DOI:
10.1016/j.crad.2017.04.014

Document Version
Peer reviewed version

Clinical Radiology

Diagnostic Yield of FDG-PET/CT in Fever of Unknown Origin: A Systematic Review, Meta-analysis and Delphi Exercise.

--Manuscript Draft--

Manuscript Number:	CRAD-D-17-00145R1
Full Title:	Diagnostic Yield of FDG-PET/CT in Fever of Unknown Origin: A Systematic Review, Meta-analysis and Delphi Exercise.
Article Type:	Original Paper
Corresponding Author:	Tehmina Bharucha University College London London, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University College London
Corresponding Author's Secondary Institution:	
First Author:	Tehmina Bharucha
First Author Secondary Information:	
Order of Authors:	Tehmina Bharucha
	Andrew Rutherford
	Sarah Skoech
	Abass Alavi
	Michael Brown
	James Galloway
Order of Authors Secondary Information:	
Abstract:	Aim: To perform a systematic review, meta-analysis and Delphi exercise to evaluate diagnostic yield of 2-[18F]-fluoro-2-deoxy-D-glucose-Positron-Emission-Tomography/Computed-Tomography (FDG-PET/CT) in Fever of Unknown Origin (FUO). Materials and Methods: Study-ID CRD42016032696. Four databases were searched for studies of FDG-PET/CT in FUO 1/1/2000-1/12/2015. Exclusions were non-English language, case reports, non-standard FDG-radiotracer and significant missing data. Quality was assessed by two authors independently using a standardised tool. Pooled diagnostic yield was calculated using a random-effects model. An iterative electronic and face-to-face Delphi generated interspeciality consensus. Results: Pooled diagnostic yield was 56% (95%CI 50-61%), I2=61%, 18 studies and 905 patients. Only 5 studies reported results of previous imaging, and sub-group analysis estimated diagnostic yield beyond conventional CT at 32% (95%CI 22-44%), I2=66%. Consensus was established that FDG-PET/CT is increasingly available with an emerging role, but there is prevailing variability in practice. Conclusion: There is insufficient evidence to support the value of FDG-PET/CT in investigative algorithms of FUO. We need a paradigm shift in research, involving prospective studies recruiting at diagnosis of FUO, with updated case definitions and hard outcome measures. While these studies will be a significant undertaking with multi-centre collaboration, their completion is vital for balancing both radiation exposure and costs against possible benefits of utilising FDG-PET/CT.

Title

Diagnostic Yield of FDG-PET/CT in Fever of Unknown Origin: A Systematic Review, Meta-analysis and Delphi Exercise.

Authors

Tehmina Bharucha[1,2] t.bharucha@doctors.org.uk

Andrew Rutherford[3,4] arutherford1@nhs.net

Sarah Skeoch[5] sarah.skeoch@manchester.ac.uk

Abass Alavi[6] Abass.Alavi@uphs.upenn.edu

Michael Brown[1,7,8] Michael.Brown@lshtm.ac.uk

James Galloway[4, 9] James.Galloway@kcl.ac.uk

The FDG-PET/CT in FUO working group

Affiliations

1 Division of Infection and Immunity, University College London, London, UK

2 Royal Free Hospital NHS Foundation Trust, London, UK

3 NIHR Guy's and St Thomas' Biomedical Research Centre, London, UK

4 Rheumatology department, King's College London, London, UK

5 Rheumatology department, University of Manchester, Manchester, UK

6 Radiology department, Hospital of the University of Philadelphia, Pennsylvania, USA

7 Faculty of Infectious Diseases and Tropical Medicine, London School of Hygiene and Tropical Medicine, London, UK

8 Hospital for Tropical Diseases, University College London Hospital NHS Foundation Trust, UK

9 Rheumatology department, King's College Hospital NHS Foundation Trust, London, UK

Corresponding author

Dr Tehmina Bharucha, University College London, Division of Infection & Immunity

London, UK. Telephone number: 02031082130

Email: t.bharucha@doctors.org.uk

Acknowledgements

With sincere thanks to the FDG-PET/CT in FUO Working Group: Robert Miller, Martin Llewelyn, Neil Jenkins, John Lambourne, Catherine Cosgrove, Elinor Moore, Chris Conlon, Coaimh NicFhogartaigh, Daniel Agranoff, Andy Ustianowski, Benjamin Parker, Nicola Gullick, Neil Snowden, David Jayne, Marwan Bukhari, Kevin Davies, Wendy Stewart, Kirit Ardeshna, Mohamedbhai Sajir, Dr Jamshed Bomanji, Haroon Athar, Wai-lup Wong, Amy Eccles, Manil Subesinghe, Neel Patel, Fahmid Chowdhury, John Buscombe, Sabina Dizdeveric and Dan Marks.

We are also indebted to the Special Trustees of the Hospital of Tropical Diseases for their generous contribution, and the exceptional support of the late Dr Manek-Phiroz Eddie Bharucha.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The face-to-face meeting was supported by a grant from the Special Trustees of the Hospital of Tropical Diseases

Conflicts of Interest

No authors have any conflicts of interest

Authors Contributions

	TB	AR	SS	AA	MB	JG
1. Guarantor of integrity of the entire study	+					
2. Study concepts and design	+	+			+	+
3. Literature research	+	+	+	+	+	+
4. Clinical studies	N/A	N/A	N/A	N/A	N/A	N/A
5. Experimental studies / data analysis	N/A	N/A	N/A	N/A	N/A	N/A
6. Statistical analysis	+	+	+		+	+
7. Manuscript preparation	+	+	+		+	+
8. Manuscript editing	+	+	+	+	+	+

Abstract

Aim: To perform a systematic review, meta-analysis and Delphi exercise to evaluate diagnostic yield of 2-[18F]-fluoro-2-deoxy-D-glucose-Positron-Emission-Tomography/Computed-Tomography (FDG-PET/CT) in Fever of Unknown Origin (FUO).

Materials and Methods: Study-ID CRD42016032696. Four databases were searched for studies of FDG-PET/CT in FUO 1/1/2000-1/12/2015. Exclusions were non-English language, case reports, non-standard FDG-radiotracer and significant missing data. Quality was assessed by two authors independently using a standardised tool. Pooled diagnostic yield was calculated using a random-effects model. An iterative electronic and face-to-face Delphi generated interspeciality consensus.

Results: Pooled diagnostic yield was 56% (95%CI 50-61%), I2=61%, 18 studies and 905 patients. Only 5 studies reported results of previous imaging, and sub-group analysis estimated diagnostic yield beyond conventional CT at 32% (95%CI 22-44%), I2=66%. Consensus was established that FDG-PET/CT is increasingly available with an emerging role, but there is prevailing variability in practice.

Conclusion: There is insufficient evidence to support the value of FDG-PET/CT in investigative algorithms of FUO. We need a paradigm shift in research, involving prospective studies recruiting at diagnosis of FUO, with updated case definitions and hard outcome measures. While these studies will be a significant undertaking with multi-centre collaboration, their completion is vital for balancing both radiation exposure and costs against possible benefits of utilising FDG-PET/CT.

Key words

Imaging, Nuclear Medicine, Fever of Unknown Origin, Diagnostics

Abbreviations

CI	Confidence Intervals
FDG-PET/CT	2-[18F]-fluoro-2-deoxy-D-glucose - Positron Emission Tomography/Computed Tomography
FUO	Fever of Unknown Origin
IQR	Interquartile Range
IUO	Inflammation of Unknown Origin
KPI	Key Performance Indicator

Introduction

Fever as an isolated clinical presentation has challenged clinicians for decades[1, 2]. In 1961 Petersdorf and Beeson provided a case definition for 'fever (or pyrexia) of unknown origin': 1) a body temperature above 38.3°C; 2) on several occasions; with 3) a duration of illness of at least three weeks; and 4) no diagnosis within one week of hospital admission[2-4]. Fifty years on, definitions of FUO and the spectrum of aetiologies have evolved, however the diagnostic challenges remain[4]. FUO represents an estimated 2.9% of hospital admissions, with morbidity associated with prolonged hospital stay, repeated cycles of invasive investigations and presumptive treatment, mortality rates between 12-35%, and cost implications[5].

2-[18F]-fluoro-2-deoxy-D-glucose (FDG)-Positron-Emission-Tomography/Computed-Tomography (PET/CT) emerged at the end of the 20th century as an amalgamation between functional and conventional anatomical imaging[6]. Its role in oncological staging has been well-defined, however in other specialities there is less clarity[7]. Specifically, in the investigation of FUO the role of FDG-PET/CT in clinical practice and diagnostic algorithms is inconsistent and unestablished. Existing guidelines suggest that FDG-PET/CT *may* be used where conventional investigations have not revealed a source[8].

FDG-PET/CT is not associated with nephrotoxicity, and standard protocols expose patients to less radiation than a conventional CT. An average FDG-PET/CT scan exposes a patient to 15mSv radiation, approximately 5-6 years background radiation, rather than 20-25mSv in a contrast-enhanced chest-abdomen-pelvis CT. Other advantages include imaging areas (e.g. head and neck, extremities) which are beyond the range of most CT scans used in this context, and detection of vascular and truncal musculoskeletal inflammation for which cross-sectional contrast CT imaging is insensitive. The main caveats are cost and accessibility, FDG-PET/CT costing £800, compared to £250 for a contrast-enhanced chest-abdomen-pelvis CT. However this could easily be remunerated by earlier definitive

treatment associated with additional diagnostic sensitivity. A marginally reduced length of inpatient stay could mitigate the cost, with an average £400 for one night hospital admission[9].

Current literature evaluating the role of FDG-PET/CT in FUO is based on observational data involving small samples, outdated case definitions, and poor generalisability. Outcomes reported by existing meta-analyses focus on sensitivity of FDG-PET/CT in FUO[10, 11]. Sensitivity refers to the proportion of cases with a diagnosis to explain the FUO for which FDG-PET/CT contributed to the diagnosis, or A/(A+B) (Table 1). This is statistically inappropriate as there is no reference standard for the investigation of FUO to enable estimates of diagnostic accuracy[12]. In comparison, 'diagnostic yield' provides a more suitable outcome measure, calculated as the proportion of all FDG-PET/CT scans (both normal and abnormal) that contribute to the diagnosis of FUO, A/(A+B+C+D) (Table 1)[13]. Strikingly, there has been limited analysis of diagnostic yield of FDG-PET/CT beyond that of conventional CT. Further, previous meta-analyses have not studied individual patient data.

Table 1

We performed an up-to-date meta-analysis of the diagnostic yield of FDG-PET/CT in all patients with FUO. Secondary outcomes included the proportion with an abnormal FDG-PET/CT, final diagnosis, false positive results and mortality. The results of the meta-analysis were used to inform two rounds of a Delphi survey and a half-day meeting, to develop a consensus on the current knowledge on the role of FDG-PET/CT in FUO and inform future research.

Materials and Methods

Systematic Review and Meta-analysis

The protocol was prospectively registered with PROSPERO, an online international database of systematic reviews (Study-ID CRD42016032696). It adhered to PRISMA guidelines. QUADAS-2, STROBE, Cochrane guidelines and MOOSE guidelines were also utilised[14-17].

Inclusion and Exclusion criteria: All patients were included irrespective of age, comorbidities or immunocompromise. Inclusion criteria for FDG-PET/CT protocols were not defined, provided they involved a standard [18]-FDG radiotracer. Exclusion criteria were case reports, significant missing data such that the primary outcome could not be calculated and non-English studies.

Search strategy and study detection: See Table 2.

Table 2

Methodological quality assessment: Two authors (TB&AR) independently performed the quality assessment and used this to identify studies to be included in the meta-synthesis. Disagreements were resolved by a third author (SS). Existing research is restricted to case series and, in the absence of comparison with a reference standard, these cannot be interpreted as diagnostic accuracy studies. For this reason a specific quality assessment tool was utilised, with nine criteria scored as 'High', 'Unclear' or 'Low' risk of bias, see Supplement[18]. Each study is given a quality rating 'Poor', 'Fair' and 'Good', and quality assessment are summarised in Figure 3. The studies included in the inter-rater agreement on the quality assessment is evaluated by a calculated kappa statistic, with

95% confidence intervals (CIs) ranging from zero (completely chance-explained agreement) and one (perfect agreement)[19].

Data extraction: A data extraction form was developed using Microsoft Excel, see Supplement, and two authors (TB&R) independently piloted the form and subsequently performed the data extraction. Disagreements were resolved by a third author (SS). Authors of included studies were contacted for missing data.

Analysis: A qualitative synthesis and summary was performed. Results for studies included in the quantitative analysis were calculated as proportions, with meta-analysis performed using a random-effects model in Stata.13 to produce a summary outcome proportion with 95% CIs, and I^2 statistic for heterogeneity. Sensitivity analyses was performed to exclude poor quality studies. Sub-group analyses were performed for immunocompetent adults.

Delphi Consensus

The Delphi technique is an accepted method for generating consensus in a wide variety of disciplines[20-22]. It involves multiple iteration questionnaire surveys with anonymous and unbiased methods. This study included 2-rounds of sequential pre-tested questionnaires, and a half-day face-face meeting. The working-group included 30 UK-based clinicians with expertise in Epidemiology, Research Methods, and Clinical Practice in the specialities of Nuclear Medicine, Radiology, Infectious Diseases, Rheumatology, Haematology and General Medicine. The questionnaires were developed, refined and administered, each consisting of single and multiple answer questions, free-text comments, and 5-point Likert agreement scales. An initial survey was performed in 2015 before the face-to-face meeting and consisted of 12 questions. After the meeting, a refined survey with 22 questions was performed. The surveys and discussion surrounded the current evidence and available guidelines, availability of FDG-PET/CT, working case-definitions of FUO, position of FDG-PET/CT in

diagnostic algorithms of FUO, and potential factors involved in improving the outcomes in the application of FDG-PET/CT. There was also a focus on the future direction of research. Consensus in surveys (Supplement) was accepted if agreement (participants responding 'Strongly agree' or 'Agree') was over 60%.

Results

Systematic review and Meta-analysis

Study Selection: 22 studies were identified for the qualitative synthesis, and the quality assessment selected 18 studies with a total of 905 patients for meta-analysis, see Figure 1. Interrater agreement between reviewers was 91% with Kappa 0.85 and P<0.001. Reasons for exclusions are displayed in Supplementary Data[23-26].

Figure 1

Quality Assessment and Study Design: The qualitative assessment demonstrated a high risk of bias across all the included studies, see Figure 2. All the studies were observational case series with no comparison group. They were largely (89%) retrospective, involving recruitment from the Nuclear Medicine Department databases of patients referred for the indication of a FUO. The studies were largely confined to tertiary care centres, and were geographically widely distributed across 15 different countries in Europe and Asia. The median sample size was 48 (Interquartile range, IQR 24-74), with a median sample size per year 22 (IQR 8-29). The year of commencement of the studies ranged from 2003-2010 (median 2007, IQR 2005-2007), with the year of publication ranging from 2008-2015 (median 2012, IQR 2010-2013). The median study duration was 35 (IQR 23-49) months.

There is insufficient data to report the proportion of children. Three studies included children and none were exclusively performed in children. 50% of the over-all population was female. 10 (56%) studies excluded immunocompromised patients.

Figure 2

Case definitions: The included studies largely reported standardised case definitions of FUO as a fever for 3 weeks with at least one documented fever over 38'c (17, 94%). There was minimal documentation on the duration of symptoms prior to admission or the length of inpatient stay. Patients were referred to the nuclear medicine department for FDG-PET/CT at the discretion of the responsible clinician. One study mandated discussion at a multidisciplinary meeting prior to referral.

Intervention: 17 (94%) studies reported details of their FDG-PET/CT protocols. The protocols demonstrate the studies utilised the same radiotracer injected at a standard interval of 60-90 mins prior to scan. 7 (39%) used IV and/or oral contrast. It was notable that at least 4 (28%) studies utilised high-dose CT. One study incorporated a 24 hour carbohydrate restricted diet prior to the scan to reduce non-specific cardiac uptake. No studies reported independent assessors interpreting the scans, however 7 (39%) reported the involvement of discussion between two assessors, usually a nuclear medicine physician and a radiologist.

Primary outcome: A meta-analysis of 18 studies suggest an overall diagnostic contribution of 56% (95% CI 50-61%), I^2 61% of FDG-PET/CT in all patients with FUO, illustrated in the forest plot in Figure 3. Sub-group analysis for diagnostic contribution in 1) adults, 2) immunocompetent patients ('classical FUO'), 3) immunocompetent adults and 4) immunocompetent adults without contrast reduced the heterogeneity in the model, however the point estimate of diagnostic yield remained largely unchanged, Forest Plots included in Supplementary Data.

Previous cross-sectional imaging and added contribution of FDG-PET/CT: There were sparse data on the documentation or results of previous imaging. Previous investigations were reported in 12 (67%) studies, with a median 51% (IQR 27-81%) receiving a CT prior to referral for FDG-PET/CT. Out of

these, 5 studies reported the results of previous imaging. A sub-group analysis of these data suggest the diagnostic yield of FDG-PET/CT over CT is 32% (95% CI 22-44%), I^2 66%.

Figure 3

Secondary outcomes

Meta-analysis of the proportion with an abnormal FDG-PET/CT produced an overall result of 69% (95% CI 63-75%), I^2 72. The higher proportion of abnormal scans was accounted for by a proportion of 'false positives', abnormal scans with no contribution to the final diagnosis, with an overall result of 9% (95% CI 5-14%), I^2 72. The overall estimate was low which is reassuring but there was striking variation across individual studies, between 0 to 33% reported false positive scans.

73% (95% CI 68-78%) had a final diagnosis, mainly corresponding with three categories: infectious diseases representing 30% (95% CI 26-35%), inflammatory causes 20% (95% CI 17-24%) and malignancy 13% (95% CI 9-17%), data included in Supplementary Text. Individual patient data extraction from 16/18 studies, totalling 749 patients facilitated stratification of diagnoses that did and did not benefit from FDG-PET/CT, illustrated in Figures 4-6.

The presence of raised inflammatory markers were reported in 7 (39%) studies, and there were insufficient data to suggest any association with contribution of FDG-PET/CT to diagnosis.

Methods for the establishment of the final diagnosis were not uniformly reported, however existing data suggests a variety of methods including bone marrow, lymph node, tissue biopsy, serology, microbiology cultures, immunology and autopsy.

There were limited data on the period of follow-up and final outcomes of patients. 12 (67%) studies reported the length of follow-up, with median 6 (IQR 6-12) months.

Figures 4-6

Delphi Consensus

31/40 (78%) participants responded to the initial Delphi survey. 22/40 (55%) attended the face-to-face meeting. 30/40 (75%) responded to the second Delphi. The initial Delphi survey consisted of three parts aiming to assess 1) availability of FDG-PET/CT for FUO, 2) clinical practice in requesting of FDG-PET/CT for FUO, and 3) decision-making in a hypothetical case of FUO, see Supplementary Data for the full questionnaire. While 100% reported access to FDG-PET/CT, there was wide-variability in reported time from referral to FDG-PET/CT ranging from 2 days to 2 weeks (UK Key Performance Indicator, KPI 5 days), and time to reporting of scans ranging from 1 day to 1 week (UK KPI 2days). There was widespread agreement (87% responders) that FDG-PET/CT does have a role in the investigation of unknown origin (suggested to be 56%), however there was little consensus on sub-groups or factors that might improve the diagnostic yield. There was also agreement in the value of re-assessing patients for developing symptoms and signs, involving other specialities during the investigation process, and involvement of nuclear medicine physicians in case discussions. The initial survey demonstrated consensus of opinions that false positives needed to be taken into account in the decision to refer, that FDG-PET/CT has a high negative predictive value and that false negatives may arise due to empirical steroids.

The face-to-face meeting involved a presentation of the results of the systematic review, meta-analysis and initial Delphi survey, with sufficient time for questions and discussion. There were focussed debate surrounding the case-definition of FUO, investigations required and priority outcomes. The meeting identified the variability in access and knowledge of FDG-PET/CT, the heterogeneity and updated working definitions of FUO and dearth of evidence but encouraging

results in clinical practice. It highlighted the need for clinicians to be aware of the deficits of FDG-PET/CT: not always imaging the brain, low sensitivity for cardiac and renal tract pathology and reduced gastrointestinal uptake with certain medication. In contrast to previous opinions, there is no evidence for poor glycaemic control as a contraindication to FDG-PET/CT. Further, the fact that low-contrast imaging is incorporated into standard protocols does reduce the resolution as compared to conventional contrast-CT. It was agreed that certain circumstances affect decision-making, e.g. renal impairment, suitability for invasive tests and recent surgery. The meeting concluded with dialogue on prospects and feasibility of future research. Current practice incorporates FDG-PET/CT late in diagnostic algorithms, however there was acknowledgement that it may have a role as a 'front-loaded' investigation in a subset of patients. This has potential to speed diagnosis, reduced radiation exposure and shorten hospital stay, maybe reduce mortality.

The second Delphi aimed to develop agreement on a case definition of FDG-PET/CT, basic investigations required and resolve disagreement to questions. The participants agreed that a febrile illness for 2 weeks and without immediate diagnostic clues worked for their practice was a clinically acceptable definition. They agreed the definition should incorporate 'Inflammation of Unknown Origin', IUO, unexplained symptoms for 2 weeks with raised inflammatory markers. Specific investigations prior to PET imaging were deemed important, including a cross-sectional CT, TTE and specific serology (see supplementary data). However there was also agreement that a front-loaded FDG-PET/CT prior to conventional imaging may have a role. There was indecision about whether antibiotics should be delayed prior to FDG-PET/CT. Priorities in the outcome of a formal analysis of the benefit of front-loaded PET/CT, in the order of importance (most to least important) were 1) Time to diagnosis, 2) Time to treatment, 3) Mortality, 4) Side-effects of investigations/ treatment and 5) Time to discharge.

Conclusion

PET is a functional imaging tool that provides added information about site and intensity of active metabolism, and so unsurprisingly has found its way into the diagnostic pathway of the febrile patient. However it is expensive, lacks specificity and needs adequate evidence for its diagnostic role. This meta-analysis suggests that a diagnostic yield was achieved in 56% (95% CI 50-61%) performed. The results are consistent with previous results of 54% 'overall helpfulness' (synonymous with diagnostic yield) in a meta-analysis of 10 studies[27]. Two meta-analyses reviewing sensitivity reported 85% (95% CI 81-88%; 15 studies) and 98% (95% CI 94-99%; 9 studies).

The results are based on results of case series, involving convenience sampling of FUO patients referred to Nuclear Medicine departments at the discretion of the responsible physician. Specifically, recruitment is not at the point of diagnosis of fever of unknown origin, and there is no control group. Patient recruitment may favour patients with renal impairment, poor fitness for invasive biopsies, and exclude patients taking metformin, recent surgery or unable to lie still. The room for bias is high and these important patient characteristics are poorly documented in the included studies.

It is also striking that reported diagnostic yield does not address contribution beyond conventional imaging as all the patients did not undergo conventional imaging, and reporting of those that did was inconsistent. 5 studies included in this meta-analysis reported results of previous imaging. A sub-group analysis of these data suggest the diagnostic yield of FDG-PET/CT beyond CT is 32% (95%CI 22-44%) with significant heterogeneity (I^2 66%).

Case definitions of FUO adhered to outdated definitions that were established based on minimal evidence. It is accepted that subsets of patients do not mount any fever, and for this reason it has been suggested that IUO be included in future research. The definition also encompasses an extensive list of diagnoses and possibilities, is geographically diverse and limited by resources.

FDG-PET/CT is perceived to be an objective intervention. However there is minimal data on inter-reporter agreement, and none of the studies involved independent reporting by more than one radiologist. Importantly the protocols frequently included nephrotoxic contrast, and high dose attenuation CTs. Not only may this bias the outcome, but it demonstrates potential risks associated with the scans. There is evidence that a special diet to reduce cardiac non-specific cardiac uptake may improve outcomes, however the only study that included this protocol did not report cardiac diagnoses.

There is no diagnostic reference standard for FUO, and many patients remain undiagnosed. Furthermore there is a level of ambiguity in final diagnoses made by clinicians, and the impression of whether the FDG-PET/CT contributed to the diagnosis. In most studies this was based on the result of the FDG-PET/CT being compatible with the final diagnosis, however it did not demonstrate a diagnostic yield over conventional imaging. Outcome measures need to be relevant to hard patient outcomes and to current health systems processes. While sensitivity is not an appropriate outcome measure, diagnostic yield may also overestimate the contribution and does not indicate the clinical impact of the scan. Other possible outcomes include evaluating time to treatment, discharge or mortality.

It is evident that studies included patients that had not had conventional cross-sectional imaging. Furthermore, a referral for FDG-PET/CT was frequently made in spite of pathology identified on cross-sectional imaging that could undergo alternative, more specific and objective investigation such as a biopsy. With this is mind, the question of diagnostic yield of FDG-PET/CT beyond abnormalities detected by cross-sectional imaging is clinically important.

The individual patient meta-analysis is limited by the low quality of included studies. It does provide suggestion of diagnoses that did and did not benefit from FDG-PET/CT, see Figures 4-6. It is rational that viral infections, urinary tract infections, bacteraemias and small vessel vasculitides are not easily detected on FDG-PET/CT. There are limitations in interpretation of FDG avidity in the brain, heart

and urinary tract. The brain and the heart have high glucose uptake and the urinary tract concentrates FDG during excreted.

This study provides a rigorous, updated and balanced insight into current evidence for the role of FDG-PET/CT in FUO. It demonstrates a lack of evidence supporting the value and positioning of FDG-PET/CT in investigative algorithms. The Delphi survey enabled the working group to interpret results in line with current practice, and explore directions for research. It highlighted the need for a paradigm shift in research, involving prospective studies recruiting at the point of diagnosis of FUO, with updated case definitions and hard outcome measures. While these studies will be a significant undertaking with multi-centre collaboration, their completion is vital for balancing both radiation exposure and costs against the possible benefits of utilising FDG-PET/CT.

Figure and Table Legends

Figure 1: Flow diagram of study selection.

Figure 2: Summary of the Quality Assessment of Included Studies Using the NIH Tool

Figure 3: Diagnostic Yield of FDG-PET/CT in Fever of Unknown Origin (n=18), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

Figure 4: Infections (n=241; 32% of final diagnosis): Diagnostic yield from PET/CT

Figure 5: Inflammatory/ Autoimmune (n=171; 20% of final diagnosis): Diagnostic yield from PET/CT

Figure 6: Malignancy (n=112; 13% of final diagnoses): Diagnostic yield from PET/CT

Table 1: 2x2 table categorising possible study outcomes.

Table 2: Search Strategy and Study Selection

References

1. Horder T. Some Cases of Pyrexia without Physical Signs. Postgrad Med J 1925;1:17-21.
2. Brown M. Pyrexia of unknown origin 90 years on: a paradigm of modern clinical medicine. Postgrad Med J 2015;91:665-669.
3. Petersdorf RG, Beeson PB. Fever of unexplained origin: report on 100 cases. Medicine (Baltimore) 1961;40:1-30.
4. Durack DT, Street AC. Fever of unknown origin--reexamined and redefined. Curr Clin Top Infect Dis 1991;11:35-51.
5. Mourad O, Palda V, Detsky AS. A comprehensive evidence-based approach to fever of unknown origin. Archives of internal medicine 2003;163:545-551.
6. Townsend DW. Combined PET/CT: the historical perspective. Seminars in ultrasound, CT, and MR 2008;29:232-235.
7. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. European Journal of Nuclear Medicine and Molecular Imaging 2015;42:328-354.
8. RCP/RCR/BSNM. Evidence-based indications for the use of PET-CT in the UK 2013. 2013.
9. Data.gov.uk. https://data.gov.uk/data-request/nhs-hospital-stay. 2015.
10. Dong M-j, Zhao K, Liu Z-f, et al. A meta-analysis of the value of fluorodeoxyglucose-PET/PET-CT in the evaluation of fever of unknown origin. European Journal of Radiology 2011;80:834-844.
11. Hao R, Yuan L, Kan Y, et al. Diagnostic performance of F-18-FDG PET/CT in patients with fever of unknown origin: a meta-analysis. Nuclear Medicine Communications 2013;34:682-688.
12. FDA. Statistical guidance on reporting results from studies evaluating diagnostic tests., 2007.
13. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991;11:88-94.
14. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Annals of internal medicine 2011;155:529-536.

15. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61:344-349.

16. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. JAMA 2000;283:2008-2012.

17. Higgins JPT GSeTCC, 2011. . Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. ; 2011.

18. Health NIf. Quality Assessment Tool for Case Series Studies. 2014.

19. Vierra A, Garrett J. Understandign interobserver agreement: The Kappa statistic. Family medicine 2005;37:360-363.

20. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. Journal of Advanced Nursing 2000;32:1008-1015.

21. Thangaratinam S, Redman C. The Delphi Technique. The Obstetrician and Gynacologist 2011;7:120-125.

22. Shankar-Hari M, Phillips GS, Levy ML, et al. Developing a New Definition and Assessing New Clinical Criteria for Septic Shock: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315:775-787.

23. Bharucha T, Brown M. Use of CT-PET imaging in investigation of pyrexia of unknown origin. British Journal of Hospital Medicine 2013;74:654.

24. Castaignea C, Tondeura M, De Witb S, et al. Clinical value of FDG-PET/CT for the diagnosis of human immunodeficiency virus-associated fever of unknown origin: a retrospective study. Nuclear Medicine Communications 2009;30:41–47.

25. Martin C, Castaigne C, Tondeur M, et al. Role and interpretation of fluorodeoxyglucose-positron emission tomography/computed tomography in HIV-infected patients with fever of unknown origin: a prospective study. HIV Medicine 2013;14:455-462.

359 26. Jasper N, Daebritz J, Frosch M, et al. Diagnostic value of F-18 -FDG PET/CT in children with
360 fever of unknown origin or unexplained signs of inflammation. European Journal of Nuclear
361 Medicine and Molecular Imaging 2010;37:136-145.
362 27. Kouijzer IJ, Bleeker-Rovers CP, Oyen WJ. FDG-PET in fever of unknown origin. Seminars in
363 Nuclear Medicine 2013;43:333-339.

364

Key words

Imaging, Nuclear Medicine, Fever of Unknown Origin, Diagnostics

Abbreviations

CI	Confidence Intervals
FDG-PET/CT	2-[18F]-fluoro-2-deoxy-D-glucose - Positron Emission Tomography/Computed Tomography
FUO	Fever of Unknown Origin
IQR	Interquartile Range
IUO	Inflammation of Unknown Origin
KPI	Key Performance Indicator

Introduction

Fever as an isolated clinical presentation has challenged clinicians for decades[1, 2]. In 1961 Petersdorf and Beeson provided a case definition for 'fever (or pyrexia) of unknown origin': 1) a body temperature above 38.3°C; 2) on several occasions; with 3) a duration of illness of at least three weeks; and 4) no diagnosis within one week of hospital admission[2-4]. Fifty years on, definitions of FUO and the spectrum of aetiologies have evolved, however the diagnostic challenges remain[4]. FUO represents an estimated 2.9% of hospital admissions, with morbidity associated with prolonged hospital stay, repeated cycles of invasive investigations and presumptive treatment, mortality rates between 12-35%, and cost implications[5].

2-[18F]-fluoro-2-deoxy-D-glucose (FDG)-Positron-Emission-Tomography/Computed-Tomography (PET/CT) emerged at the end of the 20th century as an amalgamation between functional and conventional anatomical imaging[6]. Its role in oncological staging has been well-defined, however in other specialities there is less clarity[7]. Specifically, in the investigation of FUO the role of FDG-PET/CT in clinical practice and diagnostic algorithms is inconsistent and unestablished. Existing guidelines suggest that FDG-PET/CT *may* be used where conventional investigations have not revealed a source[8].

FDG-PET/CT is not associated with nephrotoxicity, and standard protocols expose patients to less radiation than a conventional CT. An average FDG-PET/CT scan exposes a patient to 15mSv radiation, approximately 5-6 years background radiation, rather than 20-25mSv in a contrast-enhanced chest-abdomen-pelvis CT. Other advantages include imaging areas (e.g. head and neck, extremities) which are beyond the range of most CT scans used in this context, and detection of vascular and truncal musculoskeletal inflammation for which cross-sectional contrast CT imaging is insensitive [9]. The main caveats are cost and accessibility, FDG-PET/CT costing £800, compared to £250 for a contrast-enhanced chest-abdomen-pelvis CT. However this could easily be remunerated by earlier definitive

treatment associated with additional diagnostic sensitivity. A marginally reduced length of inpatient stay could mitigate the cost, with an average £400 for one night hospital admission[10].

Current literature evaluating the role of FDG-PET/CT in FUO is based on observational data involving small samples, outdated case definitions, and poor generalisability. Outcomes reported by existing meta-analyses focus on sensitivity of FDG-PET/CT in FUO[11, 12]. Sensitivity refers to the proportion of cases with a diagnosis to explain the FUO for which FDG-PET/CT contributed to the diagnosis, or A/(A+B) (Table 1). This is statistically inappropriate as there is no reference standard for the investigation of FUO to enable estimates of diagnostic accuracy[13]. In comparison, 'diagnostic yield' provides a more suitable outcome measure, calculated as the proportion of all FDG-PET/CT scans (both normal and abnormal) that contribute to the diagnosis of FUO, A/(A+B+C+D) (Table 1)[14]. Strikingly, there has been limited analysis of diagnostic yield of FDG-PET/CT beyond that of conventional CT. Further, previous meta-analyses have not studied individual patient data.

Table 1

We performed an up-to-date meta-analysis of the diagnostic yield of FDG-PET/CT in all patients with FUO. Secondary outcomes included the proportion with an abnormal FDG-PET/CT, final diagnosis, false positive results and mortality. The results of the meta-analysis were used to inform two rounds of a Delphi survey and a half-day meeting, to develop a consensus on the current knowledge on the role of FDG-PET/CT in FUO and inform future research.

Materials and Methods

Systematic Review and Meta-analysis

The protocol was prospectively registered with PROSPERO, an online international database of systematic reviews (Study-ID CRD42016032696). It adhered to PRISMA guidelines. QUADAS-2, STROBE, Cochrane guidelines and MOOSE guidelines were also utilised[15-18].

Inclusion and Exclusion criteria: All patients were included irrespective of age, comorbidities or immunocompromise. Inclusion criteria for FDG-PET/CT protocols were not defined, provided they involved a standard [18]-FDG radiotracer. Exclusion criteria were case reports, significant missing data such that the primary outcome could not be calculated and non-English studies.

Search strategy and study detection: See Table 2.

Table 2

Methodological quality assessment: Two authors (TB&AR) independently performed the quality assessment and used this to identify studies to be included in the meta-synthesis. Disagreements were resolved by a third author (SS). Existing research is restricted to case series and, in the absence of comparison with a reference standard, these cannot be interpreted as diagnostic accuracy studies. For this reason a specific quality assessment tool was utilised, with nine criteria scored as 'High', 'Unclear' or 'Low' risk of bias, see Supplement[19]. Each study is given a quality rating 'Poor', 'Fair' and 'Good', and quality assessment are summarised in Figure 3. The studies included in the inter-rater agreement on the quality assessment is evaluated by a calculated kappa statistic, with

95% confidence intervals (CIs) ranging from zero (completely chance-explained agreement) and one (perfect agreement)[20].

Data extraction: A data extraction form was developed using Microsoft Excel, see Supplement, and two authors (TB&AR) independently piloted the form and subsequently performed the data extraction. Disagreements were resolved by a third author (SS). Authors of included studies were contacted for missing data.

Analysis: A qualitative synthesis and summary was performed. Results for studies included in the quantitative analysis were calculated as proportions, with meta-analysis performed using a random-effects model in Stata.13 to produce a summary outcome proportion with 95% CIs, and I^2 statistic for heterogeneity. Sensitivity analyses was performed to exclude poor quality studies. Sub-group analyses were performed for immunocompetent adults.

Delphi Consensus

The Delphi technique is an accepted method for generating consensus in a wide variety of disciplines[21-23]. It involves multiple iteration questionnaire surveys with anonymous and unbiased methods. This study included 2-rounds of sequential pre-tested questionnaires, and a half-day face-face meeting. The working-group included 30 UK-based clinicians with expertise in Epidemiology, Research Methods, and Clinical Practice in the specialities of Nuclear Medicine, Radiology, Infectious Diseases, Rheumatology, Haematology and General Medicine. The questionnaires were developed, refined and administered, each consisting of single and multiple answer questions, free-text comments, and 5-point Likert agreement scales. An initial survey was performed in 2015 before the face-to-face meeting and consisted of 12 questions. After the meeting, a refined survey with 22 questions was performed. The surveys and discussion surrounded the current evidence and available guidelines, availability of FDG-PET/CT, working case-definitions of FUO, position of FDG-PET/CT in

105 diagnostic algorithms of FUO, and potential factors involved in improving the outcomes in the
106 application of FDG-PET/CT. There was also a focus on the future direction of research. Consensus in
107 surveys (Supplement) was accepted if agreement (participants responding ‘Strongly agree’ or
108 ‘Agree’) was over 60%.

109

Results

Systematic review and Meta-analysis

Study Selection: 22 studies were identified for the qualitative synthesis, and the quality assessment selected 18 studies with a total of 905 patients for meta-analysis, see Figure 1. Interrater agreement between reviewers was 91% with Kappa 0.85 (95% CI 0.75-0.96). Reasons for exclusions are displayed in Supplementary Data[24-27].

Figure 1

Quality Assessment and Study Design: The qualitative assessment demonstrated a high risk of bias across all the included studies, see Figure 2. All the studies were observational case series with no comparison group. They were largely (89%) retrospective, involving recruitment from the Nuclear Medicine Department databases of patients referred for the indication of a FUO. The studies were largely confined to tertiary care centres, and were geographically widely distributed across 15 different countries in Europe and Asia. The median sample size was 48 (Interquartile range, IQR 24-74), with a median sample size per year 22 (IQR 8-29). The year of commencement of the studies ranged from 2003-2010 (median 2007, IQR 2005-2007), with the year of publication ranging from 2008-2015 (median 2012, IQR 2010-2013). The median study duration was 35 (IQR 23-49) months.

There is insufficient data to report the proportion of children. Three studies included children and none were exclusively performed in children. 50% of the over-all population was female. 10 (56%) studies excluded immunocompromised patients.

Figure 2

Case definitions: The included studies largely reported standardised case definitions of FUO as a fever for 3 weeks with at least one documented fever over 38'c (17, 94%). There was minimal documentation on the duration of symptoms prior to admission or the length of inpatient stay. Patients were referred to the nuclear medicine department for FDG-PET/CT at the discretion of the responsible clinician. One study mandated discussion at a multidisciplinary meeting prior to referral.

Intervention: 17 (94%) studies reported details of their FDG-PET/CT protocols. The protocols demonstrate the studies utilised the same radiotracer injected at a standard interval of 60-90 mins prior to scan. 7 (39%) used IV and/or oral contrast. It was notable that at least 4 (28%) studies utilised high-dose CT. One study incorporated a 24 hour carbohydrate restricted diet prior to the scan to reduce non-specific cardiac uptake. No studies reported independent assessors interpreting the scans, however 7 (39%) reported the involvement of discussion between two assessors, usually a nuclear medicine physician and a radiologist.

Primary outcome: A meta-analysis of 18 studies suggest an overall diagnostic contribution of 56% (95% CI 50-61%), I^2 61% of FDG-PET/CT in all patients with FUO, illustrated in the forest plot in Figure 3. Sub-group analysis for diagnostic contribution was performed in 1) adults, 2) immunocompetent patients ('classical FUO'), 3) immunocompetent adults and 4) immunocompetent adults undergoing PET/CT without contrast enhancement. These analyses reduced the heterogeneity in the model, however the point estimate of diagnostic yield remained largely unchanged, Forest Plots included in Supplementary Data.

Previous cross-sectional imaging and added contribution of FDG-PET/CT: There were sparse data on the documentation or results of previous imaging. Previous investigations were reported in 12 (67%) studies, with a median 51% (IQR 27-81%) receiving a CT prior to referral for FDG-PET/CT. Out of

these, 5 studies reported the results of previous imaging. A sub-group analysis of these data suggest the diagnostic yield of FDG-PET/CT over CT is 32% (95% CI 22-44%), I^2 66%.

Figure 3

Secondary outcomes

Meta-analysis of the proportion with an abnormal FDG-PET/CT produced an overall result of 69% (95% CI 63-75%), I^2 72. The higher proportion of abnormal scans was accounted for by a proportion of 'false positives', abnormal scans with no contribution to the final diagnosis, with an overall result of 9% (95% CI 5-14%), I^2 72. The overall estimate was low which is reassuring but there was striking variation across individual studies, between 0 to 33% reported false positive scans.

73% (95% CI 68-78%) had a final diagnosis, mainly corresponding with three categories: infectious diseases representing 32% (95% CI 27-37%), inflammatory causes 20% (95% CI 17-24%) and malignancy 12% (95% CI 8-17%), data included in Supplementary Text. Individual patient data extraction from 16/18 studies, totalling 749 patients facilitated stratification of diagnoses that did and did not benefit from FDG-PET/CT, illustrated in Figures 4-6.

The presence of raised inflammatory markers were reported in 7 (39%) studies, and there were insufficient data to suggest any association with contribution of FDG-PET/CT to diagnosis.

Methods for the establishment of the final diagnosis were not uniformly reported, however existing data suggests a variety of methods including bone marrow, lymph node, tissue biopsy, serology, microbiology cultures, immunology and autopsy.

There were limited data on the period of follow-up and final outcomes of patients. 12 (67%) studies reported the length of follow-up, with median 6 (IQR 6-12) months.

Figures 4-6

Delphi Consensus

31/40 (78%) participants responded to the initial Delphi survey. 22/40 (55%) attended the face-to-face meeting. 30/40 (75%) responded to the second Delphi. The initial Delphi survey consisted of three parts aiming to assess 1) availability of FDG-PET/CT for FUO, 2) clinical practice in requesting of FDG-PET/CT for FUO, and 3) decision-making in a hypothetical case of FUO, see Supplementary Data for the full questionnaire. While 100% reported access to FDG-PET/CT, there was wide-variability in reported time from referral to FDG-PET/CT ranging from 2 days to 2 weeks (UK Key Performance Indicator, KPI 5 days), and time to reporting of scans ranging from 1 day to 1 week (UK KPI 2days). There was widespread agreement (87% responders) that FDG-PET/CT does have a role in the investigation of unknown origin (suggested to be 56%), however there was little consensus on sub-groups or factors that might improve the diagnostic yield. There was also agreement in the value of re-assessing patients for developing symptoms and signs, involving other specialities during the investigation process, and involvement of nuclear medicine physicians in case discussions. The initial survey demonstrated consensus of opinions that false positives needed to be taken into account in the decision to refer, that FDG-PET/CT has a high negative predictive value and that false negatives may arise due to empirical steroids.

The face-to-face meeting involved a presentation of the results of the systematic review, meta-analysis and initial Delphi survey, with sufficient time for questions and discussion. There were focussed debate surrounding the case-definition of FUO, investigations required and priority outcomes. The meeting identified the variability in access and knowledge of FDG-PET/CT, the heterogeneity and updated working definitions of FUO and dearth of evidence but encouraging

results in clinical practice. It highlighted the need for clinicians to be aware of the deficits of FDG-PET/CT: not always imaging the brain, low sensitivity for cardiac and renal tract pathology and reduced gastrointestinal uptake with certain medication. In contrast to previous opinions, there is no evidence for poor glycaemic control as a contraindication to FDG-PET/CT. Further, the fact that low-contrast imaging is incorporated into standard protocols does reduce the resolution as compared to conventional contrast-CT. It was agreed that certain circumstances affect decision-making, e.g. renal impairment, suitability for invasive tests and recent surgery. The meeting concluded with dialogue on prospects and feasibility of future research. Current practice incorporates FDG-PET/CT late in diagnostic algorithms, however there was acknowledgement that it may have a role as a 'front-loaded' investigation in a subset of patients. This has potential to speed diagnosis, reduced radiation exposure and shorten hospital stay, maybe reduce mortality.

The second Delphi aimed to develop agreement on a case definition of FDG-PET/CT, basic investigations required and resolve disagreement to questions. The participants agreed that a febrile illness for 2 weeks and without immediate diagnostic clues worked for their practice was a clinically acceptable definition. They agreed the definition should incorporate 'Inflammation of Unknown Origin', IUO, unexplained symptoms for 2 weeks with raised inflammatory markers. Specific investigations prior to PET imaging were deemed important, including a cross-sectional CT, TTE and specific serology (see supplementary data). However there was also agreement that a front-loaded FDG-PET/CT prior to conventional imaging may have a role. There was indecision about whether antibiotics should be delayed prior to FDG-PET/CT. Priorities in the outcome of a formal analysis of the benefit of front-loaded PET/CT, in the order of importance (most to least important) were 1) Time to diagnosis, 2) Time to treatment, 3) Mortality, 4) Side-effects of investigations/ treatment and 5) Time to discharge.

Conclusion

PET is a functional imaging tool that provides added information about site and intensity of active metabolism, and so unsurprisingly has found its way into the diagnostic pathway of the febrile patient. However it is expensive, lacks specificity and needs adequate evidence for its diagnostic role. This meta-analysis suggests that a diagnostic yield was achieved in 56% (95% CI 50-61%) performed. The results are consistent with previous results of 54% 'overall helpfulness' (synonymous with diagnostic yield) in a meta-analysis of 10 studies[28]. Two meta-analyses reviewing sensitivity reported 85% (95% CI 81-88%; 15 studies) and 98% (95% CI 94-99%; 9 studies).

The results are based on results of case series, involving convenience sampling of FUO patients referred to Nuclear Medicine departments at the discretion of the responsible physician. Specifically, recruitment is not at the point of diagnosis of fever of unknown origin, and there is no control group. Patient recruitment may favour patients with renal impairment, poor fitness for invasive biopsies, and exclude patients taking metformin, recent surgery or unable to lie still. The room for bias is high and these important patient characteristics are poorly documented in the included studies.

It is also striking that reported diagnostic yield does not address contribution beyond conventional imaging as all the patients did not undergo conventional imaging, and reporting of those that did was inconsistent. 5 studies included in this meta-analysis reported results of previous imaging. A sub-group analysis of these data suggest the diagnostic yield of FDG-PET/CT beyond CT is 32% (95%CI 22-44%) with significant heterogeneity (I^2 66%).

Case definitions of FUO adhered to outdated definitions that were established based on minimal evidence. It is accepted that subsets of patients do not mount any fever, and for this reason it has been suggested that IUO be included in future research. The definition also encompasses an extensive list of diagnoses and possibilities, is geographically diverse and limited by resources.

FDG-PET/CT is perceived to be an objective intervention. However there is minimal data on inter-reporter agreement, and none of the studies involved independent reporting by more than one radiologist. Importantly the protocols frequently included nephrotoxic contrast, and high dose attenuation CTs. Not only may this bias the outcome, but it demonstrates potential risks associated with the scans. There is evidence that a special diet to reduce cardiac non-specific cardiac uptake may improve outcomes, however the only study that included this protocol did not report cardiac diagnoses.

There is no diagnostic reference standard for FUO, and many patients remain undiagnosed. Furthermore there is a level of ambiguity in final diagnoses made by clinicians, and the impression of whether the FDG-PET/CT contributed to the diagnosis. In most studies this was based on the result of the FDG-PET/CT being compatible with the final diagnosis, however it did not demonstrate a diagnostic yield over conventional imaging. Outcome measures need to be relevant to hard patient outcomes and to current health systems processes. While sensitivity is not an appropriate outcome measure, diagnostic yield may also overestimate the contribution and does not indicate the clinical impact of the scan. Other possible outcomes include evaluating time to treatment, discharge or mortality.

It is evident that studies included patients that had not had conventional cross-sectional imaging. Furthermore, a referral for FDG-PET/CT was frequently made in spite of pathology identified on cross-sectional imaging that could undergo alternative, more specific and objective investigation such as a biopsy. With this is mind, the question of diagnostic yield of FDG-PET/CT beyond abnormalities detected by cross-sectional imaging is clinically important.

The individual patient meta-analysis is limited by the low quality of included studies. It does provide suggestion of diagnoses that did and did not benefit from FDG-PET/CT, see Figures 4-6. It is rational that viral infections, urinary tract infections, bacteraemias and small vessel vasculitides are not easily detected on FDG-PET/CT. There are limitations in interpretation of FDG avidity in the brain, heart

and urinary tract. The brain and the heart have high glucose uptake and the urinary tract concentrates FDG during excreted.

This study provides a rigorous, updated and balanced insight into current evidence for the role of FDG-PET/CT in FUO. It demonstrates a lack of evidence supporting the value and positioning of FDG-PET/CT in investigative algorithms. The Delphi survey enabled the working group to interpret results in line with current practice, and explore directions for research. It highlighted the need for a paradigm shift in research, involving prospective studies recruiting at the point of diagnosis of FUO, with updated case definitions and hard outcome measures. While these studies will be a significant undertaking with multi-centre collaboration, their completion is vital for balancing both radiation exposure and costs against the possible benefits of utilising FDG-PET/CT.

Lastly, there is no doubt that the application of FDG-PET/CT is a rapidly evolving field. This review did not cover emerging evidence from new modalities and tracers, such as FDG-leucocyte or Gallium-labelled imaging [29].

Figure and Table Legends

Figure 1: Flow diagram of study selection.

Figure 2: Summary of the Quality Assessment of Included Studies Using the NIH Tool

Figure 3: Diagnostic Yield of FDG-PET/CT in Fever of Unknown Origin (n=18), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

Figure 4: Infections (n=241; 32% of final diagnosis): Diagnostic yield from PET/CT

Figure 5: Inflammatory/ Autoimmune (n=171; 20% of final diagnosis): Diagnostic yield from PET/CT

Figure 6: Malignancy (n=112; 12% of final diagnoses): Diagnostic yield from PET/CT

Table 1: 2x2 table categorising possible study outcomes.

Table 2: Search Strategy and Study Selection

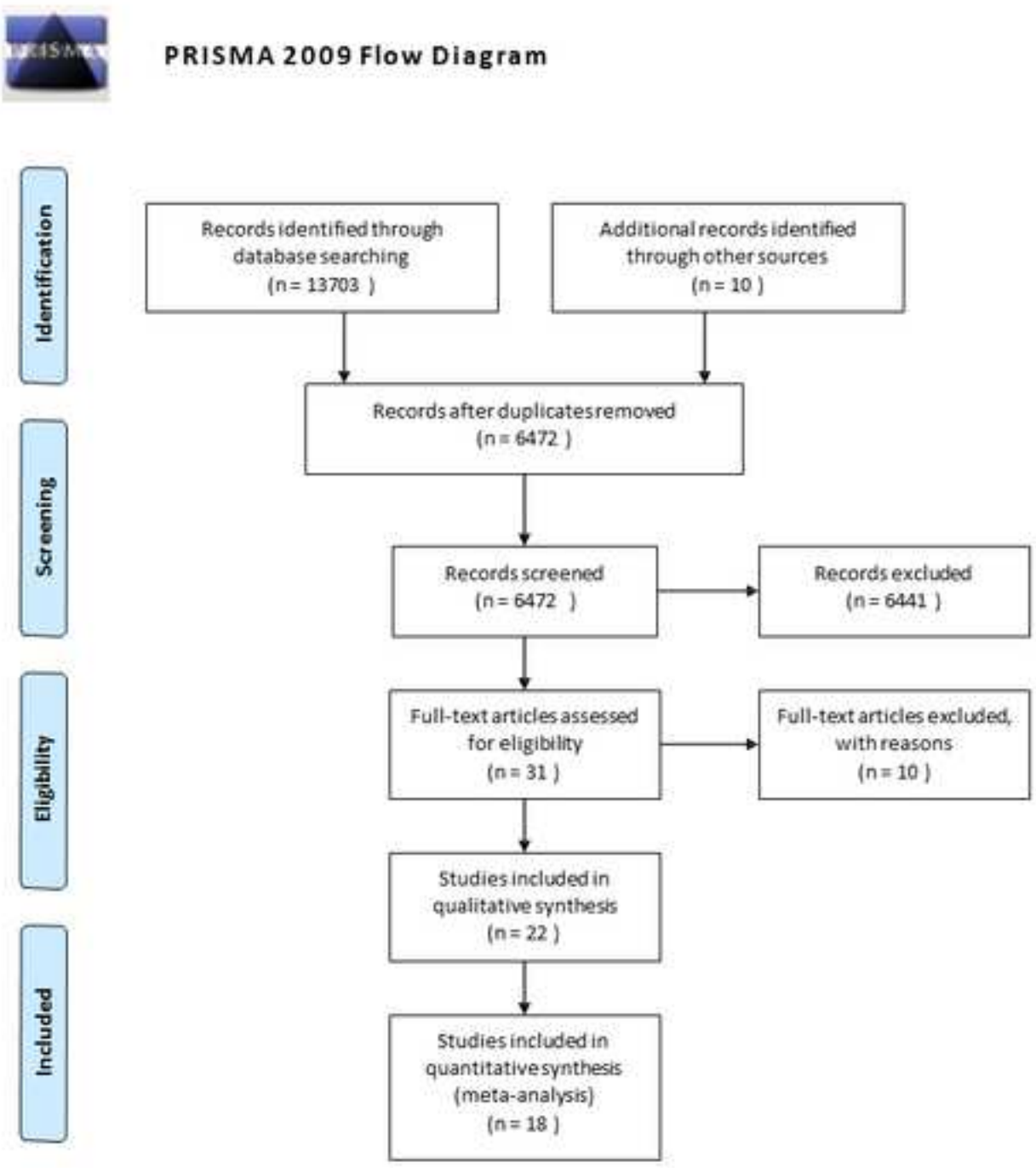
References

1. Horder T. Some Cases of Pyrexia without Physical Signs. Postgrad Med J 1925;1:17-21.
2. Brown M. Pyrexia of unknown origin 90 years on: a paradigm of modern clinical medicine. Postgrad Med J 2015;91:665-669.
3. Petersdorf RG, Beeson PB. Fever of unexplained origin: report on 100 cases. Medicine (Baltimore) 1961;40:1-30.
4. Durack DT, Street AC. Fever of unknown origin--reexamined and redefined. Curr Clin Top Infect Dis 1991;11:35-51.
5. Mourad O, Palda V, Detsky AS. A comprehensive evidence-based approach to fever of unknown origin. Archives of internal medicine 2003;163:545-551.
6. Townsend DW. Combined PET/CT: the historical perspective. Seminars in ultrasound, CT, and MR 2008;29:232-235.
7. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. European Journal of Nuclear Medicine and Molecular Imaging 2015;42:328-354.
8. RCP/RCR/BSNM. Evidence-based indications for the use of PET-CT in the UK 2013. 2013.
9. Quinn B, Dauer Z, Pandit-Taskar N, et al. Radiation dosimetry of 18F-FDG PET/CT: incorporating exam-specific parameters in dose estimates. BMC Medical Imaging 2016;16.
10. Data.gov.uk. https://data.gov.uk/data-request/nhs-hospital-stay. 2015.
11. Dong M-j, Zhao K, Liu Z-f, et al. A meta-analysis of the value of fluorodeoxyglucose-PET/PET-CT in the evaluation of fever of unknown origin. European Journal of Radiology 2011;80:834-844.
12. Hao R, Yuan L, Kan Y, et al. Diagnostic performance of F-18-FDG PET/CT in patients with fever of unknown origin: a meta-analysis. Nuclear Medicine Communications 2013;34:682-688.
13. FDA. Statistical guidance on reporting results from studies evaluating diagnostic tests., 2007.
14. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991;11:88-94.
15. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Annals of internal medicine 2011;155:529-536.
16. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61:344-349.
17. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. JAMA 2000;283:2008-2012.
18. Higgins JPT GSeTCC, 2011. . Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. ; 2011.
19. Health NIf. Quality Assessment Tool for Case Series Studies. 2014.
20. Vierra A, Garrett J. Understandign interobserver agreement: The Kappa statistic. Family medicine 2005;37:360-363.
21. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. Journal of Advanced Nursing 2000;32:1008-1015.
22. Thangaratinam S, Redman C. The Delphi Technique. The Obstetrician and Gynacologist 2011;7:120-125.
23. Shankar-Hari M, Phillips GS, Levy ML, et al. Developing a New Definition and Assessing New Clinical Criteria for Septic Shock: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315:775-787.
24. Bharucha T, Brown M. Use of CT-PET imaging in investigation of pyrexia of unknown origin. British Journal of Hospital Medicine 2013;74:654.
25. Castaignea C, Tondeura M, De Witb S, et al. Clinical value of FDG-PET/CT for the diagnosis of human immunodeficiency virus-associated fever of unknown origin: a retrospective study. Nuclear Medicine Communications 2009;30:41–47.

26. Martin C, Castaigne C, Tondeur M, et al. Role and interpretation of fluorodeoxyglucose-positron emission tomography/computed tomography in HIV-infected patients with fever of unknown origin: a prospective study. HIV Medicine 2013;14:455-462.
27. Jasper N, Daebritz J, Frosch M, et al. Diagnostic value of F-18 -FDG PET/CT in children with fever of unknown origin or unexplained signs of inflammation. European Journal of Nuclear Medicine and Molecular Imaging 2010;37:136-145.
28. Kouijzer IJ, Bleeker-Rovers CP, Oyen WJ. FDG-PET in fever of unknown origin. Seminars in Nuclear Medicine 2013;43:333-339.
29. Vaidyanathan S, Patel CN, Scarsbrook AF, et al. FDG PET/CT in infection and inflammation--current and emerging clinical applications. Clinical Radiology 2015;70:787-800.

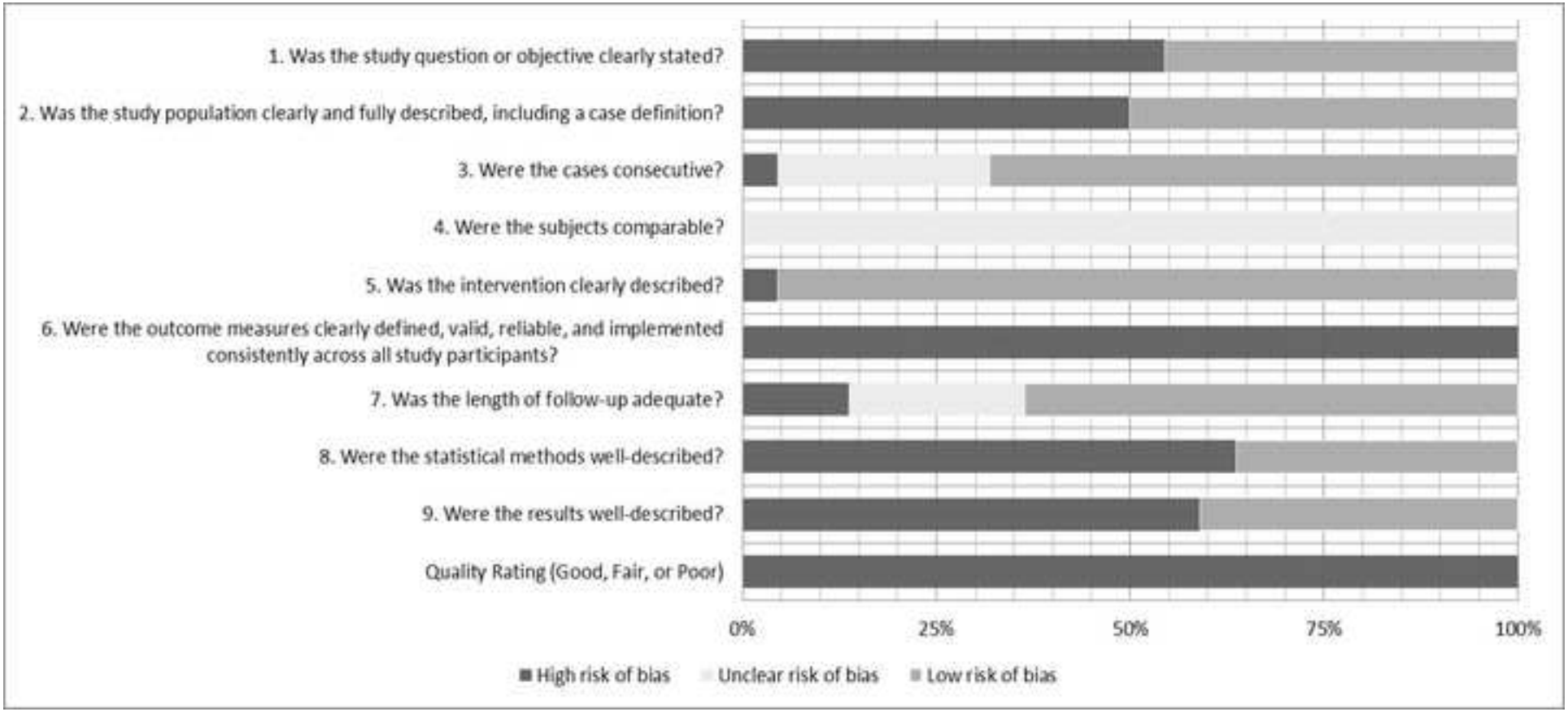
Revised Figure 1

Click here to download Revised Figure Figure 1 Prisma diagram.tif



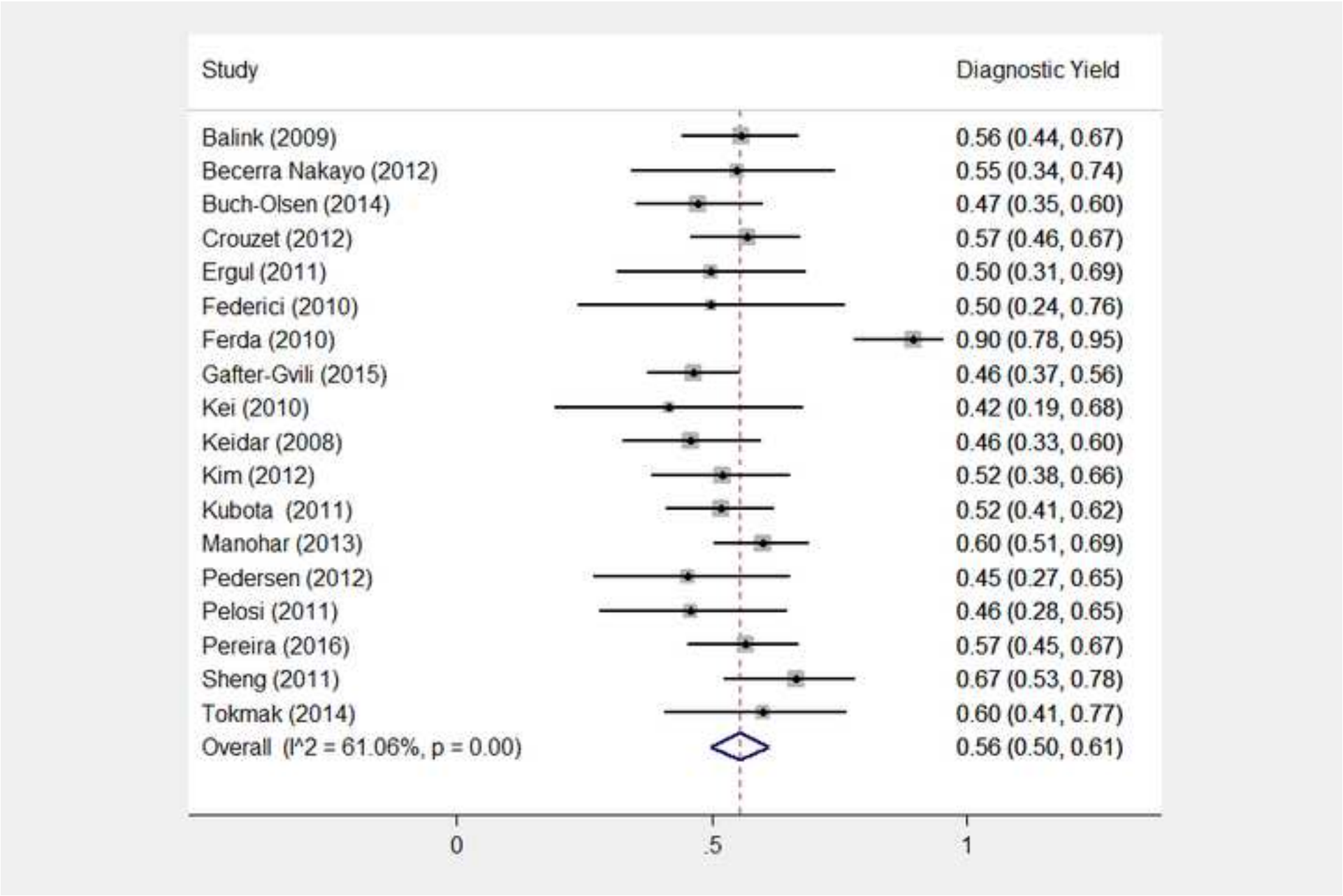
From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.* PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

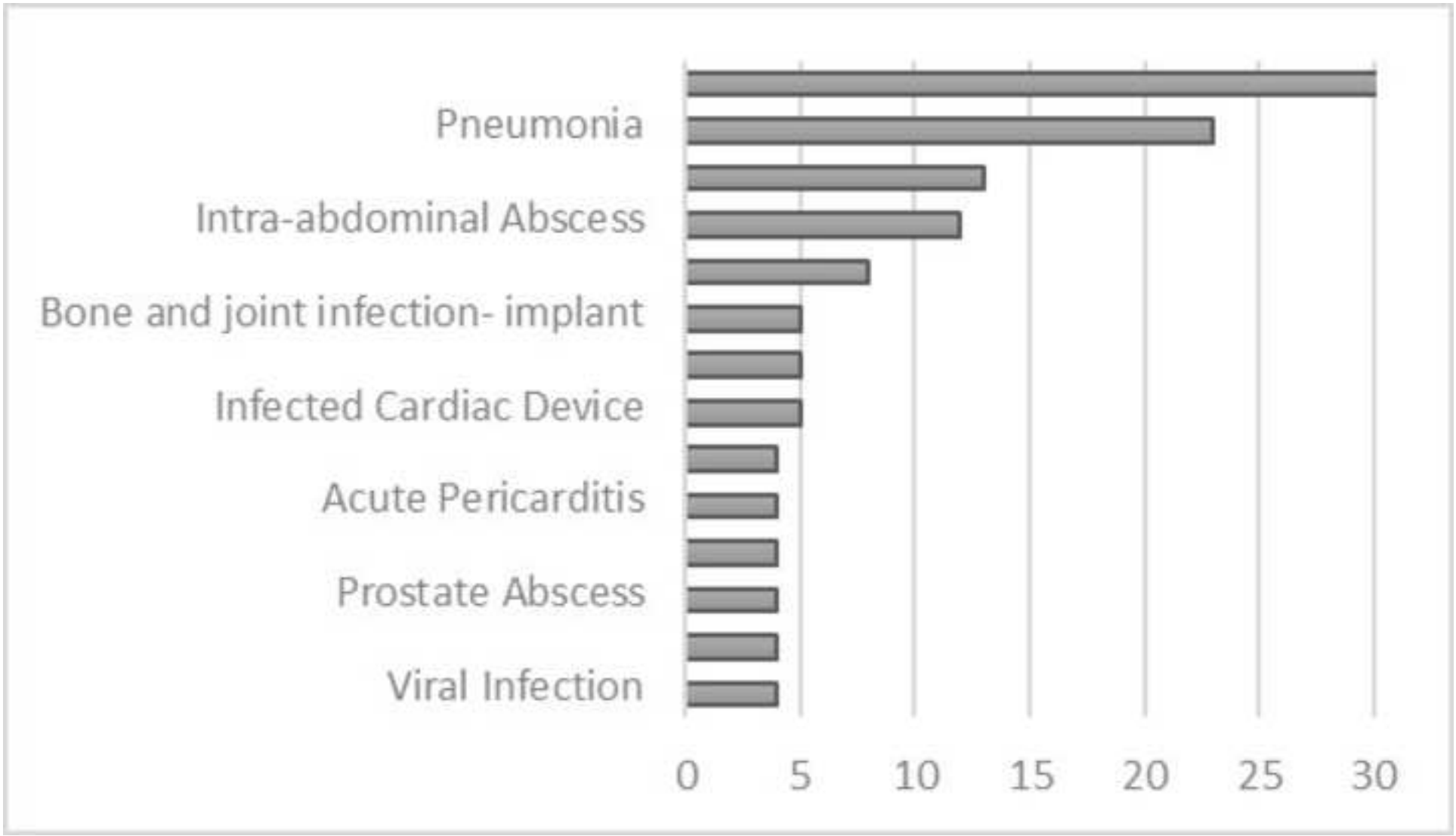
For more information, visit www.prisma-statement.org.

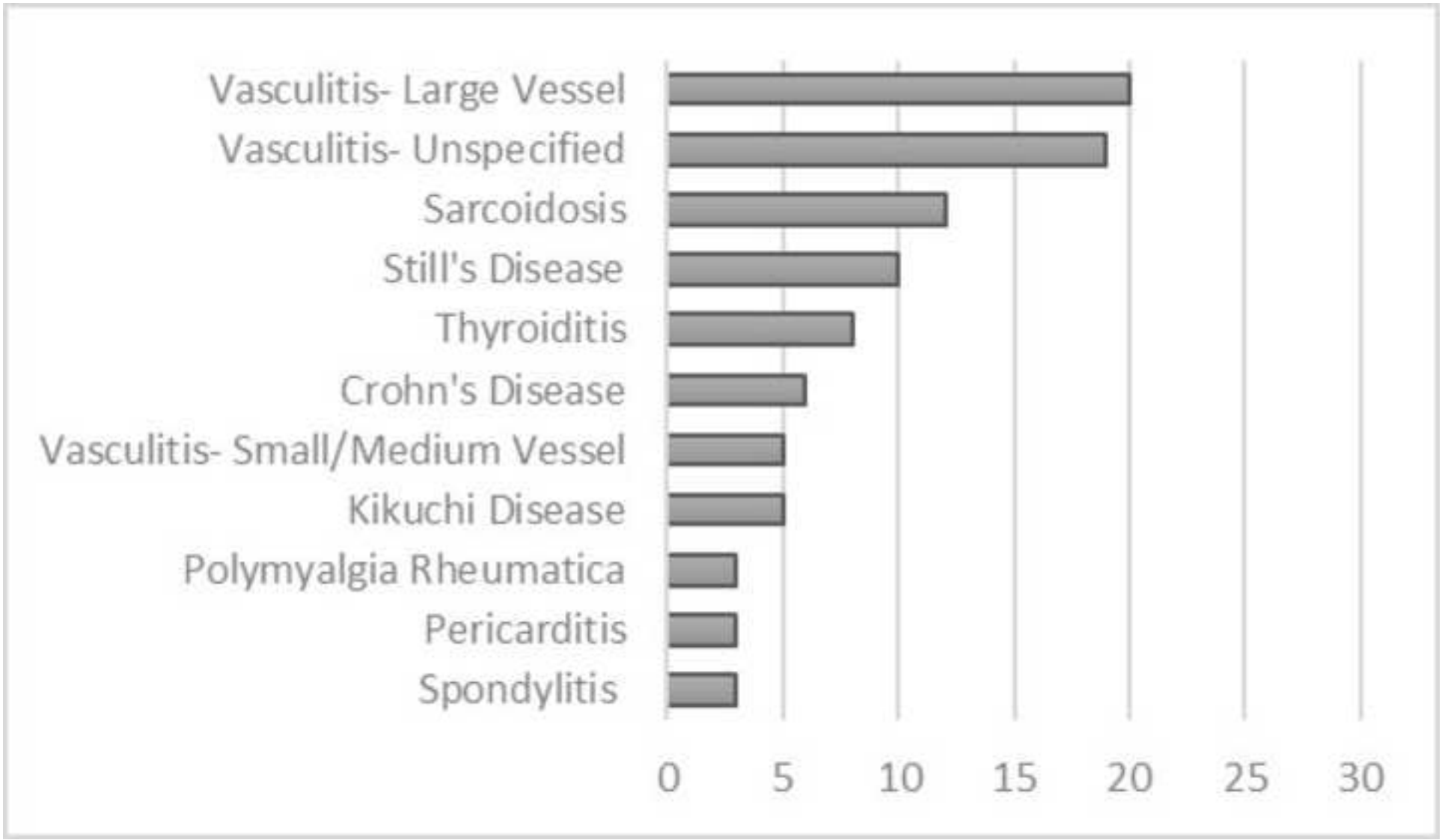


Revised Figure 3

Click here to download Revised Figure Figure 3 Diagnostic Yield.tif







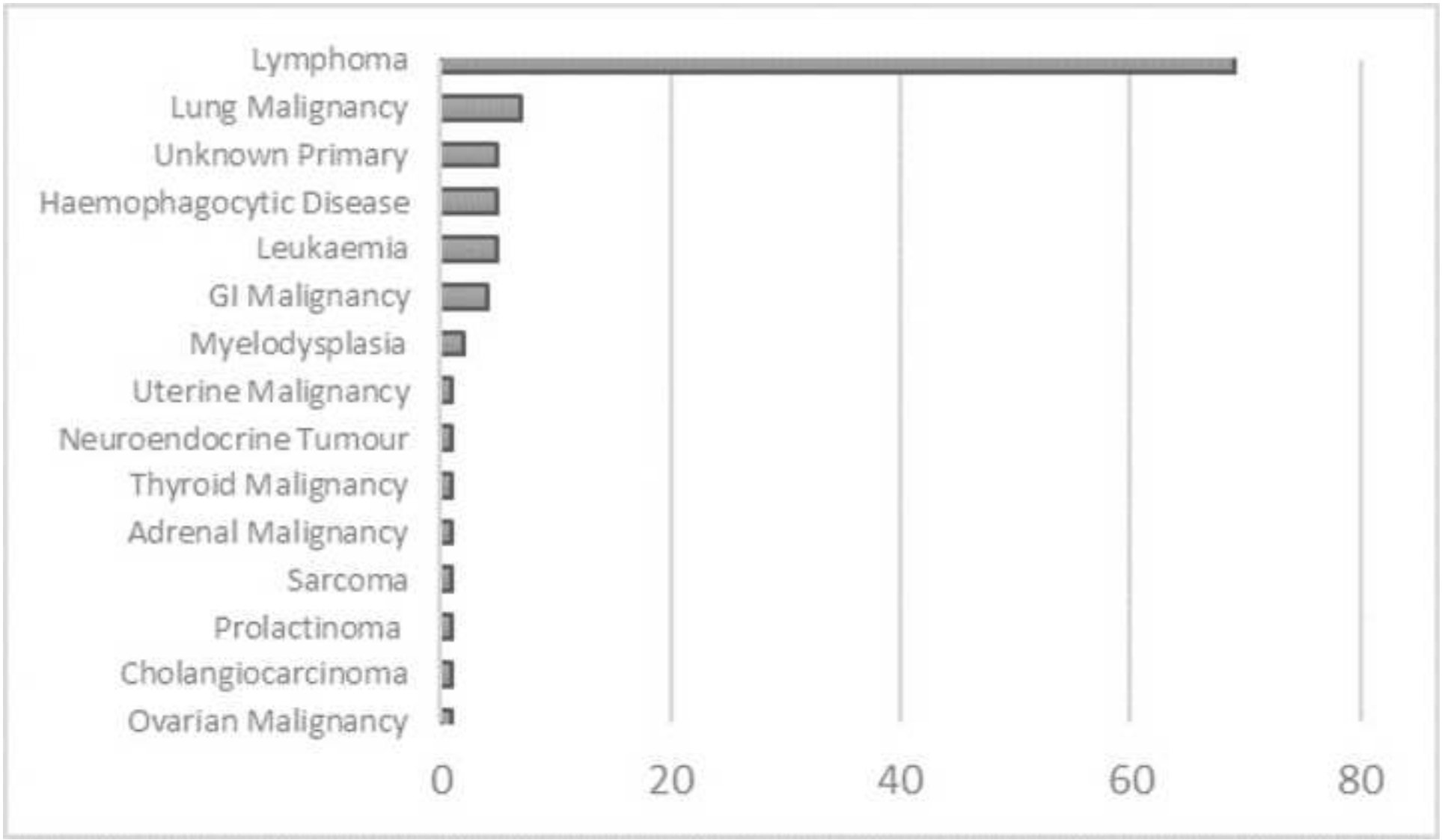


Table 1: 2x2 table categorising possible study outcomes.

[A] True Positives: Patients with an abnormal FDG-PET/CT that contributed to diagnosing the cause of the FUO.	[B] False Negatives: Patients with a normal FDG-PET/CT that received a diagnosis by other means.
[C] False Positive: Patients with an abnormal FDG-PET/CT that did not contribute to diagnosing the FDG-PET/CT.	[D] True Negative: Patients with a normal FDG-PET/CT that remained undiagnosed after investigation or follow-up.

Table 2: Search Strategy and Study Selection

Search Strategy: Electronic searches were performed 1/12/15 in Medline, Embase, Web of Science and Cochrane Central Register of Controlled Trials. All subheadings were included. Hand-searching references was performed for included studies and identification of unpublished work was attempted by contacting experts and reviewing conference abstracts. MESH terms: Ovid Medline: ('Tomography Positron-Emission' OR 'Fluorodeoxyglucose F18') AND ('Fever' exploded). EMBASE: ('Positron Emission Tomography' OR 'Fluorodeoxyglucose F18') AND ('Fever' exploded). Keyword searches for ('Positron Emission* OR 'PET*' OR 'fluorodeoxyglucose*' OR 'fludeoxyglucose*' OR '18fluorodeoxyglucose*' OR 'fdg*' OR 'ffdg*' OR '18fdg*' OR '18ffdg*' OR '(18)ffdg*' OR '(18)fdg*' OR '2fluoro2deoxyglucose*' OR '2 fluoro 2 deoxyglucose*' OR '2 fluoro 2 deoxy d glucose*') in combination with ('Fever' OR 'Pyrexia' OR 'Febrile' OR 'PUO' OR 'FUO').
Study selection: One author (TB) performed the de-duplication of records in EndNote XL, screened titles and excluded irrelevant publications. TB reviewed abstracts and/or full texts to identify eligibility for inclusion in the qualitative synthesis.

Supplements

1) Quality Assessment Tool

http://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/case_series

Quality Assessment Tool for Case Series Studies

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the study question or objective clearly stated?			
2. Was the study population clearly and fully described, including a case definition?			
3. Were the cases consecutive?			
4. Were the subjects comparable?			
5. Was the intervention clearly described?			
6. Were the outcome measures clearly defined, valid, reliable, and implemented consistently across all study participants?			
7. Was the length of follow-up adequate?			
8. Were the statistical methods well-described?			
9. Were the results well-described?			

Quality Rating (Good, Fair, or Poor)
Rater #1 initials:
Rater #2 initials:
Additional Comments (If POOR, please state why):

*CD, cannot determine; NA, not applicable; NR, not reported

2) Data extraction form

Study ID

First author

Year of Publication

Country

Sample size

Start Year

Duration (in months)

Age range and Median age

Percentage of Female patients included

Study design and inclusions:

Case definition for FUO

Duration of symptoms prior to FDG-PET/CT

Inpatient stay prior to FDG-PET/CT

Study excluded immunocompromised patients

Study design (Retrospective; Consecutive; In/outpatients)

Patients excluded due to missing data and explanation

Prior diagnostic investigations documented

Outcomes:

Primary outcome: FDG-PET/CT Diagnostic Yield

Secondary Outcomes:

- Abnormal FDG-PET/CT
- False Positives

Final Diagnosis

-Infection

-Inflammation

-Malignancy

Mortality

Prior CT

Diagnostic yield over CT

%abnormal inflammatory markers in the group with diagnostic yield

Basis of diagnosis

Outcome

Follow-up

3) Delphi survey

See attached documents

4) Studies included in the qualitative synthesis

	Author/ Year	Country	Sample size	Study design	Inclusion in Meta-analysis
1.	Balink 2009	Netherlands	68	Retrospective case series	Yes
2.	Becerra Nakayo 2012	Spain	20	Retrospective case series; Only immunocompetent	Yes
3.	Bharucha 2013	UK	33	Retrospective case series; Only immunocompetent	No- Reported dIfferent outcome.
4.	Buch-Olsen 2014	Netherlands	57	Retrospective case series	Yes
5.	Castaigne 2009	Belgium	10	Retrospective case series	No- Only HIV patients and only reviewed abnormal scans.
6.	Crouzet 2012	France	79	Retrospective case series; Only immunocompetent	Yes
7.	Ergul 2011	Turkey	24	Retrospective case series; Only immunocompetent	Yes
8.	Federici 2010	France	10	Retrospective case series; Only immunocompetent	Yes
9.	Ferda 2010	Czech Rep.	48	Retrospective case series	Yes
10.	Gafter-Gvili 2015	Israel	112	Retrospective case series	Yes
11.	Jasper 2010	Germany	30	Retrospective case series	No- Combined results for FDG-PET and FDG-PET/CT
12.	Kei 2010	Singapore	12	Retrospective case series	Yes
13.	Keidar 2008	Israel	48	Prospective case series; Only immunocompetent	Yes

14.	Kim 2012	South Korea	48	Retrospective case series; Only immunocompetent	Yes
15.	Kubota 2011	Japan	81	Retrospective case series	Yes
16.	Manohar 2013	India	103	Retrospective case series	Yes
17.	Martin 2013	Belgium	20	Retrospective case series	No- Only HIV patients and only reviewed abnormal scans.
18.	Pedersen 2012	Denmark	22	Retrospective case series; Only immunocompetent	Yes
19.	Pelosi 2011	Italy	24	Retrospective case series; Only immunocompetent	Yes
20.	Pereira 2016	Switzerland	76	Retrospective case series	Yes
21.	Sheng 2011	China	48	Prospective case series; Only immunocompetent	Yes
22.	Tokmak 2014	Turkey	25	Retrospective case series; Only immunocompetent	Yes

5) Subgroup analysis of the primary outcome, Diagnostic Yield

Figure A: Subgroup Analysis for Diagnostic Yield of FDG-PET/CT only in adults with FUO, (n=15), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

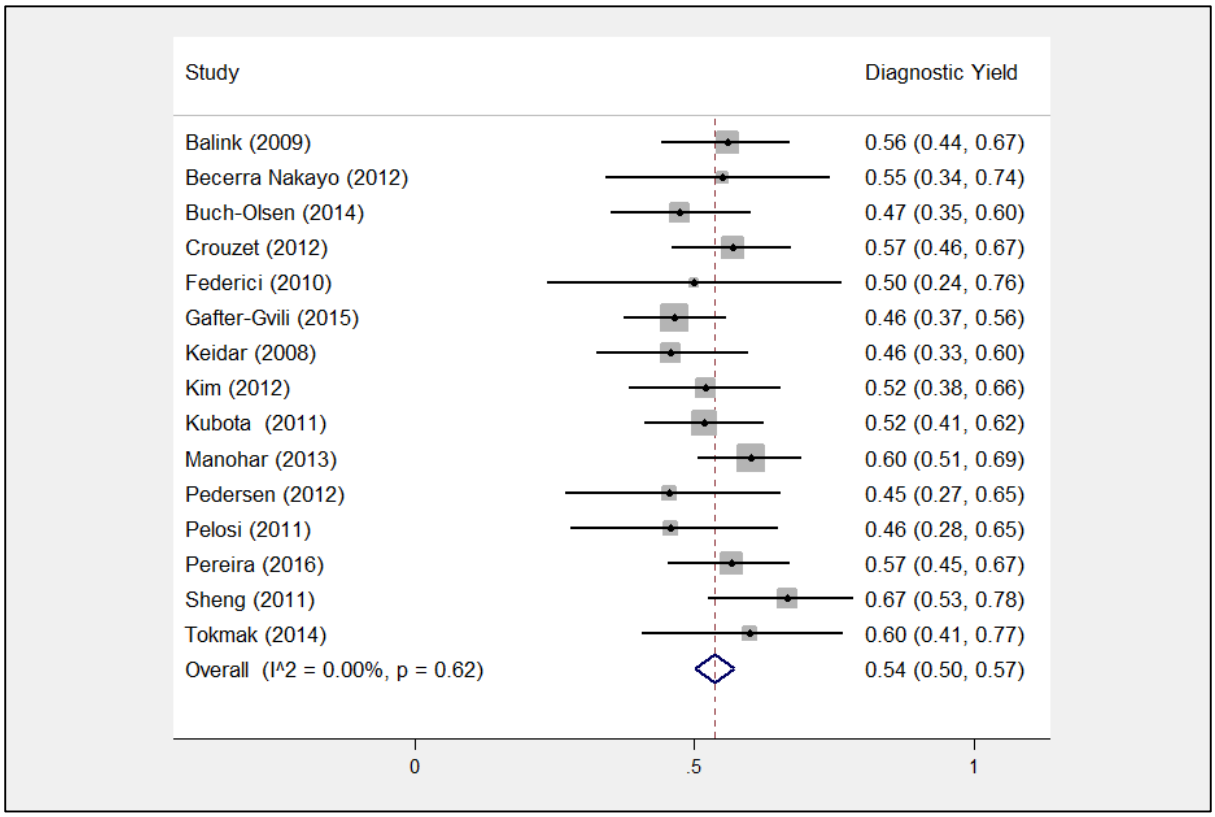


Figure B: Subgroup Analysis for Diagnostic Yield of FDG-PET/CT only in immunocompetent patients with FUO, (n=10), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

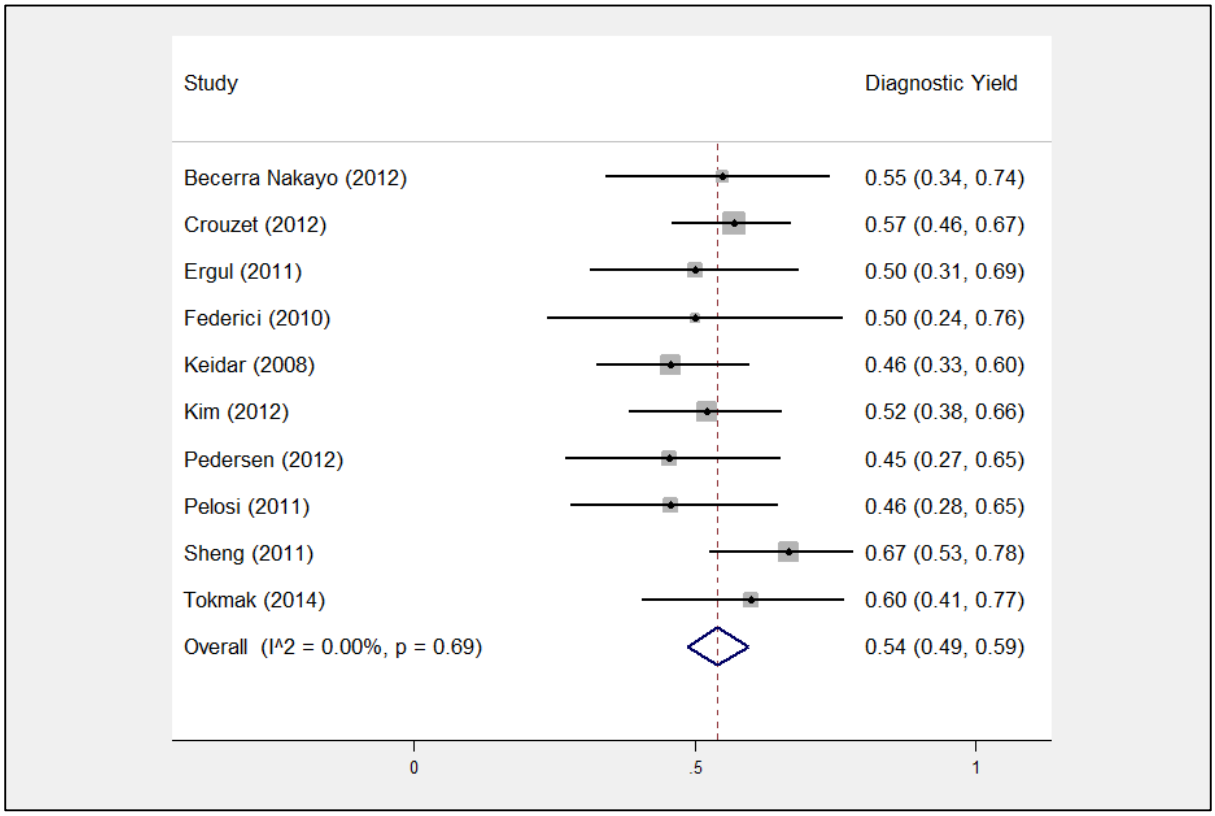


Figure C: Subgroup Analysis for Diagnostic Yield of FDG-PET/CT only in immunocompetent adults patients with FUO, (n=9), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

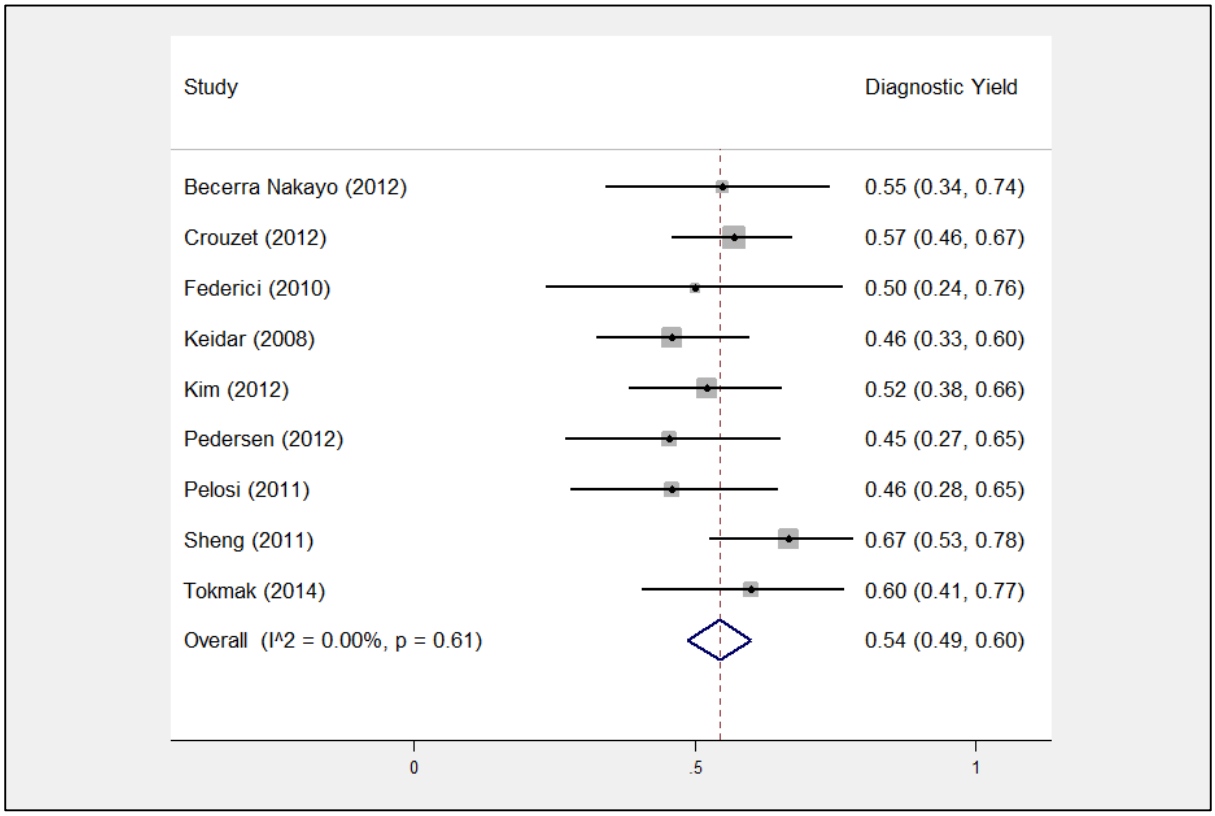
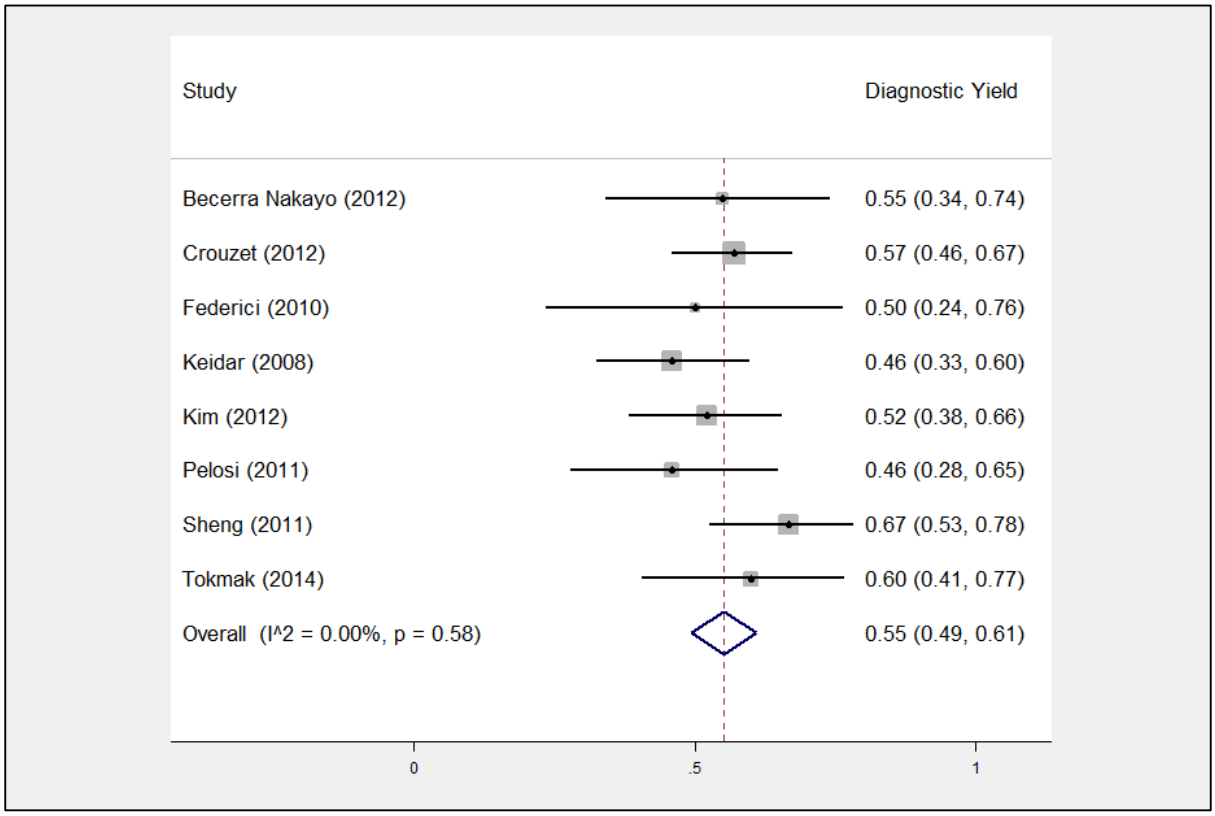


Figure D: Subgroup Analysis for Diagnostic Yield of FDG-PET/CT only in immunocompetent adults with FUO without contrast, (n=8), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.



5) Forest plots for secondary outcomes:

Figure E: Abnormal FDG-PET/CT in Fever of Unknown Origin (n=18), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

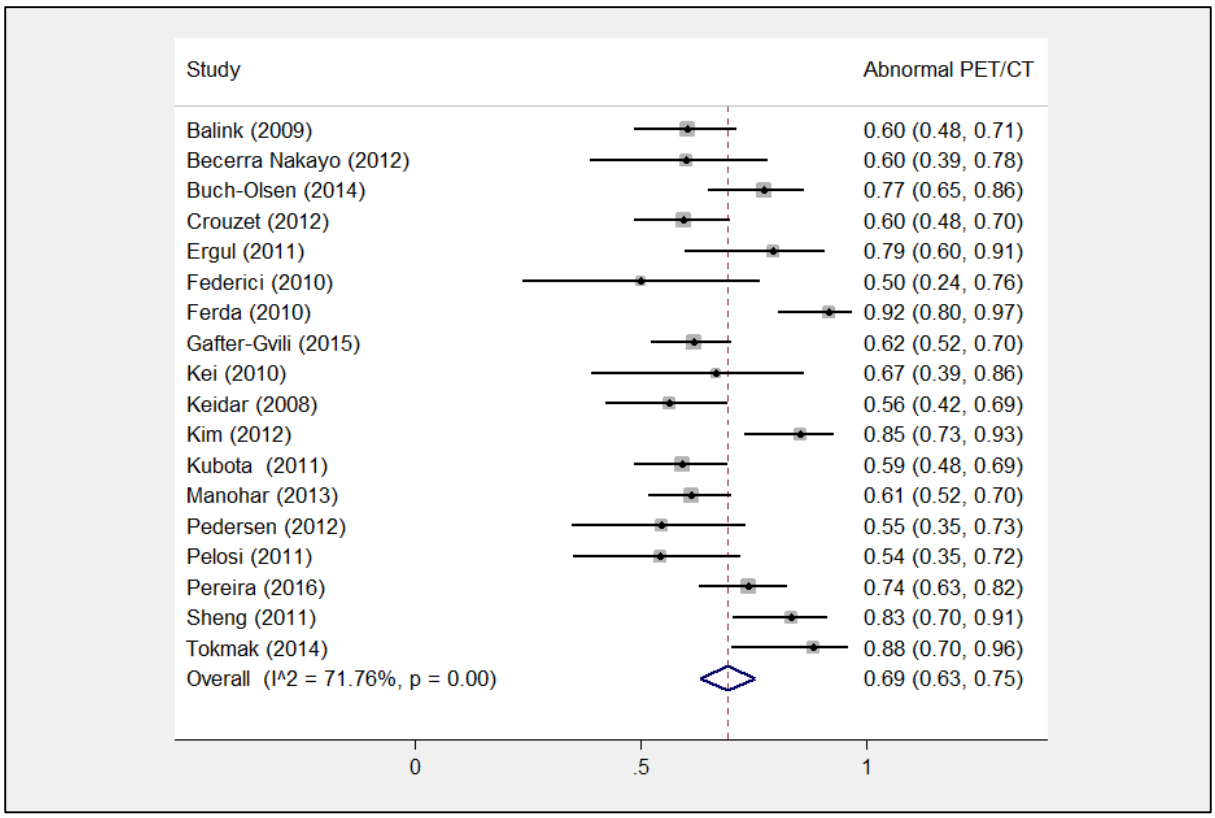


Figure F: False Positives of FDG-PET/CT in Fever of Unknown Origin (n=18), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

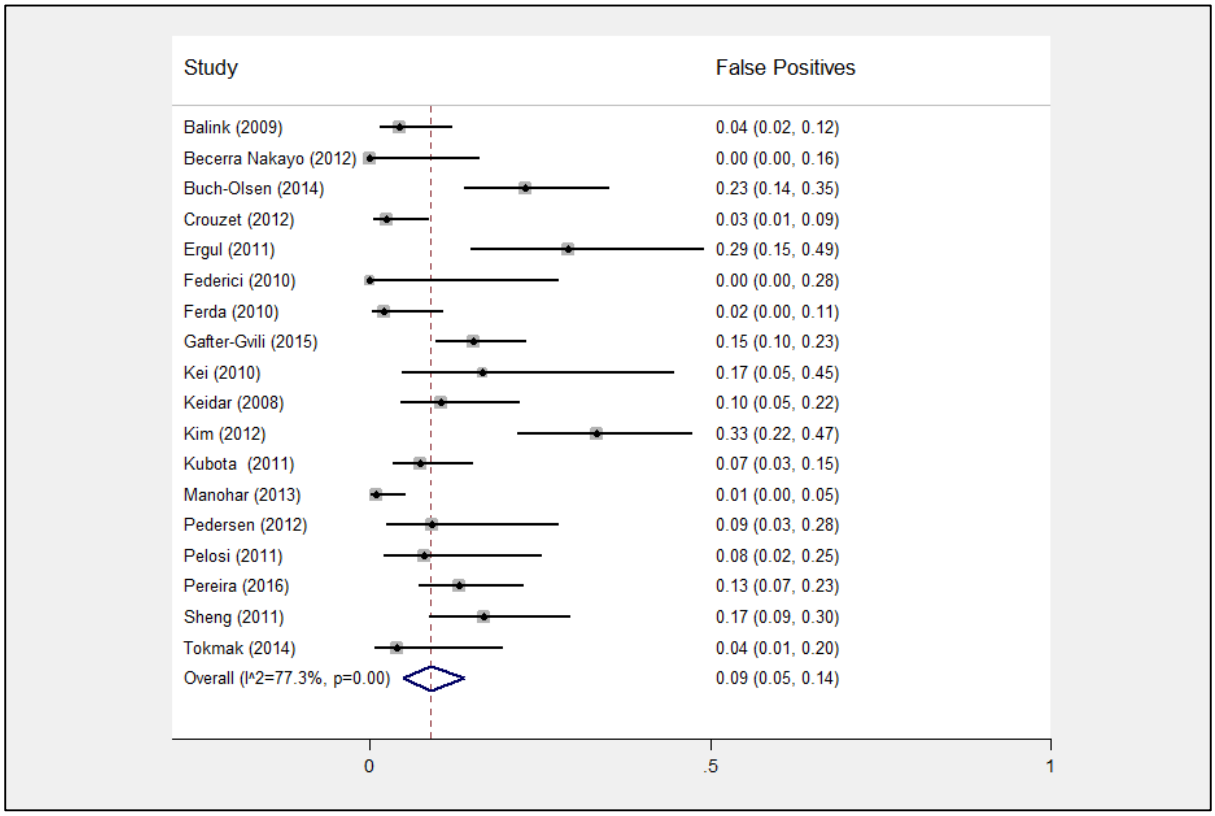


Figure G: Final Diagnosis of Fever of Unknown Origin identified (n=18), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

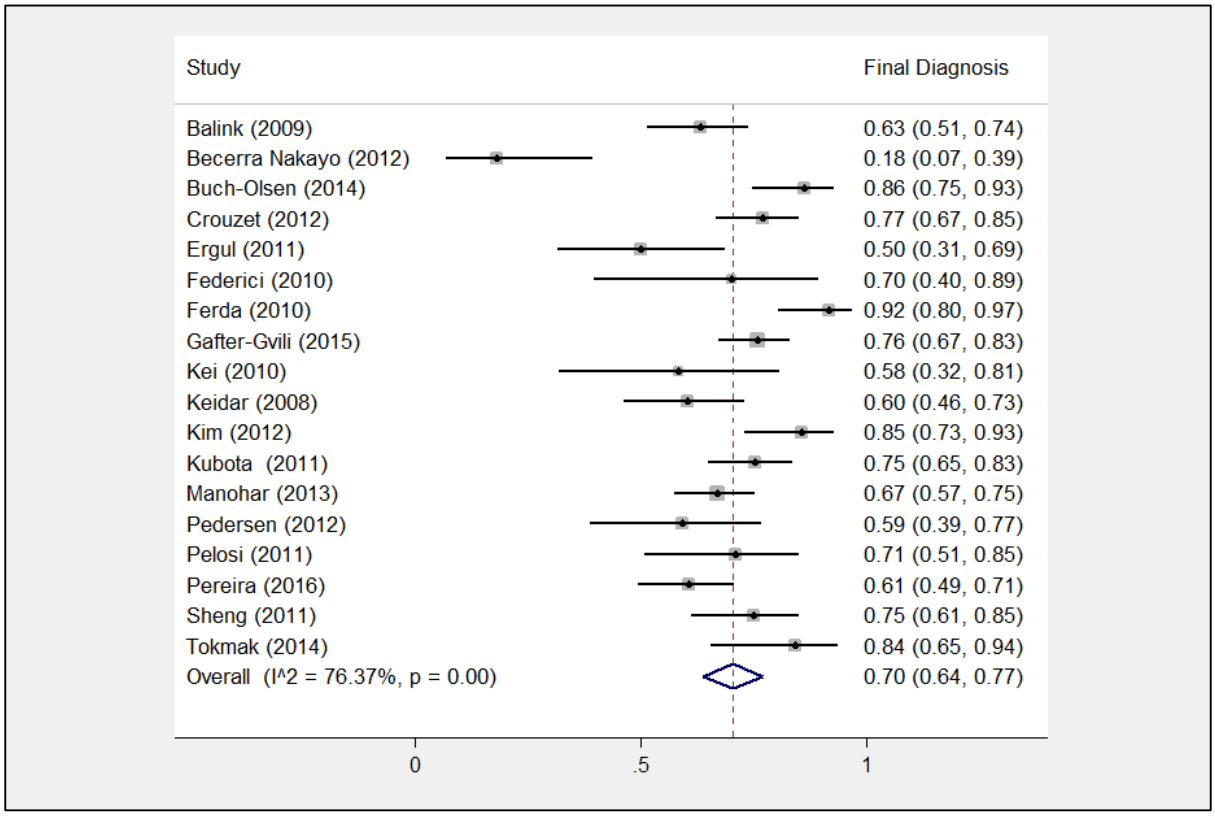


Figure H: Infectious Final Diagnoses in Fever of Unknown Origin (n=18), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

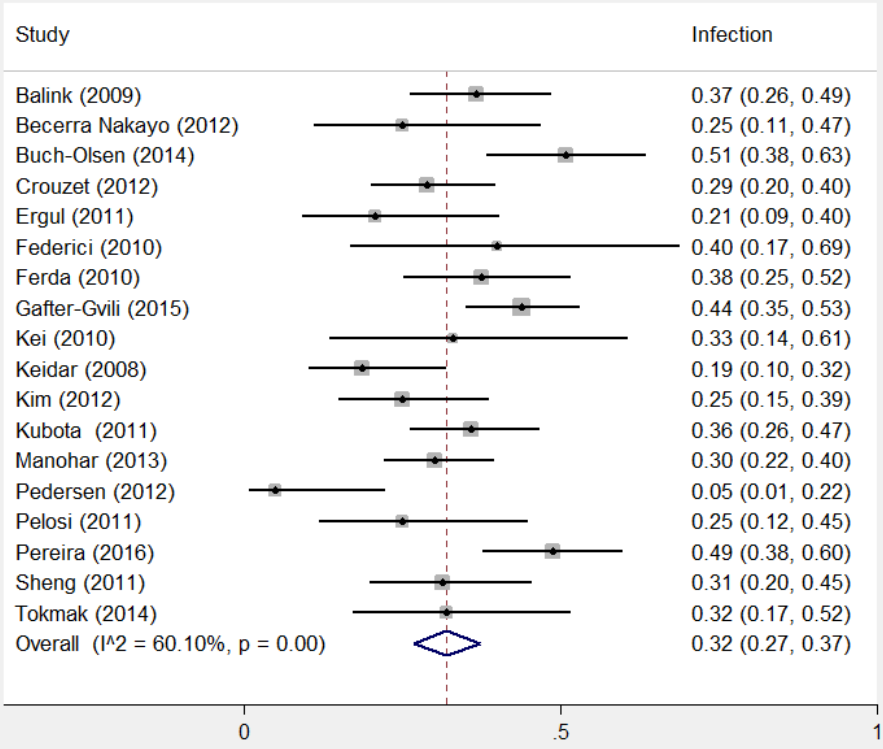


Figure I: Inflammatory Final Diagnoses in Fever of Unknown Origin (n=18), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.

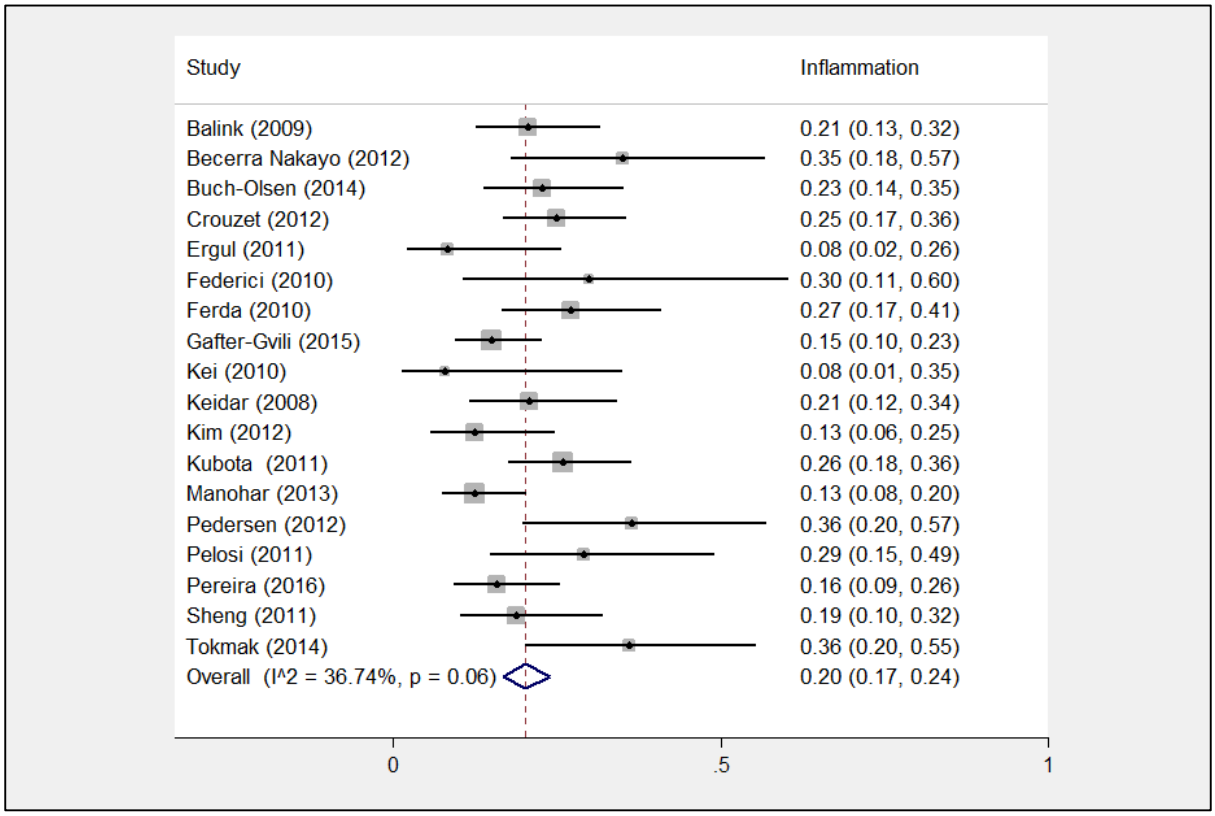
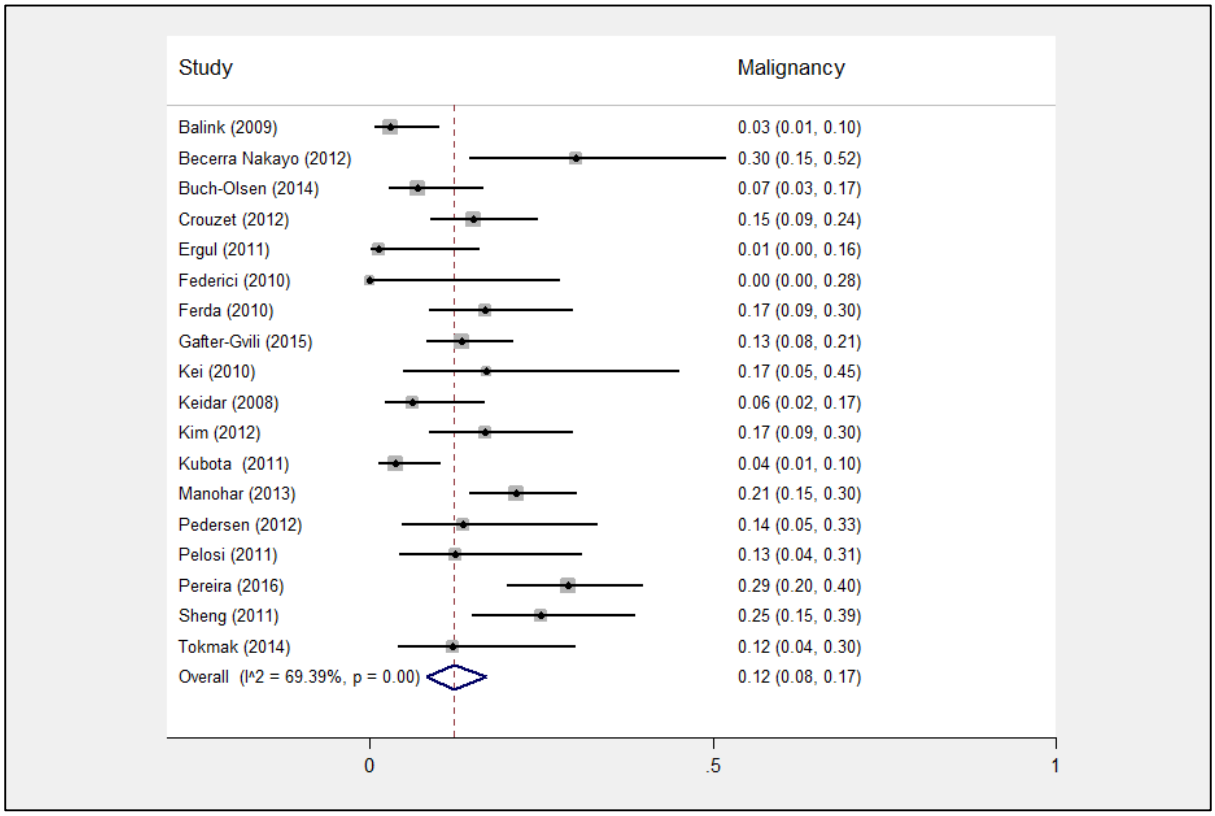


Figure J: Malignancy as Final Diagnoses in Fever of Unknown Origin (n=18), Proportion 0=0% to 1=100% +/- 95% CI. The size of the grey box provides a measure of the sample size. I^2>50% implies moderate heterogeneity.



Highlights

- A systematic review identified 18 eligible studies, 905 patients, of FDG-PET/CT in FUO
- Pooled diagnostic yield was 56% (95%CI 50-61%), I2=61%
- Sub-group analysis of diagnostic yield over conventional CT was 32% (95%CI 22-44%) I2=66%
- Iterative Delphi Surveys generated interspeciality consensus on the topic.
- There is insufficient evidence to support the value of FDG-PET/CT in investigative algorithms