

Improving Health Information Systems Data Quality in Sub-Saharan Africa

Bry SYLLA^{a,bl}, Léon SAVADOGO^c and Gayo DIALLO^a

^a BPH Inserm-U1219, Univ. Bordeaux, France

^b Public Health Team, Nazi Boni University, Burkina Faso

^c Public Health Department, Alexandria Senghor, Egypt

ORCID ID: Bry SYLLA <https://orcid.org/0000-0003-3057-8711>

Abstract. Data quality in health information systems (HIS) is essential for informed decision-making in the health sector, particularly in sub-Saharan Africa (SSA) where these systems face many challenges like resource limitations and weak infrastructure. This systematic review assessed the quality of HIS data in the region, focusing on the dimensions, and factors influencing this quality. It highlights the importance of systematic evaluation, ongoing training for data collectors in the analysis and use of data for decision-making, and the adoption of information and communication technologies in the healthcare system to improve data quality. These findings point the way to better use of health data and the need for a more integrated approach to digital health in SSA.

Keywords. Health information system, data quality dimensions, sub-Saharan Africa

1. Introduction

Data quality is crucial in decision-making and policy development, especially in the health sector, where the Health Information System (HIS) plays a key role as one of the six pillars defined by the WHO. HIS can be defined as a system made up of actors, tools, and methods that interact at different stages of the health information production process, namely data production, aggregation and storage, analysis, and information sharing [1]. It provides the tools needed to evaluate and plan health policies, networks, systems, and services [2]. Inadequate HIS can lead to resource misallocation, incorrect medical outcomes, and a loss of trust in the healthcare system [3,4]. The Routine Health Information System (RHIS), a critical part of HIS, circulates data from health facilities' routine activities, with its role varying by health system level: operational at the peripheral, strategic at the regional, and political at the national level. HIS's effectiveness relies on quality data collection using standardized physical and/or electronic data collection tools across all healthcare levels [5]. Recent efforts aimed at improving HIS globally, particularly in Africa through electronic data collection systems, the popularization of the DHIS2, the introduction of data warehouses, and tools for evaluating the quality and performance of HIS [6].

Despite these improvements, in sub-Saharan Africa (SSA), HIS faces challenges like endemic diseases, resource shortages, and infrastructure limitations, impacting data

¹ Corresponding Author: Bry SYLLA; E-mail: bry.sylla@u-bordeaux.fr.

quality, we are witnessing the production of data that is much more quantitative than qualitative [7]. This systematic literature review examines the data quality of HIS in these regions, focusing on how human resources, technology, regulations, and governance influence data quality. It aims to identify improvement strategies to enhance routine information systems, guiding public policy and interventions in SSA.

2. Methods

2.1. Choice of bibliographic databases and search strategy

The review ran from June 2023 to December 2023, and Our research question focuses on identifying data quality and related factors in SSA.

Pubmed, Scopus and Google Scholar were used to search for articles. As a strategy, we used keywords, sometimes substituting synonyms or similar words or expressions, such as "health information system" (routine health information system, health management information system, health information system), "data quality" (data quality assessment, data quality improvement, information quality), "factors" (causes, determinants), "Africa" (Sub Saharan, developing country, low- and middle-income country). An additional manual search consisted of combining keywords specifically with certain terms or following links to references cited in other studies. Once the keywords had been defined, a combination of these words was used by linking them using the Boolean operators AND and OR.

2.2. Data extraction and inclusion/exclusion criteria

The selected publications were stored in a Zotero library, with extracted data including author, year, dimensions assessed, study population, and associated factors. Quality assessment involves quantitative measures across multiple dimensions. Articles published in scientific journals assessing HIS data quality, from 2012 to 2022 about SSA are considered, regardless of methodology. Qualitative assessments of influential factors and indices on data quality were included, but narrative comments were excluded.

3. Result

From a search across three electronic bibliographic databases, 1115 articles were initially identified (632 on Google Scholar, 125 on SCOPUS, and 358 on PubMed). After removing 120 duplicates, 995 articles remained. Screening titles and abstracts for relevance to data quality in routine information systems in sub-Saharan Africa excluded 960 articles. A detailed review of the 35 remaining articles, using specific inclusion and exclusion criteria, resulted in 10 articles being selected. An additional article was identified through reference checking, totaling 11 articles for the review.

3.1. Study locations, data quality dimensions, and data collection tools

The studies included in this review were conducted between 2012 and 2022 in East Africa (n=06), West Africa (n=03), Central Africa (n=01), and South Africa (n=1).

This study deals with hybrid information systems combining paper and digital methods. Data are collected from paper consultation registers or paper patient records, then manually collated into paper monthly activity reports (MARs). These reports are sent to the health information office at the intermediate level for entry into the DHIS2 electronic system, where a digital copy is kept for each hospital.

All the studies concluded that the data quality in their HIS was poor. Our review identified eight attributes or dimensions.

- The studies revealed confusion between accuracy (S3, S6, S10, S11) and precision (S1, S2, S4, S5, S8) with authors defining them as the agreement between data in paper MAR or on DHIS2 and consultations registers. A "verification ratio" was used to measure this agreement.
- Exhaustiveness (S1-S11) refers to various aspects of the completeness of items in the register it includes checking for missing values of all items against the provided instructions and procedures, the presence of all relevant indicators in the report against the registers, and the percentage of paper MAR received from health facilities compared to the expected number of monthly paper-based MAR. The studies determined this by checking for missing values between the reports and the registers and calculating the reports' availability rate at the district level.
- Promptness (S1) and timeliness (S3, S5, S7, S9, S10) are defined by the authors as the delay in transmitting paper reports to the Health Information Office for entry into DHIS2. It was assessed during the study periods by determining the rate of paper-based MAR not transmitted within the defined deadlines.
- Reliability (S2): is characterized by authors as the alignment of each case with the defined case criteria outlined in the recommended terms of reference. Verification relied on clinical information recorded in the register without direct patient clinic verification.
- Internal (S7, S9) and external (S7, S9) consistency: Internal consistency evaluates data coherence within the DHIS2 system, while external consistency assesses DHIS2 data alignment with estimates from external sources like prior surveys and household surveys.

Table 1. Summary of studies according to collection tools, and dimensions of data quality

Authors	Country and period	HIS collection tools
Toalé Sibri SORY et al	Burkina Faso, April to May 2015	Mixed (Paper and electronic)
Yolaine Glèlè Ahanhanzo et al	Benin, October 2012	Mixed (Paper and electronic)
Binyam Tilahun et al	Ethiopia, 2020	Mixed (Paper and electronic)
Daniel Hahn et al	Kenya, march to august 2012	Mixed (Paper and electronic)
Adisu Tafari Shama et al	Ethiopia, july 2020	Mixed (Paper and electronic)
Susan F. Rumisha et al	Tanzania, Oct to Nov 2017	Mixed (Paper and electronic)
Alphonse Nshimiyiryo et al	Rwanda, July to September 2017	Mixed (Paper and electronic)
Richael O'Hagan et al	Malawi, June 2016	Mixed (Paper and electronic)
Antoinette H Bhattacharya et al	Nigeria, July to June 2017	Mixed (Paper and electronic)
Nigusu Getachew et al	Ethiopia, June to July 2021	Mixed (Paper and electronic)
Azoukalné Moukéné et al	Chad, January to December 2018	Mixed (Paper and electronic)

3.2. System factors influencing data quality

Data quality in healthcare institutions is often correlated with various factors associated with the information system. These factors can be categorized according to the five main groups outlined in the WHO Data Quality Review (DQR) framework and can be assessed during evaluations. The authors articulated these factors as deficiencies, which we have analyzed and organized into these five groups. Detailed factors can be found in the accompanying analysis file.

Table 2. Summary of factors associated with the data quality

Group of factors	Inadequate staff training	Inadequate guidelines for the collection process	Inadequate quality of data collection tools	Insufficient supervision and feedback	Insufficient analysis and use of data
Studies	S1-S11	S1-S3, S5-S7, S9-S11	S1, S2-S7, S9, S11	S1, S3, S5, S8-S10	S1, S2-S7, S9, S11

4. Discussion

4.1. Main conclusions

SSA countries often face data management challenges that affect health data quality [8]. Among the WHO's 50 data quality dimensions, only 11 are primary, with completeness, accuracy, and timeliness being the most commonly assessed [9]. The eleven reviewed articles focused on the three key attributes, but none covered all dimensions. Only, four studies assessed the three simultaneously, but inconsistencies led to the interchangeable use of terms like accuracy/precision and timeliness/promptness. Across the studies, poor data quality was consistently highlighted, exacerbated by the absence of a consensus or national benchmark for assessment. Comparisons were often drawn against country-specific thresholds, previous studies, or WHO's DQR module 1 recommendations.

Our analysis found that health worker training, high-quality data collection tools, and clear data collection procedures most often affected data quality. These factors were more important than data collection supervision and analysis and use. To improve data reporting, recommendations stressed these three areas. Field supervision was beneficial, but resource constraints limited its implementation. Healthcare staff data analysis and use competency received less attention than other factors.

Furthermore, the review also indicated low adoption of information and communication technologies (ICTs) in HIS across SSA. Primary digital tools, like computerized patient records, were underutilized, relying on intermediate electronic tools such as DHIS2 for storing health data. Typically, aggregated data was entered into monthly health facility reports [6].

4.2. Contributions and limitations

This study highlights limitations in data quality assessment across Sub-Saharan Africa (SSA). Our review identified completeness as a universal concern within these countries' health information systems (HIS) [10]. However, mechanisms to improve data quality are lacking. There's no consensus on which dimensions to monitor; some dimensions are incompatible with paper-based data collection. To address these issues, countries in SSA

need to: (i) Define and monitor critical data quality dimensions: This includes establishing thresholds for acceptable data completeness, accuracy, and timeliness within their HIS. (ii) Invest in technological solutions: Implementing electronic data capture systems like DHIS2 can significantly improve data quality [10]. (iii) Prioritize data culture: A pervasive lack of information culture among healthcare workers persists [11, 12]. Ongoing training in data utilization and technical skills development are crucial to elevating data quality standards.

5. Conclusions

Our study provides valuable insights by identifying the specific data quality attributes that require monitoring and the most effective methods to accomplish this. Our proposed solutions include the utilization of health data, the implementation of comprehensive training programs, and the provision of technical resources. The purpose of these insights is to assist countries and stakeholders in improving the management of health information. This includes enhancing the accuracy of data and promoting the use of ICT to collect routine health data at healthcare facilities.

References

- [1] World Health Organization. Everybody's business -- strengthening health systems to improve health outcomes: WHO's framework for action 2007:44.
- [2] Salam RA, Das JK, Bhutta ZA. Achieving Health-Related Sustainable Development Goals: Role of Health Systems Strengthening. In: Mataria A, Rouleau KD, Iqbal M, Siddiqi S, editors. *Making Health Systems Work in Low- and Middle-Income Countries: Textbook for Public Health Practitioners*, Cambridge: Cambridge University Press; 2022, p. 464–78. <https://doi.org/10.1017/9781009211086.031>.
- [3] Getachew N, Erkallo B, Garedeew MG. Data quality and associated factors in the health management information system at health centers in Shashogo district, Hadiya zone, southern Ethiopia, 2021. *BMC Medical Informatics and Decision Making* 2022; 22:154. <https://doi.org/10.1186/s12911-022-01898-3>.
- [4] El Sibai R, Abdo JB. *Data Quality Management for Pervasive Health Monitoring in Body Sensor Networks. 5G Impact on Biomedical Engineering*, CRC Press; 2022, p. 129–46.
- [5] Health Metrics Network, Organization WH. *Framework and standards for country health information systems*. World Health Organization; 2008.
- [6] Koumamba AP, Bisvigou UJ, Ngoungou EB, Diallo G. Health information systems in developing countries: case of African countries. *BMC Med Inform Decis Mak* 2021; 21:232. <https://doi.org/10.1186/s12911-021-01597-5>.
- [7] Glèlè Ahanhanzo Y, Ouendo E-M, Kpozèhouen A, Levêque A, Makoutodé M, Dramaix-Wilmet M. Data quality assessment in the routine health information system: an application of the Lot Quality Assurance Sampling in Benin. *Health Policy and Planning* 2015; 30:837–43. <https://doi.org/10.1093/heapol/czu067>.
- [8] Chekol A, Ketemaw A, Endale A, Aschale A, Endalew B, Asemahagn MA. Data quality and associated factors of routine health information system among health centers of West Gojjam Zone, northwest Ethiopia, 2021. *Frontiers in Health Services* 2023;3.
- [9] Alipour J, Ahmadi M. Dimensions and assessment methods of data quality in health information systems. *Acta Medica Mediterranea* 2017; 2017:313–20. https://doi.org/10.19193/0393-6384_2017_2_047.
- [10] Chen H, Hailey D, Wang N, Yu P. A Review of Data Quality Assessment Methods for Public Health Information Systems. *International Journal of Environmental Research and Public Health* 2014;11: 5170–207. <https://doi.org/10.3390/ijerph110505170>.
- [11] Mogessie YG, Ntacyabukura B, Mengesha DT, Musa MB, Wangari M-C, Claude N, et al. Digital health and COVID-19: challenges of use and implementation in sub-Saharan Africa. *Pan Afr Med J* 2021; 38:240. <https://doi.org/10.11604/pamj.2021.38.240.27948>.
- [12] Tom M, Mwanzo I, Otieno G, Kamau P. Technical factors influencing the use of data for evidence-based decision making amongst health workers at Kisumu County, Kenya. *International Journal Of Community Medicine And Public Health* 2023; 10:1362–8. <https://doi.org/10.18203/2394-6040.ijcmph20230911>.