

Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: longitudinal observational study.

Evangelos Kontopantelis, research fellow¹, David Reeves, senior research fellow¹, Jose M. Valderas, senior clinical research fellow^{2 3}, Stephen Campbell, reader¹, Tim Doran, clinical research fellow¹.

¹Health Sciences Primary Care Research Group, 5th floor Williamson Building, University of Manchester, Manchester, United Kingdom.

²Health Services and Policy Research Group, NIHR School for Primary Care Research, Department of Primary Health Care, University of Oxford, Oxford, United Kingdom

³European Observatory of Health Systems and Policies, London School of Economics, London, United Kingdom.

Correspondence to:

Evangelos Kontopantelis

e-mail: e.kontopantelis@manchester.ac.uk

Abstract

Background: The UK's Quality and Outcomes Framework (QOF) was introduced in 2004/05, linking remuneration for general practices to recorded quality of care for chronic conditions, including diabetes mellitus. We assessed the effect of the incentives on recorded quality of care for diabetes patients and its variation by patient and practice characteristics.

Methods: Using the General Practice Research Database we selected a stratified sample of 148 English general practices in England, contributing data from 2000/01 to 2006/07, and obtained a random sample of 653,500 patients in which 23,920 diabetes patients identified. We quantified annually recorded quality of care at the patient-level, as measured by the 17 QOF diabetes indicators, in a composite score and analysed it longitudinally using an Interrupted Time Series design.

Results: Recorded quality of care improved for all subgroups in the pre-incentive period. In the first year of the incentives, composite quality improved over-and-above this pre-incentive trend by 14.2% (13.7% -14.6%). By the third year the improvement above trend was smaller, but still statistically significant, at 7.3% (6.7%-8.0%). After three years of the incentives, recorded levels of care varied significantly for patient gender, age, years of previous care, number of co-morbid conditions and practice diabetes prevalence.

Conclusions: The introduction of financial incentives was associated with improvements in the recorded quality of diabetes care in the first year. These improvements included some measures of disease control, but most captured only documentation of recommended aspects of clinical assessment, not patient management or outcomes of care. Improvements in subsequent years were more modest. Variation in care between population groups diminished under the incentives, but remained substantial in some cases.

Keywords

Diabetes, General Practice Research Database (GPRD), Quality and Outcomes Framework (QOF), incentives, recorded quality of care.

Introduction

In the last 15 years the National Health Service in the UK has undergone a series of reforms aimed at improving the quality of care for people with chronic conditions. These include the creation of the National Institute for Health and Clinical Excellence (NICE), and the introduction of National Service Frameworks which set minimum standards for the delivery of health services in specified clinical areas, including diabetes mellitus [1]. The quality of primary care generally, and of diabetes care in particular, improved in the early 2000s [2], partly in response to these quality improvement initiatives [3]. In 2004 new contractual arrangements for family doctors allowed them to opt out of out-of-hours care and linked financial incentives to quality of care under the Quality and Outcomes Framework (QOF) [4], the largest and most ambitious pay-for-performance scheme ever attempted in health care [5 6].

The QOF initially included 76 clinical indicators, covering a range of processes of care (e.g. measurement of blood pressure) and intermediate outcomes (e.g. glycaemic control). A further 70 indicators covered aspects of practice organisation and patient experience of care. Eighteen of the clinical indicators related to care for patients with diabetes, reflecting in part the national importance of the disease, the recorded prevalence of which had increased by 75% in the previous decade [7]. By achieving all the diabetes targets, an average practice could earn over £7,500 in the first year of the scheme and £12,500 in subsequent years, when incentives were increased [8].

To date, most studies on the effect of the QOF on diabetes care have been conducted with aggregate data at the practice level. These studies suggest that the quality of diabetic care increased over and above the underlying trend in the first three years of the QOF scheme [9], leading to modest reductions in emergency hospital admissions [10], but this effect was not sustained [11 12]. However, Calvert et al, using patient level data, concluded that the introduction of the QOF scheme had little impact, although practices from which data were drawn were not necessarily nationally representative and the analyses did not control for potential confounders including age and co-morbidity [13]. Millett et al investigated the effect of the QOF on three of the eighteen incentivised activities – control of blood pressure, cholesterol

and HbA1c levels. They concluded that improvements in care in the first year of the scheme were greater for patients with multiple medical conditions [14] and that improvements were similar for patients from affluent and deprived areas [15].

We use data for individual patients registered with a nationally representative sample of practices to investigate quality and outcomes of diabetes care from 2000/01 to 2006/07. Our approach aimed to: a) establish the effect of the introduction of the QOF on all incentivised aspects of diabetes care at the patient level, b) assess whether there was variation in the impact depending on patient and practice characteristics, and c) identify inequalities of care before and after the intervention.

Methods

Patients and practices

Patient-level data were extracted from the General Practice Research Database (GPRD), which contains complete electronic medical records of patients attending primary care practices in the UK. Practitioners used Read codes, a hierarchical coding rule set, to enter data on their clinical computing systems. In October 2010, data were available for 593 practices and over 10 million patients.

We selected a structured sample of 148 practices. Selected practices were nationally representative in terms of patient sex and age distribution and area socio-economic deprivation, but had a relatively high average list size, reflecting a bias towards larger practices in the GPRD. All patients - registered with a selected practice for at least one day during the study period - were selected from practices with list sizes of 4,500 or below. For larger practices, a random selection of 4,500 patients was drawn from each practice. By including a larger percentage of patients from smaller practices we were able better control for the nested structure of the data in the analyses. We do not expect bias to have been introduced since practice list size seems to have no effect in quality of diabetes management under the QOF [16]. The final sample consisted of

653,500 patients. The characteristics of the sampled practices have been described elsewhere [17].

Study period

We divided the study period into financial years (1 April to 31 March the following year), to correspond to the timings of QOF years. Following QOF inclusion criteria, we selected for each year patients aged 17 or over that had been registered with their practice for the whole of that year. This created three pre- (2000/01, 2001/02 and 2002/03) and three post-implementation time points (2004/05, 2005/06 and 2006/07). We treated the financial year 2003/04 as a transition year, because information on the forthcoming pay-for-performance framework was made publicly available, although practices were not yet financially incentivised. We provide a summary of the main characteristics of the scheme and its remuneration policy in the online appendix; a more detailed description has been provided elsewhere[18].

Diabetes and co-morbidities

Patients with either type 1 or type 2 diabetes mellitus were identified using the relevant diagnostic codes. We did not limit ourselves to the standard QOF business rules on Read codes for identifying patients with diabetes, since prevalence estimates based only on these codes may underestimate the true prevalence due to changes in coding behavior following the introduction of the scheme and subsequent changes to QOF business rules [19]. We therefore identified additional diagnostic codes linked to diabetes, including codes in the older OXMIS (Oxford Medical Information System) coding scheme. As many diagnostic codes do not differentiate between type 1 and type 2 diabetes mellitus, we did not differentiate between types in the study. The same procedure was used to identify nine co-morbid conditions: asthma, coronary heart disease (including angina), chronic kidney disease, chronic obstructive pulmonary disease, depression, heart failure, hypertension, hypothyroidism and stroke. We selected conditions for which quality of reporting was verified using prevalence trends reported in previous studies [17]. We also extracted information on patient age, gender and date of diabetes diagnosis. Deprivation was measured at the practice level using the Index of Multiple Deprivation (IMD) score at the super-output area level, a low level geographical area categorisation with unchanged boundaries

over time and a relatively small and consistent population size across areas (around 1,500) [20]. Codes used to create condition registers are available from the authors upon request.

Quality of care

Quality of care was measured by achievement of the 17 diabetes quality indicators from the QOF (Table 1). For each patient in each year, an indicator was deemed to be achieved if a relevant process event or outcome was identified in the required time period (usually 15 months prior to the end of the financial year - see Table 1) and a diagnosis of diabetes pre-dated the event.

A composite quality of care score was calculated for each patient and time point as the number of indicators achieved for that patient as a percentage of the number that applied to that patient:

$$QS_{jy} = 100 \frac{\sum_{i=2}^{18} DMm_{ijy}}{\sum_{i=2}^{18} DMA_{ijy}}$$

where

QS_{jy} the composite recorded quality score for patient j in year y

DMA_{ijy} a binary variable quantifying whether indicator i was applicable to patient j in year y ($DMA_{ijy} = 1$) or not ($DMA_{ijy} = 0$)

DMm_{ijy} a binary variable quantifying whether indicator i was applicable and met for patient j in year y , in which case $DMm_{ijy} = 1$. In all other cases, i.e. applicable but unmet or not applicable, $DMm_{ijy} = 0$.

The scores had a Cronbach's alpha coefficient of 0.82, above the threshold for acceptable reliability of 0.7, and represent the degree to which each patient was recorded to have received the 'necessary care' for their diabetes, as stipulated by the QOF indicator set.

Statistical analyses

We modelled the changes over time in the patient-level composite recorded QOF care score using Interrupted Time Series (ITS) methods, to estimate the changes in level and trend associated with the intervention [21]. This quasi-experimental design uses the pre-intervention trend of the data to estimate its projection in the first post-intervention year and compare it with

the observed data. The approach is similar for following years but the post-intervention slope also needs to be taken into account.

Therefore, our model estimated three key components of the trend across time: i) the slope in patient scores in the pre-QOF years (2000/01-2002/03); ii) the change in level in scores in the first year of the intervention (2004/05); iii) the change in slope in scores from the pre-QOF to the QOF period (2000/01-2002/03 to 2004/05-2006/07). We then used these components in post-estimation tests to examine the impact of pay-for-performance on overall recorded QOF patient care at two points: (i) 2004/05 ('short-term' impact) and (ii) 2006/07 ('long-term' impact).

We conducted three main analyses. Analysis 1 examined the overall impact of pay-for-performance in the study population, regardless of any changes in the patient population over time. Analysis 2 controlled for a range of patient and practice characteristics, and examined whether the impact of the intervention varied in relation to these. In analysis 2 interaction terms were added into the model between the three ITS components and a series of covariates: age, gender, time since diagnosis (in years), number of co-morbid conditions, practice diabetes prevalence and area socio-economic deprivation (measured by the Index of Multiple Deprivation at the practice post-code). By introducing covariates at the patient level this analysis also controlled for changes in the practice samples over time. Model complexity necessitated examination of each covariate in a separate regression. Analysis 3 compared mean recorded QOF scores across the levels of each covariate in the pre-QOF (2000/01-2002/03) and post-QOF (2004/05-2006/07) periods. To aid interpretability of results, continuous covariates (e.g. age) were recoded as categorical variables (Table 2). To identify overall differences across covariate categories we performed omnibus significance tests.

Floor and ceiling effects (scores at, or approaching, 0% and 100%) can bias results based on linear models. We therefore applied a logit transformation to the data prior to analysis [9 11]. Instances where the composite score was at 100% or 0% were dealt by applying the empirical logit [22]. The non-loglinear component of the pre-QOF trend was found to be non-significant ($p>0.05$), indicating a good fit to the logit model. Residual autocorrelation was found to be negligible (Pearson's correlation coefficient between the residuals and their lagged values, across patients and over all available time points, was 0.03). Results from the logit-transformed

analyses were back-transformed to percentages for ease of interpretation, but we additionally conducted a sensitivity analysis using the untransformed scores.

The data had a complex multi-level structure. Patients were nested within practices, with each practice sample partially changing year-by-year (as some patients left or joined). We modelled patients across time using the Stata **xtmixed** procedure and allowed for random intercepts at the patient and practice levels. Stata v11.1 and an alpha level of 5% were used throughout.

Results

A total of 23,920 diabetic patients who met our inclusion criteria were identified and 23,780 were included in the regression analyses. The recorded prevalence of diabetes in the study practices increased from 3.3% to 5.0% between 2000/01 and 2006/07. The recorded incidence rate of new cases increased from 0.38% to 0.49% between 2000/01 and 2002/03, and remained relatively stable thereafter. The mean age of diabetic patients increased from 62.9 to 63.7 over the study period, and the mean number of co-morbid conditions increased from 1.4 to 1.7, from a maximum of nine monitored conditions (Table 1).

Quality of diabetes care

Recorded quality of care across the 148 practices increased for all individual indicators between 2000/01 and 2006/07 (Table 1), with absolute improvements ranging from 4.2% (control of HbA1c levels $\leq 10\%$) to 85.5% (providing smoking cessation advice). Recorded QOF care as measured by the composite quality of care score increased from 46.5% in 2000/01 to 81.0% in 2006/07 (Table 2), with scores increasing for all subgroups. On average, the highest scores across the whole period were observed for patients with three or more co-morbid conditions and patients aged 65 or over. Newly diagnosed cases and patients aged 17-29 consistently had the lowest recorded care levels. The relationship between age and recorded QOF care was not uniform across indicators, for example: control of blood pressure (indicator DM12) was, predictably, better for younger patients (see Figure 1 and online appendix).

Overall effect of the introduction of the Quality and Outcomes Framework

In the first year of the QOF (2004/05) there was an improvement in composite recorded QOF care, over-and-above that expected from the pre-intervention trend, of 14.2% (95% CI: 13.7%-14.6%; $p < 0.001$). By the third year (2006/07), the difference was smaller, but still statistically significant, at 7.3% (95% CI: 6.7%-8.0%; $p < 0.001$) (Table 2 and Figure 2).

Variations in care and in the impact of the QOF by patient and practice characteristics.

The effect of the intervention did not vary significantly by age category although older patients were recorded to have received better care throughout the study period. The composite score for patients aged 65 or over was higher than that for patients aged 17 to 39 by on average 11% in the pre-QOF and 11.7% in the post-QOF period. Recorded QOF care was marginally lower for females in both time periods, by around 2%, but the impact of the intervention did not differ by gender. There were no significant differences in the intervention effect between multi-morbidity categories, although recorded QOF care increased with number of conditions, with performance scores for patients with three or more conditions being higher on average by 6.3% in the pre-intervention and 6.1% in the post-intervention periods compared to patients with no co-morbid conditions.

In the pre-intervention period recorded QOF care varied with the number of years living with diabetes. It was highest for patients living with the condition for 1-4 years and lowest for newly diagnosed patients. The gap in recorded care between these two groups widened from an average of 4.7% in the pre-QOF period to 9.1% in the post-QOF period. This is reflected in the intervention effect: compared to new diagnoses, all other patient subgroups were more positively affected by the intervention both in 2004/05 and 2006/07, and to a similar extent (around 6% to 7%).

Recorded QOF care did not vary significantly by area deprivation before or after the introduction of the incentivisation scheme. However, the effect of the intervention did vary with area deprivation: patients attending practices from the most deprived quartile appear to have

gained less from the intervention compared to patients in the most affluent quartile of practices, by 4.9% in 2004/05 and 3.8% in 2006/07.

There was significant variation in recorded QOF care by practice diabetes prevalence rates, but the differences were small and diminished over time. The intervention effect also varied with practice diabetes prevalence. Compared to practices in the first quartile (lowest diabetes prevalence), the QOF effect was larger for practices in the second and third quartiles – by 1.4% and 2.1% respectively in the short term (2004/05) and by 3.2% and 4.8% respectively in the long term (2006/07).

Sensitivity analyses

The sensitivity analysis (based on untransformed data) agreed with the main analysis in all respects except for the relationship between the intervention and patient gender: in the sensitivity analysis both the short- and long-term impact of the pay-for-performance scheme was significantly larger, though small in scale, for female patients (1.2%, $p=0.048$ and 2.5%, $p=0.01$ respectively).

Discussion

The main aim of pay-for-performance schemes is to incentivise physicians to provide high quality care and thereby improve patient outcomes. Research to date suggests that pay-for-performance schemes have limited impact when implemented in isolation, but when supported by other quality improvement initiatives can have a positive effect on quality of care [3]. We found that recorded quality of primary care in the UK, as measured by the QOF diabetes indicators, was already improving prior to the introduction of the scheme in 2004, and that it improved at an accelerated rate in the first years of its implementation, supporting the findings of previous studies [11 23]. This accelerated improvement did not seem to benefit all population groups equally, however.

Strengths and limitations of the study

The main strength of our study was in the use of data for individual patients drawn from a nationally representative sample of practices. However, the study is subject to certain limitations. First, the QOF was introduced universally and was not implemented as part of a randomised experiment. The lack of a practice control group entails that analyses of its effect in quality of care are only possible using quasi-experimental methods. Results obtained with these methods can be method- and assumptions-sensitive, nevertheless, the interrupted time-series design is one of the most effective and powerful of all quasi-experimental designs and is routinely being used as the best possible approach when such research scenarios arise [24]. Second, we are reliant on the accurate and consistent recording of data by practices; however usage of clinical computing systems has changed over time and practices may have exaggerated their performance in response to the financial incentives. This study has investigated the quality of *recorded* diabetes care and there may be differences with care actually delivered. On the other hand, improved measurement is a necessary prerequisite for improved quality of care and one could argue that it is improved quality of care. Third, most of the measures refer to documentation levels and do not necessarily lead to the intended improvements in outcomes if not properly followed-through or the interventions are not offered in an appropriate manner (e.g. advising a patient briefly “perhaps you should consider quitting smoking” in order to ‘tick’ the relevant QOF box is not really a smoking cessation intervention). Fourth, some quality indicators are dependent on others, for example indicator DM12 (blood pressure controlled) cannot be met unless indicator DM11 (blood pressure measured) has also been met. However, we aimed to quantify and assess overall quality of care as measured by the whole diabetes domain in the QOF and to be as inclusive as possible. Fifth, the conditions we modelled to investigate the effect of co-morbidity were not an exhaustive list and only the presence or absence was modelled and not the severity of each condition. However, the number of chronic co-morbidities is a well-established marker of the overall clinical complexity of a patient [25]. Sixth, our findings assume that the observed trends in indicator achievement prior to QOF would have continued unchanged had the incentive scheme not been introduced, while only three years of data were available between the publication of the National Service Framework for diabetes[1] and the implementation of the QOF (not including the transition year), on which we based our trend estimates. Finally, certain patient characteristics, such as ethnicity and socio-economic position, were not adequately recorded in the dataset and therefore could not be directly analysed.

Quality of diabetes care and patient characteristics

Recorded quality of diabetic care continuously improved over the period of our study, against a background of increasing disease prevalence. Recorded prevalence increased from 3.3% to 5% between 2000/01 and 2006/07, which is broadly in line with national trends [26]. Overall levels of recorded quality of care improved in the pre-QOF period across all subgroups, following a general trend which has also been noted in other countries [27 28]. In the first year of the QOF scheme (2004/05), recorded quality improved above the underlying trend. However, gains in the following years were smaller, especially in the third year (2006/07), a finding consistent with previous research [11 23]. Our finding of an increase in recorded quality of care for diabetes in the pre-incentivisation period, leading to difficulties in separating the effect of the intervention, was also noted by Calvert et al [13], who found a significant QOF effect only on tight HbA1c control for type II diabetes patients.

The diminution of the gains in recorded quality of diabetes care under the QOF in the second and third years may be attributed to two factors. First, achievement for some indicators was close to 100% in 2004/05 and large improvements were not possible. Second, in 2004/05 most practices exceeded the maximum payment thresholds (the level of achievement required to secure maximum remuneration) for most indicators, hence there was limited financial incentive for further improvement. As a result, the QOF may not ultimately have resulted in higher quality of care than would have eventually been achieved in its absence, although it may have hastened practices' arrival at that end point. The high annual cost of the QOF scheme – over £1billion per year, with almost 10% (£100m) allocated to diabetes care – needs to be assessed in this light. In addition, the QOF relied on a large investment in information technology, which is likely to have contributed to the increase in recorded levels of care. However, that uniform investment across the whole of the UK primary care would be very unlikely to have taken place in the absence of the incentivisation scheme. Even if the IT investment was there, it is doubtful the GPs would choose to adapt to the new systems altogether, as they have, if not rewarded to do so.

As for differences between patient groups, women appear to have received slightly poorer QOF care both before and after the intervention. Levels of recorded care also varied substantially by patient age, with younger patients consistently receiving poorer quality care. This finding may

be partly attributable to the higher proportion of type 1 diabetic patients in the younger age groups, for whom some quality targets are more difficult to achieve. Improvements in recorded quality over time were similar across all age groups, and the absolute gap between older and younger patients therefore did not close. Hamilton et al reported similar findings, with larger improvements for women, smaller benefits for younger patients and similar improvements for patients in deprived and affluent areas alike [15].

Under the QOF scheme some activities, such as measurement and control of blood pressure, are individually incentivised for patients with different conditions – for example diabetes and coronary heart disease. Practices therefore receive greater financial rewards for achieving these targets for patients with multiple conditions. However, we found that recorded QOF care tended to be higher for patients with more co-morbid conditions throughout the entire study period, including the pre-intervention years, which suggests that practices may be more attentive towards diabetic patients with co-morbidities even in the absence of incentives.

Prior to the introduction of the scheme, newly diagnosed and long-standing (10 or more years) diabetic patients were recorded to have received worse QOF care, compared to patients who had been registered with the condition for 1-9 years. Improvements in recorded quality under the QOF were smaller for newly diagnosed patients and consequently the gap between newly diagnosed and established patients widened. One possible explanation is that cases diagnosed towards the end of the financial year (within 3 months for measurement indicators and 9 months for control indicators) are automatically excluded from QOF payment calculations for that year, and practices may therefore be more likely to defer care for these patients to the following year when they become eligible for participation in the programme. Another possible explanation is that practices may simply have insufficient time to achieve the targets for these patients, especially controlling for risk factors if they have had undiagnosed diabetes for some time.

Quality of diabetes care and practice characteristics

Practices with fewer diabetic patients generally performed better before the intervention, but less well afterwards. In the pre-intervention period, practices in deprived areas provided slightly

higher overall recorded QOF care than those in more affluent areas. Recorded improvements in the first year of the QOF scheme were greater for practices in more affluent areas, however, and practices in deprived areas therefore fell behind. In the second and third year of the scheme practices in deprived areas closed the gap again, as has been noted in previous research [18 29]. This suggests that practices in affluent areas were likely to have responded more quickly to the incentives, thus gaining a financial advantage in the early years of the scheme [23].

Conclusions

The introduction of financial incentives was associated with an acceleration in the underlying improvement in recorded quality of primary care for diabetic patients in the UK, as measured by the QOF indicators. These improvements mostly captured documentation of recommended aspects of clinical assessment and only some measures of disease control. Although Incentive schemes operating in other countries may be less successful, given the size of the financial incentives in the UK and the existence of parallel quality improvement initiatives. Most assessed diabetic patients appeared to benefit, but improvements were not as great for some sub-groups, for example newly diagnosed patients and those attending practices in deprived areas and inequalities between some population groups remain. Overall, the introduction of the QOF significantly improved recorded quality of diabetes care, although the gains from the expensive scheme seem to be diminishing over time.

What is already known on this topic

Aggregate analyses at the practice level suggest that diabetes care was increasing before the introduction of the QOF and continued to increase after the scheme came into effect, above the underlying trend.

These improvements have been linked with modest reductions in emergency hospital admissions, although this effect was not sustained.

However, more recent patient level analyses using data from a primary care database concluded that the impact of the QOF on quality of care for diabetes is unclear.

What this study adds

Although levels of quality of diabetes care were on the increase before the intervention, the QOF did have a significant impact and led to improvements in overall recorded patient-level diabetes care.

All patient subgroups seem to have benefited from the intervention but variation in care between them largely remains.

Patients with a new diagnosis of diabetes benefitted the least from the intervention and their care may be deferred to the next financial year.

Copyright

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, a worldwide licence to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

Funding

This work received no specific funding.

Declaration of competing interests

All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: EK was partly supported by a NIHR School of Primary Care Research (NSPCR) fellowship in primary health care; no other relationships or activities that could appear to have influenced the submitted work.

Authorship & contributorship

EK, TD and DR designed the study. EK extracted the data, and performed the statistical analyses. EK and TD wrote the manuscript. DR, JV and SC edited the manuscript.

EK is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Data sharing

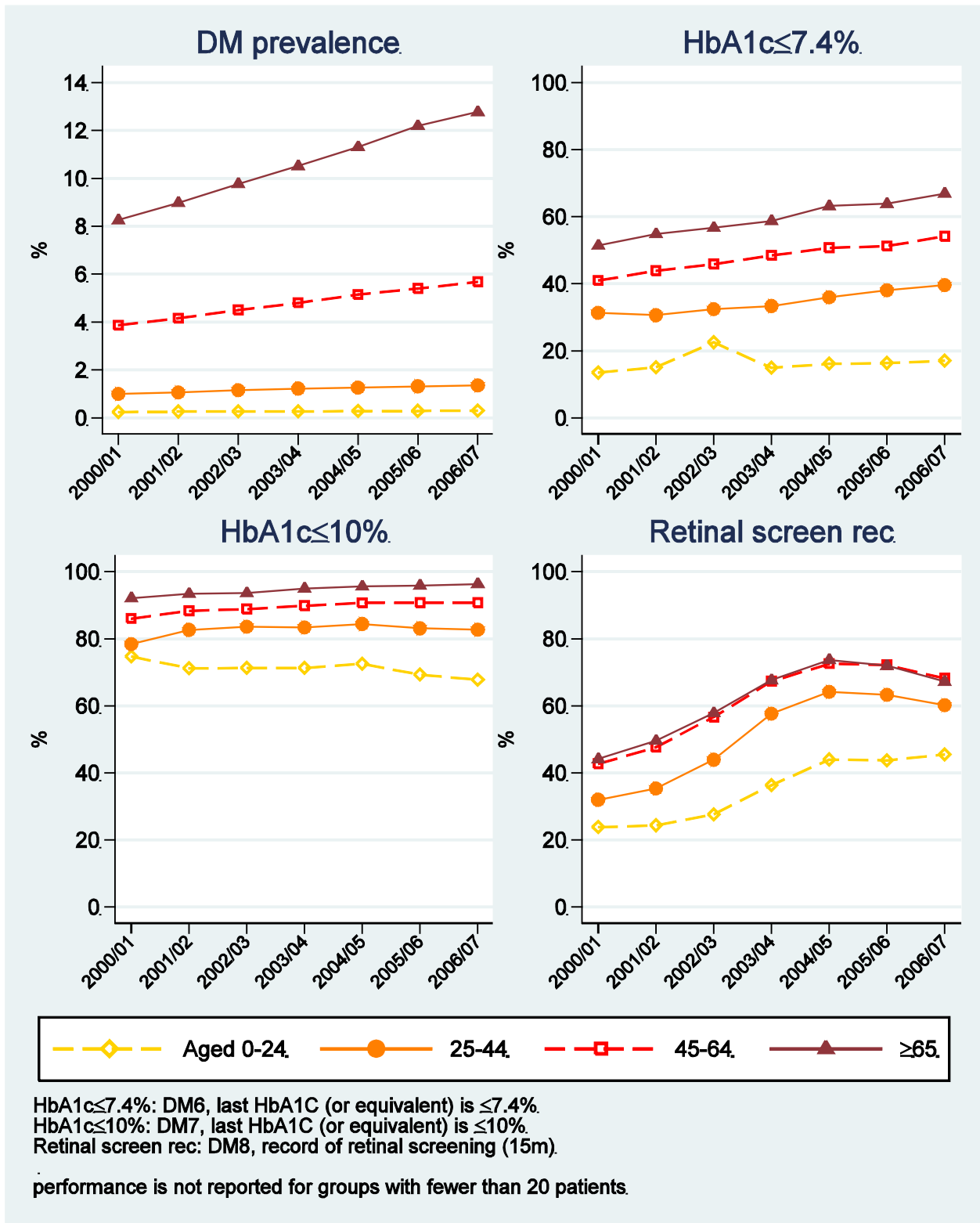
The lists of Read and OXMIS codes used are available from the corresponding author at e.kontopantelis@manchester.ac.uk. However, the General Practice Research Database data cannot be shared due to licencing restrictions.

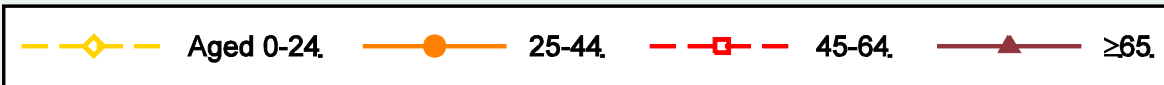
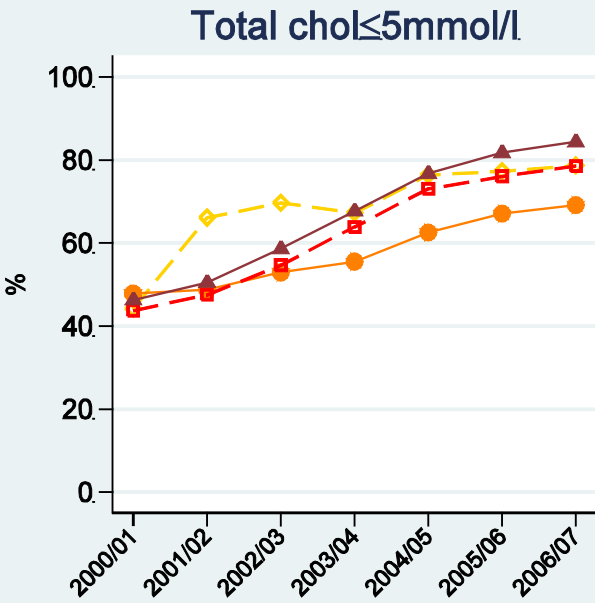
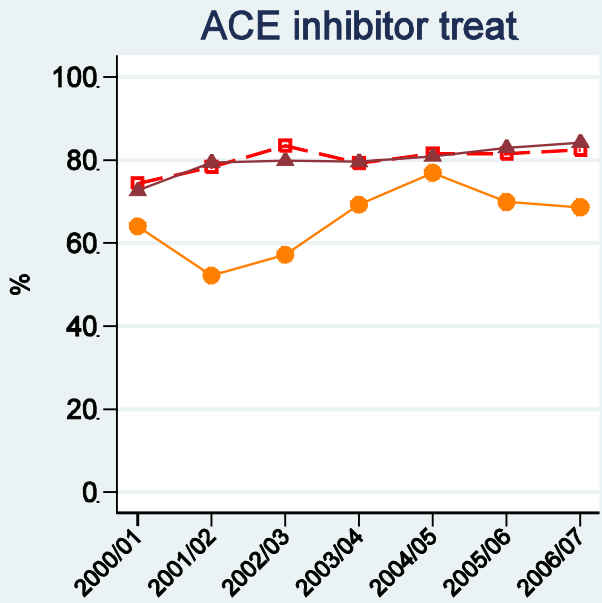
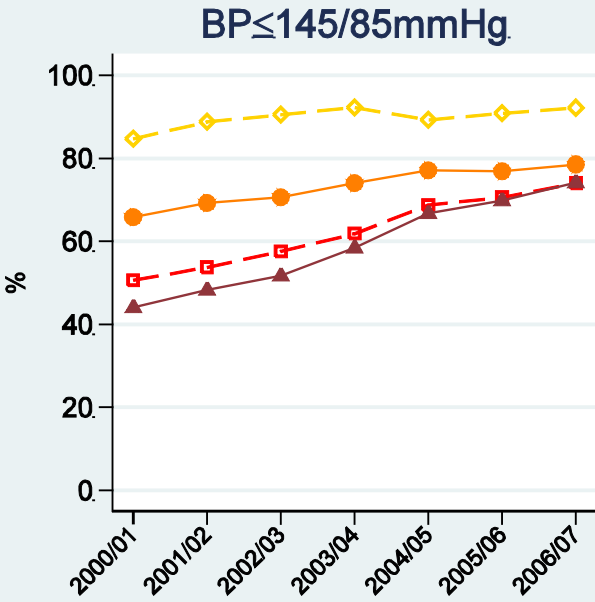
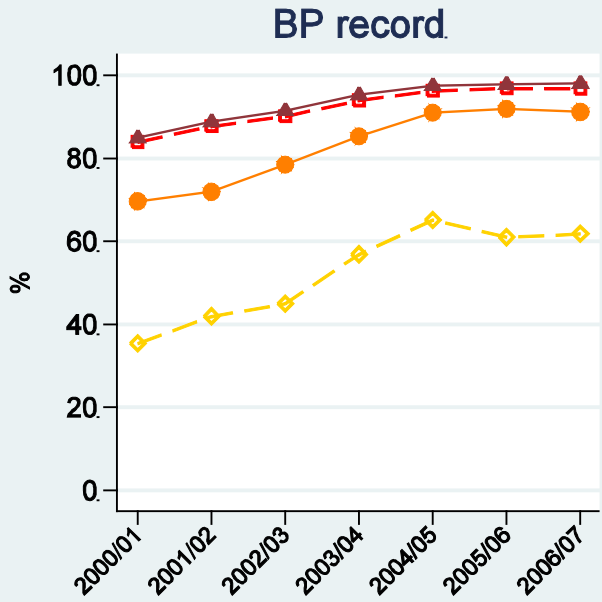
References

1. Department of Health. National service framework for diabetes: standards. Secondary National service framework for diabetes: standards 2001. http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4002951.
2. Campbell S, Steiner A, Robison J, et al. Is the quality of care in general medical practice improving? Results of a longitudinal observational study. *Br.J.Gen.Pract.* 2003;**53**:298-304
3. Doran T, Roland M. Lessons from major initiatives to improve primary care in the United Kingdom. *Health Aff.(Millwood.)* 2010;**29**:1023-29
4. Gubb J, Li G. Checking-up on Doctors: A Review of the Quality and Outcomes Framework for General Practitioners. Civitas Health Unit: Civitas: Institute for the Study of Civil Society, London, 2008.
5. Roland M. Linking physicians' pay to the quality of care-a major experiment in the United Kingdom. *N Engl J Med* 2004;**351**:1448
6. Shekelle P. New contract for general practitioners. *BMJ* 2003;**326**:457-58
7. The Information Centre for health and social care. Health Survey for England 2006, latest trends. Health Survey for England. Leeds: The Information Centre, 2008.
8. The information Centre for health and social care. National Quality and Outcomes Framework Statistics for England 2006/07. The Quality and Outcomes Framework: National Health Service, 2007.
9. Campbell S, Reeves D, Kontopantelis E, et al. Quality of primary care in England with the introduction of pay for performance. *N Engl J Med* 2007;**357**:181-90
10. Dusheiko M, Doran T, Gravelle H, et al. Does Higher Quality of Diabetes Management in Family Practice Reduce Unplanned Hospital Admissions? *Health Serv.Res.* 2010;**46**:27-46
11. Campbell S, Reeves D, Kontopantelis E, et al. Effects of pay-for-performance on the quality of English Primary care. *N Engl J Med* 2009;**361**:368-78
12. 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
13. Calvert M, Shankar A, McManus RJ, et al. Effect of the quality and outcomes framework on diabetes care in the United Kingdom: retrospective cohort study. *BMJ* 2009;**338**:b1870
14. Millett C, Bottle A, Ng A, et al. Pay for performance and the quality of diabetes management in individuals with and without co-morbid medical conditions. *J.R.Soc.Med.* 2009;**102**:369-77
15. Hamilton FL, Bottle A, Vamos EP, et al. Impact of a pay-for-performance incentive scheme on age, sex, and socioeconomic disparities in diabetes management in UK primary care. *J.Ambul.Care Manage.* 2010;**33**:336-49
16. Vamos EP, Pape UJ, Bottle A, et al. Association of practice size and pay-for-performance incentives with the quality of diabetes management in primary care. *Can Med Assoc J* 2011;**183**:E809-E16
17. Doran T, Kontopantelis E, Valderas JM, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ* 2011;**342**:d3590

18. 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
19. Hippisley-Cox J, Vinogradova Y, Coupland C. Time Series Analysis for selected clinical indicators from the Quality and Outcomes Framework 2001-2006. The Quality and Outcomes Framework: QRESEARCH and The Information Centre for health and social care, 2007.
20. Communities and Local Government. The English Indices of Deprivation 2010: Technical Report. Indices of Deprivation. London: Department for Communities and Local Government, 2011.
21. Wagner AK, Soumerai SB, Zhang F, et al. Segmented regression analysis of interrupted time series studies in medication use research. *J Clin Pharm Ther* 2002;**27**:299
22. Collett D. *Modelling binary data*. 2nd ed. Boca Raton: Chapman & Hall/CRC, 2003.
23. Christianson JB, Leatherman S, Sutherland K. Lessons from evaluations of purchaser pay-for-performance programs: a review of the evidence. *Med.Care Res.Rev.* 2008;**65**:5S-35S
24. Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont CA: Wadsworth, 2002.
25. Valderas JM, Starfield B, Sibbald B, et al. Defining comorbidity: implications for understanding health and health services. *Ann.Fam.Med.* 2009;**7**:357-63
26. O'Flaherty M, Critchley J, Wild S, et al. Diabetes: 018 Forecasting diabetes prevalence using a simple model: England and Wales 1993-2006. *J Epidemiol Commun H* 2010;**64**:A7-A7
27. Trivedi AN, Zaslavsky AM, Schneider EC, et al. Trends in the quality of care and racial disparities in Medicare managed care. *N Engl J Med* 2005;**353**:692-700
28. Mundet X, Cano F, Mata-Cases M, et al. Trends in chronic complications of type 2 diabetic patients from Spanish primary health care centres (GEDAPS study): ten year-implementation of St. Vincent recommendations. *Prim Care Diabetes* 2012;**6**:11-8
29. Ashworth M, Seed P, Armstrong D, et al. The relationship between social deprivation and the quality of primary care: a national survey using indicators from the UK Quality and Outcomes Framework. *Br.J.Gen.Pract.* 2007;**57**:441-48

Figure 1: Prevalence and recorded QOF care by age group





BP record: DM11, record of blood pressure (15m).
 BP≤145/85mmHg: DM12, last blood pressure is ≤145/85 mmHg.
 ACE inhibitor treat%: DM15, treated with ACE inhibitors or A2 antagonists (15m).
 Total chol≤5mmol/l: DM17, last measured total cholesterol is ≤5mmol/l.

performance is not reported for groups with fewer than 20 patients.

Figure 2: Aggregate patient level QOF care and predictions based on the pre-incentivisation trend

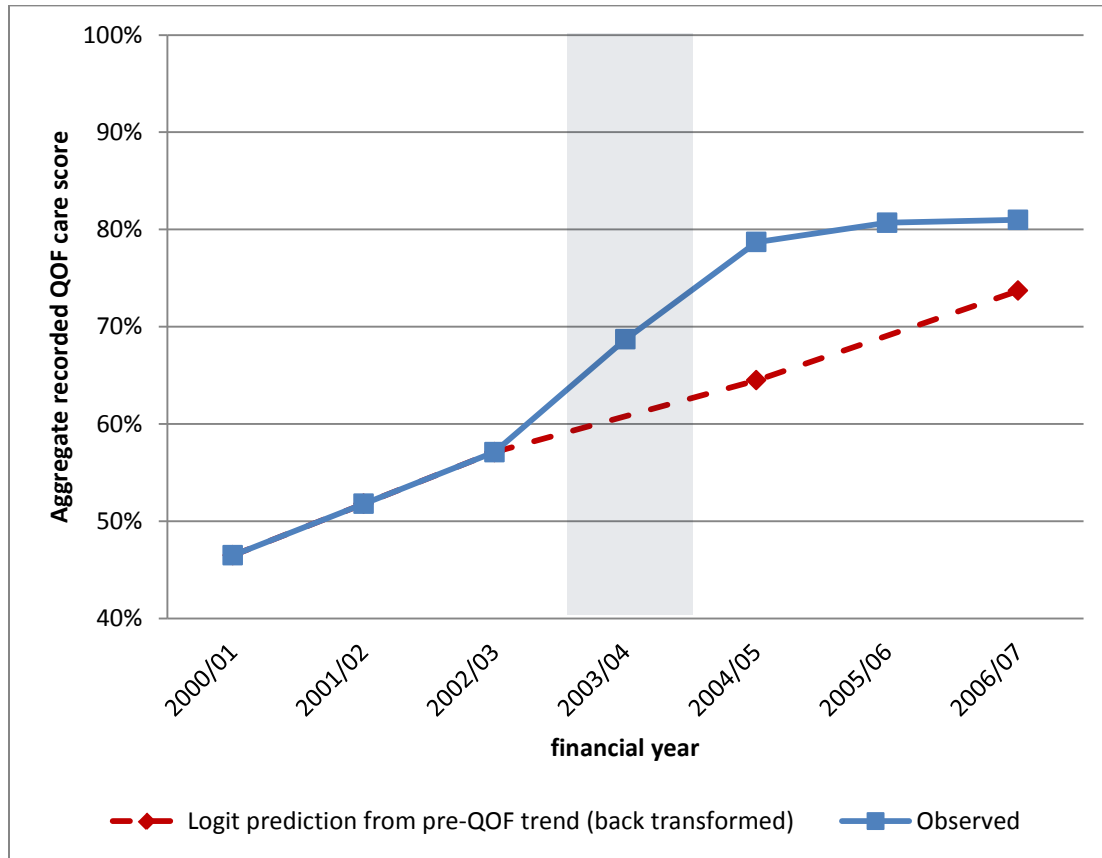


Table 1: Sample characteristics and indicator performance scores across time.

	Mean (sd), except where noted otherwise						
	00/01	01/02	02/03	03/04 ^a	04/05	05/06	06/07
<i>Sample/patient characteristics</i>							
Number of diabetic patients	11,219	12,161	13,306	14,488	15,739	17,083	18,259
Prevalence (%)	3.3	3.6	3.9	4.2	4.5	4.8	5.0
Number of new cases diagnosed	1,242	1,364	1,608	1,645	1,687	1,677	1,649
Incidence (%)	0.38	0.42	0.49	0.49	0.50	0.49	0.47
Female (%)	45.6	45.5	45.2	45.1	44.9	44.9	44.8
Age	62.9 (14.8)	63.0 (14.7)	63.1 (14.6)	63.2 (14.4)	63.4 (14.4)	63.5 (14.5)	63.7 (14.5)
Age at diagnosis	54.9 (17.2)	55.2 (17.0)	55.5 (16.7)	55.6 (16.6)	55.9 (16.5)	56.0 (16.5)	56.0 (16.4)
Years of previous diabetic care	8.0 (8.5)	7.8 (8.5)	7.7 (8.4)	7.6 (8.3)	7.5 (8.2)	7.6 (8.2)	7.7 (8.1)
Number of co-morbid conditions	1.4 (1.4)	1.5 (1.4)	1.5 (1.4)	1.5 (1.3)	1.6 (1.4)	1.6 (1.4)	1.7 (1.5)
<i>Process indicators (time window, months)^b</i>							
DM2 Record of Body Mass Index (15m)	66.6 (47.2)	71.3 (45.2)	74.4 (43.7)	83.6 (37.0)	90.9 (28.8)	92.1 (26.9)	92.6 (26.2)
DM3 Record of smoking status (15m)	35.6 (47.9)	44.1 (49.6)	53.0 (49.9)	78.7 (41.0)	91.3 (28.2)	90.2 (29.7)	91.4 (28.1)
DM4 Smoking cessation advice offered to current smokers (15m)	5.5 (22.7)	9.0 (28.7)	13.4 (34.1)	50.3 (50.0)	82.2 (38.2)	87.1 (33.5)	88.0 (32.5)
DM5 Record of HbA1c or equivalent (15m)	71.1 (45.3)	77.9 (41.5)	82.8 (37.7)	89.2 (31.1)	93.0 (25.5)	93.7 (24.3)	93.5 (24.6)
DM8 Record of retinal screening (15m)	42.2 (49.4)	47.2 (49.9)	55.7 (49.7)	66.3 (47.3)	72.1 (44.8)	71.1 (45.3)	66.9 (47.1)
DM9 Record of presence/absence of peripheral pulses (15m)	40.2 (49.0)	44.5 (49.7)	48.8 (50.0)	62.8 (48.3)	77.6 (41.7)	79.7 (40.3)	78.8 (40.8)
DM10 Record of neuropathy testing (15m)	0.6 (7.8)	0.9 (9.6)	3.6 (18.7)	28.3 (45.0)	63.6 (48.1)	70.9 (45.4)	72.9 (44.4)
DM11 Record of blood pressure (15m)	82.5 (38.0)	86.3 (34.4)	89.3 (30.9)	93.6 (24.5)	96.3 (18.9)	96.7 (17.8)	96.7 (17.8)
DM13 Record of micro-albuminuria testing (15m)	2.2 (14.8)	6.3 (24.3)	13.9 (34.6)	33.4 (47.2)	51.9 (50.0)	56.8 (49.5)	54.5 (49.8)
DM14 Record of serum creatinine testing (15m)	52.2 (50.0)	62.7 (48.4)	71.8 (45.0)	86.5 (34.2)	93.1 (25.3)	94.6 (22.6)	94.5 (22.7)
DM15 Treated with ACE inhibitors or A2 antagonists; patients with proteinuria/micro-albuminuria (15m)	71.5 (45.2)	76.4 (42.5)	78.9 (40.9)	78.4 (41.2)	80.6 (39.6)	81.3 (39.0)	82.2 (38.2)
DM16 Record of total cholesterol (15m)	57.4 (49.5)	66.7 (47.1)	75.8 (42.8)	85.7 (35.0)	92.5 (26.3)	93.7 (24.2)	93.6 (24.5)

	Mean (sd), except where noted otherwise						
	00/01	01/02	02/03	03/04 ^a	04/05	05/06	06/07
DM18	64.6	67.1	67.0	71.5	75.8	79.4	77.6
Influenza immunisation (6m) ^c	(47.8)	(47.0)	(47.0)	(45.2)	(42.8)	(40.5)	(41.7)
<i>Intermediate outcome indicators</i> ^b							
DM6	45.5	48.4	50.2	52.2	55.6	56.4	59.3
Last HbA1C (or equivalent) is $\leq 7.4\%$	(49.8)	(50.0)	(50.0)	(50.0)	(49.7)	(49.6)	(49.1)
DM7	88.5	90.4	90.8	91.8	92.6	92.5	92.7
Last HbA1C (or equivalent) is $\leq 10\%$	(31.9)	(29.4)	(28.9)	(27.4)	(26.3)	(26.3)	(26.0)
DM12	48.9	52.6	56.0	61.5	68.7	71.0	74.6
Last blood pressure $\leq 145/85$ mmHg	(50.0)	(49.9)	(49.6)	(48.7)	(46.4)	(45.4)	(43.5)
DM17	45.3	49.2	56.6	65.1	74.1	78.3	80.8
Last measured total cholesterol ≤ 5 mmol/l	(49.8)	(50.0)	(49.6)	(47.7)	(43.8)	(41.2)	(39.4)

^a Excluded from analyses.

^b Number of patients for which the process has been performed within the specified time window or the outcome was met when last measured, over the number of diabetic patients in the respective financial year.

^c Patient immunised in previous influenza season.

Table 2: Composite performance scores, intervention effects and differences in mean levels of care (overall and by patient & practice characteristics).

		Descriptive statistics						Results of regression analyses ^a								
		Mean (sd)						Intervention effect				Differences in care levels				
		00/01	01/02	02/03	03/04 ^b	04/05	05/06	06/07	Short term (2004/05 score minus prediction from pre-QOF trend)		Long term (2004/05 score minus prediction from pre-QOF trend)		Pre-QOF, 2000/01 to 2002/03		Post-QOF, 2004/05 to 2006/07	
									Coefficient (95% CI) ^c	p-value ^d	Coefficient (95% CI) ^c	p-value ^d	Coefficient (95% CI) ^c	p-value ^d	Coefficient (95% CI) ^c	p-value ^d
Overall		46.5 (23.0)	51.8 (22.5)	57.1 (22.1)	68.7 (21.3)	78.7 (19.3)	80.7 (18.2)	81.0 (17.7)	14.2 (13.7, 14.6)	<0.001	7.3 (6.7, 8.0)	<0.001	-	-	-	-
<i>Patient covariates</i>																
Gender	Male ^e	46.9 (23.5)	52.3 (23.0)	57.8 (22.4)	69.0 (21.4)	79.0 (19.3)	81.0 (18.3)	81.5 (17.6)	-	0.292 ^f	-	0.056 ^f	-	<0.001	-	<0.001
	Female	46.2 (22.4)	51.3 (22.0)	56.3 (21.7)	68.3 (21.2)	78.1 (19.3)	80.3 (18.2)	80.9 (17.1)	0.8 (-0.7, 2.2)		1.7 (-0.05, 3.4)		-1.7 (-2.3, -1.1)		-2.0 (-2.5, -1.5)	
Age	17-29 ^e	33.4 (24.1)	37.5 (24.9)	43.5 (24.7)	54.8 (25.3)	66.4 (24.4)	68.4 (24.3)	68.7 (23.0)	-	0.372	-	0.205	-	<0.001	-	<0.001
	30-64	46.2 (23.0)	51.1 (22.6)	56.4 (22.1)	67.4 (21.5)	77.3 (19.9)	79.6 (18.4)	79.8 (18.0)	1.7 (-1.3, 4.6)		3.0 (-0.5, 6.0)		9.1 (8.2, 9.9)		8.9 (8.2, 9.7)	
	65 or over	49.0 (22.2)	54.5 (21.3)	59.7 (21.0)	71.6 (19.8)	81.2 (17.3)	83.0 (16.5)	83.9 (15.2)	2.1 (-0.9, 4.8)		3.0 (-0.4, 6.0)		11.0 (10.2, 11.7)		11.7 (11.0, 12.4)	
Previous care	0 years ^e (new diagnoses)	44.7 (21.4)	50.4 (20.7)	56.5 (19.3)	65.3 (20.5)	73.4 (19.3)	74.2 (18.7)	74.3 (17.8)	-	<0.001	-	<0.001	-	<0.001	-	<0.001
	1-4 years	48.4 (23.2)	53.9 (22.0)	59.4 (21.7)	71.1 (20.4)	80.9 (18.4)	83.0 (17.1)	83.2 (16.6)	6.2 (3.7, 8.6)		6.0 (3.2, 8.5)		4.7 (4.1, 5.2)		9.1 (8.7, 9.5)	
	5-9 years	46.4 (23.4)	51.9 (23.1)	56.8 (23.2)	69.1 (21.9)	78.7 (20.0)	81.4 (18.2)	81.8 (17.8)	6.4 (3.8, 8.8)		6.1 (3.2, 8.6)		2.7 (2.0, 3.4)		7.8 (7.3, 8.3)	
	10+ years	45.4 (23.0)	50.0 (23.2)	55.1 (22.7)	66.7 (21.9)	77.6 (19.4)	79.3 (18.9)	80.4 (17.4)	6.6 (4.1, 8.9)		7.0 (4.4, 9.3)		1.3 (0.6, 2.1)		6.1 (5.5, 6.7)	
Number of co-morbid conditions	None ^e	42.3 (24.7)	47.1 (24.5)	52.2 (24.6)	62.9 (24.8)	73.5 (23.3)	75.6 (22.7)	76.2 (21.6)	-	0.472	-	0.665	-	<0.001	-	<0.001
	One	46.6 (22.7)	51.8 (22.3)	57.2 (21.9)	68.8 (21.4)	78.6 (19.4)	80.9 (17.9)	80.8 (17.7)	0.5 (-1.4, 2.4)		0.2 (-2.2, 2.4)		4.1 (3.5, 4.7)		3.7 (3.1, 4.3)	
	Two	48.5 (21.8)	53.7 (21.4)	59.2 (20.2)	70.8 (18.7)	80.6 (16.9)	82.2 (15.9)	82.7 (15.0)	0.9 (-1.3, 2.9)		-0.3 (-3.0, 2.2)		6.2 (5.6, 6.9)		5.4 (4.8, 6.0)	
	Three or more	51.2 (20.8)	56.6 (20.0)	61.3 (19.8)	73.0 (17.5)	82.1 (15.2)	83.9 (14.4)	84.4 (13.7)	1.9 (-0.5, 4.2)		1.6 (-1.4, 4.3)		6.3 (5.5, 7.1)		6.1 (5.4, 6.8)	

		Descriptive statistics							Results of regression analyses ^a							
		Mean (sd)							Intervention effect				Differences in care levels			
		00/01	01/02	02/03	03/04 ^b	04/05	05/06	06/07	Short term (2004/05 score minus prediction from pre-QOF trend)		Long term (2004/05 score minus prediction from pre-QOF trend)		Pre-QOF, 2000/01 to 2002/03		Post-QOF, 2004/05 to 2006/07	
									Coefficient (95% CI) ^c	p-value ^d	Coefficient (95% CI) ^c	p-value ^d	Coefficient (95% CI) ^c	p-value ^d	Coefficient (95% CI) ^c	p-value ^d
<i>Practice covariates</i>																
Index of Multiple Deprivation score	1 st quartile ^e	47.5 (22.5)	52.8 (21.4)	57.4 (21.2)	68.4 (20.4)	79.3 (18.8)	81.4 (17.5)	81.8 (17.3)	-	<0.001	-	0.002		0.121		0.266
	2 nd quartile	43.2 (22.3)	48.7 (22.3)	54.4 (22.0)	67.6 (21.2)	78.8 (18.7)	80.7 (17.9)	81.3 (17.5)	1.8 (-0.3, 3.8)		1.1 (-1.5, 3.4)		-3.3 (-8.0, 1.0)		-2.3 (-6.2, 1.4)	
	3 rd quartile	46.5 (24.0)	51.1 (23.3)	56.5 (23.4)	68.2 (22.6)	78.0 (20.5)	79.9 (19.2)	80.3 (17.8)	-0.8 (-3.0, 1.3)		-1.6 (-4.4, 1.0)		-2.6 (-7.2, 1.6)		-3.8 (-7.8, -0.02)	
	4 th quartile (most deprived)	49.0 (22.8)	54.8 (22.8)	60.4 (21.4)	70.6 (20.9)	78.3 (19.2)	80.7 (18.2)	81.5 (16.9)	-4.9 (-7.2, -2.7)		-3.8 (-6.8, -1.1)		1.4 (-3.0, 5.3)		-2.1 (-6.2, 1.6)	
Diabetes prevalence	1 st quartile ^e	47.8 (23.4)	53.9 (22.2)	59.2 (21.8)	69.0 (21.4)	79.6 (19.0)	80.7 (18.7)	81.0 (17.9)	-	0.004	-	<0.001		<0.001		<0.001
	2 nd quartile	44.9 (23.4)	49.0 (23.0)	54.1 (22.4)	68.0 (21.0)	78.2 (19.2)	80.5 (18.2)	81.1 (17.3)	1.4 (-0.7, 3.5)		3.2 (0.7, 5.4)		-1.8 (-2.9, -0.8)		-1.1 (-2.0, -0.3)	
	3 rd quartile	48.7 (21.6)	53.4 (21.1)	57.2 (20.7)	69.2 (21.2)	78.0 (19.3)	81.7 (17.2)	81.9 (16.9)	2.1 (-0.02, 4.1)		4.8 (2.5, 6.8)		-0.02 (-1.2, 1.1)		0.7 (-0.5, 1.7)	
	4 th quartile (highest)	44.9 (23.4)	51.3 (23.4)	58.1 (23.3)	68.5 (21.7)	78.6 (19.7)	79.8 (18.8)	81.0 (17.4)	-1.6 (-3.9, 0.5)		-0.8 (-3.6, 1.7)		-0.9 (-2.4, 0.5)		1.1 (-0.3, 2.5)	

^a 23,780 of the 23,920 identified patients were included in the analyses (for 140 data was only available for 2003/04). Results are back-transformed to percentages from logit scores.

^b Excluded from regression analyses

^c For covariates the value reported is the difference from the reference group coefficient

^d From omnibus post-regression tests when categories are three or more

^e Reference category

^f Significant in the sensitivity analysis with the untransformed scores