Quality measurement of care for people with type 2 diabetes in Tayside, Scotland: implications for the new UK general practice contract

Bruce Guthrie, Alistair Emslie-Smith, Andrew Morris, Tom Fahey, Frank Sullivan

SUMMARY

Background: *The new United Kingdom general practice contract proposes that up to a third of general practitioners' income will come from achieving quality targets.*

Aim: *To examine selected quality markers in terms of their robustness to case-mix variation and chance effects, and in the attribution of quality to practices.*

Study design and methods: *Data were extracted from a population-based diabetes clinical information system in Tayside, Scotland, for patients with type 2 diabetes registered in 67 practices with complete ascertainment.*

Results: *Most practices would have received relatively high levels of payment for the process measures examined. Outcome measures appeared more challenging. Case-mix adjustment for age, sex, and postcode-assigned deprivation altered measured performance by up to 7%, but payment by up to 14%. Despite no strong evidence of any real difference in quality, chance effects meant that there was greater apparent variability for smaller practices from year to year. Hospital attendance was common, but highly variable between practices.*

Conclusion: *Case-mix adjustment to allow fairer comparison is routine in national performance indicators, and ignoring it risks making the new contract quality framework inequitable. Because of chance effects, smaller practices may have greater year-to-year variability in income. Reflecting National Health Service structure, the new contract provides no incentives for integrated care and offers a perverse incentive to refer more patients to hospital. There are trade-offs between the validity of measures, and the cost and bureaucracy of collecting data. The planned evaluation of the new contract should examine the effectiveness and equity of the quality framework, and rapidly act on deficiencies found.*

Keywords: *diabetes mellitus; family practice; healthcare quality assurance; healthcare quality indicators.*

B Guthrie MCRGP MSc PhD, general practitioner, Department of Community Health Sciences, General Practice, University of Edinburgh. A Emslie-Smith MCRGP, general practitioner, Mill Practice, Arthurstone Medical Centre, Dundee. A Morris FRCP MSc MD, professor of diabetic medicine, Department of Medicine, Ninewells Hospital, University of Dundee. T Fahey MCRGP MSc, professor of primary care medicine; F Sullivan FRCGP FRCP PhD, professor of research and development in general practice and primary care, Tayside Centre for General Practice, Ninewells Hospital, University of Dundee.

Address for correspondence:

Bruce Guthrie, Department of Community Health Sciences, General Practice, University of Edinburgh, 20 West Richmond Street, Edinburgh EH8 9DX. E-mail b.guthrie@ed.ac.uk

Submitted: 17 February 2003; Editor's response: 28 April 2003; final acceptance: 23 July 2003.

Introduction

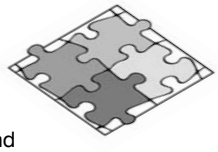
PAYMENTS for achieving quality targets for chronic disease management are a central part of the proposed new United Kingdom (UK) general practice contract, and will account for up to a third of practice outcome.[1] When calculating the level of quality that has been achieved in terms of the new contract, practices can exclude patients who do not attend review, decline particular treatments, or who have contraindications to treatment.

A core principle of the new contract is that practices should be fairly rewarded for the work they do to achieve particular levels of quality of care.[1] Achieving equitable payment partly depends on the robustness of the underlying system of quality measurement. There are several criteria by which a quality measure can be judged, although different stakeholders disagree about the purpose and methodology of measurement.[2,3] Ideally, an incentivised quality measure should be robust to differences in case mix between practices and to chance variation, and any changes in measured quality should be clearly attributable to practice work. It should provide timely and usable data to influence clinical practice and organisation, avoid perverse incentives, and be resistant to gaming or manipulation to maximise payment without improving patient care.[2,3]

In any quality measurement system linked to payment there are likely to be trade-offs between these different criteria of a 'good' quality measure, and almost inevitably any system will offer perverse incentives and have unintended consequences.[2-5] From this perspective, all quality measures are only proxies for true quality of care, and no system of measurement can be perfect. However, an understanding of the strengths and weaknesses of a particular incentive system can help to improve its effectiveness and equity.

National performance indicators routinely use standardisation or case-mix adjustment to ensure fairer comparison, and ignoring case mix risks inequitable distribution of resources.[6,7] Although the Carr-Hill allocation formula is intended to adjust the global sum or non-quality-linked practice payment for differences in workload between practices serving different populations, there is no case-mix adjustment in the final version of the quality-linked payments. Allowing exclusions may potentially mitigate this, since, for example, it might be expected that patients living in more deprived areas will be more likely to not attend for review or to have significant co-morbidity precluding some forms of care.

HOW THIS FITS IN



What do we know?

The new UK general practice contract includes an incentive mechanism to promote better quality of clinical care and practice organisation, but the effectiveness and equity of such a complex system is difficult to predict.

What does this paper add?

'Measured quality' of diabetes care and payment under the new GP contract are likely to be significantly affected by variations in case-mix, and chance effects may lead to significantly variable year-to-year income for smaller practices.

Despite care commonly being shared between primary and secondary care, the new contract provides no incentive to promote integrated care across boundaries, and potentially offers a perverse incentive to increase hospital referral for patients with diabetes.

The effectiveness and equity of the new contract quality framework should be evaluated using anonymised individual data from a range of practices. The contract should be flexible enough to accommodate changes to correct whatever deficiencies emerge.

Like all quantitative measures, clinical indicators are subject to chance variation. Smaller practices are particularly affected because they have fewer patients. For example, consider a small practice that runs one diabetic clinic every six weeks. 'Measured quality', in terms of process measures such as cholesterol testing within a defined period, will be lower the day before a clinic than the day after, although there is no real difference in underlying performance between these two days. 'Measured quality' will therefore partly depend on the arbitrary day that quality measurement is carried out. This will be less important in larger practices running weekly or twice-weekly clinics. One consequence of this and other chance variation is that, in any one year, the confidence intervals for quality measures will be wider in smaller practices. A less obvious consequence of chance effects is that 'measured quality' in smaller practices is likely to be more variable from year to year.

The aim of this paper is to examine how the quality measures for chronic disease management in the new contract might work when applied to care for people with type 2 diabetes. It takes for granted that the quality measures examined are important, timely, usable and responsive to practice intervention.[8-10] It examines the effects on 'measured quality' of adjusting for case mix, the impact of chance variation on smaller practices, the attribution of quality to individual practices, and discusses some possible unintended consequences and perverse incentives evident from these results.

Methods

Ethical approval was obtained from the Tayside Research Ethics Committee to extract anonymised data from the DARTS diabetes clinical information system in the Tayside region of Scotland.[11] DARTS is a comprehensive, population-based diabetes database with regular manual validation of primary care data, and automatic linkage to Tayside laboratory databases. Having excluded five practices where ascertainment was incomplete due to the referral of some patients outside the region, we created a dataset that consisted of all people with type 2 diabetes registered with 67 practices in Tayside. For these patients, complete laboratory, demographic (age, sex, postcode), and hospital attendance data for the years 1999–2001 was available.

Data were analysed using the Statistical Package for the Social Sciences version 11. Individual patient data were aggregated to practice level to calculate quality measures. For each practice, two measures of process were calculated (the percentages of patients with glycated haemoglobin and cholesterol measured in 2001), and two measures of intermediate outcome (the percentages of all patients with glycated haemoglobin measured in 2001 and result $\leq$7.4%, and cholesterol measured in 2001 and result $\leq$5 mmol/l). Case-mix adjusted measures were calculated by standardisation using predictions from a fully saturated logistic regression model for all patients with type 2 diabetes in Tayside (equivalent to indirect standardisation).[12]

Results

The Tayside prevalence of type 2 diabetes was 2.25% in 2001. For 2001, we studied 9,064 people with type 2 diabetes with a mean age of 66.4 years (standard deviation [SD] 12.6) and a mean duration of diabetes of 7.6 years (SD 6.7), of whom 52.7% were male. Table 1 shows the means and ranges for the four quality measures examined. Most practices would have received high levels of payment for the process measures. Many would also have received high payments for outcome measures, although these appeared more challenging, and a few practices would have received no payment for these.

Effects of adjusting for case mix

Unadjusted quality measures were compared with case-mix adjusted measures after standardisation for patient age, sex, and postcode-derived deprivation score. The maximum differences between crude and case-mix adjusted measures were 2.6% for the percentage of patients with glycated haemoglobin measured in the previous 12 months, 6.9% for cholesterol testing, 3.4% for the percentage achieving glycated haemoglobin $\leq$7.4%, and 4.7% for achieving cholesterol $\leq$5 mmol/l. However, because quality payments increase on a sliding scale within the ranges shown in Table 1,[1] the maximum changes in measured quality in this study equate to maximum changes in quality points achieved and therefore payment received of 4%, 10.6%, 13.6%, and 13.4% for glycated haemoglobin measurement, cholesterol measurement, glycated haemoglobin $\leq$7.4%, and cholesterol $\leq$5 mmol/l respectively.

Effects of chance variation on small practices

There was no evidence that smaller practices provided a lower quality of care than larger practices for any of the measures examined. For example, there were no significant differences between practices in the largest and smallest quartiles of list size in terms of the percentage of patients achieving a glycated haemoglobin $\leq$7.4% in any year from

Table 1. Distribution of quality measures in 67 Tayside practices in 2001.

Quality measure	Mean (%) (SD)	Range (%)	Interquartile range (%)	Range within which quality payment increases[a] (%)
Percentage of patients with glycated haemoglobin measurement in 2001	88.1 (6.1)	61.1–97.3	85.7–91.8	25–90
Percentage of patients with cholesterol measurement in 2001	77.8 (7.9)	61.9–97.3	72.5–83.2	25–90
Percentage of all patients with glycated haemoglobin measured in 2001 and result ≤7.4%	42.5 (8.5)	20.4–66.7	37.3–48.0	25–50
Percentage of all patients with cholesterol measured in 2001 and result ≤5 mmol/l	42.4 (7.9)	22.2–64.0	37.3–45.8	25–60

[a]Practices receive no payment for <25% of patients having a test done or achieving a target, and payment increases on a sliding scale up to the maximum shown, with no further increase thereafter.[1]

Table 2. Quality measurement in smaller and larger practices.

	Mean percentage of patients with glycated haemoglobin ≤7.4% (95% CI)[a]		
Year	Practices in smallest quartile of list size <4500 (n = 16)	Practices in largest quartile of list size >12000 (n = 17)	Mean difference (95% CI)[b]
1999	41.4 (38.5–44.2)	37.9 (35.3–40.5)	3.5 (-0.5–7.4) P = 0.086
2000	33.3 (29.3–37.4)	32.9 (30.6–35.1)	0.5 (-4.4–5.3) P = 0.847
2001	43.8 (38.3–49.3)	41.8 (39.2–44.4)	2.0 (-4.5–8.4) P = 0.518

[a]DCCT standardised values for all years. [b]t-test for equality of means

1999–2001 (Table 2). Despite this evidence of small practices performing as well or better than large practices in achieving glycaemic control targets, the measured quality of smaller practices was more variable than that of larger practices within each year (indicated by the wider confidence intervals [CI] in the second and third columns of Table 2). More importantly, due to chance effects, there were large changes in 'measured quality' from year to year for many smaller practices (Figures 1 and 2).

Attribution of care to practices

In 2001, 49.1% of patients had attended hospital, but there was striking variation between practices with a range from 6% to 88.9% of patients in each practice attending. There was no significant correlation between the proportion of patients in each practice attending hospital and any of the quality measures examined (Pearson correlation coefficients for proportion of patients in each practice attending hospital, and proportion with glycated haemoglobin measured -0.064, P = 0.605; proportion with cholesterol measured -0.132, P = 0.285; proportion with glycated haemoglobin ≤7.4% 0.020, P = 0.875; proportion with cholesterol ≤5 mmol/l 0.029, P = 0.815).

Discussion

Summary of main findings

Even using a 12-month measuring period, rather than the new contract's 15-month measuring period, and without accounting for exceptions, most practices in Tayside would have received high levels of quality payments for the process measures examined in this population-based analysis. Achieving high-quality outcomes appeared more challenging, although many practices would still have received high levels of payment.

Case-mix adjustment had only modest effects on measured quality, but because of the way that payment is structured, these equate to larger potential changes in practice income. Tayside has few areas of maximal deprivation, and few residents from ethnic minorities (not adjusted for here). Case-mix effects may therefore be more significant for areas with more variable patient populations.

Even for a common disease like diabetes, smaller practices had more variation in apparent performance, despite providing similar levels of quality of care as larger practices. Such chance-driven variation will be greater for less common diseases included in the quality framework, such as epilepsy. If this leads to significant year-to-year variation in income, it will make long-term planning more difficult in smaller practices.

Care was commonly shared with hospital clinics, although there was marked variation between practices. The new general practice contract reflects the rigid financial division between primary and hospital care in the National Health Service (NHS), and therefore ignores the interdependence of primary and secondary care. It provides no incentives for the development of integrated care or managed clinical networks across existing boundaries, even though other NHS policy casts these as essential to improve quality.[9,13] At practice level there was no clear relationship between the proportion of patients attending hospital and performance on any of the measures used here. Although such an ecological analysis of limited markers of quality should be treated with caution, this is consistent

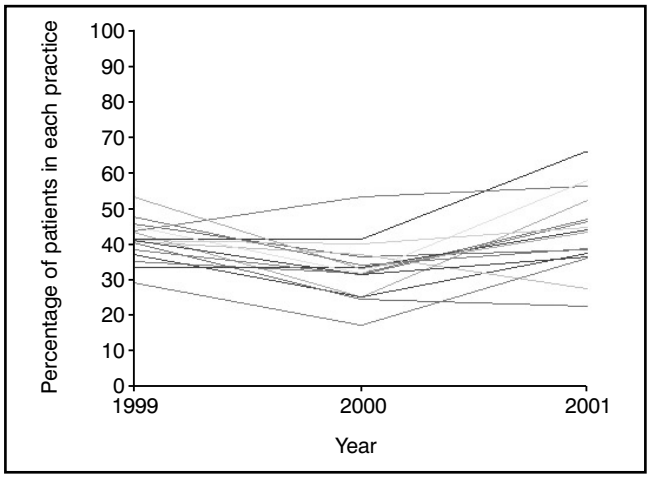


Figure 1. Percentage of patients in practices in the smallest quartile of list size (<4500) with last recorded glycated haemoglobin ≤7.4% in each year 1999–2001 (each line represents one practice).

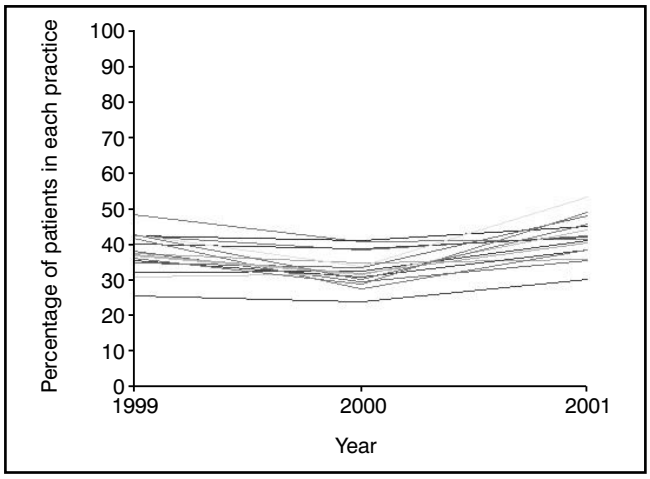


Figure 2. Percentage of patients in practices in the largest quartile of list size (>12,000) with last recorded glycated haemoglobin ≤7.4% in each year 1999–2001 (each line represents one practice).

with clinical trial evidence of the equivalence of appropriately structured general practice and hospital care for diabetes.[14]

Strengths and limitations of the study

The key strength of this study is that complete data were available on all patients with type 2 diabetes in the 67 practices studied. Potential limitations of this analysis are that Tayside is unusual in having an established diabetes managed clinical network, and only a small number of the proposed measures have been examined. Additionally, data about exclusions from measurement were not available. However, we believe the findings have wider relevance.

Implications for future research and policy

Although case-mix adjusted measures are generally accepted as necessary to allow fair comparison, and therefore fair remuneration,[6,7] their calculation would require practices to submit patient-level data to a central analyst. This would be likely to make the new contract quality system more opaque, less timely, less usable, more bureaucratic, and more expensive. A potentially simpler alternative is to use case-mix adjustment for payment by varying the payment for achieving particular levels of quality depending on the case mix of patients registered with practices. The original version of the contract proposed doing this using the Carr-Hill weighted allocation formula. However, since this is derived from measurement of current workload, it is unlikely to be adequate for the purpose of weighting chronic-disease management work, and robust alternatives may have to be developed. Deciding if reducing any inequities resulting from case-mix variation are worth the additional costs of adjusting either quality measures or payments will require examination of data from many areas.

Others have also argued that case-mix adjustment risks institutionalising variations in care and excludes the most vulnerable from the potential benefits of providing feedback on quality of care to the clinicians responsible for their care.[15] In part, this suggests that any disease-based resource allocation system is unlikely to be well suited to addressing the interacting problems of severe social deprivation.[16] Such considerations also apply to allowing practices to exclude particular patients from quality measurement, and emphasise that there may be necessary trade-offs between different criteria for judging the worth of a quality indicator.

From this analysis, the most obvious perverse incentive is for practices to refer more patients to hospital. This will maximise income for minimal extra work. The experience of fund-holding suggests that this kind of gaming will be attractive to at least some.[17] A second perverse incentive is for practices to focus effort on patients whose outcomes are already near to a treatment threshold, since achieving a glycated haemoglobin of ≤7.4% is likely to be easier for a patient with a current result of 7.5% than one with a higher result. This may improve 'measured quality', but potentially at the cost of less attention being paid to those with greater need for, and greater potential benefit from treatment. Finally, the most serious unintended consequence may be that other important areas of practice that are less amenable to measurement are relatively ignored, or that the system undermines public and professional trust in the NHS.[5,18-19]

We believe that the quality framework of the new contract is a reasoned and careful attempt to achieve multiple aims, but the effectiveness and equity of such a complex system cannot be easily predicted. The plan to evaluate the new contract is therefore to be welcomed.[1] Based on these results, such an evaluation should include an examination of the effects on resource allocation and equity of different ways of handling case mix, including how allowing exceptions to measurement actually works in practice. Additionally, it should examine the magnitude of chance-driven variation in small practices' income, and actively search for the undesirable effects of perverse incentives. This will require a reasonably large, representative sample of practices submitting anonymised, patient-level data for analysis. The contract must then be flexible enough to allow rapid change to correct whatever deficiencies emerge.

References

1. General Practitioners Committee of the British Medical Association. *Investing in general practice: the new general medical services contract.* London: British Medical Association, 2003.
2. British Medical Association Board of Science and Education. *Clinical indicators.* London: British Medical Association, 2000.
3. Pringle M, Wilson T, Grol R. Measuring 'goodness' in individuals and healthcare systems. *BMJ* 2002; **325:** 704-707.
4. Hofer TP, Hayward RA, Greenfield S, *et al*. The unreliability of individual physician 'report cards' for assessing the costs and quality of care of a chronic disease. *JAMA* 1999; **281:** 2098-2105.
5. Marshall M, Davies H. Public release of information on quality of care: how are health services and the public expected to respond? *Journal of Health Services & Research Policy* 2001; **6:** 158-162.
6. NHS Executive. *NHS performance indicators July 2000, Annex A standardisation methodology*. http://www.doh.gov.uk/nhsperformanceindicators/hlpi2000/tech_stand.html (accessed July 2003).
7. Greenfield S, Kaplan SH, Kahn R, *et al*. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. *Ann Intern Med* 2002; **136:** 111-121.
8. Clinical Standards Board for Scotland. *Clinical standards: diabetes.* Edinburgh: Clinical Standards Board for Scotland, 2001.
9. Department of Health. *National service framework for diabetes.* London: Department of Health, 2002.
10. Renders CM, Valk GD, Griffin S, *et al*. Interventions to improve the management of diabetes in primary care, outpatient, and community settings: a systematic review. *Diabetes Care* 2001; **24:** 1821-1833.
11. Morris AD, Boyle DIR, MacAlpine R, *et al*. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. *BMJ* 1997; **315:** 524-528.
12. Kendrick S, MacLeod M. *Clinical indicators support team (CIST) working paper no. 3. Adjusting outcomes for case mix: indirect standardisation and logistic regression.* CIST, 2001. http://www.show.scot.nhs.uk/indicators/Work/Papersintro.htm (accessed July 2003).
13. Scottish Office Department of Health. *Acute services review report.* Edinburgh: The Stationery Office, 1998.
14. Griffin S. Diabetes care in general practice: meta-analysis of randomised control trials. *BMJ* 1998; **317:** 390-396.
15. Hollis S, Bennett J, Khong C, *et al*. *Quality indicators for diabetes: final report.* Manchester: QUIDS Project, 2003.
16. Watt G. The inverse care law today. *Lancet* 2002; **360:** 252-254.
17. Dowell JS, Snadden D, Dunbar J. Changing to generic formulary: how one fund-holding practice reduced prescribing costs. *BMJ* 1995; **310:** 505-508.
18. O'Neill, O. *Called to account*. Reith lecture, 2002. http://www.bbc.co.uk/radio4/reith2002/ (accessed July 2003).
19. Marshall M, Roland M. The new contract: renaissance or requiem for general practice? *Br J Gen Pract* 2002; **52:** 531-532

Acknowledgements

This study was jointly funded by the Lothian and Tayside Primary Care Research Networks, and relied on the support of the DARTS/MEMO collaboration. We would like to thank Philip Thomson and Douglas Boyle for assistance in data extraction, GPs and other clinicians in Tayside, and the members of the DARTS Steering Group: D Boyle, B Brennan, K Boyle, J Broomhall, F Cargill, P Clark, A Connacher, S Cunningham, E Dow, D Dunbar, A Dutton, S Greene, K Hunter, R Jung, M Kenicer, B Kilgallon, G Leese, R Locke, T MacDonald, R McAlpine, S McKendrick, R Newton, P Slane, F Sullivan, R Walker, S Young. DARTS is supported by The Scottish Executive, Tenovus Tayside, NHS Tayside, Tayside University Hospitals NHS Trust and Tayside Primary Care NHS Trust.