# Assessing data quality issues in the Emergency Department through data and process mapping

Jing Gao
*University of South Australia*, jing.gao@unisa.edu.au

Andy Koronios
*University of South Australia*, Andy.Koronios@unisa.edu.au

# Assessing data quality issues in the Emergency Department through data and process mapping

Jing Gao
School of Computer and Information Science
University of South Australia
Adelaide, Australia
Email: jing.gao@unisa.edu.au


Andy Koronios
School of Computer and Information Science
University of South Australia
Adelaide, Australia
Email: andy.koronios@unisa.edu.au


Eui-Soo Choi
Demand and Performance Evaluation Branch, NSW Health
Sydney, Australia
Email: echoi@doh.health.nsw.gov.au

## Abstract

*Data and information quality is a well-established research topic and gradually appears on the decision-makers' top concern lists. Many studies have been conducted on how to investigate the generic data/information quality issues and factors by providing a high-level abstract framework or model. Based on these previous studies, this study tries to discuss the actual data quality issues with the operation-level and middle-level managers emerged during the emergency department data collection and reporting processes. By conduct data quality issues and business processes mapping, possible data quality issues are summarised under the well-known TOP model and the recommendations of data quality improvement are suggested.)*

## Keywords

Data Quality, Information Quality, Health Service and Emergency Department

## INTRODUCTION

In Australia, the public health system is managed by the Department of Health or a similar government organisation within each state. At the Department of Health (DoH) level, managers rely on data to report, manage and lead. Gendron and D'Onofrio (2001)'s data quality work in the healthcare industry point out that the business contribution of leadership and management activities depends on the quality of the decisions that are made and, concomitantly, the quality of the data used to make them. As a result, the level of data quality directly impacts on health service performance and funding opportunities.

Within the areas of public health services, emergency department (ED) performance is considered as a key performance benchmark, which often draws public attention. In order to meet the performance requirement, McKee (2001) asserts that data availability and data quality are critical issues for the health systems performance measures. It must be noted that a number of studies have investigated the data quality indicators in the ED such as Graff et al 2002 and Lindsay et al 2002; and other researchers (Bazarian and Schneider 1996, Lewandrowski 2004, Holland et al 2005) looked into the process improvement in ED on issues such as how to reduce the patient length of stay.

By bringing the two trends of research together, this study tries to adopt a process-oriented approach to understand how data quality issues emerged through the ED data collection and reporting processes.

### Health Service Performance Quality Indicators

The performance is measurable through key performance indicators (KPIs). KPIs are very important to management decision-making and are relied upon by all levels of an organisation to measure success in achieving outcomes. The public health services need to identify its high-value high-risk data quality issues that support each KPI. However, in practice, this process is often conducted unsatisfactorily. For example, in public

organisations, Theurer (1998) indicates that there are several "pitfalls" of performance measurement (and the use of KPIs), closely related to the data quality management:

- Data by themselves have no meaning
- There is often lack of a strong commitment from leaders to move toward measuring performance and not just collecting data on effort
- Employees may not have the capacity to develop measures, so they use whatever 'measures' are already available
- If measurement focuses on negative accountability, managers and employees will seek to avoid accountability when things go wrong

In health services, the review of both national and international health quality indicators (e.g. OECD 2006) suggests that the most common KPIs dimensions for Health services are:

- *Effectiveness*: the degree of achieving desirable outcomes, given the correct provision of evidence-based healthcare services to all who could benefit, but not to those who would not benefit;
- *safety*:  the degree to which health care processes avoid, prevent, and ameliorate adverse outcomes or injuries that stem from the processes of health care itself;
- *Responsiveness*: how a system treats people to meet their legitimate non-health expectations (or often discussed as patient-experience);
- *Accessibility*: the ease with which health services are reached. Access can be physical, financial or psychological, and requires that health services are a priori available;
- *Equity*:  defines the extent to which a system deals fairly with all concerned; and
- *Efficiency:* is the system's optimal use of available resources to yield maximum benefits or results.

These dimensions of KPIs will directly influence the data quality requirements in the ED processes and are considered in the design of this study.

## DATA QUALITY

Managers intuitively differentiate information from data, and describe information as data that has been processed. However, data and information are often used synonymously in practice, particularly when addressing quality issues. Therefore, this paper uses "data" interchangeably with "information", as well as using "data quality" (DQ) interchangeably with "information quality" (IQ).

The quality of data is multi-dimensional (Wang & Strong 1996) including accuracy, reliability, importance, consistency, precision, timeliness, fineness, understandability, conciseness, and usefulness (Ballou & Pazer 1995; Wand & Wang 1996). It is also suggested that the quality of data is dependent on how the data will be used (Ballou & Pazer 1995; Neely 2001; Strong 1997; English 1999; Salaun and Flores 2001; Orr 1998). This fitness for use can be defined as the intersection of the quality dimension being considered, the use of data (purpose) and which data fields are used in order to fulfil the purpose (Neely 2002).

When selecting appropriate data quality dimensions to investigate ED processes. This paper has adopted the four key dimensions that have been included in the Canadian Institution for Health Information data quality framework (CIHI 2009).  In the health domain this is likely to include the most important of all the dimensions of data quality, that of: **accuracy**, **timeliness**, **comparability**, and **usability** of clinical data. These dimensions of data quality, once defined need to be used consistently across units of health care and across time if comparisons and benchmarking are to be meaningful. The **relevance** dimension (the degree to which a database or registry meets the current and potential future needs of users) of the CIHI data quality framework is not fully considered as this study will not look into the actual database design of various ED systems, however, a recommendation has been made on how this can be included.

The above dimensions can be further decomposed into a number of elements. This will enable health institutions to adopt the definition that suits their data quality strategy most closely. With respects to the Canadian Institution for Health Information data quality framework and other data quality literature, the decomposed data quality dimensions for this study are:

| Dimensions | Decomposed | Descriptions |
|---|---|---|
| Accuracy | **Accuracy** | Data entered correctly according to specified formats |
| | | Data Values are within specified ranges |
| | **Completeness** | All required data fields are entered at each data entry point |
| Timeliness | **Timeliness** | Data entered on-time to meet the collection purposes |
| Comparability | **Consistency** | Data in the system is a true reflection of what actually happened |
| | **Integrity** | Changes are authorised and reasons are legitimate and recorded |
| Usability | **Conformity** | Data entered as a result of correct interpretation of polices & standards |

## RESEARCH DESIGN

This study tries to adopt a process-oriented approach to understand how data quality issues emerged through the ED data collection and reporting processes. A specific focus has been placed on obtaining an understanding of how data quality issues emerged through the ED data collection and reporting processes. Thus understanding will help better design the ED process to improve the overall service performance. Based on the literature, the following research model was developed to guide this study.
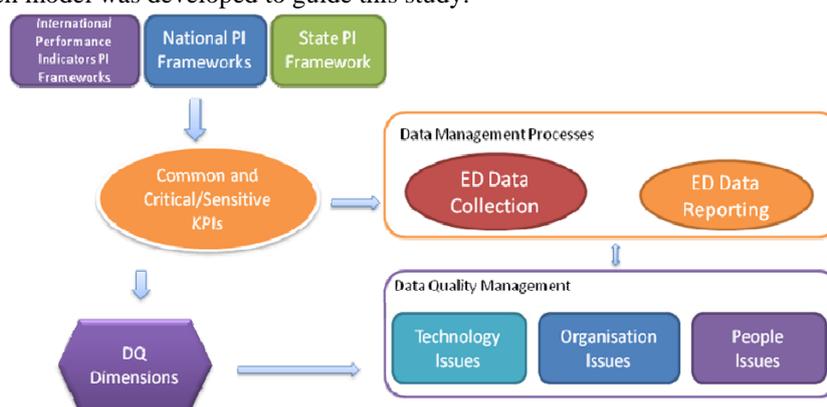


Figure 1: Research Model

The data quality issues are categorised under the well-known TOP framework. Mitroff and Linstone (1993) argue that any phenomenon, subsystem or system can be analysed from what they call a Multiple Perspective method – employing different ways of seeing, to seek perspectives on the problem. The TOP model allows analysts to look at the problem context from either Technical or Organisational or Personal points of view:

- The technical perspective (T) sees organizations as hierarchical structures or networks of interrelationships between individuals, groups, organizations and systems.
- The organisational perspective (O) considers an organization's performance in terms of effectiveness and efficiencies. For example, leadership is one of the concerns.
- The personal perspective (P) focuses on individual concerns. For example, the issues of job description and job security are some of the main concerns in this perspective.

This study was conducted jointly with an Australian State Department of Health, which serves multi-millions of its citizens. This study consists of two stages:

- Stage One: Primary data collection was conducted through observation in various ED departments, interviews with ED staff and non-ED staff who have roles of data stewards, data custodians, data users and data collectors in relation to the ED data management. Overall, more than 30 interviews were conducted through individual or group face-to-face meetings and teleconferencing.
- Stage Two: The ED data collection and submission process flow charts were created. Data quality issues were summarised against the critical tasks within the processes and presented by using the TOP model.

Based on the findings, a data quality scorecard is suggested for assessing the data quality in ED systematically.


# FINDINGS

The performance of Emergency Department is as a key performance indicator in the public health services. The waiting time in emergency department is often used for benchmarking purposes. The flow charts below provide the descriptions for major tasks of the ED visits and associated data quality issues. However, these tasks are not designed as a linear process. Thus, the process model is attached at the end to indicate the actual process.

**Data Collection Process**


**Arrival in Person**

The majority of patients arrives the hospital in person. It must be noted that some of these patients may not necessarily seek urgent assistance, but rather require simple clinical procedures due the inability to access these services from the general practitioners.

The arrival time does not start from the moment that these patients enters the ED. Instead, it starts from the moment that he / she has been registered in the system by a triage nurse.

Data Quality Issues to be assessed

| Perspectives | Issues | DQ Dimensions |
|---|---|---|
| Technical | | |
| Organisational | Does the current policy take the actual waiting time before triage into consideration? | Completeness Consistency |
| People | | |

Table 1:  Arrival in person – Data Quality Issues


**Arrive in Ambulance**

A critical patient may arrive ED in ambulance. They will be transferred to an ambulance waiting area accompanied by ambulance staff.  These patients will be seen by a triage nurse in the first instance.

Data Quality Issues to be assessed

| Perspectives | Issues | DQ Dimensions |
|---|---|---|
| Technical | Is the ambulance information system linked to the ED systems? | Consistency |
| Organisational | Does the current policy take the actual waiting time before triage into consideration? | Completeness |
| | There is an ambulance entry timestamp (when the ambulance staff clicks the button) recorded in the Ambulance system, but is this data used as the starting time for ED visits in the ED system? | Consistency |
| People | | |

Table 2: Arrive in Ambulance – Data Quality Issues


**Triage**

Triage is officially considered as the start of ED visit. A triage nurse will ask for patient details and assign a triage category for the patient. This task usually takes 2~3 minutes. The acquired data is entered into the system and a start time has been recorded. Once the triage is completed, the patient will be given a paper form to fill and asked to wait for the treatment.

It must be noted that in some instances, triage can take an average of 29 minutes in one state public hospital. Due the lack of resources, the patients cannot be seen and treated by a nurse or nurse practitioner (NP) or a doctor within the allowed time for the specific triage category that they have been assigned. Thus, instead of keeping the patient waiting, the triage nurse takes an effort to conduct an extensive examination of the patient which is beyond of the scope of his / her duties.

Data Quality Issues to be assessed

| Perspectives | Issues | DQ Dimensions |
|---|---|---|
| Technical | | |
| Organisational | Due the lack of resources or high number of presentations, can the triage take a long time to complete? | Timeliness |
| People | Is the triage conducted based on the triage staff's expertise? Human errors may occur. | Accuracy |
| | Is it possible that the patient's situation may change over time (which may require changes in the triage category) and how is the data updated in the system to reflect this change? | Timeliness<br><br>Integrity |

Table 3: Triage – Data Quality Issues

**Initial Registration**
Once the patient fills the form, they will submit the form to a data entry / admin clerk. The clerk will enter the patient information into the ED system.

Data Quality Issues to be assessed

| Perspectives | Issues | DQ Dimensions |
|---|---|---|
| Technical | Is the ED information system user-interface not well-designed to neither embed the data verification rules nor highlight critical data items? | Accuracy<br><br>Completeness<br><br>Conformity |
| Organisational | Does the current process track those (patients) who have submitted the form to be entered into the system, but have decided to leave the ED due to long waiting times? | Consistency<br><br>Completeness |
| People | Is it possible that the patient may never complete the form as they have been called for the treatment? | Accuracy<br><br>Completeness |
| | Is it possible that the clerk may misread the patient's hand-writing and not enter the correct data into the system? | Accuracy |

Table 4: Initial registration – Data Quality Issues

**Treatment**
The patient receives the treatment from nurses, NPs and doctors. Some emergency departments have allocated fast-track areas for simple procedures. During the treatment, nurses, NPS and doctors will record the procedures in the ED system (with or without the need of individual logins depending on the actual implement of ED systems).

Data Quality Issues to be assessed

| Perspectives | Issues | DQ Dimensions |
|---|---|---|
| Technical | When the ED and PAS systems are not developed by one vendor, is there any data integration issue (e.g. requires a manual process)? | Integrity<br>Consistency<br>Accuracy |
| Organisational | Different area of services / hospital may use different ED and PAS systems. | Consistency<br>Conformity |
| People | Is it possible that the patient may not provide all required data? For example, they may not have brought their insurance details. | Accuracy<br>Timeliness<br>Integrity<br>Completeness |

Table 5: Treatment – Data Quality Issues

**Admitted**
A proportion of patients will be admitted to the hospital. Once they have been physically transfer to the ward, the exit time is recorded as the end of ED visit. Their data will be transferred either manually or electronically to the Patient Admission System (PAS) systems.

Data Quality Issues to be assessed

| Perspectives | Issues | DQ Dimensions |
|---|---|---|
| Technical | The ED information system does not have individual accounts for audit trails when a change has been made. Changes can be made at any time. | Integrity<br>Duplication |
| Organisational | Is it true that the use of SNOMED CT system for clinical coding may be over-complicated in the ED environment? | Accuracy<br>Consistency |
| People | Is it possible that it is not the first priority of the individual who updates the system records, especially if the staff are busy treating other patients?<br><br>There is a possibility for individuals to change the data for various reasons - e.g. assign a lower triage category | Accuracy<br>Timeliness<br>Integrity<br><br>Accuracy<br>Integrity |

Table 6: Admitted– Data Quality Issues

**Discharge**
Some patients will be discharged from the hospital. A nurse, NP, doctor will record the departure time and discharge the patient. In some cases, admin clerks will check whether the patient data and departure time has been recorded properly.
If the patient arrives in ambulance, regardless whether he / she is admitted or discharged, the ambulance departure time is also recorded (the ambulance staff needs to click a button before leaving the hospital).
It is noted that there are inconsistency in recording the departure time in the system as it is not the exact moment that patient leaves the ED. The actual time can be when the ED staff has time to enter the data.

Data Quality Issues

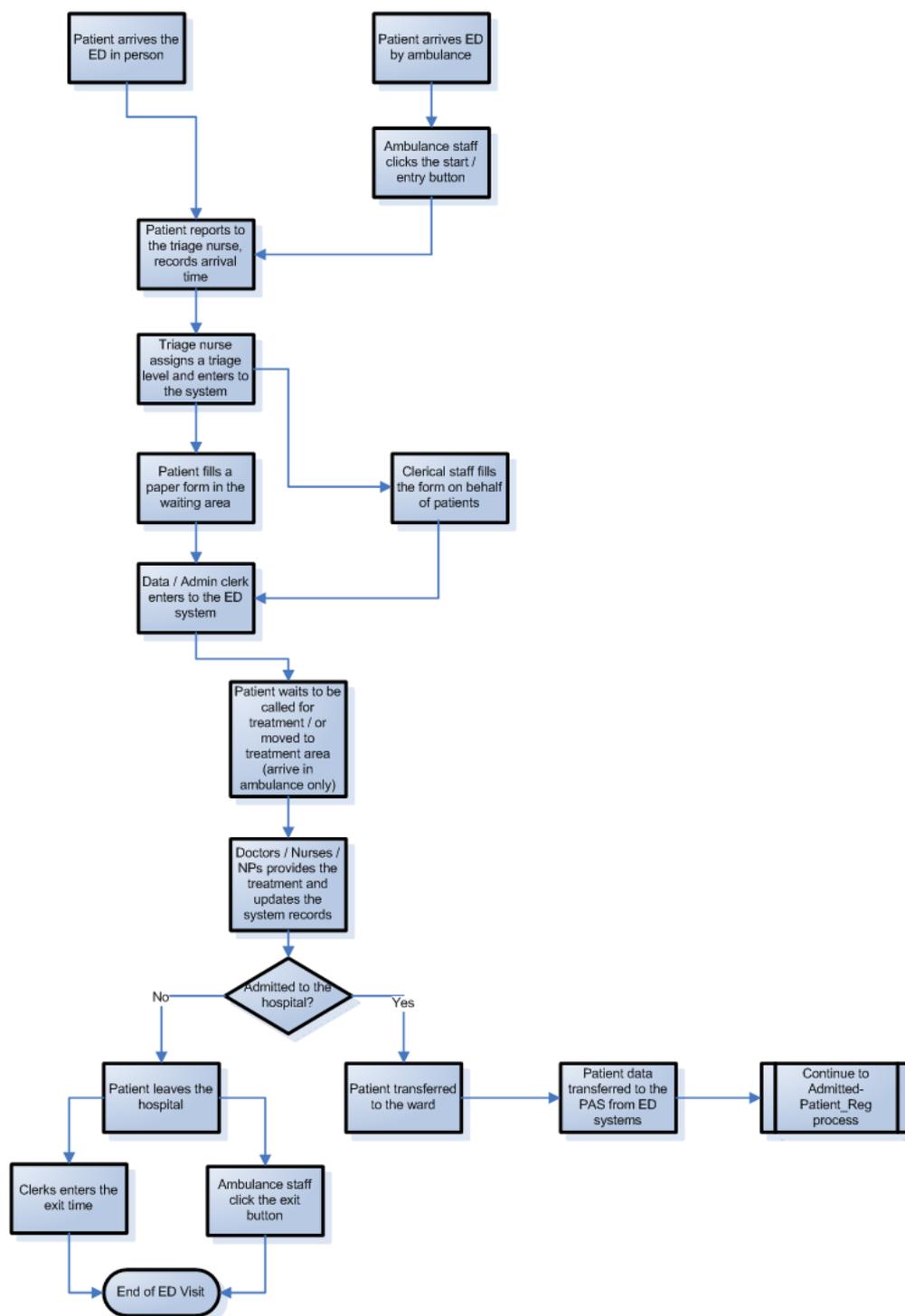| Perspectives | Issues | DQ Dimensions |
|---|---|---|
| Technical | | |
| Organisational | | |
| People | Is it possible that the departure time is not the exact moment that patient leaves the ED? The actual exit time can be the time when the ED staff has time to enter the data. | Accuracy<br><br>Timeliness |

Table 7: Discharge – Data Quality Issues

Figure 2: ED Data Collection Process

**ED Data Reporting**

**Data Extraction to Area and DoH HIES**
The data managers will firstly correct the errors found in the source systems and run the extraction to provide data from local systems into Area and DoH Health Information Exchanges (HIEs).

Data Quality Issues to be assessed

| Perspectives | Issues | DQ Dimensions |
|---|---|---|
| Technical | Depending on the extracting methods and the source systems, is it possible that data may be corrupted or truncated? | Accuracy<br><br>Consistency<br><br>Completeness<br><br>Integrity<br><br>Conformity |
|  | Currently there is no official method for local area staff to access the data within the Cerner systems in order to run data quality checks (the built-in 13 reports failed in providing these required functions). | Accuracy<br><br>Consistency<br><br>Completeness |
| Organisational | Depending on the local resources and actual system, is it possible that the quality of data verification and correction before extraction may vary between different facilities? | Accuracy<br><br>Consistency<br><br>Completeness<br><br>Integrity |
| People | Is it true that the number of data errors corrected during this early stage depends on the personal motivation and technical skills of local area staff? | Accuracy<br><br>Consistency<br><br>Completeness<br><br>Conformity |

Table 8: Data Extraction – Data Quality Issues

**Data Quality Verification and Correction by DoH**
The Department of Health will perform data quality checks and make reports to local areas of services. When data errors were found, the local area staff were asked to fix the errors in the source systems. Once the errors are fixed, a re-extraction will be performed to ensure the area and DoH HIEs are loaded with correct data.

Data Quality Issues to be assessed

| Perspectives | Issues | DQ Dimensions |
|---|---|---|
| Technical | The data extraction processes from the source systems to HIEs are not real-time. | Timeliness |
|  | During the extraction process, is it possible that data may be corrupted and truncated? | Accuracy<br><br>Completeness |
| Organisational | No data quality benchmarks are provided to local area of services | Consistency<br><br>Conformity |
| People | No reward policy is implemented to high DQ achievers. | Accuracy<br><br>Timeliness |
|  | Is it true that the data quality feedback has not been made available / accessible to the data collectors? | Conformity<br><br>Consistency<br><br>Integrity |

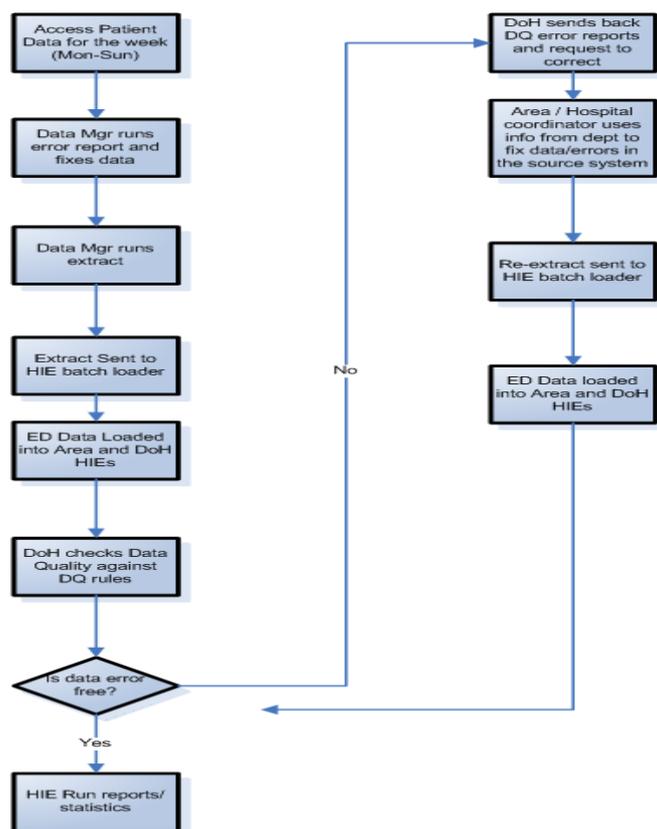Table 9: Data Quality Verification and Correction – Data Quality Issues

Figure 3: ED Data Reporting Process

## CONCLUSION

The development of the ED process maps is central to a comprehensive data quality assessment. These process models are likely to be available as part of the organisations' enterprise architecture documentation but this is not guaranteed (e.g. not found in the participating organisation). These process maps will not only serve as a roadmap of where to look for data quality problems but would also allow for possible rationalization and optimization of information resources by identifying redundancies, under-utilisation of data resources as well as in identifying conflicting definitions.

As much as possible, peoples' attitudes towards the care processes and particularly the IT systems which support these processes should also be recorded and analysed as they would provide clues about possible 'workarounds' and also the motivations that may or may not drive them to have commitment to high levels of data quality. If for example staff express dissatisfaction with IT systems as cumbersome and unresponsive to their needs, it is likely to point to them not taking the requisite care in data entry for example or point to a practice of them maintaining their own locally based systems. Although such practices may be tolerated, proper governance of such data resources would be critical.

Organisational issues such as local culture for example must also be thoroughly examined to identify vulnerabilities such as duplication of data, contradictory classifications of data or differing interpretations of mandated or advisory departmental standards.

Based on the findings, the following improvement activities are recommended to manage the data quality issues emerged through the ED data collection and reporting processes.

| Category | Data Quality Improvement Activities |
|---|---|
| Technology | • Data Quality profiling technology<br>• Metadata technology for data lineage to provide everyone in the data supply chain with the ability to see what is going where and how it is changed |
| Organisation | • Systematic processes to identify data quality errors within the data supply chain<br>• Develop data quality metrics<br>• Procedures to monitor the data quality metrics over time |
| People | • Data ownership, stewardship and other components of a data governance framework<br>• Procedures for communicating data quality activities to build the business case for winning executive commitment and sponsorship |

Table 10: The TOP model for Data Quality Improvement Activities

This study also indicates that understanding and mapping the clinical processes and associated 'data and information path' would be critical in establishing the points for assessing the data quality along the various business processes. When taking the actual data item into consideration (including the relevance dimension from the CIHI 2009's data quality framework), a data quality scorecard which links data, process and KPIs together can be established to help organisations assess their data quality performance systematically. An example of such a data score card is shown below (note: risk levels are assessed as impacts of DQ errors on the quality of actual services provided):



| Data Items | Purpose of Collection | Status | Data Collection Process / Task Impact on DQ - Risk Assessment | Accuracy 30% | Consistency 20% | Completeness 20% | Conformity 10% | Timeliness 10% | Integrity 10% |
|---|---|---|---|---|---|---|---|---|---|
| ED Visit Identifier | Unique Visit Identifier | Mandatory | Triage nurse assigns a triage level and enters to the system | N/A | N/A | N/A | N/A | N/A | N/A |
| Triage Category | Triage Benchmark KPI | Mandatory | Triage nurse assigns a triage level and enters to the system | High | Low | Low | High | High | High |
| Triage Date | Triage Benchmark KPI | Mandatory | Triage nurse assigns a triage level and enters to the system | High | Low | Low | Low | High | High |
| Triage Time | Triage Benchmark KPI | Mandatory | Triage nurse assigns a triage level and enters to the system | High | High | High | Low | High | High |
| First Seen Time | Triage Benchmark KPI | Mandatory | Doctors / Nurses / NPs provides the treatment and updates the system records | High | Low | High | Low | High | High |

Data Quality Dimensions:
| | |
|---|---|
| Accuracy | Data entered correctly according to specified format within specified range |
| Consistency | Data entered into the system is a true reflection of what happened |
| Completeness | All required data fields are entered at each data entry point |
| Conformity | Data entered as a correct interpretation of the polices and various coding standards |
| Timeliness | Data entered on-time to meet the reporting purpose |
| Integrity | Changes in data are authorised and reasons of changes are legitimate and recorded |

Data Collection Process / Task Impact on DQ
| | |
|---|---|
| High | Critical DQ risks as results of Technology, Organisation and People issues |
| Moderate | Moderate risk |
| Low | Low risk |
| N/A | Not affected |

Data Quality Dimension Weighting
| | |
|---|---|
| | Jointly decided by the committee |

Figure 4: Data Quality Scorecard

Finally it must be acknowledged that the provision of healthcare during an encounter between a patient and a clinician (nurse, NP, triage, or doctor) may involve additional issues such as the governance of service providers' interpretation of individual patient data utilizing their existing knowledge and guideline. Since these issues may or may not result in data quality problems, further study is required to explore insights.

## REFERENCES

Ballou, D.P. & Pazer, H.L. (1995) Designing Information Systems to Optimize the Accuracy-timeliness Tradeoff. *Information Systems Research*. 6(1), 51-72

Bazarian JJ, Schneider SM et al. Do admitted patients held in the emergency department impair throughput of treat and release patients? *Acad Emerg Med*. 1996; 3(12): 1113-1118.

Canadian Institute for Health Information (CIHI), (2005), *The CIHI Data Quality Framework* 2009, URL: http://www.cihi.ca/cihiweb/en/downloads/data_quality_framework_2009_e.pdf, online accessed [1/6/2010]

English, L.P., (1999) *Improving Data Warehouse and Business Information Quality: Methods for reducing costs and increasing Profits*, Willey & Sons

Gendron S. M. and D'Onofrio, J. M. (2001), Data Quality in the Healthcare Industry, *Data Quality*, 7:1

Graff L, Stevens C, et al. (2002), Measuring and Improving Quality in Emergency Medicine. *Acad Emerg Med*. 9 (11): 1091-1107.

Holland L, Smith L et al., (2005)Reducing Laboratory Turnaround Time Outliers Can Reduce Emergency

Department Patient Length of Stay. *Am J Clin Pathol.* 2005; 125(5) 672-674.

Lewandrowski K. (2004), How the clinical laboratory and the emergency department can work together to move patients through quickly. *Clinical Leadership Management Review*, 18(3): 155-9.

Lindsay P, Schull M, et al. (2002), The Development of Indicators to Measure the Quality of Clinical Care in EDs Following the Modified-Delphi Approach. *Academic Emergency Medicine*, 2002; 9(11): 1131-1139.

McKee, M. (2001): Measuring the efficiency of health systems. *British Medical Journal*, 323(7308): 295-296.

Mitroff, I. I. and H. A. Linstone, (1993), *The Unbounded Mind: Breaking the Chains of Traditional Business Thinking*, New York, Oxford University Press, 1993.

Neely, M.P. (2001) A Proposed Framework for the Analysis of Source Data in a Data Warehouse, *Proceedings of the Sixth International Conference on Information Quality*, November 2001, Cambridge, MA, USA, MIT

OECD, (2006), "Health Care Quality Indicators Project Conceptual Framework Paper", Editors: Edward Kelley and Jeremy Hurst, *OECD Health Working Paper*, DELSA/HEA/WD/HWP(2006)3

Orr, K. (1998) Data Quality and System Theory. *Communications of the ACM*. 41(2), 66-71

Salaun, Y., and Flores, K., (2001), "Information Quality: Meeting the Needs of the Consumer", *International Journal of Information Management*, 21(1), 2001, pp. 21-37.

Strong, D.M. (1997) IT Process designs for Improving Information Quality and reducing Exception Handling: A Simulation Experiment. *Information and Management*. 31, 251-263

Theurer, J., (1998), Seven pitfalls to avoid when establishing performance measures. *Public Mgt.*, 80(7), p21

Wand, Y. & Wang, R.Y. (1996) Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM*. 39(11), 86-95

Wang, R.Y., and Strong, D.M., (1996) "Beyond Accuracy: What Data Quality Means to Data Consumers", *Journal of Management Information Systems*, 12(4), 1996, pp. 5-33.

## COPYRIGHT