JECT. 2011;43:P44–P51
The Journal of ExtraCorporeal Technology

Perfusion Downunder Collaboration Database—Data Quality Assurance: Towards a High Quality Clinical Database

Sigrid C. Tuble, PhD, CCRC

Department of Cardiac and Thoracic Surgery, Flinders Medical Centre, Adelaide, Australia

Presented at Perfusion Downunder, Queenstown, New Zealand, August 5–8, 2010.

Abstract: Maintaining a high quality clinical database is critical to obtain reliable information upon which to base clinical and institutional decisions, and to preserve the public and the user's confidence in the quality of the data. The success of the Perfusion Downunder Collaboration (PDUC) Database, a dataset for cardiopulmonary bypass procedures, can only be guaranteed through the assurance of the quality of its data. This paper presents the evaluation of the data quality in the PDUC Database. Three participating centers located in Adelaide, Australia were audited: Flinders Private Hospital (FPH), Flinders Medical Center (FMC), and Ashford Hospital (AH). Ten percent of the cases submitted from the first year of data harvest were audited (2008: FPH and FMC, 2009: AH). A total of 57 variables were reviewed and rates of discrepancies (*inaccurate*, *missing*, *not entered*, *cannot be validated*) categorized as 0–25%, 25–50%, 51–75%, and 75–100% of cases (% = cases with discrepancy/total cases audited) evaluated. Sixty randomly selected cases were audited, comprising of 13 cases from FPH, 31 cases from FMC, and 16 cases from AH. Of a total of 3420 data points evaluated, 6.9% were found to be *inaccurate* and 3.2% were *missing*. For each participating center, the great majority of variables have discrepancies in few (0–25%) of the cases audited. The discrepancies found can be attributed to systematic errors (e.g., error in date difference calculation for length of stays, data transformation error for postoperative dialysis) and random errors (e.g., use of incorrect unit for creatinine, transcription error for discharge date). The PDUC Database is currently reasonably accurate and complete. This evaluation is part of a complex system of data quality assurance, and when conducted routinely, could provide a continuous feedback loop towards a high quality PDUC Database. **Keywords:** cardiopulmonary bypass, quality control, audit. *JECT. 2011;43:P44–P51*

"Data quality assurance… the whole of planned and systematic procedures that take place before, during, and after data collection, to guarantee the quality of data in a database." (1)

INTRODUCTION

The benefits of clinical databases are well recognized and over the past decade, numerous databases and registries have been established in institutions across different regions, and across the different fields of medicine. In cardiac surgery, clinical databases have been used to assess the quality of clinical performance (2,3), to determine patients' risk profiles and outcomes (4,5), and to inform the management and planning of clinical services (6). Clinical databases have the benefit of being cost-effective, having the ability to provide large amounts of information that is timely and has generalizability, and promote research activities (7). It is therefore critical to be able to maintain high quality clinical databases to obtain reliable information upon which to base clinical and hospital management decisions, and to be able to preserve the public and the user's confidence in the clinical database.

Address correspondence to: Sigrid C. Tuble, Cardiac Surgery Research and Perfusion, Level 6 Flinders Private, Flinders Medical Centre, Bedford Park, SA 5042, Australia. E-mail: Sigrid.Tuble@fmc.sa.gov.au

Data Quality

Data quality has been defined as "the totality of features and characteristics of a data set that bear on its ability to satisfy the needs that result from the intended use of the data" (1). It has been found to have two key features: accuracy or "the extent to which registered data are in conformity to the truth" and completeness or "the extent to which all necessary data that could have been registered have actually been registered". Errors in data (i.e., inaccuracies and incompleteness) have been classified as systematic (type I) and random (type II) errors (8). Causes of systematic data errors include programming errors (e.g., data transfer, calculation errors), unclear database interface design, unclear and ambiguous data definitions, and violation or even unclear data collection protocols.

Random data errors have been commonly attributed to transcription errors, lack of adherence to data definitions, and data source issues (illegible handwriting or incompleteness) (1).

Perfusion Downunder Collaboration Database

The Perfusion Downunder Collaboration (PDUC) Database (9) has been established for the purpose of collecting a dataset for cardiopulmonary bypass (CPB) procedures. The PDUC Database will assist in the aims of the PDUC, a regional network of perfusionists and researchers committed to the pursuit of excellence in perfusion and to the development of relationships with individuals, groups, and organizations to enhance their collective knowledge, capabilities, and capacity to foster research. The PDUC endeavors to encourage the development of new researchers as well as to facilitate the development of research by all members of the Collaboration. The achievement of these aims can be guaranteed by the success of the PDUC Database through the assurance of the quality of its data. Currently, the quality of the data in the PDUC Database is maintained though quality assurance and control procedures, such as the training of data managers, data validation (e.g., range checks to ensure that data lie within a specified range of values), and the use of PDUC Database Data Quality Reports which allows the review of data that is missing or have been indicated as *not entered*. However, the evaluation of the accuracy and completeness of the PDUC Database is needed to assist in the determination of additional quality assurance and control procedures.

This paper presents the evaluation of the data quality in the PDUC Database, from the development of the audit guidelines, its implementation, and the outcomes. The evaluation of data quality, through the assessment of the accuracy and completeness of the data from the different centers participating in the PDUC Database will assist in the identification of discrepancies in the database system, from data collection at the local centers to the transfer of data to the central database. This will facilitate the development and implementation of relevant quality assurance procedures that will provide continuous data quality improvement towards a high quality PDUC Database.

METHODS

Development of the Audit Guidelines

The guidelines of the audit were initially developed, fully considering the aims of the audit that have been identified above. Organizers of the PDUC Database were interviewed to gain insight and understanding of the data collection process. Audits were conducted in three of the current seven centers participating in the PDUC Database project. The centers were Flinders Private Hospital (FPH), Flinders Medical Center (FMC), and Ashford Hospital (AH), located in Adelaide, Australia. Ten percent of the cases submitted by each participating center from the first year of data harvest were audited, comprising of the 2008 cases for FPH and FMC, and the 2009 cases for AH. The list of cases and corresponding data to be audited for each site was generated using a query embedded in the PDUC Database front-end. This query is coded to randomly select 10% of the total number of cases for the date range indicated for the center to be audited. Cases were included in the audit if the procedures were performed with cardiopulmonary bypass including: isolated coronary artery bypass grafting, isolated valve, and concomitant valve/graft surgery.

The variables to be audited from the PDUC Database dataset were identified, taking into account the current data needs and focus of research of the members of the Collaboration and its partners. Of current particular interest are those variables relevant to the determination of cardiac surgery risk models such as the AusScore (10), EuroScore (11), and the Society of Thoracic Surgeons' 30-day operative mortality and morbidity risk models (12). In addition, relevant variables which have been found in the October 2009 PDUC Database Data Quality Reports to have >20% of data missing or have been indicated as *not entered* have also been reviewed. A total of 57 variables (Table 1), out of the 260 variables that currently comprise the PDUC Database dataset, were reviewed. This includes five demographic, 15 clinical, four procedure, nine perfusion, and 24 outcomes variables.

A criterion was defined for each variable to be able to evaluate the accuracy of data, similar to the methodology used by Arts et al. (1), which was peer-reviewed and based on clinical relevance and the PDUC Database data definitions. Any disagreement between the data in the PDUC Database and the information in the source document for nominal and categorical data (yes/no) was considered as inaccuracy in the data. Criteria of acceptable values for continuous data were established and are listed in Table 2. For example, if the PDUC Database value for weight deviates from the value found in the source document by ±10 kg, the data was considered to be inaccurate. However, data for continuous variables such as age, dates, number of days (e.g., length of hospital stay), laboratory results (glucose, activated clotting time), and units of blood products transfused must be exactly equal to the values found in the source documents.

The documents used as source for this audit include case notes, perfusion reports, local data collection forms for the PDUC Database, and local electronic databases (for data where the first point of entry is the database).

Planning and Implementation of the Audit Process

A flowchart of the audit process is shown in Figure 1. The audit of the identified PDUC Database participating

Table 1. PDUC Database variables for audit.

Category	Variable	Category	Variable
Demographic	Age	Procedure	Procedure type
	Sex		Coronary artery bypass graft
	Weight		Mitral valve repair
	Height	Outcomes	Death
	Ethnicity		Stroke
Clinical	Chronic pulmonary disease		Encephalopathy
	Extracardiac arteriopathy		Ventilation time
	Neurological dysfunction		ICU stay
	Previous cardiac surgery		Length of hospital stay
	Active endocarditis		Length of postoperative stay
	Critical preoperative state		Discharge status
	Unstable angina		30 day mortality
	Estimated LV function		Death location
	Recent MI		Cause of death
	Urgency		New coma
	Pulmonary hypertension		New renal failure
	Hypercholesterolaemia		Discharge date
	Ejection fraction		Postoperative maximum creatinine
	Congestive heart failure		IABP
	NYHA classification		Return to theatre
	Infarct septal rupture		ICU blood loss 4 hours
Perfusion	Procedure date		ICU blood transfusion
	CPB time		ICU platelet transfusion
	Cross-clamp time		ICU FFP transfusion
	CP induction temperature		Postoperative MI
	Preoperative glucose		Return to ICU
	Final nasopharyngeal temperature		Postoperative dialysis
	ACT baseline		
	ACT final		
	Creatinine > 200 µmol/L		

ACT, Activated clotting time; CP, Cardioplegia; FFP, Fresh frozen plasma; LV, Left ventricular; MI, Myocardial infarction.

Table 2. Audit criteria for continuous variables.

Category	Variable	Criteria for Acceptable Deviation from SD
Demographic	Weight	Must not exceed ± 10 kg
	Height	Must not exceed ± 10 cm
Clinical	Ejection fraction	Must not exceed ± 5%
Perfusion	CPB Time	Must not exceed ± 5 minutes
	Cross-clamp time	Must not exceed ± 5 minutes
	CP induction temperature	Must not exceed ± 1°C
	Final nasopharyngeal temperature	Must not exceed ± .5 kg
	Creatinine > 200 µmol/L	Must not exceed ± .005 mmol/L
Outcomes	Ventilation time	Must not exceed ± 5 minutes
	Postoperative maximum creatinine	Must not exceed ± 5 µmol/L
	ICU blood loss 4 hours	Must not exceed ± 50 mL

CP, Cardioplegia; SD, Source document.

centers was conducted from May–July 2010. A researcher independent from the PDUC conducted the audit. Audit and data change/correction forms specifically designed for the PDUC Database audit were used to ensure proper documentation. During the audit, the Ethics approval for the PDUC Project was checked, which for all the centers currently being audited, were provided by the Flinders

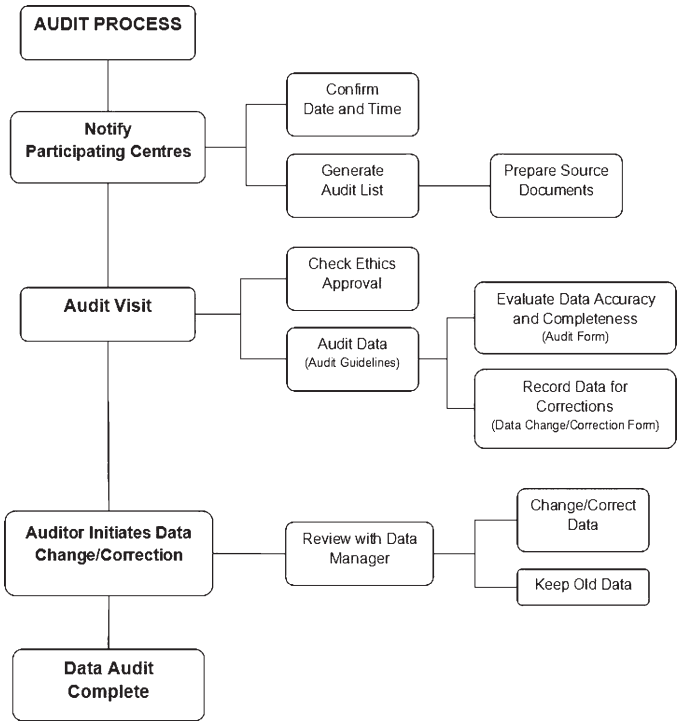

Figure 1. Perfusion Downunder Collaboration Database audit process.

Clinical Research Ethics Committee (Approval 149/09). An extensive review of the source documents was conducted to assess and verify the accuracy and completeness of the data submitted by the participating centers to the PDUC Database. Discrepancies were classified into five categories: *accurate and complete* (data was validated and found to be correct), *inaccurate* (data did not match the data in the source documents), *missing* (data was not entered, sub-classified into validated, or cannot be validated from the source documents), *not entered* (data was actively not entered, sub-classified into validated, or cannot be validated from the source documents), and *data entered* but *cannot be validated* from the source documents. Data that can *only be sighted in the local database* were also identified accordingly to be able to assess their rate of occurrence.

Analysis

Data discrepancies were evaluated for all the data audited (n = 3420). The rate of discrepancy for each of the variables audited (n = 57) was assessed based on number of discrepancies found for all participating centers, as well as for each of the individual centers. The rate of discrepancy was categorized as 0–25%, 25–50%, 51–75%, and 75–100% of cases (% = cases with discrepancy/total cases audited).

RESULTS

There were a combined total of 5 perfusionists and 5 cardiothoracic surgeons at the PDUC Database participating centers audited, which includes FPH, FMC and AH. A total of 635 cases were eligible for audit from the cases submitted to the PDUC Database by FPH (126) and FMC (319) in 2008, and AH (190) in 2009. Sixty-three cases were randomly selected for audit, comprising of 13 cases from FPH, 31 cases from FMC and 19 cases from AH. The audit of all the cases selected for FPH and FMC has been completed; however, only 16 out of the 19 cases selected for AH were audited as three case notes were unavailable. The audit of 60 case notes was completed.

Overview of Data Accuracy and Completeness

A total of 3420 data were evaluated and 76.9% (2629) of these were found to be *accurate and complete*, whilst 6.9% (237) were found to be *inaccurate*. Just over 3% (108) of the data were *missing*, 2.4% (82) of which were able to be validated whilst .8% (26) *cannot be validated*. Data actively *not entered* were 2.2% (78) of the total audited data, 1.5% (53) were validated, whilst .7% (25) *cannot be validated*. Data were available for 2.0% (70) of the variables, which cannot be validated from any source documents whilst 8.7% (298) were *only sighted in the local database*.

Table 3. Number of variables with discrepancy.

	Number of Variables with Discrepancy (n = 57)						
Discrepancy Type	Inaccurate	Missing		Not Entered		Data Entered	
% Cases (n = 60)		*V*	*CV*	*V*	*CV*	*D*	*CV*
75–100%	1	0	0	0	0	4	0
51–75%	1	1	0	0	0	1	0
25–50%	1	1	0	0	0	1	2
0–25%	54	55	57	57	57	51	55

The number of variables audited (n = 57) with discrepancies based on the number of cases (n = 60), grouped into 0–25% (0–15 cases), 25–50% (16–30 cases), 51–75% (31–45 cases), and 75–100% (46–60 cases) are presented.
CV, Cannot be validated from the source documents; D, Only sighted in local database; V, Validated from the source documents.

An overview of the discrepancy rates of the data audited are shown in Table 3 where the percentage of variables audited (n = 57) with discrepancies based on the number of cases (n = 60), grouped into 0–25% (0–15 cases), 25–50% (16–30 cases), 51–75% (31–45 cases), and 75–100% (46–60 cases), are presented.

Almost all of the 57 variables audited each have discrepancies in some (0–25%) of the cases audited. Three variables have more than 0–25% of the cases with data found to be *inaccurate*, and included ventilation time, postoperative maximum creatinine, and the requirement for dialysis postoperatively. Two variables audited have data in more than 25% of the cases audited found to be *missing*, namely the location and the cause of death at 30 days postoperatively. Six variables were not documented in the case notes, however, in more than 25% of cases audited, information was only recorded in the local databases. These variables include the New York Heart Association (NYHA) classification, the volume of blood loss in the first 4 hours postoperatively, the use of intra-aortic balloon pump (IABP) intraoperatively, cardioplegia induction temperature, the nasopharyngeal temperature at the end of CPB, and the number of hours spent in the intensive care unit (ICU). Both the cardioplegia induction temperature and the nasopharyngeal temperature were collected using the electronic Stockert Data Management System (Stockert, Munich, Germany), which was not assessed in this audit. The ICU time was difficult to validate using the case notes, in particular, the time of transfer from the ICU to the ward was often not recorded.

In terms of accuracy and completeness, the overall majority (42 variables) of data were accurate and complete in most (75–100%) of the cases audited. Three variables had accurate and complete data in 51–75% of cases and five in 25–50% of cases. There were seven variables with accurate and complete data that occurred in the 0–25% of cases cohort.

Inaccurate Data by Participating Center

As with the overall data discrepancy rates according to participating center, the majority of the data variables collected by each participating center have less than 25% of the cases with data inaccuracies. The two variables for FPH in the 51–75% of cases with inaccuracies group were related to the number of days patients were in-hospital. Investigation revealed that there was an issue with the automated date difference calculation. For FMC, data for all of the cases for the two variables post-operative creatinine and post-operative dialysis were inaccurate. Examination of the data revealed that post-operative creatinine has been entered in the unit μmol/L, when by definition it should be in mmol/L. For post-operative dialysis, none of the patients whose cases have been audited had dialysis post-operatively and this should be reflected as 0 (no); however, this has been coded as 2 (no) in the database (thus a systemic coding error). Inaccuracies in ventilation time in 51–75% of the cases audited can be attributed to the inclusion of the time when the patients were intubated, instead of indicating the "number of hours post operation for which the patient was ventilated, calculated from the date and time of exit from theatre to that of extubation." Whilst this discrepancy in ventilation time was expected to be found in all cases for FMC, some cases were not able to be validated from the source documents. Inaccuracies in ICU time for 25–50% of cases were also found, which can be attributed to transcription errors, whereby there were differences in ICU entry and exit times entered in the database compared to those recorded in the source documents. The same cause of inaccuracies was also seen for AH for the ventilation time for 25–50% of the cases. However, inaccuracies found for AH were mostly related to the lack of adherence to data definitions, particularly for variables such as pre-operative and post-operative creatinine, and the coronary artery bypass graft procedure type. It was found that data entered for creatinine levels were in a different unit (μmol/L), and post-operative creatinine entries were not always the maximum. For the coronary artery bypass graft procedure type, 25–50% of the cases audited were "off-pump" cases but were entered in the database as "on-pump" surgery.

Missing Data by Participating Center

The majority of the variables audited for FPH and FMC had 0–25% *missing* data that can be validated. Variables with *missing* data common to both centers were coronary artery bypass graft procedure type and cause of 30-day mortality. Other variables with *missing* data were post-operative dialysis (FPH) and location of 30-day mortality (FMC). A difference in data transformation has been found for the variables related to mortality (cause and location). For FPH, these variables were automatically updated as "survived" when death is "no"; however, this was not the case for FMC. For AH, there were very few variables (i.e., activated clotting time and discharge date) with *missing* data. All of the *missing* data that *cannot be validated* were only found in less than 25% of the cases audited for each of the participating centers.

Not Entered Data by Participating Center

Between 55 and 57 variables audited for each of the three participating centers have very few (0–25%) cases with data that were actively *not entered*. Data for ejection fraction (FPH) and mortality at 30-day postoperatively (AH) could not be validated in 25–50% of the cases.

For FPH, all cases audited were *not entered* for the type of mitral valve surgery (whether "repaired or converted to replacement"); however, for FMC and AH, data was actively entered in the database for this variable, regardless of the type of procedure. Discrepancies for the same variables were not found in FMC and AH, in fact for FMC, data that were actively *not entered* occurred in only 0–25% of cases of all the variables audited.

Data Not Validated in Source Documents by Participating Center

Most of the variables audited for the three participating centers have few (0–25%) cases with data that cannot be independently validated. Variables that *cannot be validated* at all in more than 25% of the cases for FMC and AH were those relating to mortality at 30 days, and for AH, this included the location and cause of death. Incompleteness (lack of information) of the source documents was found to be the cause of the inability to validate data relating to mortality at 30 days postoperatively. It is worth noting that for FPH, there were no variables found to have more than 25% of cases that cannot be validated.

Data Only Sighted in Local Database by Participating Center

For each of the participating centers audited, few (0–25%) of the cases audited can *only be sighted in the local database* for majority of the variables. Common to all the participating centers, in more than 25% of the cases audited, variables collected using electronic data management systems (e.g., cardioplegia induction temperature and the final nasopharyngeal temperature on CPB) were only recorded in the local databases. The NYHA classification, the volume of blood loss within the first 4 hours after surgery in ICU, and the ICU time were also only recorded in the local databases. Intensive care unit data were not able to be validated completely where ICU charts were unavailable for audit unless the data were found in the other source documents. The use of IABP intraoperatively was only recorded in the local databases for both FPH and FMC and was not cited in the source documents (although IABP was not used in any of the cases audited), whilst the intubation time and ejection fraction were more specific to FMC and AH, respectively.

Table 4. Data error type and data quality improvement recommendations.

Variable	Data Error	Issue at Participating Center	Data Quality Improvement Recommendations
Length of stay (hospital and postop)	Systematic	Inaccuracy in data difference calculation	Correct programming error
Creatinine	Systematic	Incorrect unit used in the data transfer	Correct programming error
	Random	Incorrect unit used due to lack of adherence to data definition	Correct data based on source documents
Coronary artery bypass graft	Random	Lack of adherence to data definition	Correct data based on source documents
Postoperative dialysis	Systematic	Incorrect code entered in the database	Implement the use of the correct code
Ventilation time	Systematic	Data transformation programming did not adhere to data definition	Correct programming error
ICU stay	Random	Transcription error	Correct data based on source documents
Cause of death and location	Systematic	Data does not automatically update based on the data of 30 day mortality variable	Correct programming/update error
Missing data (e.g., ACT, discharge date)	Random	Transcription error	Enter missing data based on source documents
Mitral valve repair	Systematic	Data transformation programming did not adhere to data definition	Correct programming error

ACT, Activated clotting time.

Data Error Type and Data Quality Improvement Recommendations

The list of variables with commonly occurring errors, the data error type, the issues found at the participating centers, and the recommended actions for data quality improvement are shown in Table 4.

DISCUSSION

Improving the quality of the PDUC Database guarantees that it would better satisfy the needs of the intended use of the data, consequently meeting the goals of the Collaboration. Currently, the quality of the data in the PDUC Database is maintained though quality assurance and control procedures, such as the training of data managers, data validation, and the use of PDUC Database Data Quality Reports. However, formal evaluation of the quality of the PDUC databases could provide valuable insight that can assist in its development. This was evident from the validation and feedback study of data from cases submitted to the Society for Cardiothoracic Surgeons of Great Britain and Ireland database (13). The percentage of missing data was reduced significantly to 9.33% (from 24.96%) ($p < .0001$), after providing feedback based on a retrospective evaluation of cases.

Overall, based on the audited variables and cases that can be validated, the PDUC Database is reasonably accurate and complete, with only 7% of data found to be inaccurate and 3% missing. This is considerably better than what was found in the retrospective evaluation by Fine et al. (13) of 17 data elements from coronary artery bypass grafting surgery cases (54 from 10 centers) submitted to the Society for Cardiothoracic Surgeons of Great Britain and Ireland database whereby a mean of 24.96% of data was found to be missing (13). However, much higher accuracy and completeness is achievable, as found by Datta et al. (14) from evaluating 100 records (14 fields) from the Alberta Trauma Registry. Assessments revealed a 98% accuracy and 99% completion rate.

While the majority of errors found in this evaluation of the PDUC Database were systematic in nature, random types of errors were also found. Systematic errors, such as data transfer and calculation errors of variables relating to mortality (location and cause of death were not updated depending on status of mortality), creatinine levels (using the incorrect unit), requirement for dialysis postoperatively (incorrect transformation of data codes), intubation time (inclusion of time from intubation), and hospital stay (incorrect number of days) were found. Discrepancies in the data for the variable relating to mitral valve surgery were found to be systematic in nature, although these were caused by the lack of adherence or due to unclear or ambiguous data definitions. Two of the centers (FMC and AH) entered data into this variable, whilst for FPH, data was actively *not entered*. The definition "was the mitral valve repaired or converted to replacement?" may have been interpreted in FPH as to be completed only if the procedure involves the mitral valve. However, discrepancies in this variable may have been caused as well by a violation of or unclear data collection guidelines.

Random types of errors such as transcription errors and data source issues (illegible handwriting or incompleteness) may have caused discrepancies in ICU time, either with inaccurate entry of date and time of entry and exit from ICU or inability to find information on these. Inaccuracies in data caused by transcription errors are characteristic of data that are manually collected, as suggested by Arts et al. (1), as is the case for ICU time. Data for variables such as ejection fraction and those relating to mortality at 30 days postoperatively (status, location, and cause) cannot be validated due to incompleteness of the source

documents. Lack of adherence to data definitions may have caused discrepancies in creatinine levels, types of coronary artery bypass graft procedure (differentiating between on-pump and off-pump surgery), and types of mitral valve surgery. Data source issues, in particular incompleteness of the source documents, can be attributed to a good number of data that were unable to be validated. Manually collected data such as blood loss in the first 4 hours after surgery (in ICU), ventilation time, use of IABP intraoperatively, and NYHA classification were only found in the local databases. The local databases may have been the first point of entry for these data.

The use of the PDUC Database front-end has shown to contribute to the assurance of data quality, as in the case of AH, whereby fewer variables were found to have missing data for the center. For FMC and FPH, data missing in audited variables were largely due to systematic errors, i.e., data translation and transfer coding issues from the local databases to the PDUC Database. With the help of the participating centers' data managers, the assessment of the different types of errors and the recommended actions to correct the errors should ensure data quality improvement. However, the effectiveness of this evaluation and the recommendations for data quality improvement, upon implementation, could only be realized upon the re-evaluation of the PDUC Database.

Additional data quality assurance procedures that are carried out before, during, and after the collection of data, can be implemented to improve further the quality of data in the PDUC Database. Good data definitions, that are clear and unambiguous, are critical for accurate data collection (15). Availability of a comprehensive documentation of data collection guidelines, as suggested by Gassman et al. (16), would assist participating centers in ensuring that data submitted from each case are collected in a standardized manner and less likely to result in data discrepancies. Designing a high quality data collection form would also ensure standardized data collection, particularly for information only entered in the local databases. The completed data collection forms can also be used as a source document during monitoring and auditing. Periodic training of database personnel on data definitions and database guidelines could also improve data accuracy, as shown by Arts et al. (15). Training improved the accuracy of data from 79–86% ($p < .01$) in the Dutch National Intensive Care Evaluation registry. Training also reduced the frequency of errors such as incompleteness, non-adherence to data definitions, and inter-observer variability.

A routine data monitoring program, carried out as close as possible to data collection, could be implemented as part of quality assurance (16). Data which are simple to monitor and relevant to the aims of database edit of data, may be selected by the custodians of the PDUC Database. Data managers could be involved in double checking the accuracy of data entered prior to signing off particular sections of the database. External audit, similar to what was implemented in this evaluation, could also conduct periodic data monitoring of a particular sample size.

This evaluation has limitations in terms of the sample size of cases audited from each participating center. Ten percent of the total number of cases in the time period selected for each participating center was selected as there was a lack of consensus in literature. The audit was also conducted on an inter-related group of hospitals from one city. Evaluation of the electronic data capture by the data management systems used will also provide better insights on the accuracy of data for variables such as cardioplegia induction temperature and nasopharyngeal temperature at the end of CPB. It is also recognized that critical source documents such as the ICU chart were not available for review for this audit and that source documents used for this evaluation, whilst used as the gold standard, may also have inaccuracies.

CONCLUSION

The PDUC Database is currently reasonably accurate and complete. In addition to the current quality assurance and control procedures (e.g., training of data managers, data validation, and the use of data quality reports) that are already in place, additional data quality assurance procedures could be implemented. These might include making available a comprehensive documentation of data collection guidelines and conducting routine data monitoring that would further improve data quality. This evaluation is part of a complex system of data quality assurance, and when conducted routinely, could provide a continuous feedback loop towards a high quality PDUC Database.

ACKNOWLEDGMENTS

The author would like to thank the support and assistance of the Perfusion Downunder Collaboration.

REFERENCES

1. Arts DGT, De Keizer NF, Scheffer G-J. Defining and improving data quality in medical registries: A literature review, case study, and generic framework. J Am Med Inform Assoc. 2002;9:600–11.
2. Fortuna D, Vizioli M, Contini A, et al. Assessing clinical performance in cardiac surgery. Does a specialised clinical database make a difference? Interact Cardiovasc Thorac Surg. 2006;5:123–7.
3. Peterson ED, Coombs LP, Ferguson TB, et al. for the STS National Cardiac Database Investigators. Hospital variability in length of stay after coronary artery bypass surgery: Results from the society of thoracic surgeon's national cardiac database. Ann Thorac Surg. 2002;74:464–73.
4. Ferguson TB Jr, Hammill BG, Peterson ED, DeLong ER, Grover FL, for the STS National Database Committee. A decade of change—risk

profiles and outcomes for isolated coronary artery bypass grafting procedures, 1990–1999: A report from the STS National Database Committee and the Duke Clinical Research Institute. Ann Thorac Surg. 2002;73:480–90.
5. Yan BP, Clark DJ, Buxton B, et al. on behalf of the Australasian Society of Cardiac and Thoracic Surgeons (ASCTS), the Melbourne Interventional Group (MIG). Clinical characteristics and early mortality of patients undergoing coronary artery bypass grafting compared to percutaneous coronary intervention: Insights from the Australasian Society of Cardiac and Thoracic Surgeons (ASCTS) and the Melbourne Interventional Group (MIG) Registries. Heart Lung Circ. 2009;18:184–90.
6. Black N, Payne M. Improving the use of clinical databases. BMJ. 2002;324:1194.
7. Black N. High-quality clinical databases: Breaking down barriers. Lancet. 1999;353:1205–6.
8. Sørensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. Int J Epidemiol. 1996;25:435–42.
9. Newland R, Baker RA, Stanley R, Place K, Willcox TW. The Perfusion Downunder Collaborative Database Project. J Extra Corpor Technol. 2008;40:159–65.
10. Reid C, Billah B, Dinh D, et al. An Australian risk prediction model for 30-day mortality after isolated coronary artery bypass: The AusSCORE. J Thorac Cardiovasc Surg. 2009;138:904–10.
11. Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg. 1999;16:9–13.
12. Shroyer AL, Coombs LP, Peterson ED, et al. The Society of Thoracic Surgeons: 30-day operative mortality and morbidity risk models. Ann Thorac Surg. 2003;75:1856–65.
13. Fine LG, Keogh BE, Cretin S, Orlando M, Gould MM. on behalf of the Nuffield-Rand Cardiac Surgery Demonstration Project Group. How to evaluate and improve the quality and credibility of an outcomes database: Validation and feedback study on the UK Cardiac Surgery Experience. BMJ. 2003;326:25–8.
14. Datta I, Findlay C, Kortbeek JB, Hameed SM. Evaluation of a regional trauma registry. Can J Surg. 2007;50:210–3.
15. Arts DGT, Bosman RB, de Jonge E, Joore JCA, de Keizer NF. Training in data definitions improves quality of intensive care data. Crit Care. 2003;7:179–84.
16. Gassman JJ, Owen WW, Kuntz TE, Martin JP, Amoroso WP. Data quality assurance, monitoring, and reporting. Control Clin Trials. 1995; 16:104S–36S.