Open access • Journal Article • DOI:10.1056/NEJMSA0807651

Effects of Pay for Performance on the Quality of Primary Care in England

Stephen Campbell, David Reeves, Evangelos Kontopantelis, Bonnie Sibbald ...+1 more authors

Institutions: University of Manchester

Published on: 10 Dec 2009 - The New England Journal of Medicine (Massachusetts Medical Society)

Topics: Health care, Quality and Outcomes Framework, Pay for performance and Medical record

SPECIAL ARTICLE

Effects of Pay for Performance on the Quality of Primary Care in England

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

ABSTRACT

BACKGROUND

A pay-for-performance scheme based on meeting targets for the quality of clinical care was introduced to family practice in England in 2004.

METHODS

We conducted an interrupted time-series analysis of the quality of care in 42 representative family practices, with data collected at two time points before implementation of the scheme (1998 and 2003) and at two time points after implementation (2005 and 2007). At each time point, data on the care of patients with asthma, diabetes, or coronary heart disease were extracted from medical records; data on patients' perceptions of access to care, continuity of care, and interpersonal aspects of care were collected from questionnaires. The analysis included aspects of care that were and those that were not associated with incentives.

RESULTS

Between 2003 and 2005, the rate of improvement in the quality of care increased for asthma and diabetes ($P<0.001$) but not for heart disease. By 2007, the rate of improvement had slowed for all three conditions ($P<0.001$), and the quality of those aspects of care that were not associated with an incentive had declined for patients with asthma or heart disease. As compared with the period before the pay-for-performance scheme was introduced, the improvement rate after 2005 was unchanged for asthma or diabetes and was reduced for heart disease ($P=0.02$). No significant changes were seen in patients' reports on access to care or on interpersonal aspects of care. The level of the continuity of care, which had been constant, showed a reduction immediately after the introduction of the pay-for-performance scheme ($P<0.001$) and then continued at that reduced level.

CONCLUSIONS

Against a background of increases in the quality of care before the pay-for-performance scheme was introduced, the scheme accelerated improvements in quality for two of three chronic conditions in the short term. However, once targets were reached, the improvement in the quality of care for patients with these conditions slowed, and the quality of care declined for two conditions that had not been linked to incentives. Continuity of care was reduced after the introduction of the scheme.

From the National Primary Care Research and Development Centre, University of Manchester, Manchester (S.M.C., D.R., E.K., B.S., M.R.); and the University of Cambridge General Practice and Primary Care Research Unit, Institute of Public Health, Cambridge (M.R.) — both in the 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, or at stephen.campbell@manchester.ac.uk.

N Engl J Med 2009;361:368-78.

In 2004, the U.K. government introduced a pay-for-performance scheme with 136 indicators for family practices. The indicators covered the management of chronic disease, practice organization, and patients' experiences with respect to care.[1] In 2006, revisions to the scheme added seven new clinical areas, including dementia and chronic kidney disease, and two new indicators of patient access to care (see the Supplementary Appendix, available with the full text of this article at NEJM.org).[2] Payments make up approximately 25% of family practitioners' income, and 99.6% of family practitioners participated in the pay-for-performance scheme, which is voluntary.

We have previously reported on the quality of clinical care in 2005, the year after the pay-for-performance scheme was introduced.[3] We found a modest acceleration in the rate of improvement in the quality of care for asthma and diabetes but not for heart disease. There had been rapid improvement in the quality of care for all three conditions before the introduction of pay for performance. This article extends these analyses to include performance data in 2007. We used an interrupted time-series analysis to examine trends in the quality of clinical care from 1998 through 2007, a period spanning the introduction of pay for performance. We also report on trends in patient reports on communication with their physician, on access to care, and on continuity of care across the same period.

METHODS

DATA COLLECTION

Trained research staff abstracted clinical data from the medical records kept by 42 nationally representative family practices. In each practice, data were collected for nonoverlapping random samples of patients (20 in 1998 and 12 each in 2003, 2005, and 2007) who had heart disease, asthma, or diabetes; the data were collected with the use of quality indicators.[4,5] The methods used to collect data in 2007 were consistent with the methods used in 1998, 2003, and 2005.[3]

For patient evaluation, a version of the General Practice Assessment Questionnaire (www.gpaq.info) was mailed, with one follow-up reminder, to a random sample of 200 registered adult patients (age, ≥18 years) in each practice.[6,7] Rapid access to any doctor within 48 hours was associated with an incentive under the pay-for-performance scheme, and our questionnaire included two items addressing the patient's ability to get an appointment within 48 hours with "any doctor" and with "a particular doctor." Because of concern that the scheme's focus on clinical indicators might lead practitioners to neglect other aspects of care,[8] we also analyzed communication with physicians and continuity of care. Communication was assessed by asking seven questions, with the answers scored on a six-point scale ranging from "very poor" to "excellent"; continuity of care was assessed with the use of the same six-point scale and a single question: "How often do you see your usual doctor?" All scores were rescaled to range from 0 to 100. The rate of response to the survey was 38% in 1998, 47% in 2003, 45% in 2005, and 38% in 2007. In all cases, higher scores indicate higher quality of care. The research protocol was approved by the North West Research Ethics Committee.

STATISTICAL ANALYSIS

As we had done previously,[3,9,10] we computed an overall clinical quality score for each patient in 1998, 2003, 2005, and 2007, which was based on the number of indicators for which appropriate care was provided, divided by the number of indicators relevant to that patient. This score represents the percentage, from 0 to 100%, of "necessary" or "indicated" care provided to the patient. Practice-level quality scores were computed as the mean of individual patient scores in each practice. We computed separate quality scores for the subgroups of indicators that were assigned incentives under the pay-for-performance scheme and for the subgroups that were not assigned incentives.

Data on quality of care had been collected in the same practices in 1998.[9] When a pay-for-performance scheme was announced for commencement in 2004, we designed an interrupted time-series study whereby data on quality of care would be collected at two points before the scheme was introduced (1998 and 2003) and at two points after its introduction (2005 and 2007). We use the term "pre-introduction period" to refer to the period from 1998 through 2003, "introduction period" for 2003 through 2005 (from the year before to the year after the implementation of pay for performance), and "post-introduction period" for 2005 through 2007.

We analyzed the data as an interrupted, or

segmented, time series. In this model, the within-practice variation was partitioned into three main components to provide independent tests of the slope in scores for the pre-introduction period (test 1); the change in level during the introduction period, allowing for the trend before pay for performance (test 2); and the change in slope from before to after pay for performance (test 3).[11] Practice was treated as a random effect, and robust standard-error estimates were used (see the Supplementary Appendix).

The analysis for each outcome measure was conducted in two steps. In step 1, we used the interrupted time-series analysis to look for evidence that pay for performance was having an effect on the trend in scores over time, as indicated by a statistically significant result with respect to either the change in level or the change in slope (tests 2 and 3). The results of these tests determined step 2: if the results of neither test were significant, there was no evidence that pay for performance had affected the preexisting trend and we conducted no further analyses; if the results of either test were significant, there was evidence of an effect and we investigated this further by using the coefficients from the time-series analysis to compare the immediate- and long-term effects of the scheme (i.e., compare the slope during the introduction period with the slope during the post-introduction period) and to estimate the size of the effect on mean quality scores in 2005 and 2007.

We compared the trends in quality scores for the subgroups of indicators associated with incentives and indicators not associated with incentives by means of interactions between indicator set and the changes in level and slope (as defined above) within a regression analysis. If either interaction was significant, we took this as evidence that the trends varied by indicator set and next tested the interaction between indicator set and the change in slope from the introduction period to the post-introduction period.

The quality scores based on medical records and those based on patient evaluation are subject to ceilings of 100%, and many practices achieved this level on at least one indicator. The ceiling necessarily limits any linear trend in improvement, since a score on quality cannot exceed 100%. Analyses were therefore conducted on scores transformed to a logit scale, which has no ceiling, as described previously.[3] The transformation increases the weight given to score changes near the ceiling or floor — for example, score changes from 97 to 98% and from 55 to 65% are numerically equivalent (0.41) after transformation. However, where possible, results are re-expressed in original units to facilitate interpretation.

To assess the sensitivity of the findings to our statistical assumptions, we varied the method of statistical inference with the use of a bootstrap method, using 1000 bootstrap samples, and we assumed a linear model for the trend by repeating the analysis on untransformed scores (for details see the Supplementary Appendix). We report any results that differ from those of the primary analysis.

Table 1. Mean Clinical-Quality Scores for 42 Family Practices in 1998, 2003, 2005, and 2007.*

Variable	Mean Clinical-Quality Score			
	1998	2003	2005	2007
Clinical care				
Coronary heart disease†	58.6±1.4	76.2±1.6	85.0±1.0	84.8±1.3
Asthma	60.2±2.5	70.3±2.5	84.3±1.8	85.0±1.4
Diabetes	61.6±1.8	70.4±1.5	81.4±0.8	83.7±0.7
Patients' perceptions				
Communication with physicians	69.4±1.0	70.5±1.4	69.1±1.6	71.3±1.2
Access to care (appointment within 48 hr)				
To see a particular physician	39.0±4.3	33.3±4.0	34.4±3.9	32.1±3.2
To see any physician	67.2±3.3	61.0±3.7	63.9±3.2	64.2±3.2
Continuity of care	70.7±1.7	70.3±1.7	66.2±1.8	66.0±1.6

* Plus–minus values are means ±SE. Data on the care of patients with asthma, diabetes, or coronary heart disease were extracted from medical records, and data on patients' perceptions of communication with physicians, access to care, and continuity of care were obtained from questionnaires. Communication was assessed by asking seven questions, with the answers scored on a six-point scale ranging from "very poor" to "excellent"; continuity of care was assessed with the use of the same six-point scale and a single question: "How often do you see your usual doctor?" Access to care was scored as the percentage of patients who reported that they were able to get an appointment within 48 hours. All scores were rescaled to range from 0 to 100.

† Scores are shown for 40 of the 42 practices.

RESULTS

CORONARY HEART DISEASE

The quality of care for coronary heart disease had been improving before the pay-for-performance incentives were introduced (Tables 1 and 2 and Fig. 1). The rate of increase was equivalent to an average of 3.5% per annum from 1998 through

2003 (95% confidence interval [CI], 2.8 to 4.2; P<0.001). In 2005, after the introduction of pay for performance, scores on quality rose slightly, but not significantly, higher than expected, as compared with the trend before the introduction of pay for performance (P=0.06). Subsequently, the rate of improvement dropped below the improvement rates for both the pre-introduction period (P=0.02) and the introduction period (P=0.001), and the overall quality score in 2007 (84.8; 95% CI, 82.2 to 87.4) was similar to that in 2005 (85.0; 95% CI, 83.0 to 87.1) (Tables 2 and 3 and Fig. 1).

ASTHMA

The quality of care for asthma was improving during the pre-introduction period, at an average rate of 2.0% per annum (95% CI, 0.9 to 3.1; P<0.001), and there was a significant change in the level of quality over and above this trend in 2005 (P<0.001) (Tables 1 and 2 and Fig. 1). However, this accelerated rate of increase was not maintained after 2005 (P=0.001). The trend after 2005 did not differ significantly from the trend before the introduction of pay for performance (P=0.16), although in absolute terms, overall quality hardly changed between 2005 (84.3; 95% CI, 80.6 to 88.1) and 2007 (85.0; 95% CI, 82.2 to 87.8) (Tables 2 and 3 and Fig. 1).

DIABETES

The quality of care for patients with diabetes was improving in the pre-introduction period, at an average rate of 1.8% per annum (95% CI, 1.1 to 2.4; P<0.001) (Tables 2 and 3 and Fig. 1). Diabetes care, like asthma care, showed a significant change in the level of improvement after the introduction of pay for performance that was well above the preexisting trend (P<0.001). As with asthma care, this accelerated rate of improvement was not maintained after 2005 (P<0.001); instead, the rate fell back to the pre-introduction level (P=0.91) (Tables 2 and 3 and Fig. 1).

EFFECT OF INCENTIVES ON QUALITY SCORES

Mean quality scores for aspects of care that were linked to incentives were higher than those for care that was not linked to incentives, and this pattern applied to all conditions at all four time points (Fig. 2). Allowing for these overall differences, there were further differences over time in the scores for aspects of care that were linked to

Table 2. Summary of Interrupted Time-Series Analysis.*

Variable	Rate of Improvement in Clinical-Quality Score, 1998 to 2003 (95% CI)‡	P Value	Interrupted Time-Series Analysis: Change in Level — 2005 Score minus Predicted Score Based on 1998–2003 Trend (95% CI)	P Value	Change in Rate of Improvement from 1998–2003 to 2005–2007 (95% CI)‡	P Value	Subsequent Test†: Change in Rate of Improvement from 2003–2005 to 2005–2007 (95% CI)‡	P Value
Clinical care								
Coronary heart disease	0.177 (0.141 to 0.213)	<0.001	0.216 (−0.010 to 0.441)	0.06	−0.143 (−0.260 to −0.025)	0.02	−0.250 (−0.401 to 0.100)	0.001
Asthma	0.116 (0.056 to 0.177)	<0.001	0.680 (0.283 to 1.078)	0.001	−0.128 (−0.305 to 0.050)	0.16	−0.468 (−0.748 to −0.187)	0.001
Diabetes	0.081 (0.053 to 0.109)	<0.001	0.445 (0.278 to 0.613)	<0.001	0.003 (−0.046 to 0.051)	0.91	−0.220 (−0.313 to −0.127)	<0.001
Patients' perceptions								
Communication with physicians	0.012 (−0.009 to 0.032)	0.26	−0.082 (−0.177 to 0.013)	0.09	0.039 (−0.004 to 0.082)	0.08	NA	NA
Access to care — appointment within 48 hr								
To see a particular physician	−0.059 (−0.124 to 0.006)	0.08	0.216 (−0.104 to 0.536)	0.19	−0.003 (−0.142 to 0.136)	0.97	NA	NA
To see any physician	−0.049 (−0.113 to 0.015)	0.14‡	0.187 (−0.185 to 0.559)	0.32	0.022 (−0.134 to 0.177)	0.79	NA	NA
Continuity of care	−0.006 (−0.027 to 0.015)	0.57	−0.191 (−0.286 to −0.096)	<0.001	−0.004 (−0.047 to 0.039)	0.85	0.091 (0.025 to 0.157)	0.007

* Communication was assessed by asking seven questions, with the answers scored on a six-point scale ranging from "very poor" to "excellent"; continuity of care was assessed with the use of the same six-point scale and a single question: "How often do you see your usual doctor?" Access to care was scored as the percentage of patients who reported that they were able to get an appointment within 48 hours. All scores were rescaled to range from 0 to 100. Scores and annual rates of improvement were measured on the logit scale. NA denotes not analyzed. Variables that were not analyzed did not meet the criterion for significant change specified in the statistical analysis plan, as stated in the Methods section.

† The subsequent test was applied only for outcomes with a significant change in level or slope.

‡ P=0.03 in the linear model.

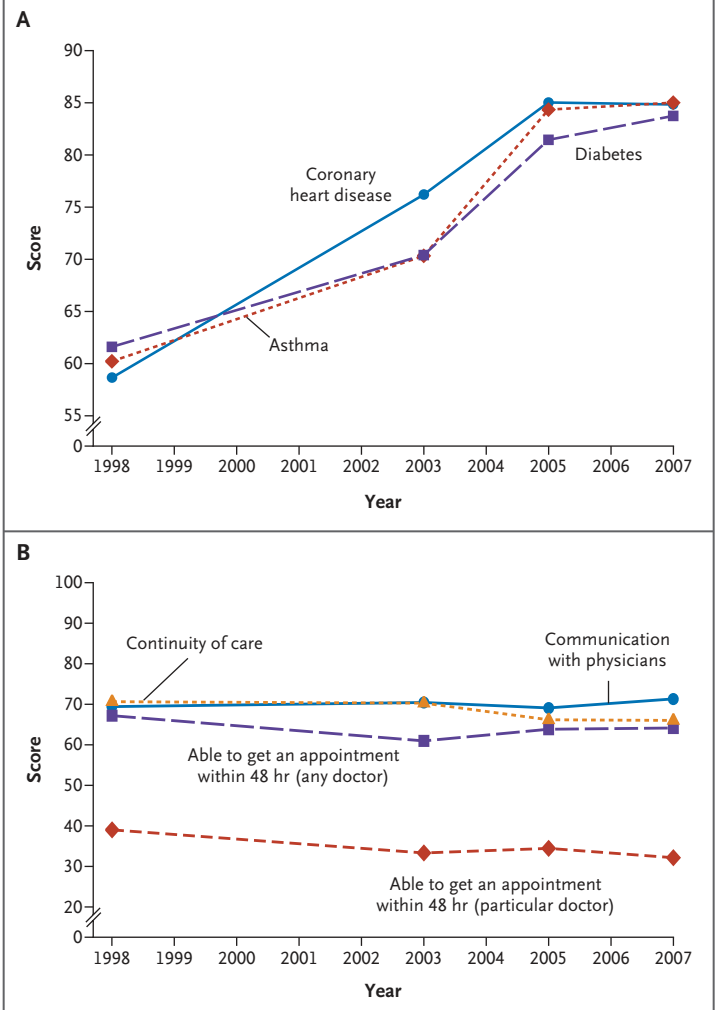

Figure 1. Mean Scores for the Quality of Care at the Practice Level, 1998–2007.

Panel A shows scores for the quality of care provided for coronary heart disease, asthma, and diabetes. Quality scores range from 0% (no quality indicator was met for any patient) to 100% (all quality indicators were met for all patients). Panel B shows scores for patients' perceptions of communication with physicians, access to care, and continuity of care. Communication was assessed by asking seven questions, with the answers scored on a six-point scale ranging from "very poor" to "excellent"; continuity of care was assessed with the use of the same six-point scale and a single question: "How often do you see your usual doctor?" Access to care was scored as the percentage of patients who reported that they were able to get an appointment within 48 hours. All scores were rescaled to range from 0 to 100.

incentives as compared with those that were not. For heart disease, the scores for aspects of care that were linked to incentives showed a bigger immediate increase when the pay-for-performance system was introduced (P=0.05), although this trend was not significant as calculated in the linear model (P=0.46). The long-term trends (scores in the post-introduction period vs. scores in the pre-introduction period) did not differ significantly (P=0.06). However, the difference was significant when calculated with the use of the bootstrapping method (P=0.05) or the linear model (P=0.03), and in absolute terms, the mean quality score for aspects of care for heart disease that were not linked to incentives declined after 2005, whereas the quality score for care that was linked to incentives increased. For asthma, the immediate effect of pay for performance did not differ between care that was and care that was not linked with incentives (P=1.00), but the trends subsequently diverged (post-introduction period vs. pre-introduction period, P=0.006; post-introduction period vs. introduction period, P=0.05), with the mean score for care that was not linked to incentives declining after 2005, and the mean score for care that was linked to incentives increasing. Trends in diabetes care did not differ at any time according to whether the care was linked to incentives.

COMMUNICATION, WAITING TIMES, AND CONTINUITY OF CARE

The percentages of patients able to see a physician within 48 hours, as well as the mean scores on the physician-communication scale, showed no significant changes in trend. Continuity of care declined significantly after the introduction of pay for performance (P<0.001) and remained at this lower level (Tables 1 and 2 and Fig. 2).

ESTIMATED OVERALL EFFECT OF PAY FOR PERFORMANCE

For outcomes in which there was evidence that pay for performance altered the trend in quality improvement, we used coefficients from the interrupted time-series analysis to compute estimates of the increase in scores beyond that expected from the trend in the pre-introduction period (back-transforming the results from the logit analysis, with estimated 95% confidence limits). As compared with the expected level of improvement based on the pre-introduction trend, the pay-for-performance scheme was associated with an improvement in the quality of care for diabetes of 7.5 percentage points in 2005 (95% CI, 4.7 to 10.4) and 6.9 percentage points in 2007 (95% CI, 3.8 to 10.0). For asthma, the increase in quality potentially attributable to pay for performance was 9.4 percentage points in 2005 (95% CI, 3.9 to 15.0)

Table 3. Mean Clinical-Quality Scores in 1998, 2003, 2005, and 2007, According to Individual Clinical Indicators.*

Condition	1998		2003		2005		2007	
	Mean Score	No. of Practices	Mean Score	No. of Practices	Mean Score	No. of Practices	Mean Score	No. of Practices
Coronary heart disease								
Clinical indicators								
Frequency or pattern of angina attacks (past 15 mo)	43.4±2.2	40	60.3±3.9	40	70.1±3.4	40	59.7±4.4	40
Blood pressure (past 15 mo)†	87.0±1.7	40	95.3±1.2	40	98.8±0.5	40	99.0±0.5	40
Exercise capacity or activity status (past 15 mo)	36.5±3.3	40	66.3±4.6	40	72.8±3.7	40	64.5±4.9	40
Cholesterol level (past 5 yr)†	63.5±2.6	40	89.7±2.0	40	98.7±0.5	40	99.4±0.4	40
Dietary advice (past 5 yr)	58.8±3.9	40	75.7±4.0	40	87.4±2.8	40	83.5±3.4	40
Smoking status (past 5 yr)†	85.1±1.9	40	86.6±2.8	40	98.3±0.6	40	99.2±0.5	40
Referral to specialist for exercise electrocardiogram (ever)†	76.2±2.0	40	88.2±1.9	40	89.2±1.9	40	93.8±1.4	40
Blood pressure controlled (≤150/90 mm Hg)†	47.9±2.2	40	72.2±2.1	40	81.5±1.8	40	82.6±2.4	40
Serum cholesterol controlled (≤190 mg/dl)†	16.9±1.8	40	60.7±3.3	40	72.9±2.9	40	79.5±2.2	40
Aspirin prescribed or recommended unless record of contraindication or intolerance†	76.2±1.9	40	80.7±2.4	40	92.5±1.5	40	88.9±1.9	40
Conditional factors‡								
Advice given to smokers (past 5 yr)†	67.3±4.8	37	77.9±6.0	35	97.9±1.6	35	95.7±2.1	29
Weight advice for overweight patients (past 5 yr)	63.7±40	40	78.8±40	40	82.5±40	40	90.0±40	40
Action taken on blood pressure if systolic pressure >160 mm Hg or if >140 mm Hg with cholesterol level >190 mg/dl†	35.7±5.8	35	60.3±7.7	29	67.5±9.8	20	78.1±9.1	16
Action taken if cholesterol >210 mg/dl (for patients <70 yr of age)†	55.1±5.9	36	76.1±6.4	23	71.7±9.3	23	84.6±10.4	13
Prescribed beta-blocker as maintenance therapy if sole therapy†	48.2±5.0	37	58.4±5.4	39	80.8±3.5	39	68.7±4.5	39
Asthma								
Clinical indicators								
Normal or predicted peak flow or record of difficulty using meter (past 5 yr)	76.1±2.9	42	85.7±2.3	42	91.6±1.9	42	95.6±1.2	42
Record of daily, nocturnal, or activity-limiting symptoms (past 15 mo)	43.7±3.3	42	58.1±4.2	42	76.5±2.8	42	71.6±3.9	42
Smoking status (past 5 yr)†	82.6±2.9	42	84.3±2.6	42	96.0±1.1	42	98.8±0.5	42
Inhaler technique (past 5 yr)	53.1±4.1	42	58.7±4.5	42	76.0±4.1	42	85.7±2.8	42
Conditional factors‡								
Advice given to smokers (past 5 yr)†	62.8±5.2	41	86.8±3.7	39	96.9±1.6	41	98.9±0.8	40
Self-management plan for patients who are receiving high-dose corticosteroids or who have had inpatient treatment for asthma (past 5 yr)	34.6±5.7	41	50.5±7.9	32	84.0±6.9	25	89.4±6.8	16

The New England Journal of Medicine

For patients with recorded exercise-induced bronchospasm, short-acting bronchodilators prescribed for use before exercise (past 5 yr)	40.5±6.4	36	84.3±6.2	26	94.7±5.3	19	80.7±8.0	18	
Peak flow during a consultation for an exacerbation (most recent exacerbation)	72.2±5.1	40	72.6±6.2	34	82.7±4.7	33	68.6±6.5	33	
Oral corticosteroids prescribed if peak flow <60% of normal value (most recent exacerbation)	88.8±4.7	27	94.4±5.6	9	100.0±0.0	10	84.6±10.4	13	
Action taken if patient has nocturnal symptoms (past 5 yr)	62.7±4.9	29	76.2±5.1	38	91.8±3.6	36	83.3±4.4	38	
Action taken if patient has activity-limiting symptoms (past 5 yr)	56.4±7.7	26	78.6±6.9	23	100.0±0.0	23	58.8±7.1	27	
Referral to respiratory physician when oral corticosteroids are used in maintenance treatment (past 5 yr)	47.1±11.7	17	53.3±22.6	5	33.3±33.3	3	16.7±16.7	6	
Diabetes									
Clinical indicators									
Glycated hemoglobin (past 15 mo)†	87.1±2.4	42	93.1±1.3	42	99.4±0.3	42	98.8±0.5	42	
Visual examination of feet (past 15 mo)	57.4±4.6	42	69.6±3.7	42	88.0±2.8	42	91.5±2.0	42	
Recording of peripheral pulses or record of ability to feel vibration (past 15 mo)†	60.0±4.6	42	62.9±4.2	42	90.2±2.1	42	91.3±2.0	42	
Serum creatinine (past 15 mo)†	79.6±2.9	42	89.9±2.1	42	96.4±1.0	42	99.0±0.8	42	
Proteinuria (past 15 mo)†	66.3±4.0	42	74.8±3.7	42	82.8±3.6	42	89.8±1.9	42	
Examination of fundi or visual acuity (past 15 mo)†	69.4±3.2	42	72.2±3.2	42	82.7±2.3	42	81.1±3.7	42	
Weight (past 15 mo)†	80.2±2.8	42	86.5±2.9	42	97.2±0.8	42	99.6±0.3	42	
Blood pressure (past 15 mo)†	92.6±1.3	42	95.8±1.8	42	99.0±0.4	42	99.0±0.5	42	
Smoking status (past 5 yr)†	86.5±2.5	42	88.5±2.8	42	98.4±0.7	42	99.6±0.3	42	
Serum cholesterol (past 5 yr)†	75.1±3.1	42	97.6±0.8	42	99.4±0.3	42	99.8±0.2	42	
Blood pressure controlled (≤140/85 mm Hg)†	21.8±1.9	42	35.4±2.6	42	49.0±3.1	42	51.6±2.5	42	
Serum cholesterol controlled (≤190 mg/dl)†	21.8±1.4	42	52.0±2.5	42	72.5±2.6	42	78.9±1.9	42	
Glycated hemoglobin controlled (≤7.4 g/dl)†	37.8±2.8	42	39.8±2.4	42	50.6±3.1	42	54.9±3.4	42	

Conditional factors‡								
Patient education provided within past 5 yr if diagnosis made during that period	85.2±3.2	41	88.0±3.4	41	100.0±0.0	42	99.0±0.7	42
Advice given to smokers (past 5 yr)†	67.1±5.6	40	79.9±5.7	36	95.8±3.1	36	100.0±0.0	34
Referral to a specialist if creatinine >200 mmol/liter (past 5 yr)	75.0±13.1	12	100.0±0.0	3	100.0±0.0	4	100.0±0.0	4
For patients <80 yr of age, treatment offered if average of past 3 readings shows diastolic pressure >100 mm Hg or systolic pressure >150 and diastolic pressure >90 mm Hg	63.3±7.1	32	82.2±7.8	22	84.4±9.1	15	60.7±13.0	14
If patient is being treated for hypertension and has proteinuria, ACE inhibitor provided	53.4±3.2	42	57.1±20.2	7	68.8±13.2	8	63.0±16.1	9
If patient is receiving ACE inhibitor, creatinine and potassium measured within 1 mo after start of treatment (past 5 yr)	37.2±4.8	41	30.3±5.0	42	39.7±4.4	41	34.3±5.5	41
Assessment for hypoglycemia symptoms if patient receiving sulfonylurea (past 15 mo)	18.2±3.1	42	8.1±1.8	42	7.8±1.5	42	10.6±2.1	42
For patients <70 yr of age, if most recent glycated hemoglobin value was >9 g/dl, therapeutic intervention to improve glycemic control offered	77.4±4.5	36	73.6±5.6	37	69.4±6.7	33	73.3±6.7	30

* Plus–minus values are means ±SE. Data on the care of patients with asthma, diabetes, or coronary heart disease were extracted from medical records. Scores are based on a scale of 0 to 100. To convert the values for cholesterol to millimoles per liter, multiply by 0.02586. To convert the values for creatinine to milligrams per deciliter, divide by 88.4. ACE denotes angiotensin-converting enzyme.
† This indicator or an equivalent clinical procedure was associated with an incentive in the pay-for-performance scheme.
‡ Conditional factors are indicators that did not apply to all patients.

and 5.5 percentage points in 2007 (95% CI, −1.0 to 12.1). For heart disease, pay for performance in 2005 was associated with a nonsignificant improvement in quality above the levels expected (2.8 percentage points; 95% CI, −0.1 to 5.8), and in 2007, it was associated with a nonsignificant reduction in quality from that expected (0.8 percentage points; 95% CI, −4.7 to 3.1). The results of patient evaluations of continuity of care were 4.1 percentage points lower than expected in 2005 (95% CI, −6.1 to −2.0) and 4.3 percentage points lower in 2007 (95% CI, −6.9 to −1.6).

DISCUSSION

We previously found that there were improvements in some aspects of clinical care over and above the underlying trend after the introduction of a pay-for-performance scheme.[3] Our current findings suggest that although these initial improvements were maintained, for two of the three conditions studied (heart disease and asthma), improvements in the quality of care reached a plateau a year after the scheme's introduction. Allowing for ceiling effects, care for diabetes continued to improve, but it did so at a rate equivalent to that of the rate in the pre-introduction period.

Within these overall trends for care, we found significant differences between aspects of care that were linked to incentives and aspects of care that were not linked to incentives. For asthma and heart disease, we found a significant difference in the effect of pay for performance on these two groups of quality indicators; for both conditions, mean quality scores for aspects of care that were not linked to incentives dropped between 2005 and 2007, whereas mean scores for aspects of care that were linked to incentives continued to increase. This widening gap in quality came on top of already lower levels of care for indicators not linked to incentives.

For all aspects of care — whether associated with incentives or not — and for all three conditions, rates of quality improvement slowed considerably after 2005. There are several possible explanations. The first is that near-maximal scores had been achieved. However, whereas achievement was high for some indicators (e.g., smoking status recorded for more than 98% of patients for all conditions), the logit transformation theoretically eliminates ceiling effects, and we ob-

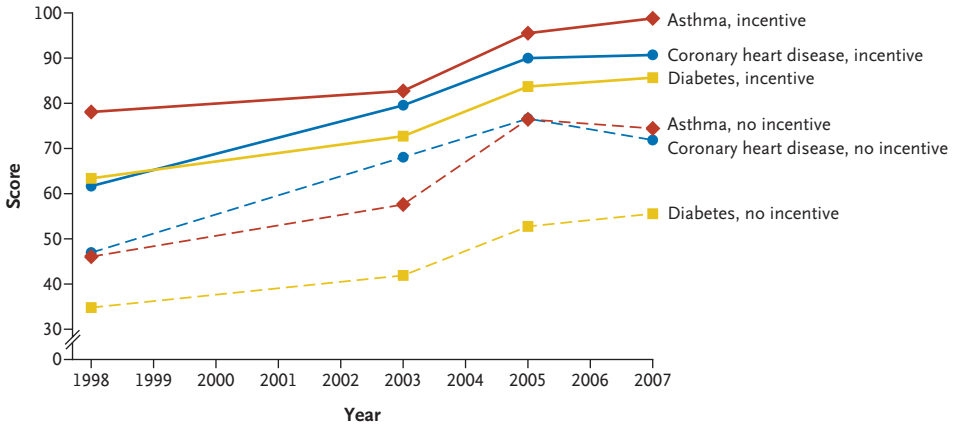

Figure 2. Mean Scores for Clinical Quality at the Practice Level for Aspects of Care for Coronary Heart Disease, Asthma, and Type 2 Diabetes That Were Linked with Incentives and Aspects of Care That Were Not Linked with Incentives, 1998–2007.

Quality scores range from 0% (no quality indicator was met for any patient) to 100% (all quality indicators were met for all patients).

served the same plateau effect for indicators reflecting lower levels of achievement. A second explanation is that once initial gains had been made, subsequent gains were more difficult to achieve. A third explanation is that the structure of the pay-for-performance scheme did not reward further improvement once targets had been attained. This explanation is supported by the fact that family practices in our study gained, on average, 96.9% of available clinical-quality payment points in 2005 and 97.8% in 2007 (which were similar to the average gains of 97.1% and 97.5%, respectively, for all family practices in England[12,13]) — that is, there was little financial incentive for further improvement. A fourth explanation is that family practitioners had sufficient income and had little personal motivation to improve performance and income further (the target-income hypothesis); this explanation would be consistent with the 30 to 40% gains in family practitioners' net income from the 2002–2003 period to the 2005–2006 period.[14]

Our data cannot be used to ascertain the relative merit of these explanations. However, government negotiators in England appear to endorse the third explanation (too many physicians achieving maximal or near-maximal payments for quality of care). Alterations of the pay-for-performance scheme in 2006 introduced higher thresholds for maximal clinical-quality payments and a wider range of indicators (see the Supplementary Appendix).

This study suggests that continuity of care declined after pay for performance was introduced. One possible explanation is that practices focused on meeting rapid-access targets in which access to any doctor in the practice within 48 hours was linked to incentives but access to a particular physician was not,[15] making it more difficult for patients to see their own doctor. This could be an unintended and perverse effect of the scheme and is a concern, since continuity is an aspect of family practice that patients value.[16] Another explanation is that there were increases in the size of practices, and many practices introduced nurse-led clinics for management of individual chronic diseases. Although this may have been an important part of improving the quality of care, it may have made continuity of care harder to achieve.

Other studies suggest that financial incentives result in small improvements in quality.[17,18] Our data suggest that the pay-for-performance scheme in England attained its quality-improvement goals but that the pace of improvement was not sustained once these goals had been reached. There may be unintended consequences for aspects of care other than those studied, which may be influenced by differences in the operational details of apparently similar incentive schemes.[19] An

unanticipated benefit of the scheme in England has been a reduction in sociodemographic inequalities in the delivery of health care.[20]

Our study has several limitations. First, the pay-for-performance scheme was introduced throughout the United Kingdom, thereby precluding a controlled trial and making the use of an interrupted time series the best evaluation method available. The only other time-series analysis of the quality of primary care in England suggests that pay for performance has had a more modest effect than that suggested by our results.[21] Second, because practices were observed at only two time points before the introduction of pay for performance, we cannot say whether the rate of improvement was already accelerating as a result of earlier but still ongoing initiatives. Third, the statistical power of our study was such that only moderate-to-large differences in trend were detectable between indicators that were and those that were not associated with incentives. Fourth, response rates for the patient questionnaire were poor (38 to 47%), although there is no reason to suspect any differences in bias at the four study time points. Finally, we focused on three diseases for which substantial efforts had been made to improve the quality of care before the introduction of the pay-for-performance scheme. Pay for performance might have a greater effect on conditions with lower profiles, including some introduced as the scheme developed (e.g., learning disabilities).

In conclusion, between 1998 and 2007, there were significant improvements in measurable aspects of clinical performance with respect to the care provided for three major chronic diseases. The initial acceleration in the underlying rate of quality improvement after the introduction of pay for performance was not sustained. If the aim of pay for performance is to give providers incentives to attain targets, the scheme achieved that aim. There may have been unintended consequences, including reductions in the quality of some aspects of care not linked to incentives and in the continuity of care.

Supported by the Department of Health for England.

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

The views presented here are those of the authors and not those of the Department of Health.

We thank the staff of all the practices that participated in the study and also Nan Bailey, Michele Bohan, Nicholas Burr, Ella Gaehl, Nicola Small, and Angela Swallow, who contributed to data collection.

REFERENCES

1. Roland M. Linking physicians' pay to quality of care — a major experiment in the United Kingdom. N Engl J Med 2004; 351:1448-54.
2. Quality and Outcomes Framework: guidance for GMS contract, 2009/10. London: NHS Employers, March 2009. (Accessed June 29, 2009, at http://www.nhsemployers.org/Aboutus/Publications/Documents/QOF_Guidance_2009_final.pdf.)
3. Campbell S, Reeves D, Kontopantelis E, Middleton E, Sibbald B, Roland M. Quality of primary care in England with the introduction of pay for performance. N Engl J Med 2007;357:181-90.
4. Campbell SM, Roland MO, Shekelle PG, Cantrill JA, Buetow SA, Cragg DK. The 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.
5. 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 Saf Health Care 2002;11:125-30.
6. Ramsay J, Campbell JL, Schroter S, Green J, Roland M. The General Practice Assessment Survey (GPAS): tests of data quality and measurement properties. Fam Pract 2000;17:372-9.
7. Mead N, Bower P, Roland M. The General Practice Assessment Questionnaire (GPAQ) — development and psychometric characteristics. BMC Fam Pract 2008;9:13. (Accessed June 29, 2009, at http://www.biomedcentral.com/1471-2296/9/13.)
8. Mangin D, Toop L. The Quality and Outcomes Framework: what have you done to yourselves? Br J Gen Pract 2007;57:435-7.
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, Middleton E, Reeves D. Improvements in the quality of clinical care in English general practice 1998-2003: longitudinal observational study. BMJ 2005;331:1121-5.
11. Wagner AK, Soumerai SB, Zhang F, Ross-Degnan D. Segmented regression analysis of interrupted time series studies in medication use research. J Clin Pharm Ther 2002;27:299-309.
12. National Quality and Outcomes Framework statistics for England 2005/06. London: National Health Service Information Centre, 2006. (Accessed June 29, 2009, at http://www.ic.nhs.uk/webfiles/publications/qof/NationalQualityOutcomesFramework280906_WORD.doc.)
13. National Quality and Outcomes Framework statistics for England 2007/08. London: National Health Service Information Centre, 2008. (Accessed June 29, 2009, at http://www.ic.nhs.uk/statistics-and-data-collections/supporting-information/audits-and-performance/the-quality-and-outcomes-framework/qof-2007/08/bulletin.)
14. GP Earnings and Expenses Enquiry 2005/06: final report. London: National Health Service Information Centre, 2008. (Accessed June 29, 2009, at http://www.ic.nhs.uk/statistics-and-data-collections/primary-care/general-practice/gp-earnings-and-expenses-enquiry-2005-2006:-final-report.)
15. Goodall S, Montgomery A, Banks J, Salisbury C, Sampson F, Pickin M. Implementation of Advanced Access in general practice: postal survey of practices. Br J Gen Pract 2006;56:918-23.
16. Cheraghi-Sohi S, Hole AR, Mead N, et

al. What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. Ann Fam Med 2008;6:107-15.
17. Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. JAMA 2005;294:1788-93.
18. Lindenauer PK, Remus D, Roman S, et al. Public reporting and pay for performance in hospital quality improvement. N Engl J Med 2007;356:486-96.
19. McDonald R, Roland M. Pay for performance in primary care in England and California: comparison of unintended consequences. Ann Fam Med 2009;7:121-7.
20. Doran T, Fullwood C, Kontopantelis E, Reeves D. Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework. Lancet 2008;372:728-36.
21. QRESEARCH and The Information Centre. Time series analysis for selected clinical indicators from the Quality and Outcomes Framework 2001-2006. Nottingham, United Kingdom: QRESEARCH, 2007. (Accessed June 29, 2009, at http://www.qresearch.org/Public_Documents/Time%20Series%20Analysis%20for%20selected%20clinical.pdf.)

POSTING PRESENTATIONS AT MEDICAL MEETINGS ON THE INTERNET

Posting an audio recording of an oral presentation at a medical meeting on the Internet, with selected slides from the presentation, will not be considered prior publication. This will allow students and physicians who are unable to attend the meeting to hear the presentation and view the slides. If there are any questions about this policy, authors should feel free to call the *Journal*'s Editorial Offices.