

# Health Data Ownership and Data Quality: Clinics in the Nyandeni District, Eastern Cape, South Africa

Graham Wright<sup>1</sup> & Anthony Odama<sup>2</sup>

<sup>1</sup> Faculty of Health Sciences, Walter Sisulu University, Mthatha, South Africa

<sup>2</sup> Regional Training Centre, Walter Sisulu University, Mthatha, South Africa

Correspondence: Graham Wright, Faculty of Health Sciences, Walter Sisulu University, Mthatha, South Africa.  
Tel: 27-718-537-370. E-mail: profwright@gmail.com

Received: August 17, 2012    Accepted: October 8, 2012    Online Published: October 26, 2012

doi:10.5539/emr.v1n2p146

URL: <http://dx.doi.org/10.5539/emr.v1n2p146>

## Abstract

**Objectives:** The aim of this study was to determine the type of relationship between data ownership and data quality at the primary health care facilities of the Nyandeni sub-district of the Eastern Cape Province in South Africa.

**Method:** A data audit was conducted to assess the quality of clinical data at primary health facilities. Structured interviews and documentary analysis methods were used to determine whether clinicians at these health facilities were using collected secondary data for decision-making.

**Results:** Of the five data quality attributes health facilities were audited on, only the timeliness of data reports was found to be satisfactory. Data quality was of a poor standard and there was no evidence to suggest data collected for secondary purposes was being used for decision-making by clinicians at the primary health facilities.

**Conclusion:** The study highlights that to improve the quality of data; clinicians need to be involved in the measurement of the quality of care that they provide. This not only serves to improve the quality of service provided but also helps clinicians appreciate the value of their work and enhances the importance of collecting quality clinical data. Clinicians as data collectors are the best placed individuals to recommend a course of action based on data they receive and are also the best placed individuals to suggest whether better ways to measure results exist.

**Keywords:** data collection, data quality, data ownership

## 1. Introduction

Information forms a key resource in any health care institution, without information a health care institution would simply not function. Examples of information use within a health care institution can involve a simple verbal communication between a patient and a member of the medical staff, to the use of health data for the administrative running of a health institution. In the Eastern Cape Province of South Africa, health information collected at health care facilities is sent to sub-district offices, from there it is sent to district then provincial offices. The provincial office in turn sends information to the national office and from there it is sent to international agencies like the World Health Organisation. In order to facilitate the movement of data through the various levels of health services, collected health data has to conform to a recognised standard. The Eastern Cape Province forms one of the nine provinces of South Africa. The Province is further divided into seven districts, each of which is further subdivided into sub-districts. In the Nyandeni sub-district of the Oliver Reginald Tambo district of the Eastern Cape Province standardisation in the primary health care setting comes in the form of a minimum dataset. The primary health care facilities collect secondary data, which is data, collected by someone other than the user. It is in the form of data elements specified in the minimum data set; these data elements are then used to calculate clinical indicators. Clinical indicators in turn serve largely as quantitative measures used to monitor and evaluate the quality of important governance, management, clinical and support functions that affect patient outcomes.

## 2. Method

The study sought to determine how health information is collected, processed, analysed and used, as well as to ascertain the systems that have been put in place to ensure the quality of information gathered is of an acceptable standard. Importantly, the task was to find out what attributes of data quality the Province valued and how these attributes related to classical data quality definitions.

The investigation focused on the primary health care facility level. The task involved conducting a data audit on information collected for epidemiological purposes as well as documenting the different types of data primary health care facilities collected for use in their daily activities. The audit sought to determine the accuracy of reported data based on data quality attributes valued by the provincial information unit. Part of the investigation took place at Provincial department of health offices as well as District and sub-district offices. The purpose was to assess how service delivery and intermediate aggregation sites (department of health offices at the sub-district and district level) were collecting and reporting data to measure the audited HIV/AIDS clinical indicators and if these were accurate, and completed on time.

A data audit structured interview schedule was developed by combining questions derived from the provincial department of health primary health assessment tool, and the President's Emergency Plan for AIDS Relief (Pepfar) data quality audit tool. Pepfar provides a Data Quality Assessment Tool, which supplies a methodology to assess the ability of systems like health facilities to collect and report quality data.

A total of 12 primary healthcare centres were conveniently selected from the Nyandeni sub-district for the investigation because of the vast geographical distances between the primary health care facilities, and the inaccessible nature of the roads. A three-month reporting period from January 2009 to March 2009 was selected for the data audit.

The audit involved interviewing two or three clinicians (mostly nurses) per facility in a structured interview session. Each interview session lasted between 30 to 45 minutes, all of which were recorded using a digital recorder.

### 2.1 Data Quality

The investigation at the provincial office level focused on how the provincial office measured data quality across a number of data quality attributes. The most prominent were data validity, reliability, timeliness, integrity and precision. To test validity or data accuracy the study compared figures reported for a particular indicator, with the ones recorded in the source tools for example clinical registers. To test reliability, the consistency of the source tools used to collect particular data element indicators were compared, as well as the consistency of the data element definitions over the three-month period. The clinical data element definitions were also compared to the official data element definitions provided by the provincial office. Timeliness was measured according to the lateness of the reports submitted to the sub-district office for aggregation. Integrity checks were measured according to the availability of reported data; any reported figures where the source could not be found were considered to have compromised integrity.

## 3. Results

### 3.1 Data Collection Tools

An outstanding feature found in the study was the number of data collection tools available at the facilities. The growing demand for accountability seems to have led to an increasing number of clinical registers at the primary health care facilities. The study identified 17 patient collection tools. Thirteen (13) of these source tools originate from the Department of Health, while others were ordinary notebooks used by all health facilities surveyed to supplement the ones from the Department. The high number of source tools is consistent with an earlier verbal communication with an information manager from another district who claimed that a survey had determined up to 25 clinical registers elsewhere. The design of these data source tools does not make it easy for clinicians to complete these registers, which are highly structured. The Voluntary Counselling and Testing (VCT) register for example has 30 columns, it accommodates 20 patients per page, the Pre-Anti Retroviral Therapy register has 32 columns and can accommodate 30 patients while the Anti Retroviral Therapy (ART) and Human Immunodeficiency virus/Acquired Immune Deficiency Syndrome /Sexually Transmitted Infection/ Tuberculosis (HAST) registers has 40 columns. Each column in a clinical register corresponds to a patient attribute like patient name, gender etc. Some patients recorded in the Pre-ART register require all forty (40) columns to be completed. In cases where extra columns are required for a patient, the patient information would either be written onto a new page or in some cases a new register all together. However sometimes clinicians would draw extra columns in the registers to cater for new columns.

### 3.2 *Quality of Data*

Comparing the data reported by clinics with the data counted from source documents revealed a lot of inaccuracies that led to the conclusion that clinic data lacked validity. It was noted that not a single facility possessed data element definitions to guide data collection methods. As a result the reasons used to calculate data elements changed from month to month depending on which clinician was doing the report write up. Inconsistency of the data also extended to the use of source tools to report services rendered. Some clinics would report certain services in one register while the next clinic would report the same service in a different register. In summary collated data lacked validity, reliability, precision and there was no evidence clinics were using their data for strategic decision-making. In essence, data quality was very poor.

These results are similar to a survey of 962 health facilities, which showed that approximately 34% of healthcare facilities provided discrepant or inconsistent data. Other data quality issues found in that study saw information provided by health facilities was derived from monthly statistics or by estimation, as some facilities did not keep records. However, the idea of quality data is not necessarily zero defects, quality is conformance to valid requirements. In defining quality it must be determined who sets the requirement, how the requirements are set and the degree of conformance that is needed.

### 3.3 *Data Reporting*

There was no evidence to suggest that clinics conducted formal meetings solely dedicated to data management issues, whether it was to provide feedback or iron out data management issues. Matters concerning data issues were usually incorporated into other meetings. Staff in most facilities claimed that they received feedback from the sub-district office, but only four of the twelve clinics could provide evidence of the feedback they received. When asked about what measures could be taken to improve their data management problems eight of the facilities pointed at the problems of staff shortages. This might sound strange but if one takes into account the fact that not a single clinic had the full complement of nursing staff available as stipulated for an 8 hour clinic, plus the many clinical registers that need to be completed, then this sentiment is understandable.

### 3.4 *Data Use*

Disease trends in the form of graphs displayed on the walls of clinics is one of the ways in which the provincial information unit assesses the level of information use in health facilities. Only three (3) of the twelve (12) health facilities visited, had graphs displayed on the walls. A non-governmental organisation group working at the clinics drew all the graphs displayed on the walls. In all the clinics the same indicators were plotted and not a single one of the graphs had been updated to include the latest data at the time of the visit. All the graphs plotted were at least three months behind schedule. Using the provincial office's tool for assessing data use at the primary health care facilities, it was concluded that all clinics were at the first level of data use. This is the lowest score on the data use score sheet. An earlier discussion with the provincial information office revealed that only two healthcare care facilities had been rated top information users in a larger survey conducted by the provincial office.

### 3.5 *Data Quality*

Medical registers can serve many purposes for example as a tool to monitor and improve quality of care or as a resource for epidemiological research. For example, the Voluntary Counselling and Testing register provides insight into effectiveness and efficiency of Voluntary Counselling and Testing services. A closer scrutiny of the registers revealed poor quality of recording all the registers. In certain health facilities, no recordings were found in the Pre-Anti Retroviral Therapy, Anti Retroviral Therapy, and HAST registers even though these services were provided. In all the facilities, these three registers were incomplete, in two (2) of the facilities the pre-ART registers were not filled at all, a further two (2) facilities did not have any information recorded in the HAST registers. Seven (7) of the health facilities were using an older version of the ART register.

In all the health facilities surveyed, not one could verify all the reported figures. There were differences in figures reported in the monthly summary reports and those counted in the clinical registers. None of the health facilities kept definitions of the data elements they collected or data trails of reported figures. In all cases the nurses responsible for data collection, were unable to provide an explanation for the discrepancy in reported figures. In three health facilities, the clinicians interviewed were unsure of the source of the reported figures.

### 3.6 *Precision*

The tally sheets where monthly clinical statistics are summarised, contain two (2) portions for the statistic compiler and data verifier to sign off. Only six (6) out of the twelve (12) health facilities had a verifier sign off the reported data for the whole three months duration. Clinicians blamed non-verification of reports on lack of

clarity as to who was responsible for the verification of reports. Some clinicians believed it was the duty of the clinic supervisor while others insisted it was the role of the operational manager. In all the sites where reports were verified it was the operational manager who signed off the reports.

### *3.7 Timeliness*

All clinicians interviewed knew exactly when reports had to be in the sub-district office. Due to the remote location of some of the health facilities and poor accessibility, most delays in reporting statistics were blamed on the unavailability of transportation. Despite the poor accessibility, health facilities showed a lot of innovation when it came to delivering statistics to the relevant offices. Some clinicians relied on taxi drivers to deliver statistics, while others gave clinical statistics to courier vehicles transporting laboratory specimens. In relatively more accessible areas, the clinic supervisor came to pick up the statistics.

When it comes to *entering* data into clinical registers or source tools five (5) of the clinics reported that exclusively nursing staff completed the clinical registers. In the rest of the clinics, recording into clinical registers was shared between nurses and lay counsellors. The lay counsellors were responsible for recording non medical data like demographic information, while nurses concentrated on medical data like symptoms, medication etc. In all facilities the completion of registers was not done in real-time because each clinic had only one type of each register that had to be shared amongst the different health workers. This implied that a nurse attending to a patient had to wait for her colleague to finish using a clinic register before recording into it. Not a single clinic had a document, which stated the clinic staff who were responsible for a particular clinical register.

### *3.8 Completeness*

Only two (2) of the clinics surveyed did not have summary reports for a whole month. The rest of the clinics consistently reported on all the services they rendered. The only problem encountered was the meaning of zero counts reported in all the clinics. Since the narrative portion of the tally sheet is too small to write much, nothing was ever recorded under this section in all facilities.

### *3.9 Reliability*

Reliability issues centred on source documents used to collate reported data. For instance, for the data element first antenatal attendees some of the clinics used data from the Voluntary Counselling and Testing (VCT) register, whilst others used the Prevention of Mother to Child Transmission (PMTCT) milk registers for the same data element. Similarly some clinics relied on VCT registers to count CD4 tests conducted, whilst others used Pre-ART and ART registers. Other clinics used only their specimen notebooks. Not a single clinic kept a report explaining the data element definitions or the source tools from which each reported data element is recorded. As a result data element source records for collated figures differed from clinic to clinic, even from month to month within the same clinic. Not a single clinic kept a data trail of reported figures. This meant that during the data audit sessions all clinics battled to recount reported figures. All the registers are full of empty fields, for instance the VCT, Pre-ART and ART registers have a field for CD4 results. Some of the clinics are situated in very remote locations making it very difficult to transport blood and receive blood results from the laboratory. As a result, blood results like CD4 counts are usually left blank in clinical registers. Other column headings that were usually left empty were the WHO staging results and drug adherence records. There was a lot of inconsistency in the data recorded in the registers. Five health facilities used the pre-ART register to record details of clients with CD4 results of below 200, while the rest recorded details of patients who tested HIV positive. Similarly in all the registers, fields titled CD4, Cotrimoxazole and Tuberculosis were inconsistently filled, for example under the column for CD4 test results, some clinicians wrote the actual CD4 result figures, others wrote the date when blood was drawn for the CD4 result, whilst others simply put a tick to show blood was taken for this purpose.

### *3.10 Role of the Clinicians*

In addition to collecting clinical data on all services provided by a facility, clinicians are required to report on issues such as infrastructure status and human resources availability. Discussions with the provincial information office revealed that in addition to clinical disease management, information collected from the care process is used for secondary purposes such as: administration, financial management, resource allocation and research. All this information originates from data health workers, namely nurses, are required to collect on a routine basis. This arrangement whereby all health data are collected by health workers is questioned by Berg and Goorman (1999) when they asked whose responsibility it was to do the additional work of data collection, collation, reporting for secondary use and where do the benefits end up? Berg and Goorman emphasise that the task of producing data for secondary use by others, other than the primary care givers, is unfairly delegated to the primary care giver. When the goal is to support secondary utilisation of data outside the context of the care

process itself, this additional burden on the actual care process is highly problematic. It might even be considered unacceptable given time constraints and the fact that this additional task will take clinicians from their primary responsibility in other words caring for the patient. This does not mean that collection of health data for secondary use is unacceptable; it only means that the collection and use of information should not impose a burden on the individuals collecting it. Moreover, the data being collected should add value to the individuals collecting it. Since clinicians are not using the data they are collecting it can be assumed that the data they are collecting is of no value to them.

### *3.11 Data Management Structure*

A look at Abate et al's (1998) definition of quality data says that data are of the required quality if they satisfy the requirements stated in a particular specification and the specification reflects the implied needs of the user. The significant point about this definition is that it highlights the viewpoint of the user. In the case of the health information system in the province, it is the identification of the data "user" that raises questions. The health information system in the province is structured in such a way that data quality management is vertically aligned with data moving from facility level to sub-district, district then provincial offices. This vertical alignment is such that data analysis and data use is mainly done at the sub-district, district, and provincial and national offices reducing clinicians to mere data collectors. Although the data management structure is vertically aligned there is little evidence at the health facilities to suggest that an opposite flow of information from the province to health facilities exists. According to the United Nations data quality management should be horizontal in nature, with the sharing of data an important step in attaining quality data. In other words for data quality management to be effective health information systems need to be able to share their data effectively. The situation at the health facilities is such that data sharing or horizontal movement of data does not exist between health facilities, as well as between clinicians working in the same facility.

To be able to facilitate horizontal movement of data, information concepts need to conform to a recognised standard. In the case of the Eastern Cape Province, reported data are based on a national dataset. These standards are supposed to enable data sharing across the health system so that data collected in one facility means the same in another facility. As seen in the development of the Irish Minimum Dataset, an important step in the development of such standards lies in the development of definitions and protocols to guide the collection of the data. This requires collaborative agreements between the various stakeholders in the health system with nursing representation essential. However, this study conducted in the Nyandeni sub-district supports the view that facilities' needs are not properly addressed in terms of how, when and what data are collected. This is unfortunate as medical information is entangled with its context of production in that the meaning, hardness and significance of a piece of information cannot be detached from the specific purpose that structured the gathering of the information (Berg & Goorman, 1999). While Paley (1996) reinforces the importance of including nurses in health information system development by arguing that the language used by nurses includes terms that have colloquial meaning, this can be problematic as key terms may be ambiguous or open to interpretations. The lack of data element definition or protocols on how to gather data at the sites shows reliance on external help to develop data management solutions.

### *3.12 Source Tools*

According to O'Nuska III, (1996) there is no more important document than the instrument that is used to acquire the data from a clinical trial with the exception of the protocol, which specifies the conduct of clinicians using the tools. O'Nuska III (1996) further argues that the quality of the data collected relies first and foremost on the quality of the instrument used to collect it and no matter how much time and effort goes into providing clinical services, if the correct data is not collected, a meaningful analysis may not be possible.

Three potential problems were pointed out with the data source tools. Interviews with the clinicians indicated that the development of clinical registers used at the health facilities emanated from outside the primary care setting. This claim tallies with the ones by other authors who claim that clinical registers were primarily created for easy extraction of data elements. In other words, clinical registers were designed to meet the needs of the information officers at government institutions and not necessarily the clinicians. This would probably explain the exclusion of clinicians from the design process. The development of data collection tools seems to be aligned to government initiatives to improve health. For example, the voluntary counselling and testing (VCT) programme has the voluntary counselling and testing register, the tuberculosis (TB) programme has the TB suspect and TB confirm registers, the prevention of mother to child programme (PMTCT) has the PMTCT and ANC register, while the anti retroviral therapy (ART) programme has the pre-ART and ART registers. This implies that the development of clinical registers is linked to government initiatives *for* monitored health

programmes. Along with these data source tools there are clinical indicators linked to health programmes. The source tools create an impression that these health statuses are mutually exclusive and that the treatment of care is a vertical process. This results in too many data collection tools for clinicians to record. Reports have been heard of clinicians having to take registers home to record because they simply do not have the time during normal working hours. Looking at the number of clinical staff available to complete the registers, there are too many registers with lots of data duplications and too few clinicians to fill them, which leads to poor recording and eventually poor data quality.

#### 4. Discussion

To improve data quality it is important firstly to define what is meant by “quality” and then establish methods of measuring that quality. *Data quality* is not a single attribute; it can be measured on many dimensions and is often perceived differently by different customers e.g. timeliness may be the most important factor for one data consumer while completeness may be important to another. For this reason it is important for the Department of Health to involve the clinicians, when defining data quality. The importance of quality data is illustrated by McGlynn’s belief that clinicians can use data to improve daily care practice (McGlynn et al., 1998). McGlynn goes on to explain that data quality measurement and improvement of care are intertwined and that it is impossible to make improvements in clinical care without measuring the quality of care. Measurement in turn depends on the availability of quality data. So to improve the quality of data, clinicians need to be involved not only in data collection but also in the measurement of the quality of care that they provide (McGlynn et al., 1998). This not only serves to improve the quality of service provided but also helps clinicians appreciate the value of their work and enhances the importance of collecting quality clinical data. Clinicians, as data collectors, are the best placed individuals to recommend a course of action based on data they collect and are also the best placed individuals to suggest whether better ways to measure results exist.

However the quality of the clinical data at the health facilities was of a poor standard. There was no evidence to suggest that data collected for secondary purposes was being used for decision making at the primary healthcare facilities.

#### 5. Discussion: Improving the Quality of Clinical Data

There is belief that it is possible that South Africa can “leap frog” directly from poorly functioning paper-based health information systems to highly sophisticated and fully integrated country-wide network solutions based on e.g. telemedicine, smart cards, electronic health records etc. The use of electronic data sources common to all healthcare settings has the potential to streamline data gathering and improve public health reporting. Given the complexities of the development of health IT projects many health IT projects fail (Heeks et al., 1999). In most instances these failures occur because of insufficient understanding of the needs of health care workers. These electronic solutions certainly have their advantages over paper based information systems. However, the introduction of an electronic health system would only be providing a computerised solution to a non-computerised problem and the result of poor data quality would still exist.

#### References

- Abate, M., Diegert, K., & Allen, H. (1998). A hierarchical approach to improving data quality. *Data Quality Journal*, 4, 365-369. Retrieved from <http://www.dataquality.com/998abate.htm>
- Berg, M., & Goorman, E. (1999). The contextual nature of medical information. *International Journal of Medical Informatics*, 56, 51-60.
- Braa, J., & Hedberg, C. (2000). *Developing district-based health care information systems: the South African experience*. In, 2000. Citeseer.
- Coiera, E., & Clarke, R. (2004). e-Consent: The design and implementation of consumer consent mechanisms in an electronic environment. *Journal of the American Medical Informatics Association*, 11, 129. <http://dx.doi.org/10.1197/jamia.M1480>
- Hearnshaw, H., Harker, R., Cheater, F., Baker, R., & Grimshaw, G. (2001). Expert consensus on the desirable characteristics of review criteria for improvement of health care quality. *British Medical Journal*, 10, 173. <http://dx.doi.org/10.1136/qhc.0100173>
- Heeks, R., Mundy, D., & Salazar, A. (1999). *Why health care information systems succeed or fail*.
- Henry, P., Mac Neela, P., Clinton, G., Scott, A., Treacy, P., Butler, M., ... Byrne, A. (2006). Developing a data dictionary for the Irish nursing minimum dataset. *Studies in health technology and informatics*, 122, 510.
- Mainz, J. (2003a). Defining and classifying clinical indicators for quality improvement. *International Journal for*

*Quality in Health Care*, 15, 523.

- Mainz, J. (2003b). Developing evidence-based clinical indicators: a state of the art methods primer. *International Journal for Quality in Health Care*, 15, 5.
- Martin, M. (2005). Measuring and Improving Data Quality. Part II: Measuring Data Quality. NAHSS Outlook.
- Mcglynn, E. A., Asch, M., & Steven, M. (1998). Developing a clinical performance measure. *American journal of preventive medicine*, 14, 14-21.
- O'Nuska III, M. J. (1996). Data Management Best Practices;cited in Data Management Handout. Retrieved from <http://casaa.unm.edu/download/Data%20Management%20Handout.pdf>.
- Paley, J. (1996). How not to clarify concepts in nursing. *Journal of Advanced Nursing*, 24, 572-576. <http://dx.doi.org/10.1046/j.1365-2648.1996.22618.x>
- Ramkissoon, A., Kleinschmidt, I., Beksinska, M., Smit, J., Hlazo, J., & Mabude, Z. (2004). STI Indicator Information from the National Baseline Assessment of STI and HIV services in SA public sector health facilities. Published by the Health Systems Trust, 83.
- WHO. (2003). *Improving Data Quality: A Guide for Developing Countries*. WHO Western Pacific Regional Publication.