Accuracy of Data in Computer-based Patient Records

William R. Hogan, MD, Michael M. Wagner, MD, PhD

Abstract Data in computer-based patient records (CPRs) have many uses beyond their primary role in patient care, including research and health-system management. Although the accuracy of CPR data directly affects these applications, there has been only sporadic interest in, and no previous review of, data accuracy in CPRs. This paper reviews the published studies of data accuracy in CPRs. These studies report highly variable levels of accuracy. This variability stems from differences in study design, in types of data studied, and in the CPRs themselves. These differences confound interpretation of this literature. We conclude that our knowledge of data accuracy in CPRs is not commensurate with its importance and further studies are needed. We propose methodological guidelines for studying accuracy that address shortcomings of the current literature. As CPR data are used increasingly for research, methods used in research databases to continuously monitor and improve accuracy should be applied to CPRs.

Data in computer-based patient records (CPRs) are used in patient care, clinical research, health-system management, health-services planning, total quality improvement, billing, risk management, and government reporting. The accuracy of these data is therefore of great importance. On the basis of inaccurate data, clinicians may make treatment errors, researchers may underestimate disease prevalence, health-system managers may underestimate compliance with standards of care such as vaccination guidelines, and alerting systems may send false alarms to physicians. It is therefore surprising that the amount of research devoted to measuring data accuracy in CPRs has been relatively small.

In contrast, there is extensive literature on data accuracy in paper-based records, disease registries, and clinical trial databases. This body of work provides a well-developed framework for the analysis of data accuracy that we can apply to CPRs. For example, Komaroff provides a comprehensive review of the complex processes by which different types of medical data are recorded into traditional paper-based medical records (and how error may be introduced at each step). His description of these processes can be expanded to describe how CPRs capture data (Fig. 1), thereby providing a model for the study of causes of inaccuracy. The literature on data accuracy in computer-based registries and clinical trial databases provides standard methods for the study of data accuracy, which are applicable to CPRs. In this literature, accuracy is calculated using two measures—one that measures the proportion of recorded observations in the system that are correct (correctness) and a second that measures the proportion of observations that are actually recorded in the system (completeness) (Fig. 2). These measures are viewed as complementary; both measures are necessary for a complete understanding of accuracy in a system. From Figure 2, we can see that completeness decreases as the number of false negatives (cell c in Figure 2) increases, and correctness decreases as the number of false positives (cell b) increases.

Affiliation of the authors: Center for Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA.

This research was partially supported by grant LM07059-10 from the National Library of Medicine.

Correspondence and reprints: William R. Hogan, MD, Center for Biomedical Informatics, Suite 8084 Forbes Tower, 200 Lothrop Street, Pittsburgh, PA 15213-2582.

E-mail: wrh3+@pitt.edu

Received for publication: 1/17/97; accepted for publication: 5/19/97.
The number of false positives (cell b in Figure 2) increases. The false positives and negatives are largely independent of one another; thus, both measures provide valuable information about the accuracy of data in any system. Finally, previous work also contributes the understanding that the ideal gold standard for data accuracy is the true state of the patient (or more generally, whatever aspect of the world the data represents).

This ideal is usually difficult, if not impossible, to achieve, and researchers have devised methods for approximating it. Different methods are preferred depending on the type of data under study and resources available to the researchers. For example, in cancer registries, biopsy data are generally preferred as a gold standard for measuring data accuracy as opposed to reports of radiographic studies or data abstracted from the paper-based patient record.

We had three major objectives in conducting this review. First, we wanted to determine the quality of the literature on data accuracy in CPRs. Second, we wanted to form a synthesis of the results reported by this literature to answer the following open questions about data accuracy in CPRs:

- How accurate are data contained in CPRs?
- What are the causes of inaccurate data?
- Which CPR characteristics influence data accuracy, and does direct clinician entry of data into the CPR result in higher rates of correctness and completeness than entry of data by third parties?
- How can we improve data accuracy in CPRs?
- Is the accuracy of CPR data higher than the accuracy of data in paper-based records?

<table>
<thead>
<tr>
<th></th>
<th>Gold standard positive</th>
<th>Gold standard negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPR positive</td>
<td>a</td>
<td>b</td>
<td>a - b</td>
</tr>
<tr>
<td>CPR negative</td>
<td>c</td>
<td>d</td>
<td>c + d</td>
</tr>
<tr>
<td>Total</td>
<td>a + c</td>
<td>b - d</td>
<td>a + b + c + d</td>
</tr>
</tbody>
</table>

Completeness = $a/(a+c)$  Correctness = $a/(a+b)$

Correctness is the proportion of observations that are a correct representation of the true state of the world and is calculated as $a/(a+b)$—equivalent to sensitivity. Negative predictive value and specificity can also be derived from this table but are not measured in data accuracy because the "d cell" may be infinitely large (it is unfeasible to count the number of observations that were not made that should not have been made).
structured a textword search that retrieved citations containing at least one of the following words related to the concept of accuracy: accuracy, accurate, inaccuracy, inaccuracies, inaccurate, reliability, reliable, unreliability, unreliable, valid, validity, invalid, invalidity, correct, correctness, incorrect, incorrectness, complete, completeness, incomplete, incompleteness, error, erroneous, quality. We generated this list of words iteratively by performing a search, adding words that we found in citations, then repeating the search. We also required that articles be indexed under the MeSH term INFORMATION SYSTEMS.† We employed a second MEDLINE strategy to retrieve articles not indexed under INFORMATION SYSTEMS. This search retrieved articles containing at least one of the following phrases: data accuracy, accuracy data, data inaccuracy, inaccuracy data, inaccuracies data, data quality, quality data, data error, data errors, and erroneous data. We also searched CURRENT CONTENTS from October 1995 to February 1996 to identify articles not yet indexed by MEDLINE. We used our first MEDLINE strategy without the INFORMATION SYSTEMS restriction. Finally, we performed a citation search using SCISEARCH to identify articles that referenced an early review of the accuracy of medical data.⁵

One author (WRH) reviewed the tables of contents of all Proceedings of the Annual Symposium on Computer Applications in Medical Care (1977–1995) and the American Association for Medical Systems and Informatics Congress (1982–1989).

This same author (WRH) reviewed the titles and abstracts of citations retrieved by the previously mentioned searches, excluded citations obviously not relevant to data accuracy in clinical information systems (e.g., articles that presented data on air quality), and obtained copies of all remaining articles. WRH reviewed the references cited by these articles for additional articles.

These searches retrieved 2,443 citations from MEDLINE, approximately 800 citations from CURRENT CONTENTS, 35 citations from the citation search, approximately 2,500 citations from conference proceedings (citations of all papers from all conference proceedings were reviewed), and approximately 500 citations from reviewing references of retrieved articles. Manual review of the titles and abstracts of approximately 6,278 citations yielded 235 articles potentially relevant to data accuracy in CPRs.

Study Selection

We obtained the 235 articles and reviewed them to exclude articles that did not satisfy the following three criteria: (1) a CPR was the object of study, which we defined as a computer-based system that contains primary patient records, defined by the Institute of Medicine (IOM) as records . . . used by health care professionals while providing patient care services to review patient data or document their own observations⁶; (2) a gold standard to which computer records were being compared was stated; and (3) correctness or completeness, or data from which we could compute at least one of them was reported for at least one type of data. If an article described multiple studies, we included only those studies that met these criteria. Each author reviewed each of the 235 articles independently using a structured form. Inter-observer agreement was 92% for whether to include a given article. Differences of opinion were resolved by subsequent joint review of the article. Twenty articles satisfied our criteria for inclusion in this review.

Study Evaluation

We could find no standard method in the literature for critiquing studies about data accuracy. Thus, we developed an ad hoc scoring scheme to rate the articles.

Three of the journal articles reported the results of multiple studies of data accuracy; we scored each study from the 20 articles individually. There was a total of 26 different studies. The two authors independently scored each of these 26 studies on a scale of 0 to 18, using the scoring system described in the following paragraphs:

1. CPR description: 5 points. We awarded one point each for a description of methods of data capture, scope (e.g., how many clinics the CPR was operating in), general data content (i.e., what types of data the CPR contained), accessibility (e.g., from which locations a user can access the CPR, how much “down” time does the CPR experience), and whether the CPR constituted the official patient record.

The reason for scoring studies based on the description of CPR is that certain CPR characteristics may influence data accuracy. Therefore knowledge of these characteristics is necessary for the interpretation of results about data accuracy. For example, the method of data capture may influence accuracy—data captured on structured encounter forms or by direct clinician input may be more accurate than transcription of clinicians’ unstructured, handwritten or dictated notes.

†Specifically, the MEDLINE search logic was ((accuracy.tw. OR accurate.tw. OR . . . OR quality.tw.) AND (exp INFORMATION SYSTEMS)).
The scope of the CPR might influence accuracy because patients often visit other clinics or health care providers who may not use the CPR, and thus the care rendered to them goes unrecorded in the CPR. The types of data contained in the CPR might influence rates of accuracy because certain types of data such as demographics and medication data are likely to be more accurate than other types, such as diagnoses or problem lists. Accessibility may influence accuracy because the times and locations from which data can be entered may be inconvenient or limited in number, and thus people may defer data entry rather than disrupt their work routine. This deferral of data entry may result in data either not being recorded or being recorded less accurately. Finally, if the CPR data are the official patient record, it is conceivable that those responsible for data entry will take more care when recording data.

2. Methodology: 12 points. We awarded two points for unbiased sampling techniques, including random selection or contiguous selection (e.g., all patients who visited the clinic during a predefined study period). We awarded four points if the members of the research team who were responsible for determining the gold standard were blinded to both the purpose of the study and the CPR data, and we awarded two points if they were blinded to either the study purpose or the CPR data. We awarded four points for gold standards that most closely approximated the true state of the patient, i.e., those determined by interview, examination, observation of patients, or an objective measurement; we awarded two points for gold standards determined by review of other patient data (e.g., the paper record). Finally, we awarded two points if both measures of accuracy—correctness and completeness—were measured.

3. Study Objective: 1 point. We awarded one point if the primary study objective was the measurement of data accuracy.

We resolved differences in the two scores for studies according to the following procedure: If our scores for a study differed by one point, we took the higher of the two. If our scores differed by two or more points, we reabstracted the information in question and, if necessary, jointly reviewed the article to resolve our differences.

Analysis

We expected significant variability in the CPRs studied and methods used to measure accuracy. Because of this variability, pooling of results would not be possible, and the techniques of formal meta-analysis would not be valid. Instead, we abstracted from each study (1) CPR characteristics, (2) methods, and (3) results about accuracy. We calculated methodological and description scores for the 26 studies and investigated the CPR characteristics and other results that pertained to the second objective of the review—answering questions about data accuracy.

Results

We found 20 articles that reported the results of 26 studies of accuracy in 19 unique CPRs (two articles described studies done in the same CPR). The 20 articles were published over a period spanning 18 years. Fourteen (70%) were published in the 5 years preceding this review. The primary objective of all but one study was the measurement of data accuracy in a CPR (Table 1).

The 26 studies varied in the quality of the description of the CPR and the quality of the methods used (overall score 8 to 14, mean 10.3 of 18; see Table 1). The descriptions of the CPRs were generally adequate (description score 2 to 5, mean 3.5 of 5); however, only one study reported all five CPR characteristics that we identified at the beginning of the review as potentially influencing data accuracy. The CPR characteristics most frequently reported were the scope and content of the CPR (26 studies) and methods of data capture (22 studies). Only 16 of the 26 studies provided the name of the CPR, or any other information about its hardware components or software versions. Finally, only three studies indicated whether CPR data constituted the official medical record.

Many studies had significant methodological weaknesses (methodological scores 4 to 10, mean 5.9 of 12). Common weaknesses were reporting only one measure of accuracy (14 studies) and failure to blind the members of the research team who were responsible for determining the gold standard to the purpose of the study or the CPR data (15 studies). The number of patient records sampled ranged from 5 to 7,717. Approaches to sampling were: inclusion of all CPR records or all records for single diagnoses (nine studies), random sampling (eight studies), and consecutive patients (nine studies). Most studies used inadequate gold standards. Thirteen studies employed unblinded review of paper-based records. In eight studies, investigators merely checked for blank fields or missing data elements rather than determining a gold standard that approximated the true state of the patient. Only three studies employed examination, interview, or observation of patients in the determination of the gold standard.

The 26 studies reported rates of correctness, complete-
<table>
<thead>
<tr>
<th>Study</th>
<th>Sampling</th>
<th>Gold Standard</th>
<th>$n^*$</th>
<th>Description</th>
<th>Methods</th>
<th>Objective</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jelovsek/Hammond, 1978</td>
<td>All for a 3-year period</td>
<td>Computer algorithm to identify blank fields</td>
<td>7717</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Fortinsky/Gutman, 1981</td>
<td>Random</td>
<td>Paper records</td>
<td>109</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Jones/Hedley, 1986</td>
<td>All for a 1-year period</td>
<td>Computer algorithm to identify blank fields</td>
<td>1307</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Maresh et al., 1986</td>
<td>Consecutive deliveries of infants</td>
<td>Paper records</td>
<td>253</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Dambro/Weiss, 1988</td>
<td>Random sample taken weekly</td>
<td>Committee review of paper records</td>
<td></td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Block/Brennan, 1989</td>
<td>Study (1): random</td>
<td>(1) Paper records</td>
<td>(1) 388, 405</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Study (2): all records for certain diagnoses</td>
<td>1925, 1565</td>
<td></td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Gouveia-Oliveira et al., 1991</td>
<td>All diagnostic reports for certain diagnoses</td>
<td>No missing reports and no missing descriptors</td>
<td></td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Johnson et al., 1991</td>
<td>All patients from selected practices</td>
<td>A well-established manual influenza surveillance system</td>
<td></td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Kuhn et al., 1991</td>
<td>All or random reports for certain diagnoses</td>
<td>No missing descriptors</td>
<td>210–642</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Barrie/Marsh, 1992</td>
<td>Random from 18-month period</td>
<td>Paper records</td>
<td>200</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Kuhn et al., 1992</td>
<td>All or random reports for certain diagnoses</td>
<td>No missing descriptors</td>
<td>50, 52</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Edsall et al., 1993</td>
<td>Consecutive knee arthroscopies under general anesthesia</td>
<td>No missing observations for 46 data items</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Payne et al., 1993</td>
<td>Study (1): not stated</td>
<td>(1) Patient observation</td>
<td>(1) 234</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Study (2): all children of a certain age</td>
<td>(2) Paper records</td>
<td>(2) 218</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Study (3): all children of a certain age</td>
<td>(3) Paper records</td>
<td>(3) 104, 542</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Ricketts et al., 1993</td>
<td>Sequential admissions</td>
<td>Paper records</td>
<td>100</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Barlow et al., 1994</td>
<td>Random from 24-month period</td>
<td>Paper records</td>
<td>200</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Hohnloser et al., 1994</td>
<td>All for 18-month period</td>
<td>Recoding of original free text report</td>
<td>1219</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Wilton/Pennisi, 1994</td>
<td>Consecutive visits of children &lt;2 years old</td>
<td>Paper records</td>
<td>2098</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Pringle et al., 1995</td>
<td>Study (1): all records for certain diagnoses</td>
<td>(1) Paper chart plus CPR medication list</td>
<td>(1) Not given</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Study (2): all records</td>
<td>(2) No missing fields</td>
<td>(2) Not given</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Study (3): not stated</td>
<td>(3) No missing fields</td>
<td>(3) 1000</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Study (4): consecutive visits</td>
<td>(4) Review of videotape of doctor taking history</td>
<td>(4) 50</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Yarnall et al., 1995</td>
<td>Random from 2-year period</td>
<td>Paper records</td>
<td>300</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Wagner/Hogan, 1996</td>
<td>All records for 3-week study period</td>
<td>Clinician interpretation of medication history + paper</td>
<td>117</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>13</td>
</tr>
</tbody>
</table>

$^a$Number of patient records that were reviewed.

$^b$Studied data accuracy before and after an intervention to improve the accuracy of data.

$^c$Study only evaluated data-entry error.

$^d$Every 20th chart in an alphabetical arrangement of paper charts.

$^e$A laboratory log book was used here as the gold standard for presence or absence of a disease.

$^f$Four types of lesions seen on endoscopy.

$^g$Study included only those practices that could provide the total number of monthly visits and recorded at least one respiratory illness.

$^h$Six diagnoses were evaluated.

$^i$Two diagnoses were evaluated.

$^j$Study included four practices known to be among the best in recording CPR data.
ness, or both for 35 types of data (Table 2). Completeness was reported for all 35 types. Correctness was reported for only 13 (37%) types. The rates of correctness and completeness showed high variability, even within types of data. Correctness and completeness of diagnoses—the most frequently studied type of data (13 studies)—ranged from 67 to 100% and from 30.7 to 100%, respectively. Certain types of data, such as immunization status, medications, and demographics, tended to be more accurate than other types, such as problem lists and complications of surgical procedures, but the lowest rates of accuracy for the “more” accurate types of data overlapped with the highest rates of accuracy for the “less” accurate types.

A study by the authors of this paper was the only one that investigated causes of errors as a primary study objective. In this study of medication data in an outpatient CPR, the most common cause of inaccuracy was the patient, who either provided incorrect information or created a discrepancy between true state and CPR data by changing medications without a clinician’s instruction (36% of inaccuracies). The second most common cause was failure to capture medication changes made by clinicians who were not part of the clinic (26%), a result of CPR scope being limited to one clinic. Other causes included clinicians recording medication changes on paper but not in the CPR (13%) and clinic physicians making changes while outside the clinic and therefore recording them in neither the paper record nor the CPR (9%). Surprisingly, transcription error was a minor cause of inaccuracy (8%). The cause of 8% of inaccuracies could not be determined.

The 19 CPRs varied in ways that may influence data accuracy (Table 3). Data-capture mechanisms included transcription of data from paper-based records (10 CPRs), direct data entry by clinicians during patient care (6 CPRs), transcription of clinician dictation into the CPR (4 CPRs), and automatic capture of data from electronic patient-monitoring systems (1 CPR). Two CPRs captured data using more than one method. The types of data captured by the CPRs ranged from only a single type, such as endoscopy reports, to a comprehensive set of patient data. The scope of the CPRs also varied widely, ranging from a single clinic to a large health-maintenance organization.

Because of the variability in methods and CPRs, and because of the small number of these studies, we could not detect relationships between data accuracy and CPR characteristics across studies. However, several individual studies provided results that are informative about the relationship between CPR characteristics and data accuracy. Three studies suggest that broadening the scope of the CPR may improve completeness of CPR data.

One study measured the accuracy of data entered directly by clinicians versus data-entry personnel in the same version of the CPR at the same time. In this study, no significant difference in the accuracy of data entered directly by clinicians versus data entered from encounter forms by licensed nurses’ aides was found. However, this study lacked statistical power to demonstrate a small difference. Kuhn and associates measured data accuracy in a new version of their CPR that used direct physician entry of data into structured, electronic forms, and they compared it to the accuracy of an earlier version of the CPR based on physicians’ dictation of unstructured reports. They showed a significant improvement in accuracy with direct physician entry; however, this result is confounded by a potential checklist effect of the structured form introduced in the new version of the CPR and by the use of a historical control.

Several studies investigated interventions designed to improve data accuracy by measuring accuracy before and after the intervention. Fortinsky and Gutman found that structured encounter forms significantly increased completeness of diagnosis recording relative to unstructured forms. Yarnall and associates found that prompting physicians with previously recorded diagnostic codes on an encounter form improved correctness and completeness of diagnosis coding over the previous system (where physicians rewrote diagnoses on the billing sheet at every patient visit). Dambr and Weiss found that periodic monitoring of data accuracy and feedback to physicians and transcriptionists improved correctness of data entry. All three of these studies used historical controls, and thus the improvements in accuracy may have been due, at least in part, to other factors.

No study compared directly the accuracy of CPR data with the accuracy of data in a paper-based record. Because many of the 26 studies in CPRs used a paper-based record as the gold standard against which CPR data were compared, we cannot even compare data accuracy in CPRs with reported rates of accuracy in paper-based records. Such a comparison is logical and free of bias only when both systems, paper based and computer based, are measured against the same gold standard. We consider this issue further in the discussion.

One study demonstrated that factors external to the CPR may have a powerful influence on data accuracy. Ricketts and associates compared data accuracy at two similar hospitals using the same CPR system and found large differences in accuracy. They attributed these differences to the presence of a systems coordi-
### Table 2

Correctness and Completeness for Various Data Types

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Study</th>
<th>Correctness (%)</th>
<th>Completeness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnoses/problem list</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>Fortinsky/Gutman, 1981&lt;sup&gt;a&lt;/sup&gt;</td>
<td>94.4, 95.2</td>
<td>84.9, 89.8</td>
</tr>
<tr>
<td>Pringle et al., 1995 (Study 3)</td>
<td>—</td>
<td>81.8</td>
<td></td>
</tr>
<tr>
<td>Pringle et al., 1995 (Study 4)</td>
<td>—</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Proteinuria</td>
<td>Maresh et al., 1986</td>
<td>95</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Jones/Hedley, 1986</td>
<td></td>
<td>90.3</td>
</tr>
<tr>
<td>Smoking status</td>
<td>Block/Brennan, 1989 (Study 1)</td>
<td>90.9</td>
<td>30.7</td>
</tr>
<tr>
<td>Pringle et al., 1995 (Study 2)</td>
<td>—</td>
<td>52.1</td>
<td></td>
</tr>
<tr>
<td>Pregnancy</td>
<td>Block/Brennan, 1989 (Study 1)</td>
<td>—</td>
<td>92.3</td>
</tr>
<tr>
<td>Anemia—adults</td>
<td>Block/Brennan, 1989 (Study 2)</td>
<td>—</td>
<td>54.0&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Anemia—children</td>
<td>Block/Brennan, 1989 (Study 2)</td>
<td>—</td>
<td>35.0&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Urinary tract infection</td>
<td>Block/Brennan, 1989 (Study 2)</td>
<td>—</td>
<td>54.8&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Orthopedic diagnoses</td>
<td>Barrie/Marsh, 1992</td>
<td>96.5</td>
<td>58.7</td>
</tr>
<tr>
<td></td>
<td>Ricketts et al., 1993</td>
<td>67, 91&lt;sup&gt;c&lt;/sup&gt;</td>
<td>53, 74&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Codes for diagnoses</td>
<td>Fortinsky/Gutman, 1981&lt;sup&gt;a&lt;/sup&gt;</td>
<td>90.5, 91.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yarnall et al., 1995&lt;sup&gt;a&lt;/sup&gt;</td>
<td>77, 88&lt;sup&gt;d&lt;/sup&gt;</td>
<td>62, 82&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Hematology diagnoses</td>
<td>Hohnloser et al., 1994</td>
<td>74.1</td>
<td>54.5</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>Pringle et al., 1995 (Study 1)</td>
<td>100</td>
<td>96.7</td>
</tr>
<tr>
<td>Glaucoma</td>
<td>Pringle et al., 1995 (Study 1)</td>
<td>100</td>
<td>92.3</td>
</tr>
<tr>
<td>Asthma</td>
<td>Pringle et al., 1995 (Study 1)</td>
<td>—</td>
<td>65.1</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>Pringle et al., 1995 (Study 1)</td>
<td>—</td>
<td>59.0</td>
</tr>
<tr>
<td>Medications/prescriptions</td>
<td>Pringle et al., 1995 (Study 3, 4)</td>
<td>—</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Wagner/Hogan, 1996</td>
<td>83</td>
<td>93</td>
</tr>
<tr>
<td><strong>Procedures/operations</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthopedic</td>
<td>Barrie/Marsh, 1992</td>
<td>97.8</td>
<td>82.0</td>
</tr>
<tr>
<td></td>
<td>Ricketts et al., 1993</td>
<td>44, 86&lt;sup&gt;c&lt;/sup&gt;</td>
<td>43, 80&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Barlow et al., 1994</td>
<td>98</td>
<td>92.5</td>
</tr>
<tr>
<td>Complications of procedures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orthopedic</td>
<td>Barrie/Marsh, 1992</td>
<td>92.9</td>
<td>45.9</td>
</tr>
<tr>
<td></td>
<td>Ricketts et al., 1993</td>
<td>50, 77&lt;sup&gt;e&lt;/sup&gt;</td>
<td>17, 66&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Demographic data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>Jelovsek/Hammond, 1978</td>
<td>—</td>
<td>90.5&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Occupation</td>
<td>Jelovsek/Hammond, 1978</td>
<td>—</td>
<td>34.4</td>
</tr>
<tr>
<td>Marital status</td>
<td>Jones/Hedley, 1986</td>
<td>—</td>
<td>96.6</td>
</tr>
<tr>
<td>Date of birth</td>
<td>Jones/Hedley, 1986</td>
<td>—</td>
<td>100</td>
</tr>
<tr>
<td>Immunization Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All vaccines</td>
<td>Payne et al., 1993 (Study 1)</td>
<td>91.6</td>
<td>99.1</td>
</tr>
<tr>
<td></td>
<td>Payne et al., 1993 (Study 2)</td>
<td>87.7</td>
<td>93.6</td>
</tr>
<tr>
<td></td>
<td>Wilton/Pennisi, 1994</td>
<td>89.8</td>
<td>88.4</td>
</tr>
<tr>
<td>MMR vaccine</td>
<td>Payne et al., 1993 (Study 3)</td>
<td>—</td>
<td>86.0</td>
</tr>
<tr>
<td>Hib vaccine</td>
<td>Payne et al., 1993 (Study 3)</td>
<td>—</td>
<td>90.2</td>
</tr>
<tr>
<td>Miscellaneous data types</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Historical data</td>
<td>Jelovsek/Hammond, 1978</td>
<td>—</td>
<td>78.7–99.1</td>
</tr>
<tr>
<td></td>
<td>Payne et al., 1993 (Study 3)</td>
<td>—</td>
<td>20.7–40.5</td>
</tr>
<tr>
<td></td>
<td>Payne et al., 1993 (Study 4)</td>
<td>—</td>
<td>1.1–5.6</td>
</tr>
<tr>
<td>Laboratory data</td>
<td>Jelovsek/Hammond, 1978</td>
<td>—</td>
<td>31.2–80.8</td>
</tr>
<tr>
<td>Influenza rates</td>
<td>Johnson et al., 1991</td>
<td>—</td>
<td>28.2</td>
</tr>
<tr>
<td>Endoscopy reports</td>
<td>Gouveia-Oliveira et al., 1991</td>
<td>—</td>
<td>81.3</td>
</tr>
<tr>
<td>Modifiers of pathologic findings</td>
<td>Gouveia-Oliveira et al., 1991</td>
<td>—</td>
<td>81.9</td>
</tr>
<tr>
<td></td>
<td>Kuhn et al., 1991</td>
<td>—</td>
<td>90.7</td>
</tr>
<tr>
<td></td>
<td>Kuhn et al., 1992</td>
<td>—</td>
<td>100</td>
</tr>
</tbody>
</table>
Correctness and Completeness for Various Data Types

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Study</th>
<th>Correctness (%)</th>
<th>Completeness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exam</td>
<td>Pringle et al., 1995 (Study 3)</td>
<td>—</td>
<td>27.2</td>
</tr>
<tr>
<td>Lower extremity exam</td>
<td>Jones et al., 1986</td>
<td>—</td>
<td>97.9–98.7</td>
</tr>
<tr>
<td>Anesthesia record</td>
<td>Edsall et al., 1993</td>
<td>—</td>
<td>87</td>
</tr>
<tr>
<td>Vital signs</td>
<td>Edsall et al., 1993</td>
<td>—</td>
<td>100</td>
</tr>
<tr>
<td>Alcohol use</td>
<td>Pringle et al., 1995 (Study 2)</td>
<td>—</td>
<td>37.5</td>
</tr>
</tbody>
</table>

*Two studies of data accuracy, one before and one after an intervention to improve the accuracy of data.*

*b* This diagnosis had laboratory criteria as the gold standard.

* Rates for two separate hospitals using the same software package.

* Median, the range of completeness for demographic data is 90.3–100%.

* Completeness of peak influenza rate as measured by taking the ratio of the influenza rate derived from CPR data to the influenza rate derived from the gold standard.

* Not specified what criterion were used, such as whether only one or all components of exam had to be present.

Discussion

Quality of the Literature

The quantity and quality of the literature on data accuracy in CPRs did not match our expectations, given the importance of accurate data for its various uses. Of particular concern are those uses such as clinical research and health-system management, where decisions made based on inaccurate CPR data can potentially affect large numbers of patients. Compared with the literature on data accuracy in disease registries and clinical trials databases, the English-language literature on data accuracy in CPRs is not extensive, comprising only 26 studies conducted in 19 distinct CPRs. The quality of the 26 studies was not uniformly high. Only seven studies achieved two-thirds or more of the total methodological and description score possible; one study achieved three-fourths or more of the total possible. The variability in the quality and methods of these studies rendered formal statistical methods of meta-analysis and correlation of results unfeasible, and made even a qualitative synthesis of the literature on data accuracy in CPRs difficult.

We do not know why so few studies of high quality have been reported in the literature. The two most likely possibilities are (1) that CPRs are still relatively new (or nonexistent) in academic settings, and (2) that the problem of data accuracy in CPRs has not received much attention in the field of Medical Informatics, which typically has had a primary interest in the design and development of CPRs. Both of these situations are changing, and as CPR data are used more often for multiple purposes, especially for clinical research and as a data source for disease registries, we expect that the Medical Informatics research community will become more aware of the problem and will apply its expertise to rigorous studies of data accuracy in CPRs.

Accuracy of Data in CPRs

Two factors made analysis and interpretation of the results about rates of accuracy across studies nearly impossible. First, the variability in methods and quality of these studies (as mentioned previously) rendered meta-analysis unfeasible. Second, although researchers reported a rate of completeness for all the types of data that they studied, they reported rates of correctness for significantly fewer than one-half of types of data. As discussed in the introduction, it is difficult to understand the level of data accuracy in a system based on only one measure of accuracy—very high correctness may be achieved at the expense of leaving many observations unrecorded (i.e., a low rate of completeness). The failure to report both measures thus results in an inadequate depiction of data accuracy. All these factors taken together make it difficult to assess whether the accuracy of data in CPRs is poor, fair, good, or excellent.

Despite these limitations, we can form an impression of data accuracy in CPRs by analysis of the few studies of high quality. Our impression, based on examination of rates of accuracy in Table 2 from the 7 studies scoring 12 points or higher, is that data accuracy in CPRs is fair to good. With the exception of a few data types such as specific diagnoses (e.g., anemia in children) and occupational history, the majority of rates of correctness and completeness from these studies are 80% and higher for the types of data studied. A key limitation of the body of literature that we reviewed is that results about rates of data accuracy in
<table>
<thead>
<tr>
<th>Study</th>
<th>Name of System</th>
<th>Time in Use</th>
<th>Data-entry Mechanism</th>
<th>Scope of System</th>
<th>Data Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jelovsek/Hammond, 1978(^1)</td>
<td>Computerized Obstetric Medical Record (COMR)</td>
<td>7 years</td>
<td>Encounter forms and patient questionnaires</td>
<td>University hospital obstetric clinics</td>
<td>Demographics, history, physical, laboratory, medications, diagnoses</td>
</tr>
<tr>
<td>Fortinsky/Gutman, 1981(^2)</td>
<td></td>
<td>2 years</td>
<td>Encounter forms</td>
<td>University hospital family practice clinic</td>
<td>Demographics, diagnoses, diagnostic codes</td>
</tr>
<tr>
<td>Jones/Hedley, 1986(^3)</td>
<td></td>
<td>3 years</td>
<td>Encounter forms</td>
<td>University hospital diabetes clinic</td>
<td>Demographics, history, physical, laboratory, complications(^c)</td>
</tr>
<tr>
<td>Maresh et al., 1986(^4)</td>
<td></td>
<td>2 years</td>
<td>Encounter forms</td>
<td>University hospital dept. of obstetrics</td>
<td>Demographics, history, physical, discharge summaries, laboratory, diagnostic codes</td>
</tr>
<tr>
<td>Dambro/Weiss, 1988(^5)</td>
<td>COSTAR</td>
<td>3 weeks</td>
<td>Encounter forms</td>
<td>University hospital family practice clinics</td>
<td>Demographics, history, physical, laboratory, medications, diagnoses, treatments</td>
</tr>
<tr>
<td>Block/Brennan, 1989(^6)</td>
<td>MediData Medical Information System</td>
<td>4 years</td>
<td>Physician dictation</td>
<td>Urban hospital family practice clinics</td>
<td>Demographics, history, diagnoses, diagnostic codes</td>
</tr>
<tr>
<td>Gouveia-Oliveira et al., 1991(^7)</td>
<td>SISCOPE</td>
<td>1 year</td>
<td>Direct physician entry</td>
<td>University hospital dept. of gastroenterology</td>
<td>Endoscopy reports</td>
</tr>
<tr>
<td>Johnson et al., 1991(^8)</td>
<td>AAH Meditel</td>
<td>—</td>
<td>Direct physician entry</td>
<td>Numerous general practices(^d)</td>
<td>Medications, diagnoses</td>
</tr>
<tr>
<td>Kuhn et al., 1991(^9)</td>
<td></td>
<td>6 years</td>
<td>Physician dictation</td>
<td>University hospital dept. of gastroenterology</td>
<td>Endoscopy and abdominal ultrasound reports</td>
</tr>
<tr>
<td>Barrie/Marsh, 1992(^10)</td>
<td>Manchester Orthopaedic Database</td>
<td>1.5 years</td>
<td>Physician dictation</td>
<td>Community hospital dept. of orthopedics</td>
<td>Demographics and orthopedic diagnoses, procedures, and complications of procedures</td>
</tr>
<tr>
<td>Authors, Year</td>
<td>System</td>
<td>Time</td>
<td>Data Entry</td>
<td>Department</td>
<td>Data Included</td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>------</td>
<td>------------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Kuhn et al., 1992&lt;sup&gt;a&lt;/sup&gt;</td>
<td>—</td>
<td>21 weeks</td>
<td>Direct physician entry</td>
<td>University hospital dept. of gastroenterology</td>
<td>Endoscopy and abdominal ultrasound reports</td>
</tr>
<tr>
<td>Edsall et al., 1993&lt;sup&gt;b&lt;/sup&gt;</td>
<td>ARKIVE</td>
<td>—</td>
<td>Automatic capture of vital signs and direct physician entry</td>
<td>University hospital dept. of anesthesiology</td>
<td>All components of anesthetic record</td>
</tr>
<tr>
<td>Payne et al., 1993&lt;sup&gt;c&lt;/sup&gt;</td>
<td>—</td>
<td>8 months</td>
<td>Transcription of paper-based records</td>
<td>Large HMO</td>
<td>Demographics, immunization records</td>
</tr>
<tr>
<td>Ricketts et al., 1993&lt;sup&gt;29&lt;/sup&gt;</td>
<td>Manchester Orthopaedic Database</td>
<td>1 year</td>
<td>Physician dictation</td>
<td>Community hospital dept. of orthopedics</td>
<td>Demographics and orthopedic diagnoses, procedures, and complications of procedures</td>
</tr>
<tr>
<td>Barlow et al., 1994&lt;sup&gt;28&lt;/sup&gt;</td>
<td>Basingstoke Orthopaedic Database</td>
<td>2.5 years</td>
<td>Direct physician entry</td>
<td>Community hospital dept. of orthopedics</td>
<td>Demographics and orthopedic diagnoses, procedures, and complications of procedures</td>
</tr>
<tr>
<td>Hohnloser et al., 1994&lt;sup&gt;29&lt;/sup&gt;</td>
<td>—</td>
<td>1.5 years</td>
<td>Entry of data by laboratory staff</td>
<td>University hospital dept. of pathology</td>
<td>Hematology biopsy reports, diagnostic codes</td>
</tr>
<tr>
<td>Wilton/Pennisi, 1994&lt;sup&gt;3&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>Transcription of paper-based records</td>
<td>University hospital pediatric clinics</td>
<td>Demographics, immunization records</td>
</tr>
<tr>
<td>Pringle et al., 1995&lt;sup&gt;30&lt;/sup&gt;</td>
<td>EMIS</td>
<td>—</td>
<td>—</td>
<td>Four general practices&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Demographics, history, physical, laboratory, medications, diagnoses, treatments, referrals</td>
</tr>
<tr>
<td>Yarnall et al., 1995&lt;sup&gt;31&lt;/sup&gt;</td>
<td>The Medical Record (TMR)</td>
<td>10 years</td>
<td>Encounter forms</td>
<td>University hospital family practice clinic</td>
<td>Demographics, laboratory, medications, diagnoses, diagnostic codes, x-ray reports</td>
</tr>
<tr>
<td>Wagner/Hogan, 1996&lt;sup&gt;32&lt;/sup&gt;</td>
<td>BGC EMR</td>
<td>1.5 years</td>
<td>Encounter forms and direct clinician entry</td>
<td>University hospital geriatrics clinic</td>
<td>Demographics, history, physical, laboratory, medications, diagnoses, referrals</td>
</tr>
</tbody>
</table>

<sup>a</sup>Refers to how long the system had been in use when the data in question was entered and is approximate.

<sup>b</sup>Empty blocks signify that the information was not available from the article.

<sup>c</sup>As pertain to diabetes and its complications.

<sup>d</sup>A total of 433 general practices linked to a mainframe were included. The system was commercial, and it is not clear whether the practices share data.

<sup>e</sup>The authors selected practices that had a history of high rates of recording patient data.
CPRs may not be representative of what we would find in modern CPRs, which tend to be more extensive in scope and data content. The 19 CPRs that these studies evaluated consist largely of single-clinic or single-hospital systems that serve specialized purposes (e.g., endoscopy reports). Additionally, these studies did not include many prominent research CPRs. Of the ten CPRs considered to be today’s state-of-the-art systems by the 1991 IOM report, only two appear in our review (COSTAR and The Medical Record). We found no published studies of data accuracy about other CPRs cited by the IOM such as the Health Evaluation through Logical Processing (HELP) system at LDS Hospital, the CPRs at Beth Israel and Brigham and Women’s Hospital, the THERESA system at Grady Memorial Hospital, and the Department of Defense’s Composite Health Care System. Studies conducted in mature, comprehensive implementations of the CPR such as these CPRs would be informative. Such studies may still underestimate the potential for CPRs to provide users with accurate data. The IOM report stresses that even these model systems are not complete CPRs. The IOM envisions future systems with seamless integration of data across hospitals, clinics, pharmacies, nursing homes, and so on, with a record of the patient’s health status and functional level; problem lists; documentation of the rationale for clinical decisions; links to local and remote knowledge, literature, and administrative systems; and provision of reminders and decision analysis tools to clinicians.

Causes of Inaccurate Data

Too few causes of error in CPR data have been identified and studied. Researchers have directed most of their attention to data entry; four studies measured the amount of error resulting from the data entry process. Contrary to conventional wisdom, three of these four studies suggest that data entry is a relatively minor cause of error and that other factors such as the scope of the CPR play a larger role. Only one study developed a classification system for categorizing inaccuracy in a CPR and discussed how several classes of errors might be addressed by CPR improvements. This method identified CPR improvements that addressed several categories of causes of error and, if implemented, had the potential to improve data accuracy.

Investigating the causes of data error in a CPR is a prerequisite for the reduction of error. Because the process of data capture is a complex system (Fig. 1), techniques from the field of continuous quality improvement (CQI) seem ideal for studying causes of error in CPR data and correcting those causes by improvements in the mechanisms of data capture in the CPR. From Figure 1, it is apparent that errors may be introduced at multiple points in the process of data capture; thus, it may be insufficient to implement only a single intervention to improve accuracy. Moreover, even after successful interventions, accuracy may not be maintained over time. Medical processes are complex and ever changing, and turnover of personnel may result in unexpected changes in procedures that result in data error. These observations suggest the need for a cycle of regular monitoring, analysis of errors, and interventions designed to improve accuracy, analogous to techniques in CQI.

Although this level of monitoring, to our knowledge, is not routine practice in CPRs, it is typical in clinical research databases because of the importance of accurate data. For example, by studying the accuracy of data entry and the causes of inaccuracies, Horbar and associates were able to implement additional routines for automatic logic, range, and consistency checking that reduced inaccuracies caused by data entry by threefold. Although some of the procedures used in research databases, such as duplicate data entry, may not be practical in CPRs, the process of continuously monitoring accuracy and taking steps to improve it seem relevant, especially if CPR data are to be used for research.

Influence of CPR Characteristics on Accuracy

Several studies reported findings that associate CPR characteristics with data accuracy. As mentioned previously, three studies suggested that improvement of the scope of the CPR to include other clinics, departments, and even hospitals may improve the completeness of data capture. Two studies present conflicting data about whether direct clinician entry improves data accuracy. Wagner and Hogan found no effect on accuracy, but their study lacked the statistical power to detect a small difference. Kuhn et al. found an improvement in data accuracy with direct physician entry over physician dictation of free text reports, but the results were confounded by the checklist effect and the use of a historical control.

Intuitively, direct physician entry should improve the accuracy of CPR data. The IOM, in their report on the CPR, advocate direct data entry by clinicians at the point of clinical care as a mechanism to reduce errors. However, the literature we review here does not provide convincing evidence that direct physician entry of data is warranted solely for the purpose of reducing inaccuracy. (There are other reasons for implementing direct clinician entry, such as providing real-time decision support during order entry.) Confirmation of the hypothesis that direct clinician entry improves data accuracy must await future studies.
Improvement of Data Accuracy

Seven of the 26 studies investigated interventions that may improve data accuracy. Besides factors already discussed (expanding the scope of the CPR and direct clinician entry), other interventions included structured data capture,17,25,31 automatic capture of data from electronic patient monitoring systems,26 monitoring of data accuracy with feedback to personnel involved in data entry,20,28 and providing clinicians with access to a CPR when outside of the clinic or hospital.32 Whether these interventions will improve data accuracy consistently in a variety of CPRs, however, requires further study.

Accuracy of CPRs versus Paper-based Patient Records

No study made a direct comparison of data accuracy in a CPR with data accuracy in a paper-based record. In theory, CPRs should attain higher levels of data accuracy than paper-based records. CPRs can employ validity checks during data entry; allow continual improvement of data by editing rather than rewriting (or dictating); and use standards for transmission of medical data to consolidate observations from disparate locations into a single logical record.

The literature on data accuracy in traditional paper-based records is summarized in the Institute of Medicine’s report on the CPR.15 These studies typically investigated the accuracy of diagnoses, and used, as a gold standard, a record of the actual clinician—patient encounter (e.g., a tape-recording of the encounter, a transcript of a tape-recording of the encounter, and a consensus of observers who viewed the encounter through a one-way mirror). In contrast, the gold standards used in studies of the accuracy of diagnoses in CPRs were largely paper-based records. Because of this major methodological difference, we are unable to compare data accuracy in CPRs with that in paper-based records. Therefore, we cannot determine whether the majority of CPRs contain data that are less, more, or as accurate as those in paper records. Further studies would be necessary to test the hypothesis that data accuracy in CPRs will surpass that of paper-based records; however, such studies are likely to be neither the driving force behind the move toward CPRs (market and other forces are already leading to widespread use of the CPR), nor completed before these other forces lead to nearly universal adoption of CPRs.

Methodological Recommendations

A standard method for future studies of data accuracy in CPRs is needed for two reasons. First, by using more rigorous methods, researchers can improve the quality of the literature on data accuracy in CPRs so that the questions posed, and largely unanswered, by this review may be resolved. Second, increased uniformity of methods should greatly assist with future syntheses of the literature and might allow researchers to apply statistical methods of meta-analysis.

We base the following recommendations on our reading of the literature on data accuracy in clinical trials databases and registries and on our perception of the shortcomings of these 20 articles about data accuracy in CPRs. Researchers should (1) report numerical measures of both correctness and completeness, (2) use an unbiased sampling technique to select patient records for inclusion in the study, (3) select a gold standard with the intention of approximating the true state of the patient as closely as possible, and (4) blind the members of the research team who are responsible for the determination of the gold standard to both the purpose of the study and the CPR data when appropriate. Ideally, studies should provide a thorough description of the CPR, including its name, hardware components, and software versions (especially if the CPR is commercially or otherwise available for implementation at other sites), what types of data it contains, how long it has been in place, its scope, and a description of its methods for data capture. These characteristics of the CPR may potentially influence data accuracy and seem germane to the interpretation of results about data accuracy. Whether CPR data are the official patient record is likely to influence the care and accuracy with which clinicians record data. Accessibility of the CPR should be described—whether data can be entered remotely, from which locations, at what times, and how often the system is “down” may influence accuracy because these factors can limit opportunities to record data. The gold standard should be determined with the goal of approximating the true state of the patient (or world).

Although it is not a methodological recommendation, we propose adding a MeSH term for the concept of data accuracy to MEDLINE. Data accuracy has become an important concept with the widespread use of computer-based systems for clinical and epidemiological research.11 CPR data collected at the point of care are being used increasingly for these purposes as well as being used for making administrative and policy decisions. As a result of these developments, data accuracy will become an even more important topic. The creation of a MeSH term for data accuracy and the indexing of future studies under this MeSH term would facilitate future research and the dissemination of its results.
Conclusions

Data collected by CPRs are used throughout the health care system. The accuracy of this data is critical to the optimal outcome of many health care activities. This review shows that our understanding of data accuracy in CPRs is not commensurate with its importance. It is imperative that we both measure and characterize the accuracy of data in CPRs and investigate ways to improve it. Moreover, as data in CPRs are used increasingly for research, the methods of continuous monitoring and improvement of accuracy used in research databases should be applied to CPRs. Achievement of these objectives will be facilitated by the use of rigorous, more uniform methods to measure accuracy and the incorporation of a MeSH term for data accuracy in MEDLINE to improve dissemination of information about data accuracy in CPRs.

Jeffrey C. Whittle, MD, Department of Medicine, Oakland Veterans Affairs Medical Center; and Charles P. Friedman, Director, Center for Biomedical Informatics, University of Pittsburgh Medical Center provided invaluable discussion.

References


practices committed to recording data on computer. Br J Gen Pract. 1995;45:537–41.