Identifying and improving unreliable items in registries through data auditing

CAMERON D. WILLIS1, DAMIEN J. JOLLEY2, JOHN J. MCNEIL2, PETER A. CAMERON2,3, AND LOUISE E. PHILLIPS2

1School of Population Health and Clinical Practice, The University of Adelaide, Adelaide, South Australia, Australia, 2Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia, and 3Emergency and Trauma Centre, The Alfred Hospital, Melbourne, Victoria, Australia

Address reprint requests to: Cameron D. Willis, School of Population Health and Clinical Practice, Mail Drop DX 650 550, The University of Adelaide, SA 5005, Australia. Tel: +61-8-8313-0612; Fax: +61-8-8303-6899; E-mail: cameron.willis@adelaide.edu.au

Accepted for publication 21 February 2011

Abstract

Objective. Assessing the reliability of clinical registries is important for ensuring the availability of credible data. Therefore, this study aimed to investigate the reliability of data collected by the Australian and New Zealand Haemostasis Registry (the registry).

Design. Data from 5% of randomly selected registry cases were re-abstracted by an independent data auditor who was blinded to the results of the original data abstraction. Categorical data were investigated for agreement between original and re-abstracted data. The mean difference and standard deviations (SD) of differences were calculated for continuous variables. We estimated a ‘prediction interval’ as the mean difference + twice the SD of differences. We computed a coefficient of variation as the SD of differences.

Setting. The registry records all cases of off-licence use of recombinant activated factor VII (rFVIIa) at participating institutions (on-licence use of rFVIIa is not recorded).

Results. Data on 76 registry cases (6% of registry) were re-abstracted. Various parameters demonstrated high levels of inter-rater reliability, including age, gender and intensive care unit admission (88, 99 and 99% agreement, respectively). Other variables were highly unreliable, including crystalloid infusion volumes (coefficient of variation 123.01%), red blood cell units (92.05%) and time from bleeding onset to administration of rFVIIa (153.06%).

Conclusions. Registry audits are useful for identifying variables with poor reliability. Repeated audits will not improve data reliability; however, they can assist in identifying and evaluating the impact of modified data collection processes on improving data reliability.

Keywords: registries, data quality, recombinant FVIIa

Introduction

Clinical registries involve the systematic collection of information from patients treated across multiple institutions [1]. They are becoming increasingly important, for quality improvement, safety monitoring of drugs and devices [2, 3] and for clinical research involving specific populations such as cancer [4, 5] or trauma [6]. Data from registries may be used in pay-for-performance schemes and may provide the key information assuring the long-term safety of drugs and devices.

The credibility of registry data requires adequate steps to ensure that they are reliable. With poor data integrity, there may be inappropriate reputational damage and/or financial loss to participating institutions. In contrast to blinded trial data, registry data may be more easily manipulated through sampling (e.g. exclusion of patients with poor outcomes), which can lead to misleading representations of individual centre performance. In addition, human error in the interpretation, selection and recording of data may introduce error into registry data. These factors necessitate an effective programme of quality control which will deter data manipulation, promote careful data abstraction and send a strong message about the importance of accuracy in the data collection process.

Quality assurance requires a focus on data ‘accuracy’ and ‘completeness’ [7, 8]. Completeness relates to the processes that guarantee collection of all cases (avoiding ‘cherry-picking’ cases) and involves a triangulation process with a third source of information. Accuracy aims to ensure that
the data collected correlates with that recorded in the source documentation [7, 8]. Where registries collect data directly from medical records, a blinded note-validation approach where data are re-abstracted and compared with the original data has been reported as the most objective means of assessing data accuracy [9–11].

In an ideal registry, data elements are easily accessible and systematically and objectively recorded using standard definitions and procedures [1]. Such data are highly reliable, so that different observers should record the same information. When these criteria are not met, it may be difficult to separate normal variation from inaccurate collection. Importantly, an item may be highly reliable yet invalid, or valid yet with poor reliability. Poor reliability may suggest an issue with validity and signals a need to review the properties of the variables being collected before considering possible deficiencies on the part of those collecting the data.

There is limited information about how to establish quality control procedures in clinical registries or the optimal ongoing audit procedures in terms of costs and the numbers of cases to be audited. This paper describes a quality control process employed by the Australian and New Zealand Haemostasis Registry (the registry) and aims to: (i) describe the methodology involved in a systematic data accuracy audit, (ii) outline the methods used to quantify the level of agreement between original and audited data and (iii) explore how results from this analysis have been used to improve data collection practices and data quality in this clinical registry.

Methods

The registry collects all cases of off-licence use of recombinant activated factor VII (rFVIIa, NovoSeven, Novo Nordisk Pharmaceuticals Pty Ltd, Denmark) at participating institutions in Australia and New Zealand (on-licence use of rFVIIa is not recorded on the registry). rFVIIa is licensed for use in patients with haemophilia A and B with inhibitors, factor VII deficiency, Glanzmann’s thrombasthenia and acquired haemophilia [12–14]. Due to its localized haemostatic effects, the use of rFVIIa to reduce bleeding in non-licensed settings such as cardiac surgery [15] and trauma [16] has become increasingly common. To monitor the safety of off-licence rFVIIa, the registry was initiated through an unrestricted educational grant from Novo Nordisk Pharmaceuticals Pty Ltd and has been established in the Department of Epidemiology and Preventive Medicine, Monash University, since 2005.

Data abstraction from medical records is performed by local data collectors, typically research nurses, trained data abstractors or medical staff, who receive comprehensive training by registry staff. The accuracy of the data abstracted by data collectors is assessed as a routine part of this training. Data collectors are provided with a data dictionary that outlines standardized definitions of each variable in the case report form (CRF). During data entry, automatic range and validity checks flag implausible values resulting from errors such as incorrect entry of decimal points.

Accuracy audit process

Cases for audit were selected on 15 August 2007, when the registry contained 1345 records. A stratified sampling approach was used. Using random number tables, 5% of the cases were selected from each hospital that had contributed ≥20 cases to the registry. Hospitals that had contributed <20 cases were pooled and 5% of the cases from this group were randomly selected. The size of the registry (and the availability of resources) determined that the maximum practical sample was ∼5%. A trained data abstractor (blinded to the results of the original data collection) re-abstracted the data using the same CRF template as original data collection. Both the original and auditing data abstractor received the same training and carried out the data abstraction process using the same approach. Neither the original nor auditing data collector may be considered a ‘gold standard’.

Data were re-abstracted from 76 patient records, representing 5.7% of registry cases at the time of audit. Eighty continuous, 86 binary/categorical and 18 date/time variables were collected from each case. Each variable was analysed for agreement between original data (N1) and audited data (N2); however, only selected variables are presented in this article (results for all variables are available upon request).

Statistical analysis

The level of agreement between original and re-abstracted binary/categorical data was investigated for the percentage agreement and analysed using κ (and 95% confidence intervals [CIs]). In accordance with the ‘DocDat’ criteria for evaluating database quality [8], κ values <0.5 were considered low, those between 0.5 and 0.8 as fair and values above 0.8 as good. In addition, each pair of categorical variables was investigated for the mean difference in measurements and standard deviations (SD) of those differences. We estimated a ‘prediction interval’ for each continuous variable as the mean difference plus or minus twice the SD of differences. This estimates the likely range of error in any single case in the registry. We computed a coefficient of variation as the SD of differences, expressed as a percentage of the mean of the observed (original and audited) values for the variable. Values above 100% suggest extremely noisy and unreliable data. All analyses were conducted using STATA v. 10.2.

Results

Table 1 presents results for 10 selected continuous variables. These variables are presented in order of the level of agreement between original and audited data. The number of doses demonstrated the highest level of agreement (95%) for all variables collected in this audit, accompanied by a narrow 95% prediction interval. Prothrombin time (PT: a measure of blood clotting time) prior to the first dose was recorded.
by both data collectors in 72% of the cases with a high level of agreement. However, despite this high level of agreement, there was wide variation in PT time likely due to the small number of audited cases that were significantly different from the original data collection.

Although pH has been noted as an important predictor of outcome in patients receiving off-licence rFVIIa [17], the present results demonstrated low levels of agreement between audited and original data, as well as the highest level of missing data for all variables in this audit (38%). However, pH also recorded the lowest coefficient of variation of all continuous variables (Fig. 1), and a narrow 95% prediction interval, suggesting that for any registry case, we can be 95% confident that the difference in pH value between audited and original data lies between $-0.06$ and $0.08$.

Like pH, size of dose 1 demonstrated low levels of agreement between audited and original data; however, variability between data sets was also relatively low, as reflected by the small coefficient of variation (Fig. 1). In contrast to both size of dose 1 and pH, the number of red blood cell (RBC) units administered before the first dose was available for audit in 97% cases. Despite this high level of completeness, results are suggestive of item reliability issues as noted by low levels of agreement, a large coefficient of variation and a wide 95% prediction interval ($-20$, 22 units). Similar results were found for the volume of crystalloid administered (e.g. normal saline) prior to dose 1, with low levels of missing data (7%), yet very poor agreement (32%, CI $[22, 45]$) and extremely wide variability between audited and original data (coefficient of variation 123.01%; Table 1 and Fig. 1). The 95% prediction interval suggests that the mean difference in crystalloid volume between audited and original data lies between $-6099$ and $7882$ ml. The Bland–Altman plot of crystalloid volume in Fig. 2 demonstrates higher agreement for cases with lower volumes of crystalloid and greater variability at higher crystalloid volumes.

This audit also examined three time-linked variables: date of dose 1, bleeding onset time and time to dose 1. Time from bleeding onset to rFVIIa administration demonstrated the poorest agreement and largest variation of all parameters (30% agreement, coefficient of variation 153.06% and 95% prediction interval $[-5731.1, 6829.7]$ min). Calculation of this variable was dependent on the two additional time-linked variables (date/time of bleeding commencement and date/time of dose 1). Although date/time of dose 1 demonstrated good agreement (noting perfect agreement for day of treatment) and a low coefficient of variation (2.6%), agreement between bleeding onset was considerably lower (39%) with wide variation (coefficient of variation 131.28%; Table 1 and Fig. 1).

Binary and categorical variables demonstrated superior agreement compared with continuous variables, with both

### Table 1 Continuous variables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>% agreement</th>
<th>95% CI for agreement</th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>Mean difference</th>
<th>SD of difference</th>
<th>95% prediction interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of doses of rFVIIa</td>
<td>95</td>
<td>87, 99</td>
<td>76</td>
<td>76</td>
<td>76</td>
<td>0</td>
<td>0.04</td>
<td>0.30</td>
<td>−0.5, 0.6</td>
</tr>
<tr>
<td>PT before dose (s)</td>
<td>91</td>
<td>80, 97</td>
<td>59</td>
<td>59</td>
<td>59</td>
<td>8</td>
<td>20.06</td>
<td>172.57</td>
<td>−318.2, 358.3</td>
</tr>
<tr>
<td>Age (years)</td>
<td>88</td>
<td>78, 94</td>
<td>75</td>
<td>76</td>
<td>75</td>
<td>1</td>
<td>0.07</td>
<td>0.79</td>
<td>−1.5, 1.6</td>
</tr>
<tr>
<td>Date of dose 1 (date/time)</td>
<td>73</td>
<td>61, 82</td>
<td>75</td>
<td>74</td>
<td>73</td>
<td>3</td>
<td>9.51</td>
<td>62.18</td>
<td>−112.4, 131.4</td>
</tr>
<tr>
<td>Blood pH prior to dose 1</td>
<td>68</td>
<td>48, 84</td>
<td>49</td>
<td>36</td>
<td>28</td>
<td>29</td>
<td>0.008</td>
<td>0.04</td>
<td>−0.06, 0.08</td>
</tr>
<tr>
<td>Size of dose 1 (µg/kg)</td>
<td>67</td>
<td>54, 79</td>
<td>64</td>
<td>66</td>
<td>58</td>
<td>14</td>
<td>2.36</td>
<td>18.78</td>
<td>−34.4, 39.2</td>
</tr>
<tr>
<td>RBC prior to dose 1 (units)</td>
<td>57</td>
<td>48, 68</td>
<td>75</td>
<td>75</td>
<td>74</td>
<td>2</td>
<td>1.03</td>
<td>10.71</td>
<td>−20, 22</td>
</tr>
<tr>
<td>Bleeding onset (date/time)</td>
<td>39</td>
<td>27, 53</td>
<td>62</td>
<td>65</td>
<td>56</td>
<td>14</td>
<td>475.36</td>
<td>3131.92</td>
<td>−5663.2, 6613.9</td>
</tr>
<tr>
<td>Crystalloid prior to dose 1 (ml)</td>
<td>32</td>
<td>22, 45</td>
<td>72</td>
<td>75</td>
<td>71</td>
<td>5</td>
<td>891.55</td>
<td>3566.42</td>
<td>−6098.6, 7881.7</td>
</tr>
<tr>
<td>Time to dose (min)</td>
<td>30</td>
<td>18, 44</td>
<td>61</td>
<td>63</td>
<td>53</td>
<td>18</td>
<td>549.32</td>
<td>3204.28</td>
<td>−5731.1, 6829.7</td>
</tr>
</tbody>
</table>

N1, number of original data collections including data for this variable. N2, number of audit data collections including data for this variable. N3, number of cases with data for this variable from both original and audit data collections. Analysis is based on these cases. N4, number of cases in which either original or audit data (but not both) are missing for this variable.

PT, prothrombin time; RBC, red blood cells. *Same day agreement 100.00%. **Same day agreement 86.11%.

Figure 1 Coefficient of variation for selected continuous variables.
gender and intensive care unit (ICU) admittance (yes/no) recording 99% agreement and good $\kappa$ values (0.96 and 0.85, respectively) (Table 2). Adverse events were complete for all cases and demonstrated fair levels of agreement (78%); however, a low $\kappa$ value was present (0.54).

Table 2 Binary/categorical variables

<table>
<thead>
<tr>
<th>Parameter</th>
<th>% agreement [95% CI]</th>
<th>$\kappa$ [95% CI]</th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>Symmetry ($P$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>99 [93, 100]</td>
<td>0.96 [0.90, 1.00]</td>
<td>76</td>
<td>76</td>
<td>76</td>
<td>0</td>
<td>0.317</td>
</tr>
<tr>
<td>ICU admitted</td>
<td>99 [93, 100]</td>
<td>0.85 [0.56, 1.00]</td>
<td>76</td>
<td>76</td>
<td>76</td>
<td>0</td>
<td>0.317</td>
</tr>
<tr>
<td>Adverse events</td>
<td>78 [67, 86]</td>
<td>0.54 [0.35, 0.73]</td>
<td>76</td>
<td>76</td>
<td>76</td>
<td>0</td>
<td>0.808</td>
</tr>
</tbody>
</table>

N1, number of original data collections including data for this variable. N2, number of audit data collections including data for this variable. N3, number of cases with data for this variable from both original and audit data collections. Analysis is based on these cases. N4, number of cases in which either original or audit data (but not both) are missing for this variable.

Discussion

Without accurate and reliable data, registry analyses may generate misleading findings, leading to poor clinical/policy decisions. Registry audits provide valuable information for determining which variables are accurate and reliable, which variables are clinically important but poorly collected and which variables are unreliable with little opportunity for improvement in data accuracy.

As expected, a number of variables were highly reproducible, such as age, gender and ICU admission. These variables are routinely documented using standardized conventions in consistent sections of the medical record. As a result, their collection by trained data abstractors is relatively straightforward, resulting in highly reliable data items. Other variables, such as the number of doses of rFVIIa, were not documented in a standardized manner; however, they are critically important for collection in this context. Despite non-standardized documentation, the number of rFVIIa doses was a reasonably reliable data element, with a coefficient of variation of 18.49%. These data support results from other registries, which have found high reliability for problematic variables including tissue ischaemia times in liver transplant patients [18]. Data collector training and automated range and validity checks likely contribute to the accuracy of these data.

This study has identified a number of variables that are clinically important yet are recorded with poor levels of reproducibility. As noted, previous publications from the registry have documented outcome predictors in patients...
receiving off-licence rFVIIa, including RBC transfusions [17, 19]. Despite the importance of this variable, the accuracy of RBC transfusions was low, with results suggesting that for a given case, the difference between audited and original RBC data may vary between −20 and 22 units of blood (Table 1). The lack of systematic data collection and documentation, or a component of carelessness, may have contributed to a high coefficient of variation (92.05%) and, therefore, an unreliable data item.

Transfusion histories are often scattered through laboratory reports, transfusion forms or nursing, anaesthetic and operation notes [20]. When relevant information is located, inaccuracies are common [20]. However, blood transfusion details are central for providing risk-adjusted outcomes in the registry as well as in trauma and cardiac surgery registries. Therefore, identifying methods to improve the reliability of these data is important. To do so, the manner of data collection can be systematized, data collectors can receive additional training or recording aids such as nation-wide blood product ordering forms may be introduced. These methods have been classified as ‘quality assurance’ (activities prior to data collection) or ‘quality control’ (efforts during and after data collection) [21]. As a quality assurance initiative, the registry is instituting a policy to standardize the collection of blood product usage through data extraction from blood bank laboratory information systems, which have been found to be more reliable sources of the time of use of blood components (unpublished data).

Like RBC transfusion histories, pH is a clinically important variable for patients receiving rFVIIa, yet as identified in this audit, often missing or inconsistently documented in patient records. Missing data in registries are common, such as missing brain imaging times in 71% of subjects in stroke registries [22, 23]. Unlike prospective data collections, registries cannot exclude a patient’s entire record based on a single missing data point. Quality control initiatives such as data imputation are available; however, different results have been reported based on the imputation method employed [24]. Quality assurance procedures to improve data accessibility prior to data collection remain the ideal method of minimizing missing data. Despite its importance, only 37% of cases had a recorded value in both audited and original data. However, when pH is documented, it appears to be highly reproducible (coefficient of variation of 0.48). Although alternative data sources such as electronic laboratory reports are being explored, other quality assurance initiatives have been introduced, including in-note stickers. Since implementing this initiative, the number of missing pH values has fallen by 18%.

Similarly, time of bleeding commencement is not a standardized parameter in medical records, leading to highly unreliable estimates as well as negative influences on other related variables such as time to rFVIIa dose. Others have reported poor reproducibility with similar time-based measures such as stroke onset time in the Paul Coverdell National Acute Stroke Registry [22]. The importance of time-based parameters in critically bleeding patients has led to exploration of alternative information sources for improving the reliability of these items, such as time of surgery (in surgical cases) or time of admission (in trauma cases). The validity of these parameters as markers of bleeding commencement requires further investigation.

This analysis has identified a number of variables that are not systematically recorded, have extremely high coefficient of variation values and may have limited opportunity for improvement, such as crystalloid infusions prior to dose 1. Differences in crystalloid volume between audited and original data for a given registry case maybe between −6000 and 7800 ml (Table 1). Like blood transfusion data, information relating to fluid therapy is not documented in a standardized manner; therefore, generating an accurate and reliable value is problematic. As specific data collector training has not improved the reliability of this parameter and systematic approaches to crystalloid data collection are unavailable, recommendations have been made to remove this variable from future CRFs.

The present results demonstrate the difficulties in interpreting statistical analyses for accuracy audits of registry data. The low agreement (67%) between audited and original dose size data (measured in μg/kg) may be attributable to differences in dose volume (μg) or weight (kg). Registries that analyse such ‘combination’ variables need to investigate the accuracy of contributory variables prior to modifying quality assurance or control practices. Similarly, PT before dose 1 was found to have high agreement (91%), but with significant variability (95% prediction interval −318.2, 358.3 s), likely linked to a small number of highly discrepant results obtained from different laboratory reports. As coagulation parameters in this population are highly volatile, laboratory results from different time points maybe highly divergent. As a result, automated extraction systems from electronic pathology records are being investigated.

There are few publications on the determination of optimal sample size for reliability studies [25–27]. Most focus on the precision of reliability measures such as ‘percent agreement’ for categorical variables and the intra-class correlation coefficient for continuous variables. ‘Precision’ is usually specified by the width of the CIs around point estimates of these measures. Precision depends on the number of disagreements found in the audit and the number of sampled cases, and very little on the size of the registry itself. Some studies [28, 29] have also considered the costs associated with sampling. Costs involved in a registry audit usually comprise a fixed data collection overhead, plus a cost proportional to the sample size. Ideally, registry audits would be an ongoing process, with costs built into general registry running costs. In some cases, a once-off audit of quality and reliability is required, where the costs of the audit may have over-riding priority over consideration of the precision of reliability estimates.

In the current audit, the size of the registry at the time of audit and the available resources demonstrated that the maximum practical sample was ~5%. The sample size can and should be chosen independently of the size of the registry. The adequacy of sample size in any survey becomes evident after the data have been collected and estimates of
reliability and their CIs calculated. CIs which are wide and fail to exclude clinically relevant alternatives for the target parameters are the hallmark of inadequate sample size. For this reason, reporting CIs for all registry reliability parameters is critical.

κ is routinely employed to measure agreement between repeated data collections; however, it can be problematic. In particular, κ is dependent on the prevalence of the characteristic in the population being studied, as low numbers or imbalances in categorical variables will increase the risk of chance agreements and therefore decrease the κ value [30]. This situation is evident for adverse event data in the registry, where agreement was relatively high (77.6%); however, the κ value was 0.54. Although an accepted κ alternative is not available, future registry audits need to be mindful of the statistical limitations of these approaches.

A comprehensive assessment of registry quality needs to consider data accuracy and completeness. Preliminary estimates suggest that the registry includes ~83% of patients receiving off-licence rFVIIa in Australia and New Zealand. A volume audit is currently being conducted that compares the volume of rFVIIa recorded on the registry against the volume of the agent purchased by each participating hospital.

This study had a number of limitations. The audit process did not seek to confirm contributory factors for low data accuracy for each parameter, nor did it aim to identify facilities/data collectors with specific data accuracy problems. However, the sampling strategy employed enables stratification by facility, allowing exploration of facility-dependent data quality issues. Moreover, this audit did not define a gold standard data collection method against which the sensitivity or specificity of original data could be compared. As no single gold standard could be specified for the collection of the wide variety of parameters in this audit, we employed a blinded, repeat data collection using two data collectors and a standardized data collection approach (noted to be the most appropriate method for audit in data collected from medical records [9]).

The utility of registries as tools to inform clinical decisions, policy directions, health service delivery and patient choice is reliant upon accurate and reliable data. Although good registry governance should include a robust quality assurance framework and a continuing focus on data quality [31], few registries have published details on the audit methodology used to ensure data standards. This article has described an accuracy audit methodology and reported the strengths and deficiencies associated with repeatedly collected parameters in a bi-national clinical registry. Many parameters were found to be accurately recorded; however, a select number was noted to be highly variable. Other registries may benefit from exploring the audit methods adopted in this analysis, and refining data collection methods to improve the reliability of key registry items (such as time-related intervention details) and maximize the use of available resources (including electronic recording systems). Although registry audits alone are not sufficient to improve the reliability of these variables, data quality audits are useful for identifying unreliable data items and evaluating strategies to modify and/or supplement existing data collection practices that may assist in improving data quality.

Acknowledgements

The authors would like to acknowledge the assistance of Ms Rosalie Clementson and all local data collectors. This manuscript is written on behalf of the Haemostasis Registry Steering Committee: Professor Peter Cameron, Dr Scott Dunkley, Professor James Isbister (Chair), Professor John McNeil, Dr Louise Phillips, Professor Hatem Salem, Ms Wendy Thomas and Dr Amanda Zatta.

Funding

This work was supported by an unrestricted educational grant from Novo Nordisk Pharmaceuticals Pty Ltd.

References


Wendy Thomas and Dr Amanda Zatta.


