South African Journal of Information Management
ISSN: (Online) 1560-683X, (Print) 2078-1865

Quality of routine health data in DHIS2 in South Africa: Eastern Cape province from 2017–2020

Authors:
Sibusiso S. Thabethe[1]
Verona E. Mathews[1]

Affiliation:
[1]School of Public Health, Faculty of Community and Health Sciences, University of the Western Cape, Cape Town, South Africa

Corresponding author:
Verona Mathews,
vmathews@uwc.ac.za

Dates:
Received: 12 July 2024
Accepted: 15 Oct. 2024
Published: 29 Nov. 2024

How to cite this article:
Thabethe, S.S. & Mathews, V.E., 2024, 'Quality of routine health data in DHIS2 in South Africa: Eastern Cape province from 2017–2020', *South African Journal of Information Management* 26(1), a1903. https://doi.org/10.4102/sajim.v26i1.1903

Background: Routine health information plays a significant role in managing the healthcare system to make informed decisions, monitor and evaluate, and take action to improve health outcomes. The Eastern Cape Department of Health (ECDoH) officially adopted the online District Health Information System (DHIS2) in 2017. However, evidence suggests that the underutilisation of routinely collected health information for management was often because of poor data quality in routine health information systems.

Objectives: This study reviewed the level of quality of the routine health data in the DHIS2 using two quality dimensions, namely data completeness and internal consistency.

Method: A retrospective study design was used to assess the quality of data for April 2017– March 2020 utilising the World Health Organization Data Quality Guidelines. Secondary data were extracted from the DHIS2 using standardised reports and captured in Microsoft Excel Office. A total of 265 health facilities and 77 data elements were included and analysed using descriptive analysis and a score grading system.

Results: A total of 365 228 data element values were reported, 121 199 missing data values, unaccounted data values were at 248 153, about 6395 data values had outliers, and 5670 data values had validation errors. The rate of data completeness was 74.6%, the internal consistency was 95.1%, and the DHIS2 data quality was 84.9%.

Conclusion: The study demonstrates the high quality of DHIS2 data in the ECDoH following the implementation of the online-based system. However, the significant number of missing data elements has impacted data completeness.

Contribution: This study contributes to the body of knowledge on the importance of role of data quality for the utilisation routine health information systems as an essential tool for management in the health system.

Keywords: data quality; data dimensions; public health facilities; health information systems; routine health data; DHIS2.

Introduction

The role of routine health information has over the years gradually become a fundamental feature in the public health systems, in particular, for planning and general management of healthcare services (Blödt et al. 2018; Lutge et al. 2016; Tilahun et al. 2018). In 2008, the World Health Organization (WHO) named health information as one of the six building blocks that must be considered to strengthen the health systems (WHO 2007). Routine health information is a potentially useful and essential tool for public health because of its footprint and availability in terms of frequency and level of disaggregating that has been seen in various countries (Agiraembabazi et al. 2021; Bhattacharya et al. 2019). In recent years, many countries have adopted and implemented routine health information systems (RHIS) to manage health information and generate health statistics, which are critical for monitoring the progress of the implementation of health programmes (Lee et al. 2021; Maïga et al. 2019).

The use of information systems in the health system emanates from the desire to improve the quality of the health facility data and the effective monitoring of the progress of

health programme performance (English et al. 2011). Health information systems (HIS) play a fundamental role in ensuring that the collection of health information from various health facilities is appropriately organised and easily available in all tier levels of the health system (Tshabalala & Taylor 2016). The utilisation of HISs varies based on specific needs, priorities and resource availability in different countries (Garavand et al. 2016; Zhou et al. 2019). Many countries adopted the District Health Information System (DHIS) software as a primary data source for routinely collected health data, to improve data reporting and utilisation (Begum et al. 2020; Hagel et al. 2020). The DHIS software is a free and open-source software platform for reporting, quality checks, analysis, visualisation, and sharing of health information at all tier levels of the healthcare system (Manoj et al. 2012).

The DHIS software traced its roots to South Africa in the 1990s, with the initially developed project by the HISs Programme and the University of Western Cape, and they were supported by various development agencies (Farnham et al. 2020). Since its inception, the DHIS software has undergone major phases of development, moving from a standalone system, which was developed using Microsoft Office Package into a modernised online platform commonly referred to as DHIS2 or WebDHIS in South Africa (DHIS2 2022; Garrib et al. 2008). Currently, the DHIS2 is used in more than 100 countries and is believed to have a population estimated to be around 3.2 billion people (DHIS2 2022). Although the DHIS2 is widely adopted, some countries opted to implement the selected modules to support a specific health programmes. The utilisation of the DHIS2 as the main management information system in many countries resulted in improvement in the completeness of reporting, data availability and data utilisation by healthcare workers (Dehnavieh et al. 2019; Hung et al. 2020; Maïga et al. 2019).

In South Africa, the National Department of Health (NDoH) initially used DHIS in the Western Cape Province, and subsequently, the implementation was extended to other eight provinces in the year 2001 (Garrib et al. 2008). Since 2001, the system has been used as a standalone system, and it was only in the year 2017 that the NDoH decided to upgrade and use DHIS2 (NDoH 2017a). The Eastern Cape province adhered to the clarion call to implement the DHIS2 software in the health facilities as per the instructions of the NDoH. Dehnavieh et al. (2019) posit that the DHIS2 like any other information system has its strengths, particularly in the technical and functionality aspects; however, these strengths face some challenges and concerns that need to be considered when it is adopted. A great trust and optimism have been placed in the DHIS2 for health data management and to support the provision of the information that is crucial for health professionals and decision-makers in all tier levels of the health system (Wagenaar et al. 2016; Wude et al. 2020).

All public health facilities in South Africa are obliged to report routine health data unless stated otherwise. The process of collecting routine health information starts with clinicians recording patient data into patient records and updating the same information into tick registers at each service point, followed by monthly data consolidation into a monthly data input form for the health facility and then the health data are captured into the district health management information system (HMIS) (Mphatswe et al. 2012; NDoH 2011). The National Indicator Dataset (NIDS), which is a list of data elements and indicators, aligned with the Department of Health-implemented health programmes (Day & Gray 2017) is used as a guide to standardise routine health data recording, reporting and interpretation (NDoH 2011, 2017b).

The DHIS2 has been hailed as a good innovation and data source for many countries (Hagel et al. 2020; Shama et al. 2021). In Ghana, the study found that the implementation of DHIS2 positively influenced the improvement in the data quality particularly the reporting rate, timeliness and completeness (Amoakoh-Coleman et al. 2015). Another study conducted in Lebanon affirmed that the implementation of DHIS software as a tool for data management improved the timeliness and completeness of aggregated data reporting (Youssef et al. 2022). However, the utilisation of DHIS2 was not spared from the challenges of data quality where a study conducted in 14 countries found the presence of extreme outliers and inconsistency of data over time as well as data discrepancies between the indicators in the health facility reports (Maïga et al. 2019). In South Africa, Nicol, Dudley and Bradshaw (2016) concurred that the reported data on the management information system had data discrepancies when verified with the data sources in the health facilities.

The frameworks to assess the RHIS vary and they are developed to evaluate different aspects of the system such as technical, sociological, economic, human and organisational. These assessment frameworks provide a set of guidelines and procedures for the evaluation of the adequacy of HIS (Andargoli et al. 2017). Over the years, various frameworks have been adopted and noticeably used to evaluate the RHIS. These frameworks include the Health Metrics Network framework (HMN) which is referred to as an evaluation tool for HIS. This framework measures the performance of the country HIS, and it focuses on health information and statistical systems as well as leadership for health information production and use (Barro et al. 2020). Another framework is the Performance of the Routine Information System Management (PRISM) framework which measures the performance and the process of RHIS by evaluating the behavioural determinants, the technical determinants, and the organisational and/or environmental determinants that form part of the system (Aqil, Lippeveld & Hozumi 2009).

The WHO framework is the WHO Toolkit for Data Quality Review which provides a method for analysing RHIS data by

using four dimensions of data quality such as completeness and timeliness of data, internal consistency of the data, external consistency of data, and external comparisons of population data (Adane et al. 2021; WHO 2017a). Furthermore, this framework comprises two separate methodologies to evaluate the RHIS data which include a desk review of the data, and a health facility assessment and is considered a complex task due to the nature of healthcare systems (Andargoli et al. 2017).

Problem statement

The South African Department of Health has a strong tradition for data collection and monitoring the performance of health programmes in all tier levels of the healthcare system. Moreover, the department has been using the DHIS software for over 20 years to collect and analyse routine data. The NDoH adopted DHIS2 as the priority information system to help reduce and prevent data quality issues and has since made considerable progress in the improvement of the RHIS, to support management and healthcare service provision. Since the introduction of DHIS2 in the year 2017, there has been limited literature on the assessment of quality in the DHIS2 data. The primary aim of this study was to evaluate the quality of routine health data for the first years of the DHIS2 implementation in the Eastern Cape province. The WHO data quality review toolkit was used as the guide, and two data quality dimensions were selected, data completeness and internal consistency, to review the level of data quality.

Research methods and design

Study setting

Eastern Cape province is one of the nine provinces in South Africa and it is in the coastal part of the country. The province has eight District Municipalities and two of them are Metropolitan Municipalities (see Figure 1). The districts are further divided into 31 sub-districts. The Eastern Cape province is home to approximately 7.2 million people and an estimated 90.5% rely on public healthcare services (Eastern Cape Department of Health [ECDoH] 2024; Statistics South Africa [Stats SA] 2024). The Eastern Cape province covers an area of 168 966 km^2 and it constitutes a share of about 11.1% of the Republic of South Africa (RSA) population, and it is the second-largest province in surface area and third in terms of population size (ECDoH 2020). The ECDoH has a total of 860 public health facilities consisting of 727 Primary Health Care (PHC) Clinics, 41 Community Health Centres, and 92 hospitals as a platform for service delivery (ECDoH 2020).

A retrospective methodological approach was utilised to collect and assess the quality of the reported health data in the DHIS2 software. All public health facilities are using DHIS2 software for the collection and reporting of routine health data in the Eastern Cape province. A multistage sampling method was used to select health facilities and data elements and the multistage sampling process had two stages. The Raosoft Sample size calculator was used to calculate the sample size, which included a 5% margin of error, a 95% confidence level, and the population of 850 health

Source: Wikipedia, 2024, *List of municipalities in the Eastern Cape*, viewed n.d., from https://en.wikipedia.org/wiki/List_of_municipalities_in_the_Eastern_Cape

FIGURE 1: Districts of the Eastern Cape province.

facilities and 95 data elements. The study sample size included 265 health facilities and 77 data elements.

Data source

The data source for secondary data collection was DHIS2. The data were extracted from the DHIS2 for a period of 3 years, starting from April 2017 to March 2020. The extracted data were all downloaded into a Microsoft Office Excel format. The facility's monthly routine health data are collected by the healthcare professionals and then captured into DHIS2 by a dedicated Information Cadre appointed or seconded personnel in the health facility who is trained in DHIS2 (NDoH 2012, 2013). The DHIS2 has built-in functionality for data quality reports, which are available for system users to check and identifies data errors in the system. The data quality reports included in this study are the Data Validation Report (Appendix 2) and the Outlier and Missing Report (Appendix 3). The health facility data were extracted from the DHIS pivot tables (Appendix 1).

Data validation report

This report shows data violations based on the pre-defined validation rules (Health Information System Program [HISP] Team 2016). The validation rules are expressed by conditions set between data elements.

Outlier and missing report

This WHO Data quality tool was used to identify data gaps and outliers (HISP Team 2016).

The District Health Information System Pivot Table

This is an analytic tool that was used to summarise data according to dimensions such as data elements and indicators, periods, and organisational units (HISP Team 2016).

TABLE 1: National indicator dataset 2017 data element group collection points.

Data element group	Clinic	Community Health Centre	Hospital
Adolescent Health	-	X	X
ART Monthly	X	X	X
CCMDD	X	X	-
Child and Nutrition	X	X	X
Communicable Diseases	X	X	X
EPI	X	X	X
Eye Care	X	X	X
HIV	X	X	X
Inpatient Management	-	X	X
Management PHC	X	X	-
Maternal and Neonatal	X	X	X
Mental Health	X	X	X
NCD	X	X	X
Oral Health	X	X	X
Quality	X	X	X
Rehabilitation	-	X	X
STI	X	X	X
TB monthly	X	X	X
Women's Health	X	X	X
Total	**16**	**19**	**17**

Source: National Department of Health (NDoH), 2017b, *National indicator dataset 2017: Data elements*, National Department of Health, Pretoria

ART, antiretroviral treatment; CCMDD, Central Chronic Medicines Dispensing and Distribution; EPI, Expanded Programme on Immunisation; PHC, primary healthcare; NCD, Non-communicable Disease; STI, sexually transmitted infections; TB, tuberculosis.

The extracted data from the DHIS2 included all public health facilities (850) as they are registered in the National Department of Health Data Dictionary (NDoH 2017b). All the included health facilities were expected to report health information every month to DHIS as per NIDS requirements. In addition, a total of 229 data elements were included that form part of NIDS (NDoH 2017b) (see Table 1).

However, exclusion and inclusion criteria were applied to ensure that there were no ambiguities with the data analysis and interpretation. The exclusion for health facilities included health facilities that were non-government-owned, health facilities that were intermittent during the period under review, specialised health facilities, mobile clinics, satellite clinics, health posts, and non-medical health facilities. The exclusion of data elements included the data elements that did not apply to all health facilities, the data elements that were collected as part of the quarterly report, campaigns, and routine data from non-facility health services.

Data analysis

This DHIS2 data assessment was guided by the WHO toolkit for the data quality review framework. The WHO toolkit was developed through collaboration between WHO, the Global Fund to Fight Acquired Immunodeficiency Syndrome (AIDS), Tuberculosis, and Malaria (The Global Fund), Gavi, the Vaccine Alliance (Gavi), and the United States Agency for International Development (USAID)/MEASURE Evaluation (Adane et al. 2021; WHO 2017a). The WHO toolkit has four data quality dimensions; however, this study focussed on the two dimensions, namely completeness of data, and internal consistency of the data (WHO 2017a). The completeness of data was measured using two data metrics, namely the health facility data reporting, and completeness of data element values reported. The internal consistency was measured using data outlier, data consistency over time, and data consistency between related data elements (Table 2).

The score grading system applied criteria that consisted of three levels: a score of 90% and above was deemed excellent performance, a score of between 75% and 89% was rated as good performance, and a score rated 74% or less was regarded as poor performance. The extracted data were assembled into one Microsoft Excel spreadsheet version 2013 wherein the statistical analysis was performed by calculating frequencies and percentages. Simple descriptive statistics were used to analyse and display the study results. The quality of data in the DHIS software was assessed as shown in Table 2.

Ethical considerations

Ethical approval was obtained from the University of Western Cape Higher Degrees and Humanities and Social Science

Research Ethics Committee (HS22/3/2) and the ECDoH Epidemiology and Research Office (EC_202205_003). Secondary data from the DHIS2 software was utilised and did not involve human participants. Although the secondary data extraction from the DHIS2 did not involve humans or individual data, the anonymity of the district names was maintained by assigning pseudonyms to districts.

Results

The data extracted from DHIS2 was combined into a single spreadsheet for data cleanup and analysis. We conducted a comprehensive assessment of the quality of data for 77 data elements representing 15 health programmes, across a total of 265 health facilities spanning all health districts in the ECDoH. The results of the analysis are presented next.

Data completeness

Data completeness was measured using two data quality metrics namely health facility reporting and data element reporting. Health facility reporting was measured as the number of monthly reports received divided by the expected number of reports in a year. Data element reporting was measured as the number of data element values reported divided by the expected number of data element values in a year.

Health facility data reporting

All public health facilities are expected to submit the monthly routine health data to the district and ultimately to the Provincial Office and National office (NDoH 2011). Health facilities reported data 9489 times out of 9540 and all three financial years were very good with an average of 99.5%. It is important to observe that District E had one clinic that did not report any data in the first and third years and another clinic did not report data for a period of 3 months.

Data element reporting

A total of 365 228 data element values were reported by the health facilities during the period under review. The reported data elements are those data elements that do not have any missing values during the period under review. The missing data report was used to quantify and identify data gaps in the reports submitted by the health facilities. A total of 121 199 data element values were found to be missing in the report during the period under review. The DHIS2 missing data report revealed that the data element values missing rate was relatively high with an average of 16.50% across all the districts. However, a gap of 248 153 data element values that were expected to be reported was not accounted for in the report. When combining the number of missing data element values and unaccounted data element values an average of 50.3% was recorded while the reported data element values accounted for 49.7% of the DHIS2 data (see Figure 2).

The overall rate of health facility reporting was 99.5% (Table 3) and data element reporting was 49.7%. The data

TABLE 2: Data analysis summary for assessing the quality of routine data reported in the District Health Information System2 Software.

Data quality metric	Analysis	Data quality review guidance	Data source
Dimension 1: Data Completeness			
Completeness of health facility reporting	The proportion of expected monthly reports submitted by health facilities	The completeness rate of reporting should be above 75%	DHIS Report: • Pivot table report
Completeness of data reported	The proportion of non-missing values for a given data element in expected monthly reports	The completeness rate of reported data should be 100%	• Outlier and missing data report
Dimension 2: Internal consistency of reported data			
Outliers	Number of extreme outliers (+3SD from the mean) of monthly values during the period under review	A data element value of above 3.5 on the modified Z-score is considered an extreme outlier.	DHIS Report: • Outlier and missing data report
Consistency over time	The consistency over time was calculated as the cumulative value of the data element for the preceding years over the mean value of the data elements for the comparison year.	Assess the reported data element values by comparing the current year to the value predicted from the trend in the preceding years. It is expected that the reported values for the reference year be within a ratio of ± 1.33 for the preceding years.	• Pivot table report
Consistency between related data elements	The relationship between two data elements at the facility level is assessed by comparing their correlation to the values reported.	This examines the extent to which two related data elements follow a predictable pattern.	• Validation analysis report

DHIS, District Health Information System.

TABLE 3: Health facility data reporting per district.

Health district	District B (%)	District D (%)	District E (%)	District F (%)	District H (%)	District L (%)	District P (%)	District T (%)	EC province (%)
2017/18	100.0	100.0	95.7	100.0	100.0	100.0	100.0	100.0	99.2
2018/19	100.0	100.0	97.3	100.0	100.0	100.0	100.0	100.0	99.5
2019/20	100.0	100.0	97.9	100.0	100.0	100.0	100.0	100.0	99.6
Total	**100.0**	**100.0**	**97.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**99.5**
Clinic	100.0	100.0	96.2	100.0	100.0	100.0	100.0	100.0	99.4
CHC	-	100.0	100.0	100.0	-	-	100.0	100.0	100.0
Hospital	-	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Total	**100.0**	**100.0**	**97.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**99.5**

CHC, Community Health Centre; EC, Eastern Cape.

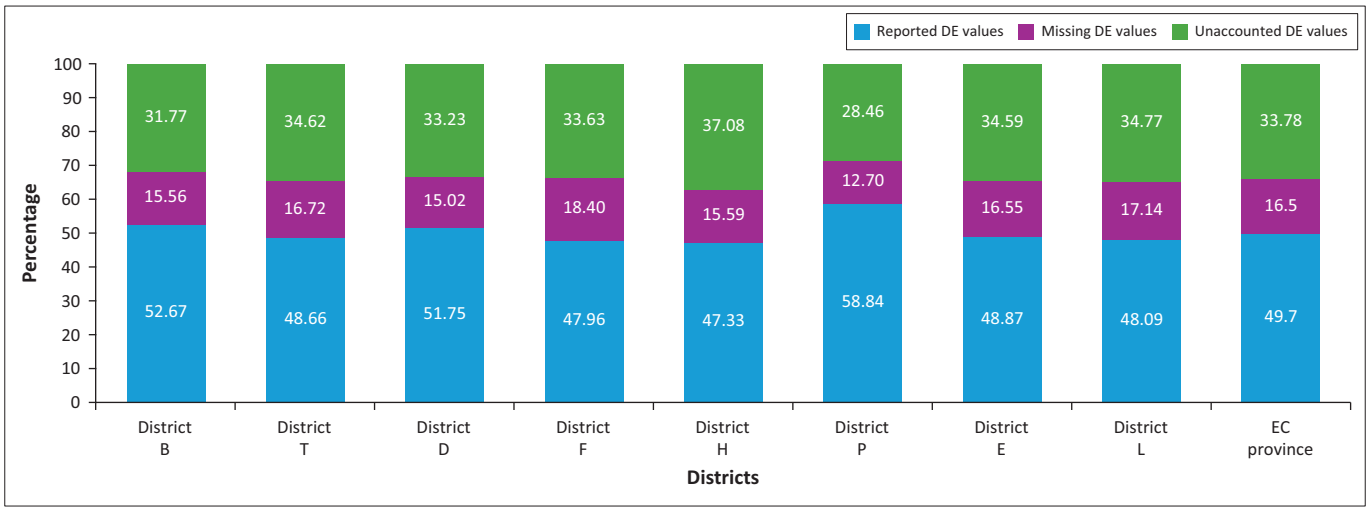
reporting rate per data element group was noticeably high for antiretroviral therapy (ART). Monthly at 99.4%, followed by the STI at 99.4% and the NCD at 85.1% (Figure 3). The data element group with the lowest reporting rate of less than 50.0% included CD, Eye Care, human immunodeficiency virus (HIV), Malaria, Maternal and Neonatal, Mental Health, Oral Health, and Quality. In terms of facility type, the CHCs recorded the highest reporting rate at 63.4%, and both the hospitals and clinics had the lowest rate for data element reporting at 41.4% and 49.7%, respectively (Figure 3).

Internal data consistency

Internal data consistency was assessed using three data quality metrics, namely outliers, data consistency over time, and consistency between related data elements. The extreme outliers were regarded as data errors and analysed to determine the extent to which data were affected. Extreme outliers were defined as those data element values that had a modified Z-score of 3.5 and above standard deviation (WHO 2017b). The second data quality metric involved the evaluation of data consistency over time, which was expected to change within the ratio of plus or minus 1.33 threshold (WHO 2017b). The third data quality metric focussed on the assessment of the consistency of the data element values between two or more data elements to show the relationship. The study results showed that the data have a low rate of outliers at 1.75%, data consistency over time was the highest data quality metric at 11.34%, and consistency between related data elements was also low at 1.55%.

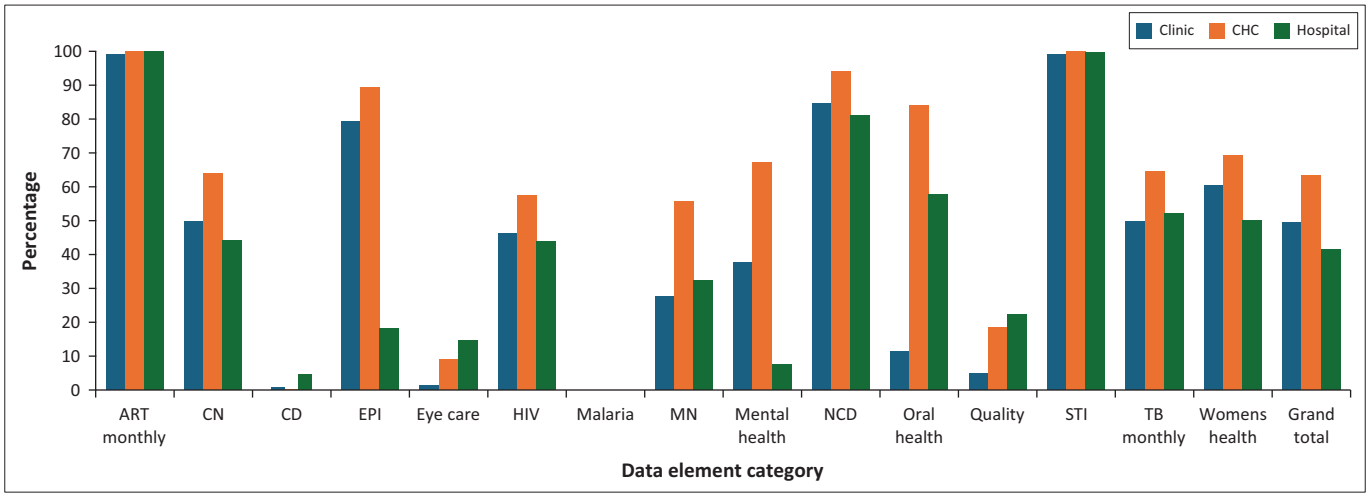
Outliers in the District Health Information System2 Data

Outlier was assessed to determine whether a data value in a series of values is out of range with the other values in the series (WHO 2017a). The study used extreme outliers only to analyse the consistency of the reported data. Table 4 presents the frequency of outliers in the reported data from the health facilities.



DE, data element; EC, Eastern Cape.

FIGURE 2: Data element reporting per district.



ART, antiretroviral treatment; CN, Child and Nutrition; CD, Communicable Diseases; EPI, Expanded Programme on Immunisation; MN, Maternal and Neonatal; NCD, Non-communicable Disease; STI, sexually transmitted infections; TB, tuberculosis; CHC, community health centre.

FIGURE 3: Data element reporting per data element group.

Table 4 shows the rate of the data element values that had outliers in the DHIS2 data. The overall rate for outliers was relatively low with an average rate of 1.75% for all eight districts. The study results showed that reported health data had a low rate of extreme outliers in the DHIS2 system.

Consistency over time in the District Health Information System2 data

The data consistency over time was assessed to determine the extent of the plausibility of reported results for health programmes in terms of the history of reporting routine data element values (WHO 2017a). The WHO framework for routine data quality recommends that data consistency over time should be measured using 4 years of data; however, the lifespan of NIDS2017 was 3 years (NDoH 2017b; WHO 2017a). To accommodate the fourth year, the data elements that overlapped with NIDS2020 were used to calculate consistency over time.

Table 5 presents the frequency of data element values that were above the tolerance level for consistency over time in the reported health facility data.

Table 5 shows the scoring rate of the data consistency over time, which was relatively acceptable as per the threshold for good performance. The overall rate for data inconsistency over time in the reported data from health facilities achieved an average of 11.34%.

Validation analysis in the District Health Information System2 Data

The validation analysis measures the data violations based on the pre-defined validation rules, which are expressed by the conditions set between data elements. The study results showed that all districts achieved a rate of less than 3% for data validation errors. The overall rate for validation errors was 1.55% (Table 6), which was significantly low in the DHIS2 data.

TABLE 4: Outliers per district.

Health district	Number of reported DE values	Number of DE with outliers	% of DE with outliers (%)
District B	27 741	533	1.92
District D	40 168	668	1.66
District E	63 664	1260	1.98
District F	77 112	1330	1.72
District H	19 681	355	1.80
District L	34 661	551	1.59
District P	29 358	588	2.00
District T	72 843	1110	1.52
EC province	**365 228**	**6395**	**1.75**

DE, Data element; EC, Eastern Cape.

TABLE 5: Data consistency over time per district.

Health district	Count of DE of ratio in the reported data	Count of DE with a ratio of 1.33 and above in Year 4 compared to Mean of Year 1–Year 3	% of DE with a ratio of 1.33 and above
District B	1197	203	16.96
District D	1764	184	10.43
District E	2961	382	12.90
District F	3654	434	11.88
District H	945	114	12.06
District L	1638	171	10.44
District P	1134	97	8.55
District T	3402	309	9.08
EC Province	**16 695**	**1894**	**11.34**

DE, Data element; EC, Eastern Cape.

Data quality in the District Health Information System2 Software

The data quality was evaluated using the WHO toolkit for data quality, and two dimensions were analysed to determine the quality of data in the DHIS2. Data completeness and internal data consistency were used, and each data quality dimension had two or more data quality metrics that were included for analysis.

The overall results of data quality in the DHIS2 were 84.9% for the period under review (Table 8). The data completeness rate of the reported data in the DHIS2 had an overall score of 74.6%, which was rated as poor performance. The rate for internal data consistency in the DHIS2 data recorded a scoring rate of 95.1% (Table 8). All districts achieved a scoring rate of 90% and above for internal data consistency, which was a good performance.

Discussions

Data completeness

To measure the data completeness, two data quality metrics were used (Table 7), and the study results on the first data quality metric showed that the health facility reporting was good with a rate of 99.5%, which was well above the accepted threshold of 75% for health facility rate (WHO 2017b). The study scoring rate showed a good performance in line with the findings of the study conducted in Ghana, which found the data reporting to be good at 94.3% (Amoakoh-Coleman et al. 2015). Another study conducted in the Democratic Republic of Congo found data reporting from health facilities to be good with a performance that ranges between 83.3% and 93.2% (Malembaka et al. 2021). In Bangladesh, the study's findings indicate that the implementation and utilisation of DHIS2 have resulted in improvements in the timeliness and completeness of data reporting over time (Begum et al. 2020). The implementation of DHIS2 helps to improve completeness in reporting routine health data from the health facility to the national level (Kiberu et al. 2014).

TABLE 6: Data validation analysis per district.

Health district	Number of reported DE values	Number of DE with validation errors	% of DE with validation errors
District B	27 741	218	0.79
District D	40 168	487	1.21
District E	63 664	1120	1.76
District F	77 112	1210	1.57
District H	19 681	315	1.60
District L	34 661	646	1.86
District P	29 358	646	2.20
District T	72 843	1028	1.41
EC Province	**365 228**	**5670**	**1.55**

DE, Data element; EC, Eastern Cape.

TABLE 7: Calculation of data quality dimensions.

Data quality dimensions	Calculations
Data completeness	**Average Rate of HF Reporting and DE Reporting**
Health Facility (HF) Reporting	(HF Reported Data Values/Data Values Expected to Be reported by HF during the period under review) × 100
Data Element (DE) Reporting	(Reported DE Values/DE Values Expected to be Reported during the period under review) × 100
Internal data consistency	**Average Rate of DE values without Outliers, DE values that are Consistent over Time, and DE values without Validations Error**
Data Element Values without Outliers	100% minus Rate of Data Element with Outliers
Data Consistency over Time	100% minus Rate of Data Inconsistency over Time
Data Elements without Validations	100% minus Rate of Data Elements with Validation Errors
Data quality	**Average Rate of Data Completeness and Internal Data Consistency**

DE, Data element; HF, Health facility.

TABLE 8: Data quality status of health facilities per health district.

Health district	Data completeness (%)	Internal consistency of data (%)	Data quality (%)
District B	76.30	93.40	84.90
District D	75.90	95.60	85.70
District E	72.90	94.50	83.70
District F	74.00	94.90	84.50
District H	73.70	94.80	84.30
District L	74.00	95.40	84.70
District P	79.40	95.70	87.60
District T	74.30	96.00	85.20
EC Province	**74.60**	**95.10**	**84.90**

EC, Eastern Cape.

For completeness of data elements, the WHO recommends a threshold tolerance level of less than 10% of the expected data to at least have missing values. However, the study results showed that the data element missing rate was found to be at 16.50%, which was above the tolerance level (WHO 2017b). The study result showed similar findings when compared to a study conducted in Ghana, which found that data reporting was around 84.6% of the reported monthly facility data (Nsiah et al. 2022). This occurrence was also identified in Nigeria where it was found that facility-reported data in the DHIS2 to be around 40% of the events documented in the facility registers (Bhattacharya et al. 2020). In addition, the study results revealed a higher rate of data element values that were not accounted for, which contributed greatly to the reduced rate of data completeness reported in the DHIS2. The substantial amount of missing data in the DHIS2 software highlights the problem of imperfect data capture, which influences the overall data quality in ECDoH. The accuracy of data reporting in health facilities relies on precise data capture and the lack of clear guidance in the NIDS document regarding data collection points for some data elements, which affects the availability and reporting of these elements in health facilities.

In this study, the results indicate that the data completeness rate in DHIS2 stands at 74.6%, slightly lower than the median rate of 95% observed in countries utilising DHIS2 as their primary data source (Githinji et al. 2017; Kebede et al. 2020; Maïga et al. 2019). Data completeness in the DHIS2 was lower than 75% for quality; this performance was lower compared to the findings of the study conducted in Guinea and Senegal. In Guinea, the study's findings showed that data completeness was at 98.5% in the DHIS2 (Reynolds et al. 2022) and in Senegal, the study indicated that the completeness of facility reporting was at 97.5% in the DHIS2 (Muhoza et al. 2022). The study results highlight the importance of striving for higher data completeness to ensure accurate and reliable information.

Internal data consistency

The overall rate for internal data consistency was found at 95.1%; the results indicated a similar trend to the study's findings in Ethiopia, showing that health programme data consistency was above 90% for almost all regions (Adane et al. 2021). The presence of extreme outliers was at 1.75% in the reported health facility data in this study and was found like the extreme outliers at 2.3% in the reported health data in the DHIS2 in Senegal (Muhoza et al. 2022). The study results on extreme outliers showed similar trends observed in Uganda that the reported data had a very low presence of extreme outliers in the national and sub-regions in a single year (Agiraembabazi et al. 2021).

The process of entering data from source documents into the system has been identified as a significant challenge (Nicol et al. 2016). However, recent studies have demonstrated that the DHIS2 exhibits very low data validation errors, with an average of just 1.55% across all districts. Minimal data errors in the DHIS2 compared to primary data sources have been reported (Amoakoh-Coleman et al. 2015). The low level of data validation errors signifies the accuracy of data reporting from health facilities. However, health facilities encountered data with validation issues in their reports (Lasim, Ansah & Apaak 2022). In addition, a data discrepancy rate of over 10% has been reported in the DHIS data (Tlale et al. 2019).

This study revealed that data consistency over time was at 11.34% during the period under review consistent with a similar study conducted in Senegal, which found data consistency over time to be at 15% in the DHIS2 (Muhoza et al. 2022). Although the WHO (2017b) does not explicitly provide a threshold rate for data consistency over time, a recommended threshold of up to 33% for data element values is suggested for routine health data changes. Cape Coast Metropolis in Ghana demonstrated good data consistency over time with an average of 7% in the reported data from health facilities (Lasim et al. 2022). In Nigeria, the study results indicate that most health indicators, which were 12 out of 19 were likely to be consistent over time (Adane et al. 2021). It is noteworthy that while some studies showed routine health data inconsistency over time in different

countries, however, maintaining a low level of data inconsistency is crucial for ensuring good data quality in DHIS2 (Bhattacharya et al. 2019).

Data quality in the District Health Information System2 Software

Data quality is an important element of the generated health information that enables authorities and healthcare workers to monitor performance and make appropriate decisions for continuous improvement (Blödt et al. 2018; Lemma et al. 2020). This study demonstrated a good rating for reporting data with an average of 84.9% of the reported data in the DHIS2 similar to Ethiopia at 83% in the HMIS (Getachew, Erkalo & Garedew 2022; Solomon et al. 2021).

Despite the good performance of 84.9% quality of data as shown in the study, other studies have not shown similar results. For example, in Ethiopia, the rate of quality of health data from the public health facilities was low at 51.4% in the Harari Region (Shama et al. 2021), whereas health centres within the West Gojjam Zone indicated a 74% data quality level (Chekol et al. 2023). The WHO (2017b) recommends a minimum threshold of 75% for routine health data to be recognised as good quality. Moreover, a literature review in South Africa identified suboptimal data quality within the Routine HIS (Roomaney et al. 2017). While the overall data quality was deemed satisfactory in the study, data quality has been identified as a challenge and often not utilised in other countries (Dadzie et al. 2021; Solomon et al. 2021).

Two data quality dimensions were assessed and the overall results for data completeness had concerns particularly the low level of reporting of data elements. The data completeness achieved a rating level of 74.6%, which was just below the acceptable level of 75% for the reported data in the RHIS (WHO 2017a). The internal data consistency scored a very good rate at 95.1% in the DHIS2 data. Although the NIDS document provides guidance for routine health data collection from health facilities and reporting in the DHIS2 (NDoH 2017b) the findings of the study reveal challenges in data element values reporting, which could either be because of poor data capturing in the DHIS2. The low data element reporting was as a result of several factors, such as inadequate data capturing from health facilities, resulting in a higher number of data elements with missing values. Also, the NIDS document lacks clarity on data collection points for certain data elements, making it difficult to allocate data element groups for reporting by health facilities in the DHIS2.

The WHO data quality toolkit provided a valuable and simple method for analysing DHIS2 data. The assessment of the reported health data in the DHIS2 can help to strengthen the awareness of data quality challenges and help improve data reliability and effective reporting in the routine health data into the DHIS2 (Adane et al. 2021; Bhattacharya et al. 2020). Limitations of the study were that timeliness and consistency of reported data compared to the original records were not evaluated as part of data quality. Further investigation is also required to determine external comparisons and external consistency of population data.

Conclusion

In summary, this study evaluated the quality of data in the DHIS2 using two data quality dimensions. The completeness and consistency of reporting of routine health data in Eastern Cape province have maintained good quality since the migration to DHIS2 in the year 2017. However, it was observed that many data element values were missing and some were not accounted for in the reported data in DHIS2. The low rate of data element reporting contributed to the reduction of the overall rate for data completeness. The district information cadres together with the DHIS2 system administrator must ensure that there is regular reviewing of data element groups assigned to the health facilities for data reporting. Also, the department should ensure that data elements are allocated to the appropriate datasets for easy data collection.

In addition, a slightly high rate of inconsistent data element values over time was found to be present across all eight districts. The inconsistency of data over time influenced the overall calculation of internal data consistency in the DHIS2. Also, the higher rate of missing data element values and data inconsistency over time indicated a concern with data quality in the DHIS2 for the ECDoH. The provincial and district information cadres should strengthen the institutionalisation of the data quality review and feedback mechanism in all tier levels of the health system. Continuous health information management training and capacity building for information cadres, health professionals, and managers to increase the level of engagement and utilisation of health data. Furthermore, there is a need to implement a programme to recognise health facilities and districts that are performing well, the award can be in the form of non-financial incentives such as awarding certificates to the best performers to encourage good practice and improved data quality in all tier levels of the health system.

Acknowledgements

This article is partially based on the author, S.S.T.'s Master's thesis entitled 'Evaluation of the data quality of routine health information system in the public health facilities in the Eastern Cape province' in the Department of Public Health at the University of the Western Cape, South Africa. The authors would also like to acknowledge the Eastern Cape Department of Health for their support.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Authors' contributions

S.S.T. is the student and executed the study and then compiled the first draft of the article. V.E.M. is the supervisor of the student and supervised the student in conceptualising and completing the study. V.E.M. revised the first draft of the article and finalised the final submission of the article.

Funding information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

The data that support the findings of this study are available from the corresponding author, V.E.M., upon reasonable request.

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency or that of the publisher. The authors are responsible for this article's results, findings and content.

References

Adane, A., Adege, T.M., Ahmed, M.M., Anteneh, H.A., Ayalew, E.S., Berhanu, D. et al., 2021, 'Routine health management information system data in Ethiopia: Consistency, trends, and challenges', *Global Health Action* 14(1), 1868961. https://doi.org/10.1080/16549716.2020.1868961

Agiraembabazi, G., Ogwal, J., Tashobya, C., Kananura, R.M., Boerma, T. & Waiswa, P., 2021, 'Can routine health facility data be used to monitor subnational coverage of maternal, newborn, and child health services in Uganda?', *BioMed Central Health Services Research* 21(suppl. 1), 1–10. https://doi.org/10.1186/s12913-021-06554-6

Amoakoh-Coleman, M., Kayode, G.A., Brown-Davies, C., Agyepong, I.A., Grobbee, D.E., Klipstein-Grobusch, K. et al., 2015, 'Completeness and accuracy of data transfer of routine maternal health services data in the greater Accra region', *BioMed Central Research Notes* 8(1), 114. https://doi.org/10.1186/s13104-015-1058-3

Andargoli, A.E., Scheepers, H., Rajendran, D. & Sohal, A., 2017, 'Health information systems evaluation frameworks: A systematic review', *International Journal of Medical Informatics* 97, 195–209. https://doi.org/10.1016/j.ijmedinf.2016.10.008

Aqil, A., Lippeveld, T. & Hozumi, D., 2009, 'PRISM framework: A paradigm shift for designing, strengthening and evaluating routine health information systems', *Health Policy and Planning* 24(3), 217–228. https://doi.org/10.1093/heapol/czp010

Barro, S.G., Dicko, A., De Lame, P.A. & Staccini, P., 2020, 'Health Metrics Network (HMN-WHO), a tool to assess the quality of the information system in Burkina hospitals', *European Federation for Medical Informatics (EFMI)* 270, 602–606. https://doi.org/10.3233/SHTI200231

Begum, T., Khan, S.M., Adamou, B., Ferdous, J., Parvez, M.M., Islam, M.S. et al., 2020, 'Perceptions and experiences with district health information system software to collect and utilize health data in Bangladesh: A qualitative exploratory study', *BioMed Central Health Services Research* 20(1), 1–13. https://doi.org/10.1186/s12913-020-05322-2

Bhattacharya, A.A., Allen, E., Umar, N., Audu, A., Felix, H., Schellenberg, J. et al., 2020', 'Improving the quality of routine maternal and newborn data captured in primary health facilities in Gombe State, Northeastern Nigeria: A before-and-after study', *British Medical Journal Open* 10(12), 1–10. https://doi.org/10.1136/bmjopen-2020-038174

Bhattacharya, A.A., Umar, N., Audu, A., Felix, H., Allen, E., Schellenberg, J.R.M. et al., 2019, 'Quality of routine facility data for monitoring priority maternal and newborn indicators in DHIS2: A case study from Gombe State, Nigeria', *PLoS One* 14(1), 1–21. https://doi.org/10.1371/journal.pone.0211265

Blödt, S., Kaiser, M., Adam, Y., Adami, S., Schultze, M., Müller-Nordhorn, J. et al., 2018, 'Understanding the role of health information in patients' experiences: Secondary analysis of qualitative narrative interviews with people diagnosed with cancer in Germany', *British Medical Journal Open* 8(3), 1–9. https://doi.org/10.1136/bmjopen-2017-019576

Chekol, A., Ketemaw, A., Endale, A. & Asemahagn, M.A., 2023, 'Data quality and associated factors of routine health information system among health centers of West Gojjam Zone, Northwest Ethiopia, 2021', *Frontiers in Health Services* 3, 1–12. https://doi.org/10.3389/frhs.2023.1059611

Dadzie, D., Boadu, R.O., Engmann, C.M. & Amma, N., 2021, 'Evaluation of neonatal mortality data completeness and accuracy in Ghana', *PLoS One*, 16(3), 1–12. https://doi.org/10.1371/journal.pone.0239049

Day, C. & Gray, A., 2017, 'Health and related indicators', in A Padarath & P Barron (eds.), *South African health rReview 2017*, Health Systems Trust, Durban, viewed 17 April 2022, from https://www.hst.org.za/publications/South%20African%20Health%20Reviews/HST%20SAHR%202017%20Web%20Version.pdf.

Dehnavieh, R., Haghdoost, A.A., Khosravi, A., Hoseinabadi, F., Rahimi, H., Poursheikhali, A. et al., 2019, 'The District Health Information System (DHIS2): A literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries', *Health Information Management Journal* 48(2), 62–75. https://doi.org/10.1177/1833358318777713

DHIS2, 2022, *About DHIS2*, viewed 30 July 2022, from https://dhis2.org/about/.

Eastern Cape Department of Health (ECDoH), 2020, *Annual performance plan 2020/21*, viewed n.d., from https://www.echealth.gov.za/index.php/document-library/annual-performance-plans?task=download.send&id=1246&catid=7&m=0.

Eastern Cape Department of Health (ECDoH), 2024, *Annual Performance Plan 2024/25*, viewed n.d., from https://www.echealth.gov.za/index.php/document-library/annual-performance-plans?task=download.send&id=4593&catid=7&m=0.

English, R., Masilela, T., Barron, P. & Schönfeldt, A., 2011, 'Health Information Systems in South Africa', in A Padarath & P. Barron (eds.), *South African Health Review 2011*, pp. 81–89, Health Systems Trust, Durban.

Farnham, A., Utzinger, J., Kulinkina, A.V. & Winkler, M.S., 2020, 'Using district health information to monitor sustainable development', *Bulletin of the World Health Organization* 98(1), 69–71. https://doi.org/10.2471/BLT.19.239970

Garavand, A., Mohseni, M., Asadi, H., Etemadi, M., Moradi-joo, M., Services, H. et al., 2016. 'Factors influencing the adoption of health information technologies: A systematic review', *Electronic Physician* 8(8), 2713–2718. https://doi.org/10.19082/2713

Garrib, A., Stoops, N., Mckenzie, A., Dlamini, L., Govender, T., Rohde, J. et al., 2008, 'An evaluation of the District Health Information System in rural South Africa', *South African Medical Journal* 98(7), 549–552.

Getachew, N., Erkalo, B. & Garedew, M.G., 2022, 'Data quality and associated factors in the health management information system at health centres in Shashogo district, Hadiya zone, Southern Ethiopia, 2021', *BioMed Central Medical Informatics and Decision Making* 22(1), 154. https://doi.org/10.1186/s12911-022-01898-3

Githinji, S., Oyando, R., Malinga, J., Ejersa, W., Soti, D., Rono, J. et al., 2017, 'Completeness of malaria indicator data reporting via the District Health Information Software 2 in Kenya, 2011–2015', *Malaria Journal* 16(1), 1–11. https://doi.org/10.1186/s12936-017-1973-y

Hagel, C., Paton, C., Mbevi, G. & English, M., 2020, 'Data for tracking SDGs: Challenges in capturing neonatal data from hospitals in Kenya', *British Medical Journal Global Health* 5(3), 1–10. https://doi.org/10.1136/bmjgh-2019-002108

HISP Team, 2016, *WebDHIS foundation: Facilitator guide*, viewed n.d., from https://www.hisp.org/?page_id=4506.

Hung, Y.W., Hoxha, K., Irwin, B.R., Law, M.R. & Grépin, K.A., 2020, 'Using routine health information data for research in low- And middle-income countries: A systematic review', *BioMed Central Health Services Research* 20(1), 1–15. https://doi.org/10.1186/s12913-020-05660-1

Kebede, M., Adeba, E., Adeba, E. & Chego, M., 2020, 'Evaluation of quality and use of health management information system in primary health care units of east Wollega zone, Oromia regional state, Ethiopia', *BioMed Central Medical Informatics and Decision Making* 20(1), 1–12. https://doi.org/10.1186/s12911-020-01148-4

Kiberu, V.M., Matovu, J.K., Makumbi, F., Kyozira, C., Mukooyo, E. & Wanyenze, R.K., 2014, 'Strengthening district-based health reporting through the district health management information software system: The Ugandan experience', *BMC Medical Informatics and Decision Making* 14(1), 1–9. https://doi.org/10.1186/1472-6947-14-40

Lasim, O.U., Ansah, E.W. & Apaak, D., 2022, 'Maternal and child health data quality in health care facilities at the Cape Coast', *BioMed Central Health Services Research* 22, 1102. https://doi.org/10.1186/s12913-022-08449-6

Lee, J., Lynch, C.A., Hashiguchi, L.O., Snow, R.W., Herz, N.D., Webster, J. et al., 2021, 'Interventions to improve district-level routine health data in low-income and middle-income countries: A systematic review', *British Medical Journal Global Health* 6(6), e004223. https://doi.org/10.1136/bmjgh-2020-004223

Lemma, S., Janson, A., Persson, L.Å., Wickremasinhe, D. & Källestål, C., 2020, 'Improving quality and use of routine health information system data in low- And middle-income countries: A scoping review', *PLoS One* 15(10), 1–17. https://doi.org/10.1371/journal.pone.0239683

Lutge, E., Moodley, N., Tefera, A., Sartorius, B., Hardcastle, T. & Clarke, D., 2016, 'A hospital based surveillance system to assess the burden of trauma in KwaZulu-Natal Province South Africa', *Injury International Journal of the Care of the Injured* 47(1), 135–140. https://doi.org/10.1016/j.injury.2015.08.020

Maïga, A., Jiwani, S.S., Mutua, M.K., Porth, T.A., Taylor, C.M., Asiki, G. et al., 2019, 'Generating statistics from health facility data: The state of routine health information systems in Eastern and Southern Africa', *British Medical Journal Global Health* 4(5), 1–9. https://doi.org/10.1136/bmjgh-2019-001849

Malembaka, E.B., Altare, C., Bigirinama, R.N., Bisimwa, G. & Banywesize, R., 2021, 'The use of health facility data to assess the effects of armed conflicts on maternal and child health: Experience from the Kivu, DR Congo', *BioMed Central Health Services Research* 7, 1–11. https://doi.org/10.1186/s12913-021-06143-7

Manoj, S., Wijekoon, A., Dharmawardhana, M., Wijesuriya, D., Rodrigo, S., Hewapathirana, R. et al., 2012, 'Implementation of District Health Information Software 2 (DHIS2) in Sri Lanka', *Sri Lanka Journal of Bio-Medical Informatics* 3(4), 109. https://doi.org/10.4038/sljbmi.v3i4.5431

Mphatswe, W., Mate, K. S., Bennett, B., Ngidi, H., Reddy, J., Barker, P.M. et al., 2012, 'Amélioration des informations sur la santé publique: Une intervention sur la qualité des données dans le Kwazulu-Natal (afrique du sud)', *Bulletin of the World Health Organization* 90(3), 176–182. https://doi.org/10.2471/BLT.11.092759

Muhoza, P., Tine, R., Faye, A., Gaye, I., Zeger, S.L., Diaw, A. et al., 2022, 'A data quality assessment of the first four years of malaria reporting in the Senegal DHIS2, 2014–2017', *BioMed Central Health Services Research* 22(1), 1–13. https://doi.org/10.1186/s12913-021-07364-6

National Department of Health (NDoH), 2011, *District Health Management Information System Policy*, National Department of Health, Pretoria.

National Department of Health (NDoH), 2012, *District health management information system standard operating procedures: Facility Level*, National Department of Health, Pretoria.

National Department of Health (NDoH), 2013, *District health management information system standard operating procedures: District level*, National Department of Health, Pretoria.

National Department of Health (NDoH), 2017a, *Annual Report 2016/17*, viewed n.d., from https://www.gov.za/sites/default/files/gcis_document/201710/national-department-health-annual-report-2016-2017a.pdf

National Department of Health (NDoH), 2017b, *National indicator dataset 2017: Data elements*, National Department of Health, Pretoria.

Nicol, E., Dudley, L. & Bradshaw, D., 2016, 'Assessing the quality of routine data for the prevention of mother-to-child transmission of HIV: An analytical observational study in two health districts with high HIV prevalence in South Africa', *International Journal of Medical Informatics* 95, 60–70. https://doi.org/10.1016/j.ijmedinf.2016.09.006

Nsiah, R., Takramah, W., Anum-Doku, S., Avagu, R. & Nyarko, D., 2022, 'Assessment of neonatal mortalities and stillbirths data quality in Offinso North District of Ghana', *International Journal of Scientific Research and Management* 10(01), 508–519. https://doi.org/10.18535/ijsrm/v10i01.mp01

Reynolds, E., Martel, L.D., Bah, M.O., Bah, M., Bah, M.B., Boubacar, B. et al., 2022, 'Implementation of DHIS2 for disease Surveillance in Guinea: 2015–2020', *Frontiers in Public Health* 9(9), 761196. https://doi.org/10.3389/fpubh.2021.761196

Roomaney, R.A., Wyk, V.P., Awotiwon, O.F., Nicol, E., Joubert, D., Bradshaw, D. et al. 2017, 'Review availability and quality of routine morbidity data: Review of studies in South Africa', *Journal of the American Medical Informatics Association* 24(e1), 194–206. https://doi.org/10.1093/jamia/ocw075

Shama, A.T., Roba, H.S., Abaerei, A.A., Gebremeskel, T.G. & Baraki, N., 2021, 'Assessment of quality of routine health information system data and associated factors among departments in public health facilities of Harari region, Ethiopia', *BioMed Central Medical Informatics and Decision Making* 21, 287. https://doi.org/10.1186/s12911-021-01651-2

Solomon, M., Addise, M., Tassew, B., Balcha, B. & Abebe, A., 2021, 'Data quality assessment and associated factors in the health management information system among health centers of Southern Ethiopia', *PLoS One* 16(10), e0255949. https://doi.org/10.1371/journal.pone.0255949

Statistics South Africa, 2024, *Mid-year population estimates2024*, viewed n.d., from https://www.statssa.gov.za/publications/P0302/P03022024.pdf

Tilahun, B., Teklu, A., Mancuso, A., Abebaw, Z., Dessie, K. & Zegeye, D., 2018, 'How can the use of data within the immunisation programme be increased to improve data quality and ensure greater accountability in the health system? A protocol for implementation science study', *Health Research Policy and Systems* 16(1), 1–6. https://doi.org/10.1186/s12961-018-0312-2

Tlale, L.B., Morake, B., Lesetedi, O., Maribe, L., Masweu, M., Faye, C. et al., 2019, 'Data quality self-assessment of child health and sexual reproductive health indicators in Botswana, 2016–2017', *PLoS One* 14(8), e0220313. https://doi.org/10.1371/journal.pone.0220313

Tshabalala, A. & Taylor, M., 2016, 'Disaggregated data to improve child health outcomes', *African Journal of Primary Health Care & Family Medicine* 8(1), 1–7. https://doi.org/10.4102/phcfm.v8i1.1221

Wagenaar, B.H., Sherr, K., Fernandes, Q. & Wagenaar, A.C., 2016, 'Using routine health information systems for well-designed health evaluations in low- and middle-income countries', *Health Policy and Planning* 31(1), 129–135. https://doi.org/10.1093/heapol/czv029

Wikipedia, 2024, *List of municipalities in the Eastern Cape*, viewed n.d., from https://en.wikipedia.org/wiki/List_of_municipalities_in_the_Eastern_Cape.

World Health Organization (WHO), 2007, *Everybody business: Strengthening health systems to improve health outcomes: WHO's framework for action*, World Health Organization, Geneva.

World Health Organization (WHO), 2017a, *Data quality review: A toolkit for facility data quality assessment: Module 1: Framework and metrics*, viewed n.d., from https://apps.who.int/iris/handle/10665/259224

World Health Organization (WHO), 2017b, *Data quality review: A toolkit for facility data quality assessment. Module 2: Desk review of data quality*, viewed n.d., from https://apps.who.int/iris/handle/10665/259225

Wude, H., Woldie, M., Melese, D., Lolaso, T. & Balcha, B., 2020, 'Utilization of routine health information and associated factors among health workers in Hadiya Zone, Southern Ethiopia', *PLoS One* 15(5), 1–11. https://doi.org/10.1371/journal.pone.0233092

Youssef, D., Yaghi, A., Jouny, A., Abou-Abbas, L., Chammaa, H. & Ghosn, N., 2022, 'Converting the existing disease surveillance from a paper-based to an electronic-based system using district health information system (DHIS-2) for real-time information: The Lebanese experience', *BioMed Central Health Services Research* 22(1), 395. https://doi.org/10.1186/s12913-022-07773-1

Zhou, L.L., Owusu-Marfo, J., Antwi, H.A., Antwi, M.O., Doris, A., Kachie, T. et al., 2019, 'Assessment of the social influence and facilitating conditions that support nurses' adoption of hospital electronic information management systems (HEIMS) in Ghana using the unified theory of acceptance and use of technology (UTAUT) mode', *BMC Medical Informatics and Decision Making* 19, 230. https://doi.org/10.1186/s12911-019-0956-z

Appendices start on the next page →

Appendix 1

TABLE 1-A1: Monthly routine data.

District Name:													
Financial Year													
Health Facility	Data Element Group	Data elements	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12

Appendix 2

TABLE 1-A2: Validation analysis report.

District Name:						
Financial Year						
Health Facility	Period	Validation rule	Importance	Left side description	Value	Operator	Value	Right side description

Appendix 3

TABLE 1-A3: Outliers and missing data report.

Province Name:																	
Financial Year																	
Health Facility	Data elements	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7	Month 8	Month 9	Month 10	Month 11	Month 12	Max Z score	Max modified Z score	Gap weight	Outlier weight	Total weight