Reputations Count: Why benchmarking performance is improving health care across the world

February 2018

Gwyn Bevan

Professor of Policy Analysis

Department of Management

London School of Economics and Political Science

London WC2A 2AE

R.G.Bevan@lse.ac.uk

Alice Evans

Lecturer in the Social Science of Development

King's College London

Department of International Development

Strand Building

Strand Campus

London WC2R 2LS

alice.evans@kcl.ac.uk

Sabina Nuti

Professor of Health Management

Laboratorio Management e Sanità, Institute of Management,

Scuola Superiore Sant'Anna, Pisa

Email: s.nuti@santannapisa.it

Forthcoming in *Heath Economics Policy & Law*.

Reputations Count: why benchmarking performance is improving health care across the world

Abstract

This paper explores what motivates improved health care governance. Previously, many have thought that performance would either improve via choice and competition or by relying on trust and altruism. But neither assumption is supported by available evidence. So instead we explore a third approach of reciprocal altruism with sanctions for unacceptably poor performance and rewards for high performance. These rewards and sanctions, however, are not monetary, but in the form of reputational effects through public reporting of benchmarking of performance . Drawing on natural experiments in Italy and the UK, we illustrate how public benchmarking can improve poor performance at the national level through 'naming and shaming' and enhance

good performance at the sub-national level and through 'competitive benchmarking' and peer learning. Ethnographic research in Zambia also showed how reputations count. Policy-makers could use these effects in different ways to improve public services.

Reputations Count: why benchmarking performance is improving health care across the world

> 'Political writers have established it as a maxim, that in contriving any system of government, and fixing the several checks and controls of the constitution, every man ought to be supposed a *knave* and to have no other end, in all his actions, than private interest. By this interest, we must govern him and, by means of it, notwithstanding his insatiable avarice and ambition, co-operate to the public good' (Miller and Hume, 1994, p42-43, italics in original).

Introduction

Under pressures of austerity, it is vital that systems of health care are governed effectively with incentives to tackle performance where it is unacceptably poor and improve on good performance. Attempts from the 1990s to create incentives to improve performance through provider competition for hospitals have been

found wanting. (And provider competition is limited in its scope for much of health care.) The reason for these attempts was the recognition that trusting providers to improve without any external pressures allowed variations in performance to continue. So, as neither governance by provider competition nor trusting providers seems to have worked effectively, we need to look for an alternative and we argue in this paper for governance based on the principles of reciprocal altruism. This form of governance follows conventional micro economics in its design of rewards and sanctions for good and bad performance but differs in that these are non monetary and based on reputation effects. We present evidence of these impacts from 'natural experiments' testing different models of governance between the different countries of the UK, following devolution in 1999, and regions in Italy, which have had autonomy in the governance of health care since their creation in 1974. In Zambia we show how reputation concerns have galvanised a rapid reduction in maternal mortality. We conclude by relating the empirical evidence we report here to developments in behavioural economics and the conceptual framework of reciprocal altruism (Gintis et al, 2005; Bowles, 2016; Oliver, 2017), using the concepts of identity (Akerlof and Kranton, 2010), and reputation effects from 'naming and shaming' for poor performance (Bevan and Fasolo, 2013) and awards for high performance (Frey, 2013).

The next section of this paper discusses two models of governance: by Trust and Altruism (T&A), and Choice and Competition (C&C). We present evidence that the latter model has not delivered the expected improvements in the English National Health Service (NHS). We discuss how reputation effects can explain why public

reporting of performance may, or may not, galvanise improvements. We examine two different kinds of reputation effects through public reporting of benchmarking performance: of 'naming and shaming' failures in England and in the form of competition for awards for high performance in the region of Tuscany in Italy. The 'natural experiments' in models of governance between the countries of the UK and the Italian regions give evidence of the power of reputation effects and the weakness of T&A. These effects were also found through ethnographic research in Zambia. The evidence from Italy also shows the weakness of C&C. The final section interprets the evidence we have presented using concepts from behavioural economics and social psychology.

Governance by Trust and Altruism or Choice and Competition

Two models of governance

This section outlines the two dominant models of governance, which were identified in the 1990s, and have been subsequently evaluated, using Le Grand's argument about 'knights' and 'knaves' in public policy (Le Grand (2003) goes back to Hume's observation, which is the epigraph at the start of this paper.

> *Trust and Altruism (T&A)*: As encapsulated by Le Grand (2003), the T&A model, unlike what Hume advocates, assumes that those who work in government and health services are 'knights', intrinsically motivated to do their best for those they serve. On this basis, performance will improve with knowledge and financial resources, without sanctions for failure or

rewards for success. Berwick et al (2003) highlighted the weakness of this model from its lack of any external incentives to overcome the inertia generated by obstacles to improving performance, which depends on understanding why performance was relatively poor; and, having gained such understanding, implementing the necessary changes. Furthermore, as has been found in the UK, the logic of this model has resulted in perverse incentives as governments reward failure with extra resources: the logic being that 'failure' cannot be due to want of effort (Bevan, 2014).

Choice and Competition (C&C): This model holds that patients act as informed consumers or insurers selectively contract or both. This model further postulates that hospital performance affects their market shares, which generates financial incentives to improve. Le Grand (2007) suggests that this model rewards high-performing 'knights' and penalises poorly-performing 'knaves'. As Berwick et al. (2003) argue, for patient choice, this model depends on a series of assumptions about patients acting as consumers of health care: that patients are aware that of differences in performance exist, have access to and can interpret such information, and act on it.

Evidence for the C&C model

A good test of the C&C model is how publishing information on risk-adjusted mortality rates for cardiac surgery in New York State by its Cardiac Surgery Reporting System (CSRS) (which uses a state of the art method of risk adjustment)

has affected hospitals' market shares. Given the common obstacles to patients acting as consumers as identified by Berwick et al. (2003) this offers the most propitious circumstances for consumer choice: the USA is one of the world's most market-driven systems; the CSRS has consistently shown significant differences in performance; this information has been widely publicised and is accessible; and, in New York City, patients have ample choice. But, as Chassin (2002) argued, public reporting has had no impact on the market share of hospitals identified as 'outliers': i.e. with statistically significant high or low risk-adjusted mortality rates. Systematic reviews on public reporting have also found that patients have not acted as consumers (Marshall et al, 2000; Fung et al, 2008). The other potential driver in the C&C model is 'commissioning': selective contracting by insurers or local health authorities. Chassin pointed out that 'Managed care companies did not use the data in any way to reward better-performing hospitals or to drive patients toward them'. Ham's review of 'commissioning', (Ham, 2008), found that 'Experience and available evidence from Europe, New Zealand and the US indicates that in no system is commissioning done consistently well'. This finding is consistent with later studies in England (Smith and Curry, 2011) and the Netherlands (Maarse et al, 2016).

Two econometric studies have found, however, that when the C&C model was reintroduced by the NHS in England from 2006, hospitals subject to greater competition recorded higher improvements in 'quality' (Cooper et al., 2011; Gaynor et al., 2013). The first of these studies was cited by the then Prime Minister in justifying a further development of governance by C&C implemented from 2014 (Pollock et al, 2011). Both econometric studies used, as their principal

measure of quality of care, reductions in mortality rates from Acute Myocardial Infarction (AMI). This is problematic as a measure of the efficacy of the C&C model because neither patients nor their carers choose hospitals: ambulances follow clear rules which hospitals are best to save that patients' lives (Bevan and Skellern 2011). A later econometric study of the impact on competition for elective surgery, for which C& C ought to apply, measured improvements in quality of life following surgery using Patient Reported Outcome Measures (PROMs). This found that hospitals subject to greater competition recorded *lower* improvements in quality for varicose veins, hip and knee replacement; with no difference for groin hernia repair surgery (Skellern, 2016).

The C&C model has high transaction costs and is limited in scope as much of health care is for those who are elderly, with chronic conditions for whom what matters is a good local integrated service. And even where the C&C model ought to have an impact, in elective surgery, there is no evidence that it has improved quality of hospital care. The C&C model has been abandoned after one attempt in New Zealand (Ashton et al, 2005), Scotland, Wales and Northern Ireland; after three attempts in England (Bevan, 2014; Anonymous, 2017); and after having been tried in Italy in the region of Lombardy only, has been substantially modified (see below). We hence argue that there is little evidence that the C&C model has been an effective model of governance.

Governance by reputation

The puzzle of public reporting: the importance of reputation

While C&C has been found wanting, we also know that public reporting can sometimes, but not always, improve performance (Fung et al, 2008). This is puzzling for two reasons. First, why – if not by influencing market shares – might public reporting motivate better health care? Second, why does the impact of public reporting vary? Hibbard et al (2003, 2005) explored these conundrums through a controlled experiment in Wisconsin. The first set of hospitals were given no information on quality; the second set were given it privately (i.e. not published); and for the third set, great efforts were made in publishing their comparative performance. Hibbard et al also found that in the third set only did hospitals make considerable efforts to improve and that the reason was that public reporting had damaged their reputations, but not their market shares.

Hibbard (2008) further suggests that for public reporting to improve poor performance by inflicting reputational damage, these reports are required to be made easily and widely accessible on a regular basis and rank performance clearly so that everyone can easily see which hospitals are performing well and poorly. And, as Bevan and Hamblin (2009) argued, Chassin (2002) found that the publication of risk-adjusted mortality rates by the Cardiac Surgery Reporting System of New York state did have an impact on providers that were publicly reported to be outliers with high rates through the damage this caused to their reputation (and not market shares).

In sum, although the C&C model has been tried and found wanting, where public reporting has been found to motivate improved performance of US hospitals, this is caused by reputational concerns. The remainder of this paper unpacks this effect through a wider evidence base: the UK, Italy and Zambia. It explores the ways and contexts in which reputational concerns can galvanise better performance, considering both sub-national and national effects.

The impact of benchmarking in the UK

Context

The 'new' Labour Government had been elected in 1997 with the promise to 'save the NHS' and abandon the C&C model of governance for that of T&A. Despite that promise, two year's later, in the winter of 1999-2000, the NHS was perceived to be in the midst of a 'crisis'. Clive Smee, Chief Economist in the Department of Health described the OECD's major review of UK health care as having 'highlighted poor cancer survival rates in the UK, suggested that other disease-specific outcomes were also poor, and noted the limited progress on waiting times and the apparent under-investment in both doctors and buildings... drew the conclusion that the NHS was underfunded'. Smee also points out in a footnote that: 'In private the authors went further and indicated that they had been unable to identify any features of the NHS that were particularly commendable' and that 'On 16 January 2000, while the OECD report was still in draft, the prime minister made his seminal commitment to match the average

health expenditure levels of the European Union by 2006/07' (Smee, 2008, p. 92). In addition to this commitment to sustained generous increases in funding the Government abandoned the T&A model for the brutal regime of annual performance 'star' ratings from 2000 to 2005. This regime combined 'naming and shaming' with 'targets and terror', sacking chief executives of poorly performing trusts (Stevens, 2004; Bevan and Hood, 2006). Each devolved government in the other countries of the UK decided to follow England's policy of increasing NHS funding substantially, but not to abandon the T&A model. The best comparison in terms of a 'natural experiment' is between England and Wales, as these countries were, prior to devolution, subject to the same legislation, and similar organisations and levels of funding (Bevan et al, 2014).

A British 'natural experiment' of models of governance

The 'star rating' system in the English NHS for acute hospitals consisted of: assessments of the implementation of clinical governance by the quality regulator (the Commission for Health Improvement); nine 'key targets' (dominated by waiting times); and about forty indicators in total, in three domains of a 'balanced scorecard', which included more targets for waiting times, clinical outcome indicators, and results of surveys of patients and staff. The targets for waiting times became progressively more demanding over time: in 2001/02 (Department of Health, 2002) and 2005 (Auditor General for Wales, 2005) these were 26 and 13 weeks for outpatients, and 78 and 26 weeks for inpatient admission; and for 2008, the target from GP referral to treatment (including diagnostic assessment) was 18 weeks (see Figure 1).

The ‘star rating’ regime satisfied the requirements of Hibbard et al (2003) for a system to inflict reputational damage on poor performance. This was a simple ranking system: to be zero-rated a trust would fail in clinical governance or more than one ‘key target’ or both; to be rated as three-star, a trust would perform satisfactorily in clinical governance, against ‘key targets’, and across the ‘balanced scorecard’. ‘Star ratings’ were published online, in national and local media, as well as professional journals (the *British Medical Journal* for physicians and the *Health Service Journal* for managers). This publicly accessible information was simple to understand: ranging from zero-rated (‘failing’) to three-star (‘high performing’). In the first year of star ratings published in the autumn of 2001, the 12 zero-rated acute trusts were ‘named and shamed’ as the ‘dirty dozen’. Six of their chief executives were sacked.

In contrast, the lax regime of the government in Wales exemplified governance by the T&A model: there was confusion over which targets were important (with over 100 in 2003-04), waiting time targets were not consistently applied and breaches were allowed, there was no ranking system, nor indeed any systematic reporting to the public of trusts’ relative performance on waiting times. Not only was the government keen to avoid any reporting system that ‘named and shamed’ failing providers, there was also a widespread perception amongst NHS managers in Wales that failure to achieve waiting time targets would be rewarded with extra resources (Auditor General for Wales, 2005).

Outcomes

In acute NHS trusts in England, waiting times were radically reduced. In England in 2004, only 37 patients were waiting more than 17 weeks before being admitted (Department of Health, 2005); but, in Wales in 2005, over 7,000 patients were waiting more than 18 months (Auditor General for Wales, 2005). Figure 1 shows the hospital waiting time targets for first being seen in outpatients after referral by a General Practitioner (GP) and for admission afterwards for England for 2001/02 (26 and 78), by December 2005 (13 and 26), for Wales in 2005 (78 and 78), and for the whole waiting time from GP Referral to Treatment (RTT) in England by 2008 (18); there was no 'similarly clear strategy' in Wales for reducing 'target waiting times over the medium term' (Department of Health, 2002; Auditor General for Wales, 2005).

Hood and Dixon (2010) suggest that improved performance of public services in England brought no political benefits in terms of public support for the government; and the relative failures in Wales brought no political costs. After just two years of top-down reform, in 1999, the Prime Minister Tony Blair famously described this painful process: 'I bear the scars on my back' (BBC News, 2007). Michael Barber (2007), leader of the Prime Minister's delivery unit from 2000 to 2005 argued that top-down reform 'done well can rapidly shift a service from 'awful' to 'adequate', as in the case of NHS waiting times in England. But, he agreed with Blair, that 'flogging' the system was insufficient to motivate excellence. From 2006, the NHS in England had another attempt at C&C. However, as we argued above that alternative has been found wanting.

Hence the evidence from the UK suggests that neither governance by T&A nor C&C has improved performance. But the whilst the combination of through 'naming and shaming' and 'targets and terror' did indeed achieve rapid results, this top-down regime (focused on penalizing failure) was not politically sustainable.

Figure 1: Hospital waiting time targets for England and Wales to go about here

This poses two questions, which we consider in the following two sections of this paper on seeking improvement through reputation effects. First, can public reporting work in quite different socio-political contexts? We examine this question by looking at benchmarking that identified poor performance for maternal mortality in Zambia in the next section. The section after that looks at how in Tuscany the system of performance reporting has developed to become one in which the different organisations compete to be high performing. Here the strength of the reputation effects is more in the form of an award, which was recognised in performance-related pay for their chief executives.

The impact of benchmarking in Zambia

Context

The story of the impact of benchmarking in Zambia has strong parallels with that of the English NHS in the 2000s. In each case international comparisons shone a spotlight on poor performance in each country and hence showed what could be achieved with extra resources coupled with performance management. For Zambia this spotlight was on its appallingly high rates of maternal mortality from

its comparative performance on the Millennium Development Goal 5 for reductions in maternal mortality (MDG 5).

The reputation effects of international and national ranking systems

As Le Grand (2010) notes, we need to be careful in deducing motivations 'by simple observation of the changes concerned'. This section thus draws on ethnographic research (interviews and observation in Zambia's Ministry of Health) to illustrate how health care managers and workers perceive and respond to public disclosure of information about other countries' performance (Evans, forthcoming). Importantly, the researcher did not set out to explore the impact of benchmarking nor introduce this topic in interviews. Participants were merely asked about the Zambian Government's health care priorities: how and why these have changed over recent decades. The subject of 'MDGs' was introduced by participants.

Historically, many Zambian health workers and managers regarded maternal mortality as inevitable. But such fatalism waned upon seeing rapid improvements in other African countries (as publicised by the MDG process). Many found this comparative data inspirational. Evidence of improved outcomes also demonstrated that other African governments were prioritising maternal health. This provided external legitimisation of their efforts to tackle this hitherto neglected health issue. Seeing peers make more rapid progress towards shared targets (indicators of 'progress' and 'Development') also induced reputational concerns. The Zambian Government did not want to lag behind. 'No, Zimbabwe can't do better than us!' exclaimed one Maternal and Child Health co-

ordinator. Likewise, once parliamentarians were shown regional statistics, they introduced a separate budget line for reproductive commodities.

When asked how government health care priorities had changed over recent decades, all participants (health care workers, district administrators and senior managers) emphasised increased attention to maternal health. Further indicators of prioritisation include the Ministry of Health institutionalising MDG Target 5.2 as its own Performance Assessment Indicator; 'a national programme to strengthen Emergency Obstetric and Neonatal Care (launched in 2007); a National Reproductive Health Policy (Ministry of Health, 2008); Maternal Death Reviews in all districts (2009); a separate budget line for reproductive health and commodities (2009); direct funding to institutions training health professionals (2009); an annual 'Safe Motherhood' week and obligatory inclusion of activities to promote Maternal Neonatal and Child Health (MNCH) in district action plans (2010); increased government expenditure on family planning commodities (Ministry of Health, 2008; Mukonka, 2012; Mukonka et al. 2014); the 'Eight-Year Integrated Family Planning Scale-Up Plan, 2013-2020' (2012); and the 'Road Map for Accelerating the Reduction of Maternal, Newborn and Child Mortality 2013-2016' (2013)' (Evans, forthcoming).

That international benchmarking can induce reputational concerns has been observed more widely. Sakiko Fukuda-Parr (2104, p 123), formerly lead author of the Human Development Reports of the United Nations Development Programme (UNDP), observes that '[c]ountries are keen to present their MDG records in international fora to bolster their standing. Countries prepare MDG progress

reports for international consumption, some for this purpose only rather than for national development planning and monitoring. The Prime Ministers of India and China have come to present and showcase their MDG reports at high-profile UN events'. Sarwar (2105) likewise argues that the Indonesian and Mexican governments sought to achieve the MDGs in order to secure their reputations as regional leaders.

These testimonies suggest that public disclosure of health outcomes shifted norm perceptions (beliefs about what others think and do). Zambian civil servants and politicians seemed especially concerned and motivated by the successes of other African countries, which they regarded as peers. This echoes insights from social psychology: people are keener to conform to the norms of a group with which they identify (Tankard and Paluck, 2016, p 196). This process of peer learning was enabled through regional events, such as the 2010 African Union Summit (which focused on MDG 5). Only by collectively deliberating and developing an 'African' agenda, to tackle common problems, did maternal health become a continental priority.

Benchmarking also seemed powerful at the subnational level. When district health officers gathered at provincial meetings (which increasingly focused on maternal health indicators), no one wanted to be at the bottom of the table. It would be embarrassing in that context, and there were also concerns about career progression. This incentivised increased attention. Meanwhile, those who had comparatively excelled took pride in recognition of their accomplishments – by provincial and central government.

Benchmarking seems to enable top performers proudly to present their success, bask in its glow and be publicly recognised. It also reveals inspirational possibilities, motivating improvements among poorly performing hospitals and countries. This finding may allay concerns that league tables might be inherently punitive and demotivating (Oliver, 2015).

Outcomes

Maternal mortality rates were reduced by 61% in Zambia from 1990 to 2015. The estimated maternal mortality ratio 'decreased to 541 (in 2000), 372 (2005), 262 (2010) and 224 (2015)' (World Health Organization (WHO), UNICEF, UNFPA, World Bank Group and the United Nations Population Division, 2015). Although Zambia missed hitting its MDG, this was nonetheless one of the largest declines and lowest contemporary ratios in Sub-Saharan Africa (ibid). Further, between 2007 and 2014, skilled birth attendance increased from 47% to 64% (CSO et al 2015:127). The percentage of women using family planning has also steadily increased over the past two decades: 15 (1992), 26 (1996), 34 (2001-2002), 41 (2007), to 49.0 (2013-2014) (CSO et al 2015:93). Additionally, the total fertility rate has reduced: from 6.2 in 2007 to 5.3 in 2013 (CSO et al 2015:70)'.

That said, international benchmarking is no magic bullet. This can be seen by comparing progress over time and across countries. In the early 2000s, internationally benchmarked data on maternal mortality was rapidly skimmed over in Zambia's national planning meetings. They focused more on indicators prioritised by cooperating partners: malaria, HIV/AIDS and TB. Clearly, health

care managers cannot prioritise all health care issues. Interest in specific publicly disclosed outcomes is clearly mediated by pre-existing ideologies and priorities (on the part of the political executive, co-operating partners, and civil servants), as well as donor relations, aid modalities and resources (Evans, 2017).

Comparisons across countries provide further evidence that international benchmarking is not inherently motivating. Relative to other African countries, Zambia's overall progress on MDG 5 was particularly rapid (WHO et al, 2015). Further, comparative research would shed light on why other countries (whose health care outcomes were also publicly disclosed) did not oversee such substantial improvements.

The impact of benchmarking in Italy

Context

The Italian National Healthcare System (NHS) follows the Beveridge model: it is a public health system providing universal coverage for comprehensive and essential health services through general taxation. Since the early 1990s, a strong decentralization policy has been adopted in Italy and the state has gradually transferred its jurisdiction to its 20 regions (France and Taroni, 2005). Each elected regional council is responsible for deciding how its system of health care is governed although nearly all the funding is from central government (strongly analogous to the arrangements for the funding and governance of systems of health in the different countries of the UK). Because Italian regions have adopted

different models of governance, they offer an interesting 'natural experiment' to see their effects on performance (Nuti et al., 2015).

An Italian 'natural experiment' of models of governance

We focus on the system of governance that has been developed in Tuscany since 2006, where more than 95% of hospital beds are public. We use data from before the recent reorganization (in 2016), when there were 12 Local Health Authorities (LHAs), financed by the regional administration under a global budget with a weighted capitation system. At the heart of the its regional system of governance is benchmarking of performance in the Performance Evaluation System (PES), which was designed and developed by the Management and Health Laboratory (MeSLab) of Sant'Anna School of Advanced Studies. The PES is grounded on benchmarking, public disclosure of results, target setting and a rewarding system for managers. It is able to put strong pressure on clinicians through reputational competition to assure clinical quality. The origins of the development of the PES were to provide information for the regional councillor to decide performance-related pay for the Chief Executive Officers of each district. This objective means that the reputation effects of benchmarking in the Tuscan PES are from not 'naming and shaming' but the very different effects of awards, as described by Frey (2013). The Tuscan PES has two other fundamental differences from the English system of 'star rating'.

First, there is no single ranking of each district; instead the complex mix of performance across six dimensions is displayed in what has become the famous Pisa 'dartboard' diagram. Figure 2 uses the dartboard at the regional level to

compare Tuscany with Marche. This shows how the 'dartboard' indicates at a glance underachievement and high performances of each Region / District. Each indicator is evaluated compared to a national or an international standard, or where that is not available, good performance in the other regions. Each District or Region is scored for each indicator, ranging from 1 (poor performance) to 5 (excellent performance). The score is associated to a colour for each score: from red for 1, then orange, then yellow, then green and dark green (5) for excellent. Within each target, disaggregated results for each evaluation measure are displayed, and the closer the evaluation indicator is to the centre of the target, the higher its performance level. Hence Figure 1 shows Tuscany to have better performance than Marche. Over the years, the power of dartboard in communicating results with such clarity has resulted in its colours becoming a common language among managers, politicians and professionals in Tuscany where it was first adopted in 2006. From December 2007, all the performance indicators presented in benchmarking and the yearly targets linked to CEO rewarding system have been available online (http://performance.sssup.it) (Nuti et al, 2013).

Second, the Tuscan PES is organized at regional, not national level, and the results are presented to meetings of the senior managers and clinicians, and heads of departments of the districts and region every six months. These managers and clinicians are closely involved in the development of the indicators and are trained by the Sant'Anna School of Advanced Studies in the use of this information. The strategy for clinical engagement is based on creating a learning environment in a community of practice with systematic meetings to discuss

comparative performance of service utilization and outcomes, and receive constructive feedback (Spurgeon et al., 2011; Clark, 2012; Wenger et al, 2002). Examples of improvements through benchmarking include how reductions were made in diabetic-related rates of foot amputations (Nuti et al., 2016) and improving the communication processes between patients and clinicians through health-professional using results from patient surveys (Murante et al., 2014).

Hence the Tuscan PES has become embedded in a social process of collegial benchmark competition, which fosters learning, on any given indicator, for those who perform poorly from those who perform well. The Tuscan system has strong similarities to the league tables used from the mid-1990s to transform quality of care in the US Veterans Health Administration (VHA), the health care system that covers honourably discharged veterans of the US armed forces. The VHA was notorious for its low quality of care but was transformed into becoming high performing by 2005 based on the reputation effects of benchmark competition (Oliver, 2007). Oliver (2017) argues such ranking systems are instruments of negative reciprocity, which provide safeguards against knavish behaviour that would, if it were prevalent, undermine the cohesion and cooperation vital for quality improvement in health care.

In Italy, 13 regions, on a voluntary basis, have adopted the Tuscan model and agreed on the same set of indicators for benchmarking. Each region is part of a network of the Inter-Regional Performance Evaluation System (IRPES). Each region is responsible for processing its own data, in order to increase the

awareness and the expertise of the regional managers and their staff. The results are shown by region and by Health Authorities. In 2015, IRPES monitored the performance of approximately 100 HAs indicators. We now compare performance of the regions of Tuscany, Marche and Lombardy.

Marche, which joined the IRPES in 2008, relies for its model of governance on T&A: it neither uses a regional planning process for benchmarking, nor shares results through public disclosure, which means no threats to the reputations of clinicians who provide poor quality of care. Indeed the region makes little use of the IRPES and focuses more on the few indicators from the Ministry of Health used to calculate National LEA (Livella Essenziali di Assistenza: Essential levels of Care) grid scores. Each Region is required achieve a minimum of 160 points in its grid score. Figure 1 above shows how the performance of Marche in 2015 was worse than Tuscany on most indicators.

Lombardy, which joined the IRPES in 2015, is the only Italian Region that has followed a C&C governance model. It adopted a quasi-open-market healthcare system, in which citizens could freely choose the providers regardless of the type of ownership (private for profit, private non profit, or public) and where the prospective payment system, based on Diagnosis-Related Groups (DRGs), is applied to reimburse hospital discharges. But, in December 2015, Lombardy approved a regional Law that fundamentally changed the healthcare system, introducing various new governance tools and promoting public disclosure of performance data.

Outcomes

We report here outcomes for the three regions, Tuscany, Lombardy and Marche, for the period 2008- 2015: for one clinical indicator, namely the percentage of femur fractures operated within 2 days; and the basket of indicators that make up the LEA grid scores (the focus of the Marche region).

We have chosen the percentage of femur fractures operated within 2 days because there is strong evidence that this indicator of process is a good indicator of outcomes and success depends on excellent management and coordination. In Italy this indicator is computed annually at the national level by the National Outcome Evaluation Programme (NOEP). The trajectory of the three Regions is shown in figure 3 from 2006 to 2015.

Figure 3: Percentage of femur fractures operated within two days from 2008 to 2015 to go about here

For Tuscany the improvement process started in 2006, when the PES was introduced together with a set of management tools (Pinnarelli et al, 2012), and the percentage of femur fractures operated within two days was below 30%. In Figure 3 Tuscany shows data from 2008 to 2015. By 2008, the region had already improved to 45%, and 2015 it achieved 70%, by far the highest percentage of the three regions. Although Marche had the highest percentage in 2008, this improved only gradually from 52% in 2008 to 57% in 2015, was overtaken by Tuscany in 2011, and actually fell in 2012. Lombardy had by far the worst performance of the three regions in 2008 with only 36%, and also shows only

gradual improvement to 39% by 2011. In 2012, Lombardy introduced a pay for performance program without public disclosure and benchmarking, which had a small impact with the percentage increasing to 47% in 2014. From 2014 the data on performance were publicized and performance improved at the fastest rate in that year ending on 57% and at virtually the same level as Marche (but lower than Tuscany). And in 2015 Lombardy joined the IRPES.

Figure 4 shows performance for the three regions between 2007 and 2014 in the LEA Grid. This shows that Tuscany steadily improved its performance. It was the best performing Region in the whole of Italy in 2013, 2014 and 2015. In 2007, Lombardy had similar performance to Tuscany and Marche was worst. By 2014, Marche had improved more than Lombardy so both regions had similar performance.

Figure 4: Performance for the three regions between 2007 and 2014 in the LEA Grid to go about here

Discussion

In the final section of this paper we outline the conceptual implications of our empirical findings, for the effective governance of health care. The forms of governance we have discussed relate to the conventional theories of micro economics based on what Thaler and Sunstein (2009) describe as individuals acting as 'econs' as follows: T&A works neither in theory, as it creates perverse

incentives, nor in practice; C&C works in theory, but not in practice; and 'reputation' does not work in theory, as there are no pecuniary incentives, but works in practice. We suggest that a framework consistent with the evidence we have observed comes from key tenets of reciprocal altruism (Oliver, 2017; Gintis, et al, 2005; Wilson, 2015), which seems appropriate, as many who choose to work in health services do so from 'knightly' motives.

Hume observes that 'it appears somewhat strange' that his political maxim '*that every man must be supposed a knave* … should be true in *politics* which is false in *fact*' (Miller and Hume, 1994, p42-43, italics in original). We interpret this paradox as Hume emphasising that *systems* must be designed to counter those who seek to pursue private interests at the expense of the public good, because to allow such behaviour has corrosive consequences by undermining a fundamental element of reciprocal altruism, which imposes sanctions on such behaviour. Hence Hume's political maxim offers a sounder starting point than, and counsel against, the T&A model, which assumes that (to paraphrase Hume) 'every man ought to be supposed a knight and to have no other end, in all his actions, than the public interest'. We now explain why governance that relies on T&A is so ineffective using ideas of identity economics.

One striking examples of the power of identity economics in understanding how to improve public services given by Akerlof and Kranton (2010, p 782) is of a headteacher knowing that he had succeeded in reforming the 'shocking' state of Baldwin Elementary school (in a blighted area in New Haven, Connecticut), when he saw a student stop a fight with the words: 'We don't do that in this school'. In

2000, the Government saw the identity of the NHS in England as still trapped in a timewarp of the rationing of the 1940s: that patients ought still to be grateful for a 'free' service after long waiting times for treatment in inadequate buildings. The aim of the combination of generous funding and the regime of 'star ratings' was to change the meaning of identity for those who worked in the English NHS by tackling its perceived symbolic 'broken windows' (Wilson and Kelling, 1982), in the form of unacceptably long waiting times. Those who worked in organisations that failed to achieve the required transformation were subjected to their identities as valued public servants being undermined by 'naming and shaming' and their chief executives were threatened with the ultimate sanction of being denied the identity of being part of the NHS through being sacked. In contrast where providers understand that governance is by T&A then there are no norms defining identity as members of a group, as anything goes, and a corrosive kind of Gresham's law is at work, which tolerates and rewards 'knavish' behaviour.

We now consider the C&C model, which Adam Smith, famously proposed as an effective way of governing knavish behaviour driven by self interest. But, as Oliver (2017) rightly points out. Smith's homely example of its efficacy using the butcher, the baker and the brewer in *The Wealth of Nations* (Smith, 2005) was one in which problems of market failure are absent: these are 'local artisans producing relatively simple, easily understood goods with limited opportunities to exploit informational asymmetries and with a bond of trust'. As Oliver argues: 'Many of the goods and services delivered by public sectors are too complex to expect competitive markets to deliver them efficiently or justly'. As we have

shown for hospitals, there are elements of market failure on the demand side: patients do not act as consumers, even when good information is available on the quality of care and where choice can be exercised. There are also elements of market failure on the supply side because there are rightly serious barriers to entry and exit. The latter point was overlooked in Enthoven's influential argument for an 'internal market' to tackle what he saw in 1985 as the problem of the gridlocked English NHS (Enthoven, 1985). He cited Schultze (2010) in his contrast between the efficacy of markets and the paralysis of government from the rule of 'do no direct harm':

> We put few obstacles in the way of a market-generated shift of industry to the South or the substitution of synthetic fibers for New England woollens, events that thrust large losses on individuals, firms and communities. But we find it extraordinarily difficult to close a military base or a post office.

This passage ignores the vital distinction between consumption where the location of production is irrelevant to a consumer (clothes) and where location matters (e.g. a post office). This is crucial as this means that the quasi market is weakened on the supply side because it is difficult for Ministers, who are expected to ensure good access to local health services, to allow 'failing' hospitals to exit the market (Tuohy, 1999, pp. 192-95). Indeed, if this quasi market were to function on the demand side, then this would result in poorly-performing hospitals losing income, e.g. for elective surgery, which would be likely to create serious financial problems in continuing to provide services that must be

provided locally. So either the hospital is bailed out financially or the local population suffers or both. Chowdry et al (2008) identify the same problem of the lack of supply-side flexibility as undermining the efficacy of the quasi market for schools. Hence the attractions of governance using reputation effects to put pressure on providers to perform satisfactorily, without reducing funding, through publishing information to make them accountable to the local populations they serve and sacking those who run them ineffectively.

For Le Grand (2007), the attraction of quasi markets is that this market mechanism tackles 'knavish' tendencies so as to reinforce 'knightly' behaviour. But, as Bowles (2016) argues, that the problem with market mechanisms is that, rather than enhance 'knightly' motives, they can destroy them as shown in a series of carefully designed experiments. So the problem is that market mechanisms are designed to appeal to self-interested motives, but in health services, which are notoriously exemplars of market failures, these mechanisms are ineffective in harnessing the pursuit of self interest for the public good. So, as Oliver (2017) argues, these may foster egoistic self-interest in ways that crowd out a desire to reciprocate; and instead a key principle in designing incentives through reputation effects is that these ought to 'encourage, rather than undermine, the obligations that the relevant members of any group ought to feel – and naturally, for the most part, do feel – towards each other'.

Oliver (2017) argues that the 'motivating force of negative reciprocity' through 'naming and shaming' ought to be exclusively reserved for performance that 'is bad in an absolute sense because unwarranted fear may undermine identity with

the group'. Further, since there is no shame in adequate performance, it may be difficult for government to use 'naming and shaming' to incentivise improvement. This echoes concerns of Le Grand (2007) and Barber (2007) that target-based reforms of public services in England in the 2000s cannot improve performance from awful to adequate. So - over ten years ago – they argued for governance based on quasi markets instead. Whilst their diagnosis was correct, their proposed remedy has not proved effective.

Oliver (2017) identifies a second strand to using reputation to improve performance once it has become adequate in complex systems such as delivering health care where people working together in effective social groups is an essential prerequisite for high performance depends. Oliver argues that this reputation effect comes from people wanting 'to signal that they are good co-operators/reciprocators'. For this second strand to be effective, he emphasises that it needs to be carefully designed 'so as to avoid demotivating poor relative performers' by 'naming and shaming'; and for this to be at a scale where for group cooperation can work effectively. So, e.g., in England, central government can use a simple ranking system that 'names and shames' 'failing' organisations, but the effects of benchmark competition to generate incentives for high performance via public reporting needs to be organised to report across multiple criteria at a regional level, as in Tuscany.

So looking back over the past thirty years the evidence of the impacts of reforms to systems of governance on measured performance are as follows. We see 'knights' and 'knaves' to be key in making sense of these systems of governance.

The evidence for health care is that neither governance by T&A nor C&C in quasi markets has proved to be effective. Governance by reciprocal altruism is based on the strategy of 'tit-for-tat', which for Ayres and Braithwaite (1992) underpins their concept of 'responsive regulation'. This is argued to be a more effective means of regulation that sticking with either the deterrence model, which assumes all providers are 'knaves', or the compliance model, which assumes all providers are 'knights'. The effective regulator 'speaks softly & carries a big stick' so when providers found out to be 'knaves' are subjected to the deterrence model until they have proved that they can be trusted to be 'knights' and regulated in the compliant model. We see responsive regulation harnessing different kinds of reputation effects. The deterrent model of 'naming and shaming', which is to be exclusively restricted to tackling 'failing' organisations, where simple rankings can be applied within a hierarchical system. An exemplar is the 'star ratings' regime for the NHS in England. The compliant model uses benchmark competition is designed to create incentives for high performance in a regional structure. An exemplar is the Tuscan PES, which is carefully designed to rank performance on multiple criteria so no single organisation can be described as 'failing' or 'high performing' and hence this system creates peer group pressures to aspire to high performance on the various criteria within each organisation.

Acknowledgements

Acknowledgements: GB is grateful for discussions with Sir Michael Barber, Professor Sir Julian Le Grand and Dr Adam Oliver in analysis of models of governance. AE is extremely grateful to her Zambian participants, who invited her to observe their work, shared their reflections with her and provided extremely useful comments on earlier drafts. The Ministry of Health (at both central and district levels) was especially generous in facilitating this ethnographic research. SN would like to thank all the MeS Lab researchers and the regional administrators of the Italian Regional Collaborative and their staff for their valuable support. The usual disclaimer applies.

References

Akerlof, G.A., Kranton, R.E. (2010), *Identity Economics: How Identities Shape Our Work, Wages, and Well Being*. Woodstock: Princeton University Press.

Anonymous (2017) The National Health Service. Policy Transplant. *Economis*t, 4 November, 56-57

Ashton, T., Mays, N., & Devlin, N. (2005), 'Continuity through change: the rhetoric and reality of health reform in New Zealand', *Social Science & Medicine*, 61(2): 253-262.

Auditor General for Wales (2005), *NHS waiting times in Wales. Volume 1 – The Scale of the problem*. *Volume 2 - Tackling the problem*. Cardiff: The Stationery Office.

Ayres, I., & Braithwaite, J. (1992), *Responsive regulation: Transcending the deregulation debate*. Oxford: Oxford University Press.

Barber M. (2007), *Instruction to deliver*. London: Politico's Publ. p335.

BBC News. (2007), Blair: In his own words. Friday, 11 May 2007 http://news.bbc.co.uk/2/hi/uk_news/politics/3750847.stm (accessed 26 January 2018).

Berwick, D. M., James, B., & Coye, M. J. (2003), 'Connections between quality measurement and improvement', *Medical care*, 41(1):I-30.

Bevan G. (2014), *The impacts of asymmetric devolution on health care in the four countries of the UK*. https://www.researchgate.net/profile/Gwyn_Bevan/publication/267864997_The_impacts_of_asymmetric_devolution_on_health_care_in_the_four_countries_of_the_UK_Research_report/links/545b841d0cf249070a7a714f.pdf

Bevan, G., Fasolo, B. (2013), 'Models of Governance of Public Services: Empirical and Behavioural Analysis of Econs and Humans'. In: A. Oliver (Ed.) *Behavioural Public Policy.* Cambridge: Cambridge University Press: 38-62.

Bevan, G., & Hamblin, R. (2009), 'Hitting and missing targets by ambulance services for emergency calls: effects of different systems of performance

measurement within the UK', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1): 161-190.

Bevan, G., & Hood, C. (2006), 'What's measured is what matters: targets and gaming in the English public health care system', *Public administration*, 84(3): 517-538.

Bevan G, Karanikolos M, Exley J, Nolte E, Connolly S, Mays N. (2014), *The four health systems of the United Kingdom: how do they compare?*. London: Nuffield Trust.

Bevan G., Skellern M. (2011), 'Does competition between hospitals improve clinical quality?: a review of evidence from two eras of competition in the English NHS'. *BMJ* 343, d6470-d6470.

Bowles S. (2016), *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens*. London: Yale University Press.

Central Statistical Office (CSO), Ministry of Health, University of Zambia Teaching Hospital, University of Zambia Department of Population Studies, Tropical Diseases Research Centre, (2015) *The DHS Program. Zambia Demographic and Health Survey 2013-14*. Lusaka: CSO.

Chassin, M. R. (2002), 'Achieving and sustaining improved quality: lessons from New York State and cardiac surgery', *Health Affairs*, 21(4): 40-51.

Chowdry, H., Muriel, A., & Sibieta, L. (2008), *Level playing field? The implications of school funding. CfBT Education Trust.* <https://www.researchgate.net/profile/Haroon_Chowdry/publication/39065923_Level_Playing_Field_The_Implications_of_School_Funding/links/5437d69d0cf2d5fa292b6dc2.pdf > (accessed 26 January 2018).

Cooper, Z., Gibbons, S., Jones, S., & McGuire, A. (2011), 'Does hospital competition save lives? Evidence from the English NHS patient choice reforms'. *The Economic Journal*, *121*(554), F228-F260.

Department of Health (2002), *NHS Performance Ratings. Acute Trusts, Specialist Trusts, Ambulance Trusts, Mental Health Trusts 2001/02*. London: Department of Health.

Department of Health. (2005), *Chief Executive's Report to the NHS. Statistical Supplement -May 2005*. London: Department of Health. p 16

Enthoven, A. C. (1985), *Reflections on the management of the National Health Service: an American looks at incentives to efficiency in health services management in the UK.* London: Nuffield Provincial Hospitals Trust.

Evans, A. (2017), 'Amplifying Accountability by Benchmarking Results at District and National Levels', *Development Policy Review*.

France, G., Taroni F. (2005), 'The evolution of health-policy making in Italy, *Journal of Health Politics', Policy and Law*, 30(1–2): 169–87

Frey B. (2013), How should people be rewarded for their work? In: A. Oliver, (Ed.) *Behavioural Public Policy.* Cambridge: Cambridge University Press: 165-183.

Fukuda-Parr, S. (2014), 'Global Goals as a Policy Tool: Intended and Unintended Consequences', J*ournal of Human Development and Capabilities,* 15(2-3): 165-183.

Fung, C. H., Lim, Y. W., Mattke, S., Damberg, C., & Shekelle, P. G. (2008), 'Systematic review: the evidence that publishing patient care performance data improves quality of care', *Annals of internal medicine*, 148(2): 111-123.

Gaynor, M., Moreno-Serra, R., & Propper, C. (2013), 'Death by market power: reform, competition, and patient outcomes in the National Health Service', *American Economic Journal: Economic Policy*, *5*(4), 134-166.

Gintis, H., Bowles, S., Boyd, R., and Fehr, E. (eds.) (2005), *Moral Sentiments and Material Interest: The Foundation of Cooperation in Economic Life.* Cambridge MA: MIT Press.

Ham, C. (2008), 'World class commissioning: a health policy chimera?', *Journal of Health Services Research & Policy*, 13(2): 116-121.

Hibbard, J.H., Stockard, J., Tusler, M. (2003), 'Does publicizing hospital performance stimulate quality improvement efforts?', *Health Affairs,* 2(2): 84-94.

Hibbard, J.H., Stockard, J., Tusler, M. (2005), 'Hospital performance reports: impact on quality, market share, and reputation', *Health Affairs,* 24(4):1150-60.

Hibbard J.H. (2008), 'What can we say about the impact of public reporting? Inconsistent execution yields variable results', *Annals of Internal Medicine*, 148(2):160-1.

Hood, C., Dixon, R. (2010), 'The political payoff from performance target systems: no-brainer or no-gainer?', J*ournal of Public Administration Research and Theory*. 1;20(suppl 2): i281-98.

Miller, E. F., & Hume, D. (1994), *Essays: Moral, Political, and Literary*. Indianapolis: Liberty Fund.

Le Grand, J. (2003), *Motivation, agency and public policy: Of knights and knaves, pawns and queens*. Oxford: Oxford University Press.

Le Grand, J. (2007), *The Other Invisible Hand: Delivering Public Services through Choice and Competition*. Woodstock: Princeton University Press.

Le Grand, J. (2010), 'Knights and Knaves Return: Public Service Motivation and the Delivery of Public Services', *International Public Management Journal*,13(1): 56-71. (p. 63)

Maarse, H., Jeurissen, P., & Ruwaard, D. (2016), 'Results of the market-oriented reform in the Netherlands: a review', *Health Economics, Policy and Law.* 11(2): 161-78.

Marshall, M. N., Shekelle, P. G., Leatherman, S., & Brook, R. H. (2000), 'The public release of performance data: what do we expect to gain? A review of the evidence', *JAMA*. 283(14): 1866-74

Ministry of Health. (2008), *National Reproductive Health Policy*. Lusaka: MoH,.

Mukonka V. (2012), *Status of maternal health in Zambia*. Geneva: Countdown to 2015.

Mukonka, V. M., Malumo, S., Kalesha, P., Nambao, M., Mwale, R., Kasonde, M. Wamulume, P. K. (2014), 'Holding a country countdown to 2015 conference on Millennium Development Goals (MDGs) – the Zambian experience'. *BMC Public Health* 14, 60.

Murante, A.M., Vainieri, M., Rojas, D. and Nuti, S. (2014), 'Does feedback influence patient - professional communication? Empirical evidence from Italy', *Health Policy,* 116(2):273-80.

Nuti, S., Vola, F., Bonini, A., & Vainieri, M. (2015), 'Making governance work in the health care sector: evidence from a 'natural experiment' in Italy', *Health Economics, Policy and Law*, 11(1): 17-38.

Nuti, S., Seghieri, C., & Vainieri, M. (2013), 'Assessing the effectiveness of a performance evaluation system in the public health care sector: some novel evidence from the Tuscany region experience', *Journal of Management & Governance*, 17(1), 59-69.

Nuti, S., Bini, B., Ruggieri, T.G., Piaggesi, A. and Ricci, L. (2016), 'Bridging the gap between theory and practice in Integrated Care: the case of the Diabetic Foot Pathway in Tuscany', *International Journal of Integrated Care*, *16*(2): 9.

Oliver, A. (2007), 'The Veterans Health Administration: An American Success Story?', *Milbank Quarterly,* 85: 5-35.

Oliver A. (2015.), 'Incentivising improvements in health care', *Health Economics, Policy and Law*, 10(03): 327-343.

Oliver, A. (2107), 'Do Unto Others: On the Importance of Reciprocity in Public Administration', *American Review of Public Administration*: DOI 10.1177/0275074016686826.

Pinnarelli, L., Nuti, S., Sorge, C., Davoli, M., Fusco, D., Agabiti, N., Vainieri, M. and Perucci, C.A. (2012), 'What drives hospital performance? The impact of comparative outcome evaluation of patients admitted for hip fracture in two Italian regions', *BMJ Quality & Safety*, 21(2): 127-134.

Pollock, A., Macfarlane, A., Kirkwood, G., Majeed, F.A., Greener, I., Morelli, C., Boyle, S., Mellett, H., Godden, S., Price, D. and Brhlikova, P. (2011), 'No evidence that patient choice in the NHS saves lives', *The Lancet*, *378*(9809): 2057-2060.

Sarwar, M. B. (2015), *National MDG implementation: lessons for the SDG era.* Working Paper 428. London: Overseas Development Institute: p.10

Schultze, C. L. (2010), *The public use of private interest*. Vol. 1976. Brookings Institution Press. Quoted by Enthoven

Skellern, M. (2016), *The hospital as a multi-product firm: Measuring the effect of hospital competition on quality using Patient-Reported Outcome Measures*. Mimeo, London School of Economics and Political Science, June, 2015.

Smee C (2008), 'Speaking truth to power', in Timmins N ed (2008) *Rejuvenate or Retire? Views of the NHS at 60*. Nuffield Trust.

Smith, A. (2005), *Wealth of Nations*. Chicago: University of Chicago Press.

Smith, J., Curry, N. (2011), 'Commissioning'. In Dixon A., Mays N. and Jones L. (Eds) *Understanding New Labour's market reforms of the English NHS*. London: King's Fund: 30-51. http://www.kingsfund.org.uk/sites/files/kf/chapter-3-commissioning-new-labours-market-reforms-sept11.pdf

Spurgeon, P., Mazelan, P. M., & Barwell, F. (2011), 'Medical engagement: a crucial underpinning to organizational performance', *Health Services Management Research*, 24(3): 114-120.

Clark, J. (2012), 'Medical leadership and engagement: no longer an optional extra', *Journal of Health Organization and Management*, 26(4): 437–443.

Stevens, S. (2004), 'Reform strategies for the English NHS', *Health Affairs*. 23(3): 37-44.

Thaler, R. H., Sunstein, C. (2009), *Nudge: Improving decisions about health, wealth, and loss aversion*. London: Penguin

Tankard, M. & Paluck, M. E. (2016), 'Norm Perception as a Vehicle for Social Change', *Social Issues and Policy Review* 10(1): 196

Tuohy, C. H. (1999), *Accidental logics: The dynamics of change in the health care arena in the United States, Britain, and Canada*. Oxford: Oxford University Press.

Wenger, E., McDermott, R. A., & Snyder, W. (2002), *Cultivating communities of practice: A guide to managing knowledge*. Boston: Harvard Business Press.

Wilson, D. S. (2015), *Does altruism exist?: culture, genes, and the welfare of others*. London: Yale University Press.

Wilson, J. Q., & Kelling, G. L. (1982), 'Broken windows'. In Dunham RG and Alpert GP (eds). *Critical issues in policing: Contemporary readings*, Long Grove, Illinois : Waveland Press, Inc.: 395-407.

World Health Organization (WHO), UNICEF, UNFPA, World Bank Group and the United Nations Population Division (2015), *Trends in maternal mortality: 1990 to 2015: estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division*. Geneva: WHO. p. 77.

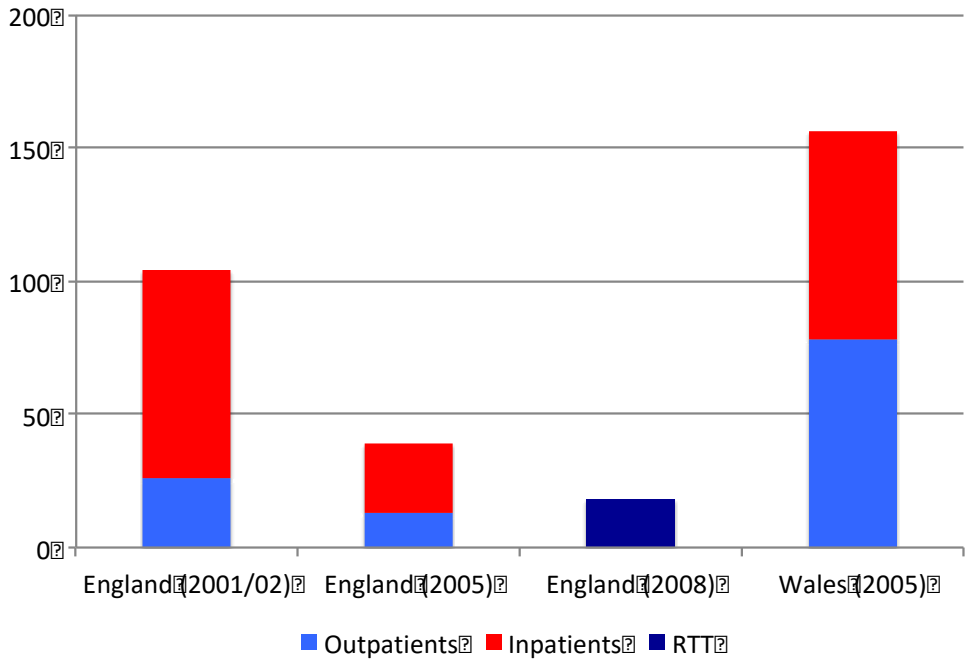

Figure 1: Hospital waiting time targets for England and Wales

Sources: Auditor General for Wales (2005) and Department of Health (2002)

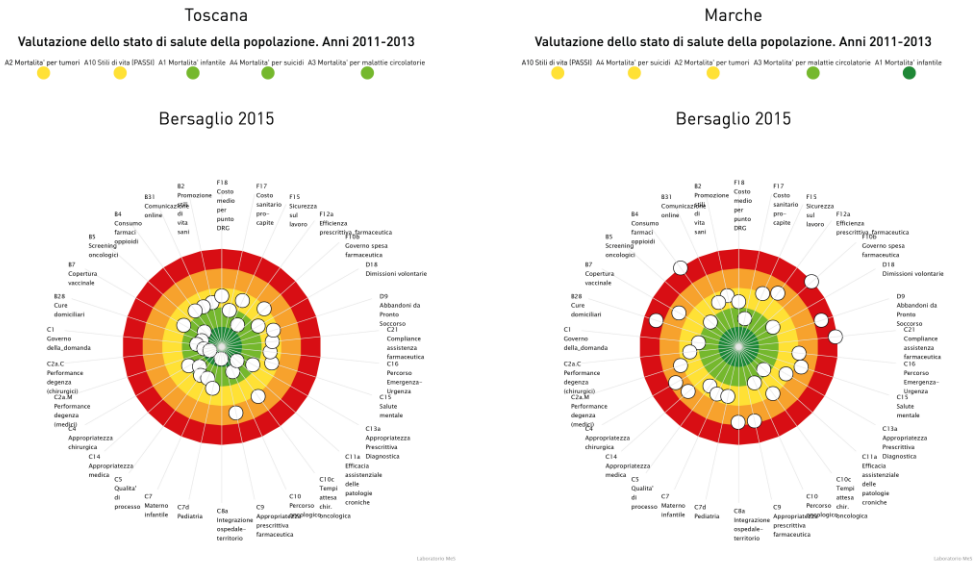

Figure 2: The dartboards for Marche and Tuscany in 2015

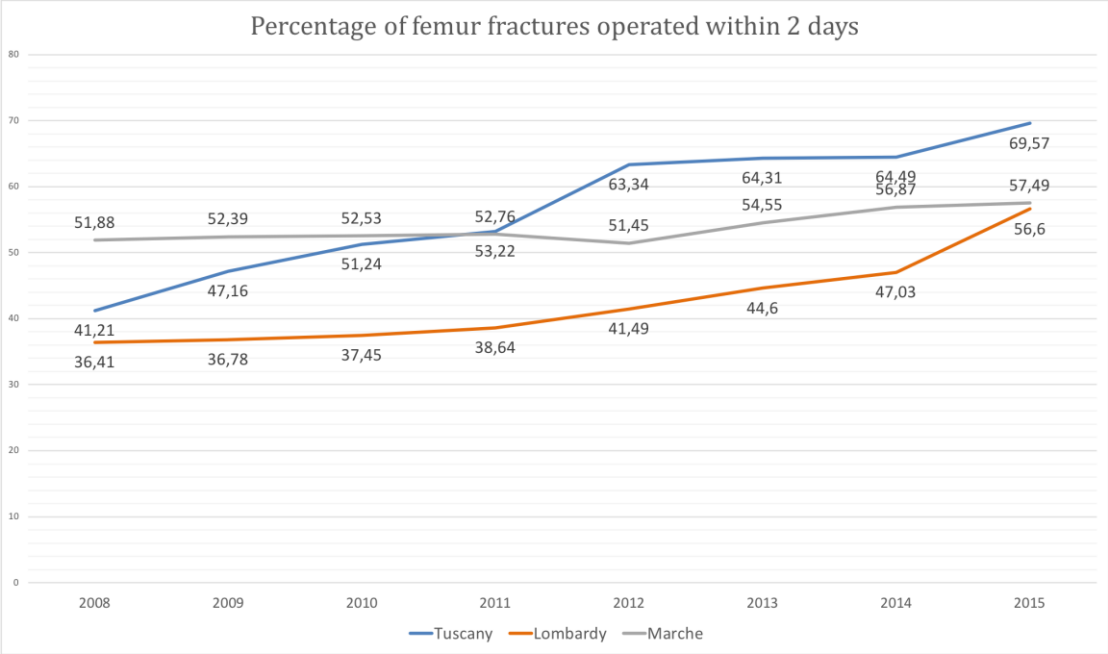

Figure 3: Percentage of femur fractures operated within two days from 2008 to 2015

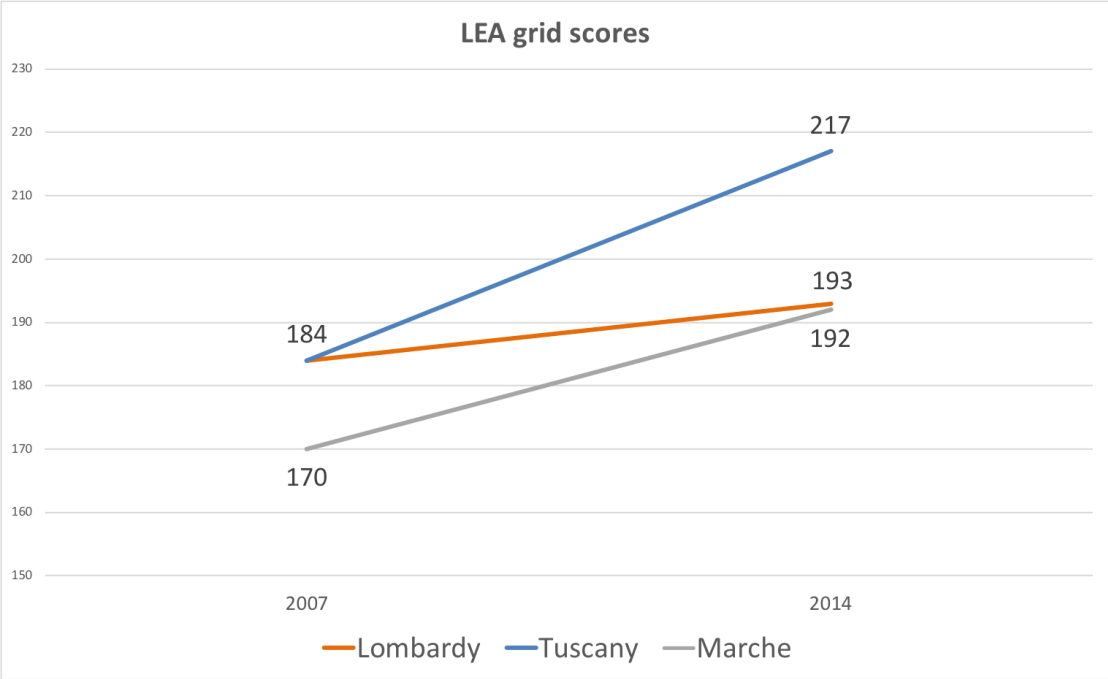

Figure 4: Performance for the three regions between 2007 and 2014 in the LEA Grid