BMJ open

Relationship between quality of care and choice of clinical computing system: retrospective analysis of family practice performance under the UK's quality and outcomes framework

Evangelos Kontopantelis,[1,2] Iain Buchan,[3] David Reeves,[1,2] Kath Checkland,[1] Tim Doran[4]

For numbered affiliations see end of article.

Correspondence to
Dr Evangelos Kontopantelis;
e.kontopantelis@manchester.ac.uk

ABSTRACT

Objectives: To investigate the relationship between performance on the UK Quality and Outcomes Framework pay-for-performance scheme and choice of clinical computer system.

Design: Retrospective longitudinal study.

Setting: Data for 2007–2008 to 2010–2011, extracted from the clinical computer systems of general practices in England.

Participants: All English practices participating in the pay-for-performance scheme: average 8257 each year, covering over 99% of the English population registered with a general practice.

Main outcome measures: Levels of achievement on 62 quality-of-care indicators, measured as: reported achievement (levels of care after excluding inappropriate patients); population achievement (levels of care for all patients with the relevant condition) and percentage of available quality points attained. Multilevel mixed effects multiple linear regression models were used to identify population, practice and clinical computing system predictors of achievement.

Results: Seven clinical computer systems were consistently active in the study period, collectively holding approximately 99% of the market share. Of all population and practice characteristics assessed, choice of clinical computing system was the strongest predictor of performance across all three outcome measures. Differences between systems were greatest for intermediate outcomes indicators (eg, control of cholesterol levels).

Conclusions: Under the UK's pay-for-performance scheme, differences in practice performance were associated with the choice of clinical computing system. This raises the question of whether particular system characteristics facilitate higher quality of care, better data recording or both. Inconsistencies across systems need to be understood and addressed, and researchers need to be cautious when generalising findings from samples of providers using a single computing system.

ARTICLE SUMMARY

Article focus

- Practice and patient-level characteristics are known predictors of quality of care, as measured by the Quality and Outcomes Framework (QOF) indicators.
- Various general practitioner (GP) clinical computer systems are used in the UK but their distribution over time and location is unknown.
- GP clinical computer systems differ in software architecture, user interface and clinical coding lists but their effect on quality of care has never been examined.

Key messages

- Seven systems were found to hold 99% of the market share, with clear geographical variation in their distribution.
- Levels of performance on the QOF differed to a small extent across clinical computer systems, even after controlling for practice and patient characteristics. Quantified differences were small but not negligible since they translate to systematic variation in recorded care for hundreds of thousands of patients nationwide.
- Researchers that utilise primary care databases, which collect data from a single clinical system, need to be cautious when generalising their findings to all English practices.

Strengths and limitations of this study

- This is the first study that investigates the effect of GP clinical computer system choice on measured quality of care.
- We used data for over 99% of all English practices and there is no risk of inductive fallacy.
- There are more aspects to quality of care than what is recorded under the QOF; this is an observational study and causality is difficult to establish and it is possible that QOF-oriented practices have particular clinical system preferences.

INTRODUCTION

Clinical computing systems have been promoted as a means to improve the quality of healthcare, holding advantages over paper-based systems of improved data recording, integration and accessibility. These advantages, combined with automated feedback and alerts, have the potential to drive improvements in efficiency, process performance, clinical decision-making and medication safety.[1] [2] The potential is greatest in primary care settings, where the activities of different providers for multiple conditions must be coordinated in order to optimise care for patients and minimise wastage of resources. In practice the results of implementing information technology systems have been mixed,[3–9] and the full promise of computer-supported healthcare has yet to be realised. This is partly attributable to the variable quality of software systems developed by multiple providers, and partly to clinicians adapting at different rates to systems that frequently do not address their specific requirements and which may challenge their approach to existing practice.

In UK primary care, health information technology developed from multiple, often amateur, systems in the 1980s into a handful of systems using a standard coding thesaurus, known as Read codes, by the 2000s.[10] From 1998, family practices were partially subsidised for the costs of installing clinical computing systems as part of a wider government programme to develop electronic patient records.[11] Full subsidies were provided from 2003 in preparation for the implementation of a national pay-for-performance scheme—the Quality and Outcomes Framework (QOF)—the following year. The QOF provides large financial bonuses to practices based on their achievement on over 100 quality of care indicators, mostly relating to processes of care for common chronic conditions.[12] Under the QOF, practices are awarded points for each quality indicator based on the proportion of patients for whom targets are achieved, between a lower achievement threshold of 40% and an upper threshold that varies by indicator from 50% to 90%. Each point scored earns the practice £126, adjusted for the relative prevalence of the disease and the size of the practice population. Practices can exclude ('exception report') inappropriate patients from achievement calculations for logistical reasons (eg, recent registration with the practice), clinical reasons (eg, a contraindication to treatment) or for patient-informed dissent. Performance data are drawn automatically from practices' clinical computing systems—which must conform to standard interoperability requirements—and collated on a national database: the Quality Management and Analysis System (QMAS). This system provides feedback to practices and is also used to calculate quality payments.

The QOF has had a substantial impact on the use of clinical computing systems by practices, and this in turn impacts on relationships with patients both within and beyond consultations.[13] Practices are required to keep disease registers, and because bonus payments increase with disease prevalence there is an incentive to case-find.

The business rules for the QOF also specify criteria and permissible Read codes for identifying patients with particular conditions, resulting in greater uniformity of code usage and in some cases changing diagnostic behaviour (eg, making clinicians more reluctant to record depression[14]). Software providers have adapted their systems to facilitate better practice performance on the QOF (usually defined in terms of practice remuneration rather than high achievement rates per se) by incorporating pop-up alerts and management tools in the software and providing QOF-oriented training programmes for practice staff. There is no independent evidence to date, however, on whether practice performance on the QOF and recorded quality of care is associated with the practice's choice of clinical computing system. We used a unique dataset to assess these relationships in English family practices operating under the national pay-for-performance scheme.

METHODS

We carried out a retrospective study of performance on the QOF by English family practices from 2007–2008 to 2010–20011, identifying practice predictors including choice of clinical computing system (only one in use within each practice) through multilevel multiple linear regression models. English practices are geographically organised in 151 Primary Care Trusts (PCTs: commissioning bodies that oversee practices operating in a locality) and those in turn into 10 Strategic Health Authorities (SHAs: responsible for fiscal policy at a regional level).

Data sources

We used data from the QMAS, the national information system supporting the QOF, which holds data for almost all English practices (over 99% of registered patients in England attend participating practices). QMAS data on QOF achievement, exception reporting and prevalence rates are freely available on the Health and Social Care Information Centre (HSCIC) website[15] but information on clinical computing systems is not publicly reported and we obtained the relevant dataset for this study from the HSCIC. Data on practice characteristics and the populations they serve were obtained from the General Medical Services (GMS) Statistics database, also provided by the HSCIC. Area deprivation, as measured by the Index of Multiple Deprivation (IMD),[16] and urban/non-urban classification,[17] at the lower super output area level (a set of geographical areas of roughly consistent size with a population of around 1500), were obtained from the Communities and Local Government[18] and the Office of National Statistics[19] websites, respectively. Lookup tables from the UKBORDERS website[20] were used to assign the measures to practices at the postcode level. Data were complete for all practices.

Clinical computing systems

Over the study period, 15 clinical computing systems from eight suppliers were active within the QOF scheme.

Practices were assigned to a group on the basis of the clinical computing system in use at each year end (March; see table 1). Computing systems with fewer than 100 users at any time point were excluded from the analyses. Seven systems from five software suppliers were thus included in the analyses, accounting for around 99% of the practices participating in the scheme.

Outcomes

We measured practice performance on the clinical quality indicators of the QOF, using non-register clinical indicators that were continually incentivised over the study period (see online supplementary table A6). Three performance measures were used: (1) reported achievement (RA), the proportion of eligible patients for whom the targets were achieved, not including exception reported patients; (2) the percentage of QOF points scored (PQ), the metric on which remuneration is based and (3) population achievement (PA), measured by the proportion of eligible patients for whom targets were actually achieved, including exception reported patients. We argue that RA and PA are proxies of quality of care and percentage of points scored is a measure of practice benefit from the scheme, although all three outcomes are very strongly correlated.

RA rates

For practice and indicator i, RA can be defined as

$$RA_{ij} = \frac{N_{ij}}{D_{ij}} \qquad (1)$$

where N_{ij} is the number of patients for whom the target was achieved and D_{ij} the number of patients meeting the criteria for the indicator who have not been an exception reported by the practice. The exception reporting provision is intended to protect patients from inappropriate care and it can be used to exclude patients for a variety of reasons, including logistical, clinical and patient dissent.[21] The RA rate is the most commonly used measure of practice performance under the QOF scheme, as in theory it focuses on patients for whom the quality targets are appropriate.

Percentage of QOF points scored

For practice j and indicator i, the number of points scored (QP_{ij}) is based on RA rates

$$
\begin{aligned}
QP_{ij} = MP_i \\
\times \max\left(\frac{\min\left[(RA_{ij} - LT_i), (UT_i - LT_i)\right]}{(UT_i - LT_i)}, 0\right)
\end{aligned} \qquad (2)
$$

where, for each indicator i, MP_i is the number of available points (ranging from 1 to 57), LT_i the lower threshold (set at 25% for two indicators and 40% for all others) and UT_i the upper threshold (ranging from 50% to 90%). For a given indicator, a practice will

secure 0 points if RA is less than or equal to the lower achievement threshold and will secure maximum points if RA equals or exceeds the upper achievement threshold. For RA rates between the thresholds, the number of points scored is calculated linearly using (2).[22] The percentage of points scored (PQ) is obtained by dividing the overall points scored by the relevant number of available points.

PA rates

The existence of upper achievement thresholds and the provision for practices to exception report patients means that maximum points can be scored (and maximum remuneration secured) without achieving the targets for all patients. Practices with markedly different achievement rates can therefore appear similar when assessed using the percentage of points scored. We therefore calculated the PA rate[23] as:

$$PA_{ij} = \frac{N_{ij}}{D_{ij} + E_{ij}} \qquad (3)$$

where E_{ij} is the number of patients' exception reported.

Across all three outcome measures, practice composite scores were calculated: (1) overall, across all 62 clinical indicators and (2) by three categories of activity: measurement activities, 35 indicators; treatment activities, 11 indicators and intermediate outcomes, 16 indicators (see online supplementary table A6). For reported and PA, scores were calculated by summing N_{ij} across all relevant indicators and dividing by the sum of eligible patients (D_{ij} for RA and $D_{ij} + E_{ij}$ for PA). Patients eligible for multiple indicators are double counted in the composite rates, and these rates therefore represent the proportion of 'opportunities' for a practice to perform an incentivised activity that resulted in a success. For percentage of achieved QOF points, we used the sum of QP_{ij} across the relevant indicators, overall and by indicator group, divided by the respective number of available points (see online supplementary appendix for details of indicators in each category).

Statistical modelling

We used multilevel mixed effects multiple linear regression models to identify population, practice and clinical computing system predictors of RA and PA rates and percentage of QOF points scored. For each outcome type, two regression models were executed, with a different summary measure used as outcome in each. In the simpler model (model 1), we used the overall summary measure across all 62 indicators, while the more complex model (model 2) included summary measures by indicator group (measurement, treatment and intermediate outcome), with their effect modelled as fixed. Using model 1 we were able to estimate the effect of each clinical system on overall achievement and percentage of points scored, and to assess whether there were

Table 1 Practice counts and percentages by system and year, sorted by popularity in 2010–2011

Provider	Product	Year			
		2007–2008	2008–2009	2009–2010	2010–2011
Included in analyses					
EMIS	LV	3811 (46.0%)	3661 (44.5%)	3498 (42.2%)	3284 (39.9%)
In Practice Systems	Vision 3	1572 (19.0%)	1599 (19.5%)	1564 (18.9%)	1492 (18.1%)
TPP	ProdSysOneX	697 (8.4%)	851 (10.4%)	1164 (14.0%)	1466 (17.8%)
EMIS	PCS	1103 (13.3%)	1160 (14.1%)	1259 (15.2%)	1216 (14.8%)
iSoft	Synergy	541 (6.5%)	487 (5.9%)	444 (5.4%)	401 (4.9%)
Microtest	Practice Manager	166 (2.0%)	164 (2.0%)	156 (1.9%)	156 (1.9%)
iSoft	Premiere	248 (3.0%)	212 (2.6%)	169 (2.0%)	132 (1.6%)
Subtotal (percentage of all practices)		8138 (98.2%)	8134 (99.0%)	8254 (99.6%)	8147 (98.9%)
Excluded from analyses					
EMIS	EMISWeb	0 (0.0%)	0 (0.0%)	0 (0.0%)	78 (0.9%)
Healthy Software	HealthyV5	0 (0.0%)	29 (0.4%)	24 (0.3%)	13 (0.2%)
Seetec	Seetec GP Enterprise	51 (0.6%)	24 (0.3%)	1 (0.01%)	1 (0.01%)
iSoft	Ganymede	8 (0.1%)	6 (0.1%)	2 (0.02%)	1 (0.01%)
Healthy Software	Crosscare	32 (0.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Protechnic Exeter Ltd	Exeter GP System	29 (0.4%)	17 (0.2%)	4 (0.05%)	0 (0.0%)
EMIS	GV	17 (0.2%)	7 (0.1%)	0 (0.0%)	0 (0.0%)
iSoft	System 6000	8 (0.1%)	3 (0.04%)	0 (0.0%)	0 (0.0%)
Subtotal (percentage of all practices)		145 (1.8%)	86 (1.0%)	31 (0.4%)	93 (1.1%)
Total		8283	8220	8285	8240

significant differences across clinical systems by using omnibus postregression χ^2 tests. In model 2, we included interaction terms (clinical system by indicator group) to estimate the effect of each clinical system in each indicator group and assessed whether there were significant differences across clinical systems, in each of the indicator groups, through postregression tests. All analyses were controlled for year, practice list size, local area deprivation, rurality, type of general practitioner (GP) contract (GMS, Personal Medical Services, Alternative Provider Medical Services or Primary Care Trust Medical Services), percentage of female patients, percentage of patients aged 65 or over, mean GP age in practice, percentage of female GPs, percentage of UK-qualified GPs and percentage of GP providers (partners, single-handed or shareholders). Practices that switched to a different system were included in both models but we excluded a few practices with fewer than 1000 patients since they were unrepresentative and would contribute data only partially (approximately 97 practices, with 0.001% of the patients).

Linear predictions, and their 95% CIs, were calculated for each clinical system from the regression models, overall (model 1) and by indicator group (model 2). These can be described as mean predicted outcome levels for each factor (clinical system and indicator group by clinical system) that are controlled for all model covariates and set to their mean values in the models, thus allowing us a 'fairer' comparison of the systems (table 2). We also performed pairwise comparisons between clinical system performance predictions, for all outcomes, with CIs adjusted using Scheffé's method for multiple comparisons.[24] The results from

these comparisons are presented in the online appendix (see online supplementary tables A3–5).

The structure of the data was three-level, with practice outcomes nested within PCTs and PCTs nested within SHAs. To account for this structure and to model variability at each level, we used mixed effects models with the xtmixed command in Stata. Owing to computational limitations, we modelled both levels as random effects. A potential complicating factor was the distributions of the outcomes, which were extremely skew-normal in some cases (eg, percentage of points achieved). Although linear regression models are robust against deviations from normality[25] we obtained bootstrap estimates of 1000 repetitions for the SE across all models, an approach that does not make any distributional assumptions about the observed data.[26] Differences in the SEs between bootstrapped and standard regression models were small and we only report the former. All statistical comparisons were made at an α level of 5%. Stata V.12.1 software was used for all analyses.[27]

RESULTS

Overall average PCT performance on RA, PA and percentage of QOF points scored are mapped in figure 1. Performance at the PCT level by type of indicator is given in the online appendix (see online supplementary figures A1–3).

Over the study period (2007–2008 to 2010–2011) 15 clinical computing systems from eight providers were in use in English family practices, but seven systems collectively accounted for approximately 99% of the market (table 1). LV by EMIS was the most widely used system,

Table 2 Regression model predictions for clinical systems, ranked on overall performance

	Overall		Measurement		Treatment		Outcome	
	Predictions (95% CI)	Rank	Predictions (95% CI)	Rank	Predictions (95% CI)	Rank	Predictions (95% CI)	Rank
Reported achievement								
Vision 3	*90.12 (90.03 to 90.21)*	*1*	*93.48 (93.40 to 93.57)*	*1*	*90.63 (90.53 to 90.74)*	*1*	81.03 (80.91 **to** 81.15)	2
Practice Manager	89.82 (89.52 to 90.13)	2	93.14 (92.87 to 93.41)	4	90.22 (89.91 to 90.52)	4	*81.36 (80.99 to 81.74)*	*1*
Synergy	89.82 (89.70 to 89.94)	3	93.40 (93.29 to 93.52)	2	90.19 (90.06 to 90.33)	5	80.41 (80.23 to 80.59)	4
Premiere	89.79 (89.54 to 90.04)	4	93.26 (93.03 to 93.48)	3	89.82 (89.54 to 90.09)	6	80.73 (80.37 to 81.08)	3
ProdSysOneX	89.57 (89.44 to 89.69)	5	92.91 (92.81 to 93.02)	6	90.43 (90.31 to 90.56)	2	80.31 (80.15 to 80.46)	5
LV	89.40 (89.35 to 89.46)	6	92.97 (92.92 to 93.02)	5	90.40 (90.34 to 90.46)	3	79.53 (79.46 to 79.60)	6
PCS	88.68 (88.58 to 88.78)	7	92.32 (92.21 to 92.43)	7	89.42 (89.30 to 89.54)	7	78.67 (78.52 to 78.82)	7
Population achievement								
Synergy	*85.57 (85.43 to 85.72)*	*1*	*90.87 (90.74 to 91.01)*	*1*	80.88 (80.72 to 81.03)	3	*75.42 (75.21 to 75.63)*	*1*
Vision 3	85.39 (85.29 to 85.49)	2	90.62 (90.52 to 90.72)	2	80.92 (80.82 to 81.03)	2	75.28 (75.15 to 75.41)	2
Premiere	85.07 (84.81 to 85.33)	3	90.39 (90.15 to 90.64)	3	80.14 (79.86 to 80.42)	6	74.92 (74.57 to 75.28)	4
LV	85.04 (84.98 to 85.10)	4	90.36 (90.30 to 90.42)	4	*80.96 (80.90 to 81.03)*	*1*	74.31 (74.23 to 74.39)	5
Practice Manager	84.71 (84.38 to 85.03)	5	90.04 (89.74 to 90.34)	5	80.01 (79.69 to 80.32)	7	75.04 (74.64 to 75.44)	3
ProdSysOneX	84.52 (84.40 to 84.64)	6	89.70 (89.58 to 89.82)	6	80.45 (80.33 to 80.57)	4	74.05 (73.88 to 74.22)	6
PCS	84.28 (84.17 to 84.39)	7	89.56 (89.45 to 89.67)	7	80.41 (80.28 to 80.53)	5	73.34 (73.19 to 73.50)	7
Percentage of points scored *†‡§								
Synergy	*98.08 (97.93 to 98.23)*	*1*	*98.00 (97.70 to 98.31)*	*1*	98.440 (98.14 to 98.74)	2	98.04 (97.74 to 98.35)	4
Premiere	98.07 (97.75 to 98.38)	2	97.75 (97.35 to 98.15)	2	*98.441 (98.04 to 98.84)*	*1*	98.27 (97.87 to 98.67)	2
Vision 3	97.81 (97.69 to 97.92)	3	97.27 (97.03 to 97.52)	4	98.27 (98.02 to 98.52)	4	*98.32 (98.07 to 98.57)*	*1*
LV	97.70 (97.63 to 97.77)	4	97.36 (97.14 to 97.59)	3	98.39 (98.16 to 98.62)	3	97.76 (97.53 to 97.98)	6
ProdSysOneX	97.59 (97.45 to 97.73)	5	97.22 (96.96 to 97.49)	5	98.07 (97.81 to 98.34)	5	97.80 (97.54 to 98.07)	5
Practice manager	97.36 (96.98 to 97.75)	6	96.83 (96.39 to 97.28)	7	97.50 (97.05 to 97.95)	7	98.21 (97.77 to 98.66)	3
PCS	97.30 (97.17 to 97.44)	7	96.88 (96.62 to 97.13)	6	98.01 (97.75 to 98.26)	6	97.39 (97.14 to 97.65)	7

*Overall, across the 62 indicators, 591 points were available in 2007–2008, 576 in 2008/2009 and 567 in 2009/2010–2010/2011 (model 1).
†Across the 35 measurement indicators, 284 points were available in 2007–2008, 279 in 2008/2009 and 274 from 2009–2010 to 2010–2011 (model 2).
‡Across the 16 treatment indicators, 135 points were available in 2007–2008, 125 in 2008–2009 and 124 from 2009–2010 to 2010–2011 (model 2).
§Across the 11 outcome indicators, 172 points were available from 2007–2008 to 2008–2009 and 169 from 2009–2010 to 2010–11 (model 2).
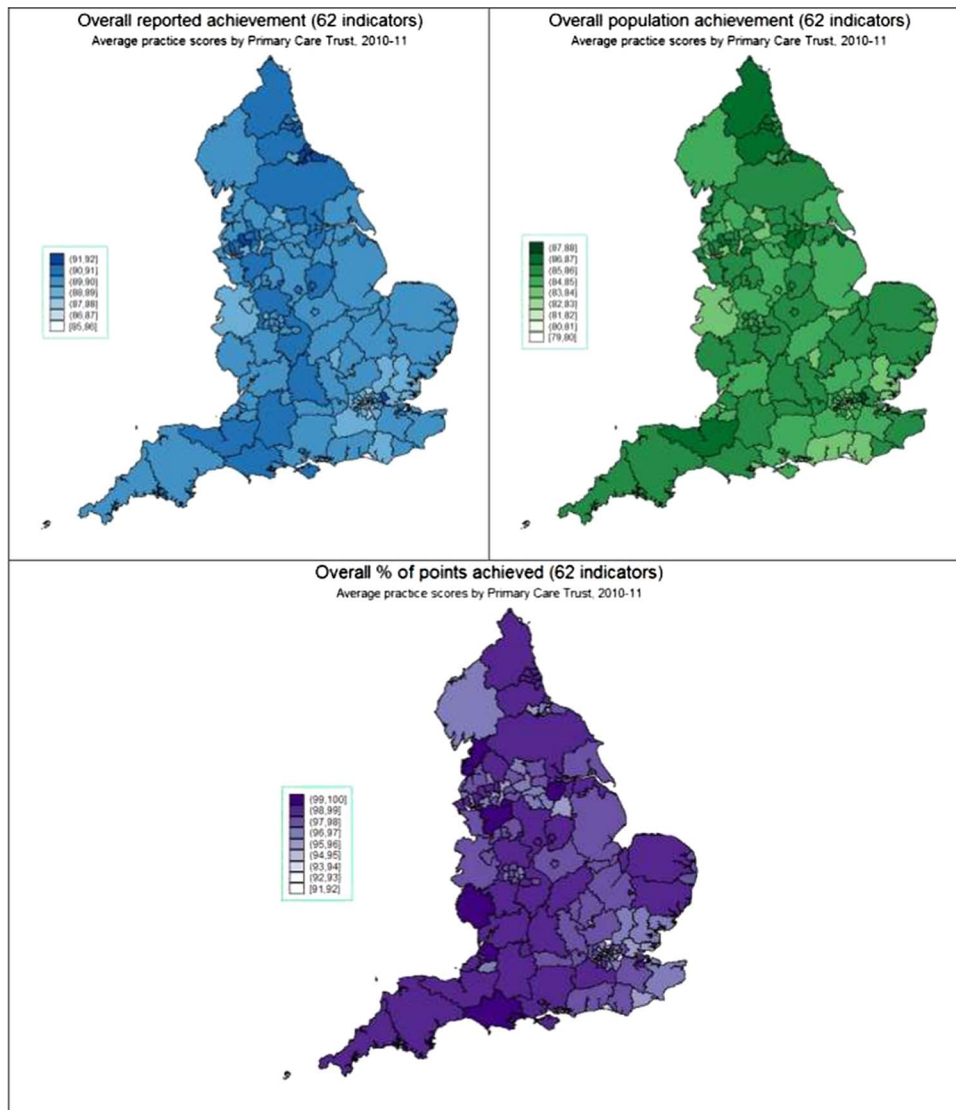Italics highlights the top performer.

Figure 1 Spatial maps of England, for each of the outcome measures.

although its use declined over time (from 46% in 2007–2008 to 39.9% in 2010–2011). Vision 3 by In Practice Systems was the second most popular choice, with a relatively stable proportion of the market (around 19%). The third most popular choice was PCS by EMIS with a stable market share of around 15%, although it has been superseded by EMISWeb and is no longer marketed. The number of practices using TPP by ProdSysOneX more than doubled in the study period, from 697 in 2007–2008 (8.4%) to 1466 in 2010–2011 (17.8%). Synergy and Premiere from iSoft had a combined market share of 9.5% in 2007–2008 but declined to 5.5% in 2010–2011. Practice Manager by Microtest was used by approximately 2% of English practices throughout the time period. EMISWeb and GV by EMIS, HealthyV5 and Crosscare by Healthy Software, Seetec GP Enterprise by Seetec, Ganymede and System 6000 by iSoft and Exeter GP System by Protechnic Exeter were used in the remaining 1% of practices.

Variation in practice characteristics by system

Practice characteristics and performance by system are presented in table 3. In 2010–2011 Practice Manager, Vision 3 and ProdSysOneX practices had the highest RA scores, with Vision 3 practices scoring higher for measurement indicators, ProdSysOneX and Practice Manager practices for treatment indicators and Vision 3 and Practice Manager practices for outcome indicators. For PA, overall Synergy practices were collectively the highest performers, with LV and Practice Manager practices achieving similar levels of performance on the treatment and outcome domains, respectively. Synergy practices also returned the highest percentage of points per practice, on average, for measurement and treatment indicators, closely followed in these domains by Premiere practices, which, along with Practice Manager practices, scored the highest for outcomes.

Practice and patient characteristics varied by system. For example, average list sizes were highest for Synergy

Table 3 Practice characteristics by system, 2010/2011: mean (SD) or percentage

	LV	Vision 3	ProdSysOneX	PCS	Synergy	Practice Manager	Premiere
Reported achievement							
Measurement*	92.9 (3.6)	93.5 (3.3)	93.2 (3.2)	92.5 (3.7)	93.4 (2.3)	93.4 (2.9)	93.2 (2.2)
Treatment†	90.6 (3.8)	90.5 (4.4)	90.7 (3.8)	89.8 (4.2)	90.3 (2.7)	90.7 (3.3)	89.9 (3.1)
Outcome‡	78.8 (4.6)	80.7 (5.2)	79.9 (5.1)	78.4 (5.4)	79.8 (3.9)	80.7 (4.8)	79.9 (4.3)
Population achievement							
Measurement*	90.3 (4.0)	90.5 (3.8)	89.8 (3.8)	89.6 (4.1)	90.9 (2.7)	90.6 (3.4)	90.5 (2.7)
Treatment†	81.2 (4.0)	80.9 (4.6)	80.6 (3.8)	80.8 (4.2)	81.1 (3.3)	80.8 (3.8)	80.5 (3.2)
Outcome‡	73.6 (5.2)	74.6 (5.5)	73.2 (6.0)	72.5 (5.8)	75.0 (4.5)	74.9 (5.2)	74.5 (4.4)
Percentage of points achieved							
Measurement* (of 274)	97.6 (5.6)	97.3 (6.8)	96.9 (6.2)	96.6 (7.2)	98.8 (3.2)	98.0 (5.9)	98.6 (3.2)
Treatment† (of 124)	98.4 (4.8)	98.0 (6.2)	97.4 (6.6)	97.3 (7.1)	99.1 (2.9)	98.5 (5.5)	98.9 (2.9)
Outcome‡ (of 169)	97.0 (5.7)	97.7 (5.6)	97.0 (5.6)	96.3 (6.9)	98.1 (3.5)	98.3 (4.4)	98.3 (4.1)
List size	6903 (4174)	6681 (4286)	6587 (4333)	5715 (3922)	8053 (4287)	6654 (3791)	7104 (3703)
IMD 2010§	25.2 (17.0)	25.8 (16.6)	28.3 (18.0)	31.1 (18.4)	23.5 (16.7)	25.1 (13.7)	21.1 (15.0)
Percentage of female patients	49.7 (2.9)	49.8 (2.6)	49.6 (3.1)	49.4 (3.5)	50.0 (2.2)	50.3 (1.7)	50.0 (1.7)
Percentage of patients aged 65 or over	15.4 (6.2)	15.0 (5.8)	15.4 (5.9)	14.3 (6.9)	16.8 (5.5)	19.6 (5.3)	16.3 (4.1)
General practitioner (GP) age	47.6 (7.5)	48.3 (7.9)	47.3 (7.8)	48.2 (8.5)	46.3 (6.4)	47.4 (6.2)	47.9 (7.3)
Percentage of female GPs	43.5 (26.3)	39.5 (27.3)	38.6 (26.5)	38.9 (28.8)	45.9 (22.8)	37.3 (24.4)	42.1 (27.5)
Percentage of UK-qualified GPs	70.6 (35.3)	62.8 (37.2)	63.0 (37.7)	60.3 (39.4)	77.7 (29.9)	84.9 (27.4)	73.5 (34.3)
Percentage of GP providers¶	74.0 (25.1)	73.4 (27.6)	73.0 (30.7)	72.2 (31.1)	72.0 (23.0)	78.1 (22.4)	79.1 (21.7)
Percentage of urban practices**	82.90%	88.80%	83.90%	89.60%	85.30%	66.70%	84.80%
Percentage of GMS practices††	59.00%	59.00%	40.90%	54.00%	46.60%	65.40%	64.40%
Strategic Health Authority‡‡							
East Midlands	49.30%	6.70%	27.70%	11.90%	3.50%	0.00%	0.20%
East England	35.80%	22.40%	2.20%	26.60%	6.00%	1.60%	2.20%
London	24.80%	2.50%	58.80%	12.00%	1.20%	0.00%	0.40%
North-East	35.40%	5.40%	48.20%	6.50%	2.50%	0.30%	1.10%
North-West	42.00%	17.90%	5.10%	22.90%	7.40%	0.90%	2.20%
South-Central	33.90%	11.70%	41.40%	6.10%	4.30%	0.80%	1.50%
South-East Coast	50.50%	26.00%	1.60%	16.60%	2.50%	0.20%	1.30%
South-West	36.80%	36.50%	6.80%	10.90%	6.00%	0.90%	1.40%
West Midlands	44.80%	27.20%	4.40%	8.90%	6.90%	2.40%	4.60%
Yorkshire-Humber	42.60%	13.80%	11.30%	8.00%	9.40%	13.50%	1.10%

*35 Indicators.
†16 Indicators.
‡11 Indicators.
§Index of multiple deprivation at the Lower Super Output Area level; range 0–85 (more deprived).
¶General practitioners who has had at least 5 years in training as a foundation doctors and specialty registrars.
**Area classed as urban if population at the Middle Super Output Area level is 10 000 or more. Non-urban areas cover towns, fringes, villages, hamlets and isolated dwellings.
††General Medical Services (GMS) contract. Personal Medical Services (PMS), Alternative Provider Medical Services (APMS) and Primary Care Trust Medical Services (PCTMS) contracts are classed together.
‡‡Clinical systems distribution within each Strategic Health Authority presented (row percentages).

practices and lowest for PCS practices, and average area deprivation was lowest for Premiere practices and highest for PCS practices. There was also clear geographical variation in system distribution, for example, the LV system was used by 24.8% of practices in London and 50.5% of practices in the South East.

Variation in practice performance

There was little change in RA, percentage of points scored or PA over time, with PA being higher by 0.42% (95% CI 0.32% to 0.52%) in 2008–2009 and 0.32% (95% CI 0.22% to 0.42%) in 2010–2011, compared with 2007–2008 (see online supplementary table A1). Most practice and patient characteristics had significant but small effects on performance, and these effects often varied by outcome (see online supplementary table A1). For example, for every additional 1000 patients on the practice list RA decreased by –0.11% (95% CI −0.12% to −0.10%), PA decreased by 0.14% (95% CI 0.13% to 0.15%), and percentage of points scored increased by 0.08 (95% CI 0.07 to 0.10). Practices located in more deprived areas performed worse across all three outcomes: by –0.01 (95% CI −0.012 to −0.007) for RA, −0.02 (95% CI −0.022 to −0.016) for PA and −0.006 (95% CI −0.01 to −0.002) for percentage of points scored, per 1 point increase on the IMD scale. Over the range of area deprivation this is equivalent to 0.8% higher RA, 1.6% higher PA and 0.48% higher percentage of points scored in the most affluent compared with the most deprived areas. Rural practices scored lower on RA and points but not on PA. Practices with a higher proportion of female patients performed better on all three outcomes.

In the regression models, overall performance (across all 62 clinical indicators) differed significantly by clinical system used, for all three outcomes (model 1: table 2 and see online supplementary table A1). The systems with the best performing practices were Vision 3 for RA and Synergy for PA and percentage of points scored.

The system with the worst performing practices was PCS, across all three measures. The clinical system rankings for each outcome, based on predictions from the linear regression models, are presented in table 2 and pairwise comparisons in the online appendix (see online supplementary tables A3–5).

Relative monetary gains by clinical system, based on the predictions of points achieved, are displayed in figure 2. Compared with PCS practices (the worst performing on points achieved), practices using Synergy were predicted to gain, on average, an additional £602/year.

Performance by type of activity varied significantly across clinical systems, for all three outcomes (model 2: table 2 and see online supplementary table A2). In the measurement domain, as in overall performance, the systems with the best performing practices were Vision 3 for RA and Synergy for PA and percentage of points scored; the systems with the worst performing practices were PCS, for reported and PA, and Practice Manager for percentage of points scored. In the treatment domain, the systems with the best performing practices were Vision 3 for RA, LV for PA and Premiere for points scored; the systems with the worst performing practices were PCS for RA and Practice Manager for PA and percentage of points scored. Finally, in the outcome domain, the systems with the best performing practices were Practice Manager for RA, Synergy for PA and Vision 3 for points scored; the worst performing system was PCS, across all three measures. The clinical system rankings for each outcome and indicator group are presented in table 2. Pairwise comparisons between systems are presented in online supplementary tables A3–5.

DISCUSSION

The UK's QOF was developed to reward high-quality primary care for a range of chronic conditions, with the ultimate aim of improving patient outcomes. The framework relies heavily on an infrastructure of clinical

Figure 2 Relative annual gains for clinical systems, overall and by indicator domain, compared with PCS. Notes: The system with the worst-performing practices on percentage of overall points scored. For the calculations we used the prediction scores from the indicator group analysis (model 2) and the average number of available points across years, attributing £126 to a point.
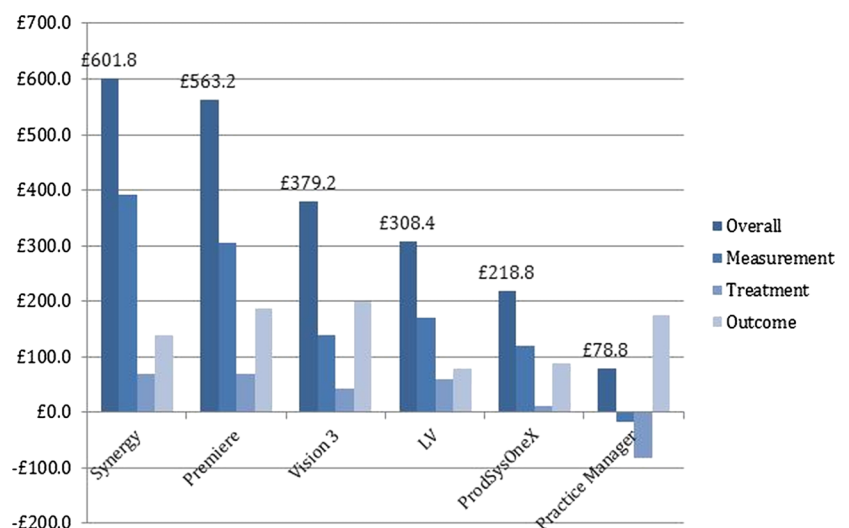
computing systems to: register and classify eligible patients; remind practice staff of the quality indicators; monitor progress towards the targets and to notify the payer of practice achievement. Subject to meeting interoperability criteria, various commercial providers were permitted to provide this infrastructure, and multiple clinical computer systems were developed with different user interfaces, mechanisms and even variations in clinical coding lists.[28] We found that practice performance on the quality indicators contained within the QOF varied significantly across clinical computer systems. This variation persisted when we controlled for a range of practice and patient characteristics, raising the possibility that differences in practice performance under the QOF may be partly attributable to architectural differences between the software systems they use.

Strengths and limitations of the study

The study uses data for all English practices participating in the QOF, covering over 99% of the population registered with Primary Care services and there is therefore no inductive fallacy risk.

However, there are several limitations to the study. First, the QOF was introduced in 2004–2005 but clinical systems information was only available from 2007 to 2008 onwards. Our findings might have been different in previous years, especially given the greater variation in practice performance in the early years of the incentivisation scheme.[23] Second, we did not investigate the mechanisms behind the clinical system variation, for example by focusing on particular indicators for which variation is greater and the recording process might differ across systems. Third, we report quality of *recorded* care and there may be differences with care actually delivered. However, improvement in measurement is a necessary pre-requisite for improved quality of care. Fourth, reported quality of care, as measured by the QOF indicators, represents only a fraction of the care provided by a practice, and the effect of clinical computing systems on the quality of care for other (non-incentivised) activities might be different. Fifth, causality is difficult to establish in observational studies, and it is possible that more QOF-oriented practices have particular clinical system preferences. Sixth, even trivial effects can be found to be statistically significant when analysing very large populations and we have therefore focused our discussion on the effects and their interpretation rather than CIs and p values. Seventh, following the incentivisation scheme, we used indicators that were not independent since some would apply to the same patient. Eighth, timing of computing system introduction to the market might be an important predictor of performance, but we did not include this as a covariate to avoid 'filtering out' the system effect we aimed to measure (ie, we do not wish to control the analyses for system characteristics). Finally, the relationship between clinical computing system and recorded quality of care is confounded by location, since system distribution varied by region. However, our analyses are controlled for two levels of geographic classification.

Findings

Our study analysed years 4–7 of the scheme, by which time variation in performance between practices was much less than in the first year.[29] Nevertheless, as in previous studies we found that several practice and population characteristics—such as list size, local area deprivation and rural location—were associated with performance on the QOF,[23][30][31] although these effects were small.

Larger practices performed worse on RA and PA and better on point scoring, a finding that might indicate that larger practices have processes in place that enable them to maximise their QOF returns despite having slightly lower levels of achievement, compared with smaller practices. The better point scoring of larger practices has been identified in the past but attributed to better performance in non-clinical aspects of the QOF.[32] The inequality gap between practices in deprived and affluent areas has been found to be diminishing over time, with the former catching up with the latter,[23] although there is evidence that the gap had been exaggerated with the introduction of the financial scheme, since more affluent practices were quicker to adapt and maximise their QOF performance.[33] In agreement with previous work, we found small practice location deprivation effects across the three outcome measures. The gap between affluent and deprived practices was wider for PA, than for RA, which is not surprising: comorbidity levels are higher[34] and education levels are lower in more deprived areas (education deprivation is one of the seven domains of the IMD score), potentially leading to higher exception rates due to contraindication or informed dissent. In the first year of the scheme, remote practices from urban centres were found more likely to score lower on points and RA but not PA,[35] and our findings are in agreement. The slightly better performance of practices serving a larger percentage of female patients might indicate that males are less receptive or cooperative in consultations, although previous work identified slightly higher levels of diabetes care for male patients.[33]

Practice characteristics also varied by computing system, for example: practices using Synergy tended to be larger than average, and practices in more deprived areas favoured PCS. The distribution of computing system usage varied by region, which may be the result of market penetration, 'critical mass' effects (with clinicians becoming familiar with particular systems as trainees and continuing to use them throughout their careers), or the influence of PCTs.

The differences in performance between groups of practices using different clinical computing systems were also small in absolute terms: for example, the modelled difference between the best and worst performing systems for overall RA was 1.4%. However, the association

between clinical computing system and QOF performance was stronger than for any other patient or practice characteristic, including list size, proportion of patients over the age of 65 and local area deprivation. The differences between computing systems were greatest for intermediate outcomes indicators (such as control of blood pressure), with a gap between the best and worst performing systems of 2.7% for RA.

These differences are not negligible, especially in light of the small overall variation in practice performance and further convergence in performance over time,[23][36] and the diminishing variation in recorded care between population groups.[33] In addition, differences become substantial at the population level. For example, a 1% difference in achievement of blood pressure control targets for hypertensive patients equates to 9 patients in the average practice and over 71 000 patients nationally.

On the other hand, remuneration differences across systems are very small, considering the average practice is awarded around £120 000/annum through the scheme. This discrepancy between remuneration equity and performance can be explained by the very high levels of performance which, for most practices, surpass the upper thresholds over which no further monetary gains are obtained.[37][38]

Overall, the best average performance across the three outcomes measures was achieved by practices using Vision 3, Synergy or Premiere systems (with one exception on RA, where practices with Practice Manager ranked second). It is notable that two of these systems— Synergy and Premiere—were installed in a diminishing minority of practices (falling from 9.5% to 7.4% over the study period) and are now likely to be withdrawn from the market.[39]

There are several reasons why particular clinical computing systems might facilitate better performance on quality schemes such as the QOF: usability and intuitiveness use of alerts and notifications, dismissability of reminders, support and training and adaptability. In particular, QOF recording relies heavily on the use of data entry templates and it is plausible that the design and usability of these may have an impact. However, the interactions between the users and developers of software systems are likely to be complex and to impact on quality of care in unpredictable ways. For example, early adopters may be more likely to be oriented towards computerised medicine and to have developed familiarity with IT systems, but may also adhere to outdated systems and heuristics, either due to habituation or lack of resources for reinvestment. System differences are also likely to vary by activity, given the different workflows and data linkage processes involved (eg, referrals to specialists as compared with obtaining laboratory results). This is an underresearched area, and the mechanisms underlying the differences in outcomes for different systems need to be examined in greater detail. It is also necessary to examine whether adapting clinical computing systems to support a pay-for-performance scheme impacts on non-incentivised aspects of care, either through neglect of other elements of software development or by reinforcing particular behaviours in clinicians.

CONCLUSIONS

In the UK, performance on the QOF, the world's largest health-related pay-for-performance scheme, is partly dependent on the clinical computing system used by practices. The raises the question of whether particular characteristics of computing systems facilitate higher quality of care, better data recording or both. This question is of interest to clinicians and to policy makers, for whom this work highlights an inconsistency across clinical computer systems which needs to be understood and addressed. For health services researchers, our findings identify an important variable to include in future studies of clinical performance, and an additional factor to consider when generalising findings from samples of providers based on a single clinical computing system. These cautionary messages are also relevant for other international healthcare systems, particularly those with multiple software providers or without stringent interoperability standards.

Author affiliations
[1]NIHR School for Primary Care Research, Centre for Primary Care, Institute of Population Health, University of Manchester, Manchester, UK
[2]Centre for Biostatistics, Institute of Population Health, University of Manchester, Manchester, UK
[3]Centre for Health Informatics, Institute of Population Health, University of Manchester, Manchester, UK
[4]Department of Health Sciences, University of York, York, UK

REFERENCES

1. Clayton PD, Hripcsak G. Decision support in healthcare. *Int J Biomed Comput* 1995;39:59–66.
2. Shortliffe EH. Computer programs to support clinical decision making. *JAMA* 1987;258:61–6.
3. Garg AX, Adhikari NK, McDonald H, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;293:1223–38.
4. Shekelle PG, Goldzweig CL. Costs and benefits of health information technology: an updated systematic review. QQUIP (Quest for Quality and Improved Performance): The Health Foundation, 2009.
5. Kawamoto K, Houlihan CA, Balas EA, *et al*. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005;330:765.
6. Delpierre C, Cuzin L, Fillaux J, *et al*. A systematic review of computer-based patient record systems and quality of care: more randomized clinical trials or a broader approach? *Int J Qual Health Care* 2004;16:407–16.
7. Bates DW, Leape LL, Cullen DJ, *et al*. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA* 1998;280:1311–16.
8. Hunt DL, Haynes RB, Hanna SE, *et al*. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA* 1998;280:1339–46.
9. Koppel R, Metlay JP, Cohen A, *et al*. Role of computerized physician order entry systems in facilitating medication errors. *JAMA* 2005;293:1197–203.
10. De Lusignan S, Chan T. The development of primary care information technology in the United Kingdom. *J Ambul Care Manage* 2008;31:201–10.
11. Department of Health. *Information for health: an information strategy for the modern NHS 1998–2005*. London: Department of Health, NHS Executive, 1998:123.
12. Roland M. Linking physicians' pay to the quality of care—a major experiment in the United Kingdom. *N Engl J Med* 2004;351:1448–54.
13. Checkland K, McDonald R, Harrison S. Ticking boxes and changing the social world: data collection and the new UK general practice contract. *Soc Policy Adm* 2007;41:693–710.
14. Mitchell C, Dwyer R, Hagan T, *et al*. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract* 2011;61:e279–89.
15. National Health Service Information Centre. The Quality and Outcomes Framework. 2012. http://www.ic.nhs.uk/qof
16. Communities and Local Government. The English Indices of Deprivation 2010: Technical Report. Department for Communities and Local Government, 2011.
17. Bibby P, Shepherd J. *Developing a new classification of urban and rural areas for policy purposes—the methodology*. London: DEFRA, 2004.
18. Communities and Local Government. *Indices of deprivation 2010*. 2012. http://www.communities.gov.uk/communities/research/indicesdeprivation/deprivation10/
19. Office of National Statistics. Neighbourhood Statistics. 2012. http://www.neighbourhood.statistics.gov.uk/
20. Edina. UKBORDERS: Boundary datasets and geographic look up tables of the United Kingdom. 2012. http://edina.ac.uk/ukborders/
21. Doran T, Kontopantelis E, Fullwood C, *et al*. Exempting dissenting patients from pay for performance schemes: retrospective analysis of exception reporting in the UK Quality and Outcomes Framework. *BMJ* 2012;344:e2405.
22. Department of Health. *GMS Statement of Financial Entitlements (SFE) 2005 onwards*. London: Department of Health, 2005:250.
23. Doran T, Fullwood C, Kontopantelis E, *et al*. Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. *Lancet* 2008;372:728–36.
24. Scheffé H. *The analysis of variance*. New York; London: John Wiley and Sons, 1959.
25. McGuinness D, Bennett S, Riley E. Statistical analysis of highly skewed immune response data. *J Immunol Methods* 1997;201:99–114.
26. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1986;1:54–75.
27. Stata Statistical software for Windows [program]. 12.1 version, 2011.
28. Tai TW, Anandarajah S, Dhoul N, *et al*. Variation in clinical coding lists in UK general practice: a barrier to consistent data entry? *Inform Prim Care* 2007;15:143–50.
29. Doran T, Fullwood C, Gravelle H, *et al*. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006;355:375–84.
30. Ashworth M, Armstrong D. The relationship between general practice characteristics and quality of care: a national survey of quality indicators used in the UK Quality and Outcomes Framework, 2004–5. *BMC Fam Pract* 2006;7:68.
31. Gray J, Millett C, Saxena S, *et al*. Ethnicity and quality of diabetes care in a health system with universal coverage: population-based cross-sectional survey in primary care. *J Gen Intern Med* 2007;22:1317–20.
32. Wang Y, O'Donnell CA, Mackay DF, *et al*. Practice size and quality attainment under the new GMS contract: a cross-sectional analysis. *Br J Gen Pract* 2006;56:830–5.
33. Kontopantelis E, Reeves D, Valderas JM, *et al*. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. *BMJ Qual Saf* 2013;22:53–64.
34. Salisbury C, Johnson L, Purdy S, *et al*. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *Br J Gen Pract* 2011;61:e12–21.
35. McLean G, Guthrie B, Sutton M. Differences in the quality of primary medical care services by remoteness from urban settlements. *Qual Saf Health Care* 2007;16:446–9.
36. Doran T, Campbell S, Fullwood C, *et al*. Performance of small general practices under the UK's Quality and Outcomes Framework. *Br J Gen Pract* 2010;60:e335–44.
37. Reeves D, Doran T, Valderas JM, *et al*. How to identify when a performance indicator has run its course. *BMJ* 2010;340:c1717.
38. Kontopantelis E, Doran T, Gravelle H, *et al*. Family doctor responses to changes in incentives for influenza immunization under the U.K. Quality and Outcomes Framework pay-for-performance scheme. *Health Serv Res* 2012;47(3 Pt 1):1117–36.
39. Todd R. CSC to withdraw from primary care. *Ehealth Insider* 17 September 2012. http://www.ehi.co.uk/news/ehi/8063/csc-to-withdraw-from-primary-care