ORIGINAL ARTICLE

The ethical problem of false positives: a prospective evaluation of physician reporting in the medical record

T R Dresselhaus, J Luck, J W Peabody

See end of article for authors' affiliations

....

Correspondence to:
Dr J W Peabody, San Francisco Veterans Affairs Medical Center, c/o Institute for Global Health, University of California, San Francisco, 74 New Montgomery St, Ste 508, San Francisco, CA 94105, USA;
peabody@psg.ucsf.edu

Revised version received 12 February 2002

Accepted for publication 11 March 2002

....

Objective: To determine if the medical record might overestimate the quality of care through false, and potentially unethical, documentation by physicians.
Design: Prospective trial comparing two methods for measuring the quality of care for four common outpatient conditions: (1) structured reports by standardised patients (SPs) who presented unannounced to the physicians' clinics, and (2) abstraction of the medical records generated during these visits.
Setting: The general medicine clinics of two veterans affairs medical centres.
Participants: Twenty randomly selected physicians (10 at each site) from among eligible second and third year internal medicine residents and attending physicians.
Main measurements: Explicit criteria were used to score the medical records of physicians and the reports of SPs generated during 160 visits (8 cases × 20 physicians). Individual scoring items were categorised into four domains of clinical performance: history, physical examination, treatment, and diagnosis. To determine the false positive rate, physician entries were classified as false positive (documented in the record but not reported by the SP), false negative, true positive, and true negative.
Results: False positives were identified in the medical record for 6.4% of measured items. The false positive rate was higher for physical examination (0.330) and diagnosis (0.304) than for history (0.166) and treatment (0.082). For individual physician subjects, the false positive rate ranged from 0.098 to 0.397.
Conclusions: These data indicate that the medical record falsely overestimates the quality of important dimensions of care such as the physical examination. Though it is doubtful that most subjects in our study participated in regular, intentional falsification, we cannot exclude the possibility that false positives were in some instances intentional, and therefore fraudulent, misrepresentations. Further research is needed to explore the questions raised but incompletely answered by this research.

Medical records are the benchmark for assessing competence and determining what clinicians do in the course of patient visits.[1–3] Despite their prominent place in quality measurement, chart abstraction is subject to important limitations, including the expense of abstraction and, for paper formats, illegibility and record unavailability.[4 5] Perhaps the most important limitation of medical records as a measure of clinical performance is that physicians do not document everything they do. This recording bias contributes to a high false negative rate, meaning that chart abstraction may underestimate the actual quality of care.[6 7]

This observation of recording bias led us to ask if the medical record might also overestimate quality due to false positive reporting by clinicians. We hypothesised that the medical record not only lacks sensitivity (due to false negatives) but also specificity (due to false positives). If present, false positives would certainly raise additional questions about the reliability of the record as a quality measure and the integrity of physician documentation. False positives also raise substantive ethical questions, including the possibility of intentional deception and fraud.[8]

Though a growing body of literature recognises the problem of recording bias and other causes of underreporting, few investigators have addressed the potential problem of erroneous inclusions in the medical record. This is primarily due to the methodological challenge inherent in measuring such errors. The use of the standardised patient (SP) encounter overcomes this obstacle, however, because SPs are a gold standard measure against which to measure not only false negatives but also false positives in the medical record.[9–13] Thus, to determine if false positives exist in the medical record, we report on a study that compares the quality of care documented by physician subjects with the quality of care reported by actor patients (case-mix controlled). We then consider the ethical concerns that emerge from such an evaluation.

METHODS

Data was collected in the general medicine clinics of two VA medical centres between December 1996 and August 1997 using methods described elsewhere.[4] All second and third year residents and attending physicians in these clinics were eligible to participate; of these, 97% consented to participate. From consenting physicians, we randomly selected 20 participants, ten at each site.

Quality of care provided by these physician participants was determined by two methods: (1) structured reports by standardised patients (SPs) who presented unannounced to the physicians' clinics and (2) abstraction of the medical records generated from these visits, in accordance with physicians' informed consent. The reports of SPs were the gold standard method.[10]

Both methods measured the process of care for four common medical conditions: low back pain, chronic obstructive pulmonary disease, diabetes mellitus, and coronary artery disease. Two detailed clinical scenarios for each diagnosis were developed, one simple and one complex, generating eight cases.

We recruited 27 experienced actors to serve as SPs. They were trained by university-based educators to present a scripted clinical scenario and to recall and record specific details of the physician encounter through completion of checklists. After training, the SPs presented unannounced to

Table 1 2 × 2 table comparing gold standard (SP) *v* index method (chart abstraction) responses

	Standardised patient	
	Documented (% total responses)	Not documented (% total responses)
Chart abstraction		
Reported	TP = 2153 (48.3%)	FP = 285 (6.4%)
Not reported	FN = 916 (20.5%)	TN = 1106 (24.8%)

TP – true positive; FN – false negative; FP – false positive; TN – true negative

Table 2 Proportion false positives and false positive rate (1 − specificity) by domain

| Domain	False positives/total responses (proportion)	False positive rate (1 − specificity)
Overall	285 / 4460 (0.064)	0.205
History	99 / 2580 (0.038)	0.166
Physical exam	127 / 940 (0.135)	0.330
Diagnosis	35 / 240 (0.146)	0.304
Treatment	24 / 700 (0.034)	0.082

the physicians' clinics; their identities were concealed from examining physicians and other clinic staff. Immediately after each visit, the SPs completed "checklist" reports to document physician performance. Simultaneously, charts generated at these visits were retrieved for purpose of abstraction by a trained nurse. In all, with ten subjects at each of the two sites, each seeing eight cases, there were a total of 160 visits. Sample size calculations were based on an estimated difference observed in earlier studies that ranged from 5–10% with standard deviation of 7%.[14] To determine if actor patients were detected, physician subjects were asked after the conclusion of the study to report encounters in which they suspected the patient was an actor.

Scoring used explicit quality criteria for each of the eight cases, derived from national guidelines and expert panels of academic and community physicians.[5 9 15–18] The number of scoring items for each case ranged from 25 to 35. These criteria explicitly and comprehensively captured the process of outpatient primary care. Identical criteria were used for both methods (standardised patient and chart abstraction). Individual scoring items were categorised into four domains of clinical performance: history, physical examination, treatment, and diagnosis.

Using the SP as the gold standard method, physician entries in the medical record were classified for each quality criteria as true positive (reported by SP, documented in record), false negative (reported by SP, not documented in record), false positive (not reported by SP, documented in record) and true negative (not reported by SP, not documented in record). As in prior analyses, individual items were treated as independent observations.[9] The proportion of total responses that were false positive and the false positive rate (1 − specificity) were determined for each of the four domains of the clinical encounter (history, physical examination, diagnosis, and treatment) and, overall, for each of the 20 physician subjects. The false positive rate was determined also for each of the 27 actor patients, for the two study sites, and for each of the four clinical conditions. A receiver-operator characteristic (ROC) curve was generated to compare physician subjects' false positive rates (1 − specificity) and true positive rates (sensitivity).

RESULTS

Compared to the gold standard of standardised patients, false positives were identified in the medical record for 6.4% of
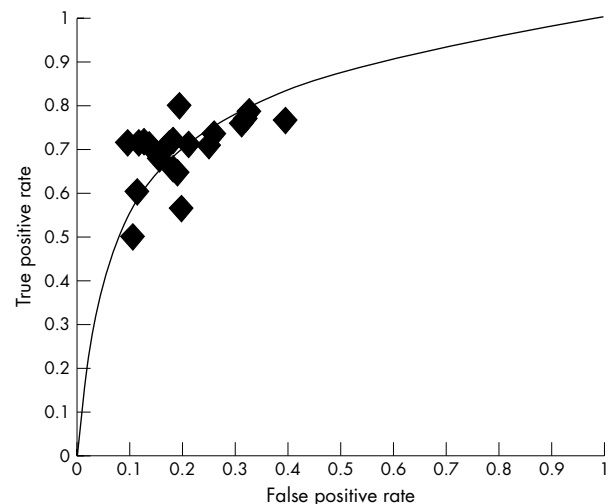
Figure 1 Receiver-operator characeristic (ROC) curve.

measured items overall and false negatives for 20.5% of measured items (see table 1). As a proportion of responses, false positives were higher for physical examination (13.5%) and diagnosis (14.6%) than for history (3.8%) and treatment (3.4%) (see table 2). Correspondingly, the false positive rate (1 − specificity) was highest for physical examination (0.330) and diagnosis (0.304).

For individual physician subjects, the proportion of false positives ranged from 2.2% to 13.0% and the false positive rate from 0.098 to 0.397. Five physicians had false positive rates above 0.25. Similarly, the false positive rates for actor patients ranged from 0.06 to 0.396. Eight actor patients had false positive rates above 0.25.

The plot of false positive rates versus true positive rates for physician subjects is typical of a receiver-operator curve, with the false positive rate rising in a curvilinear, positive relationship to the true positive rate (figure 1).

The false positive rate was similar between the two study sites (0.192 for site 1; 0.224 for site 2). The false positive rate was highest for case 2 (0.294) but comparable across the remaining cases (see table 3). Importantly, detection of standardised patients was minimal and occurred in only 5/160 (3%) visits.

Table 3 Proportion of false positives and false positive rate (1 – specificity) by condition

Case	False positives / total response (proportion)	False positive rate (1 – specificity)
Overall	285 / 4460 (0.064)	0.205
Condition 1	71/1220 (0.058)	0.180
Condition 2	98/1100 (0.089)	0.294
Condition 3	74/1160 (0.064)	0.177
Condition 4	42/980 (0.043)	0.171

DISCUSSION

Though an accepted benchmark for quality measurement, the medical record must be critically reappraised in light of emerging data. As these data indicate, the record is subject to recording bias, leading to underestimation of the actual quality of care.[7] The data presented in this analysis also indicate that the medical record is flawed by false positives. This may lead to overestimates of the quality of important dimensions of care such as the physical examination.

These results do not appear to be incidental, as they cluster around specific domains and range widely in distribution among physician subjects. Nor are they explained by under-reporting of actor patients, who have been demonstrated to be a reliable gold standard for measuring physician performance.[9 10] In this analysis, false positives did not cluster around individual actor patients.

Given time constraints and the inherent complexity of the patient-physician interaction, it might be anticipated that physician subjects would not document all that they do. Given the emphasis on truth-telling as a cornerstone of professionalism and ethical practice, however, it is perhaps surprising to observe a pattern of false documentation in the record. How might this be explained?

We observe that the false positive rate is highest in the domains of diagnosis and physical examination. For diagnosis, this suggests that physicians documented diagnostic considerations in the record that were not conveyed to the patient. One explanation is that time constraints, inherent in an increasingly cost-constrained health care settings, may limit the amount of patient-physician communication during the course of an evaluation.[19] Alternatively, latent or even overt "paternalism" on the part of physicians may further restrain information sharing.[20] Either of these explanations constitute an error of omission with important consequences: patient education is compromised, patient participation in decision making is hindered, and the process of informed consent is potentially undermined.

For the physical examination, the false positive findings are less easily rationalised. Most innocently, these false positives may be explained as careless documentation by some physician subjects or even unwitting reconstructions meant to convey anticipated rather than actual findings. The physical examination is a very specific component of the patient evaluation, however, and is likely memorable to both actor patient and physician. Thus, careless documentation by the physician or omissions by the actor patient would be inadequate to explain the high false positive rates observed among some subjects.

A more serious explanation is the possibility that these false positives are, in some instances, intentional misrepresentations of the process of care. If so, several possible motivations exist. First, false documentation of the physical examination could be used to up-code a visit for billing purposes; however, this is unlikely in this setting, as a minority of patients are billed for services. Second, the physical examination is time-consuming to perform, and a clinician might opt to falsify anticipated exam findings in order to expedite a time-pressured visit. Additionally, falsification could be a face-saving manoeuvre when an important exam element, omitted during the patient encounter, is remembered after the conclusion of the interview.

As indicated by the ROC curve, the subjects with the highest true positive rate also tended to have the highest false positive rate. In other words, physicians who provided (and documented) higher quality of care also made more false positive errors. By doing more, perhaps these physicians had greater difficulty accurately reconstructing the process of care. Alternatively, physicians who provided more comprehensive, higher quality care may have been more concerned with omissions and more likely therefore to embellish the record or fabricate specific results.

One other explanation is unlikely. Because data was collected by standardised patients, it is possible that, if unmasked, actors would be viewed differently from usual patients, perhaps as a test. However, with only three per cent of standardised patients detected, this explanation can be discarded.

These findings confirm a problem with the accuracy of the medical record. We believe it is doubtful that most subjects in our study participated in regular, intentional falsification of the record. However, we cannot exclude the possibility that false positives were in some instances intentional, and therefore fraudulent, misrepresentations. Such behaviour is not uncommon in other settings. Physicians are known to engage in deception in order to secure reimbursement from insurers, though such incentives would not pertain to the institutional setting for this study.[21 22] Surveys of house staff indicate that nearly half have witnessed actual falsification of patients' records by others and that a minority would fabricate lab values or test results to save face.[23 24]

Even if the observed false positives in this study were innocent or unintended, they nonetheless erode the integrity of the medical record in several ways. Such errors propagate misinformation to others, who expect the record to reliably reflect key historical, examination and diagnostic information at a point in time. Findings at subsequent encounters, if compared to erroneous past documentation, could lead to diagnostic and therapeutic mistakes, with consequent harm to patients. Payers who rely upon the medical record to determine reimbursement may be misled, with consequent financial implications for patients and for society. And audits of quality of care based upon physician documentation may give false impressions of individual or aggregate performance if derived from flawed records. Regardless of motive, inaccurate and false information constitutes a serious threat to the fidelity of the record and therefore the fidelity of the process of care.

Further research is needed to explore the questions raised but incompletely answered by this research. As the electronic record becomes the standard for physician documentation, new threats to the integrity of the record emerge. Templates and other time-saving mechanisms offer new possibilities for embellishing the record and propagating misinformation. The increase in documentation requirements and the growing

scrutiny of the medical record only raise the incentives to falsify it. In this context, quality of documentation must be recognised as itself an important dimension of quality. And physicians must reaffirm their historical commitment to truthtelling and accuracy in their communication with one another and their patients.

FUNDING

. . . .

Authors' affiliations

T R Dresselhaus, Veterans Affairs San Diego Healthcare System, University of San Diego, California, USA
J Luck, University of California Los Angeles School of Public Health and the Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, USA
J W Peabody, Institute for Global Health, University of California, San Francisco Veterans Affairs Medical Center, and University of California, Los Angeles School of Public Health, Los Angeles, California, USA

REFERENCES

1 **McDonald CJ**, Overhage JM, Dexter P, *et al.* A framework for capturing clinical data sets from computerized sources. *Annals of Internal Medicine* 1997;**127**:675–82.
2 **Rubin HR**, Rogers WH, Kahn KL, *et al.* Watching the doctor-watchers: how well do peer review organization methods detect hospital care quality problems? *Journal of the American Medical Association* 1992;**267**:2349–54.
3 **Gilbert EH**, Lowenstein SR, Koziol-McLain J, *et al.* Chart reviews in emergency medicine research: where are the methods? *Annals of Emergency Medicine* 1996;**27**:305–8.
4 **Peabody JW**, Luck J, Glassman P, *et al.* Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *Journal of the American Medical Association* 2000;**283**:1715–22.
5 **Lawthers AG**, Palmer RH, Edwards JE, *et al.* Developing and evaluating performance measures for ambulatory care quality: a preliminary report of the DEMPAQ project. *Joint Commission Journal on Quality Improvement* 1993;**19**:552–65.
6 **Garnick DW**, Fowles J, Lawthers AG, *et al.* Focus on quality: profiling physicians' practice patterns. *Journal of Ambulatory Care Management* 1994;**17**:44–75.
7 **Luck J**, Peabody JW, Dresselhaus TR, *et al.* How well does chart abstraction measure quality? A prospective comparison of quality between standardized patients and the medical record. *American Journal of Medicine* 2000;**108**:642–9.
8 **Leonardo JA**. Healthcare fraud: a critical challenge. *Managed Care Quarterly* 1996;**4**:67–79.
9 **Glassman PA**, Peabody JW, O'Gara E, *et al.* Using standardized patients to measure quality: evidence from the literature and a prospective study. *Joint Commission Journal on Quality Improvement* 2000;**26**:644–53.
10 **Peabody JW**, Luck J, Glassman P, *et al.* Listening in: assessing the validity of standardized patient ratings of resident and attending performance. Society of General Internal Medicine 24th Annual Meeting, San Diego, CA, May 2001.
11 **Badger LW**, deGruy F, Hartman J, *et al.* Stability of standardized patients' performance in a study of clinical decision making. *Family Medicine* 1995;**27**:126–31. Colliver JA, Swartz MH. Assessing clinical performance with standardized patients. *Journal of the American Medical Association* 1997;**278**:1164–8.
12 **De Champlain AF**, Marglois MJ, King A, *et al.* Standardized patients' accuracy in recording examinees' behaviors using checklists. *Academic Medicine* 1997;**72**(suppl 1):85–7S.
13 **See reference 11:** Colliver JA, Swartz MH.
14 **Cohen F**. *Statistical power analysis for the behavioral sciences* [2nd ed]. Hilsdale, NJ: Lawrence Earlbaum Associates, 1998: 273–406.
15 **American Diabetes Association**. Standards of medical care for patients with diabetes mellitus. *Diabetes Care* 1995;**18**:8–15.
16 **American Thoracic Society.** Standards for the diagnosis and care of patients with chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine* 1995;**15** (suppl):78–121S.
17 **Bigos SJ**, Bowyer O, Braen G**,** *et al. Acute low back pain in adults*. Clinical practice guideline no 14 (AHCPR publication no 95–0642). Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, US Department of Health and Human Services, December 1994.
18 **Ryan TJ**, Anderson JL, Antman EM, *et al.* ACC/AHA guidelines for the management of patients with acute myocardial infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Acute Myocardial Infarction). *Journal of the American College of Cardiology* 1996;**28**:1328–428.
19 **Emanuel EJ**, Dubler NN. Preserving the physician-patient relationship in the era of managed care. *Journal of the American Medical Association* 1995;**273**:323–9.
20 **Emanuel EE**, Emanuel LL. Four models of the physician-patient relationship. *Journal of the American Medical Association* 1992; **267**:2221–7.
21 **Wolf CJ**. Deception in psychiatric reimbursement. *American Journal of Forensic Medicine and Pathology* 2001;**22**:7–17.
22 **Cain JM**. Is deception for reimbursement in obstetrics and gynecology justified? *Obstetrics and Gynecology* 1993;**82**:475–8.
23 **Baldwin DC Jr**, Daugherty SR, Rowley BD. Unethical and unprofessional conduct observed by residents during their first year of training. *Academic Medicine* 1998;**73**:1195–200.
24 **Green MJ**, Farber NJ, Ubel PA, *et al.* Lying to each other: when internal medicine residents use deception with their colleagues. *Archives of Internal Medicine* 2000;**160**:2317–23.