Ensuring high accuracy of data abstracted from patient charts: the use of a standardized medical record as a training tool

Larry Pan, Dean Fergusson, Irwin Schweitzer, Paul C. Hebert

**Abstract**

**Background and Objective:** The quality and integrity of information is pivotal to the validity and reliability of inferences drawn in research. The aim of this study is to demonstrate that standardized medical records can be used as a data abstraction training tool and a quality control measure to assess the validity of medical record data abstraction.

**Methods:** Sixteen hospitals participating in a large multicenter study completed standardized data abstraction forms for three representative patient charts, one in each of the clinical areas of postoperative critical care and trauma, cardiac surgery, and repair of hip fracture. The completed forms were then compared to an established gold standard.

**Results:** The mean level of accuracy of the completed data abstraction forms in each of the above three clinical areas were 91.8, 77.5, and 91.5%, respectively. Missing data accounted for 19% of all discrepancies between the abstracted information and the gold standard. If queries and amendments were made by the study’s coordinating center, the mean level of accuracy increased to 94.5, 82.5, and 92.9%, respectively.

**Conclusion:** The present study stressed the need for quality control measures in abstracting information from medical records to ensure the accuracy and completeness of the data abstracted.

**Keywords:** Validity; Chart review; Data abstraction; Quality assurance; Quality control; Accuracy

1. **Introduction**

   The manual abstraction of data from patient medical records is a routine and efficient method of data collection for clinical databases, audits, and clinical research [1–3]. Obtaining data from charts for research offers many advantages such as accessibility, depth of information, costs, and flexibility in the time the study is conducted [3,4].

   However, there is increasing concern about the reliability and validity of data abstracted from medical records [5–9]. Medical records may contain errors introduced by medical personnel, patients, and laboratory instruments [10]. Physicians may fail to record the chief complaint or err in interpreting the physical examination findings or test results; patients may provide inaccurate histories; or laboratories may furnish erroneous results [10–14]. Furthermore, there is great variability in the amount and quality of information recorded by different practitioners [15,16]. High-quality clinical research requires a number of conditions to be met. None of which is more important than the collection of accurate information upon which to base analysis and interpretation [3]. Although study design and data analysis are covered extensively, both in clinical research textbooks and in the literature, relatively little emphasis is placed on ensuring data accuracy. The same holds for clinical care evaluation using databases or audits of medical practice using health records abstraction.

   A number of approaches may be used in ensuring the accuracy of data abstraction from source documents. The most labor intensive and expensive approach is a detailed audit of a patient’s case report form against source documents. This approach is extensively used by pharmaceutical companies, and is often mandated by regulators. This approach has the advantage of ensuring that each case report form has been exhaustively reviewed, often more than once. However, the cost of such an approach can easily exceed $1,000 per record. In addition, significant differences in individual reviewers and reviews over time may go unnoticed. Given these limitations, the authors believe that an exercise of abstracting information from standardized medical records, which offers the ability to compare the data abstracted against a gold standard reference, is a useful
teaching aid. Moreover, it is much less labor intensive and costly than detailed audits of source records.

Despite the crucial role of accurate data abstraction in clinical research, relatively little attention has been devoted to this area. In this study, we report our experience in using standardized clinical records as a data abstraction training tool and quality control measure.

2. Methods

2.1. Selection of study participants

Data abstractors from 16 academic hospitals from major centers across Canada participated in this evaluation. We made use of a large multicenter evaluation of a program of Universal Leukoreduction of red blood cells involving more than 15,000 medical records in three clinical care foci (postoperative critical care and trauma, cardiac surgery, and repair of hip fractures) from 23 centers [17]. Data abstraction staff and investigators from 16 sites were asked to participate in this learning and quality control exercise. All medical records were stripped of personal and professional identifiers prior to the circulation of any records. All chart reviewers were assigned an identification number to ensure anonymity. This evaluation was part of a larger research program approved by the Research Ethics Committees of participating sites.

2.2. Preparation of standardized medical records

The Universal Prestorage Leukoreduction of red blood cells (RBCs) was evaluated in a large before and after study involving more than 15,000 patient records from centers across Canada. This study examined records in three clinical areas: (1) postoperative and trauma critical care, (2) cardiac surgical procedures requiring extracorporeal bypass, and (3) intraoperative repair of hip fractures. All consecutive patients who received an RBC transfusion were included in the study. To assess the validity of chart data abstraction in each of these three clinical areas, one representative sample patient chart was chosen from each clinical area to be the standardized medical record from which data was abstracted. All individual identifiers were then removed from the charts. Photocopies were then made and distributed to all participants. Data abstractors abstracted the required data from the standardized records onto a standardized case report form.

To develop a gold standard case report form for comparison, five skilled and trained data reviewers from the Leukoreduction Study Coordinating Centre independently abstracted data from the standardized charts onto a standardized case report form for each of the three representative patient charts. From the data gathered, a “gold standard” case report form for each of the three standardized medical records was established by a facilitated face-to-face consensus among the five reviewers and a facilitator. The case report forms were then compared and any discrepancies were resolved by reexamining the original chart. The information to be abstracted from the standardized medical records onto the standardized case report forms consisted of objective items (Table 1). Because the items required were objective and did not require clinical judgment, the interrater reliability of the five reviewers in the determination of the gold standard was not assessed. The emphasis was to establish a gold standard through unanimity among the five reviewers.

2.3. Data collection

The standardized medical records and three blank case report forms were sent to all participating research sites to be completed by all data abstractors at each site. Once completed, the case report forms were returned and the information in each case report form was transcribed onto an electronic file. Critical data to be reviewed and analyzed were chosen, a priori, by the authors based on relevance to the Leukoreduction study. The overall total number of data

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical data from case report forms to be analyzed</td>
</tr>
<tr>
<td>Section</td>
</tr>
<tr>
<td>Demographics</td>
</tr>
<tr>
<td>Surgical information</td>
</tr>
<tr>
<td>ICU diagnosis</td>
</tr>
<tr>
<td>Red blood cell administration</td>
</tr>
<tr>
<td>Key hemoglobin concentrations</td>
</tr>
<tr>
<td>Major events during hospitalization</td>
</tr>
<tr>
<td>Outcomes</td>
</tr>
<tr>
<td>Discharge information</td>
</tr>
</tbody>
</table>

Abbreviations: Hb, hemoglobin; ICU, intensive care unit.
items for postoperative critical care and trauma, cardiac surgery, and operative repair of hip fracture were 65, 72, and 66, respectively. The groupings of information included (Table 1): demographics (4 items), surgical information (13 items), intensive care unit (ICU) diagnosis (5 items), RBC administration (4 items), key hemoglobin concentrations (7 items), major events during hospitalization (26 items), outcomes (25 items), and discharge information (2 items). The total number of items common to all three patient charts was 49.

2.4. Analysis

Completed case report forms from each data abstractor for each of the three standardized medical records were compared to the established gold standard case report forms. For data analysis, simple descriptive statistics were used. Data values that were an exact match to the gold standard were considered correct responses, and all other values were considered incorrect. In the quantitative analyses, all items were attributed equal weight. That is, a correct or incorrect answer regardless of item was treated with equal importance. Instead, the case report forms were designed to assign more weight to important data necessary for analysis and interpretation. A further analysis was performed, incorporating all errors on the case report forms and enumerating the errors that would have been detected by the study’s coordinating center during the data review process, during which queries and amendments would have been made. Assuming that all amendments made were correct, the values for the mean overall proportion correct for ICU stay, cardiac surgery, and hip surgery were determined. In the present study, an a priori ideal target accuracy level of 95% was used. The proportion of case report forms that met this target was the primary outcome of this study.

2.5. Workload related to evaluation

A medical student as a summer research project completed all preparatory work. The identification and determination of suitability of each record required approximately 1 hr. Stripping of personal identifier and all photocopying required approximately 7 hr per chart. Each case report form required an average of 1.5 hr per chart. Each case report form required an average of 1.5 hr to complete, whether completed by the abstractors or five members of the review panel. Collating and comparing responses represented 1 week of full-time work. A detailed review and comparison of the chart by the five-member panel required an additional 4 hr. Data entry and analysis represented 2 weeks of full-time work.

3. Results

3.1. Proportion correct by clinical area

Data was collected by 20 abstractors representing 16 academic hospitals. In total, 55 case report forms were received (18 cardiac surgery, 18 postoperative critical care and trauma, and 19 hip fracture). The proportion correct for each of the three clinical areas was: 1,010/1,100 = 91.8% (range 76.9–96.9%, standard deviation 5.6%) for ICU stay, 942/1,216 = 77.5% (range 59.7–87.1%, standard deviation 7.9%) for cardiac surgery, and 1,115/1,218 = 91.5% (range 80.3–100.0%, standard deviation 4.7%) for hip surgery. For the 49 items common to all of the three clinical areas, the proportion correct was 96.4, 93.3, and 94.8% for postoperative critical care and trauma, cardiac surgery, and repair of hip fracture, respectively.

3.2. Proportion correct by section

The mean proportion correct for each of the eight data collection sections was: 95.4% for demographics, 79.9% for surgical information, 55.1% for ICU diagnosis, 92.3% for RBC administration, 88.7% for hemoglobin concentration, 85.6% for major events, 97.4% for outcomes, and 95.4% for discharge information. The data indicate that ICU diagnosis and surgical information sections scored the lowest. All other sections scored above 85% (Fig. 1). The sections on demographics, outcomes, and discharge information scored greater than 95% correct. For one item within the ICU diagnosis section, the APACHE II score, the correct response rate was 8%. The APACHE II score is provided for any patient who stayed in the ICU for 24 hr or more; in the present study, this item was only relevant for the ICU and cardiac standardized medical records. For the ICU standardized patient record, responses for this score ranged between 12 and 23, compared to the gold standard APACHE II score of 18. In the cardiac surgical records, insufficient data was purposely not provided in the standardized medical record to calculate the APACHE II score. Only 11% of the data abstractors indicated “no data”; the remainder provided a range of values from 21 to 38.

3.3. Extra data and missing data

The completed case report forms were also examined for extra data and missing data. Extra data was defined as completing a data field that was not relevant to the patient chart or not required by the case report form (e.g., completing the surgical information data fields for a medical ICU patient, which is unnecessary according the procedures manual provided to the data abstractors), while missing data consisted of failing to provide relevant information by leaving the data field blank on the case report form (e.g., failure to provide patient’s discharge date). Of the 55 completed case report forms received, 31 contained no extra data, 17 provided between one and five pieces of extra data per case report form, and 7 had greater than five pieces of additional information. The majority, 70% (69/99), of the extraneous information on the case report forms consisted of sections of the case report form being completed that were not applicable to the particular patient. However, there were a few cases (eight) where inaccurate “extra data” were provided.
Missing data accounted for 19% (90/467) of all errors. The remainder of the errors consisted of incorrect data values or responses.

3.4. Amendments after review by coordinating center

In the present study, assuming that the amendments made after review were correct, the values for the mean overall proportion correct for postoperative critical care and trauma, cardiac surgery, and operative repair of hip fractures increased to 94.5% (range 90.8–98.5%, S.D. 2.2%), 82.5% (range 72.2–87.5%, S.D. 4.5%), and 92.9% (range 81.8–100.0%, S.D. 3.8%), respectively (Fig. 2). Using a threshold of 95% correct as a target accuracy level, 10 of the 18 postoperative critical care and trauma case report forms, 0 of the 18 cardiac surgery case report forms, and 5 of the 19 hip fracture case report forms met this target value.

4. Discussion

In this study, we documented an overall accuracy in each of three clinical areas of 91.8% for postoperative critical care and trauma, 77.5% for cardiac surgery, and 91.5% for repair of hip fracture. The accuracy rate increased to 96.4, 93.3, and 94.8%, respectively, when the 49 common items were examined. This exercise noted significant variability between sites and by clinical area. As a quality control measure and teaching tool at the initiation of a study, this evaluation allowed us to make critical observations that might have affected the integrity of the data and potential inferences arising from the study. Once noted, remedies were sought and instituted. This quality control exercise alerted us to the clinical areas, the type of data, and the individual data abstractors or sites that consistently yielded lower levels of accuracy. Thus, the present exercise permitted one to
suggest a number of foci that may deserve further investigation to improve the validity of the data abstracted, such as weaknesses in the case report form layout, areas that were subject to significant interpretation, variability in the quality of individual data abstractors, and deficiencies in the procedures manual. This was one of many approaches that ensured the accurate collection of data. However, in addition to a 10% audit of source records, it was the only additional method used to ensure the quality of data abstraction from source records.

When attempting to explore reasons why cardiac surgical procedures seemed to have a greater error rate, we identified a number of potential reasons for this anomaly. We explored ambiguities or a lack of clarity in the case report form and the study procedures manual, the complexity of the medical record, variation in individual reviews, the training and experience of the reviewers, and familiarity with the medical record. In this example, possible explanations for the high error rate may be a lack of familiarity with the area of cardiac surgery because some of the data abstractors do not work in a center with cardiac surgery patients (all records were sent to all centers) and individual variability. However, further investigation is warranted.

The type of data abstracted also affected the accuracy of the abstraction process. Demographic, outcome, and discharge information had the highest accuracy rating. It is evident that all of these items requested information that is readily and easily abstracted from the charts. Demographic and discharge information were clearly summarized on one or two easily identified pages in the medical record. Establishingoutcome required the abstraction of information on survival status, on antibiotic use, and the presence or absence of severe infections (pneumonia, major bacteremia, and severe sepsis/septic shock), which may also be abstracted without much difficulty.

Diagnostic information and information related to the surgical procedure scored well below accepted norms at 55.1 and 79.9%, respectively. It is clear that data in these sections was subject to a great deal of interpretation (primary vs. secondary diagnosis) and more difficult to locate in the record because of insufficient information.

The APACHE II score, a part of the ICU diagnostic section, also presented enormous difficulties for the data abstractors. Given that it represents a summative score from physiologic parameters such as vital signs, oxygenation level, arterial pH, serum biochemistry, age, and chronic health status, there were numerous opportunities to miss important information. Indeed, a number of studies [12,18–20] have demonstrated poor intrarater reliability in the reporting of similar variables. Undoubtedly, investigators should consider using more experienced staff and provide more extensive training in the calculation of APACHE II scores if such information is critical to the study’s findings.

We opted to provide further detail in the procedures manual and more audits of source records at specific sites. We also decided to use broad diagnostic groupings and categorized APACHE II scores into larger groups to minimize the effect of errors.

One of the common errors observed in this evaluation consisted of missing data, which accounted for 19% of all errors. “Missing data” was coded whenever a data field that was required to be completed but remained empty. Potential causes for these inaccuracies may include reasons such as inattention on the part of data abstractors and a lack of familiarity with procedures related to the case report form.

Evaluations against standardized medical records provide a means to ensure that all individuals abstracting information from the medical record do so with comparable rates of accuracy. We recommend that this approach be adopted in larger retrospective reviews as a method to ascertain data abstraction accuracy (at the initiation of, or during the conduct of the trial), as a diagnostic tool to determine where there may be difficulties at the initiation of a project, as a quality control measure in ongoing projects and clinical databases in concert with audits of source records, and as a teaching tool during initiation of projects. This approach may also be used in the training and evaluation of individuals performing site audits to ensure that they perform at acceptable standards.

This form of evaluation is not a substitute for 100% review of all source documents but rather a complimentary approach to less detailed reviews usually undertaken in peer-reviewed studies and as an additional quality control measure. As we outlined, it is not too labor intensive for larger projects, but would create a heavy and expensive burden for small clinical audits or research projects. This process could also be used to establish intrarater reliability, and could have been expanded to include an assessment of intrarater reliability. One of the drawbacks of sending the same representative patient charts to all sites is that some of the sites may not be familiar with the format of these “foreign” charts, as they are not from their own center. One of the ways to minimize this effect is to include records from more than one site so as to get a larger representation of charts. In this evaluation, we included charts from two sites.

In summary, evaluations using a standardized medical record serve as a useful adjunct to other quality control measures and teaching tools used by data abstractors.

References


