotting of fundal height measurements while such a serial scanning protocol is being followed.	Growth scans are variable in frequency. No customisation.	Different scan frequency recommended (4-weekly not 3-weekly)	Only offers a maximum of 2 growth scans as standard, and only if UADs are abnormal. If UADs normal, then one scan is offered at 36w for women with risk factors for SGA	Women with one major risk factor should have serial growth scans from 26-28 weeks. Women with three or more minor risk factors requires uterine artery Dopplers 20-24 weeks. If these are abnormal (PI>95th centile) then refer for serial growth scans from 26 weeks. Women in whom measurement of Fundal height is inaccurate (for example: BMI > 35, large fibroids, should be referred for serial assessment of fetal size using ultrasound	4-weekly growth scans for anyone who is high risk	Scan protocols vary according to indication but are not 3-weekly from 26-28 weeks in any instance.
Growth scan requests re abnormal FH measurements	If first fundal height measurement < 10th Centile or Static growth or Slow growth (sequential measurements not following the slope of the curve) Excessive growth/clinical suspicion of polyhydramnios: curve crosses centiles in an upward direction. Note: first plot above 90th centile is not an indication for a growth scan.	No guidance	No guidance	Referrals for additional growth scans - for fundal height<10th centile OR static growth over 4 weeks only.	If 3cms above or below GA, please commence on a growth chart. If fundal height on growth chart is >90th or <10th centile, refer for growth scan. NB. Unclear what fundal height growth chart this refers to	Also includes plots>95th centile. Uses population not customised chart.	If fundal height +/- 3cm or more discrepancy from the gestational age, refer for growth scan.

	Growth assessment protocol	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6
Referral following a growth scan	If low risk: If EFW plots between 10th-90th centiles and follows the centile curve and the liquor volume is normal then the woman will be asked to attend her next antenatal appointment as planned. If EFW does not plot within 10th and 90th centile or if it is not following a centile curve or if there are concerns regarding the liquor volume or umbilical artery Doppler: Management for LGA/SGA as per below				Only considers abnormal EFW on USS to be that which is <3rd centile. Also considers AC<3rd centile to be indicative of SGA.		Different definition of abnormal: HC/AC ratio>95th centile HC/AC ratio crossing centiles Umb a >95th centile MCA PI<5th centile
Management for a LGA fetus / polyhydramnios:	If EFW>90th centile or significantly accelerated/increased growth velocity then refer for glucose tolerance test within 1 week and to diabetic / consultant-led clinic dependent on the result.			Refer all women with fundal height >97th centile for a growth scan			

AC: abdominal circumference; APH: antepartum haemorrhage; APLS: anti-phospholipid syndrome; HC: head circumference; MCA: middle cerebral artery; MoM: multiples of median; PI: pulsatility index; PIH: Pregnancy induced hypertension; 24w: 24 weeks of gestation; UAD: uterine artery Doppler; Umb A: umbilical artery

347

10.3 DATA REQUEST FORM FOR QUANTITATIVE CLINICAL AND HEALTH-ECONOMIC OUTCOME

Data item		Expected EPR source of data			
Information required	Comments	Maternity	Ultrasound	Neonatal	Hospital activity
Hospital ID	Mandatory (to be anonymised)	X	X	X	X
NHS number	Mandatory (to be anonymised)	X	X	X	X
Demography (mother)					
Age	At start of pregnancy	X			
Date of birth	Mandatory (to be anonymised)	X			
Smoking	For 1st, 2nd, 3rd trim and delivery separately	X			
Education	i.e. highest qualification	X			
Postcode	Mandatory	X			
Ethnicity	Mandatory	X	X		
Country of birth	Or other information on ethnicity	X			
Parity	Mandatory	X	X		
Maternal height	Mandatory	X	X		
Maternal weight	Mandatory	X	X		
BMI at booking		X	X		
Previous medical history					
Hypertension		X	?		
Systemic Lupus Erythematosus		X	?		
Antiphospholipid syndrome		X	?		
Pre pregnancy diabetes		X	?		
Previous obstetric history					
Previous gestational diabetes		X			
Previous LGA		X	X		
Previous SGA		X	X		
Primary outcome data					
Neonatal sex	Mandatory	X		X	
Gestation at delivery - weeks	Mandatory	X		X	
Gestation at delivery - days	Mandatory	X		X	
Birthweight (grams)	Mandatory	X		X	
Number of babies	Mandatory	X		X	
Maternal data					
Date of delivery	Mandatory	X			
Onset of labour		X			
Maternal problem first stage		X			

Data item		Expected EPR source of data			
Information required	Comments	Maternity	Ultrasound	Neonatal	Hospital activity
Induction of labour		X			
Mode of birth		X			
Fetal presentation		X			
Length of stay in hospital		?			X
Episiotomy		X			
Estimated blood loss	Or postpartum haemorrhage	X			
Severe perineal trauma (3rd / 4th degree tear)		X			
Breast feeding at discharge		X			
Analgesia and Anaesthesia Epidural		?			
Maternal problems during pregnancy					
Pre-eclampsia		X			
Eclampsia		X			
Gestational Hypertension		X			
Gestational Diabetes		X			
Maternal infection		X			
Fetal congenital abnormality		X	X	X	
Neonatal Clinical					
Head circumference		X		X	
5-min Apgar score		X		X	
Arterial cord pH ± BE		X		X	
Respiratory support in delivery room		X		X	
Parameters related to NICU admission					
Length of stay				X	
Level of care				X	
Head cooling				X	
Hypoxic ischaemic encephalopathy				X	
Major neonatal morbidity (one or more of the following)					
Intraventricular haemorrhage				X	
Oxygen required >28 days				X	
Necrotising enterocolitis				X	
Sepsis				X	
Retinopathy of prematurity				X	

Data item		Expected EPR source of data			
Information required	Comments	Maternity	Ultrasound	Neonatal	Hospital activity
Parameters related to transitional care:					
Length of stay		X		X	
Neonatal hypothermia		X		X	
Neonatal hypoglycemia		X		X	
Neonatal nasogastric tube feeding		X		X	
Final birth outcome					
Antenatal stillbirth		X			
Intrapartum stillbirth		X			
Early neonatal death		X		X	
Late neonatal death		X		X	
Death before discharge (after 28 days of birth)		X		X	
Cause of death	To determine the non-anomalous stillbirth.	X		X	
Health economics					
Number of ultrasound scans after 24 weeks	Where available in the admin/cost system				X
Number of antenatal day unit visits (start date/time, end date/time)	Can come from hospital reports for the same period if no electronic data is recorded.	X			X
Reason for ADU visit	Please let us know if this data is only held in paper records	X			
Number of antenatal clinic appointments	Can come from hospital reports for the same period if no electronic data is recorded.				X
Length of maternal antenatal stay		X			X
Length of maternal postnatal stay		X			X
Length of overall neonatal stay		X		X	X
Ultrasound data					
Examination Date			X		
Date of birth			X		
Hospital Number			X		
Pregnancy number (on US system - not parity)			X		
Total exams (on US system)			X		

Data item		Expected EPR source of data			
Information required	Comments	Maternity	Ultrasound	Neonatal	Hospital activity
For 1st trimester screening:					
Exam number (on US system)			X		
Gestational age	Mandatory		X		
Fetus (number)			X		
EDD by scan			X		
CRL			X		
Fetal Abnormalities			X		
PAPP-A value			X		
PAPP-A MoM			X		
B-HGC value			X		
B-HCG MoM			X		
For all other scans:					
Exam number (on US system)			X		
Gestational age	Mandatory		X		
Fetus (number)			X		
Presentation			X		
BPD	Mandatory		X		
HC	Mandatory		X		
AC	Mandatory		X		
FL	Mandatory		X		
EFW	Mandatory		X		
EFW centile	From software		X		
EFW method (formula used if variation exists)			X		
Amniotic fluid (subjective - oligo, normal, poly)			X		
AFI			X		
Deepest pool (fluid)			X		
Placenta position			X		
Umb art PI			X		
Umb art RI			X		
Umb art EDF			X		
MCA PI			X		
MCA RI			X		
MCA PSV			X		
DV A-wave (categorical variable)			X		
DV PI			X		
Fetal Abnormalities			X		

10.4 STRUCTURED QUERY LANGUAGE (SQL) CODE TO EXTRACT NEONATAL DATA FROM CLEVERMED BADGERNET ELECTRONIC PATIENT RECORD SOFTWARE.

Baby data query

PARAMETERS [NHSCODE] Text (255);

SELECT

NNUEpisodes.BadgerUniqueID AS badger_id,

NNUEpisodes.NationalIDMother,

NNUEpisodes.HospitalIDMother,

NNUEpisodes.NationalIDBaby,

NNUEpisodes.HospitalIDBaby,

NNUEpisodes.BirthTimeBaby AS Birth_datetime,

NNUEpisodes.AdmitTime AS Admission_datetime,

NNUEpisodes.Sex AS Neonatal_Sex,

NNUEpisodes.GestationWeeks AS GA_weeks,

NNUEpisodes.GestationDays AS GA_days,

NNUEpisodes.Birthweight AS Birthweight_grams,

NNUEpisodes.BirthOrder AS Birth_order,

NNUEpisodes.FetusNumber AS Number_of_babies,

NNUEpisodes.BirthHeadCircumference AS Head_circumference,

NNUEpisodes.Apgar5 AS 5min_apgar,

NNUEpisodes.CordArterialpH AS Arterial_cord_pH,

NNUEpisodes.CordArterialBE AS Arterial_cord_BE,

IIf([Resuscitation] Is Null,Null,IIf([Resuscitation] Is Not Null And [Resuscitation]<>"00","Yes","No")) AS Resp_support_delivery,

IIf([DischTime]=Null,"Current   inpatient",Int([DischTime]-[AdmitTime]))   AS Length_of_stay,

 [bapm2011 days].[1] AS BAPM2011_IC,

[bapm2011 days].[2] AS BAPM2011_HD,

[bapm2011 days].[3] AS BAPM2011_SC,

IIf([Cooled]=1,"Yes","No") AS Cooling,

IIf([HIEGrade]>0,[HIEGrade],Null) AS HIE_worst_grade,

diagnoses.Intraventricular_haemorrhage,

diagnoses.[Oxygen_required>28days],

diagnoses.Necrotising_enterocolitis,

diagnoses.Sepsis,

diagnoses.Retinopathy_of_prematurity,

diagnoses.Hypothermia,

diagnoses.Hypoglycaemia,

ng_tube_feeding.NG_tube_feeding,

NNUEpisodes.DischTime AS Discharge_datetime,

Int([dischtime]-[birthtimebaby]) AS Days_at_discharge,

IIf([DischargeDestination]="3","Yes","") AS [Neonata_death<28days],

NNUEpisodes.DiedCause,

[hrg days].[HRG1 IC], [hrg days].[HRG2 HD], [hrg days].[HRG3 SC], [hrg days].[HRG4&5 NC],

NNUEpisodes.AdmitFromNHSCode

FROM (((NNUEpisodes LEFT JOIN diagnoses ON NNUEpisodes.EntityID = diagnoses.EntityID) LEFT JOIN ng_tube_feeding ON NNUEpisodes.EntityID = ng_tube_feeding.EntityID) LEFT JOIN [bapm2011 days] ON NNUEpisodes.EntityID = [bapm2011 days].EntityID) LEFT JOIN [hrg days] ON NNUEpisodes.EntityID = [hrg days].EntityID

WHERE (((NNUEpisodes.AdmitFromNHSCode) Like [NHSCODE]));

'Hrg days' query

Field	Expr1: NNUDaySum.EntityID	If([HRG]<0,"HRG4&5 NC",Choose([HRG],"HRG1 IC","HRG2 HD","HRG3 SC","HRG4&5 NC","HRG4&5 NC"))	CountOfCareDate: Count(NNUDaySum.CareDate)
Table:			
Total:	Group By	Group By	Expression
Crosstab:	Row Heading	Column Heading	Value
Sort:			
Criteria:			
Or:			

BAPM2011 days

Field	Expr1: NNUDaySum.EntityID	f([bapm2011]<0,"Unk",Choose([bapm2011],1,2,3,3))	CountOfCareDate: Count(NNUDaySum.CareDate)
Table:			
Total:	Group By	Group By	Expression
Crosstab:	Row Heading	Column Heading	Value
Sort:			
Criteria:			
Or:			

NG tube feeding

Field	Expr1: NNUDaySum.EntityID	ng_tube_feeding: IIf(Sum(IIf([FeedingMethod]="4",1,0))>0,"Yes","")
Table:		
Total:	Group By	Expression
Sort:	Ascending	
Show:	☒	☒
Criteria:		
Or:		

Diagnoses

Field	Total:	Show:
Expr1: NNUCodedItems.EntityID	Group By	⊠
intraventricular_haemorrhage: IIf(Sum(IIf([Code]="15705" Or [Code]="157	Expression	⊠
oxygen_required>28days: IIf(Sum(IIf([Code]="15605",1,0))>0,"Yes","")	Expression	⊠
necrotising_enterocolitis: IIf(Sum(IIf([Code]="1010683" Or [Code]="10708" Or [Code]="15809",1,0))>0,"Yes","")	Expression	⊠
sepsis: IIf(Sum(IIf([Code]="1015985" Or [Code]="15007" Or [Code]="15639" Or [Code]="15640" Or [Code]="15641" Or [Code]="15642" Or [Code]="15643" Or [Code]="15644" Or [Code]="15645" Or [Code]="15647" Or [Code]="15648" Or [Code]="15649" Or [Code]="15651" Or [Code]="15653" Or [Code]="15664" Or [Code]="15671" Or [Code]="15679" Or [Code]="16776",1,0))>0,"Yes","")	Expression	⊠
retinopathy_of_prematurity: IIf(Sum(IIf([Code]="11010232" Or [Code]="15218" Or [Code]="15219" Or [Code]="15220" Or [Code]="15221" Or [Code]="15222" Or [Code]="15223" Or [Code]="15224",1,0))>0,"Yes","")	Expression	⊠
hypothermia: IIf(Sum(IIf([Code]="15826",1,0))>0,"Yes","")	Expression	⊠
hypoglycaemia: IIf(Sum(IIf([Code]="15773",1,0))>0,"Yes","")	Expression	⊠

Table:
Total:
Sort:
Show:
Criteria:
Or:

10.5 EXTRACT OF THE DATA DICTIONARY TO GUIDE THE DATA MANAGEMENT AND HARMONISATION PROCESS

Variable name	Title	Variable Explanation	Variable option	Site X	Site Y
mc_agedel	Maternal age at delivery	Calculated on site from DOB and date of delivery	Numerical		
mc_ageconc	Maternal age at conception	= Date of delivery - gestational age at delivery (total days) + 14	Numerical		
mr_agebk	Age at booking - provided by the site		Numerical		
mc_age12w	Age at booking - calculated	Calculated from age at conception plus 10 weeks	Numerical		
mc_imd	IMD	Calculated on site from postcode	Numerical		
mc_lsoa	LSOA	Calculated on site from postcode			
mc_msoa	MSOA	Calculated on site from postcode			
Maternal demographics					
mr_ht	Height	Maternal height at booking	Numerical	**mheight** numerical	
mr_wt	Weight	Maternal weight at booking	Numerical	**mweight** numerical	
mr_bmi	BMI	BMI as provided by site	Numerical	**bmiatbooking** numerical	**mothercurrentriskobesitybmi** No Yes
ur_ht	Height	Maternal height (m) at booking	Numerical	**US_mat_height** Numerical (cm)	**Pre_pregnancy_height_cm** Numerical (cm)
ur_wt	Weight	Maternal weight (kg) at booking	Numerical	**US_mat_weight** Numerical (kg)	**Pre_pregnancy_weight_kg** Numerical (kg)
ur_bmi	BMI	BMI as provided by site	Numerical	**US_mat_bmi** Numerical (kg/m2)	**BMI** Numerical (kg/m2)
mh_ht	Height	Maternal height (m) at booking	Numerical		
mh_wt	Weight	Maternal weight (kg) at booking	Numerical		
mc_bmi	BMI calculated	BMI created from mh_ht & mh_wt (kg/m^2)	Numerical		
mc_bmicat	BMI category	Categorised from mc_bmi	1 = <18.5 2 = 18.5-24.99 3 = 25.0-29.99 4 = 30.0-34.99 5 = 35.0-39.99 6 = ≥40.0		
mr_ed	Education	Maternal level of education	NOT AVAILABLE		**motheroccupation** (A-Z)

10.6 STANDARDS FOR REPORTING IMPLEMENTATION STUDIES (STaRI)

CHECKLIST FOR REPORTING IMPLEMENTATION STUDIES

Checklist item		Reported in section no.	Implementation Strategy	Reported in section no.	Intervention
			"Implementation strategy" refers to how the intervention was implemented		"Intervention" refers to the healthcare or public health intervention that is being implemented.
Title and abstract					
Title	1	3	Identification as an implementation study, and description of the methodology in the title and/or keywords		Identification as an implementation study, and description of the methodology in the title and/or keywords
Abstract	2	N/A	Identification as an implementation study, including a description of the implementation strategy being implemented, and defining the key implementation and health outcomes.		Identification as an implementation study, including a description of the implementation strategy to be tested, the evidence-based intervention being implemented, and defining the key implementation and health outcomes.
Introduction					
Introduction	3	1 & 3.1	Description of the problem, challenge or deficiency in healthcare or public health that the intervention being implemented aims to address.		Description of the problem, challenge or deficiency in healthcare or public health that the intervention being implemented aims to address.
Rationale	4	Figure 2.7	The scientific background and rationale for the implementation strategy (including any underpinning theory/framework/model], how it is expected to achieve its effects and any pilot work).	1.6.2.7	The scientific background and rationale for the intervention being implemented (including evidence about its effectiveness and how it is expected to achieve its effects).
Aims and objectives	5	3.1.1	The aims of the study, differentiating between implementation objectives and any intervention objectives.		The aims of the study, differentiating between implementation objectives and any intervention objectives.
Methods: description					
Design	6	3.2.1	The design and key features of the evaluation, (cross referencing to any appropriate methodology reporting standards) and any changes to study protocol, with reasons		The design and key features of the evaluation, (cross referencing to any appropriate methodology reporting standards) and any changes to study protocol, with reasons
Context	7	3.3.2.1	The context in which the intervention was implemented. (Consider social, economic, policy, healthcare, organisational barriers and facilitators that might influence implementation elsewhere).		The context in which the intervention was implemented. (Consider social, economic, policy, healthcare, organisational barriers and facilitators that might influence implementation elsewhere).
Targeted 'sites'	8	2.1.2	The characteristics of the targeted 'site(s)' (e.g locations/personnel/resources etc.) for implementation and any eligibility criteria.	2.1.3	The population targeted by the intervention and any eligibility criteria.
Description	9	Box 3.1	A description of the implementation strategy	1.6.2 & 10.1	A description of the intervention
Sub-groups	10	2.3.4.1	Any sub-groups recruited for additional research tasks, and/or nested studies are described		Any sub-groups recruited for additional research tasks, and/or nested studies are described
Methods: evaluation					
Outcomes	11	2.3.2.1, 2.3.2.3	Defined pre-specified primary and other outcome(s) of the implementation strategy, and how they were assessed. Document any pre-determined targets	2.2	Defined pre-specified primary and other outcome(s) of the intervention (if assessed), and how they were assessed. Document any pre-determined targets
Process evaluation	12	2.3.2.4	Process evaluation objectives and outcomes related to the mechanism by which the strategy is expected to work		Process evaluation objectives and outcomes related to the mechanism by which the strategy is expected to work

Checklist item	Reported in section no.	Implementation Strategy	Reported in section no.	Intervention
Methods: evaluation (continued)				
Economic evaluation 13	2.4, 5.2.4, 5.2.5 & 5.2.7	Methods for resource use, costs, economic outcomes and analysis for the implementation strategy	2.4, 5.2.4, 5.2.5 & 5.2.7	Methods for resource use, costs, economic outcomes and analysis for the intervention
Sample size 14	3.2.3.3	Rationale for sample sizes (including sample size calculations, budgetary constraints, practical considerations, data saturation, as appropriate)		
Analysis 15	3.2.3.4 & 3.2.4.2	Methods of analysis (with reasons for that choice)		
Sub-group analyses 16	N/A	Any a priori sub-group analyses (e.g. between different sites in a multicentre study, different clinical or demographic populations), and sub-groups recruited to specific nested research tasks		
Results				
Characteristics 17	Table 3.2 & Table 3.3	Proportion recruited and characteristics of the recipient population for the implementation strategy	Table 2.16 & Table 2.17	Proportion recruited and characteristics (if appropriate) of the recipient population for the intervention
Outcomes 18	3.2.3.2	Primary and other outcome(s) of the implementation strategy	2.6.3	Primary and other outcome(s) of the Intervention (if assessed)
Process outcomes 19	3.2.4	Process data related to the implementation strategy mapped to the mechanism by which the strategy is expected to work		
Economic evaluation 20	Table 5.6	Resource use, costs, economic outcomes and analysis for the implementation strategy	5.3	Resource use, costs, economic outcomes and analysis for the intervention
Sub-group analyses 21	N/A	Representativeness and outcomes of subgroups including those recruited to specific research tasks		
Fidelity/ adaptation 22	3.3.1	Fidelity to implementation strategy as planned and adaptation to suit context and preferences	3.3.1	Fidelity to delivering the core components of intervention (where measured)
Contextual changes 23	3.3.2.1	Contextual changes (if any) which may have affected outcomes		
Harms 24	N/A	All important harms or unintended effects in each group		
Discussion				
Structured discussion 25	3.4	Summary of findings, strengths and limitations, comparisons with other studies, conclusions and implications		
Implications 26	3.4.4	Discussion of policy, practice and/or research implications of the implementation strategy (specifically including scalability)	8.3	Discussion of policy, practice and/or research implications of the intervention (specifically including sustainability)
General				
Statements 27	2.1.4 - 2.1.8	Include statement(s) on regulatory approvals (including, as appropriate, ethical approval, confidential use of routine data, governance approval), trial/study registration (availability of protocol), funding and conflicts of interest		

358

10.7 DATA COLLECTION FORM FOR THE NOTES AUDIT ON INTERVENTION COMPLIANCE

Cluster site reference	1	2	3	4	5
NHS number					
DOB					
Study ID (generate)	Automatic calculation				
Details of birth					
Date of baby's birth					
EDD (scan)					
Birthweight					
Gestational age at birth (weeks)	Automatic calculation				
Baby sex					
Maternal Customisation Data					
Age at baby's birth	Automatic calculation				
Ethnicity					
Maternal height at booking (m)					
Maternal weight at booking (kg)					
Maternal BMI	Automatic calculation				
Parity					
Risk Factors for SGA					
Age >40y	Automatic calculation				
Smoker at booking					
Drug misuse					
Previous SGA baby (see GAP chart)					
Previous stillbirth					
Chronic hypertension					
Pre-existing diabetes					
Renal impairment					
Antiphospholipid syndrome					
Large fibroids					
BMI>35	Automatic calculation				
Development of risk factors in pregnancy after booking?					

PAPP-A <0.415 MoM					
Fetal echogenic bowel					
Risk Status					
Evidence of Risk Assessment at Booking?					
Date of risk assessment					
Risk as assessed by clinician					
Risk status as classified by local policy					
Risk status as classified by GAP					
Low Risk Women					
Customised Chart in Notes					
Gestational age of first SFH plot					
Minimum expected number of SFH plots	Automatic calculation				
Number of SFH plots at least 2 weeks apart (from 26/40 onwards)					
Percentage of expected SFH plots done	Automatic calculation				
Was a growth USS performed?					
Date of USS growth					
Was the growth USS indicated by a SFH deviation?					
Was the EFW correctly plotted onto the GROW chart?					
Does the GROW chart suggest (interpretation of reviewer) presence of					
-Plot<10th centile					
-Slow growth					
-Static growth					
-Accelerative growth					
If deviation noted by reviewer, was USS growth performed?					
Was the growth USS interpreted by a clinician as demonstrating SGA?					
Change of Risk Status (Low to High Risk?)					
Development of late pregnancy risk factors in pregnancy after booking?					
Severe PIH or any PET					
Unexplained APH					
SGA diagnosed when referred for growth scan					

Was risk assessment for SGA re-assessed?					
High Risk Women					
Customised Chart in Notes					
Gestation at determination of high-risk status (if at booking = 12)					
First USS growth					
Minimum expected number of EFW measurements	Automatic calculation				
Number of EFWs measured (minimum interval of 2 weeks)					
Percentage of expected USS EFWs done	Automatic calculation				
Number of EFWs plotted on customised chart (minimum interval of 2 weeks)					
Percentage of expected USS EFWs plotted	Automatic calculation				
Was the growth USS interpreted by a clinician as demonstrating SGA?					
Does the GROW chart suggest (interpretation of reviewer) presence of					
-Plot<10th centile					
-Slow growth					
-Static growth					
-Accelerative growth					
If there is evidence of growth deviation, was this noted by a clinician?					

10.8 TOPIC GUIDE FOR SEMI-STRUCTURED INTERVIEWS WITH GAP LEADS IN IMPLEMENTING SITES

Establish purpose of interview: ensure opportunity to read PIS and gain informed consent before commencing interview.

Establish practice for confidentiality of content (e.g., use of pseudonyms for participants/colleagues, and others mentioned, including Trust site/ obscuring of roles).

Clarify participants understanding of the term 'GAP' and specify that when we refer to this, we refer to the entire intervention (baseline audit, training, protocols, risk assessment, GROW charts, missed case audit).

Context and preparation

1. How did you/your Trust come to hear about the GAP approach?
2. What are your thoughts about these interventions? (Did you think there was a need for these; explore evidence/policy/political driver contexts)
3. Other issues; Context –priorities, politics, other things happening in the Trust at the same time, or in the outside world and affecting the Trust/maternity department
   Issues around whether SGA or stillbirth are considered problematic here.

Early implementation

As you know, (your) Trust is an early implementer in the DESiGN trial and is in the process of or has introduced the GAP approach, which is designed to increase antenatal detection of SGA babies. We are interested to find out about your experience of how these changes have been implemented at your Trust/hospitals within your Trust.

1. Could you tell me, from your perspective, how the GAP programme has been implemented here?
   a. **Who is your GAP team** - do you have representatives from each area (clinicians, midwives, sonographers)?
   b. **Timeline and 'go live' date – when did you go live/when are you planning to go live?**
   c. **What activities, promotion materials, posters, meetings or other steps have been taken to raise awareness of the GAP protocols? Who has done these?**
   d. **Has this Trust made any additional policy changes since implementing GAP (i.e., policy/resources regarding other Stillbirth Care Bundle elements - smoking cessation, reduced fetal movements or fetal monitoring in labour?)**
   e. **Baseline audit**
      - Who is doing it?
      - How much time will take to finish it?
      - Should the audit be done retrospectively x prospectively?
      - If completed, do you know what the baseline SGA detection rate was here?
2. **Explore which staff have been trained in the new protocol and in the use of customised growth charts; when this started, and how it has been going, whether there have been any difficulties with staff training.**
   a. Train the trainer: provided by PI.

- Invite comments on the usefulness of this training, sufficiency of materials provided, and ongoing support by PI
    b. Face-to-face training:
        - The PI requires Trusts to reach 75% face to face 'front line' staff training before GAP can be implemented; is this a useful target? What figure would you propose before implementing GAP? (if you were advising another Trust, for example).
        - Have you encountered barriers or facilitators in relation to reaching this target?
        - How is face to face training done?
        - How long does it take - time allocation?
        - Who is in charge of cascading the training?
        - Training GPs
    c. E-Learning
        - The PI required trusts to reach 75% e-learning before 'go live' and this was altered to 75% within a year of receiving initial training. What barriers and facilitators have you encountered in reaching this target?
        - How much time does it take to complete e-learning?
        - What arrangements has your Trust made for staff to undertake this training? (Is it mandatory? Is it done in staff's paid time, or in paid study time, or own time?)
        - Is it useful? What are your views on the materials and information provided?

Full Implementation

1. **Risk assessment for SGA**
    a. In your trust, who does the initial risk assessment for SGA?
    b. What approach is used? (GAP/RCOG/Other, or variation on these)
    c. What decisions have been made around BMI, smoking, referral pathways/scan frequency?
    d. If risk factors are identified, are women referred to an obstetrician for review?
2. **Customised growth charts (generated when, by whom?)**
    a. Who generates the chart?
    b. When are the charts created?
    c. Are there enough clinics/ultrasonographers/appointments available?
    d. Does generating the chart lengthen appointment times?(If so, with whom, and by how much?)
        - If 'yes' - how is your Trust responding to this? What impact has it had on staff? i.e., longer clinics, less time for lunch, finishing late, seeing fewer patients in one clinic
    e. Have you rearranged any services in order to accommodate GAP?
    f. Has the protocol lead to new issues (such as following up patients who don't attend appointments, or being asked for additional scans which are outside the protocol?)
    g. Who does the scan referrals - doctors or midwives?
    h. Any other resource and capacity issues affecting this Trust during implementation (appointment/clinic availability, USS, AL, training, etc)?
3. **Management strategies for suspected SGA (plot below line/abnormal growth trajectory)**

    a. Are you aware of any differences between the trust protocol and perinatal institute guidance?

    b. Are there plans to audit these differences and review?

Reflection

1. Sustainability: in your view, is there resource to continue with GAP after the implementation period ends?
2. Do you think that it will be possible to either continue with GAP, if the trial shows a benefit, or to return to the approach used previously, if the trial does not show benefit?

Is there anything else you would like to add? (Provide an opportunity for participants to discuss any issue in relation to GAP implementation not covered by questions).

Thank participant and invite any questions about research or what happens next.

10.9 Topic guide for semi-structured interviews with frontline clinicians in implementing sites

Establish purpose of interview: ensure opportunity to read PIS and gain informed consent before commencing interview.

Establish practice for confidentiality of content (e.g., use of pseudonyms for participants/colleagues, and others mentioned, including Trust site/ obscuring of roles).

Clarify participants understanding of the term 'GAP' and specify that when we refer to this, we refer to the entire intervention (baseline audit, training, protocols, risk assessment, GROW charts, missed case audit).

Context/Preparation:

What are your thoughts about the GAP interventions? (Did you think there was a need for these; explore evidence/policy/political driver contexts)

> Other issues; Context – priorities, politics, other things happening in the Trust at the same time, or in the outside world and affecting the Trust/maternity department
> Issues around whether SGA or stillbirth are considered problematic here.

Early implementation

As you know, (your) Trust is an early implementer in the DESiGN trial and is in the process of or has introduced the GAP approach, which is designed to increase antenatal detection of SGA babies.

We are interested to find out about your experience of how these changes have been implemented at your Trust/hospitals within your Trust. Could you tell me, from your perspective, how the GAP programme has been implemented here?

1. **Who is your GAP team** - do you have representatives from each area (clinicians, midwives, sonographers)?
2. **How was awareness raised** amongst staff? Education about local protocols?
3. **Awareness of any additional policy changes since implementing GAP.** (i.e., policy/resources regarding other Stillbirth Care Bundle elements – smoking cessation, reduced fetal movements or fetal monitoring in labour?)
4. **Were you involved in conducting the baseline audit?**
    a. Who is doing it?
    b. How much time will take to finish it?
    c. Should the audit be done retrospectively x prospectively?
    d. If completed, do you know what the baseline SGA detection rate was here?
5. **Involvement in cascading training or had training cascaded?**
    a. Train the trainer: provided by PI.
        - Invite comments on the usefulness of this training, sufficiency of materials provided, and ongoing support by PI
    b. Face-to-face training:
        - The PI requires Trusts to reach 75% face to face 'front line' staff training before GAP can be implemented; is this a useful target?

        - Have you encountered barriers or facilitators in relation to reaching this target?

- How is face to face training done?
- How long does it take - time allocation?

    c. E-Learning
- The PI required trusts to reach 75% e-learning before 'go live' and this was altered to 75% within a year of receiving initial training.
- What barriers and facilitators have you encountered in reaching this target?
- How much time does it take to complete e-learning?
- What arrangements has your Trust made for staff to undertake this training? (Is it mandatory? Is it done in staff's paid time, or in paid study time, or own time?)
- Is it useful? What are your views on the materials and information provided?

Full implementation

1. **Risk assessment for SGA**
   a. In your trust, who does the initial risk assessment for SGA?
   b. What approach is used?
   c. Has your local protocol been adapted in any way? Thoughts about this.
   d. If risk factors are identified, are women referred to an obstetrician for review, or directly to scans?

2. **Customised growth charts (generated when, by whom?)**
   a. Who generates the chart?
   b. When are the charts created?
   c. How does your trust deal with third trimester late bookers/transfers of care?
   d. Are there enough clinics/ultrasonographers/appointments available?
   e. Does generating the chart lengthen appointment times? (If so, with whom, and by how much?)
      - If 'yes' - how is your Trust responding to this? What impact has it had on staff? i.e., longer clinics, less time for lunch, finishing late, seeing fewer patients in one clinic
   f. Has the protocol lead to new issues (such as following up patients who don't attend appointments, or being asked for additional scans which are outside the protocol?)
   g. Who does the scan referrals - doctors or midwives?
   h. If a midwife or doctor makes a scan referral, can they be confident that the scan they have asked for will be done?
   i. Any other resource and capacity issues affecting this Trust during implementation (appointment/clinic availability, USS, AL, training, etc.)?

3. **Management strategies for suspected SGA (plot below line/abnormal growth trajectory)**
   a. Are you aware of any differences between the trust protocol and perinatal institute guidance?
   b. If a midwife or doctor makes a scan referral, can they be confident that the scan they have asked for will be done?

Reflection

1. Do you think that it will be possible to either continue with GAP, if the trial shows a benefit, or to return to the approach used previously, if the trial does not show benefit?
   a. How easy would it be to return to previous standard practice? Unlearning knowledge/skills.
2. Is there anything else you would like to add? (Provide an opportunity for participants to discuss any issue in relation to GAP implementation not covered by questions).

Thank participant and invite any questions about research or what happens next.

10.10   TOPIC GUIDE FOR SEMI-STRUCTURED INTERVIEWS WITH GAP LEADS IN NON-IMPLEMENTING SITES

Establish purpose of interview: ensure opportunity to read PIS and informed consent before commencing interview.

Clarify participants understanding of the term 'GAP' and specify that when we refer to this, we refer to the entire intervention (baseline audit, training, protocols, risk assessment, GROW charts, missed case audit).

Outline interview scope and establish ground rules for confidentiality of content (e.g., use of pseudonyms for participants/colleagues, and others mentioned, including Trust site/ obscuring of role titles).

Context / preparation

As you know, your Trust is part of the DESiGN trial, and is in the delayed implementation arm. DESiGN is a cluster RCT being conducted to explore whether the GAP approach improves AN detection of SGA babies

1. How did you/your Trust come to hear about the GAP approach? What are your thoughts about these interventions?
2. Did you think there was a need for these; explore evidence/policy/political driver contexts
3. What was the response here when you found out the Trust was in the delayed implementation group?
4. Other issues; Context –priorities, politics, other things happening in the Trust at the same time, or in the outside world and affecting the Trust/maternity department
Issues around whether SGA or stillbirth are problematic here.

Current practice

1. Can you tell me, from your perspective, what it has been like to be in the delayed arm of this study?
2. At the moment, what aspects of the Stillbirth Care bundle is your trust using in current routine care, and how are you approaching these targets?
   a. Reducing smoking in pregnancy
   b. Detecting fetal growth restriction
   c. Raising awareness of reduced fetal movement
   d. Improving effective fetal monitoring in labour
3. Do you have a Trust lead for the Stillbirth Care Bundle?
4. What is your Trust's current routine practice for detection of SGA babies?
   a. Include any booking risk assessment (what approach is used? RCOG/local protocol?)
   b. How are midwives asked to assess fetal growth during routine care and at what points during pregnancy this happens?
   c. Training: Does any current mandatory Trust training cover detection of SGA babies? Are there any other ways (such as newsletters/in-house presentations) that awareness of SGA detection is raised within the Trust?

Reflection
1. When the DESiGN trial ends, does your Trust plan to implement GAP?

    a. If yes:
- **Does your Trust have, or plan to have, a GAP team? If so** - do you plan to have representatives from each area (clinicians, midwives, sonographers)?
- **Is there a plan to have a designated lead clinician to implement GAP**?
- **Are there any current plans to instigate staff training from Perinatal Institute?**
- **Are there anticipated resource issues? (staff time, ultrasound clinics, local issues)**
- **Timeline - what is your current understanding of what this might be?**

    b. If no, or unsure: What are the considerations for your Trust in relation to implementing GAP?
- Is your Trust planning an alternative?

Is there anything else you would like to add?

(Provide an opportunity for participants to discuss any issue in relation to current practice in SGA detection or anticipated GAP GROW implementation not covered by questions).

Thank participant and invite any questions about research or what happens next.

	Site 7	Site 8	Site 9	Site 10	Site 11
Epidemiology	Local high rate of stillbirth related to high-risk population of recent migrants, women with obesity and smokers. Knowledge that babies at risk of stillbirth need to be screened for SGA but that locally, the SGA detection rates were poor. "We definitely have a problem in terms of adverse outcomes like stillbirths, miscarriages and neonatal deaths." "we get a lot of asylum seekers, a lot of late bookers, so we do have women who are considered extremely high risk who come into pregnancy with medical complexities, loads of safeguarding" "We deal with a really high-risk population which includes asylum seekers, refugees, many women who don't speak English at all who come from countries where they haven't had any health care, who haven't got full medical history." "looking at stillbirth in neonatal deaths and tracking that back and seeing that a lot of them, [...] the babies were small for gestational age, and putting those things together and saying if we identified this small for gestational age earlier we might have been able to intervene." "So yes, there were babies being missed and the outcomes were not good for those babies, so definitely needed to be implemented."	Belief that locally there was no problem with SGA detection or high stillbirth rates despite high rates smoking who were not being adequately screened for SGA. Stillbirth rate locally had improved following implementation of a reduced fetal movement policy. "I don't think this particular hospital has had... Compared nationally, our stillbirth rate was quite low, actually." "I:And is there a high smoking rate? I didn't ask. R:There is actually. We're well above the national average." "it was the smokers I was really worried about and the raised BMI ladies, because I thought this evidence is there and we're not doing it and we should be doing it" "I feel we've had quite an impact on our stillbirth rate here at [maternity unit] and I feel that a lot of that could be contributed, mostly, to the fact that we've been looking at reduced fetal movements"	Uncertainty regarding local stillbirth rate in comparison to the National problem and previously not acknowledged that stillbirth associated with SGA, although SGA detection rates thought to be low. Maternity unit serves a high-risk population due to high rates of smoking and referrals for fetal medicine specialist opinion. "maybe at our hospital we had a poor detection rate of SGA babies" "I think in this trust particularly we are doing well. Again, I would say GAP would be more beneficial in other trusts or other areas." "I do not know what the dashboard statistics are on how we are compared to nationally, so I do not know" "So, it wasn't really something that was being looked at in, in, in that way, the two comparables, you know, the two, the stillbirth, or the IUD, relating to the weight. That wasn't, those two components were not being put together." "we have a very high-risk population and we, and, because our [Trust fetal medicine service], we got referrals from all over the country" "if we implemented um growth scans for all women who smoked within this Trust, they believe that our um, our ultrasound department would collapse under the weight"	Belief that local stillbirth rates are either on par or below national rates, but that these rates are high compared to other European countries. Maternity unit serves a high-risk population due to provision of specialist maternal and fetal medicine services, and older age in mothers of local population. Belief that in order to reduce stillbirth rates it is necessary to do more than just screen for SGA. "Again, I'm not sure of our exact percentage, but I'm pretty sure that we are on par, if not below national" "I believe some of the more European countries, that they do have a lower stillbirth rate." "I think sometimes we think that fetal growth is the only cause of stillbirth. So I think if you are trying to reduce stillbirth you need to do more" "Yes, and also we have a lot of referrals from other places, and because we are a tertiary hospital we have a very well-known and recognised Fetal Medicine Unit and a lot of patients come here because of that. And we have a well-established Maternal Medicine Clinic led by [name] and everybody knows him, so a lot of patients with complex conditions and some of these conditions come with a higher risk of stillbirth, so are seen here. So yes, our population might have a slightly higher risk of stillbirth that others because it's slightly more complex, a bit older than average."	Uncertainty regarding local stillbirth rate, but knowledge that London and UK have high rates. Local population has a high rate of Asian ethnicity, which are possibly more affected by constitutionally small babies. Knowledge that GAP-GROW has been effective at reducing stillbirth rate nationally. "well I don't think our stillbirth rate is higher than anywhere else" "so the UK has a problem with the stillbirth rate, London in particular, you know, we are not identifying the babies as we should" "So I know that the stillbirth rate since they've been introduced has reduced, which is the first time in years and years" "Yeah...yeah, because in our population here, we do lots of Asians... and I think that our constitutionally... small babies are fine as opposed to the babies that really ought to have more intervention and observation."

	Site 7	Site 8	Site 9	Site 10	Site 11
Ethical	Distress regarding capacity issues for consultant and scan appointments.	Stillbirth considered to be morally distressing.	Busy clinical roles mean that clinical activities and training must be prioritised.	Noted a desire to perform better at work.	Agreement that interventions which save babies' lives are worthwhile but disagreement that GAP is necessarily the right intervention (not time or cost effective).
	"It's how you then balance the need for that woman to have the scan and also with the availability of slots.."	*"It's an awful thing that happens at the time"*	Distress and reaction following adverse incidents involving SGA babies.	*"You are going to perform better because you are aware about, for example, the ethnicity of the woman."*	*"I agree with the principle that if you even save one more baby than you would ordinarily then yes it's well worth it. I don't know whether I believe that it's true."*
	Concern that cases of SGA are missed due to the high pressure of working	Concern from one respondent about the risk to the developing fetus from serial ultrasound scans.	*"It is me being super cautious because of past experience."*		*"That's my issue. It's the cost effectiveness, because all of that extra time that we are using up, whether it be generating the charts or... At the beginning of this interview we talked about how at the study day we learnt that the whole process, the premise of the day was if we save one more baby it's worth it, but my issue is, okay, so you might save one more baby with the idea of the GAP protocol, but actually in the time that we are using up implementing it are there other things that our time could be more effectively spent on to save another baby?"*
	"occasionally things are missed just because of the pressure of work, if they are both feeling rushed and all of that, that might be a reason for it being missed."	*"The one thing, it's scanning these developing fetuses. In the future, may there be a problem with some of these children and it comes that it's from the cause of having all these scans when they were developing? That's something that does worry me sometimes."*	Concern that charts discriminate unfairly against people from non-White racial origins.		Feeling of responsibility to patients to ensure that the baby's 'make it' through the pregnancy.
		Concern about offending groups of women who smoke or are obese by labelling them as 'high risk'.	*"they thought that they were addressing this by using the population-based chart, saying well OK, well look we are measuring, and we are plotting on a chart. But it is, is it an effective tool? And that, that's the thing with looking at that. But yes we were doing it. But actually it's not an effective tool, if you look at it that way. Um. Because it was based on this white, Caucasian, woman in the 1980s, and that's not the clientele that we have here. So, you know, that's only a very small number of our, our clients would meet that criteria. So therefore we're, you know, it makes sense that we need something more customised."*		*"I have a vested interest as a community midwife in identifying the low risk ladies whose babies unfortunately don't make it. I felt sort of a responsibility that you know, as part of the low risk midwifery role to actually be better at doing this..."*
		Distress as a result of previous adverse events.			Strongly held beliefs regarding the diagnostic methods for fetal growth restriction.
		"We were a little bit reactive, we had an IUFD a few years ago and when we looked into it the lady had been referred for small for gestational age and didn't get seen probably for four days. So I think we were a little bit reactive and changed our guideline because of that."			*"So we still say if there's a reduction in the AC on our graph, we will say it and we just basically have been told to write the estimated fetal weight, see the GAP chart, but we still need to report, because actually AC is the most important on a growth scan."*
					Conflicting interest of allocating resources fairly between obstetric and gynaecology ultrasounds.
					"Our growth scan rate has gone through the roof and that in turn has affected our gynae waiting list, so we have been breaching our gynae waiting list and I've had one permanent member of staff resign and that was because... well one of the reasons, was mentioned when she told me, was GAP."
Geographical	No relevant references				

371

	Site 7	Site 8	Site 9	Site 10	Site 11
Legal	Midwives have autonomy to make decisions regarding referral for fetal growth ultrasound. "So we have an extra page that we attach to their notes that you then highlight that they will need a series of growth scans. They will also need a consultant referral so that, again, together we can all manage that."	Previously, lack of agreement on who and how to refer for serial growth scans. "I think it's just nice having like a protocol to follow, because before GAP and GROW, the women that were classed as high risk would all get different care depending on what consultant they were under" Professional autonomy: Only consultants/doctors make antenatal referrals for serial growth scans where indicated. New local guidance for the management of women presenting with reduced fetal movements. "So we now have, if somebody comes in with their first episode of reduced fetal movements, we now have a criteria to follow, of who needs a scan."	Midwifery staff have autonomy to make decisions regarding referral for fetal growth "when we have, um, women where we have to give the, their growth scan, n-, routine growth scan, so extra growth scan, er, the midwife just refer to the, to the scan department, and they just book their own appointment." New local guidance for the management of women presenting with reduced fetal movements. "So, if they have reduced foetal movements, they are encouraged to come into the day assessment unit and if they are less than, I think, twenty-eight weeks, the midwife just listens in. If they are more than twenty-eight weeks, then they do the CTG and if they come for their second presentation of reduced foetal movements, then they get referred for a scan." Concern regarding legal action from families of babies in whom SGA is missed. "there'll be a, a very good legal question as well, you know, if a lady comes along, with an undiagnosed SGA, this [referring to GAP] has been taken away, and could have been picked up using that methodology, where would people stand?"	[Quote identifies site so removed] "So, they have, er, because our sonographers are obstetricians."..." Um, where the protocol kind of unde-, identifies who is to go in what pathways, when we've got four pathways. Er, and the sonographer decides that." Professional autonomy: All scanning requests are screened by the doctor in charge of triage and may be refused. "if the midwives wanted to organise a scan for any reason then it would just normally go for our MFAU lead."..." Normally it's triaged because sometimes... Or the MFAU lead might disagree. I was going to say it probably doesn't happen that often, but just to make sure that the scans are absolutely required because, as you can imagine, a lady can be referred a scan for any reason. So to make sure that the scans are appropriate, to make sure that they are happening at the right time. Normally, it will go for our MFAU lead who can triage and just make sure that the scan is appropriate and happening in a timely manner." Previously, the departmental guideline was to offer a third trimester scan to all women, which reassured staff. "Um, so, what happened here is that we traditionally had um a third trimester scan, any-, for, for all women. Um, and then I was obviously lin-, when I first started working here I was obviously feeling reassured that, you know, we're scanning everybody. Then times changed, then we couldn't provide that scan any more, so, for us, it was a period of going from high intervention to nothing." New local guidance for the management of women presenting with reduced fetal movements and plans to re-write the guidelines for electronic fetal monitoring. "we created a um, a working group on um re-, er, changing the current guidelines, er, for intrapartum CTG, um, and therefore that will involve as well new training"	Professional autonomy: lead sonographer vets forms to ensure that all referrals are properly indicated "from an ultrasound perspective, I vet every single form that comes in as superintendent. And I make sure that they're valid requests. If they're not, I'll bounce them back" Concern regarding inherent human error of hand-plotting on GAP charts. "then you are hand plotting them, which I don't find very accurate myself because we've been told to do an empty circle which is if you plot it more accurately it should be a dot"

	Site 7	Site 8	Site 9	Site 10	Site 11
Setting	Some community clinics don't have computers and/or printers. Most community midwives are now being provided with laptops. Trust is going paperlight which has further impacted on the access to printers.	Site short on consultation and scanning space.	[Quote identifies site so removed]	Printers in clinic often don't work.	Some staff work in outlying clinics where there is no access to computers
	"I:And are there lots of cases where your midwives don't have computers in their clinics?	"Because we need more scan rooms and more capacity and if you talk to any of our staff that's what they'll say, because it was built to meet a service 30 years ago, not the current service."	"We sometimes need to ask the patient to come between different hospitals, there are two sisters, but that's as well not very convenient for the patient because if they live next to one hospital they actually want the appointments to be there, but sometimes we offer them appointments in a different hospital and they are a little bit not happy."	"And if the printer works, fine; if the printer doesn't work like this one, then you need to find a printer."	"So when we go to a clinic or outreach area, we have to take the information from the computer, so blood results, scan results and document them in their…in their hospital notes which we take to the clinic"
	R:There's, yes, a fair percentage. It's things like children's centres where they don't. Now, we have got laptops, but laptops you can't print from, so they are the cases you have to come back and print it at a later date."	Sufficient computers for staff to access but slow to load software.	Lacking or unreliable computer and/or printer access in community clinics.	Local shortage of ultrasound referral slots.	"The ultrasound department has needed to expand down the corridor, this has impacted on antenatal clinic rooms. New clinic 'pods' have been built but another clinic room is still required."
	"I think where we are going paper-light we have become light on printers, so I think then we are stuck, isn't it, because we want the printers because we want to print this and the trust is like, "We no longer need that many printers because we are going paper-light"."	"I: And you've got enough computer systems and everything else? R: Yes."	"The IT access can be quite shoddy sometimes out in the community"	"Sometimes it can be hard because sometimes we do have a shortage of slots available"	"ultrasound has expanded down the corridor, so antenatal has gone up, we've had all these extra pods added. Um. And, so, we've sort of run out of space"
	Shortage of midwifery staff due to resignations and maternity leave.	"My biggest problem is getting onto the system at the beginning of the day because the computer system here is very slow"	"They are clinics in children's centres where we have our computers that connect to our remote access servers, which is not very good."		National and local shortage of sonographers. Local department has trained 2 each year to keep up with the need but has otherwise been unable to recruit to vacant posts.
	"Cause I think, we just had a period where we couldn't recruit, because, I think we have like about 8 people, all on maternity leave at the same time. So they were trying to back-recruit, but the paperwork was going through. People, we had people leaving."	Midwives were trained to perform growth scans in order to increase departmental capacity.	"SR:How does that work in setting up or organising your clinic, printing the charts? R:There is not much you can do. They are busy clinics. If you haven't got internet access, you just can't print it. It is quite hard actually."		"there is a national and international shortage of sonographers. Um. And so, trying to staff departments properly, and trying to expand at the rate that you have to do all this extra scanning, er, is very very difficult, because there just aren't the sonographers out there. I've sort of coped with it, 'cause I've trained two sonographers every year. So, all sonographers I've got are sort of, you know, ones I've grown myself, so to speak. But trying to get, recruit, I've had a rolling advert for the last 4 months out, and I haven't, not one single applicant. And we advertise at band 8a, so."
		"SR: So how is the department managing the increase in scans? R1: So we have two midwives doing scanning R2: - So we were trained. R1: - on a Monday, Wednesday and Friday. We have a full day, from half past eight 'til half past four, where the girls will scan these patients."	Obstetric ultrasound scans can be conducted within three distinct departments within the hospital setting.		

	Site 7	Site 8	Site 9	Site 10	Site 11
Political	Awareness of the Secretary of State for Health ambition to reduce stillbirths & neonatal deaths nationally. Awareness of the aims of Each Baby Counts, MBRRACE-UK & the Stillbirth Care Bundle, although sometimes confuses the different aims. *"So Jeremy Hunt made up a figure and then we all tried to achieve it....But actually, I do think it was wonderful because it was such an ambitious figure which resulted in some really ambitious work to achieve it. So I think he said he wanted to reduce it by 25% or 50% by 2025, the number of stillbirths and neonatal deaths, then we've had RCOG with each baby counts, the GAP protocol. That was one of the things, isn't it? It's one of the pillars. And I think they have the smoking cessation and they've got some other bits going on there as well. Yes."* Local implementation of all 4 aspects of Stillbirth Care Bundle. *"KC:OK, so you mentioned you've got CO monitoring, and that's kind of important part of the um stillbirth care bundle element, isn't it? So, um, can I just ask about the other elements of the stillbirth care, care bundle, what, what [Trust] is doing around smoking cessation, or identifying reduced foetal movement? SC28: So, we have Dawes/Redman CTG, we are training more and more women, more and more midwives. Because they don't understand Dawes/Redman, and sometimes, and we had stillbirths which the CTG stayed pre-terminal trace in the Dawes/Redman and still nobody believed the Dawes/Redman. So I have taken the initiative to teach everybody, the DAU have Dawes/Redman for women."* Local borough offers smoking cessation therapies free on prescription. *"I think we're the only borough now who, we offer smoking cessation to the women and their partners, and their family. Um, and it's on prescription, so they can get a certain level free. Um, which is I think unheard of in the other sites."*	A CQC inspector suggested the trust not rush into GAP implementation because of the resultant impact on services. Commissioners were keen for GAP to be implemented. *"I remember a CQC assessor coming who had been the lead midwife at an organisation and she said to me 'Don't rush into it, you will not realise what the impact is on your service', and that echoed with me"* *"then there was the Perinatal Institute's GAP GROW, and our commissioners were very keen"* Recent implementation of the other 3 components of the Stillbirth Care Bundle. *"We've got Saving Babies' Lives, so obviously that's come into place."*	Some clinical leads not keen implement GAP due to a belief that it was not effective and was costly. *"I know that our head of midwifery was not keen to take part in the trial...Um, because he did not believe that it would make any difference, and he didn't want to spend any money, and it cost money."* Belief that Perinatal Institute policy for scan frequency every 2-3 weeks between 24 to 42 weeks is excessive. *"So, we think that the Perinatal Institute protocol is excessive and I do not know a unit that scans people every two weeks all the way up until delivery, but that is what the protocol says, so if somebody is on a scanning protocol, they are meant to be scanned every two to three weeks from twenty-four weeks until forty-two weeks."* Awareness amongst some (not all) staff of the Saving Babies' Lives care bundle and the National ambition to reduce stillbirth rates. *"I think around the same time that we were doing the training last year was when I heard about the Saving Babies' Lives policy document"* Local implementation of the other three elements of the Saving Babies' Lives Care Bundle. *"NHS England Saving Babies live care bundle, so, we start introducing that, probably, it was like a year, a year ago"*	Belief that data examining customised growth charts is biased as it has all come from the same group. *"a lot of the data looking at the customised charts so far have come from the same group which in itself can create some bias."* Awareness of MBRRACE-UK, Saving Babies' Lives care bundle, the National ambition to halve stillbirth by 2020 and the Panorama TV documentary on the Perinatal Institute and GAP-GROW. *"So I know there was a care bundle and I can't remember the exact name of it now but I think there were four main areas they were trying to target"* *"Yes. So there's a national ambition, isn't there, to halve stillbirth by 2020. 2020? I don't think we are there."* *"I also saw the documentary on one TV channel, I can't remember which one, and then they were talking about the customised version and people from the Perinatal Institute"* Trust clinical director is also the chair of [organisation] and was very keen for GAP-GROW to be implemented. *"our clinical director, who has a special interest into foetal grow, er, is also the, um, Chair of the [city] network, and therefore there was a debate within the [city] network of potentially starting, er, the DESiGN trial, er, specifically"* *"And as a Trust I think our, our clinical director was very keen to, for us to um get involved."*	Incentivised by CNST to be compliant with all 4 elements of the Saving Babies' Lives care bundle but variable knowledge of its four elements *"So they are offering a 10% reduction in your CNST or a maximum of 10% it's going to be graded um... uh.. um and so one of the ten elements is about being compliant with the same baby loss care bundle now"* *"I think we have tackled the NHS care bundle quite significantly, I mean we have been doing the carbon monoxide testing on women for quite a while."* *"Obviously, there's the Saving Babies' Lives, March 2016, and I implemented here, because it was relevant to the Day Assessment Units, one of the elements which was reduction in fetal movements. So I was involved with bringing that to the forum."* Pressures from the trust board to implement GAP-GROW quickly, motivated by low compliance with the Saving Babies' Lives care bundle. *"Trust-wide to implement that we really didn't really want to implement until DESIGN was up and recruiting because it was um... really important that it is evaluated and that this was a great opportunity to do that. Um... so that, that has caused a little bit of um... tension I guess within the Trust."* Awareness of the Panorama documentary on the GAP-GROW programme. *"Well going back a few years ago I remember seeing a Panorama programme, I think it was about Gardosi, about his findings and the implementation or the beginning of the GAP charts in the West Midlands and it's purpose to reduce the stillbirth rate."* A senior midwife was employed full-time to support the implementation of GAP. *"we were talking about doing the trial here and I said 'please'. Actually, I was the only one who put my name forward."* *"I was on secondment full time and it was the only thing I need to really be actively doing at that time"*

	Site 7	Site 8	Site 9	Site 10	Site 11
Socio-cultural	Staff who have any doubt about referrals for growth scans feel able to ask for advice from the staff in the antenatal clinic or the sonography department. *"we have got a group WhatsApp and we can message and somebody who is generally at a computer will look it up for you. If not, just ring the clinic and they'll advise you."*	Staff feel comfortable to discuss issues with immediate colleagues and to ask for help. *"And even now they do tend to run things past us, don't they? Which is fine."* *"if we were unsure, we'd just ask one of our colleagues, "Is this the way we do it?" or, "Can you tell me...?" That kind of thing."*	Staff are friendly with midwives using GAP at other trusts and therefore felt that the local trust was falling behind national recommendations. *"I can remember when I was training, other midwives would speak about different trusts that had more personalised growth charts and that maybe at our hospital we had a poor detection rate of SGA babies or IUGR because we were just using the standardised chart and we were quite behind on other trusts by doing that."* *"I have helped other staff. I haven't trained them, but I've helped them when we were first doing it, to say, "Oh it's really easy, this is what you do, you just put this in".* Where decisions to be made are unclear, midwives are happy to refer to immediate colleagues, doctors or the project leader. *"I think it's just specific cases are quite hard to see whether it's something that needs referral or not and I think definitely in those cases it's beneficial to speak to a doctor."*	Variable knowledge re. GAP-GROW prior to implementation. *"I come from a different country. In my country we didn't use it so it was a totally new experience when I joined this trust."* Knowledge that there is conflicting evidence for and against the potential effectiveness of GAP-GROW. *"I know the studies that have been performed so far, some of them show that population might be better at picking up SGA babies, whether others show that customised might be better and it's difficult because a lot of the data looking at the customised charts so far have come from the same group which in itself can create some bias."* *"But, you know, there is a little bit, I can sense that, you know, the difference opinio-, differences in opinions in, um, our obstetric body in terms of the, you know, the accuracy of it."* Junior midwives approaching senior midwives for support when scan referrals are rejected. *"two, um, midwives who've come to me to say I've requested a scan and it's been rejected"*	Variable level of knowledge re. GAP prior to implementation. Those who were aware knew of its potential to reduce stillbirth but were also concerned re. the impact on services *"I didn't know anything about it. I've only worked at this trust since I've qualified. I started in 2008. I wasn't aware of it being used in any other hospitals."* *"We had heard quite a bit about it, yes, because we've been to quite a few study days where hospitals had implemented it and they were giving us their feedback on it".* *"So I know that the stillbirth rate since they've been introduced has reduced, which is the first time in years and years. So yes, we know that they have some effect, but I'm also aware of the impact that that has on services as well because I've heard from colleagues and friends about how it's affected their service."* Early feeling amongst sonographers that they didn't want to implement GAP. *"Generally, I think the majority of us don't really want to. We don't really understand why we are doing it. Yes, we don't feel how it's benefitting... We don't see the huge benefit it is to patients and we just weren't clear on it. I think everyone has different ideas of what it does as well."* Where a midwife is uncertain in making a referral decision, they refer to another colleague (or the lead midwife) for second opinion. *"Usually what we do is, because our clinics run and there's two midwives, so we get a second opinion for measurements and then if we both think that this lady needs an extra growth scan then we just complete the form"*
Socio-economic	Slow implementation in the site due to a lack of funding to support it. *"they just couldn't justify taking a clinical midwife out of the picture to do implementation of a study that was gonna take a while to get off the ground."* Insufficient finances to pay for additional sonographers or to increase current scanning capacity, despite local need for both. *"if the workload increases, it will be, demonstrate that this, we need more sonographers, and we don't have the money!"*		Service limitations mean that the trust cannot offer women a scan every 3 weeks on the serial growth pathway *"we had to change of course er most of the criteria for the, for the Perinatal Institute, to adapt to our, our service, so we can't offer scans every er, every 3 weeks."*	Overbooked scanning appointments can mean that urgent referrals are not seen urgently. Combatted by an offer for sonographers to be paid to do additional sessions to meet demand. *"We are doing some extra hours to help the department so recently we manage with everything."* Trust not willing to pay for the more expensive, electronic version of GAP software. *"I believe it's definitely more expensive to have that electronic version because I know that we have looked into it but I think financially it's just not something that this trust I think is willing to do at the moment."*	There is a local reliance on agency sonographers due to post vacancies and insufficient budget to overstaff the department to ensure that all leave can be covered. *"So ideally we would like to overrecruit so we have plenty of people to cope with things like annual leave and people leaving, maternity leave, that sort of thing, and currently there isn't the budget to do that"*

10.12 PREFERRED REPORTING ITEMS FOR SYSTEMATIC REVIEWS AND META-ANALYSES (PRISMA) STATEMENT FOR REPORTING SYSTEMATIC REVIEWS

	Item no.	Recommendation	Section no.
TITLE			
Title	1	Identify the report as a systematic review.	4.2
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	N/A
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	4.1
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	4.1.1
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	4.2.1
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	4.2.1
Search strategy	7	Present the full search strategies for all databases, registers, and websites, including any filters and limits used.	4.2.1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	4.2.1
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	4.2.2
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g., for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Table 4.2
	10b	List and define all other variables for which data were sought (e.g., participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Table 4.3
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	4.2.3
Effect measures	12	Specify for each outcome the effect measure(s) (e.g., risk ratio, mean difference) used in the synthesis or presentation of results.	4.2.4
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g., tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	4.2.4
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	4.2.4
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	4.2.4
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	4.2.4
	13e	Describe any methods used to explore possible causes of heterogeneity	4.2.5

		among study results (e.g., subgroup analysis, meta-regression).	
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure 4.1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	Table 4.3
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Table 4.4
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g., confidence/credible interval), ideally using structured tables or plots.	4.3
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	4.3
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g., confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	4.3
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	4.3.1
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	4.4.1 & 4.4.2
	23b	Discuss any limitations of the evidence included in the review.	4.4.3
	23c	Discuss any limitations of the review processes used.	4.4.3
	23d	Discuss implications of the results for practice, policy, and future research.	4.4.4
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	4.2
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	4.2
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	4.2.5
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	N/A
Competing interests	26	Declare any competing interests of review authors.	N/A
Availability of data, code, and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	N/A

10.13 CONSOLIDATED HEALTH ECONOMIC EVALUATION REPORTING STANDARDS (CHEERS) STATEMENT FOR REPORTING ECONOMIC EVALUATIONS

Section/Item	Item no.	Recommendation	Reported in section no.
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as 'cost-effectiveness analysis', and describe the interventions compared.	5
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	N/A
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	5.1
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	2.1.3
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	2.1.2
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	2.4.3.1 & 5.2.2
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	2.1.1
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	5.2.2
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	N/A
Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	2.2.1.1 & 5.2.3
Measurement of effectiveness	11a	*Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	2
	11b	*Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	N/A
Measurement and valuation of preference-based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	N/A
Estimating resources and costs	13a	*Single study-based economic evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	2.4.3, 5.2.4 & 5.2.5
	13b	*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	N/A
Currency, price data and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of	2.4.3.3

		reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	5.2.7 & **Figure** 5.1
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	0 & 5.2.7
Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	5.2.7 & 5.2.8
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	Table 5.7
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	Table 5.8
Characterising uncertainty	20a	*Single study-based economic evaluation:* Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	Table **5.9** & Figure 5.2
	20b	*Model-based economic evaluation:* Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	N/A
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	5.2.8
Discussion			
Study findings, limitations, generalisability and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	5.4
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	2.1.8
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	N/A

10.14   THE STRENGTHENING THE REPORTING OF OBSERVATIONAL STUDIES IN EPIDEMIOLOGY (STROBE) COMPLETED CHECKLIST

	Item no.	Recommendation	Reported in section no.
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	6
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	N/A
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	6.1
Objectives	3	State specific objectives, including any prespecified hypotheses	6.1.1
Methods			
Study design	4	Present key elements of study design early in the paper	6.2.1
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	2.1
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls	2.1.3, 6.2.3
		(*b*) For matched studies, give matching criteria and the number of controls per case	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6.2.4, 6.2.5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	2.2.2 & 6.2.6
Bias	9	Describe any efforts to address potential sources of bias	6.3.9
Study size	10	Explain how the study size was arrived at	2.2.1.3
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6.2.5 & 6.2.7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	6.2.7
		(*b*) Describe any methods used to examine subgroups and interactions	6.2.7
		(*c*) Explain how missing data were addressed	6.2.6
		(*d*) If applicable, explain how matching of cases and controls was addressed	N/A
		(*g*) Describe any sensitivity analyses	6.2.7.1
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	0
		(b) Give reasons for non-participation at each stage	Figure 6.1
		(c) Consider use of a flow diagram	Figure 6.1
Descriptive data	14*	(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders	Table 6.3 - Table 6.4
		(b) Indicate number of participants with missing data for each variable of interest	Table 6.2
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure	Table 6.3
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (egg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 6.5 - Table 10.8

		(*b*) Report category boundaries when continuous variables were categorized	6.2.5
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A
Other analyses	17	Report other analyses done—egg analyses of subgroups and interactions, and sensitivity analyses	6.3.6

Discussion

Key results	18	Summarise key results with reference to study objectives	6.3.7
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	6.3.9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	6.3.8
Generalisability	21	Discuss the generalisability (external validity) of the study results	6.3.10

Other information

Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	2.1.8

10.15 THE CONSOLIDATED STANDARDS OF REPORTING TRIALS (CONSORT) CLUSTER EXTENSION CHECKLIST FOR REPORTING RESULTS OF RANDOMISED CONTROL TRIALS

Section/Topic	Item No	Standard Checklist item	Extension for cluster designs	Section No.
Title and abstract				
	1a	Identification as a randomised trial in the title	Identification as a cluster randomised trial in the title	7
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	See table 2	N/A
Introduction				
Background and objectives	2a	Scientific background and explanation of rationale	Rationale for using a cluster design	7.1
	2b	Specific objectives or hypotheses	Whether objectives pertain to the cluster level, the individual participant level or both	7.1.1
Methods				
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Definition of cluster and description of how the design features apply to the clusters	2.1
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons		N/A (see primary trial report)
Participants	4a	Eligibility criteria for participants	Eligibility criteria for clusters	2.1.2
	4b	Settings and locations where the data were collected		2.1.2
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	Whether interventions pertain to the cluster level, the individual participant level or both	2.1.1
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	Whether outcome measures pertain to the cluster level, the individual participant level or both	2.2.1
	6b	Any changes to trial outcomes after the trial commenced, with reasons		N/A (see primary trial report)
Sample size	7a	How sample size was determined	Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intracluster correlation (ICC or k), and an indication of its uncertainty	2.2.1.3
	7b	When applicable, explanation of any interim analyses and stopping guidelines		2.1.6
Randomisation:				
Sequence generation	8a	Method used to generate the random allocation sequence		2.1.2

	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details of stratification or matching if used	2.1.2
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level or both	2.1.2
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Replace by 10a, 10b and 10c	2.1.2
	10a		Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions?	2.1.2
	10b		Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling)	2.1.3
	10c		From whom consent was sought (representatives of the cluster, or individual cluster members, or both), and whether consent was sought before or after randomisation	2.1.2
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how		N/A
	11b	If relevant, description of the similarity of interventions		N/A
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	How clustering was taken into account	2.2.6
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses		7.2
Results				
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome	Figure 2.10 & Table 7.2
	13b	For each group, losses and exclusions after randomisation, together with reasons	For each group, losses and exclusions for both clusters and individual cluster members	N/A
Recruitment	14a	Dates defining the periods of recruitment and follow-up		2.6.1
	14b	Why the trial ended or was stopped		N/A
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Baseline characteristics for the individual and cluster	

				levels as applicable for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	For each group, number of clusters included in each analysis	Table 7.2
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or k) for each primary outcome	Table 7.5 - Table 7.6
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended		-
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory		7.3
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)		N/A
Discussion				
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses		7.4.3
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	Generalisability to clusters and/or individual participants (as relevant)	7.4.4
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence		7.4.2
Other information				
Registration	23	Registration number and name of trial registry		2.1.7
Protocol	24	Where the full trial protocol can be accessed, if available		See primary trial report
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders		2.1.8

10.16 CHAPTER 6: SUPPLEMENTARY RESULTS TABLES

Figure 10.1 - Consort diagram detailing the construction of the study population (available case data)
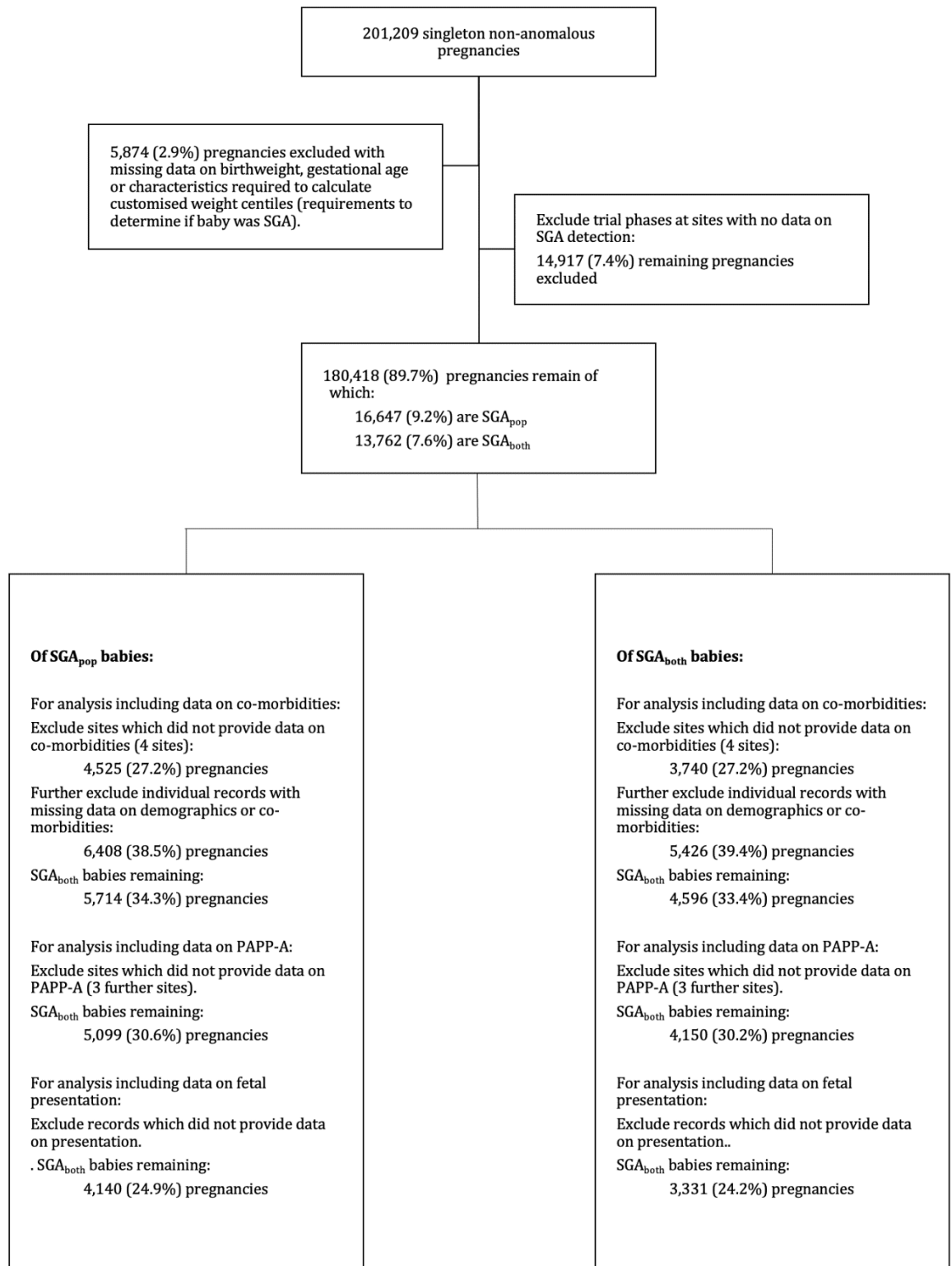
Table 10.1 - Characteristics of the included women, presented for all SGA$_{both}$ pregnancies, and stratified by detection status (complete case analysis)

		All SGA$_{both}$ babies (n=4,596)	Missed SGA$_{both}$ (n=3,438)	Detected SGA$_{both}$ (n=1,158)
Age (years), n(%)	Mean (SD)	31.35 (5.56)	31.38 (5.48)	31.25 (5.79)
	Age over 40y	231 (5.0%)	161 (4.7%)	70 (6.0%)
IMD, n(%)	1=least deprived	466 (10.1%)	329 (9.6%)	137 (11.8%)
	2	514 (11.2%)	368 (10.7%)	146 (12.6%)
	3	1,092 (23.8%)	816 (23.7%)	276 (23.8%)
	4	1,552 (33.8%)	1,194 (34.7%)	358 (30.9%)
	5=most deprived	972 (21.1%)	731 (21.3%)	241 (20.8%)
Ethnicity, n(%)	White	2,109 (45.9%)	1,612 (46.9%)	497 (42.9%)
	Black	808 (17.6%)	601 (17.5%)	207 (17.9%)
	Asian	1,102 (24.0%)	773 (22.5%)	329 (28.4%)
	Mixed	100 (2.2%)	70 (2.0%)	30 (2.6%)
	Other	477 (10.4%)	382 (11.1%)	95 (8.2%)
BMI (kg/m^2), n(%)	Mean (SD)	25.5 (5.32)	25.56 (5.19)	25.30 (5.70)
	Under 18.5	171 (3.7%)	105 (3.1%)	66 (5.7%)
	18.5-24.9	2,328 (50.7%)	1,744 (50.7%)	584 (50.4%)
	25.0-29.9	1,299 (28.3%)	993 (28.9%)	306 (26.4%)
	30.0-34.9	526 (11.4%)	393 (11.4%)	133 (11.5%)
	35.0-39.9	185 (4.0%)	142 (4.1%)	43 (3.7%)
	40.0 or above	87 (1.9%)	61 (1.8%)	26 (2.2%)
Parity, n(%)	0	2,339 (50.9%)	1,747 (50.8%)	592 (51.1%)
	1	1,407 (30.6%)	1,039 (30.2%)	368 (31.8%)
	2	486 (10.6%)	374 (10.9%)	112 (9.7%)
	3	218 (4.7%)	162 (4.7%)	56 (4.8%)
	4 or above	146 (3.2%)	116 (3.4%)	30 (2.6%)
Smoking, n(%)	Smoker	471 (10.2%)	330 (9.6%)	141 (12.2%)
Co-morbidities, n(%)	Hypertension	98 (2.1%)	61 (1.8%)	37 (3.2%)
	Diabetes	71 (1.5%)	53 (1.5%)	18 (1.6%)
	APLS	5 (0.1%)	3 (0.1%)	2 (0.2%)
Antenatal complications, n(%)	Pre-eclampsia	158 (3.4%)	96 (2.8%)	62 (5.4%)
	PIH	102 (2.2%)	68 (2.0%)	34 (2.9%)
	GDM	264 (5.7%)	171 (5.0%)	93 (8.0%)
PAPP-A, n/N(%)	<0.300MoM	111/4,150 (2.7%)	65/3,115 (2.1%)	46/1,035 (4.4%)
	0.3-0.415MoM	187/4,150 (4.5%)	111/3,115 (3.6%)	76/1,035 (7.3%)
	>0.415MoM	2,105/4,150 (50.7%)	1,557/3,115 (50.0%)	548/1,035 (52.9%)
	Missing data	1,747/4,150 (42.1%)	1,382/3,115 (44.4%)	365/1,035 (35.3%)
Indication for serial fetal scans,[+] n/N(%)	Any indication	1,446/4,150 (34.8%)	987/3,115 (31.7%)	459/1,035 (44.3%)
	No indication	2,704/4,150 (65.2%)	2,128/3,115 (68.3%)	576/1,035 (55.7%)

No recorded indication, complete case data except that information on PAPP-A may be missing.

Table 10.2 - Characteristics of the included babies, presented by all pregnancies, and stratified by detection status (complete case analysis)

		All SGA$_{both}$ babies (n=4,596)	Missed SGA$_{both}$ (n=3,438)	Detected SGA$_{both}$ (n=1,158)
Fetal presentation at birth, n/N(%)	**Non-cephalic**	190/3,331 (5.7%)	117/2,491 (4.7%)	73/840 (8.7%)
Gestational age at birth (weeks), n(%)	**Mean (SD)**	275.2 (17.7)	279.1 (14.7)	263.6 (20.7)
	<28^{+0}	43 (0.9)	20 (0.6%)	23 (2.0%)
	28^{+0} – 33^{+6}	117 (2.5%)	36 (1.0%)	81 (7.0%)
	34^{+0} – 36^{+6}	301 (6.5%)	115 (3.3%)	186 (16.1%)
	37^{+0} – 37^{+6}	373 (8.1%)	174 (5.1%)	199 (17.2%)
	38^{+0} – 38^{+6}	611 (13.3%)	376 (10.9%)	235 (20.3%)
	39^{+0} – 39^{+6}	870 (18.9%)	673 (19.6%)	197 (17.0%)
	40^{+0} or above	2,281 (49.6%)	2,044 (59.5%)	237 (20.5%)
Birthweight customised centile, n(%)	**Mean (SD)**	3.8 (2.8)	4.2 (2.7)	2.5 (2.5)
	<3rd centile	2,070 (45.0%)	1,307 (38.0%)	763 (65.9%)
	3rd – 5th centile	1,043 (22.7%)	848 (24.7%)	195 (16.8%)
	5th-10th centile	1,483 (32.3%)	1,283 (37.3%)	200 (17.3%)
Birthweight population centile, n(%)	**Mean (SD)**	4.75 (2.79)	5.1 (2.8)	3.8 (2.7)
	<3rd centile	1,451 (31.6%)	924 (26.9%)	527 (45.5%)
	3rd – 5th centile	1,035 (22.5%)	772 (22.5%)	263 (22.7%)
	5th-10th centile	2,110 (45.9%)	1,742 (50.7%)	368 (31.8%)
Birthweight allocated centile, n(%)	**Mean (SD)**	4.7 (2.8)	5.0 (2.8)	3.63 (2.69)
	<3rd centile	1,525 (33.2%)	962 (28.0%)	563 (48.6%)
	3rd – 5th centile	1,016 (22.1%)	764 (22.2%)	252 (21.8%)
	5th-10th centile	2,055 (44.7%)	1,712 (49.8%)	343 (29.6%)

Table 10.3 - Unadjusted and adjusted odds ratios comparing demographic or clinical characteristics of women with missed SGA$_{both}$ to women in whom SGA$_{both}$ was antenatally detected (complete case analysis).

		Missed SGA$_{both}$ (3,438)	Detected SGA$_{both}$ (1,158)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)	Adjusted p value
Age (years), n(%)	**≤40y**	3,277 (75.1%)	1,088 (24.9%)	Ref	Ref	Ref
	>40y	161 (69.7%)	70 (30.3%)	0.76 (0.57-1.02)	0.77 (0.56-1.05)	0.10
IMD, n(%)	**1=least deprived**	329 (70.6%)	137 (29.4%)	Ref	Ref	Ref
	2	368 (71.6%)	146 (28.4%)	1.05 (0.80-1.38)	0.95 (0.71-1.28)	0.74
	3	816 (74.7%)	276 (25.3%)	1.23 (0.97-1.57)	1.14 (0.87-1.50)	0.34
	4	1,194 (76.9%)	358 (23.1%)	1.39 (1.10-1.75)	1.20 (0.91-1.57)	0.19
	5=most deprived	731 (75.2%)	241 (24.8%)	1.26 (0.99-1.62)	1.08 (0.81-1.46)	0.60
Ethnicity, n(%)	**White**	1,612 (76.4%)	497 (23.6%)	Ref	Ref	Ref
	Black	601 (74.4%)	207 (25.6%)	0.90 (0.74-1.08)	0.83 (0.67-1.04)	0.11
	Asian	773 (70.1%)	329 (29.9%)	0.72 (0.62-0.85)	0.68 (0.56-0.83)	<0.001
	Mixed	70 (70.0%)	30 (30.0%)	0.72 (0.46-1.12)	0.77 (0.49-1.23)	0.28
	Other	382 (80.1%)	95 (19.9%)	1.24 (0.97-1.59)	0.84 (0.64-1.11)	0.23
BMI (kg/m²), n(%)	**<18.5**	105 (61.4%)	66 (38.6%)	0.53 (0.39-0.73)	0.56 (0.40-0.78)	0.001
	18.5-24.9	1,744 (74.9%)	584 (25.1%)	Ref	Ref	Ref
	25.0-29.9	993 (76.4%)	306 (23.6%)	1.09 (0.93-1.27)	1.15 (0.97-1.36)	0.12
	30.0-34.9	393 (74.7%)	133 (25.3%)	0.99 (0.80-1.23)	1.08 (0.85-1.37)	0.51
	35.0-39.9	142 (76.8%)	43 (23.2%)	1.11 (0.78-1.58)	1.18 (0.81-1.73)	0.38
	≥40.0	61 (70.1%)	26 (29.9%)	0.79 (0.49-1.26)	0.95 (0.57-1.57)	0.83
Parity, n(%)	**0**	1,747 (74.7%)	592 (25.3%)	Ref	Ref	Ref
	1	1,039 (73.8%)	368 (26.2%)	0.96 (0.82-1.11)	0.89 (0.75-1.06)	0.18
	2	374 (77.0%)	112 (23.0%)	1.13 (0.90-1.43)	1.12 (0.87-1.45)	0.36
	3	162 (74.3%)	56 (25.7%)	0.98 (0.71-1.35)	0.89 (0.63-1.26)	0.53
	4 or above	116 (79.5%)	30 (20.5%)	1.31 (0.87-1.98)	1.27 (0.81-1.99)	0.29
Smoking, n(%)	**Non-smoker**	3,108 (75.3%)	1,017 (24.7%)	Ref	Ref	Ref
	Smoker	330 (70.1%)	141 (29.9%)	0.77 (0.62-0.94)	0.72 (0.56-0.91)	0.01

Table 10.4 - Unadjusted and adjusted odds ratios or mean differences comparing co-morbidities or obstetric factors of women with missed SGA$_{both}$ to women in whom SGA$_{both}$ was antenatally detected (complete case analysis).

		Missed SGA$_{both}$ (3,438)	Detected SGA$_{both}$ (1,158)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/ mean diff (95% CI)	Adjusted p value
Co-morbidities, n(%)	No hypertension	3377 (75.1%)	1,121 (24.9%)	Ref	Ref	Ref
	Hypertension	61 (62.2%)	37 (37.8%)	0.55 (0.36-0.83)	0.64 (0.40-1.00)	0.049
	No diabetes	3,385 (74.8%)	1,140 (25.2%)	Ref	Ref	Ref
	Diabetes	53 (74.6%)	18 (25.4%)	0.99 (0.58-1.70)	1.01 (0.56-1.83)	0.96
Antenatal complications, n(%)	No pre-eclampsia	3,342 (75.3%)	1,096 (24.7%)	Ref	Ref	Ref
	Pre-eclampsia	96 (60.8%)	62 (39.2%)	0.51 (0.37-0.70)	0.61 (0.43-0.86)	0.01
	No PIH	3,370 (75.0%)	1,124 (25.0%)	Ref	Ref	Ref
	PIH	68 (66.7%)	34 (33.3%)	0.67 (0.44-1.01)	0.75 (0.48-1.18)	0.22
	No GDM	3,267 (75.4%)	1,065 (24.6%)	Ref	Ref	Ref
	GDM	171 (64.8%)	93 (35.2%)	0.60 (0.46-0.78)	0.60 (0.45-0.80)	<0.001
PAPP-A, n/N(%)±	<0.3 MoM	65 (58.6%)	46 (41.4%)	0.50 (0.34-0.73)	0.69 (0.42-1.14)	0.15
	0.3-0.415 MoM	111 (59.4%)	76 (40.6%)	0.51 (0.38-0.70)	0.70 (0.46-1.51)	0.08
	>0.415MoM	1,557 (74.0%)	548 (26.0%)	Ref	Ref	Ref
	Missing data	1,382 (79.1%)	365 (20.9%)	2.84 (2.58-3.13)	1.12 (0.80-1.56)	0.52
Any indication for serial growth scans+, n/N(%)	No indication	2,128 (78.7%)	576 (21.3%)	Ref	Ref	Ref
	Any indication	987 (68.3%)	459 (31.7%)	0.58 (0.50-0.67)	0.60 (0.51-0.70)	<0.001
	Missing data	323	123	N/A	N/A	N/A
Fetal presentation at birth, n/N(%)§	Cephalic	2,374 (75.6%)	767 (24.4%)	Ref	Ref	Ref
	Non-cephalic	117 (61.6%)	73 (38.4%)	0.52 (0.38-0.70)	0.56 (0.41-0.78)	<0.001
Allocated centile at birth, mean (SD) *		5.0 (2.8)	3.6 (2.7)	1.29 (1.25-1.33)	1.22 (1.18-1.25)	<0.001

Change in OR with a one centile increase (<10th centile).

+Adjusted only for characteristics not included in the composite (IMD, parity, ethnicity, and allocated birthweight centile).

± Population restricted to 3,115 in cases and 1,035 in controls.

§ Population restricted to 2,491 in cases and 840 in controls.

Table 10.5 - Patterns of ultrasound utilisation for pregnant women and their SGA_{both} babies, by presence of a recorded indication for serial fetal growth scans and detection status of SGA

	Pregnancies with any indication for serial growth scans					Pregnancies with no known indication for serial growth scans*				
	Missed SGA_{both} (n=987)	Detected SGA_{both} (n=459)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/mean diff (95% CI)	Adjusted p value	Missed SGA_{both} (n=2,128)	Detected SGA_{both} (n=576)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/mean diff (95% CI)	Adjusted p value
Frequency of screening scans – one scan every n weeks, mean n (SD)*	6.7 (3.4)	3.9 (2.0)	2.8 (2.3-3.3)	2.8 (2.3-3.3)	<0.001	6.6 (3.5)	4.0 (2.6)	2.7 (2.2-3.1)	2.7 (2.2-3.2)	<0.001
Scan frequency for pregnancies with at least two scans, n(%): ≤3-weekly	85 (48.3%)	91 (51.7%)	0.88 (0.56-1.38)	0.81 (0.50-1.30)	0.375	119 (51.7%)	111 (48.3%)	0.89 (0.59-1.33)	0.83 (0.54-1.28)	0.40
4-weekly	66 (51.6%)	62 (48.4%)	Ref	Ref	Ref	87 (54.7%)	72 (45.3%)	Ref	Ref	Ref
≥5-weekly	393 (86.2%)	63 (13.8%)	5.86 (3.79-9.07)	5.96 (3.80-9.34)	<0.001	434 (85.1%)	76 (14.9%)	4.73 (3.18-7.02)	5.20 (3.44-7.84)	<0.001
If scan performed, gestation at the time of the first in weeks, mean (SD)	30.6 (3.9)	29.6 (3.8)	1.0 (0.6-1.5)	1.2 (0.7-1.6)	<0.001	31.9 (4.5)	31.4 (4.4)	0.6 (0.1-1.0)	0.6 (0.2-1.1)	0.01
Duration from the last scan until birth in days Mean (SD)	25.8 (19.2)	9.8 (10.5)	16.0 (14.1-17.9)	15.2 (13.3-17.1)	<0.001	29.4 (23.5)	10.8 (13.3)	18.6 (16.5-20.7)	18.5 (16.4-20.5)	<0.001
Duration by gestational age at birth: <28+0	N/A	5.4 (5.1)	N/A	N/A	N/A	1.5 (2.1)	4.9 (4.0)	-3.4 (-10.1-3.3)	-3.4 (-12.5-5.6)	0.38
28+0 - 30+6	N/A	9.3 (7.4)	N/A	N/A	N/A	37.0 (0.0)	4.1 (4.9)	32.9 (20.6-45.2)	34.5 (19.7-49.3)	0.002
31+0 - 33+6	14.6 (9.5)	5.7 (4.8)	8.8 (4.0-13.7)	5.8 (-1.3-12.9)	0.104	18.7 (14.4)	10.9 (13.0)	7.8 (-5.4-21.0)	9.8 (-3.2-22.84)	0.13
34+0 - 36+6	13.3 (13.7)	5.1 (6.8)	8.1 (4.5-11.8)	6.7 (3.1-10.4)	<0.001	16.4 (18.8)	5.8 (6.1)	10.6 (5.9-15.3)	12.1 (6.8-17.3)	<0.001
37+0 - 38+6	17.3 (14.7)	8.6 (6.4)	8.7 (6.4-11.1)	8.6 (6.1-11.0)	<0.001	17.6 (16.0)	7.7 (7.5)	10.0 (7.5-12.4)	10.4 (7.9-12.9)	<0.001
≥39+0	29.5 (19.6)	16.7 (15.1)	12.8 (9.1-16.5)	12.2 (8.6-15.8)	<0.001	32.2 (24.0)	15.3 (17.0)	16.9 (13.7-20.0)	16.1 (13.0-19.2)	<0.001

*Includes records in which PAPP-A was either ≥0.415 MoM, or missing.
- Too few observations in either group, unable to generate mean difference.
* Negative values suggest that fetal scan occurred after birth, these are caused by imputation estimates of gestational age at birth in a small sample.

390

Table 10.6 - Unadjusted and adjusted odds ratios comparing demographic or clinical characteristics of women with missed SGA$_{pop}$ to women in whom SGA$_{pop}$ was antenatally detected (sensitivity analysis).

		Missed SGA$_{both}$ (66.3%)	Detected SGA$_{both}$ (33.7%)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)	Adjusted p value
Age (years), %	≤40y	66.4%	33.6%	Ref	Ref	Ref
	>40y	65.1%	34.9%	0.96 (0.75-1.24)	1.00 (0.76-1.31)	0.98
IMD, %	1=least deprived	65.1%	34.9%	Ref	Ref	Ref
	2	64.5%	35.5%	0.97 (0.77-1.22)	0.97 (0.76-1.24)	0.80
	3	67.9%	32.1%	1.15 (0.94-1.40)	1.15 (0.92-1.44)	0.21
	4	66.9%	33.1%	1.09 (0.90-1.31)	1.04 (0.84-1.29)	0.72
	5=most deprived	64.9%	35.1%	0.99 (0.81-1.22)	0.97 (0.76-1.23)	0.78
Ethnicity, %	White	68.0%	32.0%	Ref	Ref	Ref
	Black	66.6%	33.4%	0.95 (0.81-1.11)	1.00 (0.83-1.20)	0.995
	Asian	64.2%	35.8%	0.85 (0.75-0.97)	0.78 (0.67-0.92)	0.002
	Mixed	64.8%	35.2%	0.87 (0.57-1.31)	0.94 (0.61-1.45)	0.79
	Other	66.1%	33.9%	0.93 (0.77-1.13)	0.81 (0.65-1.02)	0.07
BMI (kg/m²), %	<18.5	60.8%	39.2%	0.78 (0.63-0.98)	0.77 (0.61-0.98)	0.03
	18.5-24.9	66.1%	33.9%	Ref	Ref	Ref
	25.0-29.9	67.1%	32.9%	1.05 (0.91-1.20)	1.12 (0.97-1.30)	0.12
	30.0-34.9	67.4%	32.6%	1.08 (0.88-1.32)	1.19 (0.96-1.49)	0.12
	35.0-39.9	69.1%	30.9%	1.16 (0.86-1.55)	1.25 (0.92-1.70)	0.16
	≥40.0	69.3%	30.7%	1.16 (0.75-1.78)	1.31 (0.84-2.04)	0.24
Parity, %	0	66.5%	33.5%	Ref	Ref	Ref
	1	65.8%	34.2%	0.98 (0.86-1.11)	0.88 (0.76-1.01)	0.06
	2	67.3%	32.7%	1.06 (0.86-1.30)	0.98 (0.78-1.22)	0.83
	3	62.5%	37.5%	0.84 (0.63-1.13)	0.77 (0.56-1.06)	0.11
	4 or above	69.5%	30.5%	1.15 (0.80-1.64)	1.14 (0.78-1.68)	0.50
Smoking, %	Non-smoker	66.6%	33.4%	Ref	Ref	Ref
	Smoker	63.5%	36.5%	0.87 (0.73-1.03)	0.84 (0.69-1.03)	0.09

Table 10.7 - Unadjusted and adjusted odds ratios or mean differences comparing co-morbidities or obstetric factors of women with missed SGA$_{pop}$ to women in whom SGA$_{pop}$ was antenatally detected (sensitivity analysis).

		Missed SGA$_{both}$ (66.3%)	Detected SGA$_{both}$ (33.7%)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/ mean diff (95% CI)	Adjusted p value
Co-morbidities, %	No hypertension	66.5%	33.5%	Ref	Ref	Ref
	Hypertension	57.4%	42.6%	0.68 (0.49-0.97)	0.70 (0.48-1.02)	0.06
	No diabetes	66.4%	33.6%	Ref	Ref	Ref
	Diabetes	59.7%	40.3%	0.75 (0.51-1.10)	0.77 (0.51-1.17)	0.22
Antenatal complications, %	No pre-eclampsia	67.3%	32.7%	Ref	Ref	Ref
	Pre-eclampsia	45.3%	54.7%	0.42 (0.33-0.53)	0.45 (0.34-0.59)	<0.001
	No PIH	66.7%	33.3%	Ref	Ref	Ref
	PIH	53.0%	47.0%	0.56 (0.41-0.76)	0.59 (0.42-0.83)	0.002
	No GDM	66.1%	33.9%	Ref	Ref	Ref
	GDM	68.7%	31.3%	1.13 (0.92-1.39)	0.99 (0.79-1.23)	0.92
PAPP-A, %	<0.3 MoM	53.9%	46.1%	0.55 (0.40-0.75)	0.63 (0.45-0.89)	0.01
	0.3-0.415 MoM	56.1%	43.9%	0.60 (0.45-0.79)	0.66 (0.49-0.88)	0.01
	>0.415MoM	68.2%	31.8%	Ref	Ref	Ref
	Missing data	66.4%	33.6%	0.93 (0.82-1.05)	0.90 (0.76-1.07)	0.24
Any indication for serial growth scans+, %	No indication	68.8%	31.2%	Ref	Ref	Ref
	Any indication	62.3%	37.7%	0.76 (0.67-0.86)	0.78 (0.68-0.89)	<0.001
Fetal presentation at birth, %	Cephalic	67.9%	32.1%	Ref	Ref	Ref
	Non-cephalic	49.8%	50.2%	0.47 (0.37-0.60)	0.55 (0.43-0.71)	<0.001
Population centile at birth, mean (SD)		5.7 (2.7)	4.0 (2.8)	1.24 (1.21-1.26)	1.24 (1.22-1.27)	<0.001

*Change in OR with a one centile increase (<10th centile).

+Adjusted only for characteristics not included in the composite (IMD, parity, ethnicity and allocated birthweight centile).

Table 10.8- Patterns of ultrasound screening for fetal growth anomalies, by detection status of SGA_{pop} (sensitivity analysis)

		Missed SGA$_{both}$ (65.8%)	Detected SGA$_{both}$ (34.2%)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/mean diff (95% CI)	p value
Frequency of screening scans - one scan every *n* weeks, mean *n* (SD)		6.6 (3.3)	3.8 (2.2)	2.8 (2.6-3.0)	2.8 (2.6-2.9)	<0.001
Scan frequency for pregnancies with at least two scans:	**≤3-weekly**	45.7%	54.3%	0.70 (0.59-0.82)	0.69 (0.58-0.81)	<0.001
	4-weekly	54.7%	45.3%	Ref	Ref	Ref
	≥5-weekly	87.7%	12.3%	5.90 (5.04-6.91)	6.01 (5.12-7.06)	<0.001
If scan performed, gestation at the time of the first in weeks, mean (SD)		31.8 (4.2)	30.9 (4.3)	0.9 (0.7-1.1)	0.9 (0.-7-1.1)	<0.001
Duration from the last scan until birth in days	**Mean (SD)**	28.2 (21.4)	10.2 (12.5)	18.0 (17.2-18.7)	17.9 (17.1-18.6)	<0.001
	Median (IQR)	25 (13-37)	7 (2-14)	N/A	N/A	N/A
Duration by gestational age at birth:	**<28^{+0}**	0.8 (23.4)	3.8 (3.7)	N/A	N/A	N/A
	28^{+0} – 30^{+6}	11.8 (13.1)	4.6 (5.3)			
	31^{+0} – 33^{+6}	15.6 (15.1)	5.3 (8.5)			
	34^{+0} – 36^{+6}	15.6 (15.6)	5.9 (7.2)			
	37^{+0} – 38^{+6}	18.5 (16.4)	8.2 (7.9)			
	≥39^{+0}	30.7 (21.6)	15.8 (16.5)			

Table 10.9 - Characteristics of the SGA$_{both}$ pregnancies, stratified by whether the woman had an anomaly scan recorded at the same site at which she later gave birth (imputed data)

		Anomaly scan (78.6%)	No recorded anomaly scan (21.4%)
Age (years)	Mean (SD)	31.1 (5.6)	30.4 (5.8)
	Age over 40y, %	4.5%	4.3%
IMD, %	1=least deprived	9.8%	8.1%
	2	11.6%	11.9%
	3	24.3%	22.7%
	4	34.8%	35.6%
	5=most deprived	19.4%	21.6%
Ethnicity, %	White	44.0%	41.9%
	Black	16.2%	18.1%
	Asian	25.9%	22.8%
	Mixed	2.0%	2.1%
	Other	11.9%	15.1%
BMI (kg/m²)	Mean (SD)	25.4 (5.4)	26.0 (5.9)
	Under 18.5, %	4.4%	4.1%
	18.5-24.9, %	50.9%	45.3%
	25.0-29.9, %	27.8%	30.6%
	30.0-34.9, %	11.0%	13.2%
	35.0-39.9, %	3.9%	4.5%
	40.0 or above, %	1.8%	2.3%
Parity, %	0	54.1%	57.1%
	1	29.1%	24.4%
	2	10.0%	10.5%
	3	4.0%	4.4%
	4 or higher	2.8%	3.6%
Smoking, %	Smoker	10.0%	11.6%
Co-morbidities, %	Hypertension	2.1%	3.3%
	Diabetes	1.5%	1.4%
	APLS	0.1%	0.0%
Antenatal complications, %	Pre-eclampsia	4.0%	5.6%
	PIH	2.7%	2.6%
	GDM	5.7%	3.6%
PAPP-A, %	<0.300MoM	3.0%	0.1%
	0.3-0.415MoM	4.7%	0.3%
	>0.415MoM	51.2%	5.0%
	Missing data	41.0%	94.6%
Indication for serial fetal scans,+ %	Any indication	65.4%	66.1%
	No indication	34.6%	33.9%
Fetal presentation, %	Non-cephalic	5.8%	8.2%
	Cephalic	94.2%	91.8%
Birthweight allocated centile	Mean (SD)	4.6 (2.8)	4.1 (2.9)
	<3rd centile, %	34.8%	42.0%
	3rd – 5th centile, %	21.8%	20.0%
	5th-10th centile, %	43.3%	38.0%

*No recorded indication, complete case data except that information on PAPP-A may be missing.

Table 10.10 - Unadjusted and adjusted odds ratios comparing demographic or clinical characteristics of women with missed SGA$_{both}$ to women in whom SGA$_{both}$ was antenatally detected, restricted to women with a record of an anomaly scan (sensitivity analysis).

		Missed SGA$_{both}$ (74.1%)	Detected SGA$_{both}$ (25.9%)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)	Adjusted p value
Age (years), %	≤40y	74.3%	25.7%	Ref	Ref	Ref
	>40y	67.9%	32.1%	0.74 (0.56-0.98)	0.77 (0.57-1.03)	0.08
IMD, %	1=least deprived	72.0%	28.0%	Ref	Ref	Ref
	2	72.6%	27.4%	1.03 (0.81-1.31)	0.94 (0.73-1.22)	0.66
	3	75.3%	24.7%	1.19 (0.97-1.47)	1.12 (0.89-1.42)	0.34
	4	75.0%	25.0%	1.16 (0.95-1.42)	1.04 (0.83-1.31)	0.74
	5=most deprived	72.7%	27.3%	1.04 (0.83-1.29)	0.99 (0.77-1.28)	0.93
Ethnicity, %	White	76.5%	23.5%	Ref	Ref	Ref
	Black	71.8%	28.2%	0.78 (0.67-0.93)	0.81 (0.66-0.98)	0.03
	Asian	70.1%	29.9%	0.72 (0.63-0.83)	0.70 (0.60-0.83)	<0.001
	Mixed	67.6%	32.4%	0.65 (0.42-0.99)	0.68 (0.44-1.06)	0.09
	Other	77.9%	22.1%	1.10 (0.90-1.35)	0.80 (0.64-1.01)	0.06
BMI (kg/m²), %	<18.5	60.4%	39.6%	0.53 (0.40-0.70)	0.54 (0.40-0.71)	<0.001
	18.5-24.9	74.0%	26.0%	Ref	Ref	Ref
	25.0-29.9	76.0%	24.0%	1.11 (0.96-1.28)	1.19 (1.03-1.39)	0.02
	30.0-34.9	75.2%	24.8%	1.07 (0.87-1.31)	1.19 (0.95-1.50)	0.12
	35.0-39.9	74.6%	25.4%	1.03 (0.76-1.40)	1.16 (0.83-1.62)	0.38
	≥40.0	71.2%	28.8%	0.86 (0.55-1.35)	1.04 (0.64-1.67)	0.88
Parity, %	0	73.9%	26.1%	Ref	Ref	Ref
	1	73.9%	26.1%	1.01 (0.88-1.15)	0.89 (0.77-1.02)	0.10
	2	75.4%	24.6%	1.10 (0.90-1.35)	1.02 (0.82-1.28)	0.85
	3	71.1%	28.9%	0.86 (0.65-1.15)	0.80 (0.58-1.09)	0.16
	4 or above	77.3%	22.7%	1.19 (0.82-1.73)	1.16 (0.77-1.74)	0.47
Smoking, %	Non-smoker	74.4%	25.6%	Ref	Ref	Ref
	Smoker	70.5%	29.5%	0.82 (0.68-0.98)	0.75 (0.61-0.92)	0.01

Table 10.11 - Unadjusted and adjusted odds ratios and mean differences comparing co-morbidities or obstetric factors of women with missed SGA$_{both}$ to women with detected SGA$_{both}$, restricted to women with a record of an anomaly scan (sensitivity analysis).

		Missed SGA$_{both}$ (74.1%)	Detected SGA$_{both}$ (25.9%)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/ mean diff (95% CI)	Adjusted p value
Co-morbidities, %	No hypertension	74.3%	25.7%	Ref	Ref	Ref
	Hypertension	61.3%	38.7%	0.56 (0.39-0.80)	0.62 (0.42-0.91)	0.02
	No diabetes	74.2%	25.8%	Ref	Ref	Ref
	Diabetes	62.4%	37.6%	0.57 (0.37-0.88)	0.53 (0.33-0.85)	0.01
Antenatal complications, %	No pre-eclampsia	74.9%	25.1%	Ref	Ref	Ref
	Pre-eclampsia	54.5%	45.5%	0.41 (0.32-0.54)	0.45 (0.35-0.60)	<0.001
	No PIH	74.3%	25.7%	Ref	Ref	Ref
	PIH	63.4%	36.6%	0.59 (0.43-0.81)	0.66 (0.47-0.94)	0.02
	No GDM	74.6%	25.4%	Ref	Ref	Ref
	GDM	65.6%	34.4%	0.65 (0.52-0.82)	0.61 (0.48-0.78)	<0.001
PAPP-A, %	<0.3 MoM	53.5%	46.5%	0.40 (0.29-0.56)	0.42 (0.29-0.60)	<0.001
	0.3-0.415 MoM	60.4%	39.6%	0.53 (0.40-0.70)	0.56 (0.41-0.75)	<0.001
	>0.415MoM	74.4%	25.6%	Ref	Ref	Ref
	Missing data	76.3%	23.7%	1.12 (0.98-1.28)	0.88 (0.72-1.07)	0.19
Any indication for serial growth scans+, %	No indication	77.9%	22.1%	Ref	Ref	Ref
	Any indication	66.3%	33.7%	0.57 (0.50-0.64)	0.57 (0.50-0.66)	<0.002
Fetal presentation at birth, %	Cephalic	75.5%	24.5%	Ref	Ref	Ref
	Non-cephalic	60.6%	39.4%	0.50 (0.39-0.63)	0.55 (0.42-0.71)	<0.001
Allocated centile at birth, mean (SD)		4.9 (2.8)	3.5 (2.6)	1.3 (1.26-1.32)	1.2 (1.2-1.3)	<0.001

*Change in OR with a one centile increase (<10th centile).

+ No recorded indication, complete case data except that information on PAPP-A may be missing. Adjusted only for characteristics not included in the composite (IMD, parity, ethnicity and allocated birthweight centile).

Table 10.12- Patterns of ultrasound screening for fetal growth anomalies by detection status of SGA_both, restricted to women with a record of an anomaly scan (sensitivity analysis).

		Missed SGA_both (72.4%)	Detected SGA_both (27.6%)	Unadjusted OR/mean diff (95% CI)	Adjusted OR/mean diff (95% CI)	p value
Frequency of screening scans - one scan every *n* weeks, mean *n* (SD)		3.9 (1.4)	2.3 (1.1)	1.8 (1.7-1.8)	1.7 (1.7-1.8)	<0.001
Scan frequency for pregnancies with at least two scans, %:	**≤3-weekly**	50.0%	50.0%	0.14 (0.11-0.17)	0.13 (0.11-0.16)	<0.001
	4-weekly	87.9%	12.1%	Ref	Ref	Ref
	≥5-weekly	95.6%	4.4%	3.01 (2.06-4.40)	3.28 (2.22-4.83)	<0.001
If scan performed, gestation at the time of the first in weeks, mean (SD)		31.6 (4.2)	30.7 (4.3)	0.9 (0.7-1.1)	1.0 (0.8-1.1)	<0.001
Duration from the last scan until birth, days	**Mean (SD)**	26.9 (20.3)	9.8 (12.0)	17.3 (16.5-18.1)	17.2 (16.4-18.0)	<0.001
	Median (IQR)	24 (12-36)	7 (2-13)	N/A	N/A	N/A
Duration by gestational age at birth, mean days (SD):	**<28^{+0}**	-1.9 (30.2)	4.1 (4.0)	2.7 (-0.9-6.2)	1.6 (-2.3-5.6)	0.41
	28^{+0} – 30^{+6}	12.8 (14.0)	5.0 (5.5)	10.9 (6.6-15.3)	10.2 (5.8-14.7)	<0.001
	31^{+0} – 33^{+6}	16.0 (15.6)	5.3 (8.9)	11.9 (8.3-15.6)	10.4 (6.7-4.1)	<0.001
	34^{+0} – 36^{+6}	14.1(14.3)	5.6 (6.9)	8.6 (7.0-10.2)	9.0 (7.4-10.6)	<0.001
	37^{+0} – 38^{+6}	17.6 (15.2)	8.0 (7.7)	9.8 (8.7-10.8)	9.7 (8.7-10.7)	<0.001
	≥39^{+0}	29.9 (20.6)	15.5 (16.2)	14.5 (13.2-15.9)	14.2 (12.9-15.5)	<0.001

10.17 CHAPTER 7: SUPPLEMENTARY RESULTS TABLES

Table 10.13 - Number and proportion of babies who were LGA by population, customised or both centile definitions at birth, presented by trial arm and phase (available case data)

		Pre-randomisation phase				Outcome comparison phase			
		Standard Care (n=29,404)		Intervention (n=26,546)		Standard Care (n=13,810)		Intervention (n=11,096)	
LGA$_{both}$, n/%		1693	5.76%	1395	5.26%	665	4.82%	532	4.79%
LGA$_{pop}$, n/%		369	1.25%	303	1.14%	203	1.47%	146	1.32%
LGA$_{cust}$, n/%		852	2.90%	726	2.73%	357	2.59%	307	2.77%
LGA (defined by intervention allocation)*, n/% born during each gestational period.	24^{+1}-27^{+6}	1	0.05%	7	0.33%	1	0.12%	4	0.48%
	28^{+0}-31^{+6}	11	0.53%	14	0.66%	7	0.81%	10	1.19%
	32^{+0}-35^{+6}	57	2.76%	69	3.25%	31	3.57%	27	3.22%
	36^{+0}-37^{+6}	273	13.24%	282	13.30%	136	15.67%	111	13.23%
	38^{+0}-39^{+6}	1038	50.34%	1010	47.62%	466	53.69%	411	48.99%
	40^{+0}-42^{+6}	682	33.07%	738	34.79%	227	26.15%	275	32.78%
	Missing gestational age	0	0.00%	1	0.05%	0	0.00%	1	0.12%

LGA by population centiles in standard care arm and by customised centiles in intervention arm

Table 10.14 - Clinical and sociodemographic characteristics of pregnancies in which the baby was born LGA during the pre-randomisation trial phase, presented by trial arm

		Standard care			Intervention		
		LGA$_{pop}$ (n=396)	LGA$_{cust}$ (n=841)	LGA$_{both}$ (n=1599)	LGA$_{pop}$ (n=323)	LGA$_{cust}$ (n=819)	LGA$_{both}$ (n=1364)
Imputed data							
Age at estimated conception, years, median [IQR]		32.25 (28.38-35.35)	32.04 (27.71-35.26)	32.84 (29.16-35.99)	31.73 (27.36-35.97)	31.23 (27.59-35.13)	32.48 (28.39-36.14)
Ethnicity, n (%)	White	85.53%	51.57%	65.06%	80.77%	38.83%	62.76%
	Black	5.58%	17.23%	16.92%	4.85%	14.74%	14.00%
	Asian	1.85%	25.94%	10.88%	1.94%	39.50%	14.14%
	Mixed	2.21%	0.88%	1.56%	1.91%	0.28%	1.38%
	Other	4.83%	4.39%	5.58%	10.52%	6.65%	7.71%
Index of Multiple Deprivation Quintiles, n (%)	1 (Least deprived)	17.99%	17.82%	20.01%	9.01%	6.33%	9.06%
	2	14.42%	13.01%	13.47%	11.16%	10.08%	10.79%
	3	16.01%	17.69%	15.43%	20.35%	23.02%	23.10%
	4	27.14%	27.07%	28.11%	34.70%	37.09%	31.60%
	5 (Most deprived)	24.44%	24.41%	22.98%	24.78%	23.48%	25.46%
Body Mass Index, kg/m², median [IQR]		28.60 (24.91-33.35)	23.44 (20.96-26.35)	26.06 (22.86-30.52)	28.45 (24.26-33.22)	22.86 (20.70-25.59)	25.56 (22.53-29.74)
n (%)	<18.5	0.90%	5.72%	2.25%	0.82%	6.65%	2.17%
	18.5-25	24.78%	59.46%	40.00%	27.84%	63.33%	44.00%
	25-30	33.38%	24.89%	29.94%	31.60%	22.04%	29.89%
	30-35	19.12%	6.08%	16.98%	22.66%	6.00%	15.40%
	35-40	14.32%	1.64%	7.31%	9.49%	0.96%	5.21%
	≥40	7.50%	2.21%	3.53%	7.58%	1.03%	3.34%
Parity, n (%)	Nulliparous	35.88%	36.96%	29.46%	46.13%	50.72%	45.54%
	1	42.28%	38.79%	42.18%	32.73%	31.99%	32.71%
	2	13.70%	14.33%	15.49%	13.59%	11.39%	13.42%
	3	4.52%	6.12%	7.09%	4.91%	3.84%	4.90%
	≥4	3.62%	3.79%	5.78%	2.64%	2.06%	3.43%

	Standard care			Intervention		
	LGA$_{pop}$ (n=396)	LGA$_{cust}$ (n=841)	LGA$_{both}$ (n=1599)	LGA$_{pop}$ (n=323)	LGA$_{cust}$ (n=819)	LGA$_{both}$ (n=1364)
Non-imputed data						
Smoking in pregnancy, n (%)*	17 (4.93%)	26 (3.29%)	50 (3.17%)	16 (7.14%)	10 (1.62%)	27 (2.82%)
Missing smoking	15 (4.17%)	40 (4.81%)	56 (3.43%)	64 (22.22%)	90 (12.71%)	363 (27.50%)
Pre-existing comorbidities, n (%)*						
Diabetes	7 (3.17%)	11 (2.12%)	32 (3.08%)	6 (2.84%)	19 (3.69%)	43 (4.12%)
Missing diabetes	139 (38.61%)	311 (37.42%)	593 (36.31%)	77 (26.74%)	193 (27.26%)	276 (20.91%)
Hypertension	3 (1.29%)	7 (1.29%)	26 (2.40%)	3 (1.42%)	7 (1.36%)	16 (1.53%)
Missing hypertension	128 (35.56%)	288 (34.66%)	551 (33.74%)	77 (26.74%)	193 (27.26%)	276 (20.91%)
Antenatal complications, n (%)*						
Gestational diabetes (GDM)	26 (8.93%)	25 (3.68%)	88 (6.75%)	20 (9.57%)	36 (7.03%)	104 (10.00%)
Missing GDM	69 (19.17%)	152 (18.29%)	330 (20.21%)	79 (27.43%)	196 (27.68%)	280 (21.21%)
Gestational hypertension	14 (6.33%)	8 (1.47%)	31 (3.06%)	14 (20.59%)	7 (4.55%)	40 (8.60%)
Missing gestational hypertension	139 (38.61%)	285 (34.30%)	621 (38.03%)	220 (76.39%)	554 (78.25%)	855 (64.77%)
Infant sex, male, n (%)*	58 (16.11%)	686 (82.55%)	879 (53.83%)	38 (13.19%)	592 (83.62%)	742 (56.21%)
Missing Infant sex	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)

*Mothers and babies with missing data have been excluded from the denominator

Table 10.15 - Clinical and sociodemographic characteristics of pregnancies in which the baby was born LGA during the pre-randomisation trial phase, presented by trial arm (available case data where not already presented in main text)

	Standard care			Intervention		
	LGA$_{pop}$ (n=396)	LGA$_{cust}$ (n=841)	LGA$_{both}$ (n=1599)	LGA$_{pop}$ (n=323)	LGA$_{cust}$ (n=819)	LGA$_{both}$ (n=1364)
Age at estimated conception, years, median [IQR]*	32.25 (28.29-35.33)	32.23 (27.85-35.27)	32.83 (29.16-35.95)	32.07 (27.61-36.33)	31.15 (27.63-35.00)	32.43 (27.95-35.95)
Missing age, n(%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Ethnicity, n (%)* White	301 (86.74%)	410 (50.31%)	1024 (64.73%)	183 (81.33%)	215 (33.39%)	596 (58.55%)
Black	19 (5.48%)	146 (17.91%)	272 (17.19%)	10 (4.44%)	103 (15.99%)	166 (16.31%)
Asian	5 (1.44%)	217 (26.63%)	172 (10.87%)	3 (1.33%)	276 (42.86%)	158 (15.52%)
Mixed	7 (2.02%)	8 (0.98%)	26 (1.64%)	5 (2.22%)	1 (0.16%)	15 (1.47%)
Other	15 (4.32%)	34 (4.17%)	88 (5.56%)	24 (10.67%)	49 (7.61%)	83 (8.15%)
Missing ethnicity	13 (3.61%)	16 (1.93%)	51 (3.12%)	63 (21.88%)	64 (9.04%)	302 (22.88%)
Index of Multiple Deprivation Quintiles, n (%)* 1 (Least deprived)	62 (17.22%)	138 (16.69%)	328 (20.21%)	24 (8.39%)	37 (5.27%)	117 (8.92%)
2	52 (14.44%)	105 (12.70%)	220 (13.56%)	33 (11.54%)	72 (10.26%)	144 (10.98%)
3	54 (15.00%)	149 (18.02%)	253 (15.59%)	61 (21.33%)	164 (23.36%)	294 (22.43%)
4	99 (27.50%)	228 (27.57%)	454 (27.97%)	98 (34.27%)	262 (37.32%)	420 (32.04%)
5 (Most deprived)	93 (25.83%)	207 (25.03%)	368 (22.67%)	70 (24.48%)	167 (23.79%)	336 (25.63%)
Missing IMD	0 (0.00%)	4 (0.48%)	10 (0.61%)	2 (0.69%)	6 (0.85%)	9 (0.68%)

*Mothers and babies with missing data have been excluded from the denominator

	Standard care			Intervention		
	LGApop (n=396)	LGAcust (n=841)	LGAboth (n=1599)	LGApop (n=323)	LGAcust (n=819)	LGAboth (n=1364)
Body Mass Index, kg/m², median [IQR]*	29.02 (25.08-34.18)	23.84 (21.34-26.75)	26.66 (23.34-31.12)	28.37 (23.91-33.22)	22.90 (20.94-25.59)	25.58 (22.64-29.73)
n (%)*						
<18.5	1 (0.35%)	25 (3.96%)	9 (0.76%)	1 (0.45%)	25 (4.55%)	5 (0.60%)
18.5-25	69 (24.47%)	368 (58.23%)	448 (37.77%)	67 (30.04%)	361 (65.76%)	383 (45.70%)
25-30	92 (32.62%)	168 (26.58%)	362 (30.52%)	69 (30.94%)	125 (22.77%)	249 (29.71%)
30-35	53 (18.79%)	42 (6.65%)	221 (18.63%)	51 (22.87%)	30 (5.46%)	132 (15.75%)
35-40	46 (16.31%)	12 (1.90%)	100 (8.43%)	20 (8.97%)	4 (0.73%)	41 (4.89%)
≥40	21 (7.45%)	17 (2.69%)	46 (3.88%)	15 (6.73%)	4 (0.73%)	28 (3.34%)
Missing BMI	78 (21.67%)	199 (23.95%)	447 (27.37%)	65 (22.57%)	159 (22.46%)	482 (36.52%)
Parity, n (%)*						
Nulliparous	123 (39.94%)	290 (40.28%)	473 (33.71%)	108 (44.08%)	308 (48.20%)	467 (44.95%)
1	116 (37.66%)	246 (34.17%)	525 (37.42%)	83 (33.88%)	220 (34.43%)	348 (33.49%)
2	43 (13.96%)	108 (15.00%)	219 (15.61%)	37 (15.10%)	73 (11.42%)	139 (13.38%)
3	14 (4.55%)	47 (6.53%)	102 (7.27%)	12 (4.90%)	26 (4.07%)	51 (4.91%)
≥4	12 (3.90%)	29 (4.03%)	84 (5.99%)	5 (2.04%)	12 (1.88%)	34 (3.27%)
Missing parity	52 (14.44%)	111 (13.36%)	230 (14.08%)	43 (14.93%)	69 (9.75%)	281 (21.29%)

*Mothers and babies with missing data have been excluded from the denominator

Table 10.16 - Clinical and sociodemographic characteristics of pregnancies in which the baby was born LGA during the trial outcome comparison phase, presented by trial arm (available case data where not already presented in main text)

	Standard care			Intervention		
	LGApop (n=221)	LGAcust (n=367)	LGAboth (n=618)	LGApop (n=153)	LGAcust (n=340)	LGAboth (n=506)
Age at estimated conception, years, median [IQR]*	33.09 (29.55-35.90)	32.12 (28.33-35.29)	33.07 (29.28-36.26)	32.69 (29.03-35.87)	31.71 (28.61-34.76)	32.58 (28.76-36.43)
Missing age, n(%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Ethnicity, n (%)*						
White	148 (77.89%)	181 (53.24%)	407 (66.50%)	117 (88.64%)	106 (37.06%)	286 (62.17%)
Black	16 (8.42%)	57 (16.76%)	102 (16.67%)	5 (3.79%)	46 (16.08%)	63 (13.70%)
Asian	2 (1.05%)	78 (22.94%)	54 (8.82%)	3 (2.27%)	116 (40.56%)	74 (16.09%)
Mixed	8 (4.21%)	5 (1.47%)	9 (1.47%)	1 (0.76%)	2 (0.70%)	4 (0.87%)
Other	16 (8.42%)	19 (5.59%)	40 (6.54%)	6 (4.55%)	16 (5.59%)	33 (7.17%)
Missing ethnicity	6 (3.0%)	11 (3.13%)	21 (3.32%)	9 (6.38%)	12 (4.03%)	37 (7.44%)
Index of Multiple Deprivation Quintiles, n (%)*						
1 (Least deprived)	32 (16.41%)	70 (20.17%)	143 (22.92%)	11 (7.80%)	29 (9.86%)	43 (8.83%)
2	25 (12.82%)	56 (16.14%)	89 (14.26%)	17 (12.06%)	28 (9.52%)	60 (12.32%)
3	29 (14.87%)	52 (14.99%)	92 (14.74%)	40 (28.37%)	77 (26.19%)	116 (23.82%)
4	47 (24.10%)	96 (27.67%)	160 (25.64%)	45 (31.91%)	108 (36.73%)	155 (31.83%)
5 (Most deprived)	62 (31.79%)	73 (21.04%)	140 (22.44%)	28 (19.86%)	52 (17.69%)	113 (23.20%)
Missing IMD	1 (0.51%)	4 (1.14%)	9 (1.42%)	0 (0.00%)	4 (1.34%)	10 (2.01%)

*Mothers and babies with missing data have been excluded from the denominator

	Standard care			Intervention		
	LGA$_{pop}$ (n=221)	LGA$_{cust}$ (n=367)	LGA$_{both}$ (n=618)	LGA$_{pop}$ (n=153)	LGA$_{cust}$ (n=340)	LGA$_{both}$ (n=506)
Body Mass Index, kg/m², median [IQR]*	29.22 (25.88-34.37)	24.17 (21.33-27.32)	27.2 (23.62-32.04)	28.58 (24.70-33.40)	23.06 (20.72-24.93)	26.89 (23.77-31.81)
n (%)*						
<18.5	0 (0.00%)	11 (4.04%)	3 (0.64%)	0 (0.00%)	7 (2.89%)	5 (1.32%)
18.5-25	33 (20.00%)	148 (54.41%)	155 (33.19%)	34 (26.15%)	177 (73.14%)	124 (32.63%)
25-30	60 (36.36%)	71 (26.10%)	155 (33.19%)	42 (32.31%)	46 (19.01%)	123 (32.37%)
30-35	37 (22.42%)	30 (11.03%)	95 (20.34%)	30 (23.08%)	9 (3.72%)	71 (18.68%)
35-40	24 (14.55%)	8 (2.94%)	33 (7.07%)	15 (11.54%)	2 (0.83%)	41 (10.79%)
≥40	11 (6.67%)	4 (1.47%)	26 (5.57%)	9 (6.92%)	1 (0.41%)	16 (4.21%)
Missing BMI	31 (15.82%)	79 (22.51%)	166 (26.22%)	11 (7.80%)	56 (18.79%)	117 (23.54%)
Parity, n (%)*						
Nulliparous	71 (40.34%)	111 (38.01%)	171 (33.33%)	52 (36.88%)	119 (40.75%)	176 (36.36%)
1	59 (33.52%)	114 (39.04%)	207 (40.35%)	55 (39.01%)	106 (36.30%)	191 (39.46%)
2	26 (14.77%)	37 (12.67%)	74 (14.42%)	24 (17.02%)	41 (14.04%)	71 (14.67%)
3	10 (5.68%)	18 (6.16%)	36 (7.02%)	4 (2.84%)	20 (6.85%)	26 (5.37%)
≥4	10 (5.68%)	12 (4.11%)	25 (4.87%)	6 (4.26%)	6 (2.05%)	20 (4.13%)
Missing parity	20 (10.20%)	59 (16.81%)	120 (18.96%)	0 (0.00%)	6 (2.01%)	13 (2.62%)

*Mothers and babies with missing data have been excluded from the denominator

Table 10.17 - Rate of detection of LGA by each definition, presented by trial arm and phase, using available case data

	Pre-randomisation phase		Comparison phase		Intervention effect size - unadjusted (95%CI)	Intervention effect size - adjusted (95%CI)	p-value
	Standard Care	GAP	Standard Care	GAP			
Primary outcome							
LGA at birth by customised and population centiles, %(n/N)	5.8% (1,633/28,298)	5.6% (1,081/19,302)	4.8% (633/13,111)	4.9% (388/7,941)	0.5% (-1.3, 2.2)	-0.3% (-2.8, 2.2)	0.77
Antenatal detection, %(n/N)	24.1% (393/1,633)	40.8% (441/1,081)	47.9% (303/633)	39.4% (153/388)	-5.6% (-23.0, 11.8)	-5.2% (-29.6, 19.3)	0.61
Test positive rate, %(n/N)	4.8%	4.2%	3.7%	3.9%			
Secondary outcomes*							
LGA at birth by customised centiles (<10th centile) , %(n/N)	8.7% (2,464/28,301)	8.3% (1,611/19,341)	7.5% (984/13,111)	7.5% (598/7,976)	0.2% (-1.8,2.3)	-0.2% (-3.9, 3.4)	0.87
Antenatal detection, %(n/N)	19.2% (474/2,464)	32.3% (521/1,611)	39.0% (384/984)	37.8% (226/598)	1.1% (-16.9, 19.0)	-0.8% (-23.8, 22.3)	0.94
False positive rate, %(n/N)	3.1% (805/25,837)	2.6% (394/15,274)	6.7% (817/12,127)	4.0% (264/6,638)	-3.1% (-7.2, 1.0)	-2.6% (-6.2, 1.0)	0.12
LGA at birth by population centiles (<10th centile) , %(n/N)	7.0% (1,993/28,302)	6.7% (1,301/19,305)	6.3% (829/13,111)	6.3% (497/7,941)	0.6% (-1.9, 3.1)	-0.5% (-3.3, 2.3)	0.65
Antenatal detection, %(n/N)	23.3% (464/1,993)	39.2% (510/1,301)	45.7% (379/829)	34.8% (173/497)	-9.6% (-22.5, 3.3)	-7.1% (-26.5, 12.4)	0.39
False positive rate, %(n/N)	3.1% (816/26,309)	2.6% (404/15,542)	6.7% (822/12,282)	4.7% (318/6,718)	-2.4% (-6.3, 1.5)	-2.3% (-5.9, 1.3)	0.15

*One site did not contribute data on screening outcomes for the pre-randomisation phase, so excluded from all secondary outcomes

Table 10.18 - Ultrasound utilisation amongst pregnancies in which the baby was born LGA$_{both}$ at or after 36^{+0} weeks of gestation, presented by trial arm and phase (available case data)

	Pre-randomisation phase		Comparison phase		Intervention effect size - unadjusted (95%CI)	Intervention effect size - adjusted (95%CI)	p-value
	Standard Care (n=1633)	Intervention (n=1320)	Standard Care (n=633)	Intervention (n=497)			
Proportion of pregnancies with at least one scan, n (%)	1177 (72.08%)	1218 (92.27%)	598 (94.47%)	471 (94.77%)	0.88% (-6.22, 7.97)	-14.57% (-41.68, 12.18)	0.23
Number of all scans, mean (SD)	3.17 (2.96)	3.66 (2.18)	4.68 (2.89)	3.95 (2.14)	-0.75 (-2.24, 0.74)	-0.86 (-1.27, -0.45)	0.002
Number of scans ≥34^{+0} weeks, mean (SD)	0.92 (1.17)	0.89 (0.83)	1.22 (1.15)	1.01 (0.96)	-0.15 (-0.86, 0.56)	-0.13 (-0.58, 0.31)	0.50
Proportion of pregnancies with scans ≥34^{+0} weeks, n (%)	645 (39.50%)	788 (59.70%)	461 (72.83%)	324 (65.19%)	-6.53% (-24.27, 11.21)	-14.17% (-34.74, 6.40)	0.14
Number of scans ≥34^{+0} with EFW, mean (SD)*	0.48 (0.67)	0.75 (0.70)	0.9 (0.72)	0.75 (0.66)	-0.15 (-0.41, 0.10)	-0.13 (-0.43, 0.18)	0.31
Proportion of pregnancies with scans ≥34^{+0}, including EFW, n (%)*	620 (37.97%)	522 (61.27%)	449 (70.93%)	211 (63.36%)	-0.26% (-30.72, 14.21)	-12.03% (-39.68, 15.63)	0.29
Of women with EFW recorded ≥34+0 weeks, median duration between the last growth scan and the birth	18.24 (12.92)	17.71 (10.84)	17.76 (11.46)	18.24 (10.44)	0.87 (-4.20, 5.94)	-0.53 (-5.93, 4.87)	0.80

*Values for the intervention arm exclude two cluster sites at which data on estimated fetal weight was not available during both trial phases for non-SGA babies.

Table 10.19 - Secondary outcomes for mothers who gave birth to LGA$_{both}$ babies at or after 36^{+0} weeks of gestation, presented by trial arm and phase (available case data).

	Pre-randomisation phase		Comparison phase		Intervention effect size - unadjusted (95%CI)	Intervention effect size - adjusted (95%CI)	p-value
	Standard Care (n=1633)	Intervention (n= 1320)	Standard Care (n= 633)	Intervention (n= 497)			
Induction of Labour, n/N(%)	396/1629 (24.32%)	322/1151 (27.98%)	158/632 (25.00%)	148/479 (30.90%)	8.22% (-5.03, 21.47)	0.78% (-9.41, 10.97)	0.857
Mode of birth, n/N (%)							
Spontaneous vaginal delivery	747/1632 (45.77%)	603/1231 (48.98%)	274/633 (43.29%)	200/472 (42.37%)	1.67% (-11.84, 15.19)	-8.14% (-22.42, 6.14)	0.213
Instrumental delivery	155/1632 (9.50%)	142/1231 (11.54%)	62/633 (9.79%)	43/472 (9.11%)	0.37% (-5.59, 6.34)	-3.75% (-7.90, 0.40)	0.069
Elective caesarean section	412/1632 (25.25%)	308/1231 (25.02%)	187/633 (29.54%)	143/472 (30.30%)	-2.38% (-17.71, 12.94)	-5.56% (-20.59, 9.48)	0.401
Emergency caesarean section	317/1632 (19.42%)	176/1231 (14.30%)	110/633 (17.38%)	86/472 (18.22%)	0.33% (-4.30, 4.97)	-3.02% (-7.31, 1.26)	0.135
Estimated blood loss, mls mean (SD)	625.45 (483.45)	639.97 (486.01)	651.40 (551.20)	647.25 (455.62)	-26.74 (-124.78, 71.30)	2.13 (-84.21, 88.47)	0.954
Post-partum haemorrhage (>1500mls), n/N(%)	90/1621 (5.55%)	58/1291 (4.49%)	38/630 (6.03%)	20/475 (4.21%)	-2.37% (-5.46, 0.73)	-2.40% (-4.77, -0.03)	0.048
3rd/4th degree tears, n/N (%)	35/1633 (2.14%)	40/1320 (3.03%)	8/633 (1.26%)	11/497 (2.21%)	1.42% (-1.46, 4.30)	0.68% (-1.41, 2.76)	0.457
Epidural, n/N (%)	458/1435 (31.92%)	375/1318 (28.45%)	177/556 (31.83%)	141/496 (28.43%)	-5.20% (-26.66, 16.27)	-5.68% (-18.66, 7.30)	0.326
Episiotomy, n/N (%)	165/1358 (12.15%)	167/909 (18.37%)	76/558 (13.62%)	46/319 (14.42%)	13.97% (-11.18, 39.13)	-5.70% (-12.18, 0.78)	0.075

Table 10.20 - Secondary outcomes for LGA$_{both}$ babies born at or after 36^{+0} weeks of gestation, presented by trial arm and phase (available case data).

	Pre-randomisation phase		Comparison phase		Intervention effect size - unadjusted (95%CI)	Intervention effect size - adjusted (95%CI)	p-value
	Standard Care (n=1633)	Intervention (n=1320)	Standard Care (n=633)	Intervention (n=479)			
Gestational age at birth, weeks , mean (SD)	39.31 (1.30)	39.21 (1.34)	39.19 (1.25)	39.26 (1.27)	0.07 (-0.10, 0.24)	0.12 (-0.08, 0.32)	0.197
Babies born before 39^{+0} weeks, n/N (%)	594/1633 (36.37%)	513/1320 (38.86%)	248/633 (39.18%)	185/497 (37.22%)	-2.13% (-10.74, 6.48)	-6.44% (-19.56, 6.68)	0.275
Birthweight, g mean (SD)	4207.22 (351.77)	4180.39 (372.31)	4183.94 (336.98)	4196.82 (334.28)	17.26 (-27.25, 61.76)	32.67 (-17.98, 83.32)	0.166
Apgar score < 7 at 5 minutes, n/N (%)	28/1633 (1.71%)	22/1320 (1.67%)	15/633 (2.37%)	5/497 (1.01%)	-1.12% (-2.55, 0.31)	-1.06% (-2.97,0.85)	0.225
Arterial cord pH < 7.1, n/N (%)	38/1404 (2.71%)	42/1320 (3.18%)	16/552 (2.90%)	15/497 (3.02%)	-0.05% (-2.26, 2.16)	-0.30% (-2.03, 1.43)	0.672
Neonatal unit admission, n/N (%)	216/1333 (16.20%)	146/1320 (11.06%)	104/514 (20.23%)	46/497 (9.26%)	-10.31% (-31.76,11.15)	-6.38% (-14.36, 1.60)	0.095
Hypoxic-Ischaemic encephalopathy, n/N (%)	1/1633 (0.06%)	2/1320 (0.15%)	2/633 (0.32%)	2/497 (0.40%)	0.07% (-0.52, 0.65)	0.32% (-0.46, 1.11)	0.349
Hypoglycaemia, n/N (%)	37/1633 (2.27%)	41/1320 (3.11%)	20/633 (3.16%)	10/497 (2.01%)	-0.61% (-3.22, 2.00)	-0.40% (-3.66, 2.86)	0.773
Nasogastric tube feeding, n/N (%)	16/1333 (1.20%)	38/1320 (2.88%)	8/514 (1.56%)	13/497 (2.62%)	0.88% (-1.35, 3.11)	0.76% (-1.39, 2.92)	0.404
Stillbirth, n/N (%)	0/1633 (0.00%)	2/1320 (0.15%)	1/633 (0.16%)	0/495 (0.00%)	-0.15% (-0.54, 0.23)	#	0.389*
Neonatal death, n/N (%)	0/1633 (0.00%)	0/1320 (0.00%)	0/633 (0.00%)	0/497 (0.00%)	0.00% (0.00, 0.00)	#	1.000*
Perinatal mortality, n/N (%)	0/1633 (0.00%)	2/1320 (0.15%)	1/633 (0.16%)	0/495 (0.00%)	-0.15% (-0.54, 0.23)	#	0.389*

Cannot be calculated due to too few events

*Unadjusted p-value