

## SPECIAL REPORT

## Quality of Primary Care in England with the Introduction of Pay for Performance

Stephen Campbell, Ph.D., David Reeves, Ph.D., Evangelos Kontopantelis, Ph.D.,  
Elizabeth Middleton, M.Sc., Bonnie Sibbald, Ph.D., and Martin Roland, D.M.

In 2004, the United Kingdom committed £1.8 billion (\$3.2 billion) to a new pay-for-performance contract for family practitioners.<sup>1</sup> During the first year, the levels of achievement exceeded those anticipated by the government, with an average of 83.4% of the available incentive payments claimed.<sup>2</sup> However, the quality of care in English family practices had already begun to improve in response to a wide range of initiatives,<sup>3-6</sup> including national standards for the treatment of major chronic diseases and a national system of inspection (Table 1). Family practitioners already had some experience with financial incentives from the limited use of incentive programs that were initiated in 1990.<sup>7,8</sup> It is therefore unclear whether the high levels of quality attained after the pay-for-performance contract was introduced in 2004 reflect improvements that were already under way or whether existing trends toward improvement were accelerated. The effect of the incentive program must be understood in the context of the comprehensive quality-improvement strategy within which the contract was introduced.

This report presents data from a longitudinal cohort study that measured the quality of care in a representative sample of primary care practices in England at two time points (1998 and 2003)

before the pay-for-performance program was introduced and at one time point (2005) after its introduction. A validated set of criteria was used to assess quality in the management of three chronic conditions: asthma, coronary heart disease, and type 2 diabetes. Because some clinical indicators — the measures of the quality of clinical care — were not rewarded with financial payments in the 2004 pay-for-performance program, the study design also permitted a comparison of trends in the quality of care for indicators for which financial incentives were provided and for those for which they were not provided in the management of these three conditions.

### METHODS

In 1998, we measured the quality of care in a stratified, random sample of 60 primary care practices in six geographic areas of England. These practices were nationally representative in terms of size, whether the practice was approved for residency training, and the sociodemographic characteristics of their populations.<sup>9</sup> We followed up 42 of these practices in 2003 and 2005. The reduction in the number of practices was due partly to attrition and partly to the retirement of solo physi-

**Table 1. Examples of Key Initiatives in the Broad National Quality-Improvement Strategy.**

National standards for the treatment of major chronic diseases, such as the National Service Frameworks for coronary heart disease (1999) and diabetes (2003)
Contractual requirement for practitioners to undertake a clinical audit (initially a requirement in the 1990 contract)
Financial incentives for cervical cytologic testing and immunization (early 1990s)
Widespread use of audit and feedback by the Primary Care Trusts
Release of comparative data for quality of care to practitioners (common) and the public (rare) by the Primary Care Trusts
Annual appraisal of all primary care physicians (by the Primary Care Trusts and including discussion of some audit data)
National system of inspection and monitoring of performance (by the Healthcare Commission)

cians and the closing of other practices. The 42 practices for which data were available for the longitudinal analysis were still nationally representative in terms of socioeconomic status, but solo practitioners were underrepresented.<sup>4</sup> However, these 42 practices have values close to the national averages for socioeconomic status, population density, and type of housing of the patient population, and their performance was also typical of English family practices during the first year of the pay-for-performance program.<sup>2</sup> The analysis was restricted to the 42 practices for which data were available for all three time points (1998, 2003, and 2005). The research protocol was approved by the ethics committee of the multicenter Manchester National Health Service.

#### DATA COLLECTION

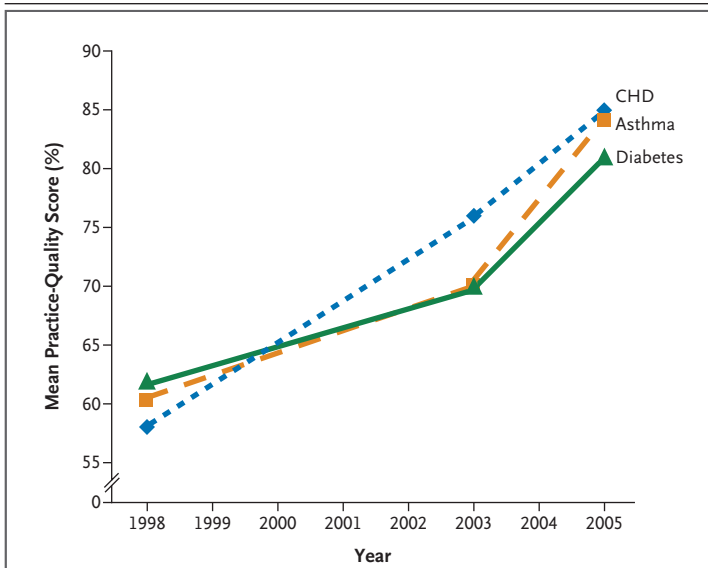
Trained research staff extracted the data to assess the quality of clinical care for the categories of coronary heart disease (15 clinical indicators), asthma (12 clinical indicators), and type 2 diabetes (21 clinical indicators). Data were collected from both computerized and handwritten medi-

cal records with the use of evidence-based review criteria<sup>10,11</sup> developed with the RAND–UCLA appropriateness method.<sup>12</sup> Patients with these three conditions were randomly selected from lists of those receiving the relevant drugs (see the Supplementary Appendix, available with the full text of this article at [www.nejm.org](http://www.nejm.org)) according to repeat prescriptions within the previous 6 months, and separate samples of patients treated in 1998, 2003, and 2005 were selected. In 1998, for two practices, there were no eligible patients who had coronary heart disease because of the young age of the patient population, so for that time point, the results for this condition are based on only 40 practices. Data were collected for up to 20 patients for each of the three conditions in each practice in 1998 (some small practices did not have 20 patients for each of the conditions) and for up to 12 patients for each condition in each practice in 2003 and 2005. Data were collected for a total of 2300 patients in 1998, for 1495 patients in 2003, and for 1482 patients in 2005. These data are presented as a pooled analysis across practices.

Although the study did not include conditions that were not rewarded with financial incentives in the pay-for-performance program, there were clinical indicators for coronary heart disease, asthma, and type 2 diabetes for which financial incentives were not provided in 2004. We compared 30 indicators for which financial incentives were provided with 17 indicators for which financial incentives were not provided. In this analysis, we excluded three clinical indicators for which this distinction was unclear — that is, it was not clear whether there were financial incentives provided for the indicator at all three time points.

#### STATISTICAL ANALYSIS

An overall score for the quality of care was computed for each patient included for 1998, 2003, and 2005. For each patient with asthma, coronary heart disease, or type 2 diabetes, the score was computed as a ratio: the number of clinical indicators for which appropriate care was provided, divided by the number of indicators relevant to that patient. Expressed as a percentage, this score represents the percentage of “necessary care”<sup>10</sup> provided to each patient, within a range from 0 to 100. We adopted this measure for consistency with our previous investigation of this sample.<sup>4</sup> Scores for the quality of care at the practice level were



**Figure 1.** Mean Scores for Clinical Quality at the Practice Level for Coronary Heart Disease, Asthma, and Type 2 Diabetes, 1998 to 2005.

The quality of care for coronary heart disease (CHD), asthma, and type 2 diabetes was improving between 1998 and 2003, before the introduction of pay for performance. The rate of improvement in quality of care increased significantly for diabetes and asthma between 2003 and 2005, after the introduction of pay for performance; the rate for coronary heart disease, which was increasing most rapidly before pay for performance, continued at the same rate after pay for performance was introduced.

computed as the simple average of the scores for individual patients within each practice.

We also performed analyses for individual clinical indicators. Data were available for at least five patients in a practice for each indicator analyzed and for geographic areas where the number of practices meeting this criterion was at least 10. In 1998, data were collected for indicators requiring an activity to be undertaken annually on the basis of data recorded during the previous 14 months, whereas in 2003 and 2005, data were collected for indicators requiring an activity to be undertaken annually on the basis of data recorded during the previous 15 months, in line with the pay-for-performance contract. For consistency, we generated 15-month versions of the clinical indicators for 1998 from our original data for use in the present analysis.

We compared scores for observed quality in 2005 with the scores predicted on the basis of the trend between 1998 and 2003. Practice scores for individual indicators, computed as a percentage of patients receiving appropriate care as the indicator, were subject to a ceiling effect of 100%. In calculating the expected scores for 2005, it was inappropriate to use a simple linear model, because such a model would fail to account for ceiling effects (i.e., some predicted scores would have exceeded 100% if the previous linear trend had been extrapolated from 2003 to 2005). Probit and logit models are most commonly used to model binary data. We adopted the logit model a priori for the current analysis, calculating expected values for 2005 on the basis of the logit curve that the scores from 1998 to 2003 followed and extrapolating this curve to 2005. After per-

forming the analysis, we computed probit predictions for comparative purposes and found these all to be within 1 percentage point of their logit equivalent. For quality-of-care scores between 20 and 80%, the logit and probit curves are essentially linear.

For each practice, we therefore computed a predicted score for 2005 using a logit projection from 1998 and 2003. We computed predicted values for overall scores and for individual clinical indicators. The predicted scores were then compared with the actual scores for the practices in 2005. However, because of floor and ceiling effects, the differences between actual and predicted scores are not equivalent across the scale: the difference between an observed score of 54 and a predicted score of 50 does not have the same import as the difference between a score of 99 and a score of 95. To adjust for this difference, observed and predicted scores were converted into their logit equivalents before the analysis. Under the transformation, a proportion,  $P$ , is transformed into a log odds, as  $\text{Logit}(P) = \ln[P \div (1 - P)]$ . Since  $\text{Logit}(P)$  cannot be computed where the value of  $P$  is 0 or 1, we used the empirical logit,  $\text{Logit}(P) = \ln[(P + 0.5 \div n) \div (1 - P + 0.5 \div n)]$ , in these cases, where  $n$  is the number of observations over which  $P$  is calculated.<sup>13</sup> The logit transformation maps the scale of 0 to 100% to a scale of less than infinity to infinity.<sup>13</sup> The transformation therefore “stretches” the scores at the extremes, which increases the effect on the results of the analysis of practices for which scores are close to the floor or the ceiling.

The transformed observed and predicted scores for 2005 were then compared by means of a

**Table 2. Changes in Mean Scores at the Practice Level for Quality of Care for Coronary Heart Disease, Type 2 Diabetes, and Asthma, 1998 to 2005.\***

Variable	Coronary Heart Disease	Diabetes	Asthma
Mean score for 1998 — %	58.6	61.6	60.2
Mean score for 2003 — %	76.2	70.4	70.3
Mean score for 2005 — %	85.0	81.4	84.3
Mean predicted score for 2005 (logit model) — %	80.7	73.2	72.3
Mean difference between transformed observed score and predicted score for 2005 — % (95% CI)	0.22 (−0.02 to 0.45)	0.68 (0.27 to 1.1)	0.44 (0.27 to 0.62)
P value	0.07	0.002	<0.001

\* CI denotes confidence interval. P values are for the comparison between transformed observed and predicted scores for 2005.

**Table 3. Scores for Quality of Care at the Practice Level in 1998, 2003, and 2005.\***

Condition	Clinical Indicator	1998	2003 percent	2005	Mean Difference between Transformed Observed Score and Predicted Score in 2005 (95% CI)†‡	P Value
<b>Coronary heart disease</b>	Frequency or pattern of angina attacks recorded during the previous 15 mo	43.4	60.3	70.1	0.22 (−0.42 to 0.86)	0.49
	Blood pressure recorded during the previous 15 mo	87.0	95.3	98.8	0.26 (0.02 to 0.51)	0.04
	Exercise capacity recorded during the previous 15 mo	36.5	66.3	72.8	0.21 (−0.50 to 0.91)	0.55
	Cholesterol level recorded during the previous 5 yr	63.5	89.7	98.7	0.54 (0.26 to 0.82)	<0.001
	Dietary advice recorded during the previous 5 yr	58.8	75.7	87.4	0.64 (−0.06 to 1.34)	0.07
	Smoking status recorded during the previous 5 yr	85.1	86.6	98.3	0.87 (0.47 to 1.27)	<0.001
	Referral to specialist for exercise stress testing or assessment (ECG) ever recorded	76.2	88.2	89.2	−0.01 (−0.40 to 0.38)	0.95
	Prescription or advice to take aspirin recorded unless record of contraindication or intolerance	76.2	80.7	92.5	0.71 (0.27 to 1.15)	0.002
	Blood pressure controlled to ≤150/90 mm Hg	47.9	72.2	81.5	0.15 (−0.23 to 0.52)	0.44
	Serum cholesterol controlled to 190 μg/dl	16.9	60.7	72.9	0.10 (−0.41 to 0.61)	0.69
<b>Conditional factors</b>	Smoking advice to smokers recorded during the previous 5 yr	66.5	76.8	99.1	NA	NA
	Weight advice for overweight patients recorded during the previous 5 yr	63.7	78.8	82.5	0.28 (−0.72 to 1.29)	0.56
	Action taken on blood pressure if systolic pressure >160 mm Hg, or if >140 mm Hg and total serum cholesterol >190 mg/dl	34.3	44.4	58.3	NA	NA
	Action taken if cholesterol >190 mg/dl in patient age <70 yr	53.8	73.6	79.2	NA	NA
<b>Asthma</b>	Beta-blocker prescribed as maintenance therapy if sole therapy	47.4	58.1	79.5	NA	NA
	Normal or predicted peak expiratory flow or record of difficulty using meter recorded during the previous 5 yr	76.1	85.7	91.6	0.28 (−0.07 to 0.64)	0.12
	Daily, nocturnal, or activity-limiting symptoms recorded during the previous 15 mo	43.7	58.1	76.5	0.91 (0.31 to 1.52)	0.004
	Smoking status recorded during the previous 5 yr	82.6	84.3	96.0	0.59 (0.16 to 1.01)	0.008
Inhaler technique recorded during the previous 5 yr	53.1	58.7	76.0	0.71 (0.13 to 1.28)	0.02	

Conditional factors	Smoking advice to smokers recorded during the previous 5 yr	64.2	86.6	96.6	NA	NA	
	Self-management plan for patients receiving high-dose corticosteroids or those who had inpatient treatment for asthma recorded during the previous 5 yr	22.5	56.0	77.8	NA	NA	
	Short-acting bronchodilators prescribed for use before exercise for patients with exercise-induced bronchospasm recorded during the previous 5 yr	40.1	74.2	90.9	NA	NA	
	Peak expiratory flow during a consultation for an exacerbation recorded for most recent episode	100	100	100	NA	NA	
	Oral corticosteroids prescribed if peak expiratory flow <60% of normal recorded for most recent episode	61.7	73.6	93.6	NA	NA	
	Action taken if patients had nocturnal symptoms recorded during the previous 5 yr	30.6	80.8	100	NA	NA	
	Action taken if patients had activity-limiting symptoms recorded during the previous 5 yr	50.0	0	0	NA	NA	
	Referral to respiratory physician if oral steroids were used in maintenance treatment, recorded during the previous 5 yr	76.1	85.7	91.6	NA	NA	
	<b>Type 2 diabetes</b>						
	Glycated hemoglobin recorded during the previous 15 mo	87.1	93.1	99.4	0.54 (0.29 to 0.80)	<0.001	
	Visual examination of feet recorded during the previous 15 mo	57.4	69.6	88.0	0.99 (0.45 to 1.53)	0.001	
	Peripheral pulses or vibration sense recorded during the previous 15 mo	60.0	62.9	90.2	1.58 (0.98 to 2.17)	<0.001	
	Serum creatinine recorded during the previous 15 mo	79.6	89.9	96.4	0.33 (-0.08 to 0.73)	0.11	
	Urine proteinuria recorded during the previous 15 mo	66.3	74.8	82.8	0.43 (-0.12 to 0.98)	0.12	
Examination of fundi or visual acuity recorded during the previous 15 mo	69.4	72.2	82.7	0.58 (0.10 to 1.06)	0.02		
Weight recorded during the previous 15 mo	80.2	86.5	97.2	0.60 (0.20 to 1.01)	0.005		
Blood pressure recorded during the previous 15 mo	92.6	95.8	99.0	0.15 (-0.10 to 0.40)	0.22		
Smoking status recorded during the previous 5 yr	86.5	88.5	98.4	0.58 (0.13 to 1.03)	0.01		
Serum cholesterol recorded during the previous 5 yr	75.1	97.6	99.4	0.05 (-0.08 to 0.18)	0.47		
Blood pressure controlled to $\leq 140/85$ mm Hg (recorded in previous 15 mo)	21.8	35.4	49.0	0.49 (0.04 to 0.94)	0.03		
Total serum cholesterol controlled to $\leq 190$ mg/dl (recorded in previous 5 yr)	21.8	52.0	72.5	0.42 (0.03 to 0.81)	0.03		
Glycated hemoglobin controlled to $\leq 7.4\%$ (recorded in previous 15 mo)	37.8	39.8	50.6	0.40 (0 to 0.81)	0.05		

Conditional factors	84.8	87.7	100	0.29 (-0.14 to 0.73)	0.17§
Documentation of patient education if diabetes diagnosed <5 yr recorded during the previous 5 yr	84.8	87.7	100	0.29 (-0.14 to 0.73)	0.17§
Advice given to smokers recorded during the previous 5 yr	70.0	78.3	95.2	NA	NA
Referral to a specialist if creatinine >2.26 µg/dl recorded during the previous 5 yr	100	100	100	NA	NA
In patients age <80 yr, treatment offered if average of the three previous blood-pressure readings was diastolic >100 or systolic >150 and diastolic >90 mm Hg	53.6	90.5	66.7	NA	NA
In patients treated for hypertension, if proteinuria present, patients should receive ACE inhibitor	68.8	100	75.0	NA	NA
In patients starting ACE inhibitor, measurement of creatinine and potassium within 1 mo before treatment started recorded during the previous 5 yr	38.1	29.9	38.8	NA	NA
In patients receiving sulfonyleurea, hypoglycemia symptoms recorded during the previous 15 mo	18.2	8.1	7.8	-0.12 (-0.57 to 0.33)	0.60
In patients ≤70 yr of age, if previous glycosylated hemoglobin measured was >9%, patients offered a therapeutic intervention to improve glycemic control	81.3	71.9	66.0	NA	NA

\* Conditional indicators are indicators that did not apply to all patients. To convert values for cholesterol to millimoles per liter, multiply by 0.026. To convert values for creatinine to micromoles per liter, multiply by 88.4. CI denotes confidence interval, ECG electrocardiography, NA not analyzed (analysis was performed only when there were at least 10 practices with a minimum of five patients each that provided data for an indicator at each of the three time points), and ACE angiotensin-converting enzyme.  
 † The results are based on all practices for which data were available.  
 ‡ The results are based on practices with at least five patients who had the condition at each of the three time points. Statistical tests were not performed for geographic areas in which fewer than 10 practices met this criterion.  
 § Results became significant (P=0.004) using the bootstrap method.

matched-pairs two-tailed t-test. In view of the relatively small number of practices included in the analysis and the distributional assumptions made by this test, a bootstrap procedure based on 1000 bootstrap samples was used to confirm the significance of the tests.

Although the logit model is appropriate to an analysis of individual binary indicators, the overall scores for quality used here may not conform to the logit curve as the scores approach the ceiling. Therefore we repeated the analysis, using a linear model applied to the untransformed scores. The linear model makes no adjustment for ceiling effects, except that we did not allow predictions greater than 100%. Hence the results of the sensitivity analysis are likely to be conservative.

For comparisons of clinical indicators for which financial incentives were provided with those for which they were not provided, we derived observed and predicted scores for clinical quality at the practice level separately for the groups of indicators in the management of each condition for which financial incentives were provided and for those indicators for which they were not provided. The logit transformation was then applied to these scores. A matched-pairs t-test was used to compare the difference between the observed scores for indicators for which financial incentives were provided and those for which financial incentives were not provided with the predicted difference. The significance of the test was confirmed with the use of the bootstrap procedure.

---

## RESULTS

---

The quality of care in the categories of coronary heart disease, asthma, and type 2 diabetes improved between 2003 and 2005, continuing the earlier trend (Fig. 1). However, the increase in the rate of improvement between 2003 and 2005 was significant for asthma ( $P<0.001$ ) and diabetes ( $P=0.002$ ) (Table 2). Scores for coronary heart disease also increased, but the change in the rate of improvement was not significant ( $P=0.07$ ). Similarly, the sensitivity analysis performed with the use of the more conservative linear model showed significant increases in the rate of improvement for asthma and diabetes, as compared with the rates for coronary heart disease.

On the basis of the conservative linear model, the annual rate of increase in the quality of care

between 2003 and 2005 for diabetes was faster than the annual rate between 1998 and 2003 in 34 practices and was slower in 8 practices (significantly so for 13 and 0 practices, respectively;  $P<0.05$ ). For asthma, the annual rate of improvement in the quality of care was unchanged between 1998 and 2003 in 1 practice, faster in 28 practices, and slower in 13 practices (significantly so for 10 and 2 practices, respectively;  $P<0.05$ ). For coronary heart disease, although the annual rate of improvement between 1998 and 2003 was faster in 23 practices, it was slower in 17 practices (significantly so for 7 and 5 practices, respectively;  $P<0.05$ ).

Observed scores for clinical quality for 1998, 2003, and 2005 and the predicted score for 2005 for each indicator in each condition are shown in Table 3. The table includes examples of changes in individual clinical indicators that are likely to be particularly important for improving patient outcomes, such as control of cholesterol and blood pressure in the management of coronary heart disease.<sup>14</sup>

We then compared changes in the quality of care between clinical indicators for which financial incentives were provided in 2004 and those for which financial incentives were not provided. The quality of performance for indicators with incentives in all three conditions was substantially higher at all three time points than for those without incentives. However, in all conditions, the rate of improvement between 2003 and 2005 for clinical indicators for which financial incentives were provided, as compared with those for which they were not, did not differ significantly from the rate predicted on the basis of the trend between 1998 and 2003 (Table 4).

---

## DISCUSSION

---

Although the quality of care in the categories of asthma, coronary heart disease, and type 2 diabetes was improving before the introduction of the 2004 contract, our results suggest that the introduction of pay for performance was associated with a modest acceleration in improvement for two of these three conditions: diabetes and asthma. In most of the 42 practices for which data were available, the annual improvement for both was accelerated. The results are based on care reported in the medical records but not necessarily on care provided, and it is a common

**Table 4. Mean Difference in Improvement for Indicators with and without Incentives.\***

Category	Mean Difference (95% CI)	P Value
Coronary heart disease	0.53 (-0.01 to 1.08)	0.054
Asthma	0.03 (-0.45 to 0.51)	0.904
Type 2 diabetes	0.08 (-0.32 to 0.49)	0.682

\* The mean difference is the amount by which the observed difference between transformed overall scores for clinical indicators for which financial incentives were or were not provided exceeded the predicted difference. CI denotes confidence interval.

criticism of pay-for-performance programs that their main effect is to promote better recording of care rather than better care. However, the panels used to develop the indicators judged that to provide good care, it was necessary both to provide the care and to record the processes and intermediate outcomes assessed in terms of the indicators we used.<sup>10</sup>

Because some details of the new contract were publicized before the contract was introduced in 2004, it is possible that some practices were preparing for the incentives during our second round of data collection in 2003. If this is true, we might have overestimated improvements in quality made before pay for performance was introduced,<sup>4</sup> so that the results reported here could represent a conservative estimate of the actual improvement resulting from pay for performance.

The key question is whether the increased rate of improvement in quality of care after the new contract was introduced can be attributed to pay for performance or to other factors. The pay-for-performance program was the only major national policy implemented in primary care in England in 2004 that targeted the types of care processes evaluated in this study. However, since practices were observed at only two time points before the introduction of pay for performance, we were unable to determine whether the rate of improvement had already accelerated as a result of earlier but still ongoing initiatives. No control group could be recruited, because financial incentives were applied simultaneously across the whole of the United Kingdom. A final concern is that the patients included in the study were selected on the basis of the presence or absence of treatment with relevant drugs, and those who were untreated or

who did not comply with treatment were excluded from the analysis. This selection bias could have resulted in overestimation of the quality of care at all three time points, although the trends in quality should have been unaffected.

The study focuses on three chronic conditions — asthma, coronary heart disease, and type 2 diabetes — for which financial incentives were provided under the pay-for-performance program and which had also been subject to considerable quality-improvement activity in the United Kingdom as part of a national quality-improvement strategy. The finding of a significant increase in the rate of improvement for asthma and diabetes but not for coronary heart disease may reflect the fact that in 2003 scores for quality for coronary heart disease were already higher than those for the other two conditions. Coronary heart disease had been a particular target of earlier quality-improvement initiatives, with 98% of the Primary Care Trusts reporting coronary heart disease initiatives in 2001 and 2002.<sup>15</sup>

The finding of no significant difference in the rate of improvement between clinical indicators for which financial incentives were provided and those for which they were not provided suggests that the pay-for-performance program may not necessarily have been responsible for the acceleration in improvement that we found between 2003 and 2005. However, the study was not designed or powered for this analysis, and the broad confidence limits for many of the clinical indicators shown in Table 3 reflect the uncertainty associated with the small sample available for the analysis. In addition, there may have been a “halo effect,” as a result of which some indicators for which financial incentives were not provided may have been indirectly rewarded. For example, the clinical indicator “control of total serum cholesterol in coronary heart disease to 190 mg per deciliter (5 mmol per liter) or less,” which in the 2004 contract became an indicator for which a financial incentive was provided, is likely to have influenced performance on “evidence of action being taken if cholesterol was raised,” a clinical indicator for which a financial incentive was not specifically provided. Improvements may therefore have spilled over onto other aspects of care that were not subject to performance monitoring. This effect has previously been noted in the Department of Veterans Affairs quality-improvement programs.<sup>16</sup> The study

was not able to assess what is perhaps a more important question, namely, what the effect of financial incentives was on care for conditions for which no financial incentives were provided at all.

The introduction of the pay-for-performance program has been associated with a general trend in the National Health Service away from placing implicit trust in health care professionals and toward more active monitoring of their performance than before the program was introduced.<sup>17</sup> Financial incentives are most likely to be an effective means of influencing professional behavior when performance targets and rewards are aligned to the values of the staff being rewarded.<sup>18,19</sup> Professional motivation alone may not be sufficient to improve the quality of care, especially when physicians have to make financial investments in their practices — for example, by employing more staff to achieve gains in quality. Sustained improvement in quality of care, which involves a range of health care providers (e.g., physicians, nurses, and administrative staff), requires a combination of other factors, including clear goals, good teamwork, and effective leadership.<sup>20</sup>

Owing to the inherent limitations of our study design and data, it was not possible to determine whether improvements in quality of care resulted only from the pay-for-performance program; our findings are consistent with previous work, which suggested that financial incentives can change professional behavior<sup>21,22</sup> and that patients receive higher-quality care in geographic areas where performance measures and monitoring have been established.<sup>16</sup> However, there are also potential, unintended consequences of such schemes.<sup>1,23</sup> These include the possible neglect of geographic areas where financial incentives for improvements in care are not provided and of “myopia” (the pursuit of short-term targets at the expense of legitimate long-term objectives) or “misrepresentation” (deliberate manipulation of data so that reported behavior differs from actual behavior).<sup>24</sup> In addition, external incentives may crowd out motivation — the desire to do a task well for its own sake.<sup>25,26</sup> In the United Kingdom, family practitioners have predicted that among the adverse consequences of financial incentives may be a reduction in the continuity of care, fragmentation of care as a result of specialization within practices, and neglect of conditions for

which financial incentives are not provided.<sup>27</sup> Despite these concerns, overall job satisfaction among family physicians was higher in 2004 than in 2001.<sup>28</sup> Moreover, a recent report from the United States suggests that targeted quality-improvement programs have not resulted in a deterioration in the quality of care in untargeted disease areas.<sup>29</sup> Our results generally support the view of the Institute of Medicine that pay-for-performance programs can make a useful contribution to improving quality,<sup>30</sup> particularly when such programs are part of a comprehensive quality-improvement program.<sup>31</sup>

The size of the gains in quality in relation to the costs of pay for performance remains a political issue in the United Kingdom,<sup>32</sup> and the government now accepts that it paid more than it had expected to pay for the improvements in performance.<sup>33</sup> The proportion of practice income taken as profit by general practitioners appears to have increased after the new contract was introduced, suggesting that gains in quality could have been achieved at a lower cost. For the years 2006 through 2007, the pay-for-performance framework has been amended to introduce higher payment thresholds, new targets, and new disease areas<sup>34</sup> without increasing physicians' maximum available income from incentive payments. Physicians in the United Kingdom may now need to work harder or employ more staff to earn the same rewards that they had received before 2006.

Supported by the U.K. Department of Health. The views presented here are those of the authors and not necessarily of the U.K. Department of Health.

Dr. Roland reports serving as an academic advisor to the government and the British Medical Association negotiating teams during the development of the United Kingdom pay-for-performance scheme during 2001 and 2002. No other potential conflict of interest relevant to this article was reported.

From the National Primary Care Research and Development Centre, University of Manchester, Manchester, United Kingdom. Address reprint requests to Dr. Campbell at the National Primary Care Research and Development Centre, University of Manchester, Oxford Rd., Manchester M13 9PL, United Kingdom.

1. Roland M. Linking physician pay to quality of care: a major experiment in the United Kingdom. *N Engl J Med* 2004;351:1448-54.
2. Doran T, Fullwood C, Gravelle H, et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006;355:375-84.
3. Campbell S, Steiner A, Robison J, Webb D, Raven A, Roland M. Is the quality of care in general medical practice improving? Results of a longitudinal observational study. *Br J Gen Pract* 2003;53:298-304.
4. Campbell SM, Roland MO, Middleton E, Reeves D. Improvements in the quality of clinical care in English general practice

- 1998-2003: longitudinal observational study. *BMJ* 2005;331:1121-3.
5. Campbell S, Wilkin D, Roland M. Primary care groups: improving quality of care through clinical governance. *BMJ* 2001;322:1580-2.
  6. Campbell SM, Roland MO. The experience of the United Kingdom. In: Garcia-Pena C, Munoz O, Duran L, Vazquez F, eds. *Family medicine at the dawn of the 21st century*. Mexico City: CAMS, 2005:391-416.
  7. Baker D, Middleton E. Cervical screening and health inequality in England in the 1990s. *J Epidemiol Community Health* 2003;57:417-23.
  8. Middleton E, Baker D. Comparison of social distribution of immunisation with measles, mumps, and rubella vaccine, England, 1991-2001. *BMJ* 2003;326:854.
  9. Campbell SM, Hann M, Hacker J, et al. Identifying predictors of high quality care in English general practice: observational study. *BMJ* 2001;323:784-7.
  10. Campbell SM, Roland MO, Shekelle PG, Cantrill JA, Buetow SA, Cragg DK. Development of review criteria for assessing the quality of management of stable angina, adult asthma and non-insulin dependent diabetes mellitus in general practice. *Qual Health Care* 1999;8:6-15.
  11. Campbell SM, Hann M, Hacker J, Durie A, Thapar A, Roland MO. Quality assessment for three common conditions in primary care: validity and reliability of review criteria developed by expert panels for angina, asthma and type 2 diabetes. *Qual Health Care* 2002;11:125-30.
  12. Brook RH, Chassin MR, Fink A, Solomon DH, Koseoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *Int J Technol Assess Health Care* 1986;2:53-63.
  13. Collett D. *Modelling binary data*. London: Chapman and Hall, 1991.
  14. McElduff P, Lyratzopoulos G, Edwards R, Heller RF, Shekelle P, Roland M. Will changes in primary care improve health outcomes? Modelling the impact of financial incentives introduced to improve quality of care in the UK. *Qual Saf Health Care* 2004;13:191-7.
  15. Campbell S, Wilkin D. Clinical governance. In: Wilkin D, Coleman A, Dowling B, Smith K, eds. *The National Tracker Survey of primary care groups and trusts 2001/2002: taking responsibility?* Manchester, United Kingdom: University of Manchester, National Primary Care Research and Development Centre, University of Manchester, 2002. (Accessed June 21, 2007, at [http://www.npcrdc.ac.uk/Publications/TRACKER\\_REPORT\\_2002.pdf](http://www.npcrdc.ac.uk/Publications/TRACKER_REPORT_2002.pdf).)
  16. Asch SM, McGlynn EA, Hogan MM, et al. Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. *Ann Intern Med* 2004;141:938-45.
  17. Checkland K, Marshall M, Harrison S. Re-thinking accountability: trust versus confidence in medical practice. *Qual Health Care* 2004;13:130-5.
  18. Marshall M, Smith P. Rewarding results: using financial incentives to improve quality. *Qual Saf Health Care* 2003;12:397-8.
  19. Spooner A, Chapple A, Roland M. What makes British general practitioners take part in a quality improvement scheme? *J Health Serv Res Policy* 2001;6:145-50.
  20. Campbell S, Steiner A, Robison J, et al. Do Personal Medical Services contracts improve quality of care? A multi-method evaluation. *J Health Serv Res Policy* 2005;10:31-9.
  21. Gosden T, Forland F, Kristiansen IS, et al. Impact of payment system on behaviour of primary care physicians: a systematic review. *J Health Serv Res Policy* 2001;6:44-55.
  22. Epstein AM, Lee TH, Hamel MB. Paying physicians for high-quality care. *N Engl J Med* 2004;350:406-10.
  23. McGlynn EA. Intended and unintended consequences: what should we really worry about? *Med Care* 2007;45:3-5.
  24. Smith P. On the unintended consequences of publishing performance data in the public sector. *Int J Publ Admin* 1995;18:277-310.
  25. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol Bull* 1999;125:627-68.
  26. Gagné M, Deci EL. Self-determination theory and work motivation. *J Organ Behav* 2005;26:331-62.
  27. Roland MO, Campbell SM, Bailey N, Whalley D, Sibbald B. Financial incentives to improve the quality of primary care in the UK: predicting the consequences of change. *Prim Health Care Res Dev* 2006;7:18-26.
  28. Whalley D, Bojke C, Gravelle H, Sibbald B. GP job satisfaction in view of contract reform: a national survey. *Br J Gen Pract* 2006;56:87-92.
  29. Ganz DA, Wenger NS, Roth CP, et al. The effect of a quality improvement initiative on the quality of other aspects of health care: the law of unintended consequences? *Med Care* 2007;45:8-18.
  30. Fisher ES, Davis K. Pay for performance — recommendations of the Institute of Medicine. *N Engl J Med* 2006 (Web only). (Available at <http://content.nejm.org/cgi/content/full/NEJMp068216/DC1>.)
  31. Committee on Redesigning Health Insurance Performance Measures, Payment, and Performance Improvement Programs. *Rewarding provider performance: aligning incentives in Medicare*. Washington, DC: National Academies Press, 2007. (Accessed June 21, 2007, at <http://www.nap.edu/catalog/11723.html>.)
  32. Galvin R. Pay-for-performance: too much of a good thing? A conversation with Martin Roland. *Health Aff (Millwood)* 2006;25:w412-w419.
  33. GP pay raise 'was mistake.' *BBC News Health*. (Accessed June 21, 2007, at [http://news.bbc.co.uk/player/nol/newsid\\_6280000/newsid\\_6280000/6280077.stm?bw=bb&mp=rm](http://news.bbc.co.uk/player/nol/newsid_6280000/newsid_6280000/6280077.stm?bw=bb&mp=rm).)
  34. British Medical Association. *Quality and outcomes framework guidance: summary of indicators — clinical domain*. (Accessed June 21, 2007, at <http://www.bma.org.uk/ap.nsf/Content/qof06~summclinical#SecondaryPreventionofCoronary>.)

Copyright © 2007 Massachusetts Medical Society.