Frequency of Inappropriate Medical Exceptions to Quality Measures

Stephen D. Persell, MD, MPH; Nancy C. Dolan, MD; Elisha M. Friesema, BA; Jason A. Thompson, BA; Darren Kaiser, MS; and David W. Baker, MD, MPH

Background: Quality improvement programs that allow physicians to document medical reasons for deviating from guidelines preserve clinicians' judgment while enabling them to strive for high performance. However, physician misconceptions or gaming potentially limit programs.

Objective: To implement computerized decision support with mechanisms to document medical exceptions to quality measures and to perform peer review of exceptions and provide feedback when appropriate.

Design: Observational study.

Setting: Large internal medicine practice.

Participants: Patients eligible for 1 or more quality measures.

Measurements: A peer-review panel judged medical exceptions to 16 chronic disease and prevention quality measures as appropriate, inappropriate, or of uncertain appropriateness. Medical records were reviewed after feedback was given to determine whether care changed.

Results: Physicians recorded 650 standardized medical exceptions during 7 months. The reporting tool was used without any medical reason 36 times (5.5%). Of the remaining 614 exceptions, 93.6% were medically appropriate, 3.1% were inappropriate, and 3.3% were of uncertain appropriateness. Frequencies of inappropriate exceptions were 7 (6.9%) for coronary heart disease, 0 (0%) for heart failure, 10 (10.8%) for diabetes, and 2 (0.6%) for preventive services. After physicians received direct feedback about inappropriate exceptions, 8 of 19 (42%) changed management. The peer-review process took less than 5 minutes per case, but for each change in clinical care, 65 reviews were required.

Limitation: The findings could differ at other sites or if financial incentives were in place.

Conclusion: Physician-recorded medical exceptions were correct most of the time. Peer review of medical exceptions can identify myths and misconceptions, but the process needs to be more efficient to be sustainable.

Primary Funding Source: Agency for Healthcare Research and Quality.

Ann Intern Med. 2010;152:225-231. www.annals.org
For author affiliations, see end of text.

Performance measurement coupled with feedback (that is, audit and feedback) to health care providers has been shown to improve the quality of care (1). Computerized clinical decision support can also improve quality (2, 3). However, even when these methods are successful, the gains have generally been minimal, and quality usually remains far less than ideal.

Feedback to providers seems less effective when performance is already fairly good (1), and this may be true even when financial incentives are awarded on the basis of performance (4). As quality improves, an increasing proportion of apparent quality deficits—identified by clinical decision-support systems or on performance reports—may actually be incorrect. Valid medical exceptions explain these deficits (5–10). If physicians believe that the tools used to measure their performance are incorrect, they may not heed decision support or they may dismiss the results of performance reports, even if they are not achieving the desired results.

As financial incentives associated with performance measurement increase, measurement errors could produce undesired effects. Physicians may prescribe tests and treatments to patients for whom the benefits are insignificant because of concern that it will hurt their remuneration or publicly reported quality performance (5, 11, 12).

A potential solution is to incorporate standard ways for clinicians to record medical exceptions or to indicate when patients decline treatment into performance measurement systems as part of their routine workflow by using electronic health record (EHR) systems. The United Kingdom's primary care pay-for-performance system (13, 14) has done exception reporting for quality measurement, and some performance measures developed in the United States (15) have included this system. Exception reporting preserves clinicians' judgment but still enables them to strive for high levels of performance and avoid financial or other penalties for doing so. Furthermore, if standardized exceptions are recognized by clinical decision-support systems, the number of false alerts provided by these systems could be reduced.

Asking physicians to record exceptions to performance measures, however, raises some concerns. Clinicians could record invalid medical exceptions because of misconceptions. They could falsely enter medical exceptions to improve their measured performance (gaming).

See also:

Print
Editors' Notes 226

Web-Only
Appendix Table
Conversion of graphics into slides

Context

Some electronic medical record systems identify when a physician's order does not meet quality guidelines, notify the physician of a discrepancy, and allow the physician to indicate whether the situation is an exception to the guideline.

Contribution

Investigators studied 650 situations in which the physician indicated that the action taken was an exception to the guideline. They found that 94% of the exceptions were appropriate, 3% were inappropriate, and 3% were of uncertain appropriateness.

Caution

The study was done at a single institution that did not pay physicians for quality performance.

Implication

When physicians indicate that their actions are exceptions to guidelines, they are correct most of the time.

—The Editors

Both scenarios would result in falsely inflated performance measurements.

Building a review process to assess the validity of medical exceptions may address these concerns. This process could be used to identify an individual provider's misconception so that it could be corrected. The process could also identify common clinical situations in which physicians were uncertain about the best clinical strategy so that these could be addressed in future guidelines or investigated in future research.

As part of the UPQUAL project (Utilizing Precision Performance Measurement for Focused Quality Improvement), we created clinical decision-support tools in our EHR that allowed providers to enter medical or patient reasons for not following point-of-care alerts. In this analysis, we aimed to determine the frequency of medical misconceptions judged by a physician peer-review process and how often peer-to-peer feedback for inappropriate exceptions led to changes in clinical care.

Methods

Setting and Patients

We used data from a large, urban, academic internal medicine practice with an EHR (EpicCare, version spring 2007, Epic Systems Corporation, Verona, Wisconsin). Northwestern University's institutional review board approved the study. We included all patients if a physician or nurse practitioner recorded a medical exception to a quality measure into the EHR from 9 February to 10 September 2008. A total of 39 attending physicians, 1 nurse practitioner, and 121 residents worked at the clinic during this time.

Quality Improvement Context

Peer review of medical exceptions was one part of the larger UPQUAL study. This multicomponent quality-improvement system also included point-of-care clinical decision support, mechanisms for recording medical and patient exceptions to care standards, audit and feedback of performance to physicians, monthly reports of individual patients who were not receiving essential medications, and outreach to patients who declined recommended tests or treatments. When the study began, the practice had used this EHR for more than 10 years. Attending physicians had received reports indicating their performance on quality metrics approximately once each quarter for the preceding 2 years. On 9 February 2008, we initiated decision support for 16 chronic disease and preventive care quality measures. **Table 1** lists these measures. We provided brief training in person and through e-mail to encourage clinicians to use the decision support to order indicated services or to record exceptions. We informed clinicians that medical exceptions would be subject to peer review and that patient exceptions (records of patients who declined) would be used to direct outreach to individual patients.

Description of Decision Support and Medical Exceptions

Clinicians were alerted whenever a patient had 1 or more unaddressed quality measures when they opened a patient chart for an office visit or to document a telephone call. A single yellow highlighted bar on the side of the screen indicated that 1 or more outstanding quality alerts needed to be addressed. Clicking on this bar displayed the alerts to clinicians. Clinicians were not required to address the alerts. For each disease measure, clinicians could check a box on the alert to indicate that there was a medical reason why the measure was not met and describe the reason in a free-text comment. These comments could then be viewed on a summary page of the EHR so other providers could see why a patient had not received an intervention that was typically indicated. Preventive care alerts included a link to jump to the tracking system for health maintenance within the EHR in which clinicians could document reasons why the preventive service was not provided.

Sometimes clinical decision support was wrong because a patient did not have the medical condition in question. We built methods for physicians to indicate that a medical diagnosis previously entered in the EHR (coronary heart disease, diabetes, heart failure, or atrial fibrillation) was not present.

Some patients were not candidates for multiple services that would otherwise be indicated because they had an advanced life-limiting illness (for example, metastatic cancer or advanced dementia). In cases in which the care goals were palliation near the end of life, physicians could indicate that all decision support should be inactivated.

We used all of these recorded exceptions to remove patients from the eligible patient populations when we

Table 1. **Quality Measures and Measured Performance**

Quality Measure	Eligible Patients, *n**	Patients Satisfying Performance Measure (95% CI), %
CHD		
Antiplatelet drug prescribed in CHD	1202	89.8 (87.9–91.4)
Lipid-lowering drug prescribed or LDL-C level <2.6 mmol/L (<100 mg/dL) in CHD†	1202	87.4 (85.3–89.1)
β-Blocker prescribed for previous myocardial infarction	235	90.2 (85.7–93.7)
ACE inhibitor or ARB prescribed for CHD with diabetes	443	84.4 (80.7–87.7)
Heart failure		
β-Blocker prescribed for heart failure with LVSD	276	83.0 (78.0–87.2)
ACE inhibitor or ARB prescribed for heart failure with LVSD	276	85.1 (80.4–89.1)
Anticoagulation prescribed for atrial fibrillation and heart failure	106	65.1 (55.2–74.1)
Diabetes mellitus		
Diabetes control HbA_{1c} level <8.0% within the past year	1814	64.7 (62.5–66.9)
LDL-C level control <2.6 mmol/L (<100 mg/dL) within the past year	1595	53.3 (50.8–55.8)
Aspirin prescribed for primary prevention of CHD in patients aged ≥40 y	1695	78.2 (76.2–80.2)
Nephropathy screening or ACE inhibitor or ARB prescribed†	1814	81.4 (79.6–83.2)
Prevention and screening		
Mammography screening within the past 2 y for women aged 50–69 y	3539	81.1 (79.8–82.4)
Cervical cancer screening within the past 3 y for women aged 21–64 y	7462	84.7 (83.9–85.5)
Colon cancer screening within the past 1–10 y for patients aged 50–80 y†	7067	57.7 (56.5–58.9)
Pneumococcal vaccine for patients aged ≥65 y	2966	81.0 (79.5–82.4)
Osteoporosis screening or pharmacologic treatment for women aged ≥65 y‡	1816	78.3 (76.3–80.2)

ACE = angiotensin-converting enzyme; ARB = angiotensin-receptor blocker; CHD = coronary heart disease; HbA_{1c} = hemoglobin A_{1c}; LDL-C = low-density lipoprotein cholesterol; LVSD = left ventricular systolic dysfunction.
* Patients were eligible for quality measures if they had 2 or more office visits in the 18 months before 1 February 2008 and met the disease-specific or demographic criteria required for inclusion.
† Dependent on the procedure performed and the interval recorded in the electronic medical record indicating when repeated testing was due.
‡ Central bone density test completed once any time after age 60 years or treatment with a bisphosphonate, systemic estrogen, selective estrogen receptor modulators, parathyroid hormone, or calcitonin.

calculated performance measures for physicians' quality reports.

Physician Peer-Review Process and Feedback

We performed peer review for all exceptions entered into the EHR beginning on 9 February 2008, with the goal of performing at least 600 reviews. Every 1 to 2 weeks, we extracted all medical exceptions recorded in the EHR since the previous review. One physician reviewed medical records to collect the reason for the exception and additional clinical information needed to judge the validity of the exception. When the clinical reasoning was unclear, the peer reviewer would request clarification from the treating clinician. Three board-certified internists met regularly to review the exceptions. For some recorded medical exceptions, no real medical reason was noted (for example, "Cervical cancer screening was not done because I will do it at the next visit"). The internists judged these exceptions as having no medical reason present. In these cases, we notified physicians that they had used the exception reporting improperly, and we removed these exceptions from the clinical information system. The group reviewed the remaining medical exceptions and judged them as appropriate, inappropriate, or of uncertain appropriateness by consensus. When consensus was not reached or appropriateness was uncertain, 1 physician reviewed the medical literature and requested advice from specialists when needed, and the group discussed the case again until consensus was reached. Once the group had some experience with common appropriate medical exceptions, straightforward, appropriate exceptions were classified as such after single-physician review. We gave feedback directly to the treating clinician in cases in which the medical exception was judged to be inappropriate and gave recommendations to change management. We provided information from the medical literature or expert opinion in cases that were judged as uncertain when the peer-review panel felt there was valuable information for the treating physician to consider. We recorded all time spent performing the peer-review process.

Assessment of Uncertain and Inappropriate Cases After Feedback

We reviewed medical records after 3 months in cases in which feedback was given to determine the frequency of changes in subsequent clinical care. We considered a change in response to feedback to have occurred if the clinician ordered the test or treatment or documented that it was recommended. For cases in which physicians did not follow the measure because they stated that the diagnosis triggering the measure was not present but the peer-review panel believed it actually was, we considered that a change occurred if a new documentation of the diagnosis in the record (for example, a diagnosis of diabetes mellitus or coronary heart disease was appropriately added to the patient's problem list) was noted.

Analysis Plan

We calculated the proportion of eligible patients in the practice satisfying each of the 16 quality measures at the beginning of the study period. Patients were eligible if they were at least 18 years of age, had 2 or more visits to the clinic in the preceding 18 months, and met the characteristics that defined a particular measure. We calculated the percentage of patients who satisfied a performance measure as the percentage of eligible patients who either met the numerator criteria for the measure (received the service) or had an exception to the measure. We report descriptive statistics for all medical exceptions recorded during the study period. We did analyses by using SAS software, version 9.1 (SAS Institute, Cary, North Carolina).

Role of the Funding Source

The Agency for Healthcare Research and Quality funded the study. The funding source had no role in the design, conduct, or analysis of the study or in the decision to submit the manuscript for publication.

Results

Table 1 shows the number of patients eligible for each measure and baseline rates of performance. A total of 87 physicians (49 resident and 38 attending physicians) recorded 650 medical exceptions from 9 February to 10 September 2008.

Physicians used the medical-exception–reporting tool 36 times (5.5% [95% CI, 3.9% to 7.6%]) when the reason for not following the decision support was not due to any medical reason. Of the remaining 614 medical exceptions, 93.6% (CI, 91.4% to 95.4%) were judged as appropriate, 3.1% (CI, 1.9% to 4.8%) inappropriate, and 3.3% (CI, 2.0% to 5.0%) of uncertain appropriateness. Frequencies of inappropriate and uncertain exceptions were 7 (6.9% [CI, 2.8% to 13.6%]) and 10 (9.8% [CI, 4.8% to 17.3%]) for coronary heart disease, respectively; 0 ([CI, 0.0% to 4.3%]) and 2 (2.4% [CI, 0.3% to 8.4%]) for heart failure, respectively; and 10 (10.8% [CI, 5.3% to 18.9%]) and 8 (8.6% [CI, 3.8% to 16.2%]) for diabetes, respectively. For preventive service, nearly all medical exceptions were judged appropriate: 334 (99.4% [CI, 97.9% to 99.9%]). Only 2 (0.6% [CI, 0.1% to 2.1%]) were inappropriate, and none was of uncertain appropriateness. Of all medical exceptions recorded by physicians, 78 (12.7% [CI, 10.2% to 15.6%]) were instances in which a clinician recorded that a diagnosis that triggered a quality alert was not present. Peer reviewers disagreed with these exceptions 10.2% (CI, 4.5% to 19.2%) of the time and were uncertain 2.6% (CI, 0.3% to 9.0%) of the time. The **Appendix Table** (available at www.annals.org) shows frequencies of appropriate, inappropriate, and uncertain exceptions for each measure.

Of the cases with inappropriate exceptions in which physicians received direct feedback, 8 of 19 (42% [CI, 20% to 67%]) had a change in management after feedback. Two of 12 uncertain cases in which feedback was given resulted in a management change.

The peer-review process took 50 hours and 32 minutes (an average of 4 minutes and 40 seconds per case). Because physicians changed management in response to feedback in only 10 cases, an average of 5 hours and 3 minutes of physician-reviewer time was required to produce 1 management change.

Table 2 shows examples of medical exceptions that were judged appropriate, inappropriate, or of uncertain appropriateness. The topics that we identified included treatment of lipid disorders in patients who did not tolerate statins; treatment of lipid disorders in patients with chronic liver disease; evaluation of the risks and benefits for an angiotensin-receptor blocker in patients with angioedema from an angiotensin-converting enzyme inhibitor for different clinical indications; use of angiotensin-converting enzyme inhibitors or angiotensin-receptor blockers in patients with advanced renal disease; evaluation of the risks and benefits for antiplatelet or anticoagulant therapies in patients with particular bleeding risks; use of aspirin for cardiovascular risk reduction in patients receiving daily nonsteroidal anti-inflammatory drugs; and clarification of the health risks of diet-controlled diabetes.

Discussion

Clinical performance measures and clinical decision support based on practice guidelines have been criticized because the recommendations may be inappropriate for individual patients or because they may create undue burden for patients with many medical problems (5, 12). We designed clinical decision-support tools to create a system that would respect physician judgment about what is best for an individual patient and enable clinicians to record that information in an EHR during routine workflow. With these tools available in the EHR, we approached performance measurement with the expectation that physicians should provide guideline-recommended care or state why they did not. Our study shows that the reasons physicians entered are valid most of the time. We can use this information to deliver more accurate performance measurement to physicians, make subsequent clinical decision support more accurate, and make the rationale for deviating from guidelines visible to providers so it may be used in clinical care.

Concerns about gaming and fears that having physicians report exceptions would only codify myths and misconceptions were generally unfounded in this context in which measurement was used for internal quality improvement. Many exceptions were unequivocal contraindications or previous adverse drug events. However, some exceptions were highly nuanced assessments of the risks and benefits for an individual patient (for example, the decision not to prescribe a β-blocker in a patient with a remote myocardial infarction, low-risk clinical features, and the

Table 2. **Examples of Medical Exceptions**

Exception	Justification and Major Considerations
Appropriate	
Antiplatelet therapy or anticoagulation not prescribed in patients at increased risk for bleeding due to portal hypertensive gastropathy, esophageal varices, recent gastrointestinal hemorrhage, or ongoing blood loss from angiodysplasia of the gastrointestinal tract	Risks of therapy increased; bleeding risk may outweigh benefits
β-Blocker not prescribed after MI because of pronounced symptomatic bradycardia; no obstructive CHD, normal LVEF	Inability to tolerate the treatment; risks of therapy increased, and benefits may not justify risks
HbA_{1c} level in patients with uncontrolled diabetes due to labile blood sugar with severe episodes of hypoglycemia	Risks of intensifying therapy increased; risk may outweigh benefits
Mammography not done because of life-limiting comorbid condition	Benefits of screening are reduced
Inappropriate	
Aspirin not prescribed to a patient with high risk for CHD due to a remote retinal hemorrhage from diabetic retinopathy	Aspirin does not seem to increase the frequency and severity of retinal bleeding from diabetic retinopathy (16, 17)
Lipid-lowering drug not prescribed in CHD due to stable compensated cirrhosis	Generally safe to prescribe a statin in noncholestatic liver disease (18, 19); greater risk for muscle injury due to reduced drug metabolism, so low doses and more caution should be used; potential benefit is high in a patient with CHD
Diabetes not treated according to guidelines because diagnosis of diabetes was not present	Patients who previously met diagnostic criteria for type 2 diabetes who now have normal glucose and HbA_{1c} levels after making lifestyle changes still have type 2 diabetes
Uncertain appropriateness or complex medical decision making	
Lipid-lowering drug not prescribed in patient with CHD who developed muscle symptoms while receiving 2 statins (atorvastatin and simvastatin)	Several other statins may have a lower incidence of muscle adverse effects (20–22); nonstatin lipid-lowering drugs could be considered
β-Blocker not prescribed after MI, no LVSD, minimal coronary atherosclerosis, and concern for vasospasm (treated with an ARB and diltiazem)	Balance of risk and benefits for a regimen including a β-blocker vs. one without a β-blocker is uncertain in this clinical setting

ARB = angiotensin-receptor blocker; CHD = coronary heart disease; HbA_{1c} = hemoglobin A_{1c}; LVEF = left ventricular ejection fraction; LVSD = left ventricular systolic dysfunction; MI = myocardial infarction.

concern for vasospasm [**Table 2**]). Nevertheless, we found clear myths and misconceptions (for example, a patient with proliferative diabetic retinopathy having a contraindication to aspirin). Some misconceptions were cases with relative contraindications (for example, a patient with coronary disease and stable chronic liver disease receiving a statin) but were not applicable to the patient under review.

A previous study showed that requiring physicians to indicate why they were not following recommendations (that is, a "hard stop" requiring data entry) for preventive care improved adherence to recommended protocols (10, 23). This was not acceptable to physicians in our practice, so we did not require physicians to enter exceptions when they did not complete an indicated task nor did we provide the option to acknowledge that the recommended action was being deferred. Providers could only indicate when there was a medical reason why the action was not appropriate or that a patient declined for financial or nonfinancial reasons. Instead, our softer approach encouraged providers to enter exceptions because they would see their measured performance increase and the frequency of subsequent inappropriate alerts decline.

Our finding that physicians usually did not record inappropriate exceptions is generally consistent with reports from the United Kingdom's primary care pay-for-performance program. Initial findings from the United Kingdom suggested that the frequency of recorded exceptions had an important effect on observed performance, and few practices recorded exceptions at very high rates (13). A subsequent report, however, suggested that widespread gaming did not occur, and after the first year of the program, practices no longer reported exceptions for large numbers of patients (14). Additional studies of the frequency and distribution of exceptions in the U.K. program and others may be needed to fully understand and optimize performance measurement systems that permit exceptions.

The knowledge gained through this kind of review process could be useful to educators and the developers of guidelines or performance measures. However, our findings suggest that this approach is not practical for quality improvement in routine practice. The process was very labor-intensive and led to few changes in clinical care. Although most cases were reviewed quickly, in aggregate substantial physician effort was required to detect a few errors. The time required to yield 1 change in management seems prohibitive. Although the prevalence of misconceptions could vary greatly across different groups of physicians, the review process would need to be made much more efficient to be sustainable locations similar to this study site. Two potential ways to improve efficiency would be to build questions into EHR-based clinical decision-support mechanisms so that clinicians could rapidly record when a common appropriate exception was present (and peer review could be avoided) or to omit peer review for exceptions to

quality measures in which physicians rarely recorded inappropriate exceptions.

These findings should be viewed in the context of the size of the effect that accounting for medical exceptions may have on measured performance. The number of medical exceptions recorded during this study was often minimal in relation to the total number of quality deficiencies at the beginning of the study. Several additional limitations should be kept in mind when interpreting these results. Only 1 physician collected data directly from the medical records, and a single group of 3 physicians judged the appropriateness of medical exceptions through a consensus process that relied on review of the medical literature and expert opinion. A different chart abstractor or different physician-reviewers may have reached different conclusions. The generalizability of these findings could be limited by 2 factors. First, we studied a single physician group affiliated with an academic center. As previous research has shown, the range and type of physician misconceptions could vary greatly across practices or specialties (24–26). Second, the findings could also have been different even in the same practice if the physicians were working under different circumstances. During this study, no financial incentives were associated with physicians' measured performance, and physicians knew that peers would examine their exceptions. Physicians may be more likely to enter questionable or equivocal exceptions if they stand to gain financially and their exceptions are not subjected to peer review. Future work should examine whether applying financial incentives to measured performance increases the proportion of invalid or inappropriate exceptions that physicians record.

Our findings have important implications for public performance-reporting programs. If exceptions to quality measures can be recorded within the normal clinical workflow, then the level of performance that can reasonably be achieved becomes higher. This is in contrast to quality-measurement systems that depend on automated data sources alone that may not capture the clinical detail needed to determine when exceptions are present (8–10, 27–36). Because of limitations in these measurement systems, quality "benchmarks" are typically far less than 100% (4, 37–39). Even high-performing physician groups may only reach benchmarks in the 80% to 90% range. The gap between the benchmark and 100% is attributed to unidentified exceptions, patient preferences, or measurement error (38). The true failure rate for recommended tests or therapies remains obscured. In contrast, if exceptions to the guidelines can be recorded easily, clinicians and those who would hold them accountable through performance measurement programs could expect a level of performance that is much closer to 100%, allowing us to raise sights toward achieving the highest quality of care possible.

From Northwestern University, Chicago, Illinois.

Grant Support: By the Agency for Healthcare Research and Quality (grant 1R18HS17163-01). Dr. Persell was supported by a career development award (1K08HS015647-01) from the Agency for Healthcare Research and Quality.

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M09-1382.

Requests for Single Reprints: Stephen D. Persell, MD, MPH, Division of General Internal Medicine, Northwestern University, 750 North Lake Shore Drive, 10th Floor, Chicago, IL 60611; e-mail, spersell@nmff.org.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD.** Audit and feedback: effects on professional practice and health care outcomes. Cochrane Database Syst Rev. 2006:CD000259. [PMID: 16625533]

2. **Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al.** Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA. 2005;293:1223-38. [PMID: 15755945]

3. **Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al.** Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med. 2006;144:742-52. [PMID: 16702590]

4. **Rosenthal MB, Frank RG, Li Z, Epstein AM.** Early experience with pay-for-performance: from concept to practice. JAMA. 2005;294:1788-93. [PMID: 16219882]

5. **Boyd CM, Darer J, Boult C, Fried LP, Boult L, Wu AW.** Clinical practice guidelines and quality of care for older patients with multiple comorbid diseases: implications for pay for performance. JAMA. 2005;294:716-24. [PMID: 16091574]

6. **Wachter RM.** Expected and unanticipated consequences of the quality and information technology revolutions. JAMA. 2006;295:2780-3. [PMID: 16788133]

7. **Tinetti ME, Bogardus ST Jr, Agostini JV.** Potential pitfalls of disease-specific guidelines for patients with multiple conditions. N Engl J Med. 2004;351:2870-4. [PMID: 15625341]

8. **Persell SD, Wright JM, Thompson JA, Kmetik KS, Baker DW.** Assessing the validity of national quality measures for coronary artery disease using an electronic health record. Arch Intern Med. 2006;166:2272-7. [PMID: 17101947]

9. **Baker DW, Persell SD, Thompson JA, Soman NS, Burgner KM, Liss D, et al.** Automated review of electronic health records to assess quality of care for outpatients with heart failure. Ann Intern Med. 2007;146:270-7. [PMID: 17310051]

10. **Litzelman DK, Tierney WM.** Physicians' reasons for failing to comply with computerized preventive care guidelines. J Gen Intern Med. 1996;11:497-9. [PMID: 8872790]

11. **Werner RM, Asch DA.** The unintended consequences of publicly reporting quality information. JAMA. 2005;293:1239-44. [PMID: 15755946]

12. **Walter LC, Davidowitz NP, Heineken PA, Covinsky KE.** Pitfalls of converting practice guidelines into quality measures: lessons learned from a VA performance measure. JAMA. 2004;291:2466-70. [PMID: 15161897]

13. **Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al.** Pay-for-performance programs in family practices in the United Kingdom. N Engl J Med. 2006;355:375-84. [PMID: 16870916]

14. **Doran T, Fullwood C, Reeves D, Gravelle H, Roland M.** Exclusion of patients from pay-for-performance targets by English physicians. N Engl J Med. 2008;359:274-84. [PMID: 18635432]

15. **American Medical Association.** PCPI measures. Accessed at www.ama-assn.org/ama/pub/category/4837.html on 7 January 2009.

16. **Chew EY, Klein ML, Murphy RP, Remaley NA, Ferris FL 3rd.** Effects of aspirin on vitreous/preretinal hemorrhage in patients with diabetes mellitus. Early

Treatment Diabetic Retinopathy Study report no. 20. Arch Ophthalmol. 1995; 113:52-5. [PMID: 7826294]
17. **Bergerhoff K, Clar C, Richter B.** Aspirin in diabetic retinopathy. A systematic review. Endocrinol Metab Clin North Am. 2002;31:779-93. [PMID: 12227131]
18. **Lewis JH, Mortensen ME, Zweig S, Fusco MJ, Medoff JR, Belder R; Pravastatin in Chronic Liver Disease Study Investigators.** Efficacy and safety of high-dose pravastatin in hypercholesterolemic patients with well-compensated chronic liver disease: results of a prospective, randomized, double-blind, placebo-controlled, multicenter trial. Hepatology. 2007;46:1453-63. [PMID: 17668878]
19. **Cohen DE, Anania FA, Chalasani N; National Lipid Association Statin Safety Task Force Liver Expert Panel.** An assessment of statin safety by hepatologists. Am J Cardiol. 2006;97:77C-81C. [PMID: 16581333]
20. **Jacobson TA.** Toward "pain-free" statin prescribing: clinical algorithm for diagnosis and management of myalgia. Mayo Clin Proc. 2008;83:687-700. [PMID: 18533086]
21. **Glueck CJ, Aregawi D, Agloria M, Khalil Q, Winiarska M, Munjal J, et al.** Rosuvastatin 5 and 10 mg/d: a pilot study of the effects in hypercholesterolemic adults unable to tolerate other statins and reach LDL cholesterol goals with non-statin lipid-lowering therapies. Clin Ther. 2006;28:933-42. [PMID: 16860175]
22. **Stein EA, Ballantyne CM, Windler E, Sirnes PA, Sussekov A, Yigit Z, et al.** Efficacy and tolerability of fluvastatin XL 80 mg alone, ezetimibe alone, and the combination of fluvastatin XL 80 mg with ezetimibe in patients with a history of muscle-related side effects with other statins. Am J Cardiol. 2008;101:490-6. [PMID: 18312764]
23. **Litzelman DK, Dittus RS, Miller ME, Tierney WM.** Requiring physicians to respond to computerized reminders improves their compliance with preventive care protocols. J Gen Intern Med. 1993;8:311-7. [PMID: 8320575]
24. **Baker DW, Hayes RP, Massie BM, Craig CA.** Variations in family physicians' and cardiologists' care for patients with heart failure. Am Heart J. 1999; 138:826-34. [PMID: 10539812]
25. **Mosca L, Linfante AH, Benjamin EJ, Berra K, Hayes SN, Walsh BW, et al.** National study of physician awareness and adherence to cardiovascular disease prevention guidelines. Circulation. 2005;111:499-510. [PMID: 15687140]
26. **Chin MH, Friedmann PD, Cassel CK, Lang RM.** Differences in generalist and specialist physicians' knowledge and use of angiotensin-converting enzyme inhibitors for congestive heart failure. J Gen Intern Med. 1997;12:523-30. [PMID: 9294785]
27. **MacLean CH, Louie R, Shekelle PG, Roth CP, Saliba D, Higashi T, et al.** Comparison of administrative data and medical records to measure the quality of medical care provided to vulnerable older patients. Med Care. 2006;44:141-8. [PMID: 16434913]
28. **Fowles JB, Lawthers AG, Weiner JP, Garnick DW, Petrie DS, Palmer RH.** Agreement between physicians' office records and Medicare Part B claims data. Health Care Financ Rev. 1995;16:189-99. [PMID: 10151888]
29. **Fowles JB, Fowler E, Craft C, McCoy CE.** Comparing claims data and self-reported data with the medical record for Pap smear rates. Eval Health Prof. 1997;20:324-42. [PMID: 10183327]
30. **Fowles JB, Fowler EJ, Craft C.** Validation of claims diagnoses and self-reported conditions compared with medical records for selected chronic diseases. J Ambul Care Manage. 1998;21:24-34. [PMID: 10181337]
31. **Jollis JG, Ancukiewicz M, DeLong ER, Pryor DB, Muhlbaier LH, Mark DB.** Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. Ann Intern Med. 1993;119: 844-50. [PMID: 8018127]
32. **Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P.** Assessing the accuracy of administrative data in health information systems. Med Care. 2004; 42:1066-72. [PMID: 15586833]
33. **Solberg LI, Engebretson KI, Sperl-Hillen JM, Hroscikoski MC, O'Connor PJ.** Are claims data accurate enough to identify patients for performance measures or quality improvement? The case of diabetes, heart disease, and depression. Am J Med Qual. 2006;21:238-45. [PMID: 16849780]
34. **Hebert PL, Geiss LS, Tierney EF, Engelgau MM, Yawn BP, McBean AM.** Identifying persons with diabetes using Medicare claims data. Am J Med Qual. 1999;14:270-7. [PMID: 10624032]
35. **Newton KM, Wagner EH, Ramsey SD, McCulloch D, Evans R, Sandhu N, et al.** The use of automated data to identify complications and comorbidities of diabetes: a validation study. J Clin Epidemiol. 1999;52:199-207. [PMID: 10210237]
36. **Tang PC, Ralston M, Arrigotti MF, Qureshi L, Graham J.** Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures. J Am Med Inform Assoc. 2007;14:10-5. [PMID: 17068349]
37. **Kiefe CI, Allison JJ, Williams OD, Person SD, Weaver MT, Weissman NW.** Improving quality improvement using achievable benchmarks for physician feedback: a randomized controlled trial. JAMA. 2001;285:2871-9. [PMID: 11401608]
38. **Kiefe CI, Weissman NW, Allison JJ, Farmer R, Weaver M, Williams OD.** Identifying achievable benchmarks of care: concepts and methodology. Int J Qual Health Care. 1998;10:443-7. [PMID: 9828034]
39. **Wu B, Pope GC.** Left ventricular ejection fraction test rates for Medicare beneficiaries with heart failure. Am J Med Qual. 2002;17:61-6. [PMID: 11941996]

Current Author Addresses: Dr. Persell: Division of General Internal Medicine, Northwestern University, 750 North Lake Shore Drive, 10th Floor, Chicago, IL 60611. Dr. Dolan: Northwestern Medical Faculty Foundation, 675 North Saint Clair Street, Chicago, IL 60611.
Ms. Friesema, Mr. Thompson, and Dr. Baker: Northwestern University, 750 North Lake Shore Drive, 10th Floor, Chicago, IL 60611.
Mr. Kaiser: Northwestern Medical Faculty Foundation, 680 North Lake Shore Drive, 11th Floor, Chicago, IL 60611.

Author Contributions: Conception and design: S.D. Persell, N.C. Dolan, E.M. Friesema, D. Kaiser.
Analysis and interpretation of the data: S.D. Persell, N.C. Dolan, E.M. Friesema, J.A. Thompson, D.W. Baker.
Drafting of the article: S.D. Persell, N.C. Dolan, D.W. Baker.
Critical revision of the article for important intellectual content: N.C. Dolan, D. Kaiser, D.W. Baker.
Final approval of the article: S.D. Persell, N.C. Dolan, D.W. Baker.
Provision of study materials or patients: N.C. Dolan.
Statistical expertise: S.D. Persell, N.C. Dolan.
Obtaining of funding: N.C. Dolan, D.W. Baker.
Administrative, technical, or logistic support: S.D. Persell, N.C. Dolan, E.M. Friesema, D. Kaiser, D.W. Baker.
Collection and assembly of data: S.D. Persell, N.C. Dolan, E.M. Friesema, J.A. Thompson, D. Kaiser.

Appendix Table. Results of Peer Review of Medical Exceptions*

Medical Exception, by Category	Medical Exceptions
CHD	
Antiplatelet drug not prescribed for medical reason, *n*	24
Appropriate	19
Uncertain appropriateness	2
Inappropriate	3
Lipid-lowering drug not prescribed for medical reason, *n*	9
Appropriate	7
Uncertain appropriateness	1
Inappropriate	1
β-Blocker not prescribed for medical reason, *n*	14
Appropriate	11
Uncertain appropriateness	2
Inappropriate	1
ACE inhibitor or ARB not prescribed for medical reason, *n*	23
Appropriate	20
Uncertain appropriateness	3
Clinical recommendation not followed because CHD diagnosis was not present, *n*	31
Appropriate	27
Uncertain appropriateness	2
Inappropriate	2
Clinical recommendation not followed because myocardial infarction diagnosis was not present, *n*	1
Appropriate	1
All CHD medical exceptions, *n*	102
Appropriate [95% CI], *n (%)*	85 (83.3 [74.7–90.0])
Uncertain appropriateness [95% CI], *n (%)*	10 (9.8 [4.8–17.3])
Inappropriate [95% CI], *n (%)*	7 (6.9 [2.8–13.6])
Heart failure	
β-Blocker not prescribed for medical reason, *n*	32
Appropriate	32
ACE inhibitor or ARB not prescribed for medical reason, *n*	21
Appropriate	19
Uncertain appropriateness	2
Anticoagulation not prescribed for medical reason, *n*	13
Appropriate	13
Clinical recommendation not followed because heart failure diagnosis was not present, *n*	8
Appropriate	8
Clinical recommendation not followed because atrial fibrillation diagnosis was not present, *n*	9
Appropriate	9
All heart failure medical exceptions, *n*	83
Appropriate [95% CI], *n (%)*	81 (97.6 [91.6–99.7])
Uncertain appropriateness [95% CI], *n (%)*	2 (2.4 [0.3–8.4])
Diabetes mellitus	
Uncontrolled diabetes HbA_{1c} level ≥8.0% due to medical reason, *n*	5
Appropriate	4
Inappropriate	1
Uncontrolled LDL-C level due to medical reason, *n*	6
Appropriate	4
Uncertain appropriateness	2
Aspirin for primary prevention not prescribed for medical reason, *n*	39
Appropriate	32
Uncertain appropriateness	4
Inappropriate	3
Nephropathy screening not done due to medical reason, *n*	14
Appropriate	12
Uncertain appropriateness	2
Clinical recommendation not followed because diabetes diagnosis was not present, *n*	29
Appropriate	23
Inappropriate	6
All diabetes medical exceptions, *n*	93
Appropriate [95% CI], *n (%)*	75 (80.6 [71.1–88.1])
Uncertain appropriateness [95% CI], *n (%)*	8 (8.6 [3.8–16.2])
Inappropriate [95% CI], *n (%)*	10 (10.8 [5.3–18.9])

Appendix Table—Continued

Medical Exception, by Category	Medical Exceptions
Prevention and screening*	
Mammography screening not done due to medical reason, *n*	17
Appropriate	17
Cervical cancer screening not done due to medical reason, *n*	273
Appropriate	271
Inappropriate	2
Colon cancer screening not done due to medical reason, *n*	15
Appropriate	15
Pneumococcal vaccine not given due to medical reason, *n*	1
Appropriate	1
Stop all reminders due to a medical reason, *n*	30
Appropriate	30
All prevention and screening medical exceptions, *n*	336
Appropriate [95% CI], *n (%)*	334 (99.4 [97.9–99.9])
Inappropriate [95% CI], *n (%)*	2 (0.6 [0.1–2.1])
Total, *n*	614
Appropriate [95% CI], *n (%)*	575 (93.6 [91.4–95.4])
Uncertain appropriateness [95% CI], *n (%)*	20 (3.3 [2.0–5.0])
Inappropriate [95% CI], *n (%)*	19 (3.1 [1.9–4.8])

ACE = angiotensin-converting enzyme; ARB = angiotensin-receptor blocker; CHD = coronary heart disease; HbA_{1c} = hemoglobin A_{1c}; LDL-C = low-density lipoprotein cholesterol.

* No medical exceptions were recorded for the osteoporosis screening or treatment measure.