SPECIAL ARTICLE

Exclusion of Patients from Pay-for-Performance Targets by English Physicians

Tim Doran, M.D., Catherine Fullwood, Ph.D., David Reeves, Ph.D., Hugh Gravelle, Ph.D., and Martin Roland, D.M.

ABSTRACT

BACKGROUND

In the English pay-for-performance program, physicians use a range of criteria to exclude individual patients from the quality calculations that determine their pay. This process, which is called exception reporting, is intended to safeguard patients against inappropriate treatment by physicians seeking to maximize their income. However, exception reporting may allow physicians to inappropriately exclude patients for whom targets have been missed (a practice known as gaming).

METHODS

We analyzed data extracted automatically from clinical computing systems for 8105 family practices in England (96% of all practices), data from the U.K. Census, and data on practice characteristics from the U.K. Department of Health. We determined the rate of exception reporting for 65 clinical activities and the association between this rate and the characteristics of patients and medical practices.

RESULTS

From April 2005 through March 2006, physicians excluded a median of 5.3% of patients (interquartile range, 4.0 to 6.9) from the quality calculations. Physicians were most likely to exclude patients from indicators that were related to providing treatments and achieving target levels of intermediate outcomes; they were least likely to exclude patients from indicators that were related to routine checks and measurements and to offers of treatment. The characteristics of patients and practices explained only 2.7% of the variance in exception reporting. We estimate that exception reporting accounted for approximately 1.5% of the cost of the pay-for-performance program.

CONCLUSIONS

Exception reporting brings substantial benefits to pay-for-performance programs, providing that the process is used appropriately. In England, rates of exception reporting have generally been low, with little evidence of widespread gaming.

From the National Primary Care Research and Development Centre, University of Manchester, Manchester (T.D., C.F., D.R., M.R.); and the National Primary Care Research and Development Centre, Centre for Health Economics, University of York, York (H.G.) — both in the United Kingdom. Address reprint requests to Dr. Doran at the National Primary Care Research and Development Centre, Williamson Bldg., University of Manchester, Manchester M13 9PL, United Kingdom, or at tim.doran@manchester.ac.uk.

N Engl J Med 2008;359:274-84.

There is widespread international interest in pay-for-performance programs for health professionals. New programs are being developed in the United States,[1-3] Australia,[4] Canada,[5] Germany,[6] the Netherlands,[7] and New Zealand.[8] An ambitious program that was initiated in April 2004 in the United Kingdom, called the Quality and Outcomes Framework, ties 25% of the income of primary care practitioners to the quality of delivered care.[9,10]

A key consideration for pay-for-performance programs is avoiding inappropriate treatment of patients to whom a quality indicator does not apply or for whom other considerations take precedence.[11-14] There are three approaches to this problem. The first is to risk-adjust the quality indicators[15] — for example, by constructing indicators that allow for as many combinations of coexisting illnesses as possible.[16] This approach makes indicators very complex, and it is never possible to allow for all eventualities. The second approach sets maximum-achievement thresholds below 100% — in other words, physicians can earn the maximum financial rewards without achieving the targets for all patients. In the United States, this approach has been used for the Bridges to Excellence program, which sets thresholds of 25 to 80% for aspects of cardiac and diabetic care,[17] and the Medicare Physician Quality Reporting Initiative.[18] However, determining which proportion of patients would not meet the criteria for each quality indicator is extremely difficult. The third approach, adopted in some U.S. programs,[19,20] allows physicians to use their clinical judgment to remove inappropriate patients from calculations of quality achievement. In the United Kingdom, the pay-for-performance program adopts the second and third approaches. The removal of inappropriate patients in the United Kingdom is called exception reporting (Table 1).

Exception reporting has attracted controversy, including calls for its abolition. Although many physicians regard it as an essential safeguard against inappropriate treatment, others are concerned that it might be used as an excuse for substandard care of patients. Physicians could also exploit exception reporting for financial gain by excluding patients for whom the targets had been missed rather than because of a genuine clinical reason, a practice known as gaming. An apparent association between high rates of exception reporting and high levels of performance in the first year of the U.K. pay-for-performance program heightened fears of gaming.[10] Our study is based on the first full set of data with respect to exception reporting for all practices in England, collected during the second year of the pay-for-performance program, from April 2005 through March 2006. The data contribute to the debate regarding whether physicians should have the option of excluding individual patients from pay-for-performance programs and regarding what the threshold for maximum achievement should be in the absence of exception reporting.

METHODS

STUDY DESIGN

In the United Kingdom, primary care is provided by primary care practices, comprised of up to 10 primary care physicians (mean, 3.5) who work from common premises. The contract for providing medical care and all quality payments relate to the practice rather than to the individual physician. There is a single common electronic medical record for all practices, and data on quality of care are extracted automatically from clinical computing systems of practices and are collated in the central National Health Service Quality Management and Analysis System (QMAS) database.

The analyses that we describe here relate to all 65 clinical indicators that were used in the second year of the U.K.'s pay-for-performance program. Data were available for the prevalence of each of the 10 diseases listed in the program, since practices are required to maintain registers of patients with relevant diagnoses and the quality points earned for each indicator. Quality points translate into physicians' income through a formula that takes into account the prevalence of each disease.[21] These data are published nationally by the National Health Service Information Centre,[22] which also provided data on the number of patients who were excluded for each indicator according to medical practice. Although the reasons for exception reporting are recorded on the computing systems of practices, they are not collated in the QMAS database and hence were not available.

For each practice, we calculated the rate of exception reporting, which was defined as the number of patients who were excluded for each indicator as a proportion of the number of pa-

Table 1. Reasons for the Exclusion of Patients from Quality Assessment.

The patient has received at least three invitations for a review during the preceding 12 mo but has not attended.
The indicator is judged to be inappropriate for the patient because of particular circumstances, such as terminal illness, extreme frailty, or the presence of a supervening condition that makes the specified treatment clinically inappropriate.
The patient has recently received a diagnosis or has recently registered with the practice.
The patient is taking the maximum tolerated dose of a medication, but the levels remain suboptimal.
The patient has had an allergic or other adverse reaction to a specified medication or has another contraindication to the medication.
The patient does not agree to the investigation or treatment.
A specified investigative service is unavailable to the family practitioner.

tients who were eligible for the target. We constructed summary scores for each disease and each type of clinical activity using data for the relevant indicators (Table 2). An overall score was constructed as an unweighted mean across all indicators.

Information on characteristics of medical practices was taken from the 2004 General Medical Statistics database, which is maintained by the Department of Health. These characteristics included the number of patients, the number of physicians per 10,000 patients, and the age, sex, and place of qualification (medical school within or outside the United Kingdom) of the physicians. We included an indicator for the 34% of practices that operate alternative Personal Medical Services (PMS) contracts. These contracts are minor variants on the standard national contract and are negotiated at the local level to address local priorities. Virtually all PMS practices participated in the national pay-for-performance program. We attributed socioeconomic characteristics to each practice on the basis of the electoral district in which the practice was located (average population, 5500), using data from the 2001 U.K. Census and the Index of Multiple Deprivation, a nationally recognized measure of socioeconomic deprivation.[23]

The QMAS database contains data from 8409 family practices in England. Practices were excluded from the study if they had fewer than 1000 patients (44 practices), if one or more disease registers were missing (107 practices), or if data regarding exception reporting were missing or incomplete (175 practices). Our main results are drawn from 8105 practices, which provide care for more than 49 million patients. Complete socioeconomic data were not available for 476 practices (5.9%), so these practices were excluded from the regression analyses. As a group, these practices did not differ significantly in terms of the rates of exception reporting. Since this study was based on publicly available data, it did not require approval from institutional ethics committees.

STATISTICAL ANALYSIS

We analyzed the effect of the characteristics of patients and medical practices on rates of exception reporting, using a multiple linear regression model reporting least-square means with robust estimates of error variance. Area indicator variables were included to allow for unobserved effects of local National Health Service policies. All explanatory variables were divided by their standard deviation, so reported coefficients show the increase in the standard deviation of rates of exception reporting for each 1-SD increase in predictor variables.

We also analyzed the financial gain from exception reporting (i.e., the number of quality points earned by practices that could be attributed to exception reporting). Analyses were performed with the use of Stata software, version 9. Further details regarding the methods used for calculating rates of exception reporting and financial gain are reported in the Supplementary Appendix, available with the full text of this article at www.nejm.org.

RESULTS

CLINICAL TARGETS

Overall, medical practices achieved clinical targets for a median of 84.6% of eligible patients (interquartile range, 81.5 to 87.1) on their chronic-disease registers. Of the patients who were not excluded, practices achieved clinical targets for a median of 90.4% of patients (interquartile range, 87.4 to 92.5).

RATE OF EXCEPTION REPORTING

The median rate of exception reporting for all practices in 2005–2006 was 5.3% (interquartile range, 4.0 to 6.9) (Table 3). This rate ranged from 0 to 28.3% according to practice (Fig. 1). The rates of exception reporting for all 65 individual indicators are listed in the Supplementary Appendix.

Table 2. Quality and Outcome Indicators for Clinical Activities, According to Disease and Type of Activity.*

Disease	Type of Activity				
	Diagnosis and Referral (40 points)	Measurement and Review (183 points)	Offer of Treatment (33 points)	Provision of Treatment (79 points)	Intermediate Outcomes (157 points)
Asthma (65 points)	Diagnosis confirmed by spirometry or peak flow measurement (age, >8 yr) (15 points)	Smoking status (age, 14–19 yr or >20 yr); annual review (32 points)	Smoking-cessation advice (6 points)	Influenza vaccine (age, >16 yr) (12 points)	NA
Cancer (6 points)	NA	Review of care coordination and support needs (within 6 mo after diagnosis) (6 points)	NA	NA	NA
Coronary heart disease (111 points)	Referral for exercise testing or specialist assessment; confirmation of diagnosis by echocardiography (13 points)	Smoking status; blood pressure; total cholesterol (21 points)	Smoking-cessation advice (4 points)	Influenza vaccine; aspirin or alternative antiplatelet or anticoagulant agent; beta-blocker; ACE inhibitor (for myocardial infarction); ACE inhibitor or A2 antagonist (for left ventricular dysfunction) (38 points)	Blood pressure ≤150/90 mm Hg; total cholesterol ≤193 mg/dl (5 mmol/liter) (35 points)
Chronic obstructive pulmonary disease (40 points)	Performance of spirometry and testing of reversibility (both for newly diagnosed cases and in all patients) (10 points)	Smoking status; forced expiratory volume in 1 sec (12 points)	Smoking-cessation advice (6 points)	Monitoring of use of inhaler; influenza vaccine (12 points)	NA
Diabetes (93 points)	NA	Body-mass index; smoking status; glycated hemoglobin level; retinal function; peripheral pulses; presence of neuropathy; blood pressure; microalbuminuria screening; serum creatinine level; total cholesterol level (32 points)	Smoking-cessation advice (5 points)	Influenza vaccine; ACE inhibitor or A2 antagonist (for proteinuria or microalbuminuria) (6 points)	Glycated hemoglobin, ≤7.4% or ≤10%; blood pressure, ≤145/85 mm Hg; total cholesterol, ≤193 mg/dl (5 mmol/liter) (50 points)
Epilepsy (14 points)	NA	Seizure frequency; medication review (age, >16 yr) (8 points)	NA	NA	No convulsions for 12 mo (age, >16 yr) (6 points)
Hypertension (96 points)	NA	Smoking status; blood pressure (30 points)	Smoking-cessation advice (10 points)	NA	Blood pressure, ≤150/90 mm Hg (56 points)
Hypothyroidism (6 points)	NA	Thyroid function (6 points)	NA	NA	NA
Mental illness (34 points)	NA	Lithium therapy (levels of lithium, serum creatinine, and thyrotropin); review of clinical status, medication, and support needs (29 points)	NA	For lithium therapy, lithium levels in the therapeutic range (5 points)	NA
Stroke (27 points)	Referral for CT or MRI scan (2 points)	Smoking status; blood pressure; total cholesterol level (7 points)	Smoking-cessation advice (2 points)	Influenza vaccine; aspirin or alternative antiplatelet or anticoagulant drug (for nonhemorrhagic stroke) (6 points)	Blood pressure, ≤150/90 mm Hg; total cholesterol, ≤193 mg/dl (5 mmol/liter) (10 points)

* The maximum number of points available varies according to the indicator (see the Supplementary Appendix). The number of points available for each set of indicators is provided. The total number of points for all indicators is 492. ACE denotes angiotensin-converting enzyme, A2 angiotensin II receptor, CT computed tomography, and NA not applicable.

Table 3. Rates of Exception Reporting in Family Practices in England, 2005–2006, According to Disease and Type of Activity.*

Disease	Type of Activity					Total
	Diagnosis and Referral	Measurement and Review	Offer of Treatment	Provision of Treatment	Achievement of Intermediate Outcome	
	percent exception reporting					
Asthma						
Mean	3.6±4.6	3.0±4.0	2.4±4.4	18.0±10.7	NA	7.0±4.7
Median	2.3	1.5	0.0	16.0	NA	5.9
Interquartile range	0.0–5.1	0.7–3.6	0.0–3.0	10.0–24.0	NA	3.7–9.1
Range	0.0–60.0	0.0–40.5	0.0–50.0	0.0–68.8	NA	0.0–36.5
Cancer						
Mean	NA	9.1±8.9	NA	NA	NA	9.1±8.9
Median	NA	7.6	NA	NA	NA	7.6
Interquartile range	NA	2.9–12.7	NA	NA	NA	2.9–12.7
Range	NA	0.0–83.3	NA	NA	NA	0.0– 83.3
Coronary heart disease						
Mean	11.1±10.5	2.1±2.0	2.7±4.6	12.8±5.5	7.1±4.2	7.5±3.2
Median	9.1	1.5	0.0	12.4	6.3	7.1
Interquartile range	3.0–16.3	0.7–2.9	0.0–4.2	8.8–16.2	4.1–9.1	5.2–9.3
Range	0.0–80.0	0.0–17.6	0.0–50.0	0.0–37.9	0.0–47.8	0.0–31.1
Chronic obstructive pulmonary disease						
Mean	8.6±8.4	5.7±5.4	2.5±5.0	8.5±5.2	NA	7.2±5.3
Median	6.6	4.3	0.0	7.8	NA	6.0
Interquartile range	3.0–11.8	1.9–7.9	0.0–3.4	5.1–11.1	NA	3.6–9.5
Range	0.0–89.4	0.0–50.0	0.0–66.7	0.0–61.1	NA	0.0–57.7
Diabetes						
Mean	NA	4.1±3.0	3.2±4.6	12.6±6.1	9.4±5.0	6.1±3.4
Median	NA	3.4	1.4	11.9	8.4	5.4
Interquartile range	NA	2.0–5.3	0.0–5.0	8.5–15.7	5.9–11.5	3.8–7.5
Range	NA	0.0–34.5	0.0–66.7	0.0–50.0	0.0–51.2	0.0–40.0
Epilepsy						
Mean	NA	3.5±5.3	NA	NA	17.4±15.3	8.1±7.2
Median	NA	1.7	NA	NA	13.8	6.7
Interquartile range	NA	0.0–5.0	NA	NA	5.9–25.0	2.8–11.6
Range	NA	0.0–75.8	NA	NA	0.0–95.7	0.0–82.4
Hypertension						
Mean	NA	1.3±1.2	0.7±1.4	NA	5.4±3.9	2.6±1.8
Median	NA	1.0	0.0	NA	4.4	2.1
Interquartile range	NA	0.6–1.5	0.0–1.0	NA	3.0–6.6	1.4–3.1
Range	NA	0.0–16.9	0.0–28.6	NA	0.0–54.4	0.0–21.1
Hypothyroidism						
Mean	NA	0.7±1.4	NA	NA	NA	0.7±1.4
Median	NA	0.0	NA	NA	NA	0.0
Interquartile range	NA	0.0–0.9	NA	NA	NA	0.0–0.9
Range	NA	0.0–31.6	NA	NA	NA	0.0–31.6

Table 3. (Continued.)

Disease	Type of Activity					Total
	Diagnosis and Referral	Measurement and Review	Offer of Treatment	Provision of Treatment	Achievement of Intermediate Outcome	
	percent exception reporting					
Mental illness						
Mean	NA	4.6±7.0	NA	8.1±13.4	NA	5.1±7.1
Median	NA	2.3	NA	0.0	NA	2.8
Interquartile range	NA	0.0–6.5	NA	0.0–12.5	NA	0.0–7.1
Range	NA	0.0–76.2	NA	0.0–80.0	NA	0.0–72.7
Stroke						
Mean	10.8±12.0	3.7±3.7	3.5±7.0	10.2±5.9	10.7±6.8	7.4±4.3
Median	8.3	2.7	0.0	9.5	9.5	6.6
Interquartile range	0.0–16.5	1.0–5.3	0.0–4.8	6.4–13.1	6.1–13.9	4.4–9.5
Range	0.0–85.7	0.0–39.2	0.0–66.7	0.0–57.1	0.0–62.9	0.0–36.3
Overall						
Mean	7.6±5.3	3.0±2.0	2.0±2.1	13.1±5.2	7.9±3.9	5.7±2.5
Median	6.5	2.5	1.4	12.6	7.1	5.3
Interquartile range	4.0–10.0	1.7–3.8	0.6–2.6	9.6–16.0	5.3–9.6	4.0–6.9
Range	0.0–50.8	0.0–25.6	0.0–28.6	0.0–47.1	0.0–42.2	0.0–28.3

* Plus–minus values are means ±SD. NA denotes not applicable.

In terms of disease groups, the lowest median rate of exception reporting was 0% for hypothyroidism (on the basis of a single indicator relating to thyroid-function tests during the preceding 15 months) and the highest was 7.6% for cancer (on the basis of a single indicator relating to review of care coordination within 6 months after diagnosis). Since each disease had a different set of indicators, direct comparisons of overall rates can be misleading; however, certain activities were replicated across disease groups, thereby facilitating some comparisons. For example, there were modest variations in rates of exception reporting for smoking-cessation activities (Table 3) and wider variations for the administration of influenza vaccine (from 9.8% for patients with coronary heart disease to 16.0% for patients with asthma) (for details, see the Supplementary Appendix).

In terms of type of activity, the lowest median rate of exception reporting was for offering treatment (1.4%), and the highest was for providing treatment (12.6%). Almost half the indicators involved regular review of patients with chronic diseases (e.g., routine measurements of blood pressure and cholesterol). This group of indicators had relatively low rates of exception reporting (median, 2.5%). The exception was the cancer indicator, which required a comprehensive review

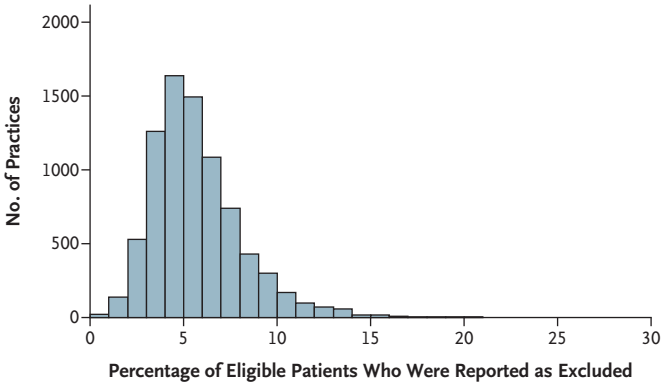

Figure 1. Overall Rates of Exception Reporting by 8105 Family Practices in England.

From April 2005 through March 2006, physicians excluded a median of 5.3% of patients (interquartile range, 4.0 to 6.9) from quality calculations in the pay-for-performance program (mean [±SD], 5.3±2.5). The percentages of exclusion reporting in the practices ranged from 0% to 28.3%.

Table 4. Regression Analyses Showing the Association of the Characteristics of Geographic Area, Patients, and Medical Practice with the Rates of Exception Reporting by 7629 Family Practices, According to Disease and Type of Activity.*

Variable	Disease									
	Asthma	Cancer	CHD	COPD	Diabetes	Epilepsy	Hyper-tension	Hypo-thyroidism	Mental Illness	Stroke
Geographic area										
Population density (persons/hectare)	0.02	–0.02	0.04†	0.01	0.05†	0.07‡	0.05‡	0.05†	0.00	0.01
Living in income-deprived household§	0.06‡	0.00	0.04	0.02	0.02	0.06‡	0.00	0.01	0.02	–0.02
Self-rated health status poor or fair	0.03	0.03	0.01	0.01	–0.02	0.08‡	0.01	–0.01	0.04†	–0.03
Patients										
Age										
≤15 yr	–0.09‡	–0.04	–0.09‡	–0.01	–0.03	0.00	–0.02	–0.04	–0.03†	0.00
≥65 yr	–0.13‡	–0.06‡	–0.11‡	–0.02	–0.13‡	–0.08‡	–0.13‡	–0.07‡	–0.04†	–0.04
Female sex	0.07‡	–0.00	0.07‡	0.03	0.00	0.02	–0.02	0.00	0.04‡	0.06‡
Member of racial or ethnic minority	–0.12‡	0.05	–0.03	0.02	–0.01	0.02	0.02	0.01	0.02	0.01
Practice										
Size of practice population	0.15‡	–0.01	0.00	0.04‡	0.05‡	–0.01	0.00	–0.04‡	0.09‡	0.04‡
No. of family practitioners per 10,000 patients	–0.01	0.01	0.00	0.00	0.00	0.00	0.03†	0.01	–0.01	0.03†
Personal Medical Services contract	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.00	0.03†	0.02
Proportion of family practitioners who are ≥55 yr of age	0.00	0.01	0.05‡	0.01	0.01	0.00	0.00	0.04†	–0.02	0.01
Proportion of family practitioners who are women	–0.03‡	0.01	0.01	0.02	–0.01	0.02	–0.01	–0.02	–0.01	–0.02
Medical education received in United Kingdom	0.01	0.01	0.02	–0.05‡	0.00	–0.01	0.04†	0.00	–0.03	0.00
Achievement in year 1	0.08‡	–0.10‡	0.12‡	–0.14‡	–0.03	0.04‡	–0.12‡	–0.05‡	–0.02	0.01
Model R^2¶	0.05	0.02	0.03	0.03	0.03	0.04	0.05	0.02	0.01	0.01

* Values are beta coefficients. The beta coefficient is the change in the standard deviation of rates of exception reporting for each 1-SD increase in the predictor variables. Complete socioeconomic data were not available for 476 practices, so these practices were excluded from the regression analyses. CHD denotes coronary heart disease, and COPD chronic obstructive pulmonary disease.
† The regression coefficient is significant at $P<0.05$.
‡ The regression coefficient is significant at $P<0.01$.
§ These households received means-tested government benefits for low-income people.
¶ Model R^2 refers to the proportion of variability in the data set that is explained by the statistical model.

of patients who potentially could be seriously ill. Rates of exception reporting for diagnostic and referral activities were generally higher than average (median, 6.5%), the exception being peak flow measurements for patients with asthma, which could be carried out in practice without referral to an external agency. The fifth group of activities — achieving intermediate outcomes — was the only one that involved outcomes rather than processes (e.g., achieving target levels of cholesterol and blood pressure), and rates of exception reporting for these activities were moderately high (median, 7.1%).

There was no correlation between the mean rate of exception reporting and the number of points (and hence financial rewards) available for individual indicators (Spearman's correlation coefficient, 0.06). The individual indicators with the highest median rates of exception reporting were prescribing beta-blockers for patients with coronary heart disease (24.7%), vaccinating patients with asthma against influenza (16.0%), and keep-

Type of Activity					Total
Diagnosis and Referral	Measurement and Review	Offer of Treatment	Provision of Treatment	Achievement of Intermediate Outcome	
0.00	0.05‡	0.02	0.04	0.05†	0.05†
0.03	0.04	0.01	0.04†	0.01	0.04
0.01	0.00	0.01	0.01	0.01	0.02
0.00	–0.01	0.00	–0.13‡	–0.02	–0.05†
0.04†	–0.08‡	0.00	–0.26‡	–0.15‡	–0.15‡
0.02	–0.02	–0.04	0.12‡	0.00	0.03
0.00	–0.01	–0.01	–0.08‡	0.03	–0.04
0.03†	0.07‡	0.09‡	0.06‡	0.02	0.07‡
0.01	0.01	0.00	–0.01	0.03†	0.01
0.03†	0.01	0.02	0.03†	0.01	0.02
0.01	0.00	0.00	0.03	0.02	0.01
0.01	–0.02	–0.04‡	0.00	0.00	–0.01
–0.02	0.02	0.05‡	–0.03	0.01	0.02
–0.18‡	–0.05‡	–0.04‡	0.16‡	–0.06‡	0.00
0.04	0.02	0.02	0.07	0.05	0.03

ing patients with epilepsy free from convulsions (13.8%).

CHARACTERISTICS OF PATIENTS AND PRACTICES

Increased overall rates of exception reporting were associated with practices located in densely populated areas and those with relatively small proportions of patients under the age of 16 years or over the age of 64 years (Table 4). Large practices also excluded a higher proportion of patients. Logistic-regression analysis identified the latter two variables — having a large practice or one with a skewed age range — as the only significant predictors of practices being in the top 1% of those reporting exceptions. All these effects were modest. For example, an increase of 1000 patients in the practice population was associated with an increase of 0.04% in the rate of exception reporting, and the variables that were included in the multiple regression model explained only 2.7% of the variance (R^2). Although there were some significant deviations for individual diseases and types of activity, these differences did not conform to any meaningful pattern.

Increased rates of exception reporting in 2005–2006 were associated with higher achievement rates in 2004–2005 for certain conditions (asthma, coronary heart disease, and epilepsy) and for activities that involve providing treatment; increased rates of exception reporting were associated with lower achievement rates for other conditions (cancer, chronic obstructive pulmonary disease, hy-

pothyroidism, and hypertension) and for all other types of activities. All these associations were weak in real terms. For example, for recording and diagnosing activities, a decrease in achievement rate of 1% in 2004–2005 was associated with an increase in the rate of exception reporting of 0.05% in 2005–2006.

FINANCIAL GAIN

Out of a maximum 492 points for the clinical-activity indicators, practices gained a median of 13.9 points (interquartile range, 8.6 to 21.2) from exception reporting. As a proportion of the total points available for these indicators, this score translated into a median of 2.8% (interquartile range, 1.7 to 4.3). The percentage gain ranged from 0% percent (in 15 practices) to 25.4% (in 1 practice) (Fig. 2). Given an average reward of £125 (approximately $250, in U.S. dollars) per point, practices gained a median of £1,738 ($3,476) as a result of exception reporting, with a maximum financial gain for an average-size practice of £15,500 ($31,000).

DISCUSSION

When the United Kingdom's pay-for-performance program was introduced, there was uncertainty about the effects of allowing physicians to exclude patients from the reporting system. Data were not available for exception reporting during the first year of the program, and rates could be estimated for only 30 of the 76 clinical indicators.[10] Our study describes a large-scale analysis of actual data on exception reporting, although the data are limited by the absence of reasons for exclusions and are available only at the practice level and not at the patient level. We also cannot differentiate between patterns of reporting by physicians, nurses, and practice managers, all of whom have the power to exclude patients.

Overall levels of exception reporting were generally modest in 2005–2006, averaging less than 6%. This rate is similar to the levels we estimated for the first year of the pay-for-performance program in 2004–2005 for a subgroup of clinical indicators.[10] However, the maximum rate of exception reporting appears to have decreased substantially, from an estimate of 86% in 2004–2005 to 28% in 2005–2006. This reduction in rates mainly affected practices with very high rates of exception reporting, since the 75th percentiles were similar in 2004–2005 and 2005–2006 (7.7% and 6.9%, respectively).

We estimate that exception reporting increased the income in practices by an average of £1,738 ($3,476) and the income of individual physicians by about £500 ($1,000), which is less than 0.5% of the average net income of a family practitioner. The total cost to the Department of Health in England was approximately £17.2 million ($34.4 million) in 2005–2006. Given that the overall cost of the pay-for-performance program in that year was approximately £1.15 billion ($2.3 billion) (Vickerman M: personal communication), all exception reporting accounted for 1.5% of the cost of the program.

The generally low overall rates of exception reporting masked some wide variations — for example, in rates of exception reporting for the same activity in different groups of patients. These variations were greatest for the targets for influenza vaccination, with an increase of 60% in rates of exception reporting for patients with asthma, as compared with patients with coronary heart disease. This difference might be explained in part by the different age profiles of these two groups of patients. There were also high exclusion rates for targets on reviewing the support needs of patients with mental-health issues or cancer, differences that might reflect long stays in secondary care or transfers to hospices and informal care providers outside the practice locale.

There was even greater variation in rates of exception reporting according to the type of activity. More straightforward activities, such as performing measures of disease and offering advice on smoking cessation, had very low rates for most disease groups, whereas rates for activities associated with administering treatments and achieving intermediate outcomes were increased by a factor of 6. Gaming is a possible explanation for these differences, since the activities with the highest rates of exception reporting tended to be more demanding of physicians. However, there are legitimate reasons for higher rates of exception reporting for treatment and intermediate outcome targets. Extreme frailty, for example, is a greater impediment to the successful treatment of hypertension than it is to simple monitoring of the condition.

The potential for gaming remains an argument against allowing exception reporting. We found only minor and inconsistent associations between

practice performance in the first year and exception reporting the following year, but any tendency to inflate exclusions fraudulently would be affected by the maximum-achievement thresholds. Analysis of data from Scotland has shown that rates of exception reporting in 2005–2006 were higher for practices that had levels of achievement below the maximum thresholds in the previous year than for those that had levels above the maximum thresholds.[24,25]

This finding raises the possibility that some practices may have gamed exceptions in order to receive maximum remuneration. However, we found no association between the rate of exception reporting and the size of the financial incentive available for individual indicators. Nor was it the case that activities for which it would be more difficult to fraudulently inflate achievement rates (due to the involvement of an external agency, such as a pathology laboratory or pharmacy) always had higher rates of exception reporting. These findings, along with the generally low levels of exclusions, suggest that extensive gaming of exception reporting is unlikely to have occurred.

Monitoring arrangements may have contributed to this finding. In England, Primary Care Trusts (statutory bodies that are entrusted with monitoring) annually inspect all practices as part of the pay-for-performance program and can access the reasons used for exception reporting. However, available resources currently limit these trusts to carrying out more detailed inspections on statistical outliers.

Another argument against allowing exception reporting is that excluded patients might be those in greatest need. In earlier studies, which involved smaller numbers of practices and either a limited number of quality indicators[26] or indirect measures of exception reporting,[27] practices in financially deprived areas were more likely to exclude patients. However, we found that the characteristics of practices and patients, including whether they were in an area of socioeconomic deprivation, had only marginal effects on rates of exception reporting. As compared with our analyses for 2004–2005,[10] this model explains less of the variation between practices. However, after accounting for extraneous error due to missing data, we found that the levels of variance explained by sociodemographic predictors were similar in both the first year and the second year of the program (for details, see the Supplementary Appendix). Our findings indicate that the high rates of achievement that were reported by practices across the sociodemographic spectrum[10] were not achieved because practices in deprived areas excluded many more patients.

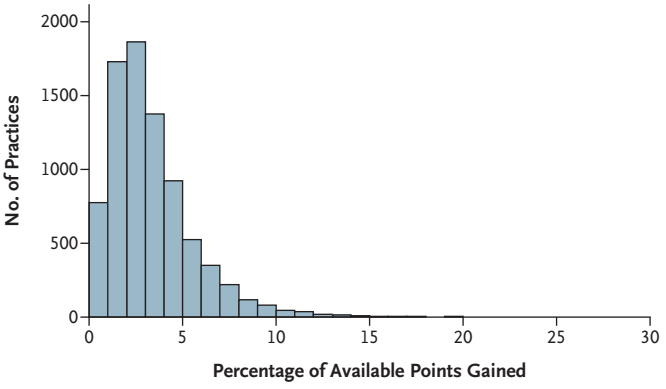

Figure 2. Percentage of Available Quality and Outcome Points Gained through Exception Reporting by 8105 Family Practices in England.

As a proportion of the total points available for these indicators, the score for the number of points gained through exception reporting translated into a median of 2.8% (interquartile range, 1.7 to 4.3), with a mean (±SD) of 3.3±2.3. The percentage gain ranged from 0% (in 15 practices) to 25.4% (in 1 practice).

An argument in favor of exception reporting is that when the provision is properly used, only patients for whom the targets are inappropriate are excluded. Maximum thresholds lack this precision and are fixed for all practices. If arbitrary thresholds are set too high, they may result in inappropriate treatment for some patients; if they are set too low, they allow physicians to receive maximum remuneration without treating all eligible patients. If exception reporting is allowed, maximum thresholds are superfluous for the purpose of avoiding inappropriate treatment, although they might still be required for indicators for which 100% achievement is unrealistic. If exception reporting is not allowed, then the adoption of maximum-achievement thresholds below 100% would be necessary to avoid inappropriate treatment. The results of our study give some indication of what the thresholds should be in this case.

The United Kingdom's experience with pay-for-performance programs provides some important lessons for other countries. As a system for safeguarding against inappropriate treatment, exception reporting has substantial benefits: it is precise, it increases acceptance of the program by physicians, and it ameliorates perverse incentives to refuse care to "difficult" patients.[28,29] The

feared drawbacks of exception reporting — that the provision would substantially increase the cost of the program and would be widely abused — appear to be unfounded. This finding may be related to peculiarities of the program in the United Kingdom, but there is no reason to believe that exception reporting would fail in other pay-for-performance programs, provided that effective systems to monitor for inappropriate exclusions and penalties for gaming are in place. Further longitudinal research is now required to study the complex links between exception reporting and levels of achievement, maximum thresholds, and size of incentives, as well as to explain why, rather than simply how often, physicians exclude certain patients from their reporting.

The National Primary Care Research and Development Centre receives core funding from the U.K. Department of Health. In 2002 and 2003, Dr. Roland was a member of a small team of academics who advised the government and professional negotiating teams that developed the U.K. pay-for-performance program. No other potential conflict of interest relevant to this article was reported.

The views expressed in this article are those of the authors and not necessarily those of the U.K. Department of Health.

REFERENCES

1. Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. JAMA 2005;294:1788-93.
2. Epstein AM. Pay for performance at the tipping point. N Engl J Med 2007;356:515-7.
3. *Idem.* Paying for performance in the United States and abroad. N Engl J Med 2006;355:406-8.
4. Medicare Australia. Practice Incentives Program (PIP). (Accessed June 23, 2008, at http://www.medicareaustralia.gov.au/provider/incentives/pip/index.shtml.)
5. Pink GH, Brown AD, Studer ML, Reiter KL, Leatt P. Pay-for-performance in publicly financed healthcare: some international experience and considerations for Canada. Healthc Pap 2006;6:8-26.
6. Greb S, Focke A, Hessel F, Wasem J. Financial incentives for disease management programmes and integrated care in German social health insurance. Health Policy 2006;78:295-305.
7. Custers T, Arah OA, Klazinga NS. Is there a business case for quality in The Netherlands? A critical analysis of the recent reforms of the health care system. Health Policy 2007;82:226-39.
8. Perkins R, Seddon M. Quality improvement in New Zealand healthcare. Part 5: measurement for monitoring and controlling performance — the quest for external accountability. N Z Med J 2006;119:U2149.
9. Roland M. Linking physicians' pay to quality of care — a major experiment in the United Kingdom. N Engl J Med 2004;351:1448-54.
10. Doran T, Fullwood C, Gravelle H, et al. Pay for performance programs in family practices in the United Kingdom. N Engl J Med 2006;355:375-84.
11. Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff 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.
12. Starfield B. New paradigms for quality in primary care. Br J Gen Pract 2001;51:303-9.
13. Yip WC. Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft (CABG) surgeries in the Medicare and private sectors. J Health Econ 1998;17:675-700.
14. Giuffrida A, Gravelle H. Inducing or restraining demand: the market for night visits in primary care. J Health Econ 2001;20:755-79.
15. Casalino LP, Elster A, Eisenberg A, Lewis E, Montgomery J, Ramos D. Will pay-for-performance and quality reporting affect health care disparities? Health Aff (Millwood) 2007;26:w405-w414.
16. Fisher ES. Paying for performance — risks and recommendations. N Engl J Med 2006;355:1845-7.
17. Bridges to Excellence home page. (Accessed June 23, 2008, at http://www.bridgestoexcellence.org/.)
18. The Physician Quality Reporting Initiative (PQRI): measures/codes. Baltimore: Centers for Medicare & Medicaid Services. (Accessed June 23, 2008, at http://www.cms.hhs.gov/PQRI/15_MeasuresCodes.asp.)
19. The Joint Commission home page. (Accessed June 23, 2008, at http://www.jointcommission.org/.)
20. Physician Quality Reporting Initiative: overview. Baltimore: Centers for Medicare & Medicaid Services. (Accessed June 23, 2008, at http://www.cms.hhs.gov/pqri/.)
21. Guthrie B, McLean G, Sutton M. Workload and reward in the Quality and Outcomes Framework of the 2004 general practice contract. Br J Gen Pract 2006;56:836-41.
22. NHS Information Centre. Quality Outcomes Framework data, 2005/06. London: National Health Service. (Accessed June 23, 2008, at http://www.ic.nhs.uk/servicesnew/qof06/ICfolder_view.)
23. Indices of deprivation 2004. London: Office of the Deputy Prime Minister. (Accessed June 23, 2008, at http://www.communities.gov.UK/archived/publications/communities/indicesdeprivation.)
24. Gravelle H, Sutton M, Ma A. Doctor behaviour under a pay for performance contract: evidence from the Quality and Outcomes Framework. CHE research paper 28. York, United Kingdom: Centre for Health Economics, University of York, 2007. (Available at http://www.york.ac.uk/inst/che/publications/publicationsbyyear.htm.)
25. *Idem.* Doctor behaviour under a pay for performance contract: further evidence from the Quality and Outcomes Framework. CHE research paper 34. York, United Kingdom: Centre for Health Economics, University of York, 2008. (Available at http://www.york.ac.uk/inst/che/publications/publicationsbyyear.htm.)
26. Sigfrid L, Turner C, Crook D, Ray S. Using the UK primary care Quality and Outcomes Framework to audit health care equity: preliminary data on diabetes management. J Public Health (Oxf) 2006;28:221-5.
27. McLean G, Sutton M, Guthrie B. Deprivation and quality of primary care services: evidence for persistence of the inverse care law from the UK Quality and Outcomes Framework. J Epidemiol Community Health 2006;60:917-22.
28. Casalino LP, Alexander GC, Jin L, Konetzka RT. General internists' views on pay-for-performance and public reporting of quality scores: a national survey. Health Aff (Millwood) 2007;26:492-9.
29. McMahon LF Jr, Hofer TP, Hayward RA. Physician-level P4P — DOA? Can quality-based payment be resuscitated? Am J Manag Care 2007;13:233-6.

CORRECTION

Exclusion of Patients from Pay-for-Performance Targets by English Physicians

Exclusion of Patients from Pay-for-Performance Targets by English Physicians . The reference list was inadvertently altered during the final processing of the file before publication. The corrected reference list is available with the article on the *Journal*'s Web site at www.nejm.org. We regret the error.