Performance of small general practices under the UK's Quality and Outcomes Framework

*Tim Doran, Stephen Campbell, Catherine Fullwood,
Evangelos Kontopantelis and Martin Roland*

ABSTRACT

Background
Small general practices are often perceived to provide worse care than larger practices.

Aim
To describe the comparative performance of small practices on the UK's pay-for-performance scheme, the Quality and Outcomes Framework.

Design of study
Longitudinal analysis (2004–2005 to 2006–2007) of quality scores for 48 clinical activities.

Setting
Family practices in England (*n* = 7502).

Method
Comparison of performance of practices by list size, in terms of points scored in the pay-for-performance scheme, reported achievement rates, and population achievement rates (which allow for patients excluded from the scheme).

Results
In the first year of the pay-for-performance scheme, the smallest practices (those with fewer than 2000 patients) had the lowest median reported achievement rates, achieving the clinical targets for 83.8% of eligible patients. Performance generally improved for practices of all sizes over time, but the smallest practices improved at the fastest rate, and by year 3 had the highest median reported achievement rates (91.5%). This improvement was not achieved by additional exception reporting. There was more variation in performance among small practices than larger ones: practices with fewer than 3000 patients (20.1% of all practices in year 3), represented 46.7% of the highest-achieving 5% of practices and 45.1% of the lowest-achieving 5% of practices.

Conclusion
Small practices were represented among both the best and the worst practices in terms of achievement of clinical quality targets. The effect of the pay-for-performance scheme appears to have been to reduce variation in performance, and to reduce the difference between large and small practices.

Keywords
incentives; quality; primary care.

INTRODUCTION

Small general practices in the UK, particularly those that are single handed, are often accused of providing poor-quality care. The 2000 *NHS Plan* cited a need to 'confirm that single-handed (solo) practices are offering high standards'.[1] The Shipman Inquiry identified advantages and disadvantages associated with single-handed practice, and described an implicit unwritten policy to reduce the numbers of solo practices in the UK; noting that concern about single-handed practitioners was not recent.[2] In 2002, the Audit Commission concluded that there were good arguments for preserving a diversity of practice sizes and types: 'One challenge is to ensure that the trend towards larger practices does not mean that patients lose out on some of the advantages that smaller practices currently offer'.[3] However, in 2008, the *NHS Next Stage Review* continued the pressure on small practices by suggesting that they be congregated as franchised practices, and also advocated groupings of larger numbers of doctors in new GP-led health centres.[4]

A new general medical services contract was introduced in the UK in 2004 incorporating a pay-for-

T Doran, MD, *clinical research fellow*; *S Campbell*, PhD, *senior research fellow*; *C Fullwood*, PhD; *E Kontopantelis*, PhD, *research associate, National Primary Care Research and Development Centre, University of Manchester, Manchester*. *M Roland*, DM, *professor of health services research, Institute of Public Health, University of Cambridge, Cambridge*.

Address for correspondence
Dr Tim Doran, National Primary Care Research and Development Centre, Williamson Building, University of Manchester, Manchester M13 9PL.
E-mail: tim.doran@manchester.ac.uk

performance scheme for family doctors: the Quality and Outcomes Framework (QOF). The scheme awards points, which are converted into financial rewards, for meeting quality targets on clinical, organisational, and patient-experience indicators.[5] The framework document stated that the contract would 'allow GPs in small practices to continue as before, but with the opportunity and the incentive to demonstrate that they can provide high quality care and be rewarded appropriately'.[6] In the first year of the scheme, quality scores appeared to increase with the size of the clinical team.[7] However, high points scores do not always correlate with high rates of achievement. Maximum achievement thresholds of between 50% and 90% exist for each clinical indicator; hence, it is possible to attain maximum points while missing the targets for between 10% and 50% of patients.

This article describes the performance, in terms of both points scored and actual levels of achievement, of small practices on 48 clinical activity indicators in the QOF compared to practices of other sizes over the first 3 years of the scheme.

METHOD

The QOF awards points to family practices on the basis of the proportion of eligible patients for whom they achieve clinical targets between a minimum threshold of 25% (that is, the target must be achieved for at least 25% of patients for the practice to earn any points) and a maximum threshold that varies according to the indicator (Table 1). There is no additional reward for achievement above the maximum threshold. Practices may exclude ('exception report') patients they deem to be inappropriate for an indicator, and these patients are removed from the achievement calculation. The maximum number of points awardable varies by indicator. In year 1 (2004–2005), each point earned the practice £76, adjusted for the relative prevalence of the disease and the size of the practice population. This was increased to £126 for years 2 and 3 (2005–2006 and 2006–2007). For year 3, most minimum achievement thresholds were raised to 40%, maximum thresholds were raised for some indicators, 17 new indicators were introduced, 32 existing indicators were combined or revised, and three were dropped. The analyses in this article relate to the 48 clinical activity indicators, covering measurement and treatment activities as well as intermediate outcomes, that remained substantially unchanged or underwent only minor revisions (Table 1).

Data on practice performance and points scored were derived from the Quality Management and Analysis System operated by the NHS Information Centre.[8] This system automatically extracts data

How this fits in

Small practices are often suspected of providing poor quality care and face pressures to congregate into larger practice groups. Under the Quality and Outcomes Framework, the quality of care provided by smaller practices is more variable than that provided by larger practices, but on average is comparable. The payment system, however, disadvantages smaller practices as a group.

from practices' clinical computing systems, including:

- the number of patients deemed appropriate for every indicator:
- those who were in the subgroup specified by the indicator and were not excluded by the practice (D_i);
- the number of patients for whom the indicator was met (N_i);
- the number of points scored (P_i); and
- for year 2 onwards, the number of patients who were excluded by the practice (E_i).[9]

Data on practice and patient characteristics were taken from the 2006 general medical statistics database, maintained by the Department of Health. Practices were grouped on the basis of the number of registered patients in each year, from group 1 (1000–1999 patients) to group 8 (≥12 000 patients; Table 2). Most group 1 practices were single handed (92.5% in year 1).

For each indicator in every practice, points scored (P_i) and reported achievement (the proportion of patients deemed appropriate by the practice for whom the targets were achieved — N_i/D_i), were measured. For years 2 and 3, when exclusion data were available, the rate of exclusions ($E_i/[D_i + E_i]$) and population achievement (the proportion of all eligible patients, including those exception reported by the practice, for whom the targets were achieved — $N_i/[D_i + E_i]$), were also measured. An example of how achievement rates are calculated and points are scored for the indicators is given in Box 1. Summary outcome scores were constructed as unweighted means of the scores for every indicator, following the method of Doran *et al*.[10] All statistical analyses were performed with Stata software (version 9).

Achievement and patient-population data for 2004–2005, 2005–2006, and 2006–2007 were available for 8277 general practices in England. Practices were excluded from the study if they had fewer than 1000 patients in any year (49 practices), one or more disease registers were missing (47 practices), complete exclusion data were not available (172 practices), complete practice characteristic data were not available (210 practices),

Table 1. Quality and Outcomes Framework clinical activity indicators included in study.

Disease	Code	Indicator[a]	Points[b] Years 1–2	Points[b] Year 3	Payment range, %[c] Years 1–2	Payment range, %[c] Year 3
Asthma	ASTHMA 2 (8)	Diagnosis confirmed by spirometry or peak flow measurement (ages ≥8 years)	0–15	0–15	25–70	40–80
	ASTHMA 3	Smoking status recorded (ages 14–19 years)	0–6	0–6	25–70	40–80
	ASTHMA 6	Have had an asthma review	0–20	0–20	25–70	40–70
Cancer	CANCER 2	Reviewed in practice (newly diagnosed patients)	0–6	0–6	25–90	40–90
Coronary heart disease	CHD 2	Referred for exercise testing and/or specialist assessment	0–7	0–7	25–90	40–90
	CHD 5	Blood pressure recorded	0–7	0–7	25–90	40–90
	CHD 6	Blood pressure ≤150/90 mmHg	0–19	0–19	25–70	40–70
	CHD 7	Total cholesterol recorded	0–7	0–7	25–90	40–90
	CHD 8	Total cholesterol ≤5 mmol/l (193 mg/dl)	0–16	0–17	25–60	40–70
	CHD 9	Taking aspirin or alternative antiplatelet/anticoagulant	0–7	0–7	25–90	40–90
	CHD 10	Taking beta-blocker	0–7	0–7	25–50	40–60
	CHD 11	Taking ACE inhibitor (history of myocardial infarction)	0–7	0–7	25–70	40–80
	CHD 12	Received influenza vaccination	0–7	0–7	25–85	40–90
Heart failure	LVD 2 (HF 2)	Diagnosis confirmed by echocardiogram	0–6	0–6	25–90	40–90
	LVD 3 (HF 3)	Taking ACE inhibitors or A2 antagonists	0–10	0–10	25–70	40–80
COPD	COPD 3 (9)	Spirometry and reversibility testing (all patients)	0–5	0–10	25–90	40–80
	COPD 8	Received influenza immunisation	0–6	0–6	25–85	40–85
Diabetes mellitus	DIABETES 2	BMI recorded	0–3	0–3	25–90	40–90
	DIABETES 5	HbA$_{1c}$ recorded	0–3	0–3	25–90	40–90
	DIABETES 6 (20)	HbA$_{1c}$ ≤7.4% (7.5% in year 3)	0–16	0–17	25–50	40–50
	DIABETES 7	HbA$_{1c}$ ≤10%	0–11	0–11	25–85	40–90
	DIABETES 8 (21)	Retinal screening recorded	0–5	0–5	25–90	40–90
	DIABETES 9	Peripheral pulses recorded	0–3	0–3	25–90	40–90
	DIABETES 10	Neuropathy testing recorded	0–3	0–3	25–90	40–90
	DIABETES 11	Blood pressure recorded	0–3	0–3	25–90	40–90
	DIABETES 12	Blood pressure ≤145/85 mmHg	0–17	0–18	25–55	40–60
	DIABETES 13	Micro-albuminuria testing recorded	0–3	0–3	25–90	40–90
	DIABETES 14 (22)	Serum creatinine recorded	0–3	0–3	25–90	40–90
	DIABETES 15	Taking ACE inhibitors/A2 antagonists (proteinuria or micro-albuminuria)	0–3	0–3	25–70	40–80
	DIABETES 16	Total cholesterol recorded	0–3	0–3	25–90	40–90
	DIABETES 17	Total cholesterol 5 mmol/l (193 mg/dl) or less	0–6	0–6	25–60	40–70
	DIABETES 18	Received influenza immunisation	0–3	0–3	25–85	40–85
Epilepsy	EPILEPSY 2 (6)	Seizure frequency recorded (ages ≥16 years)	0–4	0–4	25–90	40–90
	EPILEPSY 3 (7)	Medication reviewed (ages ≥16 years)	0–4	0–4	25–90	40–90
	EPILEPSY 4 (8)	Convulsion-free for 12 months (ages ≥16 years)	0–6	0–6	25–70	40–70
Hypertension	BP 4	Blood pressure recorded	0–20	0–20	25–90	40–90
	BP 5	Blood pressure ≤150/90 mmHg	0–56	0–57	25–70	40–70
Hypothyroidism	HYPOTHYROID 2	Thyroid function tests recorded	0–6	0–6	25–90	40–90
Severe mental health	MH 2 (9)	Reviewed in practice	0–23	0–23	25–90	40–90
	MH 4	Serum creatinine and TSH recorded (on lithium therapy)	0–3	0–1	25–90	40–90
	MH 5	Lithium levels in the therapeutic range (on lithium therapy)	0–5	0–2	25–70	40–90
Stroke	STROKE 2 (11)	Referred for CT or MRI scan	0–2	0–2	25–80	40–80
	STROKE 5	Blood pressure recorded	0–2	0–2	25–90	40–90
	STROKE 6	Blood pressure ≤150/90 mmHg	0–5	0–5	25–70	40–70
	STROKE 7	Total cholesterol recorded	0–2	0–2	25–90	40–90
	STROKE 8	Total cholesterol ≤5 mmol/l (193 mg/dl)	0–5	0–5	25–60	40–60
	STROKE 9 (12)	Taking aspirin, or alternative antiplatelet/anticoagulant (non-haemorrhagic)	0–4	0–4	25–90	40–90
	STROKE 10	Received influenza immunisation	0–2	0–2	25–85	40–85

[a]*Some indicator codes changed in year 3. Updated codes are given in brackets.* [b]*Number of points that can be awarded for the indicator. Each point earned the average practice £76 in year 1 and £126 in years 2 and 3. Total points for all indicators = 392 in years 1 and 2, and 396 in year 3.* [c]*Points are awarded on a sliding scale within the stated range. For example: for ASTHMA 6 in years 1 and 2 the practice must have reviewed at least 25% of asthma patients to earn any points, and must have reviewed 70% or more to have earned the maximum 20 points. A2 = angiotensin II receptor. ACE = angiotensin-converting enzyme. BMI = body mass index. CT = computed tomography. COPD = chronic obstructive pulmonary disease. HbA$_{1c}$ = glycated haemoglobin. MRI = magnetic resonance imaging. TSH = thyroid-stimulating hormone.*

Table 2. Number of patients and physicians in study practices, 2004–2005 to 2006–2007.

Number of patients	Group	Year 1 (2004–05)			Year 2 (2005–06)			Year 3 (2006–07)		
		Number of practices	% of practices	Number of FTE physicians	Number of practices	% of practices	Number of FTE physicians	Number of practices	% of practices	Number of FTE physicians
1000–1999	1	372	5.0	399	391	5.2	429	410	5.5	467
2000–2999	2	1143	15.2	1441	1129	15.1	1444	1096	14.6	1463
3000–3999	3	975	13.0	1816	945	12.6	1799	940	12.5	1868
4000–5999	4	1471	19.6	4065	1475	19.7	4177	1460	19.5	4363
6000–7999	5	1248	16.6	5063	1233	16.4	5153	1230	16.4	5445
8000–9999	6	973	13.0	5225	979	13.1	5347	970	12.9	5531
10 000–11 999	7	640	8.5	4202	650	8.7	4348	668	8.9	4692
≥12 000	8	680	9.1	5735	700	9.3	6065	728	9.7	6674
Total		7502	100.0	27 946	7502	100.0	28 762	7502	100.0	30 503

FTE = full-time equivalent.

the practice relocated to a more or less affluent area during the period (164 practices), or the practice population changed in size by 25% or more (258 practices). Some practices had two or more exclusion criteria. The main analyses therefore relate to 7502 practices, providing care for more than 49 million patients. Subanalyses were undertaken for excluded practices (Appendices 1–3).

RESULTS

The 7502 general practices in the main analysis provided primary-care services for 46.7 million patients in 2004–2005, and had a mean practice population of 6226 (standard deviation [SD] 3869) patients. The total number of full-time equivalent family practitioners increased from 27 946 to 30 503 over the period (Table 2), and the mean number in each practice increased from 3.72 (SD 2.56) in year 1, to 3.83 (SD 2.63) in Year 2 and to 4.07 (SD 2.84) in Year 3.

QOF points scored

The maximum number of points available for the 48 analysed indicators was 392 in years 1 and 2, and 396 in Year 3. The median proportion of available points scored was 96.6% in year 1, 99.2% in year 2, and 99.4% in year 3. In year 1 there was a clear progression in the median percentage of points scored, from 92.6% for group 1 (practices with 1000–1999 patients) to 97.6% for group 8 (practices with ≥12 000 patients; Figure 1). The gap between groups 1 and 8 decreased over time, from 5.1% in year 1 to 2.5% in years 2 and 3.

Variation in the percentage of points scored decreased with increasing number of patients; hence, the interquartile range was greatest for group 1 (13.7%) and smallest for group 8 (4.9%). Variation decreased for all groups in year 2, but there was little further change in year 3.

In year 1, the proportion of practices scoring maximum points varied according to the number of patients, from 7.0% for group 2 (2000–2999 patients) to 10.5% for group 5 (6000–7999 patients, Table 3). The number of practices scoring maximum points generally increased over time, from 8.6% overall in year 1 to 23.7% in year 3. However, the rate of increase was slowest for group 1, and by year 3 the proportion of practices attaining maximum scores

Box 1. Example of calculating achievement and allocating points under the Quality and Outcomes Framework.

Under the Quality and Outcomes Framework, indicator DM5 rewards practices for measuring the glycated haemoglobin (HbA$_{1c}$) levels of patients with diabetes. The minimum achievement threshold for this indicator was 40% in year 3, and the maximum achievement threshold was 90%. A maximum of 3 points was available for the indicator. After excluding exception-reported patients, practices achieving the target for fewer than 40% of patients earned 0 points, and practices achieving the target for 90% or more earned the maximum 3 points. Practices achieving the target for 40% to 90% of patients earned between 0 and 3 points according to the proportion of patients for whom the target was achieved.

For example: if a practice had 100 eligible patients on its diabetes register, of whom 10 were exception reported and 60 had their HbA$_{1c}$ levels measured, then:

number of exclusions (E_i) = 10

number of appropriate patients (D_i) = 100 − 10 = 90

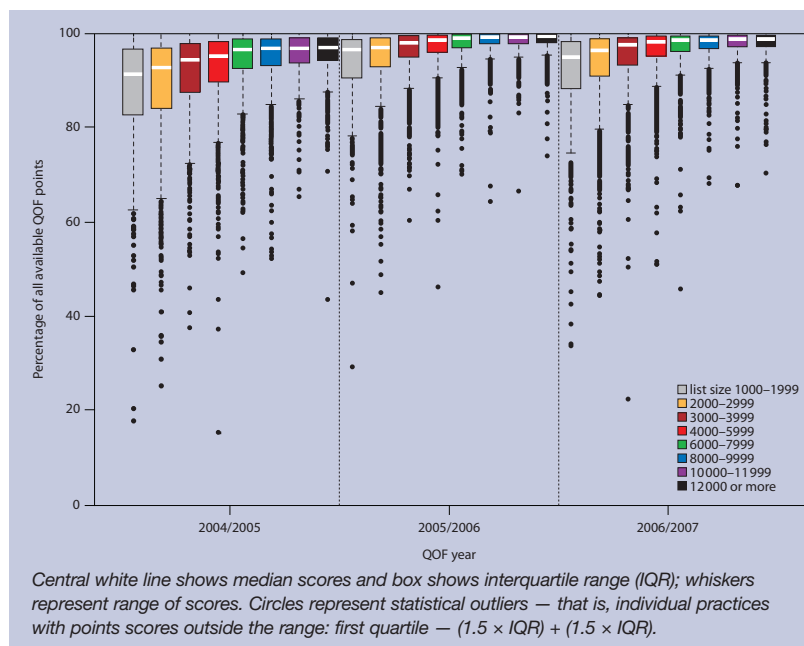number of patients for whom the target was met (N_i) = 60

and:

reported achievement = N_i/D_i = 60/90 = 66.7%

exclusion rate = $E_i/(D_i + E_i)$ = 10/(90 + 10) = 10.0%

population achievement = $N_i/(D_i + E_i)$ = 60/(90 + 10) = 60%

As reported achievement is between the minimum and maximum thresholds, the number of points scored (P_i) is calculated as:

([reported achievement — minimum threshold] × maximum points) / (maximum threshold — minimum threshold)

= (66.7 − 40) × 3/(90 − 40) = 1.6

Central white line shows median scores and box shows interquartile range (IQR); whiskers represent range of scores. Circles represent statistical outliers — that is, individual practices with points scores outside the range: first quartile — (1.5 × IQR) + (1.5 × IQR).

Figure 1. Distribution of practice scores for percentage of total QOF points scored by number of patients, year 1 (2004–2005) to year 3 (2006–2007).

Figure 2. Distribution of practice scores for overall reported achievement by number of patients, year 1 (2004–2005) to year 3 (2006–2007).

Central white line shows median scores and box shows interquartile range (IQR); whiskers represent range of scores. Circles represent statistical outliers — that is, individual practices with points scores outside the range: first quartile — (1.5 × IQR) + (1.5 × IQR).

was over two times higher for group 8. In year 1, the proportion of practices among the worst performing 5% ranged from 13.4% for group 1 to 0.7% for group 8. There was little absolute change in these proportions over time.

Reported achievement

The median overall reported achievement, the proportion of patients deemed appropriate by the practice for whom the targets were achieved, was 85.2% in year 1, 89.3% in year 2, and 90.9% in year 3. Increases in achievement between years were statistically significant (P<0.005 in all cases). Median reported achievement in year 1 varied with

number of patients, from 83.8% for group 1 to 85.9% for group 5 (Figure 2). As was the case for points scored, variation in achievement decreased with increasing number of patients; hence, the interquartile range was greatest for group 1 (1000–1999 patients, 16.2%) and smallest for group 8 (≥12 000 patients, 6.6%). However, in contrast to the findings for points scored, both the highest and the lowest achievement rates were attained by smaller practices: 13.2% of practices from group 1 were among the highest achieving 5% in year 1, and 12.1% were among the lowest achieving 5%. In contrast, the corresponding proportions for group 8 were 1.5% and 0.9% (Table 3).

By year 3 there was little difference in average achievement rates between practices of different size; a spread of 1.1% covered the median achievement of all groups. However, by this time, group 1, which had the lowest median achievement in year 1, had the highest median achievement (91.5%), and group 8 the lowest (90.4%). Variation in achievement between practices decreased for all groups in year 2 and again in year 3, with the greatest reduction for group 1; to 9.3% in year 2 and 7.2% in year 3. These patterns were consistent across all 48 individual indicators. Despite the changes over time, smaller practices remained more likely to be both very high or very low achievers in year 3. Practices with fewer than 3000 patients (groups 1 and 2) represented 20.1% of all practices, but 46.7% of the highest-achieving and 45.1% of the lowest-achieving practices.

Exception reporting and population achievement

Practices' reported achievement rates depend, in part, on the number of patients they exclude. Practices excluded a median of 6.6% of patients in year 2 and 7.4% in year 3. In both years there was a trend for higher exception reporting rates in practices with more patients, ranging from 6.3% for group 1 (1000–1999 patients) to 6.8% for group 8 (≥12 000 patients) in year 2, rising to 6.5% and 7.7% respectively in year 3 (Figure 3). As with achievement rates, there was greater variation in exception reporting rates in smaller practices, with a higher proportion of both the highest and lowest exception reporters in group 1 (Table 3).

The median overall population achievement, the proportion of all patients for whom the targets were achieved, including those exception reported by the practice, was 83.0% in year 2 and 83.8% in year 3 (Figure 4). The distributions were similar to those for reported achievement, with group 1 having the highest median population achievement in year 3 (84.6%) but also the greatest variation (interquartile
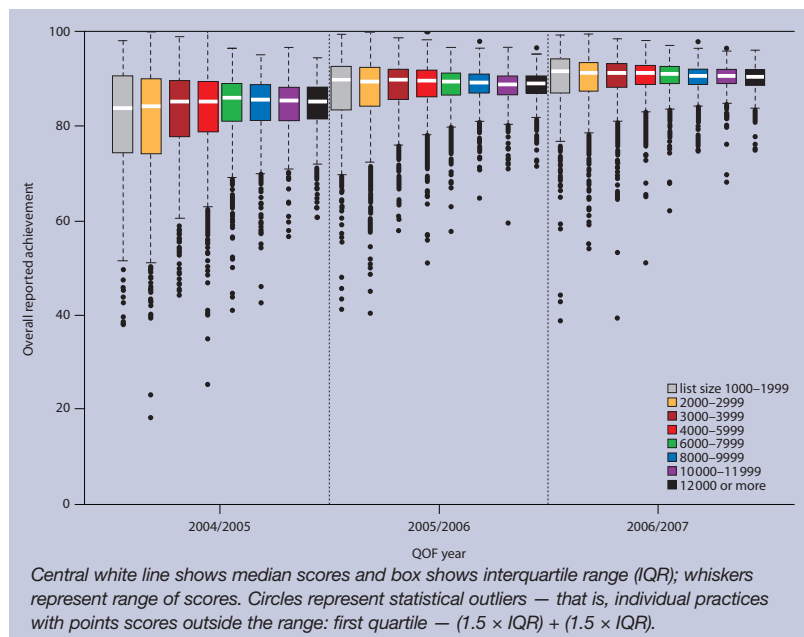
range 7.7%). The finding that group 1 practices were more likely to have both very high and very low levels of reported achievement was not, therefore, explained by their rates of exception reporting.
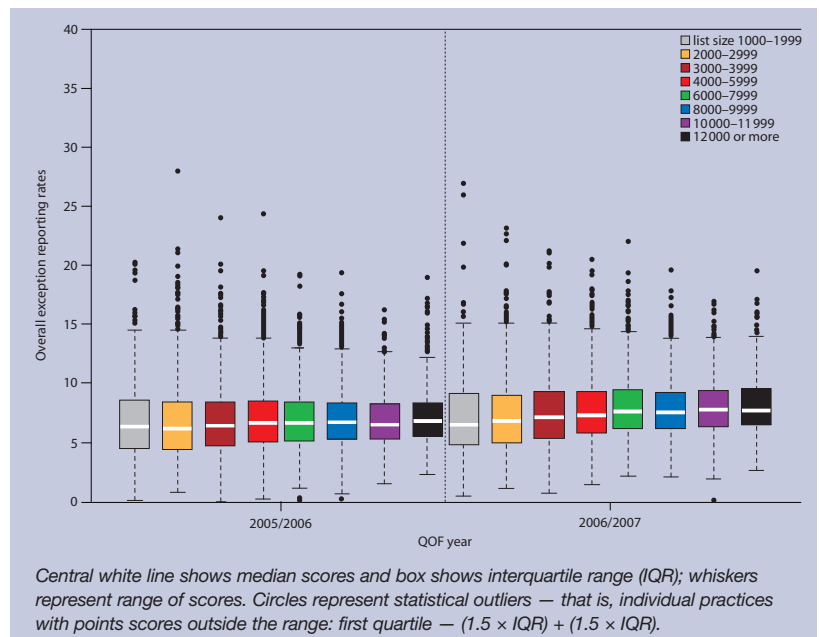
DISCUSSION

Summary of main findings

Despite a lack of clear evidence, doubts over the quality of care delivered by small, and particularly solo, practices have frequently been raised in the past. Under the QOF, this study shows that differences in performance between practices of different size depend on the measure used. When measuring quality in terms of points scored, practices with fewer patients performed worse, on average, than those with more. However, when measuring quality in terms of achievement rates, the average performance of the smallest practices was only marginally worse than that in larger practices in year 1 (2004–2005), and the gap had closed by year 3. This discrepancy is due to the maximum achievement thresholds, which make it impossible to discriminate between higher-performing practices, whose actual rates of achievement may vary by up to 50%, when using points scored as the measure of quality.

Smaller practices face several disadvantages under the pay-for-performance scheme. They tend to have relatively more patients with chronic disease, and until 2009 received less remuneration per patient under the scheme because of the formula used to adjust payments for disease prevalence.[11] Given that practices are remunerated on the basis of points scored, with achievement beyond the maximum thresholds not rewarded, the payment system does not adequately recognise the achievements of high-performing practices, many of which are small. Smaller practices therefore had to work harder for relatively less financial reward under the scheme, and yet collectively their performance improved at the fastest rate.

A particular concern with physicians working alone is that they have greater opportunity to defraud the system, for example by inappropriately exception reporting patients or falsely claiming a target has been achieved, as they do not have colleagues to directly monitor their behaviour. Fraud is difficult to monitor, but this study found that the smallest practices (predominantly single handed) had the lowest average rates of exception reporting. Patterns of population achievement, which includes exception-reported patients, were also similar to those for reported achievement, which suggests that single-handed practitioners were no more likely to have inflated their achievement scores through inappropriate

Central white line shows median scores and box shows interquartile range (IQR); whiskers represent range of scores. Circles represent statistical outliers — that is, individual practices with points scores outside the range: first quartile — (1.5 × IQR) + (1.5 × IQR).

exception reporting than their peers in larger practices. In addition, patterns of achievement for activities measured externally, such as control of glycated haemoglobin (HbA$_{1c}$) levels in patients with diabetes, were similar to those for indicators measured and reported internally.

Despite the generally high performance of small practices, many had very low achievement rates. Practices with fewer than 3000 patients represented one-fifth of all practices but nearly half of the worst-performing 5% in terms of reported achievement. Although these poorly performing practices improved at the fastest rate, in year 3 it remained the case that a significant minority of small practices

Figure 3. Distribution of practice scores for overall exception reporting by number of patients, year 2 (2005–2006) to year 3 (2006–2007).
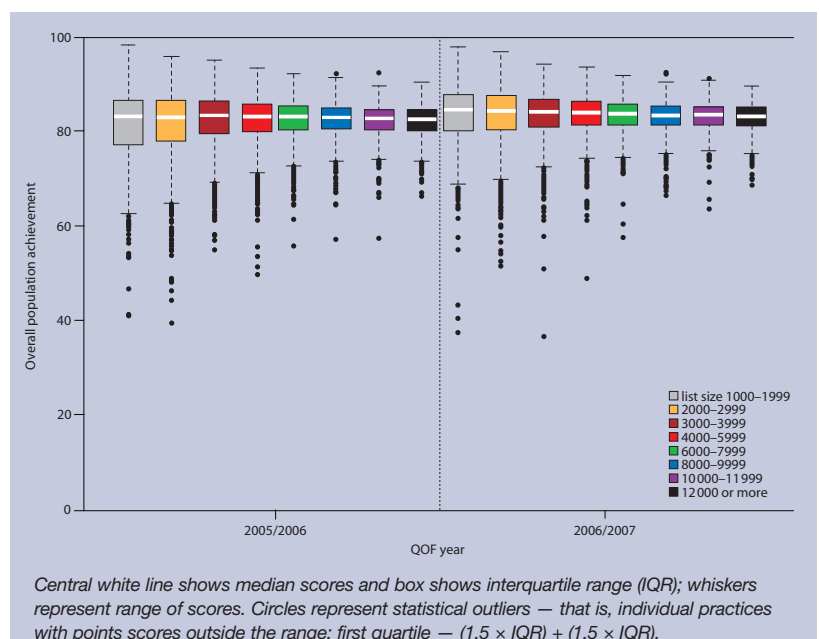
Figure 4. Distribution of practice scores for population achievement by number of patients, year 2 (2005–2006) to year 3 (2006–2007).

Central white line shows median scores and box shows interquartile range (IQR); whiskers represent range of scores. Circles represent statistical outliers — that is, individual practices with points scores outside the range: first quartile — (1.5 × IQR) + (1.5 × IQR).

Table 3. Proportion of practices from each group among the best and worst performing 5% of practices for QOF points scored, reported achievement and exception reporting from Year 1 (2004–2005) to Year 3 (2006–2007).

	Number of patients								
	1000–1999	2000–2999	3000–3999	4000–5999	6000–7999	8000–9999	10 000–11 999	≥12 000	
Group	1	2	3	4	5	6	7	8	All
Percentage of QOF points scored, practices scoring 100%[a]									
Year 1	7.26	7.00	7.90	9.25	10.50	8.84	7.34	8.82	8.58
Year 2	12.02	13.99	19.68	23.53	24.90	25.43	25.38	23.14	21.89
Year 3	13.66	15.60	19.89	24.52	27.15	26.49	29.64	29.95	23.71
Percentage of QOF points scored, worst-performing 5% of practices									
Year 1	13.44	11.11	6.87	5.37	2.24	1.44	0.78	0.74	5.00
Year 2	15.34	12.05	7.09	3.80	2.35	1.23	0.92	1.29	5.00
Year 3	17.80	11.59	7.45	3.36	1.87	1.86	1.05	1.10	5.00
Reported achievement, best-performing 5% of practices									
Year 1	13.17	10.06	5.64	4.62	3.69	1.75	2.34	1.47	5.00
Year 2	13.55	9.65	9.21	4.00	3.24	1.02	1.23	1.29	5.00
Year 3	14.88	10.40	7.02	4.79	2.85	1.44	1.20	0.96	5.00
Reported achievement, worst-performing 5% of practices									
Year 1	12.10	11.02	6.56	5.71	2.24	1.64	0.94	0.88	5.00
Year 2	13.55	10.63	6.67	4.68	2.84	1.74	1.54	1.14	5.00
Year 3	13.66	10.31	7.55	4.59	2.11	2.27	1.20	1.65	5.00
Exception reporting, best-performing 5% of practices									
Year 2	8.95	6.82	5.61	5.69	3.33	4.09	3.38	3.29	5.00
Year 3	5.85	6.11	5.85	5.34	4.63	4.33	3.89	3.57	5.00
Exception reporting, worst-performing 5% of practices									
Year 2	12.79	10.36	6.56	4.34	3.24	2.25	1.54	1.43	5.00
Year 3	12.93	11.59	7.02	4.86	2.20	1.44	1.35	1.10	5.00
Population achievement, best-performing 5% of practices									
Year 2	11.25	11.16	7.83	5.08	2.51	1.33	1.08	0.71	5.00
Year 3	13.17	11.50	6.81	5.34	2.85	1.24	0.60	0.27	5.00
Population achievement, worst-performing 5% of practices									
Year 2	14.07	10.98	6.56	4.61	2.60	1.84	1.23	1.14	5.00
Year 3	13.66	10.22	6.70	3.90	2.52	2.87	1.80	2.34	5.00

[a]*More than 5% of practices scored the maximum number of points. QOF = Quality and Outcomes Framework.*

were apparently providing substantially poorer care than the national average. Small practices are more likely to be located in deprived areas and to be poorly organised,[12] but these factors are only weakly associated with performance under the scheme, and practices with these characteristics are capable of high levels of achievement.[13,14] Other factors must, therefore, be involved.

In 2004, one solution to the problem of poor quality of primary care would have been to close small practices. However, in addition to removing many of the worst-performing practices it would have removed many of the best: over 45% of the practices in the top 5% for reported achievement had fewer than 3000 patients. A more logical approach would be to reduce variation, by bringing the worst-performing practices towards the level of the best. This appears to have occurred under the pay-for-performance scheme, through incentivising a systematic approach to a limited range of chronic diseases and publicly reporting performance. However, variation in achievement, and in exception reporting rates, remains strongly associated with practice size. This is partly a mathematical relationship: variation in all patient and practice characteristics, and in patient outcomes, will be inversely related to list size regardless of the actual quality of care provided. This has consequences for how pay-for-performance schemes measure performance, as each additional patient for whom a target is achieved or missed, whether due to the quality of clinical care provided or due to factors beyond practices' control, has greater consequences for practices with fewer patients.

Strengths and limitations of the study
The introduction of the QOF in 2004 provided the first opportunity to measure aspects of quality systematically. This study was able to measure achievement on 48 indicators across 7502 practices. However, limitations of the data restricted it to practice-level analyses, and it was not therefore possible to control for differences in age, sex, and comorbidity of patients.

Comparison with existing literature

Small practices in general, and single-handed practitioners in particular, come under regular pressure to join their colleagues in larger practices, and that pressure has intensified with the developments in general practice over the last decade, culminating in the *2008 Next Stage Review*.[4] Between 2004–2005 and 2006–2007, the number of single-handed practices in England decreased, particularly in more deprived areas,[15] as over 2500 additional full-time equivalent physicians entered general practice. Despite this, previous research suggests that there is little relationship between the size of a practice and its ability to provide high quality care.[16] Overall, some aspects of quality are associated with smaller practices, such as patient ratings of access or continuity of care, and others with larger practices, such as data recording or organisation of services.[16–18] There is also no consistent association between practice size and differences in outcome, for example number of hospital admissions for asthma, epilepsy, or diabetes; avoidable admissions;[19] or quality of care for patients with ischaemic heart disease.[20]

Implications for future research

The particular problems associated with single-handed practice — lack of peer review, risk of clinical isolation and of abuse of trust[2] — need to be addressed. The principal question is whether they are soluble without abolishing single-handed status. Single-handed practitioners are subject to the same clinical governance and appraisal arrangements as those in group practices, receive the same monitoring from primary care trusts, and, since 2004, have been measured against the same clinical quality targets under the QOF. The present results suggest that small practices, most of which are single handed, achieve, on average, similar levels of performance to larger group practices, despite an arrangement that systematically disadvantages them, but a significant minority do have low rates of achievement and the reasons for this require more attention. However, if we ask questions about why the smallest practices often appear to be the worst we should also be asking why they often appear to be the best.

REFERENCES

1. Secretary of State for Health. *The NHS Plan: a plan for investment. A plan for reform.* Cm 4818–1. London: The Stationery Office, 2000.

2. Smith J. Shipman Inquiry Fifth Report. *Safeguarding patients: lessons from the past — proposals for the future.* Cm 6394. London: HMSO, 2004.

3. Audit Commission. *A focus on general practice in England.* London: Audit Commission, 2002.

4. Department of Health. *High quality care for all: NHS next stage review.* London: Department of Health, 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(14):** 1448–1454.

6. General Practitioners Committee. *Your contract your future.* London: British Medical Association, 2002.

7. Sutton M, McLean G. Determinants of primary care medical quality measured under the new UK contract: cross sectional study. *BMJ* 2006; **332(7538):** 389–390

8. NHS The Information Centre. *Quality and Outcomes Framework Online GP Database.* http://www.qof.ic.nhs.uk/ (accessed 19 Apr 2010).

9. NHS The Information Centre. *The Quality and Outcomes Framework 2006/07 exception report.* http://www.ic.nhs.uk/statistics-and-data-collections/primary-care/general-practice (accessed 4 Aug 2010).

10. 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(4):** 375–384.

11. Guthrie B, McLean G, Sutton M. Workload and reward in the Quality and Outcomes Framework of the 2004 general practice contract. *Br J Gen Pract* 2007; **56(532):** 836–841.

12. Wang Y, O'Donnell C, Mackay D, Watt G. Practice size and quality attainment under the new GMS contract: a cross-sectional analysis. *Br J Gen Pract* 2006; **56(532):** 830–835.

13. Gulliford M, Ashworth M, Robotham D, Mohiddin A. Achievement of metabolic targets for diabetes by English primary care practices under a new system of incentives. *Diabet Med* 2007; **24(5):** 505–511.

14. Doran T, Fullwood K, Kontopantelis E, Reeves D. The effect of financial incentives on inequalities in the delivery of primary care in England. *Lancet* 2008; **372(9640):** 728–736.

15. Ashworth M, Seed P, Armstrong D, *et al.* The relationship between social deprivation and the quality of care: a national survey using indicators from the Quality and Outcomes Framework. *Br J Gen Pract* 2007; **57(539):** 441–448.

16. Campbell S, Hann M, Hacker J, *et al.* Identifying predictors of high quality care in English general practice: an observational study. *BMJ* 2001; **323(7316):** 784–787.

17. Campbell J, Ramsay J, Green J. Practice size: impact on consultation length, workload, and patient assessment of care. *Br J Gen Pract* 2001; **51(469):** 644–50.

18. Van den Hombergh P, Engels Y, *et al.* Saying 'goodbye' to single-handed practices; What do patients and staff lose or gain? *Fam Pract* 2005; **22(1):** 20–27.

19. Hippisley-Cox J, Pringle M, Coupland C, *et al.* Do single handed practices offer poorer care?: Cross sectional survey of processes and outcomes. *BMJ* 2001; **323(7308):** 320–323.

20. Majeed A, Gray J, Ambler G, *et al.* Association between practice size and quality of care of patients with ischemic heart disease; cross sectional study. *BMJ* 2003; **326(7385):** 371–372.

Appendix 1. Practices excluded from the main study.

Results for the 775 family practices excluded from the main analyses are presented below. Results are also presented for a further 507 practices that were missing data on numbers of patients for one or more years. The 775 practices for which patient numbers were available provided primary care services for a mean population of 4356 (SD 3464) patients in 2004–2005. The mean number of full-time equivalent physicians increased from 2.49 (SD 2.42) in year 1, to 2.64 (SD 2.48) in year 2 and to 3.17 (SD 2.89) in year 3 (Appendix 2). The total number of full-time equivalent physicians increased from 2563 to 2477 over the period (Appendix 3). The number of excluded practices with 1000 to 1999 patients decreased from 154 (12.0%) to 111 (8.7%).

Excluded practices generally performed worse than included practices on the pay-for-performance scheme, although the gap narrowed over time. The patterns of performance, in terms of points scored and reported achievement, were similar to those found in the included practices. The patterns for exception reporting were different, in that smaller excluded practices exception reported more patients than larger excluded practices.

Points scored

Excluded practices scored fewer Quality and Outcomes Framework points on average than included practices (mean 81.7% and 92.7% of available points respectively in year 1; Appendix 2). There was also greater variation in scores for excluded practices. As with included practices, there was a clear progression in the median percentage of points scored by excluded practices, ranging from 82.1% for group 1 (practices with 1000 to 1999 patients) to 97.7% for group 8 (practices with 12 000 or more patients). All groups improved over time, but the smallest practices improved at the fastest rate, and the gap between groups 1 and 8 decreased from 15.6% in year 1 to 5.6% in year 3. Despite this, small practices remained more likely to perform badly in year 3, with 12.2% of the worst-performing 5% of practices coming from group 1, compared with 0% from group 8.

Achievement

The median overall reported achievement for excluded practices was 80.9% in year 1, 87.8% in year 2, and 89.9% in year 3. Median reported achievement in year 1 varied with practice size, from 76.5% for group 1 to 85.2% for group 8. As was the case for included practices, variation in achievement decreased with increasing number of patients (interquartile range for group 1 of 21.7% compared with 5.8% for group 8), and the smallest practices were more likely to have either very high or very low achievement rates: 7.9% of practices from group 1 were among the highest achieving 5% in year 1, and 4.6% were among the lowest-achieving 5% (the latter figure is relatively low because 30% of the lowest-performing practices had missing list-size data). In contrast, the corresponding proportions for group 8 were 2.3% and 0%. By year 3, there was little difference in average achievement rates between groups: the difference in median achievement between group 1 and group 8 having fallen from 8.7% to 1.9%. Variation in achievement also decreased for all groups over time.

Exception reporting and population achievement

Excluded practices excluded a median of 7.0% of patients in year 2 and 8.0% in year 3. In contrast to included practices, there was a trend for lower exception reporting rates in practices with more patients, ranging from 7.5% for group 1 to 6.5% for group 8 in year 2, rising to 8.4% and 7.9% respectively in year 3. As with achievement rates, there was greater variation in exception reporting rates in smaller practices, with a higher proportion of both the highest and lowest exception reporters in group 1.

The median overall population achievement, the proportion of all patients for whom the targets were achieved including those exception reported by the practice, was 80.3% in year 2 and 82.1% in year 3. There was a clear progression in the median population achievement of excluded practices in year 2, ranging from 76.9% for group 1 to 82.3% for group 8. By year 3 the gap between these groups had reduced from 5.4% to 3.0%.

Appendix 2. Summary data for included and excluded practices.

Parameter	Year	Included practices	Excluded practices
Mean (SD) FTE GPs	1	3.72 (2.56)	2.49 (2.42)
	2	3.83 (2.63)	2.64 (2.48)
	3	4.07 (2.84)	3.17 (2.89)
Mean (SD) QOF points scored	1	92.68 (8.73)	81.71 (20.25)
	2	96.76 (5.20)	91.88 (11.91)
	3	95.93 (6.38)	91.33 (12.33)
Mean (SD) reported achievement	1	83.00 (8.90)	76.94 (13.84)
	2	88.01 (5.74)	84.83 (9.34)
	3	90.01 (4.55)	87.86 (7.67)
Mean (SD) exception reporting	2	6.97 (2.82)	7.94 (3.94)
	3	7.72 (2.84)	8.52 (3.61)
Mean (SD) population achievement	2	82.06 (5.53)	77.96 (9.13)
	3	83.19 (4.68)	80.50 (7.27)

FTE = full-time equivalent. QOF = Quality and Outcomes Framework. SD = standard deviation.

Appendix 3. Number of full-time equivalent physicians in excluded practices, 2004–2005 to 2006–2007.

Number of patients	Group	Year 1 (2004–2005)			Year 2 (2005–2006)			Year 3 (2006–2007)		
		Number of practices	% of practices	Number of FTE GPs	Number of practices	% of practices	Number of FTE GPs	Number of practices	% of practices	Number of FTE GPs
Missing	0	359	28.0	397	424	33.1	268	457	35.7	124
1000–1999	1	154	12.0	132	125	9.8	111	107	8.4	89
2000–2999	2	257	20.1	281	202	15.8	226	171	13.3	167
3000–3999	3	137	10.7	205	120	9.36	175	106	8.3	151
4000–5999	4	148	11.5	347	165	12.9	378	173	13.5	387
6000–7999	5	89	6.9	341	92	7.2	341	97	7.6	369
8000–9999	6	58	4.5	261	70	5.5	322	70	5.5	351
10 000–11 999	7	37	2.9	230	31	2.4	207	41	3.2	259
≥12 000	8	43	3.4	369	53	4.1	480	60	4.7	580
Total		1282	100.0	2563	1282	100.0	2508	1282	100.0	2477

FTE = full-time equivalent.