Workload and reward in the Quality and Outcomes Framework of the 2004 general practice contract

Bruce Guthrie, Gary McLean and Matt Sutton

ABSTRACT

Background

The Quality and Outcomes Framework (QOF) of the 2004 UK General Medical Services (GMS) contract links up to 20% of practice income to performance measured against 146 quality indicators.

Aim

To examine the distribution of workload and payment in the clinical domains of the QOF, and to compare payment based on true prevalence to the implemented system applying an adjusted prevalence factor. We aimed also to assess the performance of the implemented payment system against its three stated objectives: to reduce variation in payment compared to a system based on true prevalence, to fairly link reward to workload, and finally, to help tackle health inequalities.

Design of study

Retrospective analysis of publicly available QOF data.

Setting

Nine hundred and three GMS general practices in Scotland.

Method

Author query 1; please provide text here for this section

Results

Variation in total clinical QOF payment per 1000 patients registered is significantly reduced compared to a payment system based on true prevalence. Payment is poorly related to workload in terms of the number of patients on the disease register, with up to 44 fold variation in payment per patient on the disease register for practices delivering the same quality of care. Practices serving deprived populations are systematically penalised under the implemented payment system, compared to one based on true prevalence.

Conclusions

The implemented adjustment for prevalence succeeds in its aim of reducing variation in practice income, but at the cost of making the relationship between workload and reward highly inequitable and perpetuating the inverse care law.

Keywords

Author query 2: please provide 5–6 keywords here that are MESH terms

INTRODUCTION

The Quality and Outcomes Framework (QOF) of the 2004 UK General Medical Services (GMS) contract for GPs radically reforms payment systems for UK general practice[1,2] and has been described by a US commentator as 'an initiative to improve the quality of primary care that is the boldest such proposal on this scale ever attempted anywhere in the world'.[1] QOF links up to 25% of practice income to performance measured against 146 clinical and organisational quality indicators.

For each indicator, measured performance is transformed to 'points', with the number of points allocated to each indicator varying according to perceived workload and importance. Of the 1050 points, 550 are allocated to the clinical indicators, and in 2004–2005 each point was worth £75 to an average sized practice (approximately 5400 patients in Scotland) with average prevalence.[3] This represents a major departure from previous incentive schemes in UK primary care which paid lump sums for achieving a small number of quality targets for cervical smears and childhood immunisation.[4,5]

A major source of contention was that the original proposals did not take account of varying prevalence of disease between practices. Instead payment was

Author query 3: please provide author qualifications

***B Guthrie**, health foundation/chief scientist office postdoctoral research fellow, Tayside Centre for General Practice, University of Dundee; **G McLean**, research fellow, Department of General Practice and Primary Care, University of Glasgow; **M Sutton**, professor of health economics, Health Economics Research Unit, University of Aberdeen.*

Address for correspondence
Bruce Guthrie, Tayside Centre for General Practice, MacKenzie Building, Kirsty Semple Way, Dundee, DD2 4BF.
E-mail: b.guthrie@chs.dundee.ac.uk

Submitted: 14 July 2006; **Editor's response:** 14 July 2006; **final acceptance:** 20 September 2006.

adjusted using the global sum weighted allocation formula with an assumption that this would broadly equate with workload. This meant that practices could receive the same payment despite having large differences in the number of patients with each disease. During negotiations immediately before implementation in April 2004, it was decided to adjust for prevalence to better match payment to workload in terms of the number of patients in each practice with a particular disease.

Rather than use true prevalence, an Adjusted Disease Prevalence Factor (ADPF) was implemented, in which the 5% of practices with the lowest prevalence practices were protected by being treated as if their prevalence is higher than it actually is, and the truncated distribution was further narrowed by applying a square root transformation (Box 1).[6] The main aim of the ADPF was to 'reduce variation (in payment) and relatively protect the losers, while at the same time providing fair rewards to those who have the highest prevalence'.[6] Two further justifications were advanced. Firstly, to target resources at areas of high morbidity 'and thereby help tackle health inequalities'. Secondly, it was thought that using true prevalence 'would seriously destabilise those contractors with the lowest relative prevalence' because 'even practices with low prevalence have significant fixed costs in identifying morbidity and establishing quality systems'.[6]

How this fits in

The 2004 General Medical Services (GMS) contract links up to 20–25% of practice income to performance measured against 146 quality measures. The payment system for clinical quality measures used an Adjusted Disease Prevalence Factor rather than true prevalence, with the intention of reducing variation in overall payment compared to true prevalence, while maintaining a fair link between workload and reward, and helping tackle health inequalities. Compared to using true prevalence, the ADPF does reduce variation in total clinical QOF payment, but produces up to 44-fold variation in payment per patient with disease for practices delivering the same level of quality, and systematically penalises more deprived practices. These problems were predictable from data available before implementation. Future radical changes to payment systems should be better modelled before implementation.

Box 1. Calculating payment using implemented adjusted disease prevalence factor[6] and an alternative true disease prevalence factor.

Each indicator numerator/denominator translates into a number of points, centrally set to reflect perceived importance, and the workload involved in implementing them

- Implemented adjusted disease prevalence factor (ADPF)

Calculating the ADPF

a) Truncation: the 5% of practices with the lowest prevalences are assumed to have the same prevalence as the cut-off point (that is, they are treated as having a higher prevalence than they actually do)

b) Square root transformation is applied to truncated prevalence and mean of this calculated

c) Rebasing: each practice's truncated, square root transformed prevalence is divided by the mean of the truncated, square root transformed distribution so that the 'average' practice has an ADPF of 1

Clinical pounds per point 2004–2005 under ADPF

= £75 x ADPF x Relative List size

$$= £75 \times \frac{\text{Practice truncated, square root transformed prevalence}}{\text{Mean of truncated, square root transformed prevalences}} \times \frac{\text{Practice list size}}{\text{Mean list size}}$$

Since practice prevalence = number of patients on disease register/list size, payment under ADPF therefore varies with:

1. The number of patients on the disease register (transformed by truncation and square rooting)

2. Practice list size, which appears twice (square root transformed in ADPF, untransformed in relative list size)

- True disease prevalence factor (TDPF)

Clinical pounds per point 2004/5 under TDPF

= £75 x TDPF x relative list size

$$= £75 \times \frac{\text{Practice prevalence}}{\text{Mean prevalence}} \times \frac{\text{Practice list size}}{\text{Mean list size}}$$

Since practice prevalence = number of patients on disease register/list size, practice list size cancels out and payment under TDPF therefore only varies with register size.

Table 1. Total payment per 1000 patients registered under TDPF and ADPF.

Total payment/1000 patients registered	Mean (SD) £	Range £	Variance	F (*P*-value[a])
Based on actual points achieved (*n* = 785)				
TDPF	6763.33 (1326.02)	1361 to 12020	1758337.03	
ADPF	6653.56 (763.85)	2789 to 9055	583464.38	3.01 (<0.001)
Assuming all practices achieve maximum points (*n* = 781)				
TDPF	7899.44 (1503.43)	1448 to 14759	2260302.35	
ADPF	7761.57 (699.72)	5742 to 10611	489601.55	4.62 (<0.001)

[a]*F-test for null hypothesis of no difference between variances. Degrees of freedom for numerator and denominator = (n of practices in each domain - 1). ADPF = adjusted disease prevalence factor. TDPF = true disease prevalence factor.*

The method for calculating the ADPF is the same in all four UK countries, and is shown in Box 1. Two features are particularly notable. Firstly, using the ADPF, payment is related to the number of patients with the disease in each practice in the way intended. However, for practices with the same number of patients on the register, payment additionally increases with list size. Secondly, because prevalence and list size are systematically related to the deprivation of the population served, using ADPF rather than true prevalence is likely to have systematic effects on the distribution of resources between practices serving different populations.

Box 1 also describes a payment calculation using a True Disease Prevalence Factor (TDPF), which is the system rejected in favour of ADPF. Under TDPF, payment per point only depends on the number of patients a practice has on its disease registers.

In this paper, we compare payment based on ADPF with the rejected system based on a True Disease Prevalence Factor (TDPF), to examine whether the implemented ADPF succeeds in its stated aims of:

- reducing variation in overall payment compared to TDPF;
- maintaining a fair link between payment and workload; and
- helping tackle health inequalities.

METHOD

Prevalence and list size data for Scottish practices contracted under GMS were extracted from publicly available sources.[7] All analyses were conducted in SPSS version 10. For each practice, pounds earned per point were calculated from reported prevalences using the Adjusted Disease Prevalence Factor (ADPF)[6] and True Disease Prevalence Factor (TDPF) described in Box 1. We recalculated ADPFs rather than using reported ADPFs, because calculations depend on national values for truncation and mean prevalence. Actual payment for 2004–2005 depended on reported prevalences on 14 February, but these were incomplete. The prevalences reported on 31 March include practices which had not submitted data on 14 February, and are therefore a better guide to future performance of the payment system. For each practice, gains and losses from using ADPF compared to TDPF were calculated.

Variation in total payment

Total payment per 1000 patients registered was calculated for both TDPF and ADPF, using both actual points achieved and maximum possible points. Differences in variances of payment under different assumptions were examined with an F-test.

The relationship between workload and payment

For each of the ten clinical domains, the relationship between payment and workload was examined by calculating pounds earned per point achieved per patient on the disease register.

Systematic effects of using ADPF instead of TDPF

Systematic effects were examined by one-way analysis of variance (ANOVA) of mean potential gains/losses by deprivation decile across all clinical domains assuming that all practices achieved maximum points. Practice deprivation was defined in terms the income domain of the Scottish Index of Multiple Deprivation (SIMD) 2004, calculated on the basis of the geographical distribution of the place of residence of their registered populations.[8] We chose the income domain rather than the overall score or other domains, because it receives the highest weight in the calculation of the overall index and is highly correlated with it, is available for small geographical areas representing between 500 and 1000 people, and does not undergo transformations making it easily interpretable.

RESULTS

Data was complete for 813 to 903 practices depending on clinical domain (to protect patient confidentiality, full data is not published where there are five or fewer patients per indicator or register). Mean income deprivation was not significantly different for practices with and without complete data (mean total SIMD income score 15.87 versus 16.11 respectively, difference in means 0.25 (95% confidence interval [CI] = 1.32 to 1.82), $t_{903df} = 0.308$, $P = 0.758$).

Variation in payment

Table 1 shows income per 1000 registered patients for the QOF clinical domains. As intended, the distribution under implemented ADPF is considerably narrower than under TDPF.

Relationship between pay and workload

Under TDPF, each practice earns the same income per point achieved per patient on the register. However, there is considerable variation in pounds earned per point achieved per patient on a disease register under the implemented ADPF (Table 2). At the extreme of the distribution, there is up to 44 fold variation in payment per patient for the same level of achievement. Excluding the 10% of outlying practices by comparing the 95th and 5th centile practices, there is still 1.5 to 2.7 fold variation in payment per patient for the same level of achievement.

Systematic effects of using ADPF instead of TDPF

Supplementary Table 1 shows the relationship between income deprivation, list size and prevalence for the 10 disease areas. Total financial gains and losses across all 10 clinical domains are shown in Table 3, assuming that all practices achieved maximum points. Similar patterns are found using actual points achievement on 31 March 2004 (data not shown). The means differ significantly with the five most affluent decile gaining and the five most deprived decile losing from using ADPF instead of TDPF. The total mean difference between the most affluent and least affluent decile is £6012 per practice (95% CI = £4633 to £7392, post-hoc least significant difference $P<0.000$). The mean gain in the most affluent decile is approximately 5% of income from the QOF clinical domains, and the mean loss in the least affluent decile approximately 2.5%. However, there are winners and losers in every decile of deprivation with 22% of practices gaining or losing more than 5% of their QOF income, and 4.3% of practices gaining or losing more than 10%.

DISCUSSION

The ADPF succeeds in the first of its aims by reducing variation in total practice income. This is achieved at the cost of up to 44-fold variation in payment for practices treating the same number of patients to the same level of quality. Even excluding the outlying 10% of practices, variation is approximately twofold, and we judge that the second aim of the ADPF of fair pay for workload is therefore not achieved. Finally, the ADPF will not help tackle health inequalities in Scotland since it institutionalises the inverse care law where resource distribution favours the affluent despite systematic chronic disease care being harder to implement for more deprived populations.[9,10] The implementation of the ADPF therefore does not deliver two of its three stated objectives. It effectively prioritises equality of total practice income, over ensuring that resources are proportional to chronic care work actually done.

The potential financial impact of the inconsistent link between workload and reward can be illustrated

Table 2. Variation in payment per point per patient with disease under ADPF.[a]

Clinical domain	(*n* of practices with complete data)	ADPF mean (£ per point per patient on register)	ADPF range[a] (£ per point per patient on register)	Ratio max/min	ADPF 5th to 95th centile (£ per point per patient on register)	Ratio 95th/5th centile
Coronary heart disease	902	0.33	0.19 to 8.31	43.7	0.25 to 0.40	1.6
Asthma	903	0.27	0.17 to 1.30	7.6	0.22 to 0.33	1.5
Cancer	843	2.96	1.28 to 16.29	12.7	2.11 to 4.18	2.0
COPD	899	0.85	0.37 to 8.80	23.8	0.52 to 1.27	2.4
Diabetes	902	0.44	0.28 to 2.52	9.0	0.36 to 0.53	1.5
Epilepsy	871	2.08	1.09 to 9.73	8.9	1.60 to 2.70	1.7
Hypertension	904	0.13	0.08 to 1.71	21.4	0.10 to 0.16	1.6
Mental health	813	2.95	1.18 to 17.17	14.6	1.75 to 4.64	2.7
Stroke	898	0.85	0.43 to 19.05	44.3	0.62 to 1.14	1.8
Hypothyroidism	901	0.54	0.28 to 3.90	13.9	0.39 to 0.71	1.8

[a]*Under TDPF, payment per point achieved per patient is the same in all practices. ADPF = adjusted disease prevalence factor. COPD = chronic obstructive pulmonary disease.*

Table 3. Variation in payment per point per patient with disease under ADPF.[a]

Deprivation decile[a]	Number of practices	Mean SIMD income score[a]	Mean gain or loss (£)	95% CI of mean (£)
1 (affluent)	77	5.3	3561	2483 to 4637
2	72	8.1	1129	-232 to 2491
3	77	9.8	1012	242 to 1784
4	83	11.4	388	-834 to 1610
5	81	13.2	182	-786 to 1150
6	81	15.5	-1538	-2489 to -587
7	79	17.9	-1641	-2310 to -973
8	80	20.2	-1037	-1906 to -169
9	75	23.8	-1311	-2224 to -400
10 (deprived)	75	33.9	-2452	-3223 to -1681
ANOVA[b] F = 12.931, 9 degrees of freedom, $P<0.001$				

[a]Each one point increased in SIMD income score represents a 1% increase in the percentage of the practice population in the receipt of benefits on the grounds of low income. [b]Null hypothesis of no difference in means. ADPF = adjusted disease prevalence factor. SIMD = Scottish Index of Multiple Deprivation.

by two practices, both of which have 30 patients on their coronary heart disease (CHD) register. One has a list size of 560, a higher than average prevalence (5.4%), and achieves 101 points in the CHD domain. The other has a list size of 23 324 with a very low prevalence (0.13%, which is plausible since it almost exclusively serves university students), and achieves 100.5 points. Both practices therefore have similar fixed costs in running a register, have the same number of patients to care for, and deliver CHD care of very similar overall quality. Under TDPF they would earn the same. Under ADPF, the smaller practice is (modestly) penalised by square root transformation and is paid £850. The larger practice benefits from truncation, square root transformation and inflation by relative list-size and is paid £25 063. This 29-fold variation is not the most extreme in the CHD domain, and similar examples can be found in other disease areas and for other practices.

This also highlights the inconsistencies of truncating prevalence, where the rationale is that there are fixed costs to running register and recall systems that fall as equally on practices with few patients as those with many. However, truncating prevalence fails to deliver benefits to most practices with small registers. The largest CHD register in a practice benefiting from truncation was 243 (prevalence 2.65%, list size 9154). Fixed costs will be a greater proportion of total costs in practices with <243 patients with CHD. Four hundred and sixty-three (51.3%) practices had <243 patients on their CHD register, but because their prevalence was above the truncation level, did not receive any recognition of their relatively greater fixed costs. To our knowledge there are no estimates of how large the fixed costs of running a register are, but we believe that they are likely to be small compared to non-fixed costs directly related to register size (regularly checking diagnoses and data completeness, writing letters, seeing patients in clinic). If fixed costs can be shown to be important, they would be better addressed by paying a fixed amount for registration and recall in each disease area independent of register size, or applying a truncation to register size rather than prevalence.

A further rationale for using the ADPF was to avoid financially destabilising practices. However, the initial risk of this was minimal, since existing practice income was largely guaranteed under the Minimum Practice Income Guarantee and QOF money represented new resources. Amending the payment system now would carry a risk of financial destabilisation of some practices, and any changes might therefore require transitional arrangements for practices with very large potential losses.

Our conclusion is that the ADPF makes the relationship between workload and reward in Scotland significantly variable, and in our opinion, inequitable. Additionally, it helps perpetuate the inverse care law. Although the analysis has used Scottish data, the results on variation in payment and the relationship between workload and reward are generalisable to the rest of the UK because ADPF and payment calculations are the same. However, whether resources will be systematically redistributed away from practices serving more deprived populations will depend on the relationships between prevalence, list size, and deprivation elsewhere in the UK. We believe that a system such as TDPF, where payment varies only with register size and points achieved, would be both more transparent and fairer.

More generally, the study highlights the importance of carefully modelling the effects of complex payment systems before implementation. The distributions of true prevalence and the ADPF were compared before implementation.[6] However, since ADPF implementation fails to deliver on two of its explicit aims, the overall effects of adding a prevalence adjustment to a payment system that already included a list size adjustment do not appear to have been adequately modelled. This probably reflects the fact that prevalence adjustment was rapidly developed as a modification of an already complex contract in the period between the contract being agreed, and its implementation in April 2004.[6] Similar problems arose with the implementation of the weighted allocation formula (the 'global sum') that determines the majority of practice income.[6,11] In both cases, more comprehensive modelling before implementation using existing data would have identified major, soluble problems. Radical change to payment systems will always risk perverse and unintended consequences, but at least some of

these can and should be avoided in the future by more rigorous testing before implementation.

Supplementary information

Additional information accompanies this article at http://www.rcgp.org.uk/bjgp-supp-info

Funding body

At the time this work was undertaken Gary McLean was funded by the Platform Project. Matt Sutton is funded by the Chief Scientist Office of the Scottish Executive Health Department. Bruce Guthrie was funded by the Health Foundation and the Chief Scientist Office of the Scottish Executive Health Department. The authors accept full responsibility for the views expressed

Ethics committee

The analysis used publicly available data, and no approval was required

Competing interests

The authors have stated that there are none. During the course of the study, Bruce Guthrie worked as a GP in a practice which would gain approximately £1300 from a change to a true prevalence adjustment, but was a full-time university employee with no financial stake in the practice

Acknowledgements

AQ4 - Author: would you like to include acknowledgements to others involved in the study who were not authors?

REFERENCES

1. Shekelle P. New contract for general practitioners: a bold initiative to improve quality of care, but implementation will be difficult. *BMJ* 2003; **326:** 457–458.
2. Marshall M, Harrison S. It's about more than money: financial incentives and internal motivation. *Qual Safe Health Care* 2005; **14:** 4–5.
3. NHS Confederation and British Medical Association. *Investing in general practice: the new GMS contract.* London: British Medical Association, 2003.
4. Giufrida, A, Gosden, I, Forland, F, *et al.* Target payments in primary care: effects on professional practice and health care. outcomes *Cochrane Database Syst Rev* 1999: 4: CD000531.
5. Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. *JAMA* 2005; **294(14):** 1788–1793.
6. Department of Health. *Delivering investment in general practice: implementing the new GMS contract.* London: Department of Health, 2003.
7. Scottish Health Statistics. *Quality and Outcomes Framework — Information*. http://www.isdscotland.org/isd/3313.html (accessed 6 Oct 2006).
8. Office of the Chief Statistician. *Scottish Index of Multiple Deprivation 2004 Summary Technical Report.* Edinburgh: Scottish Executive, 2004.
9. Hart JT. The inverse care law. *Lancet* 1971; **1:** 405–412.
10. Watt G. The inverse care law today. *Lancet* 2002; **360:** 252–254.
11. Chisholm J, Farrar M. *Joint Letter on the Minimum Practice Income Guarantee (MPIG).* 2003. London, British Medical Association and NHS Confederation. **Author query 5: I cannot find this reference - please check that all details are correct.**