Reporting Practices and Data Quality in Health Information Systems in Developing Countries: An Exploratory Case Study in Kenya

Ayub Manya[1][a][1] and Petter Nielsen[1]

[1] University of Oslo

Abstract. Despite increased attention paid to health information systems and their key role for improving health systems in in Low and/or Middle income countries (LMICs), it is believed that data from majority of the health information systems contribute little to the decision-making processes due to poor data quality. We carried out an exploratory assessment of the health information system in Kenya with the main objective of determining the quality of data in terms of accuracy, timeliness and completeness. The study also considered the reasons for the observed data quality status. Data quality audits were carried out in selected health facilities. Data from the source documents at health facilities were compared to the data in the national health information systems for the same period. Key informant interviews were conducted and focus group discussions conducted during quarterly review meetings at regional levels. The study revealed that the completeness rate for the monthly reports was 86.9 percent while the timeliness of the reports was 78.7 percent. In terms of accuracy of the reports, the study showed that while there was a significant amount of low accuracy in many reports evaluated, there was a surprisingly high accuracy of reports coming from the maternity units of all health facilities visited. The accuracy of the number of deliveries could be associated with the financial incentives provided by the government to health facilities as part of the country's free maternity care project in which mothers deliver free of charge in health facilities. While most health information systems are plagued with poor data quality, a simple and practical incentive can improve accuracy, timeliness and completeness.

Keywords: health information systems, data quality assessments, incentives, data quality, accuracy of data, timeliness of data, completeness of data.

---

[a] Corresponding Author: Ayub Manya, ayubmanya@gmail.com

1. INTRODUCTION

In recent years we have seen an increasing focus on health information systems globally, with significant human and financial resources being invested at various levels to improve them (Arts, De Keizer, & Scheffer, 2002; Kihuba et al., 2014). Adoption of modern Information and Communication Technology (ICT) has greatly improved access to health information systems (Kimaro & Nhampossa, 2007). Access to timely and high quality data is very important for strengthening health systems (WHO, 2003, 2007). Such data are vital for facilitating long term planning, policy formulation and day-to-day resource allocation within the health sector (AbouZahr & Boerma, 2005). In order to run an efficient health information system, data are derived from several sources including census, surveys, service delivery reports, registration of births and deaths, (Ledikwe et al., 2014). Data from service delivery are gathered by health workers, particularly nurses, clinicians and laboratory staff, who document their service delivery work. Regularly, after the data are collected and entered in paper registers, the providers fill out summary forms and send them up to the next reporting level. Research has shown that data often get lost or become erroneous during these processes of data collection, collation and transfer; leading to mistrust in routine health information systems. It is argued that unless there are adequate mechanisms to provide high quality of data as well as for developing a culture for using data to inform decisions, health information systems have no value (Tara Nutley, 2006).

Despite the increased attention paid to health information systems in developing countries, some researchers argue that data from most of these systems contribute less to decision-making processes due to poor quality (Arts et al., 2002; Braa, Heywood, & Sahay, 2012; Kihuba et al., 2014; Sahay, 2001). When health managers are asked how many women delivered in health facilities or how many maternal deaths occurred at health facilities, they must answer them with facts, rather than beliefs and opinions. At the same time, data are too often used uncritically from health information systems without checking the data for any errors; often resulting in misleading interpretations, unwise decisions and increased costs (Chapman, 2005). Of the three challenges identified above (poor data quality, non-use of data and non-critical use of data), this paper focuses mainly on data quality.

Determining the quality of data is complex. This is because data quality encompasses multiple dimensions. Juran and Godfrey(1999) posit that data quality is determined by how fit the data are for specific uses (Juran & Godfrey, 1999). Other researchers view data quality as "the measure of the agreement between the data views presented by an information system and that same data in the real-world" (Orr, 1998) p.3. Other data quality dimensions include accuracy, completeness (having all relevant information recorded), timely data, consistency and relevance (Ballou & Pazer, 1985; Herzog, Scheuren, & Winkler, 2007; Ledikwe et al., 2014). This paper focuses on accuracy, completeness and timeliness. In order to assess data quality, there is a need to compare the data in question with the equivalents from the real-world in order to correct any deviations.

Even though most stakeholders keep raising the issue of poor data quality in health information systems, our starting point is that not all data are of equally poor quality. Thus, we should explore and learn from data of good quality, and leverage upon this to improve the information systems. This paper is an attempt to answer the following research questions: How can we leverage on high quality data and the related data collection and processing best practices to enhance health information systems? To answer this question, we evaluated the data in the Kenyan health

information system; with the main objective of determining the quality of data in terms of accuracy, timeliness and completeness. The study also considered the reasons for the observed data quality status.

The specific objectives included:

1. To determine the accuracy, timeliness and completeness of data for selected data items reported at selected health facilities, compared against what was reported for the same facilities in the national health information system for a particular period.

2. To establish possible reasons for observed differences in data quality and identify the good practices behind the high quality data that can be used for strengthening health information systems.

2. REVIEW OF COMMON DATA QUALITY PROBLEMS

There are many possible problems that undermine the quality of data in health information systems. Some of these problems relate to the set-up of the organization and the technical knowledge of those who work with the data. In a study conducted in Botswana, it was found that data quality was not only affected by the limited capacity to collect, analyse, and interpret but also by the lack of skills to ensure that data being collected are of good quality (Ledikwe et al., 2014). Data quality assessments conducted in Kenya in 2010 and 2014 consistently found many discrepancies between data in primary tools and those in the national system. The assessments outlined the following reasons for the discrepancies: failure to use registers as per instructions, failure for health workers to understand the indicators, the use of multiple tools to aggregate the data and lack of data collection tools (GOK, 2010b, 2014). Researchers have also observed that there is a relationship between data quality and use (Braa & Sahay, 2012; Orr, 1998). Braa (2012) further hypothesised that, "poor data quality will not be used, and because they are not used, the data will remain of poor quality" (Braa et al., 2012) p.379. It has also been observed that when data remain in the system for too long without use, they lose their quality (Orr, 1998). The practice of collecting large amounts of data that are not of immediate use is therefore not recommended since "it is clear that if an organization is not using data, then, over time, real-world changes will be ignored and the quality of the data in the system will decline" (Orr, 1998, p. 7).

2.1. Data Quality in Health information system of Kenya

In 2010, Kenya adopted the use of District Health information software(DHIS2), which is used countrywide (Karuri, Waiganjo, Orwa, & Manya, 2014; Manya et al., 2012). One critical program in Kenya that requires accurate data is the free maternal care project. Faced with the problems of women dying during delivery (maternal mortality ratio 362 deaths per 100,000 live births in 2014(Obonyo, 2010)), the government of Kenya decided to provide free maternity care, starting from 2013. In this program, pregnant women are allowed to deliver in any government or faith based health facility, free of charge. The accrued costs to health facilities are then reimbursed by the government according to agreed rates. The country's Health information system is used as a means of verification for these reimbursements.

Kenya adopted a decentralized system of governance (GOK, 2010a) through a constitutional change In 2010. The new constitution abolished the eight administrative provinces and instead created 47 counties (Kanyinga & Long, 2012; Nyanjom, 2011). The change was part of a negotiated agreement to stop ethnic based chaos that had erupted, following the disputed presidential election of 2007 (Kanyinga & Long, 2012). The main rationale of adopting devolution was to increase autonomy and political power to the lower levels and thereby diffuse ethnic and regional tensions (Sihanya, 2012). The counties were semi-autonomous or federated with powers to elect their own political leaders. Each administrative County was further subdivided into sub-counties, which in most cases became the parliamentary constituencies. Apart from powers to collect revenue in their area of jurisdiction, the counties were allocated formulae driven grants from the national government for recurrent and development activities, including paying workers' salaries. This was a new reality for the health sector, affecting most aspects of governance. Devolution has the potential to promote data demand and use for planning at the local levels. Local use of data represents a change to the practice that data are seen only as for onward transmission to higher levels and hence leads to better data acquisition, storage, and sharing.

Based on data from their paper registers, health facilities summarise data into predefined summary reporting forms and send to sub-counties (formerly districts), for entry into the electronic system (DHIS2). In terms of governance, the previously strong leadership at the Ministry of Health headquarters has been diluted. The county management has taken a central role in utilizing the data for implementation of various programmes. Authorized officers from the central government are able to access data for their reporting needs from DHIS2 since it is a web based system. Other functions formerly conducted by the central government like preparation and printing of paper reporting tools have been assigned to the county. Based on this concept of local data use, we argue that devolution has the potential to drastically enhance health systems.

3. METHODOLOGY

We conducted a longitudinal study in selected administrative counties between April 2015 and December 2015 using a case study methodology. The Assessment team comprised representatives of the local health administrative office and the research assistants. Team composition included officers with mixed skills including public health (to support in the disease area and indicators), program officers (e.g., health records and information officers) and relevant regional administrators. Quantitative data were collected from data collecting tools health unit and from the online database (DHIS2). Qualitative data were obtained through meetings, participant observations and key informant interviews to contextualize and support our interpretation of the quantitative. Field visits by the first author to health facilities further provided first-hand information on the processes of data management through participant observation. Findings from the field visits were presented during quarterly review meetings attended by stakeholders including senior Ministry of Health officials in the county and development partners. Discussions during the meetings provided a more in-depth understanding of data management. Thereafter the county came up with data quality improvement work plans to be implemented in the following three months.

3.1. Study Site

Following consultations with the authorities, four counties were selected for the study, based on geographical proximity (Busia, Kisumu, Siaya and Uasin Gishu). Similarly, the assessment team in consultation with sub county managers selected health facility sites using purposive sampling methodology with bias towards health facility type (County Referral Hospital/ County Hospital, Health Centre (high volume), Dispensary). Table 1 shows the targeted participants for the study.

Table 1: Target Study Participants at Each Selected Study Site

Level	Study Participants
County	County Director of Health, County Health records and information officers (CHRIOs), Directors of the county referral hospitals
Health facility	Facility manager, data clerks, nurses and clinicians in various departments

The focus was on the tools and procedures used related to health data on the county and the health facility level as summarised in table 2 below.

Table 2: Levels, tools and procedures

Level of health system	Tools	Procedures
County	DHIS2 Summary forms	Enter data from summary forms into DHIS2 Use data from facilities for local decision making
Facility	Paper registers Summary forms	Enter data in paper registers Summarise data from registers into summary forms Transfer summary forms to next level for computerization

3.2. Selection of Data Items for Verification

Three data items were selected for verification, representing data on vaccine preventable diseases (immunization), reproductive health and outpatient curative data. Specifically the following were chosen: a) Fully immunized children under one year of age b) Pregnant women who delivered in health facilities and c) The number of children under five years who were treated for malaria in the outpatient department. In coming up with the list, the following criteria were applied:

• Information needs in terms of data that have been seen to attract interest from the managers e.g. hospital deliveries in view of the free maternity program.

• Information gaps and issues e.g. misreporting, omissions, under reporting etc.

• The impact of interventions and the importance of selected data in service delivery

Availability of research resources also served to limit the number of items to be assessed. Table 3 below shows the tools from which the selected data were sourced.

Table 3: The Source of Data for the Selected Indicators

Name of indicator	Name of register	Name of summary reporting form
The number of children under one year who were fully immunized	Immunisation Register (MOH 510)	Immunization summary (MOH 710)
The number of pregnant women who delivered in health facilities	Maternity (MAT) Register (MOH 333)	MOH 711 Integrated RH, MCH, Social Work & Rehab Summary (MOH 711)
The number of children under five years who were treated for malaria in the outpatient department.	Outpatient Register: Under 5 years (MOH 204A)	Outpatient summary under 5years(MOH 705A)

3.3. Assessment of Data Accuracy: Data Quality Assessments

To assess data accuracy three main procedures were followed. First, paper registers at each health facility were evaluated in order to identify common problems related to data generation. Second, the study investigated the accuracy of transferring data from the registers to summary reporting forms. For this study the registers and summary reporting forms were considered to be source documents since they were the primary data entry documents. Accuracy of data transfer from the source documents to DHIS2 was then investigated using a tool customised from world Health Organization; the Routine Data Quality Assessment (DQA) tool (WHO, 2011). It was structured in a way that when the responses were keyed-in, data accuracy was automatically calculated. In this study, accuracy was presented as the ratio of transmitted data items compared to the data in the source document. However, the study team did not assess the accuracy of the clinical data as written by the clinicians.

3.4. Assessment of Completeness and Timeliness of Reports

Data completeness was investigated using two criteria. First, the paper tools i.e. registers and summary reporting forms were inspected to see if all parts were completely filled. The second method of assessing data completeness was done using the DHIS2. The DHIS2 has an inbuilt reporting rate functionality which provides the completeness and timeliness of reports based on the number of facilities expected to report. Timely reports were defined as those that were in the DHIS2 by 15th of the month and were presented as percentages. In addition, the team assessed the general aspects of the health information system in terms of personnel, availability of appropriate data collection and reporting forms/tools.

3.5. Process of Data Analysis

Data analysis was both quantitative and qualitative. Quantitative analysis was particularly applied during Data Quality Assessments while qualitative analysis was mainly for the views gained from quarterly review meetings and participant observation. Analysis for the qualitative data began by summarizing the notes from the meetings and the notes from the field.

Analysis for the Data Quality Assessments was done by comparing data from registers with data in duplicate copy of summary reporting forms and in the DHIS2. Verification Ratios for the indicators were calculated by dividing the recounted figures from the registers with the figures in the summary reporting forms. The same was calculated for the data in summary reporting forms compared to those in the DHIS2. These quantitative data (counts, percentages) were entered in a computer database using Microsoft Excel. Results were compared between indicators, showing which indicators had high data quality.

4. RESULTS

A total of 63 health facilities in 23 sub counties and 4 counties (Busia, Kisumu, Siaya and Uasin Gishu) were visited. Nearly all the facilities (95 percent) in the study were managed by Ministry of Health while only 3 were run by Faith Based Organizations (FBOs). Privately owned facilities were not represented. In terms of type of facilities, hospitals comprised 40 percent as shown in table 4 below.

Table 4: Type of Health Facilities Visited During the Study

Type of health facility visited	Number of facilities	% of facilities
Dispensary	19	30 %
Health Centre	19	30 %
Hospital	25	40 %
Total	63	100 %

4.1. Results on Completeness and Timeliness of Reports

Reviewing the paper registers, we noted that most of them were large with many columns which were not being filled. This was in particular the case if a certain service was not being offered in the facility. Due to many blank columns, some health workers decided to record other findings in the empty spaces, against the guidelines given. This lack of standardization meant that a different person would have problems understanding the data in the registers. And it made it difficult to count and summarize data unless the people who maintain the registers were available. Despite this, all the services that were conducted had data in the registers. Completeness of reports was also retrieved from DHIS2. This was calculated by the system as the proportion of facilities that actually sent their reports to the DHIS2 compared to the total facilities expected to report. The

online system (DHIS) revealed that the average completeness rate for the study group was 86.9 percent.

Timeliness of the reports was determined by the number of reports that were entered in the DHIS2 before the agreed deadline of 15th of day of the Month. On average 78.7 percent of the facilities evaluated submitted their reports on time as shown in figure 1.
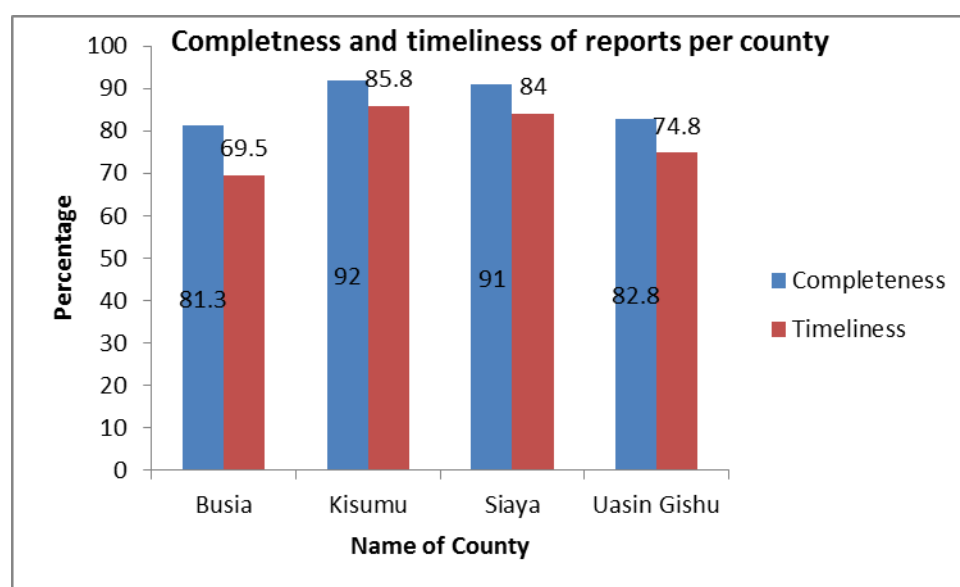

Figure 1: Completeness and Timeliness of Reporting for Selected Counties in 2015 – source: https://hiskenya.org

4.2. Accuracy of Reports

In an ideal situation where the parameters in the recorded data were found to be the same during the verification, the data accuracy value would be 100 percent. If the value of the verified data exceeded 100 percent there this was interpreted as over-reporting whereas less than 100 percent was interpreted as under-reporting.

Data verification for the different indicators revealed that there were discrepancies between the paper forms and DHIS2, albeit to varying degrees. There was over-reporting for children under five being treated for malaria (149 percent), and fully immunized children (119 percent). The number of deliveries was however nearly accurate (101 percent). This means that the data for maternal deliveries was the most accurate while that for a fully immunized child was the least accurate as shown in figure 2.
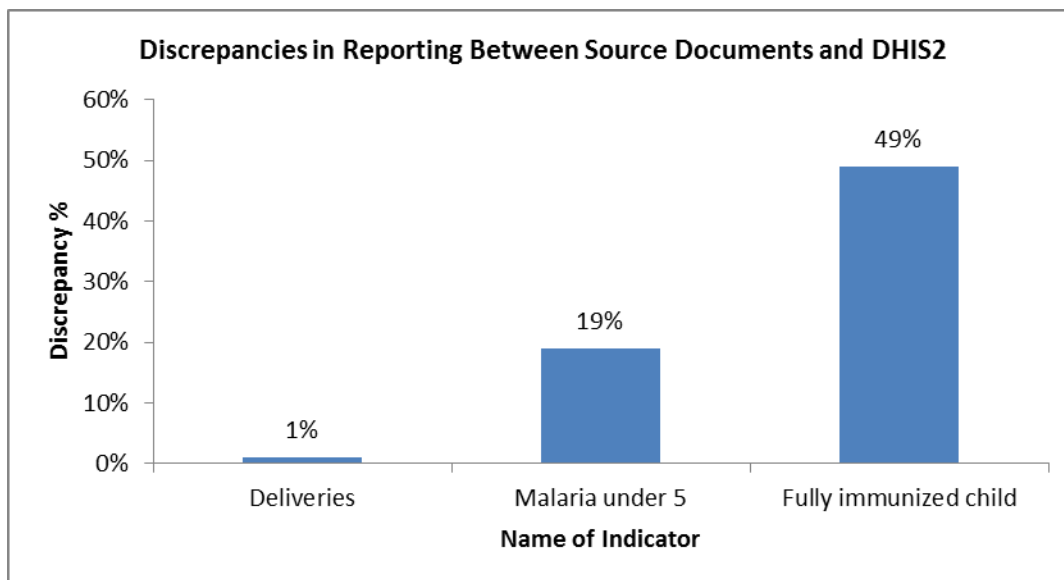
Figure 2: Discrepancies in Reporting between Source documents and the electronic database, DHIS2

4.3 Comparison of Accuracy of Data Transfer

The accuracy of transfer of was not equal. Defining 100 percent as the desired result for the compared elements figure 3 shows that transfer of data from summary tools to DHIS2 is more accurate than transfer of data from registers to summary tools.
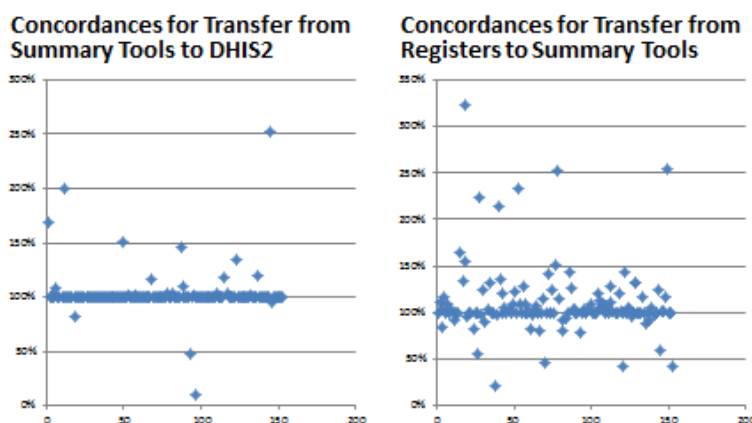

Figure 3: Comparison of Accuracy of Data Transfer between Register to Summary Tools and Summary Tools to DHIS2

5. DISCUSSION AND IMPLICATIONS

Health managers require reliable data for planning purposes. It has been shown that when data are of bad quality, their demand also drops, thereby affecting program effectiveness (Braa et al., 2012; Foreit, Moreland, & LaFond, 2006; Mavimbe, Braa, & Bjune, 2005).

The findings in this paper reflect common problems challenging health information systems. While this paper highlights poor data quality, it also shows a surprisingly high accuracy of reports coming from the maternity units of all health facilities visited. The accuracy of the numbers of deliveries was attributed to the free maternity care program. This program requires accurate data to enable reimbursement of the costs of deliveries by the national government. While most of the money is used for budgetary support to the facility, some of it is directly ploughed back to the maternity ward to improve the general working environment like staff tea and provision of vital equipment like gloves. The number of deliveries then becomes important to everyone in the health facility, the county and at the national level. For all levels, the reimbursements act as incentives to improve data quality. Due to this incentive, we expect health workers to establish their own ways of getting each day's work done and give accurate reports. It was also found that unlike in other departments, the maternity registers and summary reporting forms are managed by the same staffs who conduct the deliveries. This reduces the number of errors as would have likely happened if external clerks managed the data. These practices become intricately tied to other routinized activities, becoming part of the valued practices of the health facility. This is an example where incentives can give rise to better practices, something that should be leveraged upon to improve the health system.

From participant observation in the field, we found that counting fully immunized children from the registers and tally sheets were problematic. While data from all immunizations were recorded in registers, the records for fully immunized children were captured by tally sheets. Some health workers even forgot to tally and some tally sheets were torn. There was also confusion in counting malaria cases as some clinicians did not separate confirmed malaria and suspected malaria. So those collecting data just counted any malaria cases yet the indicator was for confirmed cases only. We further observed that the loss of quality when transferring data from registers to summary forms was more challenging compared to when transferring from the summary forms to the DHIS2. This could be due to the fact that the definitions of indicators are not clear and the way registers are maintained is not uniform. Counting from register may have peculiar problems especially since the people counting may not understand the work. In this case, the people who managed the registers were never trained while those doing data entry into the DHIS2 had undergone substantial training. Besides training, entering data from summary tools to the DHIS2 is similar to doing a copy typist work, with minimal typing errors as shown by the relatively good accuracy in this activity. Thus, the deployment of DHIS2 and related capacity building should be seen as supporting the improvement of data quality.

The decentralisation of Kenya had some effects on the observed data quality. For instance, reports were used at the county level for immediate planning. Data managers at this level were working with data not just for sending to another level but for local purposes. The findings in this paper also suggest that financial incentives may have positive effects on data quality. This should encourage more research on data quality related to mechanisms such as Results-Based Financing and their effects on data quality. Such studies should also be critical to the potential of adverse effects of incentives such as over-reporting.

The research design in this study, based on comparing different data sources and triangulation of qualitative and quantitative data, was very appropriate. This approach is useful to reveal and highlight the quite common lack of data quality in the health information systems in developing countries. But more importantly, it can also be used to identify data of relatively high quality and further to scrutinise its underlying practices. An understanding of these practices can be used to inform actions towards improving data quality across the health system.

6. **CONCLUSION AND RECOMMENDATIONS:**

This study shows that while most health information systems are plagued with poor data quality, a simple and practical incentive can influence data accuracy, timeliness and completeness. It is recommended that similar research should be done on large scale with the aim of scaling up the good practices.

REFERENCE

AbouZahr, C., & Boerma, T. (2005). Health information systems: the foundations of public health. *Bulletin of the World Health Organization, 83*(8), 578-583.

Arts, D. G., De Keizer, N. F., & Scheffer, G.-J. (2002). Defining and improving data quality in medical registries: a literature review, case study, and generic framework. *Journal of the American Medical Informatics Association, 9*(6), 600-611.

Ballou, D. P., & Pazer, H. L. (1985). Modeling data and process quality in multi-input, multi-output information systems. *Management science, 31*(2), 150-162.

Braa, J., Heywood, A., & Sahay, S. (2012). Improving quality and use of data through data-use workshops: Zanzibar, United Republic of Tanzania. *Bulletin of the World Health Organization, 90*(5), 379-384.

Braa, J., & Sahay, S. (2012). *Integrated Health Information Architecture: Power to the Users: Design, Development and Use*.

Chapman, A. D. (2005). *Principles of data quality*: GBIF.

Foreit, K., Moreland, S., & LaFond, A. (2006). Data demand and information use in the health sector: conceptual framework. *Chapel Hill, NC: MEASURE Evaluation, Carolina Population Center*.

GOK. (2010a). The Constitution of Kenya, 2010. *Attorney General Nairobi*.

GOK. (2010b). *National Data Quality Assessment Report, Division of Health Information System, Ministry of Public Health & Sanitation and Ministry of Medical Services*.

GOK. (2014). *Data Quality Audit Report, August 2014, Nairobi, Kenya Division of Health Informatics Monitoring and Evaluation, Ministry of Health, AfyaInfo Project.*

Herzog, T. N., Scheuren, F. J., & Winkler, W. E. (2007). *Data quality and record linkage techniques*: Springer Science & Business Media.

Juran, J., & Godfrey, A. B. (1999). Quality handbook. *Republished McGraw-Hill*.

Kanyinga, K., & Long, J. D. (2012). The political economy of reforms in Kenya: the post-2007 election violence and a new constitution. *African Studies Review, 55*(01), 31-51.

Karuri, J., Waiganjo, P., Orwa, D., & Manya, A. (2014). DHIS2: The Tool to Improve Health Data Demand and Use in Kenya. *Journal of Health Informatics in Developing Countries, 8*(1).

Kihuba, E., Gathara, D., Mwinga, S., Mulaku, M., Kosgei, R., Mogoa, W., . . . English, M. (2014). Assessing the ability of health information systems in hospitals to support evidence-informed decisions in Kenya. *Global health action, 7*.

Kimaro, H., & Nhampossa, J. (2007). The challenges of sustainability of health information systems in developing countries: comparative case studies of Mozambique and Tanzania. *Journal of Health Informatics in Developing Countries, 1*(1).

Ledikwe, J. H., Grignon, J., Lebelonyane, R., Ludick, S., Matshediso, E., Sento, B. W., . . . Semo, B.-w. (2014). Improving the quality of health information: a qualitative assessment of data management and reporting systems in Botswana. *Health research policy and systems, 12*(1), 1.

Manya, A., Braa, J., Øverland, L. H., Titlestad, O. H., Mumo, J., & Nzioka, C. (2012). National roll out of District Health Information Software (DHIS 2) in Kenya, 2011–Central server and Cloud based infrastructure. *IST-Africa 2012 Conference Proceedings, 5*.

Mavimbe, J. C., Braa, J., & Bjune, G. (2005). Assessing immunization data quality from routine reports in Mozambique. *BMC Public Health, 5*(1), 108.

Nyanjom, O. (2011). *Devolution in Kenya's new constitution*: Society for International Development.

Obonyo, B. (2010). Maternal Health; Kenya National Bureau Of Statistics (KNBS) And ICF Macro.

Orr, K. (1998). Data quality and systems theory. *Communications of the ACM, 41*(2), 66-71.

Sahay, S. (2001). Special issue on "IT and health care in developing countries". *The Electronic Journal of Information Systems in Developing Countries, 5*.

Sihanya, B. (2012). Constitutional implementation in Kenya, 2010-2015: Challenges and prospects. *FES Kenya Occasional Paper*(5).

Tara Nutley, E. S., Nicole Judice. (2006). Data Demand and Use: An Introduction to Concepts and Tools.http://www.cpc.unc.edu/measure/resources/publications/ms-12-54/at_download/document.

WHO. (2003). *Improving data quality: a guide for developing countries*: Manila: WHO Regional Office for the Western Pacific.

WHO. (2007). Everybody's business--strengthening health systems to improve health outcomes: WHO's framework for action.

WHO. (2011). Manual on use of routine data quality assessment (RDQA) tool for TB monitoring.