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

T R Dresselhaus, J Luck, J W Peabody

M edical records are the benchmark for assessing competence and determining what clinicians do in the course of patient visits. 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. 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.

This observation of recording bias led us to ask if the medical record might overestimate the quality of care reported by physicians. We hypothesised that the medical record would overestimate 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. 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 during these visits.

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 the physicians’ clinics, and the reports of SPs were the gold standard for determining whether each item was recorded. The reports of SPs were the gold standard for determining whether each item was recorded.

RESULTS

False positives were identified in the medical record for 6.4% of measured items. The false positive rate was higher for history (0.330) and diagnosis (0.304) than for physical examination (0.300) 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.
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%.

Compared to the gold standard of standardised patients, false positives were identified in the medical record for 6.4% of observed in earlier studies that ranged from 5–10% with standard deviation of 7%.

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 domains of clinical conditions. A receiver-operator characteristic (ROC) curve was generated to compare physician subjects’ false positives with the true positive rate (1 – specificity).

### RESULTS
Compared to the gold standard of standardised patients, false positives were identified in the medical record for 6.4% of 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 1 2 × 2 table comparing gold standard (SP) v index method (chart abstraction) responses

<table>
<thead>
<tr>
<th></th>
<th>Standardised patient</th>
<th>Chart abstraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reported</td>
<td>Not reported</td>
</tr>
<tr>
<td>FN</td>
<td>916 (20.5%)</td>
<td>1106 (24.8%)</td>
</tr>
<tr>
<td>TP</td>
<td>2153 (48.3%)</td>
<td>285 (6.4%)</td>
</tr>
<tr>
<td>FN – false negative</td>
<td>916 (20.5%)</td>
<td>1106 (24.8%)</td>
</tr>
<tr>
<td>TP – true positive</td>
<td>2153 (48.3%)</td>
<td>285 (6.4%)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Domain</th>
<th>False positives/total responses (proportion)</th>
<th>False positive rate (1 – specificity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>285 / 4460 (0.064)</td>
<td>0.205</td>
</tr>
<tr>
<td>History</td>
<td>99 / 2580 (0.038)</td>
<td>0.166</td>
</tr>
<tr>
<td>Physical exam</td>
<td>127 / 940 (0.135)</td>
<td>0.330</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>35 / 240 (0.146)</td>
<td>0.304</td>
</tr>
<tr>
<td>Treatment</td>
<td>24 / 700 (0.034)</td>
<td>0.082</td>
</tr>
</tbody>
</table>

### Figure 1 Receiver-operator characteristic (ROC) curve.
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.1 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.8-10 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 setting, may limit the amount of patient-physician communication during the course of an evaluation.10 Alternatively, latent or even overt “paternalism” on the part of physicians may further restrain information sharing.11 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

Table 3

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

www.jmedethics.com
References


12. See reference 11: Colliver JA, Swartz MH.


