

SPECIAL ARTICLE

Quality of Care in the United Kingdom after Removal of Financial Incentives

Mark Minchin, M.B.A., Martin Roland, D.M., Judith Richardson, M.P.H.,
Shaun Rowark, M.Sc., and Bruce Guthrie, Ph.D.

ABSTRACT

BACKGROUND

The benefits of pay-for-performance schemes in improving the quality of care remain uncertain. There is little information on the effect of removing incentives from existing pay-for-performance schemes.

METHODS

We conducted interrupted time-series analyses of electronic medical record (EMR) data from 2010 to 2017 for 12 quality-of-care indicators in the United Kingdom's Quality and Outcomes Framework for which financial incentives were removed in 2014 and 6 indicators for which incentives were maintained. We estimated the effects of removing incentives on changes in performance on quality-of-care measures.

RESULTS

Complete longitudinal data were available for 2819 English primary care practices with more than 20 million registered patients. There were immediate reductions in documented quality of care for all 12 indicators in the first year after the removal of financial incentives. Reductions were greatest for indicators related to health advice, with a reduction of 62.3 percentage points (95% confidence interval [CI], -65.6 to -59.0) in EMR documentation of lifestyle counseling for patients with hypertension. Changes were smaller for indicators involving clinical actions that automatically update the EMR, such as laboratory testing, with a reduction of 10.7 percentage points (95% CI, -13.6 to -7.8) in control of cholesterol in patients with coronary heart disease and 12.1 percentage points (95% CI, -13.6 to -10.6) for thyroid-function testing in patients with hypothyroidism. There was little change in performance on the 6 quality measures for which incentives were maintained.

CONCLUSIONS

Removal of financial incentives was associated with an immediate decline in performance on quality measures. In part, the decline probably reflected changes in EMR documentation, but declines on measures involving laboratory testing suggest that incentive removal also changed the care delivered.

From the National Institute for Health and Care Excellence (NICE), Manchester (M.M., J.R., S.R.), the Department of Public Health and Primary Care, University of Cambridge, Cambridge (M.R.), and the Population Health and Genomics Division, University of Dundee, Dundee (B.G.) — all in the United Kingdom. Address reprint requests to Dr. Guthrie at the Population Health and Genomics Division, University of Dundee, Mackenzie Bldg., Kirsty Semple Way, Dundee DD2 4BF, United Kingdom, or at b.guthrie@dundee.ac.uk.

N Engl J Med 2018;379:948-57.

DOI: 10.1056/NEJMsa1801495

Copyright © 2018 Massachusetts Medical Society.

PAY-FOR-PERFORMANCE SCHEMES ARE IN-creasingly used by health care payers to support improvement in the quality of care and have become widespread in many health systems, including those in the United States, the United Kingdom, Canada, Germany, Israel, Taiwan, Thailand, New Zealand, and Australia.¹⁻³ Despite widespread implementation, questions persist about the ability of these schemes to improve the quality of care or patient outcomes,⁴⁻⁶ and there is ongoing debate about the sustainability of performance once financial incentives are removed.⁷⁻¹¹

In 2004, the United Kingdom's National Health Service (NHS) implemented the Quality and Outcomes Framework (QOF). At the time, the QOF was the world's largest pay-for-performance health care scheme, accounting for around 25% of the income of family practices in the United Kingdom.^{12,13} The number of quality-of-care indicators and the proportion of income dependent on pay for performance have been reduced over time. Scotland abolished the QOF altogether in 2016, and reductions in the scope of the QOF in England seem likely. The benefits and limitations of the QOF and other pay-for-performance schemes have been extensively discussed, including the premise that pay for performance is likely to work best when aligned with other improvement efforts such as public reporting.^{6,12,14-17} However, little is known about the effects of removing pay-for-performance incentives, and the few studies in this area have had conflicting results.^{8-11,18} There is therefore considerable uncertainty about whether any gain from pay for performance is sustained once financial incentives are removed. It is plausible that incentivized care becomes so standard and routine that the continued use of financial incentives is unnecessary, but it is equally plausible that any performance gains associated with pay-for-performance schemes are dependent on continued payment.⁸

As part of a regular review of the QOF, financial incentives were removed from 40 of the 121 quality-of-care indicators at the end of March 2014 (the end of the 2013–2014 fiscal year). Incentive payments were instead used to increase capitation payments in all practices. However, performance data for a number of these indicators continued to be collected and made publicly available (although without meaningful public reporting) at the individual-practice level as part

of the Indicators No Longer in QOF (INLIQ) data set. The aim of this study was to examine the effect of the removal of financial incentives for these indicators on overall documented quality of care, variation between practices, and differences in quality between practices serving more affluent areas and those serving less affluent areas.

METHODS

STUDY DESIGN

The design was an interrupted time-series analysis of observational data for 12 quality-of-care indicators for which incentives were removed and 6 indicators for which incentives were maintained. Methods are briefly described here, with complete details provided in the Supplementary Appendix, available with the full text of this article at NEJM.org. The study did not require formal ethical review because it involved analysis of only publicly available practice-level data that do not identify individuals.

DATA SOURCES

The QOF indicators and the INLIQ are defined nationally by NHS Digital and are extracted automatically from the electronic medical records (EMRs) of English family practices annually.¹⁹ We defined the socioeconomic status of the population served by each practice using the 2015 English Index of Multiple Deprivation score for the area where the practice was located.²⁰

INDICATOR SELECTION AND DEFINITION

The primary outcomes examined were 12 quality-of-care indicators for which financial incentives were removed at the end of March 2014. For these 12 indicators, the INLIQ database included a minimum of three annual measurements before and three after removal. (Three measurements are the minimum required for time-series analysis.²¹) There were no major changes in the definitions of these 12 indicators (Table 1, and Tables S1 and S2 in the Supplementary Appendix). We defined performance as the percentage of patients on the relevant disease register who were not excluded by automatic criteria such as recent practice registration and who received the specified care. The indicators that we were able to include covered a range of clinical processes and intermediate outcomes

Table 1. Quality-of-Care Indicators Included in the Analysis.*

Indicator	Indicator Type	Period with Incentives	Maximum Payment per Average-Size Practice (U.S. Dollars)†	Median No. of Patients Eligible for Care per Practice (IQR)‡
Incentives removed				
Smoking status documented in all adults	Clinical process	2010–2011 to 2013–2014	£1,726 (\$2,313)	5260 (3293–7992)
Diabetes retinopathy screening documented	Clinical process	2010–2011 to 2013–2014	£785 (\$1,051)	319 (205–474)
Glycated hemoglobin testing documented in patients with serious mental illness§	Clinical process	2011–2012 to 2013–2014	£785 (\$1,051)	29 (17–46)
Body-mass index documented in patients with serious mental illness§	Clinical process	2011–2012 to 2013–2014	£628 (\$841)	47 (29–76)
Thyroid function tested in patients with hypothyroidism	Clinical process	2010–2011 to 2013–2014	£942 (\$1,262)	192 (110–308)
Cholesterol tested in patients with stroke or TIA	Clinical process	2010–2011 to 2013–2014	£314 (\$421)	98 (51–164)
Cholesterol controlled in patients with stroke or TIA	Intermediate outcome	2010–2011 to 2013–2014	£785 (\$1,051)	62 (30–109)
Cholesterol controlled in patients with CHD	Intermediate outcome	2010–2011 to 2013–2014	£2,668 (\$3,575)	190 (106–306)
Patients with epilepsy documented as being seizure-free	Intermediate outcome	2010–2011 to 2013–2014	£942 (\$1,262)	36 (21–58)
Lifestyle advice documented in patients with hypertension	Health advice	2010–2011 to 2013–2014	£1,569 (\$2,103)	178 (111–275)
Preconception advice documented in patients with epilepsy	Health advice	2011–2012 to 2013–2014	£471 (\$631)	9 (5–15)
Advice about using long-acting, reversible contraceptives documented	Health advice	2010–2011 to 2013–2014	£471 (\$631)	333 (187–553)
Incentives maintained				
Blood pressure documented in patients with serious mental illness§	Clinical process	2011–2012 to 2016–2017	£628 (\$841)	47 (28–74)
Alcohol consumption documented in patients with serious mental illness§	Clinical process	2011–2012 to 2016–2017	£628 (\$841)	47 (28–74)
Cholesterol controlled in patients with diabetes	Intermediate outcome	2010–2011 to 2016–2017	£942 (\$1,262)	327 (203–492)
Blood pressure controlled in patients with CHD	Intermediate outcome	2010–2011 to 2016–2017	£2,668 (\$3,575)	206 (116–335)
Blood pressure controlled in patients with stroke or TIA	Intermediate outcome	2010–2011 to 2016–2017	£785 (\$1,051)	108 (55–179)
Smoking-cessation advice documented in smokers with chronic conditions¶	Health advice	2010–2011 to 2016–2017	£3,923 (\$5,257)	223 (138–349)

* CHD denotes coronary heart disease, IQR interquartile range, and TIA transient ischemic attack.

† Shown is the maximum payment in 2013–2014 for a practice with an average number of registered patients (6911), typically with approximately three to four full-time-equivalent physicians, and an average prevalence of each condition. U.S. dollar amounts are based on the exchange rate in December 2017.

‡ The median number of patients and IQR per practice are shown for the 2013–2014 period for practices included in the main analysis (total number of practices, depending on the indicator, 2723 to 2819 for indicators with incentives that were removed and 7187 to 7230 for indicators with incentives that were maintained).

§ Serious mental illness was defined as schizophrenia, bipolar affective disorder, or another psychotic disorder.

¶ Chronic conditions included coronary heart disease, peripheral arterial disease, stroke or TIA, hypertension, diabetes, chronic obstructive pulmonary disease, chronic kidney disease, asthma, and serious mental illness.

for diabetes, coronary heart disease (CHD), stroke or transient ischemic attack (TIA), epilepsy, hypothyroidism, and serious mental illness, as well as health-advice indicators (Table 1, and Table S1 in the Supplementary Appendix). For comparison, we examined changes in performance with respect to 6 QOF indicators for which incentives were maintained over the entire period of analysis, selected to cover the same processes, intermediate outcomes, and health advice as those for diabetes, CHD, stroke or TIA, serious mental illness, and smoking. The purpose of this comparison was to evaluate whether any observed changes in documented quality for indicators with incentives that were removed could have been due to other events, such as minor changes in indicator definitions over time and recent increases in workload pressure on family doctors in the United Kingdom (Tables S2 and S3 in the Supplementary Appendix).²²

STATISTICAL ANALYSIS

We checked the validity of the data and plotted the time series to confirm assumptions of linearity before and after the removal of financial incentives. For the primary analysis, we used segmented regression of interrupted time-series data to examine changes in the quality of care that were associated with the removal of incentives at the end of fiscal year 2013–2014. Depending on when the indicators were introduced, there were either three or four time points before 2013–2014. We used the method recommended by the Cochrane Effective Practice and Organisation of Care group for short time series.²¹ INLIQ data submission was strongly encouraged, but there were no specific penalties for not submitting data.

Practices were included in the primary time-series analysis of indicators with incentives removed if the practices reported data in each of the years examined. The main analysis of incentive removal included data from 2723 to 2819 practices (totaling >20 million registered patients and constituting approximately one third of all English practices), depending on the indicator examined (Table S4 in the Supplementary Appendix). We also carried out a sensitivity analysis of incentive removal that included data from all practices submitting data at any time point. The sensitivity analysis included a maximum of 8245 practices in 2010–2011 and a mini-

mum of 4223 practices in 2014–2015. The analysis of indicators for which incentives were maintained included data from all practices submitting QOF data, which fell from 8245 practices in 2010–2011 to 7291 in 2013–2014, mainly because of practice mergers. Full model results are presented in Tables S5 and S6 in the Supplementary Appendix. The focus of this report is on the estimated change in quality-of-care performance 1 year and 3 years after 2013–2014, as compared with the performance expected on the basis of trends before incentive removal.

To examine whether variation between practices changed after the removal of incentives, we constructed box plots of practice-level quality and evaluated differences in variance between 2013–2014 (the year before removal) and 2016–2017 (3 years after removal), using the Brown–Forsythe test.²³ To examine whether incentive removal differentially affected the performance of practices serving the most affluent and least affluent populations, we used segmented regression analysis to examine changes associated with incentive removal according to the level of affluence in the area served, comparing the mean quality in the quintile of practices serving the most affluent areas with the mean quality in the quintile serving the least affluent areas. All analyses were carried out with SPSS software, version 22.

RESULTS

CHANGES IN QUALITY-OF-CARE INDICATORS AFTER INCENTIVE REMOVAL

Before the removal of financial incentives, the documented quality of care was significantly increasing for 4 of the 12 indicators (glycated hemoglobin testing in patients with serious mental illness, cholesterol control in patients with stroke or TIA, lifestyle counseling for patients with hypertension, and preconception advice for patients with epilepsy) and was significantly decreasing for 2 indicators (thyroid-function testing in patients with hypothyroidism and cholesterol testing in patients with stroke or TIA), although absolute changes from year to year were small. As compared with expected values on the basis of previous trends, there were significant, moderate-to-large reductions in documented quality for all 12 indicators in the year after incentives were

removed, with small, mainly negative changes in trend in subsequent years (Fig. 1, and Table S5 and Fig. S2 in the Supplementary Appendix).

Reductions in the documented quality of care the year after removal of financial incentives, as compared with the quality that would be expected on the basis of previous trends, ranged from 5.8 percentage points (95% confidence interval [CI], -8.9 to -2.7) for documentation of smoking status to 62.3 percentage points (95% CI, -65.6 to -59.0) for documentation of lifestyle counseling in patients with hypertension (Table 2). Three years after removal of the incentives, there were still significant reductions in documented quality for all 12 indicators, with the largest reductions in 2 of the 3 health-advice indicators (-71.6 percentage points for lifestyle counseling for patients with hypertension and -65.9 percentage points for preconception advice for patients with epilepsy) and in 1 of the intermediate-outcome measures (-53.6 percentage points for documentation of seizure-free status in patients with epilepsy). Reductions in clinical-process measures were generally smaller, ranging from a reduction of 9.2 percentage points for thyroid-function testing in patients with hypothyroidism to a reduction of 37.5 percentage points for glycosylated hemoglobin testing in patients with serious mental illness, with similar reductions for intermediate clinical outcomes (apart from documentation of freedom from seizures in patients with epilepsy), ranging from a reduction of 10.1 percentage points for cholesterol control in patients with CHD to a reduction of 16.8 percentage points for cholesterol control in patients with stroke or TIA.

In contrast, analysis of the six indicators for which incentives were maintained showed no changes in documented quality 3 years after 2013–2014 for four indicators (blood-pressure measurement in patients with serious mental illness, documentation of alcohol consumption in patients with serious mental illness, cholesterol control in patients with diabetes, and documentation of advice regarding smoking cessation in patients with chronic illness), with significant but small absolute decreases of 1.4 percentage points (95% CI, -2.5 to -0.3) and 2.4 percentage points (95% CI, -3.5 to -1.3) for blood-pressure control in patients with CHD and in patients

Figure 1. Documented Quality of Care for Indicators with Financial Incentives That Were Either Removed or Maintained at the End of 2013–2014.

Shown are the percentages of patients in whom quality-of-care indicators were documented during the period from 2010–2011 to 2016–2017. Indicators for which incentives were removed are shown in Panels A, C, and E; the vertical red lines indicate the year of removal. Indicators for which incentives were maintained are shown in Panels B, D, and F. Panels A and B show clinical-process indicators, Panels C and D show intermediate-outcome indicators, and Panels E and F show health-advice indicators. CHD denotes coronary heart disease, and TIA transient ischemic attack.

with stroke or TIA, respectively (Table 2 and Fig. 1). The results of a sensitivity analysis that included data from all practices at every time point were consistent with the results of the primary analysis of practices with complete data (Fig. S3 and Tables S6 and S7 in the Supplementary Appendix).

DIFFERENCES IN EFFECTS OF INCENTIVE REMOVAL AMONG PRACTICES

Variation among practices increased for 7 of the indicators for which incentives were removed (Fig. S4 and Table S8 in the Supplementary Appendix), was unchanged for cholesterol control in patients with stroke or TIA and thyroid-function testing in patients with hypothyroidism, and decreased for documentation of preconception advice given to patients with epilepsy and documentation of seizure-free status in patients with epilepsy (reflecting that the majority of practices ceased all documentation of these indicators). At baseline, performance on 6 of the 12 indicators was significantly better in the most affluent areas and performance on 2 of the indicators was significantly better in the least affluent areas. Absolute differences were generally small, although the three largest baseline differences favored the most affluent areas (4.9 percentage points for diabetic retinopathy screening, 7.4 percentage points for glycosylated hemoglobin testing in patients with serious mental illness, and 16.6 percentage points for documentation of seizure-free status in patients with epilepsy). After removal of the financial incentives, there were significant changes in differences between practices serving the most affluent areas and those

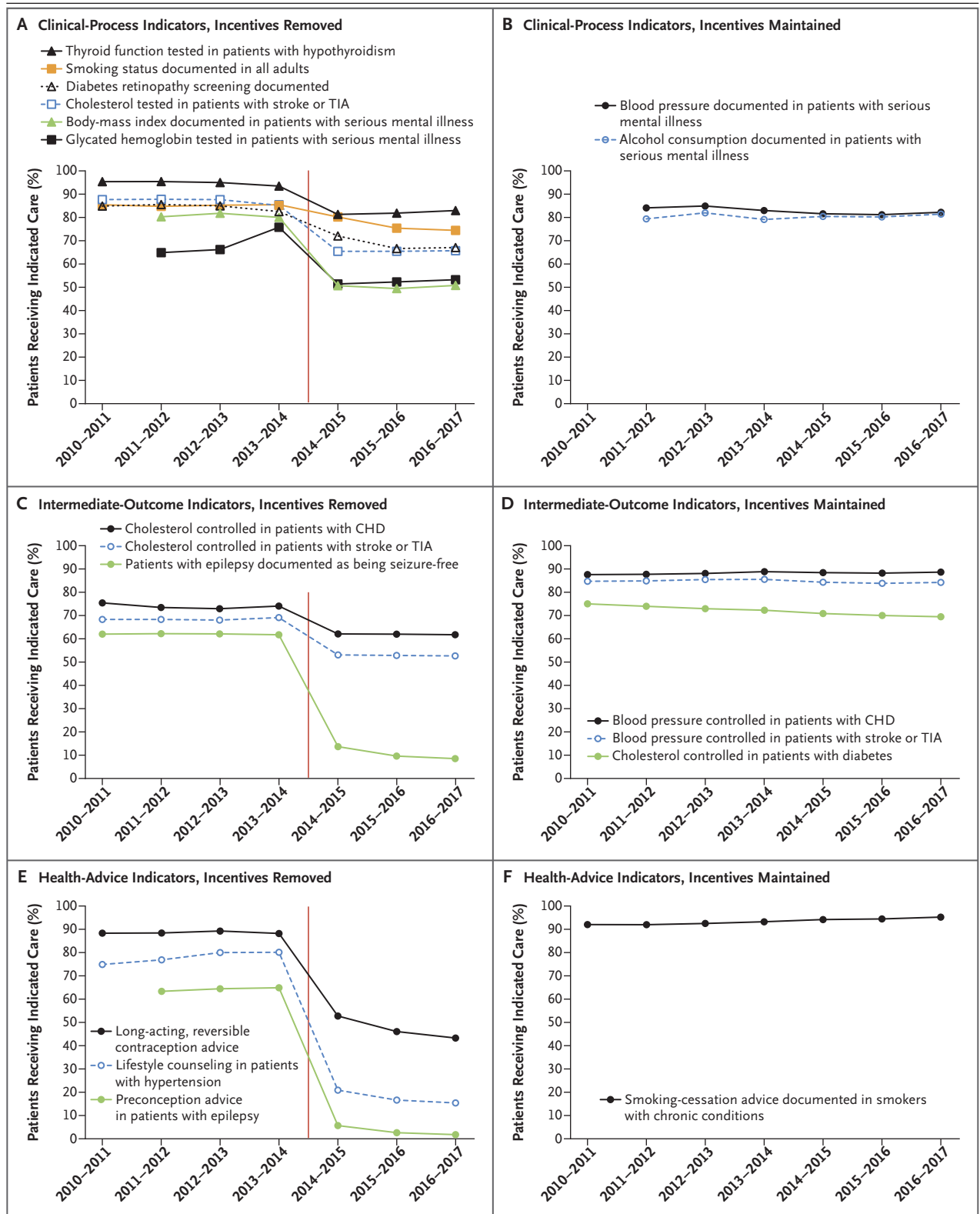


Table 2. Changes in Documented Quality of Care According to Whether Incentives Were Removed or Maintained.

Indicator	Patients Receiving Indicated Care before Incentive Removal*	Absolute Change in Expected Trend after Incentive Removal	
		1 Yr after Removal	3 Yr after Removal
		<i>percentage points (95% CI)</i>	
Incentives removed			
Smoking status documented in all adults	85.5 (85.5 to 85.5)	-5.8 (-8.9 to -2.7)	-11.7 (-16.3 to -7.1)
Diabetes retinopathy screening documented	82.6 (82.5 to 82.6)	-11.5 (-16.7 to -6.3)	-14.9 (-22.7 to -7.1)
Glycated hemoglobin tested in patients with serious mental illness	75.8 (75.5 to 76.1)	-28.5 (-39.0 to -18.0)	-37.5 (-55.4 to -19.6)
Body-mass index documented in patients with serious mental illness	80.1 (79.9 to 80.3)	-30.2 (-35.5 to -24.9)	-29.9 (-38.9 to -20.9)
Thyroid function tested in patients with hypothyroidism	93.5 (93.5 to 93.6)	-12.1 (-13.6 to -10.6)	-9.2 (-11.4 to -7.0)
Cholesterol tested in patients with stroke or TIA	85.2 (85.1 to 85.3)	-19.8 (-22.4 to -17.2)	-18.0 (-21.8 to -14.2)
Cholesterol controlled in patients with stroke or TIA	69.1 (68.9 to 69.3)	-15.9 (-17.1 to -14.7)	-16.8 (-18.0 to -15.0)
Cholesterol controlled in patients with CHD	74.1 (74.0 to 74.2)	-10.7 (-13.6 to -7.8)	-10.1 (-14.3 to -5.9)
Patients with epilepsy documented as being seizure-free	61.7 (61.4 to 62.0)	-48.7 (-51.8 to -45.6)	-53.6 (-58.2 to -49.0)
Lifestyle counseling documented in patients with hypertension	80.1 (80.0 to 80.2)	-62.3 (-65.6 to -59.0)	-71.6 (-76.5 to -66.7)
Preconception advice documented in patients with epilepsy	64.9 (64.3 to 65.4)	-60.5 (-63.5 to -57.5)	-65.9 (-71.0 to -60.8)†
Advice about using long-acting, reversible contraceptives documented	88.2 (88.1 to 88.3)	-36.6 (-39.9 to -33.3)	-46.1 (-51.8 to -45.6)
Incentives maintained‡			
Blood pressure documented in patients with serious mental illness	83.0 (82.9 to 83.1)	-1.5 (-5.4 to 2.4)	0.2 (-6.5 to 6.9)
Alcohol consumption documented in patients with serious mental illness	79.1 (79.0 to 79.2)	0.3 (-6.9 to 7.5)	1.5 (-10.7 to 13.7)
Cholesterol controlled in patients with diabetes	72.3 (72.3 to 72.4)	-0.4 (-0.8 to 0.05)	0.0 (-0.7 to 0.7)
Blood pressure controlled in patients with CHD	88.9 (88.8 to 88.9)	-0.8 (-1.6 to -0.04)	-1.4 (-2.5 to -0.3)
Blood pressure controlled in patients with stroke or TIA	85.6 (85.6 to 85.6)	-1.7 (-2.5 to -0.9)	-2.4 (-3.5 to -1.3)
Smoking-cessation advice documented in smokers with chronic conditions	93.2 (93.2 to 93.2)	0.6 (-0.2 to 1.4)	0.9 (-0.2 to 2.0)

* Values are population means for the year before incentive removal.

† The estimated reduction is greater than the value in the year before incentive removal because there was a small upward trend in this indicator before incentive removal, so the expected value 3 years after removal was slightly larger.

‡ In the analysis of indicators for which incentives were maintained, we examined whether there was a change in performance after an assumed interruption on the date of incentive removal for other indicators (at the end of 2013–2014).

serving the least affluent for 7 indicators: 2 favoring the most affluent areas and 5 favoring the least affluent areas (Fig. S5, S6, and S7 and Table S9 in the Supplementary Appendix).

DISCUSSION

Our analysis of information documented in EMRs showed that there were immediate reductions in

quality-of-care measures for all 12 indicators in the first year after the removal of financial incentives, with only small additional changes in the following 2 years. Reductions were generally largest for indicators related to documented provision of health advice (Table 1), with absolute reductions ranging from 46.1 to 71.6 percentage points 3 years after removal of financial incentives, and for the indicator related to docu-

mentation of seizure-free status in patients with epilepsy, with an absolute reduction of 53.6 percentage points (Table 2). These were indicators for which the physician was required to check boxes in the EMR to indicate that care had been delivered, and the large reductions observed could indicate either that care was no longer given or that it was no longer documented. Changes were smaller, although still substantial, for performance on clinical-process and intermediate-outcome indicators, for which data such as blood pressure and smoking status are routinely recorded in coded form and laboratory test values are automatically entered into the EMR. The smallest change in performance at 3 years was a reduction of 9.2 percentage points in thyroid-function testing in patients with hypothyroidism. There were no large changes in documented quality on indicators for which incentives were not removed. Removal of incentives was usually associated with increased variation in documented quality among practices, but socioeconomic disparities narrowed rather than widened after incentive removal.

Five previous studies of financial-incentive removal in health care had conflicting results.^{8-11,18} One study showed that incentive removal had little effect on documented quality for eight QOF indicators, but seven of them were process indicators (e.g., measurement of blood pressure) for which linked outcome indicators continued to incentivize measurement.⁹ Similarly, performance was sustained after incentive removal for six of seven measures of quality of care in 128 Veterans Health Administration (VA) hospitals in the United States¹⁰ and for nine primary care prescribing-safety indicators in the United Kingdom.¹⁸ However, in both cases, financial incentives were part of more comprehensive improvement interventions, including blends of goal setting, comparative feedback, facilitation, and education. In contrast, there were reductions in diabetic retinopathy screening (by approximately 8%) and cervical cancer screening (by approximately 4%) after incentive removal in 35 Kaiser Permanente facilities, although the reductions occurred gradually over a period of several years rather than abruptly.⁸ Similarly, quality declined after incentive removal in VA facilities participating in a randomized, controlled trial of incentives to improve hypertension control.¹¹ In both

cases, reductions after removal of the incentives were similar to gains associated with the introduction of the incentives.

Our study also showed moderate reductions in documented clinical-process and outcome quality, which were of a scale similar to that of observed increases in documented quality when the QOF was introduced.^{12,24} However, we observed much larger reductions in clinical-process documentation in patients with serious mental illness, in documentation of whether patients with epilepsy were seizure-free, and in documentation of health advice. Further research is needed to examine the reasons for these larger reductions. One possible explanation is that most practices do not offer nurse-led clinics for the management of serious mental illness or epilepsy, whereas such clinics are offered for more common chronic conditions such as diabetes, so that care processes for the less common chronic diseases may not be routinely embedded in clinical practice. The simultaneous removal from EMRs of pop-up reminders to opportunistically deliver care or document activity may also have contributed, particularly for aspects of incentivized care, such as giving health advice,²⁵ which clinicians may value less than care for more established chronic diseases. The introduction of the QOF was associated with the narrowing of quality differences between practices serving more affluent populations and those serving less affluent populations,²⁶ but previous studies of incentive removal did not examine such disparities. Despite reductions in documented quality for all indicators, it is reassuring that socioeconomic disparities more often narrowed than widened after incentive removal.

The strengths of this study include the use of interrupted time-series analysis to examine incentive removal in routine care in approximately one third of all English practices, with a sensitivity analysis that included all available data showing findings that were consistent with the primary analysis. A weakness is that the time series has relatively few data points, but this is inevitable with annual reporting. As with all observational studies, residual confounding cannot be excluded as an explanation for the observed changes in documented quality, although the stability of indicators for which incentives were maintained reduces the likelihood of residual confounding.

The key limitation of the study is that on the basis of the available data, we cannot distinguish between changes in clinical activity and changes in documentation of clinical activity in the EMR. The four serious mental illness indicators provide the most direct test of this. The two indicators for which incentives were maintained did not change (which is consistent with patients still being regularly reviewed by practices), but there were marked reductions in documentation of glycated hemoglobin measurement (automatically imported to the EMR from the laboratory) and body-mass index (automatically calculated if weight is entered, although measurement of weight and other vital signs is not routine when patients visit primary care practices in the United Kingdom). These findings are consistent with a true change in clinical practice, although proof of a true change would require a detailed record review to examine whether care was still being delivered but was documented in free text rather than in the coded fields used for measurement. Our overall interpretation is that observed reductions in quality for core clinical care (e.g., blood-pressure management, retinopathy screening, and laboratory measurements for processes and outcomes) do reflect changes in clinical practice but that the much larger reductions in documentation of clinical advice and seizure-free status in patients with epilepsy should be more cautiously interpreted. For example, family doctors in the United Kingdom may now be giving less advice about long-acting, reversible contraception, may simply not be documenting that advice, or both. We know that the introduction of incentives to give advice was associated with increases in the use of long-acting contraceptives.²⁷ What we need to know now is whether that increased use persists once incentives to give advice are removed and, more broadly, whether any gains with the

introduction of incentives are maintained or lost when those incentives are removed.

If pay for performance is to contribute widely to quality improvement, then it is inevitable that incentives will be removed from some indicators to allow resources for quality improvement to be redeployed.²⁸ A key implication of this study is that although the effect of incentive removal probably depends on the context, reductions in quality are likely, and several studies show that what is gained on incentive introduction is essentially lost on incentive withdrawal. Therefore, at a minimum, payers planning to remove incentives should monitor the quality of care after removal. In doing so, they face the same conundrum involved in introducing incentives: the uncertainty about whether changes in documented quality represent true changes in patient care. Options include using chart review or examining data that directly measure clinical practice, such as laboratory claims or prescribing data. However, collecting such data will add to the cost and potentially the viability of pay-for-performance programs. Better still, randomized removal of incentives or removal from some practices or individual physicians would provide a more robust means of evaluating the effects of incentive removal.²⁹ More generally, such effects will be mediated by the wider context of quality improvement, including public reporting, the underlying quality-of-care infrastructure, and other interventions used alongside incentives to improve quality.^{8-11,18} Financial incentives that simply pay providers to deliver specified levels of quality therefore seem unlikely to deliver sustainable improvement unless they are aligned with more comprehensive interventions that change the organization of care and future clinical practice.^{8,10,11}

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

REFERENCES

1. Eijkenaar F. Pay for performance in health care: an international overview of initiatives. *Med Care Res Rev* 2012;69:251-76.
2. Developing health care quality indicators and improving the QOF program for the Thai Universal Health Coverage – Part 1. Nonthaburi, Thailand: Health Intervention and Technology Assessment Program (<http://www.hitap.net/en/research/163364>).
3. Cashin C, Chi Y-L, Smith P, Borowitz M, Thomson S. Paying for performance in health care: implications for health system performance and accountability. Maidenhead, United Kingdom: Open University Press, 2014.
4. Scott A, Sivey P, Ait Ouakrim D, et al. The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database Syst Rev* 2011;9:CD008451.
5. Eijkenaar F, Emmert M, Scheppach M, Schöffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy* 2013;110:115-30.
6. Mendelson A, Kondo K, Damberg C, et al. The effects of pay-for-performance programs on health, health care use, and processes of care: a systematic review. *Ann Intern Med* 2017;166:341-53.
7. Guthrie B, Huby G, Davies HTO. What changes when incentives change in primary medical care? Swindon, United Kingdom: Economics and Social Research Council, 2005.

8. Lester H, Schmittiel J, Selby J, et al. The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators. *BMJ* 2010;340:c1898.
9. Kontopantelis E, Springate D, Reeves D, Ashcroft DM, Valderas JM, Doran T. Withdrawing performance indicators: retrospective analysis of general practice performance under UK Quality and Outcomes Framework. *BMJ* 2014;348:g330.
10. Benzer JK, Young GJ, Burgess JF Jr, et al. Sustainability of quality improvement following removal of pay-for-performance incentives. *J Gen Intern Med* 2014;29:127-32.
11. Petersen LA, Simpson K, Pietz K, et al. Effects of individual physician-level and practice-level financial incentives on hypertension care: a randomized trial. *JAMA* 2013;310:1042-50.
12. Doran T, Kontopantelis E, Valderas JM, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ* 2011;342:d3590.
13. Roland M. Linking physicians' pay to the quality of care — a major experiment in the United Kingdom. *N Engl J Med* 2004;351:1448-54.
14. Gillam SJ, Siriwardena AN, Steel N. Pay-for-performance in the United Kingdom: impact of the Quality and Outcomes Framework: a systematic review. *Ann Fam Med* 2012;10:461-8.
15. Guthrie B, Tang J. What did we learn from 12 years of QOF? Glasgow, United Kingdom: Scottish School of Primary Care, 2016 (http://www.sspc.ac.uk/media/media_486342_en.pdf).
16. Rud Kristensen S, Meacock R, Turner AJ, et al. Long-term effect of hospital pay for performance on mortality in England. *N Engl J Med* 2014;371:540-8.
17. Roland M, Dudley RA. How financial and reputational incentives can be used to improve medical care. *Health Serv Res* 2015;50:Suppl 2:2090-115.
18. Dreischulte T, Donnan P, Grant A, Hapca A, McCowan C, Guthrie B. Safer prescribing — a trial of education, informatics, and financial incentives. *N Engl J Med* 2016;374:1053-64.
19. The Quality and Outcomes Framework — QOF business rules and INLIQ business rules. NHS Digital, 2017 (<http://content.digital.nhs.uk/qof>).
20. Department for Communities and Local Government. The English Indices of Deprivation 2015. London: Department for Communities and Local Government, 2015 (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/465791/English_Indices_of_Deprivation_2015_-_Statistical_Release.pdf).
21. Interrupted time series (ITS) analyses. Oxford, England: Cochrane Collaboration EPOC Group, 2017 (https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/what_study_designs_should_be_included_in_an_epoc_review.pdf).
22. Roland M, Everington S. Tackling the crisis in general practice. *BMJ* 2016;352:i942.
23. Brown M, Forsythe A. Robust test for the equality of variances. *J Am Stat Assoc* 1974;69:364-7.
24. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *N Engl J Med* 2009;361:368-78.
25. Dewhurst A, Peters S, Devereux-Fitzgerald A, Hart J. Physicians' views and experiences of discussing weight management within routine clinical consultations: a thematic synthesis. *Patient Educ Couns* 2017;100:897-908.
26. 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.
27. Arrowsmith ME, Majeed A, Lee JT, Saxena S. Impact of pay for performance on prescribing of long-acting reversible contraception in primary care: an interrupted time series study. *PLoS One* 2014;9(4):e92205.
28. Reeves D, Doran T, Valderas JM, et al. How to identify when a performance indicator has run its course. *BMJ* 2010;340:c1717.
29. Hocking JS, Temple-Smith M, van Driel M, et al. Can preventive care activities in general practice be sustained when financial incentives and external audit plus feedback are removed? ACCEPT-able: a cluster randomised controlled trial protocol. *Implement Sci* 2016;11:122.

Copyright © 2018 Massachusetts Medical Society.

SPECIALTIES AND TOPICS AT NEJM.ORG

Specialty pages at the *Journal's* website (NEJM.org) feature articles in cardiology, endocrinology, genetics, infectious disease, nephrology, pediatrics, and many other medical specialties.