Towards High-quality, well-understood, auditable data in Gauteng's Primary Healthcare Facilities

Data Quality Baseline Assessment Report

Data Quality Assessment of Gauteng Department of Health Facilities for the period July, August and September 2015

12 September 2016

GAUTENG PROVINCE
HEALTH
REPUBLIC OF SOUTH AFRICA

This report has been created in collaboration with the Gauteng Department of Health supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation–Strategic Information for South Africa associate award AID-674-LA-13-00005. MEASURE Evaluation–SIFSA is implemented by the Carolina Population Center at the University of North Carolina at Chapel Hill, in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of PEPFAR, USAID, or the United States government.

Contents

List of Tables 5

List of Figures 6

List of Abbreviations 8

Acknowledgements 9

Executive Summary 10

The Assessment 11

Summary of Key Findings 12

Conclusion 15

Introduction 16

1 Purpose and Objectives 16

1.1 Overview of the Report Structure 17

1.2 Background 17

1.3 Examples of Data Improvement Interventions 23

2 Methodology—Primary Health Care Facilities 25

2.1 Research Questions 25

2.2 Data Collection Tools 26

2.3 Field-Site Selection 28

2.4 Data Collection Strategy 28

2.5 Analysis Strategy 29

2.6 Response Rates 30

2.7 Data Limitations 31

3 Baseline Findings at Health Facilities 34

3.1 Recount Results 34

3.2 Level of Technical Resourcing 60

3.3 Human Resource Capacity at Health Facilities to Carry out Data Management 71

3.4 Implementation of DHMIS Policy Status 88

4 Summary of Baseline Findings 99

5 Recommendations 104

6 References 108

7 Annexure A: Assessment Tools 110

8 Annexure B: PHC Recount Tables per District and Sub-District 110

9 Annexure C: PHC Systems and Staffing Tables per District and Sub-District 110

10 Annexure D: Hospital Recount Tables 110

11 Annexure E: Hospital Systems and Staffing Findings 110

List of Tables

Table 1: Health indicators and data elements included in the recount tool 27
Table 2: Number of PHC Facilities successfully reached during the Data Quality Baseline 30
Table 3: Number of PHC facilities in each of the data quality categories—PHC headcount 35
Table 3: Number of assessed PHC facilities in each size or headcount category 51
Table 4: Under- and over-reporting on select health indicators at large PHC facilities 53
Table 5: Types of reasons provided at PHC facilities for registries or all documentation not being provided for PHC headcount 55
Table 6: Types of reasons provided by the data-capturer why the summary and the count data do not match PHC headcount 56
Table 7: Technology resources table (included in systems tool) 61
Table 8: Facilities with no computers 62
Table 9: List of clinics with only one computer that is not fully functional 63
Table 10: Number of facilities with fully functional computers per 5000 PHC headcount 64
Table 11: Percentage of facilities with fully functional computers per PHC headcount 65
Table 12: List of sub-districts with a high percentage of facilities 65
Table 13: Number of computers with specific dataset 67
Table 14: DHIS versions at facilities 68
Table 15: Total number of clinic and administrative employees across 286 sampled Gauteng PHC facilities 73
Table 16: Median number of administrative employees in Gauteng PHC facilities 73
Table 17: Level of data-related training amongst 636 employees at 286 Gauteng PHC facilities 75
Table 18: Total and median number of consulting rooms across all sampled facilities 78
Table 19: Percentage of work time devoted to the following tasks 80
Table 20: Time burden: Percentage of work time devoted to the following tasks 86
Table 21: Number of clinics: Spontaneous vs. prompted mention by facility managers of tools and resources used 89
Table 22: Data collection tools that were observed at PHC facilities at the time of the audit 96

List of Figures

Figure 1: Paper and electronic instances of data elements. 33
Figure 2: Data quality issues at various steps in the data consolidation process. 33
Figure 3: Percentage of facilities with mostly accurate data on facility headcount 36
Figure 4: Percentage of facilities with data quality issues on facility headcount 37
Figure 5: Percentage of facilities with accurate data for MCWH data elements 38
Figure 6: Percentage of facilities with data quality issues for MCWH data elements 39
Figure 7: Data Quality for TB elements (Percentage of facilities) 40
Figure 8: Percentage of facilities with mostly accurate data for HIV elements 41
Figure 9: Percentage of facilities with data quality issues for HIV elements 42
Figure 10: Data Quality for NCD and Headcount elements (Percentage of facilities) 43
Figure 11: Percentage of facilities with mostly accurate data by district - PHC headcount 45
Figure 12: Percentage of facilities with data quality issues by district – PHC Headcount 45
Figure 13: Percentage of facilities with data quality issues by district – Immunised 46
Figure 14: Percentage of facilities with mostly accurate data by district – Immunised 47
Figure 15: Data Quality for TB suspect 5 years and older initiated on treatment for September 48
Figure 16: Data Quality for Antenatal Client Initiated on ART for September 49
Figure 17: Percentage of facilities with mostly accurate data by district – Hypertension 50
Figure 18: Percentage of facilities with data quality issues by district – Hypertension 50
Figure 19: Percentage of facilities with mostly accurate data by facility size 52
Figure 20: Number of computers per facility 62
Figure 21: Number of computers not fully functional 63
Figure 22: Percentage of computers used for multiple purposes 66
Figure 23: Percentage of computers used by more than one person 67
Figure 24: Status of Antivirus software 68
Figure 25: Computer backup status 69
Figure 26: Internet connectivity of computers 70
Figure 27: Level of data-related training amongst 636 employees at 74
Figure 28: Education level of respondents in Gauteng clinics 76
Figure 29: Number of clinics where at least one interviewed person has a post graduate degree 76
Figure 30: Types of employment 77
Figure 31: Data systems most used 82
Figure 32: Source of evidence about routine discussion on data quality per month 84
Figure 33: Source of evidence about routine discussion on data quality by district 85
Figure 34: Percentage clinics: Spontaneous vs. prompted mention by facility managers of tools and resources used 89
Figure 35: Tools and resources used to guide data collection and management at the facilities: Percentage of facilities: Spontaneous mention vs. prompted mention by senior data person 90
Figure 36: Percentage of facilities where fieldworkers observed the use of data management tools and resources 91

Figure 37: The frequency of data verification across facilities surveyed by the data personnel, facility manager, sub-district and other personnel 92
Figure 38: Availability of PHC Registers at facilities during the months of July, August and September 2015 94
Figure 39: Availability of registers at facilities during the months of July, August and September 94
Figure 40: Availability of TB screening Registers at facilities across districts 95
Figure 41: The extent to which data collection tools are maintained in accordance to SOP requirements across facilities 96
Figure 42: The extent to which calculations on data collection tools were correct 97

List of Abbreviations

ANC	Antenatal care
ARV	Antiretrovirals
CSIR	Centre for Scientific and Industrial Research
DHMIS	District Health Information Management System
DOH	Department of Health
ETR.NET	ETR.Net is an electronic tuberculosis register designed for TB/HIV surveillance, program monitoring and evaluation. The system consists of a database which is accessed by the user via the user-friendly software interface, custom developed for the Microsoft Windows environment
GDOH	Gauteng Department of Health
HIS	Health Information Systems
HIV	Human Immunodeficiency Virus
ICTs	Information and communication technology
IMR	Infant Mortality Rate
MCH	Maternal and Child Health
MCWH	Maternal, Child and Women's Health
MDR-TB	Multi-drug-resistant tuberculosis (also known as Vank's disease) is defined as a form of TB infection caused by bacteria that are resistant to treatment with at least two of the most powerful first-line anti-TB drugs, isoniazid (INH) and rifampicin (RMP).
NCD	Non communicable Diseases
NDOH	National Department of Health
NHA	National Health Authority
NHI	National Health Insurance
NHISSA	National Health Information Systems Committee
NGO	Non- Government Organisation
NIDS	National Indicator data set
NSDA	Negotiated Service Delivery Agreement
PCR	Polymerase Chain Reaction, a test method used to detect the genes of the virus
PEPFAR	U.S. President's Emergency Plan for AIDS Relief
PHC facilities	Primary health care facilities
PIDS	Provincial Indicator data set
SASQAF	South African Statistical Quality Assessment Framework
SIFSA	Strategic Information for South Africa
SOP	Standard Operating Procedure
TB	Tuberculosis
UNICEF	United Nations Children s Fund
WHO	World Health Organization

Acknowledgements

A special word of thanks to the following individuals who gave time and valuable inputs to make this study possible:

- Executive Management of Gauteng Department of Health (GDOH)
- Officials in particular Health Information Management and Monitoring & Evaluation personnel at districts and health facility level
- District, Sub-district and Facility Managers for supporting the fieldwork
- The GDOH team leading this study – Tsholofelo Adelekan, Jacob Kgalabi Ngako
- The fieldwork team from Benita Williams Evaluation Consultants

Executive Summary

This *Data Quality Assessment Baseline Report* for 2015 reviews the Gauteng Provincial Department of Health's (DOH) current data quality performance and prospects, and examines in-depth the determinants for fostering good data management in the province. The report also captures key milestones reached by Gauteng Primary Health Care (PHC) facilities in 2015 towards the realisation of the national Ten Point Plan to improve the public health information system. While more effort is needed to guarantee high quality health information, PHC facilities have made significant progress in the following areas:

- Growing awareness around national data management policies and processes within the Gauteng Department of Health
- Rationalizing health registers;
- Employing essential data-personnel;
- Providing employees with data-related training;
- Having technical resources and hardware;
- Digitizing primary health data;
- Streamlining the data capturing process; and
- Integrating different electronic data systems.

Although this report focuses on Gauteng Health facilities, districts, and sub-districts, it simultaneously reflects on larger developments across a country that is carefully navigating the strong currents of a new Information and Communication Technology (ICT)-enabled and knowledge-based economy. In this context, provincial departments of health are striving to provide essential needs-based services through re-engineering aspects of primary health care, and at the same time, developing strategies for digitizing and integrating health information into a single system.

A strong institutional culture that supports demand for quality data which are used in evidence-based decision making is vital for improving the health information system, and its subsequent health outcomes. Successfully advancing such a culture, and strengthening the information system performance, is dependent upon a trifecta of *technical, organisational*, and *behavioural* determinants. These determinants provide the foundation for this report.

The Gauteng provincial DOH has affirmed its commitment to building an information-oriented culture by calling for an independent assessment of the quality of its routinely collected health data. This report shows that, despite certain shortfalls in data quality and usage, the province is generally meeting its targets in terms of making health information readily accessible to the users

who need it most. These include managers at the national, provincial, district and facility levels of the health system.

The Assessment

The assessment reviewed: (a) accuracy of data recording, (b) technical resources capacity, (c) human resources and systems capacity, and (d) implementation of DHMIS policy.

Assessment strategy

At the start of 2015, representatives from the Gauteng DOH, with support from MEASURE Evaluation SIFSA, designed a holistic assessment of the technical, organisational, and behavioural issues and elements related to data management and data quality at the provinces' Healthcare facilities. An assessment team conducted at 287 PHC facilities and 7 Hospitals between November 2015 and February 2016. To check data quality, the assessment team recounted data on selected health indicators for the period July to September 2015 and cross-referenced these data with District Health Information System (DHIS) records. To review data management, the team conducted interviews with facility managers and senior data officials, which yielded information on data use, data capturing procedures, and systems and resource capacity.

Key findings

The key findings are structured into two parts: Findings around data quality which emerged from a recount, and findings around the *trifecta*: technical, organisational and behavioural determinants of data quality.

The recount process systematically identified data quality issues which is reported first. In this section findings about the accuracy of data are based on a recount of selected DHIS data for a three-month period.

Using the PRISM Framework,[1] the assessment team next explored the reasons for these issues through the use of a systems and staffing interview.

- *Technical.* Technical determinants relate to established processes, norms, and procedures regarding data. This part presents the technical determinants of data quality and contains a review of the contextual and policy background. Germane literature, key informant interviews, and relevant findings from the assessment provided evidence for these findings.

1 Developed under MEAUSURE evaluation, the PRISM framework consists of four tools to assess Routine Health Information System (RHIS) performance; identify technical, behavioral and organizational factors that affect RHIS; aid in designing priority interventions to improve performance; and improve quality and use of routine health data, http://www.cpc.unc.edu/measure/resources/webinars/publications/ms-11-46-d

- *Organisational.* Organisational determinants refer to ways in which leadership supports having the necessary information for data-informed decisions or policy making, and facilitates the process of, and resources for, accessing this data. This part addresses the organisational determinants of data quality.

- *Behavioural.* Behavioural determinants refer to the motivations, values, attitudes, collaboration, knowledge, skills, and competence amongst those who produce and use data to improve performance, and for data-informed decision making. This assessment focused mainly on and therefore presents data for the latter three determinants of knowledge, skills and competence. Evidence for this section are drawn from the systems and staffing survey.

Summary of Key Findings

Data Quality

- *Data quality challenges*

A recount of selected 2015 primary health data revealed that the data quality is lacking in several respects. The principal concerns are over-reporting and poor record keeping, whilst missing data, and under reporting, also occurred in a minority of facilities. Although the incidence of these problems differed from district-to-district, month-to-month and indicator-to-indicator, 53% of the assessed PHC facilities over-reported headcount data in September, 5% had missing data, 6% under reported, and approximately 37% had irregular ticks in the paper records. PHC facilities in the City of Johannesburg struggled with missing data more than others (10% missing on PHC headcount for September), while more sites in Sedibeng underreported their results on indicators such as the number of Antenatal Clients Initiated on treatment (42%). Missing data were a problem in smaller clinics (35% of those facilities with fewer than 1800 clients had missing data on the data element related to the number of antenatal clients initiated on ART), while over reporting was predominantly evident in large clinics (10 out of the 14 clinics with more than 2100 clients had a significant over report for PHC headcount in September). Data quality was also problematic for healthcare issues that required effective clinical laboratory interface or long-term tracking (e.g. diabetes).

Problems with data quality were not specific to any particular heath indicator. Although the quality for facility headcount, maternal health indicators, HIV screening, and hypertension was generally better, the quality of HIV, TB and diabetes data were generally the poorest.

Technical determinants of data quality

Established national-level policies and processes are in place that aim to ensure that health information is available and accessible for informing key management practices. Interview data indicate that PHC facility managers and senior data officers are knowledgeable about many of these policies and procedural guidelines for the data management processes. Quantitative data (the system and staffing survey) indicate that the DHMIS policy and Standard Operating Procedures is implemented at about 65% of PHC facilities. The data further indicate that the PHC facility managers use these guidelines, with interview data providing examples of thoughtful adaptations and adjustments being made at facilities which resulted in data being captured.

Other positive results from efforts to improve data quality are also clearly evident. For instance, the initiative that began in August 2015 to rationalise the health indicator registers yielded immediate positive results in the September data; September data were generally more accurate and complete than that of previous months. The assessment team also identified the successful implementation of policies and processes, such as in the use of tools and resources to guide data collection; the frequency of data verification; the availability of tools, registers and forms at health facilities; and the maintenance of data collection tools in line with standard operating procedures.

Organisational determinants of data quality

The Gauteng DOH is exemplary—leadership at provincial, district and facility levels generally understand the importance of, and need for, health information. Accommodations have been made to provide basic levels of essential technical, electronic and human resources to facilitate data recording, capturing, collation and management processes.

Common reasons for data quality challenges

- Human Resources

Reasons for most data quality problems encountered during the recount were related to poor record keeping which is often linked to human resource challenges. PHC clinics had appropriately appointed employees dedicated to data-related tasks, but there were shortfalls in the number of employees at many clinics. Usually data capturers, administrative clerks, facility managers, and volunteers or intern clerks were involved in the routine data capturing and processing. More than three quarters of the interviewed officials had data quality training on DHIS and TIER, and almost 80% were employed on a full-time basis. At a few facilities, this task fell to clinical or support personnel who were not necessarily trained in data processing, and had other competing duties. The worst case was identified at the Tshwane Metro Municipality which had twenty clinics with no data capturers or dedicated administrative staff.

- Data Capturing Process

The data capturing process provided challenges at various steps. Health data are first captured on paper registers. Challenges with improper completion of paper registers included ticks not properly done in registers and registers having illegible entries or missing information – this was a problem at more than a third of assessed facilities. These data are then manually summarised on summary sheets where errors occurred frequently. Data are then entered manually into an electronic database. During the process of translating data from paper to computer, more data capturing and checking issues emerged. The recount results showed that typically, headcount registers and data collected for MCWH indicators had the most data quality issues.

- Tools and resources

PHC facilities (71%) generally reported using the relevant tools and resources prescribed by the DHMIS policy for data collection. However, official registers were not used in about 5% of facilities, which compromises the overall quality of data. The supply of health registers was not consistent and negatively influenced data quality efforts. For instance, roughly 60% of facilities ran out of the PHC daily register in July 2015, and 12% of facilities ran out of the PHC headcount and comprehensive registers in August and September. Further, evidence showed challenges with the supply of TB screening registers to facilities.

- ICT

On average, there were two fully functional computers per facility for data-related tasks. This suggested that most Gauteng PHC facilities were reasonably equipped with the necessary computers needed to record, capture, and digitally generate summaries of health-related data. However, challenges existed. Almost 80% of computers at PHC facilities are shared by more than one primary user, and used to capture numerous datasets, which created barriers to accessing computers for data entry. Software issues also existed, with different versions of the DHIS database running in different facilities, and only basic antivirus software installed, leaving the system open to potential viruses. Only 46% of sites surveyed indicated that they backed up their data weekly, leaving data vulnerable to loss or corruption.

Behavioural determinants of data quality

The Systems and Staffing survey suggested that the majority of employees responsible for data-related tasks have the necessary basic skills and education. Specifically, approximately 73% had received DHIS or TIER training, and 36% had received ETR training. Also, most brought experience managing data, as 83% had worked at the clinic for a year or more. However, not all the employees interviewed had been trained on the DHMIS policy and standards - fewer than 55% had received formal training on DHMIS policy, NIDS and PIDS.

While the survey results show that the data personnel were capable of maintaining the collection tools in line with DHMIS standard operating procedures, the challenge was in having data

collectors and senior data officers consistently double checking the accuracy of the information and calculations, minimising missing data, and correctly signing-off on the data and the tools at approximately 20% of the facilities.

Evidence show that a culture of concern for the integrity of information is starting to emerge in PHC clinics, with many implementing practices of reviewing data. For example, many respondents noted routine discussions took place regarding data processing and data quality. Nevertheless, data use at the facilities is limited. Reasons for the lack of use warrant further exploration, as it likely undergirds the challenges identified with data quality.

Conclusion

This *Data Quality Baseline Assessment Report* for Gauteng PHC facilities provides a foundation for further examination of how data are captured, processed, managed and used at all levels of the public health system. The assessment results offer a clear baseline and understanding of the data quality for select health indicators, which can be used to inform provincial data quality improvement plans. Gauteng PHC facilities can, and are, capturing health data as required. Nevertheless, more effort is needed to build capacity for ensuring that data quality standards are met.

This report also provides an opportunity for other provincial departments of health to learn from Gauteng's experiences. Managers and policymakers, as well as stakeholder partners, can draw upon the findings and the success-cases of PHC facilities when determining priorities and approaches for future interventions—such as the improvement of information legislation, the creation of incentives for achieving excellent data quality, and increasing facilities access to computers and the Internet.

This study shows that Gauteng facilities are making progress in cultivating a culture of respect for data quality, and towards improving the type and quantity of health data that are digitally recorded and electronically available. This is a positive step in the direction towards strengthening the public health information system.

Introduction

In an era of management by numbers, Ted Friedman's remark captures the motivation underlying this report well:

> Data is useful. High-quality, well-understood, auditable data is priceless.
>
> – Ted Friedman, Gartner

This *Data Quality Assessment Report* for 2015 reviews the Gauteng Provincial Department of Health's (DOH) current data quality performance and prospects, and examines in-depth the determinants for fostering good data management in the province. The report also captures key milestones achieved by Gauteng Primary Healthcare (PHC) facilities in 2015 towards the realisation of the national Ten Point Plan to improve the public health information system. It recommends some courses for correction, and provides a useful baseline against which future improvements may be measured. It is hoped that this work, together with that of other players in the health information field, may someday soon reveal the truly priceless nature of good quality health data.

1 Purpose and Objectives

The Gauteng Department of Health, with support from MEASURE Evaluation SIFSA (Strategic Information for South Africa) embarked upon a data quality baseline assessment in order to improve the data quality of its non-financial indicators. The assignment's purpose was to:

(a) Get a clear baseline understanding of the data quality for select indicators at Gauteng Province health facilities so as to inform provincial data quality improvement plans,

(b) Determine staffing capacity needs at each of the facilities,

(c) Conduct a data management system assessment for each facility, including but not limited to, rationalised register impact, and uptake and compliance with the District Health Management Information System (DHMIS) Policy and its associated standard operating procedures,

(d) Build the capacity of government and PEPFAR partner organisations to ensure health-information quality—i.e. by including government Department of Health officials in any training and data gathering activities related to this data quality baseline assessment

The need for a data quality baseline assessment was first articulated, at the Gauteng Department of Health's provincial quarterly review, held on 14 August 2015. At the time of the meeting, pressure was mounting to generate quality data and strategic performance information to guide health programmes. There was also an increased need to demonstrate how the use of current

resources was maximised and was producing health-related results. The Gauteng Department of Health had already received three consecutive years of qualified audit outcomes on non-financial data, but it was agreed that a baseline study could lead to an evidence based data improvement plan and might help improve public confidence in the Department's capabilities. It was envisioned that following the baseline data quality assessment, the Auditor General would conduct a sampled mid-term readiness-assessment, which in turn would be followed by post measurement of the select facilities towards the end of 2016 or in early 2017.

For MEASURE Evaluation SIFSA, the data assessment also provided an opportunity to make progress on a strategic area regarding an existing health systems strengthening intervention project in Gauteng. This intervention coincides with PEPFAR's "Focusing on Impact Plan." It includes four out of Gauteng's five districts that were selected for scale-up, and to receive full packages of care, treatment, prevention and health system strengthening support. This support is part of PEPFAR's geographical and population prioritisation drive. Gauteng was identified as the ideal "proof of concept" province in which to do a full data quality baseline assessment, whereby the focus would be on data verification, as well as system and health information system staffing components.

1.1 Overview of the Report Structure

This report documents the findings from the data quality assessment conducted at select Gauteng Primary Health Care facilities. It forwards recommendations for improving data quality that were developed by Gauteng Department of Health officials. The report contains a review of the contextual and policy background, an explanation of the methodology used in this assessment, and descriptions of the findings based on recount data and on the systems and staffing data. It concludes with a set of key observations and suggested recommendations for future data quality improvement strategies. Annexes with the findings per district and sub-district are provided together with the assessment tools, and information about the formulas used for key calculations.

1.2 Background

- ### Data Quality in Health Care Management

In 2015, Gauteng Department of Health officials began to discuss the need for a data quality baseline assessment. This conversation was born out of a wider desire to see improvements in the health system and in health outcomes. It was also a part of a broader national health promotion plan, a summary of which provides the context for this report.

Strengthening the health system, and subsequently health outcomes, is a top priority for South Africa's national agenda and it is reflected as a core objective in the Government's Negotiated

Service Delivery Agreement (or NSDA for short) (NDOH 2010). In pursuit of this endeavour, the availability of quality and timely data from the Health Information Systems (HIS) is foundational (WHO 2007). A robust HIS is prerequisite for the generation of good-quality data and is a core building block of any modern health system (English et al. 2011). The need to be responsive to emerging health issues also drives the requirement for better quality health-related information that is relevant, comprehensible, and timely (WHO 2005). Such information is used to inform decision-making during program monitoring, evaluation, planning, and improvement; during advocacy; and during policy development and review (Nutley and Reynolds 2013). This is in line with the District Health Management Information System (DHMIS) Policy, which calls for enhanced data analysis and use "for decision making, planning, and monitoring in the health sector" (Gauteng DOH 2013a: 13). However, users need to trust that the data they are utilising are accurate, complete, and up to date. Without this confidence, demand for data drops, evidence-informed decision making does not occur, and program efficiency and effectiveness suffers (Garrib et al. 2008, Gauteng DOH 2013b).

Historically, the South African HIS has been characterised by a lack of coordination and by fragmented manual systems that relied on too much paperwork and poor automation (Mphatswe et al. 2012, Nutley and Reynolds 2013). To address these problems, in 2000 the National Department of Health (NDOH) adopted an improved Health Management Information System in which routine data are collected at the district primary health-care level, and entered electronically into the District Health Information System's software (DHIS) (NDOH 2011a). With this routine health information system (RHIS), the DOH generates large volumes of information. Data on about 150 indicators from roughly 3,800 health facilities nationwide are collected on paper-based registers, tally sheets, and monthly data collation forms which are then captured into the DHIS (JSI, ESI and NDOH 2011, NDOH 2011a). This RHIS process also ensures that data generation, and entry of data into the DHIS software, are separated "for the application of data validation and for data analysis and utilisation at the clinic" (JSI, ESI and NDOH 2011, Schönfeldt 2014).

Data quality is addressed through functions within the DHIS software, such as validation rules to detect impossible or improbable relationships between data elements; and by incorporating feedback mechanisms into the data collection process, such as supervisors checking data accuracy (Garrib et al. 2008). While these checks may exist in principle, in practice they fall short of ensuring quality data standards are met. Recent studies of the South African HIS report that the quality of the data is suboptimal, which hinders efforts to strengthen service delivery (Garrib et al. 2008, Mate et al. 2009). One review found that major quality defects primarily concern the completeness and accuracy of data: errors occurred roughly 9% of the time and complete data was available at the clinic level less than 60% of the time, and from the DHIS between 30-80% of the time (Mate et al. 2009). In 2013, then Auditor General, Terence Nombembe, stated that the national audit of performance information found that data are seldom used because they are not of a high quality (Auditor General 2010, Nombembe 2013). The DOH's own annual performance plan for 2014 also reveals known data limitations concerning new client treatment success of TB and TB MDR confirmed treatment initiation rates (NMOH 2014).

The SAHR 2013/2014 described Persistent Challenges relating to eHealth to include:

-The lack of a national eHealth strategy (until July 2012);

Widely differing levels of eHealth maturity across and within provinces;

-A large number of disparate systems between which there is little or no inter-operability and communication;

-Inequity of eHealth services provided and expenditure on eHealth across Provincial and National Departments of Health;

-Expensive broadband connectivity;

-The absence of a national master patient index;

-The absence of a national unique identification system of patients; and

-Limited capacity within the public sector for implementation.

The growing awareness amongst various healthcare stakeholders of the limitations to data collection and data use begs the question of why data quality remains a persistent issue in South Africa's HIS. Research points to the combination of numerous factors. Typically, the primary reasons data quality falls short of expectations have to do with shortcomings in human resources, the data collection process, technical resources, and skill or comprehension about data use. These aspects of concern are highlighted below.

Human resources

- Shortage of dedicated health information personnel or permanent government information technology officers (Auditor General 2010)
- Insufficiently adequate and on-going training at different levels of the DHIS process (NDoH 2014)

- Roles and responsibilities of HIS staff are often not standardised or clearly defined (Loveday, Smith and Monticelli 2006)

Data collection processes

- Late submission of reports (Sibuyi 2014)
- Incomplete or missing data (Mate et al. 2009)
- Errors in accuracy of data (Mate et al. 2009)
- Data collection tools not used as intended (Garrib et al. 2008)
- Multiple steps in the collection process increase risk of data capture error (JSI, ESI and NDOH 2011)
- Lack of internal quality control—tendency to rely on auditors and external evaluators to identify errors (Auditor General 2010, Mate et al. 2009)
- Frequent changes in the names and definitions of the data elements used for monitoring (Mphatswe et al. 2012)
- High data collection burden, in part because of the low trust in the quality of DHIS data, which causes multiple parties to collect their own information—duplicating efforts instead of strengthening existing systems (Garrib et al. 2008, Mphatswe et al. 2012)

Technical hardware and software

- Chronic shortage of access to computers, printers, fax lines and the Internet connection, particularly at the upper district and provincial level (JSI, ESI and NDOH 2011)
- Limited or no software version-control—more than one third of districts using out-dated DHIS software (JSI, ESI and NDOH 2011, Vital Wave Consulting 2009)
- Poor integration between fragmented, different electronic systems (Vital Wave Consulting 2009)
- Inadequate technical support at all levels in the health system (English et al. 2011)

Skill and comprehension of data use

- Low awareness amongst those who generate and input data about the importance of data quality because they do not use the data—i.e. there is a disconnect between data producers and data users (Mate et al. 2009, Nutley and Reynolds 2013, Sibuyi 2014)
- Two thirds of upper district and provincial level managers have had no training in the use of information for management purposes (JSI, ESI and NDOH 2011)

This status of data quality limitations was confirmed in 2013 during a rapid assessment conducted by the MEASURE Evaluation Strategic Information for South Africa Project (MEval–SIFSA). Results of this assessment showed that the main causes of limited demand and

under-utilisation of data were: data use is not part of clinicians' in-service training; managers and clinicians focus on service provision or "pushing queues" and they consider data an extra burden/diversion from core business; data reporting is erroneously considered data use; and managers, clinicians, and other health care workers lack knowledge, skills, and confidence in calculating targets for indicators, data analysis, interpretation, and use (MEval and SIFSA 2015).

Although data quality remains an issue, research has also found that the DHIS is well integrated into the clinic routine, that there is awareness about the information system, and that it has strong district management and national leadership support (Mate et al. 2009). Furthermore, numerous efforts to support the collection, integration, and use of quality data are part of the broader national strategy to improve health systems outcomes. This enabling, strategic framework is discussed in the following section.

- **Strategic Context**

South Africa has made many gains in establishing and implementing a post-1994 HIS (English et al. 2011). This is accompanied by a re-engineering of the Primary Health Care system, which decentralises management to the district level. The DHIS collects crucial health service delivery data at this level in support of decentralised management. Data at this the primary level permit district and facility managers to make decisions about their service delivery based on locally available information (Mate et al. 2009). Thus, the availability of quality data at all levels of the DOH is vital to attaining the outcomes set forth by government's 2010 Service Delivery Agreement.

As early as 2001, commitments were made to improve the functionality of information systems. Government released a *Public Service Regulation* framework that requires departments to put in place mechanisms to appropriately manage and develop information management, including an information plan and the supporting infrastructure (Public Service Regulation Act 2001). In 2003 legislation was passed concerning the governance of health information, which assigns stewardship to the National Department of Health and which outlines the structure for use of information in the health sector. In particular, section 74 of the *National Health Act* of 2003 charges the Minister of Health with responsibility for eHealth systems in pursuit of health outcomes, including data quality, and facilitating and coordinating health information at all levels of government (see Republic of South Africa 2001 and 2003).

Further laws governing official statistics (see the Statistics Act of 1999) and civil registration (see the Births and Deaths Registration Act of 1992) are in place, as are policies that provide guidance on improving data quality and the monitoring and evaluation of performance—found in the *Policy Framework for the Government-Wide Monitoring and Evaluation System* of 2007 (Republic of South Africa 1992, 1999 and 2007). Through the Public Audit Act of 1999 the Auditor General of South Africa has been constitutionally mandated with the responsibility for monitoring

accountability and governance in the public sector by auditing financial and performance data. Additionally, the Electronic Communications and Transactions Act of 2002 frames all information, communication and technology initiatives in the country by making provisions for the development of a five year national e-strategy that facilitates electronic transactions in the public interest (Republic of South Africa 2002).

Aside from supportive legislation, there is also strong national backing for the development of electronic health systems (eHealth) that will provide good-quality, reliable and timely evidence, useful for tracking and improving health service delivery. eHealth is a priority for the Presidential National Commission on Information Society and Development (Vital Wav Consulting 2009). The National DOH exercises its coordination and facilitation role through the National Health Information System Committee, which was established with the broad objective to ensure the availability of information for the management of health services, and to coordinate countrywide health information systems. The Minister, DOH, and National Health Information System Committee work together to leverage eHealth systems and to strengthen healthcare transformation. To this end, the National eHealth Strategy for 2012/13 to 2016/17 was developed with ten strategic priorities that address strengthening systems and processes for collecting and disseminating quality data (NDOH 2012).

Despite these various mandates for implementing a national HIS that ensures the availability and access to quality data, barriers to the establishment of a strong HIS persist. A 2009 study on ICTs in Health conducted by the National eHealth Steering Committee found that since 1994, the considerable yet uncoordinated investment in eHealth has yielded limited tangible benefits and return on investment (NDOH 2012). The Health Systems Trusts suggests that the key challenges to establishing a strong HIS and the adoption of a "culture of information" include: limits to legislation and policy, challenges in governance and leadership, problems with data sources and storage, and concerns about data management and feedback (English et al. 2011).

Poor quality data undermines not only the HIS but also the achievement of government's strategic NSDA objective of "A Long and Healthy Life for All South Africans" (NDOH 2010). Similarly it is a concerning issue that affects Gauteng Provincial DOH Strategic Plans. The annual District Health Barometer (DHB), which focuses specific attention on DHIS data quality, analysis, feedback and use at the provincial and district level found that at district level there were major data quality problems (Barron, Day and Monticelli 2007). The report shows that outliers and missing information were used to portray trends in a misleading way—low data quality negatively affected district reports on average length of stay, bed utilisation, immunization coverage, clinic supervision, proportion of antenatal clients tested for HIV by rural districts, Nevirapine uptake rate of pregnant HIV positive women, and many more (Barron, Day and Monticelli 2007). Another independent evaluation of the HIS found that "the culture of

information use" at clinics and sub-districts, which is essential to an information system having an impact at the local level, is very weak and needs much improvement (Garrib et al. 2008). The Gauteng DOH's 2013 Annual Performance report also explicitly recognised that "data collection and reporting remain a challenge, as data submission from facilities is often delayed and this results in under-reporting" (Gauteng DOH 2013a). Clearly, further training, support, and improvements in data collection and utilisation processes are required for the DHIS to function as intended.

1.3 Examples of Data Improvement Interventions

In spite of the varied issues limiting data quality, numerous strategic processes are in place to secure improvements. An evaluation of the HIS reveals five important steps towards improvement. First, the DOH acknowledges the existence of key HIS challenges and policy translation gaps, and is attempting to remedy these through the National Health Information Systems Committee (NHISSA), which is setting data quality standards (English et al. 2011). Second, in 2011 the District health information systems policy (DHMIS) was developed to provide an overarching regulatory framework in terms of the National Health Act of 2003 (NDOH 2011a). This policy aims to standardise implementation of the district HIS and creating uniformity across the country. Third, several data cleaning workshops have been held in each province to "ensure alignment between data submitted to provincial level and that stored at district level" (English et al. 2011). Fourth, NHISSA has put in place regular monitoring of progress towards compliance with Auditor-General requirements. Fifth, a Health Data Advisory and Co-ordination Committee has been established to assist in improving the quality of health outcomes data and to advise on NSDA indicators (English et al. 2011, NDOH 2011b).

Compared to other provinces, , Gauteng does well. Upon the recommendation of the Auditor General and various Portfolio Committees, the Gauteng DOH revealed in its 2015 Annual Report numerous efforts to improve the health information systems for efficient data capturing (Gauteng DOH 2015). One such effort is the implementation of the Electronic Health Record, which provides a paperless system at hospitals and clinics in order to seamlessly integrate all the systems used in a hospital environment. In addition, by the end of 2015, the MEDICOM software upgrade was installed in eight institutions, and all hospitals running PAAB software had new Health Information System installed. Scanning of medical records to convert them into electronic documents commenced in January 2015 for all hospitals in order to Gauteng DOH 2015).

Furthermore, in 2015 the Gauteng DOH resolved to follow the National Health Council's advice to reduce the number of registers used in Primary Health Care facilities for the collection of health programme data (Adelekan, Makhubele and Ngcobo 2015). This entailed the roll out of a rationalised set of registers that would reduce duplications in data collection thereby minimising

administrative burden on facilities. With support from the Health Systems Trust, the province initiated the rationalisation process from March 2015, it identified key task team members to serve on a Rationalisation Committee, and Primary Health Care facilities began using the rationalised tools as of 1 August 2015 (Adelekan, Makhubele and Ngcobo 2015).

At the macro level, the new National Development Plan for 2030 sets out nine long-term health goals for South Africa (Republic of South Africa 2012). Five of these goals relate to improving the health and well-being of the population, and the other four deal with aspects of health systems strengthening. These moves are indicative of general system changes to the pipeline of information flow, and of a gradual shift towards the need for more "real time" data. Moreover, in 2013, the Centre for Scientific and Industrial Research (CSIR) was commissioned to produce a normative standards framework for health information systems in South Africa and the final result the National Health Normative Standards Framework for Interoperability in eHealth in South Africa (HNSF) was recently approved by the National Health Council (CSIR and NDOH 2013).

This framework ensures that base standards and interoperability specifications are in place, and that unique system identifiers are implemented to work across many systems and devices from a broad range of implementers and vendors in the health system (CSIR and NDOH 2013). It conforms to the South African Statistical Quality Assessment Framework (SASQAF) released in 2008, and which provides clear criteria and transparent procedures for evaluation of official statistics and other data.

Other data quality initiatives come from the private sector. For example, the annually published *District Health Barometer* is designed to assist the DOH in making health related information available for monitoring progress in health service delivery at district level (Massyn et al. 2015). It combines data from DHIS as well as the national TB register, and systems data from Statistics South Africa and the National Treasury (Kareithi et al. 2015, Vital Wave Consulting 2009). Another example comes from *MEASURE Evaluation Strategic Information for South Africa* (MEval-SIFSA), a USAID-funded project, which in 2013 developed and implemented a "Data Demand, Quality, and Use Facilitated Group Self-Assessment Tool." The tool examines the determinants of health information system performance in order to enhance the capacity of the DOH to identify data needs, improve data quality and its use for evidence-based decision making (Kareithi et al. 2015, MEval and SIFSA 2014, MEval and SIFSA 2015).

There are also multiple experiments in developing and using open source health applications for mobile phones and personal digital assistants (PDAs). One such company is Cell-Life, which has developed a variety of HIV/AIDS health management solutions for mobile phones and PDAs (they are a part of the OpenROSA/JavaROSA consortium) (Kareithi et al. 2015). Another example is an initiative called MomConnect, which is a South African National Department of

Health (NDoH) initiative which provides a means for pregnant women to receive stage-based relevant information about their pregnancy and also allows them to provide feedback on the services that they receive at clinics via cellphone. Furthermore, the DOH and its development partners provide a number of training workshops that cover various aspects of data quality. Some of these courses include Measure Evaluation-SIFSA's courses on Data Quality, DHIS Foundation and Intermediate, Monitoring and Evaluation, Evidence Based Health Management (accredited at 30 CPD points) and Communicating Data for Decision Making (accredited at 30 CPD points, level 3) (MEval and SIFSA 2014).

At the provincial and district level, several commitments to improve data quality have also been made. For instance, the Gauteng DOH approved, as part of its 2015 strategic plan and in alignment with Government' National Development Plan for 2030, the medium-term goal to complete a reform of the health system (Gauteng DOH 2013b). This sets out three proprieties including increasing the "number of health facilities with Health Information Systems to all 36 hospitals and 372 PHC facilities" and enhancing "broadband network access to all health facilities" by 2020 (Gauteng DOH 2013b). The province is introducing the *3TIER* system for HIV-ART in all clinics, which will improve data quality going forward, and it has set a goal to employ additional data capturers to improve the flow of information (Gauteng DOH 2013b).

This background reveals that there are numerous legislative, management, leadership, training, and technical resource initiatives underway to improve the quality of and access to information in health care facilities across the nation. Within this context, the current data quality assessment of Gauteng Primary Health Care facilities is not only necessary but also timely and pioneering. The following section describes the methods and techniques used to conduct this assessment.

2 Methodology—Primary Health Care Facilities

2.1 Research Questions

The data quality baseline set out to answer the following questions:

1. What is the data quality for selected indicators in all of Gauteng Province's health facilities, and what are the causes for poor data quality? (**Section 3.1 Recount Results** responds to this question)
2. What is the impact of the rationalisation of registers on data quality? (**Section 3.1 Recount Results** responds to this question)
3. What technical capacity is available for supporting data collection, reporting and use at Gauteng health facilities, and how does this affect data quality? (**Section 3.2 Level of Technical Resourcing** responds to this question)

4. What human resource capacity is available for supporting data collection, reporting and use at Gauteng health facilities, and how does this affect data quality? (**Section 3.3 Human Resource Capacity at Health Facilities to Carry out Data Management** responds to this question)
5. What is the status of DHMIS policy implementation at the sites? (**Section 3.4 Implementation of DHMIS Policy Status** responds to this question)

2.2 Data Collection Tools

The assessment teams conducted two types of assessments that collected data on: (1) how accurately data are captured, and (2) data systems and staffing capacity. This necessitated the development of two different research instruments—a recount tool and a systems and staffing survey tool. These data collection tools were collaboratively developed with officials from the Gauteng Department of Health, piloted at three sites, and further refined after fieldworker training. The tools are provided in Annexure A, and a brief description of each tool is provided here.

The field work teams used the **recount tool** to recount data elements for the selected health indicators from the source data for a three month period2. A detailed fieldwork manual supported fieldworkers to identify the correct source documents and count columns, and provided examples of typical errors. In additional fieldworkers also had access to a range of counting sheets to assist them in counting from source registers.

Trained teams of fieldworkers captured the information directly onto portable computer tablets using DOOBLO tablet survey software, where the DHIS data was imported into the backend of the software. Monthly summary data, as calculated and reported per the facility records, were recorded, and compared with DHIS data. The fieldwork teams could immediately identify discrepancies between their recounted data, the summary data, and reported values found on the DHIS.

The strategy encouraged teams to count actual values, and double check their own work. However, while teams could identify discrepancies, they did not know the absolute DHIS values. When the team identified a discrepancy, the fieldwork team was required to give feedback to a facility representative. The representative had to sign-off on the recounted data and provide a possible reason for the error. The recount tool identified whether ticks were recorded correctly as per the Auditor General's guidance, and whether official registers were used.

The recount was conducted for nineteen indicators and approximately twenty-seven data elements, for the period July, August and September 2015. The recount data were compared to a

2 Paper registers and summary sheets were the source data in PHC facilities. In hospitals, the data are directly captured onto relevant database systems

DHIS extract for the period July, August and September 2015, on 13 November 2015. The tables below show which health indicators and data elements were recounted3.

Table 1: Health indicators and data elements included in the recount tool

Facility Headcount		
1	PHC Utilization Rate	PHC Headcount total
Maternal, child and women's health		
2	Immunization coverage under 1 year	Immunized fully under 1 year new
3	Measles second dose coverage	Measles second dose coverage
4	Antenatal 1st visit before 20 weeks rate	Antenatal 1st visit before 20 weeks
		Antenatal 1st visit total
5	Mother postnatal visit within 6 days	Mother postnatal visit within 6 days after delivery
		Delivery within facility total
6	Cervical cancer screening coverage	Cervical cancer screening in women 30 yrs and older
TB Indicators		
7	TB symptom 5yrs and older screen rate	Client 5yrs and older screened for TB symptoms
		PHC headcount 5yrs and older
8	TB client lost to follow up rate	TB client lost to follow up
		TB client started on treatment
9	TB client treatment success rate	TB client successfully completed treatment
		TB client started on treatment
10	TB client death rate	TB client died during treatment
		TB client started on treatment
11	TB MDR Confirmed Treatment Start Rate	TB MDR confirmed client start on treatment
		TB MDR confirmed client
12	TB MDR Treatment Success Rate	TB MDR confirmed client successfully treated
		TB MDR confirmed client start on treatment
HIV Indicators		
13	Clients tested for HIV (including ANC)	Clients tested for HIV (including ANC)
14	Rate of HIV infections in young men and women aged 15-49	HIV Client tested positive age 15-49
		Client tested HIV age 15-49
15	Infant 1st PCR test positive rate around 6 weeks	Infant 1st PCR test positive around 6 weeks
		Infant 1st PCR test around 6 weeks
16	Antenatal client initiated on ART rate	Antenatal client start on ART
		Antenatal client eligible for ART initiation
17	Medical male circumcision performed	Sum of males who are circumcised under medical supervision
Non-communicable Diseases		
18	Hypertension clients treatment new	Hypertension clients treatment new
19	Diabetes clients treatment new	Diabetes clients treatment new

The **system and staffing tool** (included in Annexure A) is a structured interview that covers a range of topics, including technical resources, human resources, and the status of DHMIS policy implementation. At each facility, the fieldworkers interviewed the facility manager or their designee, and the most senior staff member in charge of data capturing. Responses were then

3 In hospitals, it was not feasible to recount all of the indicators, and teams recounted as much of the data on as many of the indicators as possible, within the allotted fieldwork time for each site

captured onto tablet computers which used computer assisted personal interviewing (CAPI) software.

2.3 Field-Site Selection

The data quality baseline assessment intended to cover all Primary Health Care (PHC) facilities in Gauteng. Gauteng DHIS identified 377 Primary Health Care facilities across the five Gauteng districts that would be the target of the assessment. This list excluded satellite clinics, health posts and mobile clinics. In total, the systems and staffing assessment was completed at 286 sites, and the recount assessment was completed at 287 sites. In addition, 8 hospitals were selected for participation in this research.

Reasons for drops in numbers

Some site visit data, and therefore sites, were not used in analysis when they were later identified as not eligible (e.g. satellite clinic or not operational for the full data collection period). The other excluded sites included those that were unreachable, uncontactable, or sites with effort expended that yielded no data.

A site was considered unreachable when, after three attempts at establishing contact, there was no response. A site was considered uncontactable when the contact information could not be identified through provincial, district office or supervisors. The site was deemed "effort expended" if a site had a confirmed appointment, but the team was unable to commence with fieldwork or complete fieldwork as planned due to unforeseen circumstances (e.g. the facility manager was not present at the agreed upon time).

2.4 Data Collection Strategy

A team of thirty-two fieldworkers and five supervisors conducted the fieldwork over an eleven-week period between November 2015 and March 2016. After initial fieldwork suggested that the estimated time allocation for clinics were inaccurate, MEASURE Evaluation SIFSA also deployed a team of five interns to support the fieldwork. The fieldworkers who conducted the recount and interviews were not health specialists, but had previous experience in conducting business-to-business data collection. Fieldwork supervisors provided oversight and quality assurance during site visits.

In November 2015 the fieldworkers participated in a one-week training workshop. The first two days consisted of office-based training with support from Gauteng Department of Health staff from each of sub-district and key experts on HIV and TB indicators. The second two days were in the field, with training taking place at a sample of facilities, followed by a debriefing session to address any fieldworker concerns or questions, and adaption of instruments and guidance

materials as needed. If a fieldworker resigned and had to be replaced, the fieldwork supervisor conducted in-field, individual training for the new recruit.

While a high-level fieldwork plan guided the teams, each fieldwork team scheduled their own fieldwork dates and organized their daily logistics.

The Assessment Team provided the provincial office staff with a weekly update of the fieldwork schedule who then then communicated to the appropriate district facility staff. Provincial officials and district representatives assisted fieldworkers to obtain contact details of clinic managers, and notify the clinics of the impending fieldwork.

After initial difficulties in communicating with facilities on time, the team implemented a bi-weekly meeting. This meeting, hosted by the Provincial office, included the fieldwork supervisors and key officials in each of the Health Districts. The purpose was to share the fieldwork plan and recruit the help of relevant officials in informing clinics of the preparations required. Between meetings, communication was facilitated through WhatsApp groups involving fieldworkers, fieldwork supervisors, and district representatives.

Fieldwork supervisors in each of the five districts addressed fieldworkers' data quality. They did this by: conducting sample checks by telephone to verify that fieldwork took place, accompanying fieldwork teams together with the district representatives (in some cases the district manager), and checking for data consistencies against the recount sheets. Fieldwork supervisors also conducted in-office data quality checks. Where data quality challenges were identified, or where feedback from the facilities, district offices or provincial offices prompted intervention, supervisors revisited the clinic.

2.5 Analysis Strategy

Facilities received two sets of feedback; informal and formal. Fieldworkers provided immediate, informal feedback to facilities while at the sites. Then, following the in-office data checking and cleaning, the Team produced a facility report using E-Tabs software. Each facility then received this formal report, indicating their score per-element or per-programme, and the overall data quality score, and the absolute values of recounted data, summary data and DHIS data.

The team conducted quantitative analysis using statistical software for social science such as PSPP, SPSS and Stata. Tables and graphs for reporting purposes were produced using Microsoft Excel software. Most analysis is descriptive and non-inferential, and where inferential statistics were conducted it was confined to anova tests, t-tests, chi-square tests and pairwise correlations. The team conducted qualitative analysis using manual open, axial and selective coding, as applicable.

2.6 Response Rates

Table 2: Number of PHC Facilities successfully reached during the Data Quality Baseline

Districts	Completed	Sample	% Completeness
City of Johannesburg Metro Municipality	**91**	**119**	**76%**
Johannesburg A Health sub-District	7	13	54%
Johannesburg B Health sub-District	10	11	91%
Johannesburg C Health sub-District	11	13	85%
Johannesburg D Health sub-District	21	29	72%
Johannesburg E Health sub-District	7	9	78%
Johannesburg F Health sub-District	14	15	93%
Johannesburg G Health sub-District	21	28	75%
City of Tshwane Metro Municipality	**60**	**75**	**80%**
Tshwane 1 Health sub-District	19	22	86%
Tshwane 2 Health sub-District	9	13	69%
Tshwane 3 Health sub-District	11	12	92%
Tshwane 4 Health sub-District	5	6	83%
Tshwane 5 Health sub-District	4	6	67%
Tshwane 6 Health sub-District	8	9	89%
Tshwane 7 Health sub-District	4	7	57%
Ekurhuleni Metropolitan Municipality	**56**	**94**	**60%**
Ekurhuleni East 1 Health sub-District	8	14	57%
Ekurhuleni East 2 Health sub-District	6	17	35%
Ekurhuleni North 1 Health sub-District	6	12	50%
Ekurhuleni North 2 Health sub-District	9	17	53%
Ekurhuleni South 1 Health sub-District	11	15	73%
Ekurhuleni South 2 Health sub-District	16	19	84%
Sedibeng District Municipality	**33**	**39**	**85%**
Emfuleni Local Municipality	25	26	96%
Lesedi Local Municipality	4	8	50%
Midvaal Local Municipality	4	5	80%
West Rand District Municipality	**47**	**50**	**94%**
Merafong City Local Municipality	14	15	93%
Mogale City Local Municipality	18	18	100%
Randfontein Local Municipality	8	9	89%
Westonaria Local Municipality	7	8	88%
Total	**287**	**377**	**76%**

Seventy-six percent of the sampled PHC facilities have been assessed in this study. The percentage of sampled sites completed in the districts varied with a high percentage namely; 94% achieved in the West Rand district and only 60% in Ekurhuleni. It should be noted that a total

number of 56 PHC facilities have been assessed in Ekurhuleni and 47 in the West Rand. The completion rates were higher where the district staff was very supportive an actively communicated with their facilities, where facilities were more accessible to the field team, and updated contact details were available.

Seven hospitals were included in the data quality baseline – mainly to test the methodology for conducting a data quality baseline at facilities that are larger, more complex and more reliant on technology. These were Charlotte Maxeke, Chris Hani, Dr G Mukhari, Helen Joseph, Kalafong and Steve Biko.

2.7 Data Limitations

All assessments and evaluation studies have data limitations. The limitations to this study are noted in this section.

- **Coverage of PHC facilities**

  This report reviewed data quality assessed at 287 of the 377 facilities in Gauteng constituting 76% of the list of sampled facilities. Since most DQAs usually only assess data in a sample of facilities, the plan to cover all facilities in this assessment was commendably ambitious, and resulted in a higher the usual coverage of facilities than is customary.

- **Sample of health indicators**

  The time frame and budget limited the study to a sample of indicators during a three-month period. The table of indicators and the data elements assessed are included in Table 1: Health indicators and data elements included in the recount tool). Furthermore, some of the TB indicators and data elements although assessed could not be suitably used in the analysis since the ETR.NET4 data were not available for comparative use in the study.

- **Level of comparison is often at data element level, not necessarily at indicator level**

  Findings and recommendations in this report pertains to data at a data element level or at indicator level. If an indicator consists of more than one data element each data element is used as a separate data quality entity in this report. Thus many findings and

4 ETR.Net is an electronic tuberculosis register designed for TB/HIV surveillance, program monitoring and evaluation. The system consists of a database which is accessed by the user via the user-friendly software interface, custom developed for the Microsoft Windows environment.

recommendations pertain to the separate data elements of an indicator and not the indicator. Direct comparisons with indicator level data should not be made.

- **A full data quality analysis not implemented**
  Given the aggressive deadline limitations from the Gauteng DOH, it was not possible to attempt a full data quality analysis for all indicators at all facilities in the Gauteng province. The team analysed a representative sample of indicators and elements.
- **Compares recount to a version of the DHIS extracted three months after the period has passed.**
  The recount data were compared to a DHIS extract for the period July, August and September 2015 (extracted 13 November 2015). Data quality updates applied to the DHIS after 13 November (due to routine checking and updating at facilities) may not be adequately reflected in this analysis. It does, nonetheless, give an indication of the typical discrepancies between source and DHIS data that would be found after the closing of the data set for the preceding quarter.
- **This analysis compares the recount to two other data Instances, despite their being many more instances that may be of interest**

  The data collection and update processes to the DHIS are done across various data instances. Each individual data elements or ticks is presented in many data sets (up to eight) as illustrated in Figure 1.

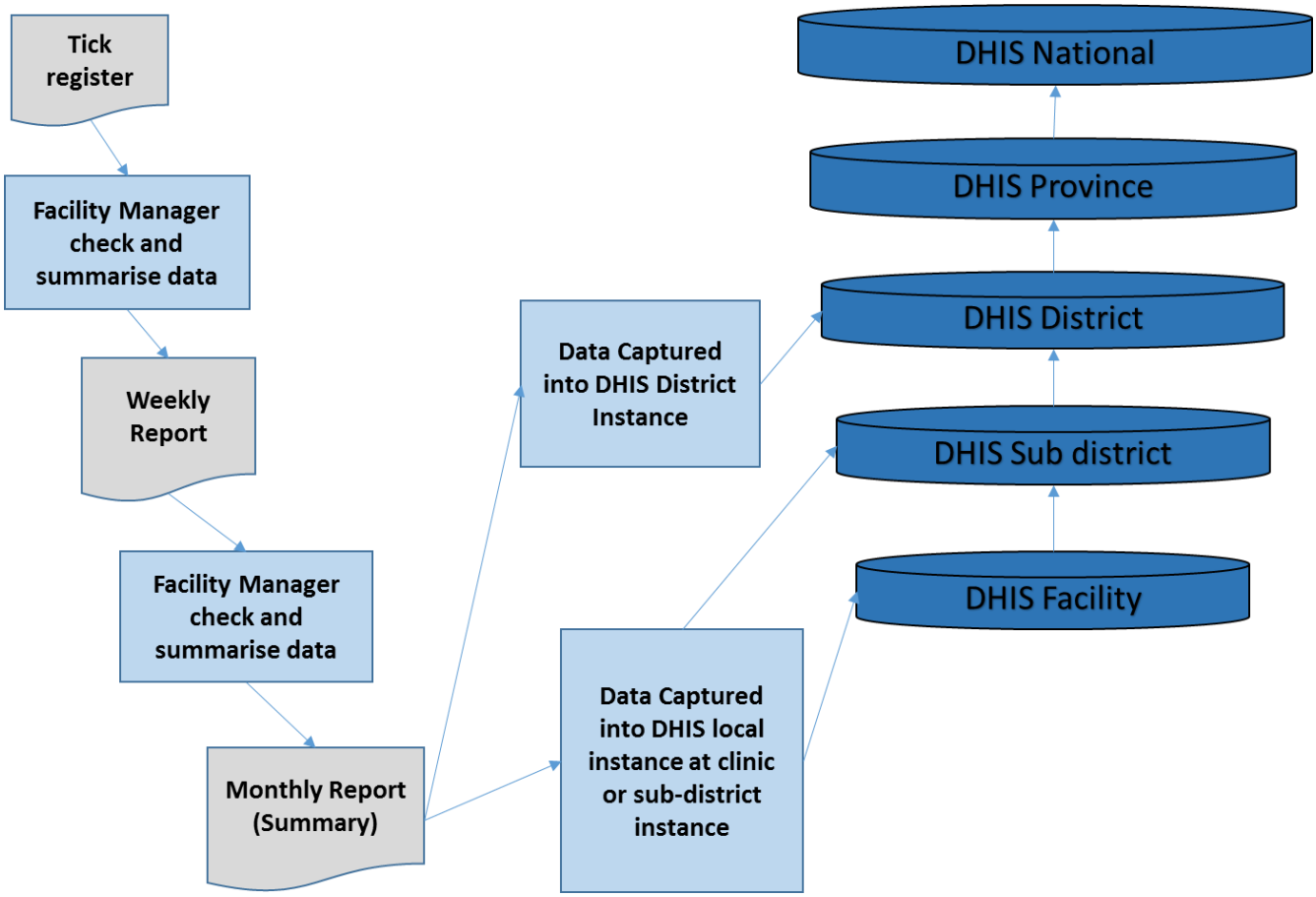


Figure 1: Paper and electronic instances of data elements.

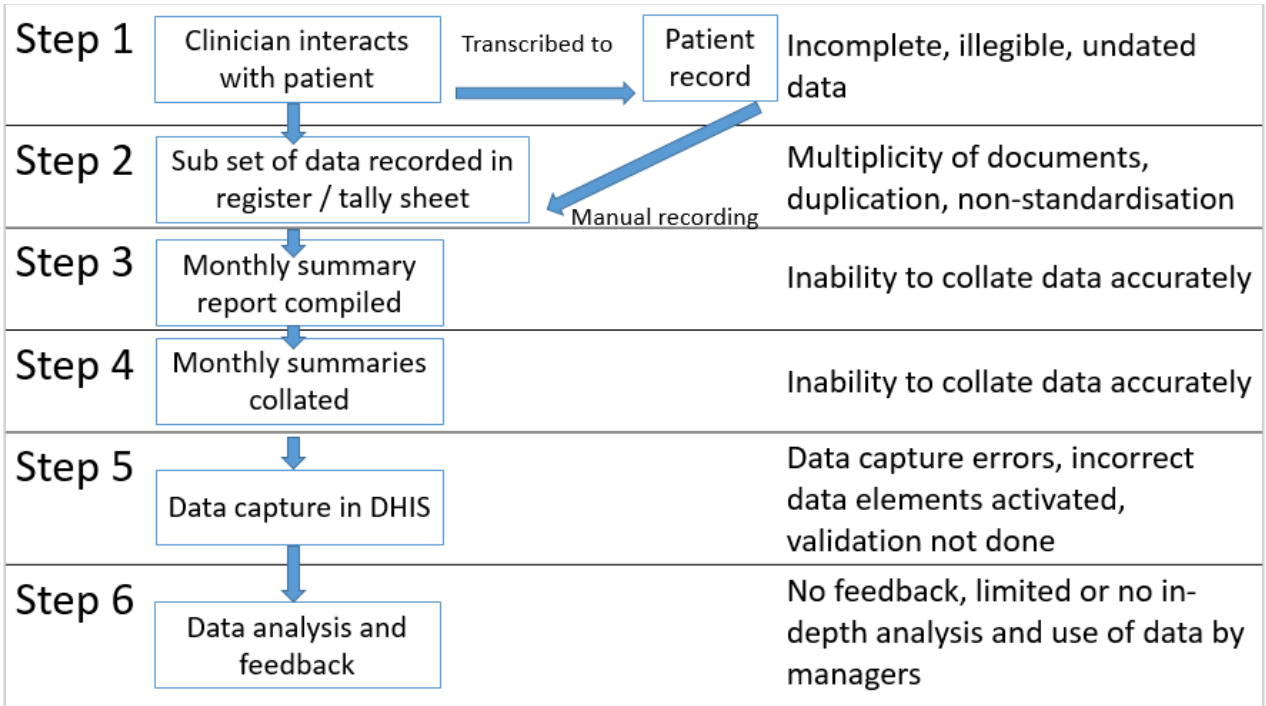


Figure 2: Data quality issues at various steps in the data consolidation process.

It is possible that human or programming errors may be made between each of the data instances. These errors include, for example, human errors of capturing, and interpretation as well as potential aggregation inconsistencies. This report only compared the source data against the monthly summary and one instance of the DHIS. Therefore, no comments are made regarding data issues pertaining to errors at each DHIS update or aggregation level.

- **Bias inherent to reductionist quantitative methods, and qualitative methods that rely on self-report and observation data**

The quantitative data represented in this report is a reduction of the complex situation of data management at facility level, to a simple measure of agreement between a recount of source, summary and DHIS data. This is supplemented by qualitative data to provide insight into the nature and reasons for discrepancies. These data rely on self-reporting from facility staff and observations made by fieldworkers during a short period of interaction at facilities. Self-report bias may skew the findings if facility staff answer questions where they do not actually know the answer or because either consciously or unconsciously they wish to present themselves in a socially acceptable manner.

Despite the limitations outlined above, the findings in this report can be regarded as accurate enough to base decisions about data improvement initiatives on.

3 Baseline Findings at Health Facilities

The baseline findings presented in this section rely on information collected from 287 Primary Health Care facilities in Gauteng, and 7 hospitals. The results were obtained by recounting various data elements at the facilities. Data come from source documents and are supplemented with information collected from an interview on the Systems and Staffing at the health care facilities.

3.1 Recount Results

- **Finding 1: Most PHC facilities (>90%) need to improve the quality of patient headcount data. Only around 10 percent of facilities have accurate headcount data, and over-reporting and missing data is the biggest problem. The accuracy of headcount data improved over time with the rationalization of registers.**

The recount from source document data are compared to the DHIS extract of 13 November 2015. According to guidelines set by the Auditor General, a discrepancy of 5 percent over or under the reported values is acceptable, meaning that data should be at least 95 percent accurate.

In this assessment, only a small percentage of the PHC facilities actually met this accuracy threshold – For September 2015, only 11% of the PHC headcount data was found to be within the acceptable range.

In this assessment fieldworkers recounted various data elements from source documents available at each PHC facility for July, August and September 2015. When comparing the result of this recount to the values as per the DHIS extract for the same period, it becomes apparent that data quality is in dire need of improvement at most facilities.

Table 3 below depicts the number of PHC facilities that fall into each of these data quality categories. These findings are also visually presented in figures 3 and 4, which show the percentage of PHC facilities with good and poor quality data, respectively. What is evident from figure 4 is that the percentage of sites with missing data or significant over reporting of data is very high. However, there appears to be some improvement in the quality of headcount data over the three-month period observed. In July, about 84% of PHC facilities had problems, but this was reduced to 68% in August and to 64% in September. Over reporting data may result from using multiple patient registers, whereby a patient's information gets counted more than once. Ultimately, as the Department of Health makes progress towards integrated patient care (where a patient is seen for all of his/her problems in one consulting room) the impact of over-counting is likely to decrease. The percentage of sites with minor data errors or totally accurate data was significantly higher for September (figure 3), possibly due to the rationalisation of registers initiative which was implemented in August.

Table 3: Number of PHC facilities in each of the data quality categories—PHC headcount

		Under-report >5%	Under-report <5%	Accurate	Over-report <5%	Over-report >5%	Missing Data5	Total
PHC headcount total	July	34	7	14	28	177	27	287
	August	52	13	27	43	118	34	287
	September	54	10	32	35	127	29	287

5 "Missing data" refers to data that is not entered in the DHIS, or data that is missing because the headcount registers were not available on the day of the recount

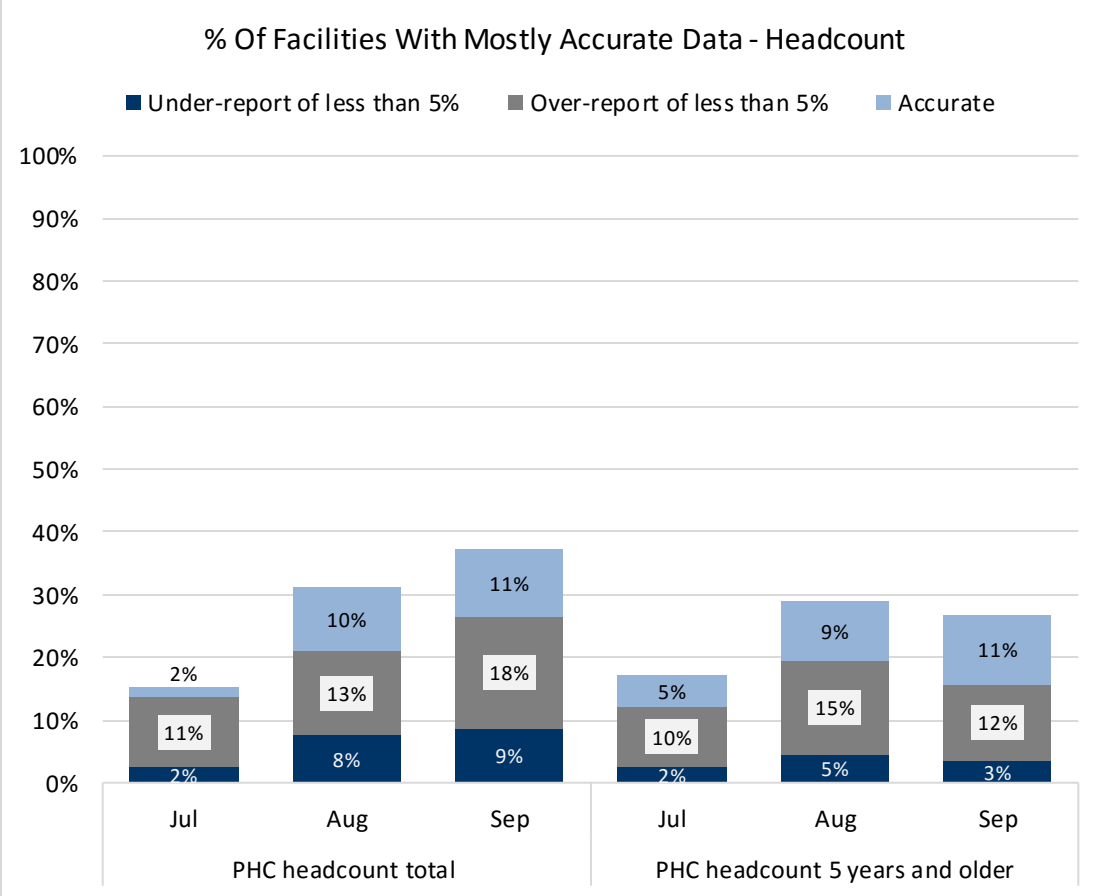


Figure 3: Percentage of facilities with mostly accurate data on facility headcount

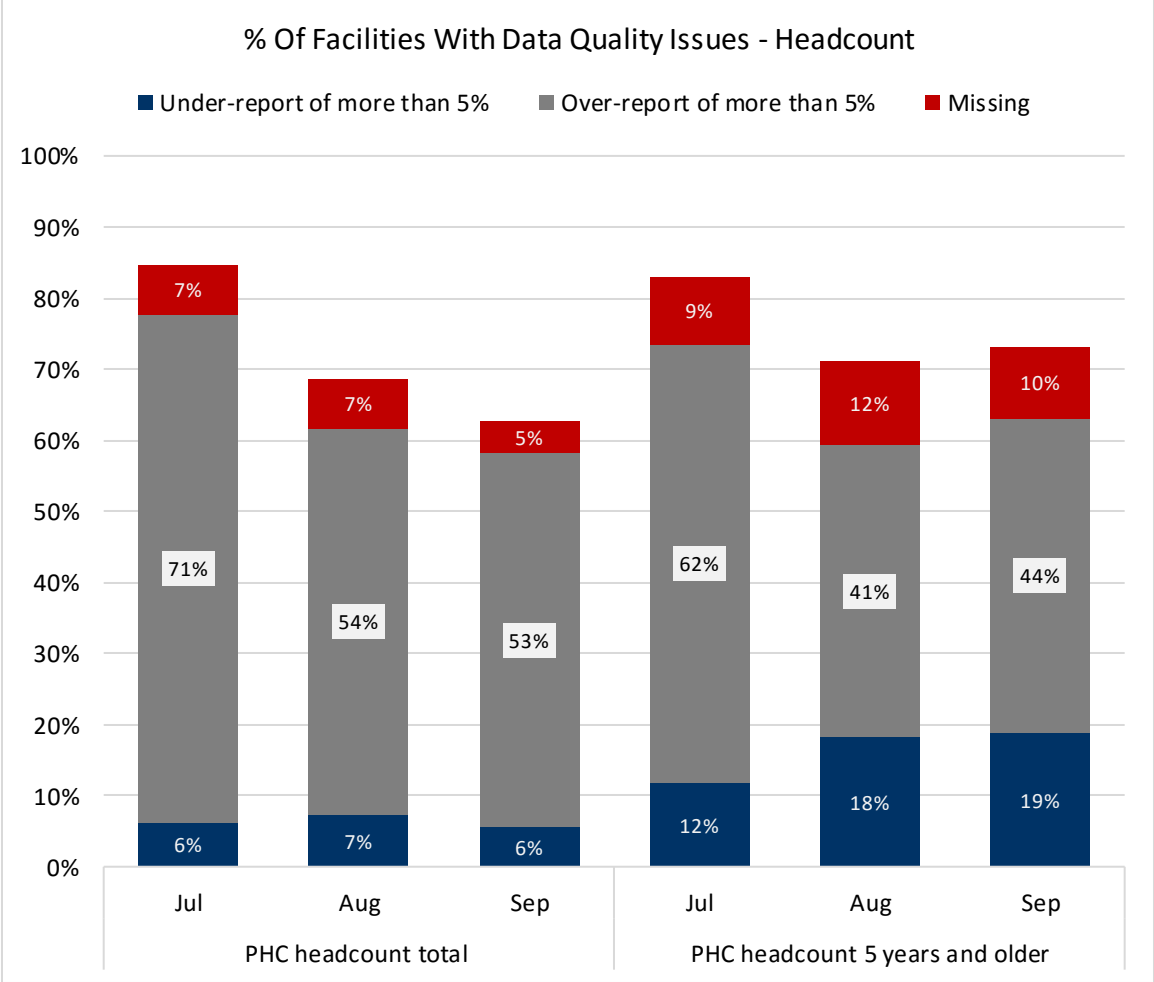


Figure 4: Percentage of facilities with data quality issues on facility headcount

- **Finding 2: Only one-third of PHC facilities produced high-quality health-related data. The highest quality was found for Maternal, Child and Women's Health (MCWH), antenatal HIV care, and Hypertension data elements. Data quality was poorest for other HIV, Tuberculosis (TB) and Diabetes elements.**

The next eight figures demonstrate the percentage of Primary Health Care facilities with accurate or problematic data on health-related data elements that were recounted, including: a) maternal, child, and women's health, b) tuberculosis, c) HIV/AIDS, and d) diabetes and hypertension.

A) MATERNAL, CHILD, AND WOMEN'S HEALTH (MCWH) ELEMENTS

Figures 5 and 6 show the findings for all MCWH elements. Data quality was generally best for immunization records, and tracking cervical cancer screenings. The quality of data for antenatal and postnatal visits was poor, usually over or under reported, or entirely missing. There was little

variation of data quality over the observed time. Fewer than 25 percent of PHC facilities could show accurate data for child immunisation and measles vaccinations.

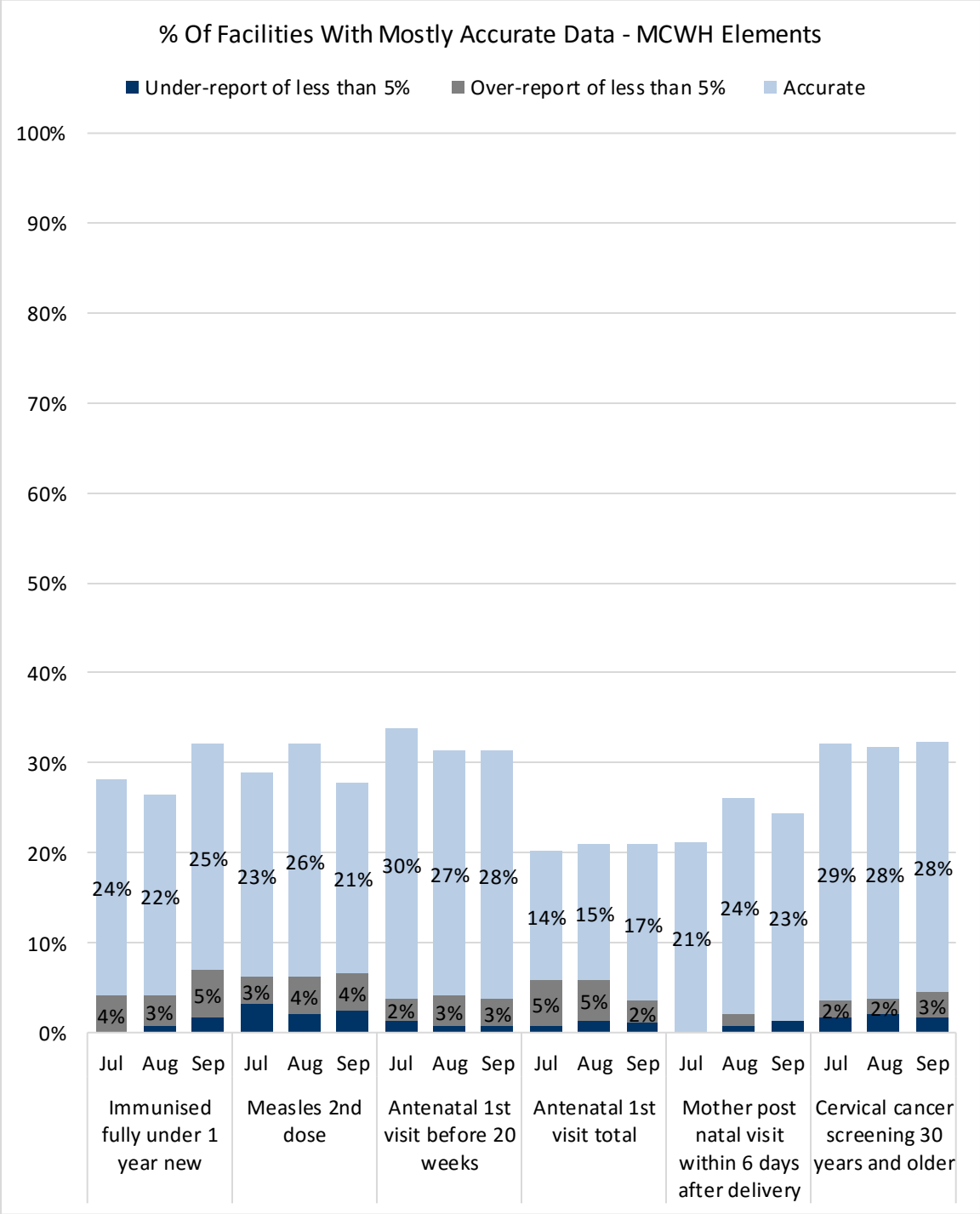


Figure 5: Percentage of facilities with accurate data for MCWH data elements

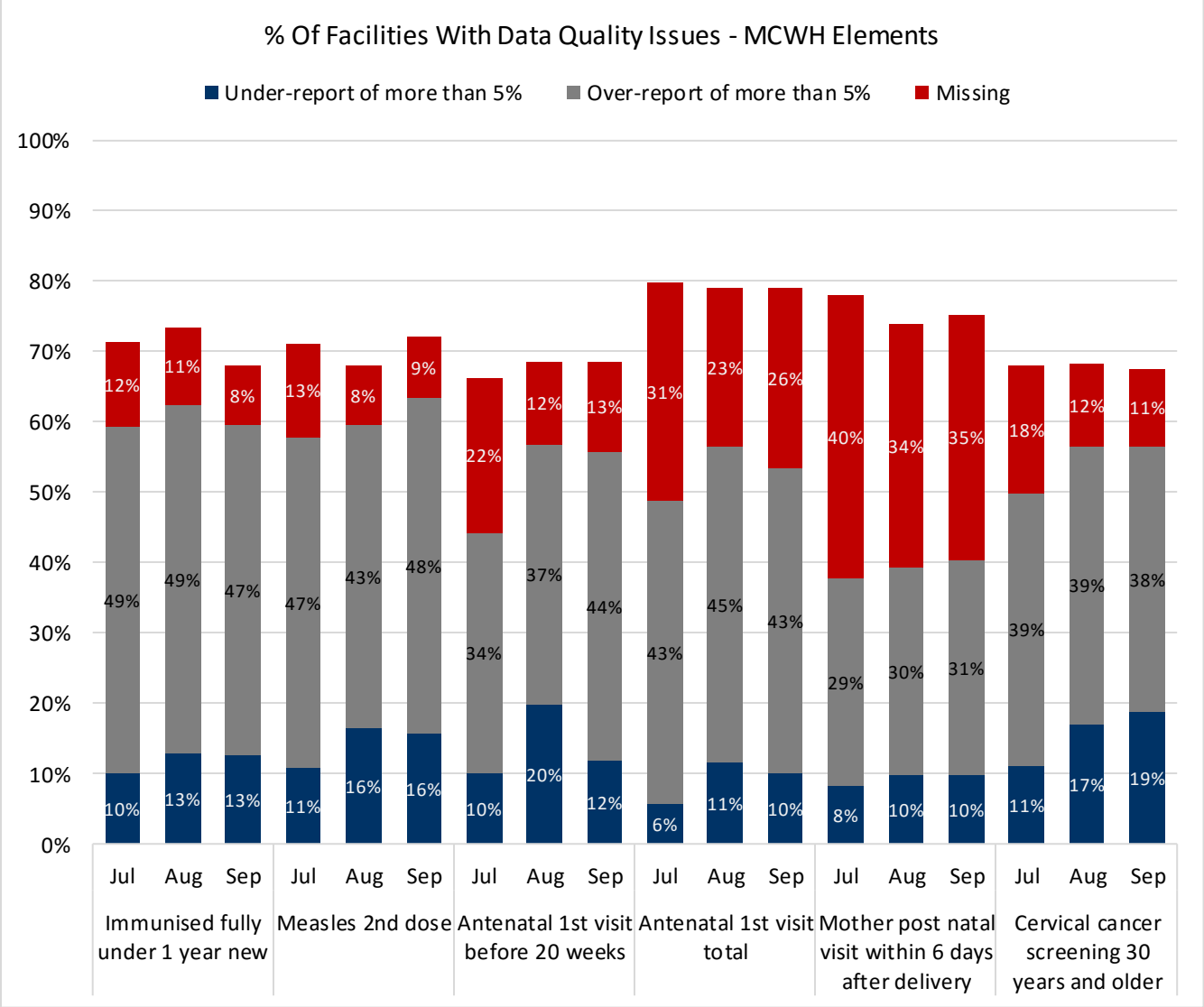


Figure 6: Percentage of facilities with data quality issues for MCWH data elements

B) TB ELEMENTS

Most of the recounted TB data elements could not be compared to provincial level data, since the ETR.net data was not made available to the research team. Recounting TB data was challenging due to the technicalities of calculating certain data elements. For example, to determine the "Lost to follow up" rate for July, August and September 2015 the cohort of patients who started treatment in July, August and September of 2014 had to be identified. As figure 7 below shows, "lost to follow up" data were generally missing, as were large amounts of data for on initiating treatment. Only about a third of the facilities had accurate data, which is far from adequate. There was little variation in quality over the three observed months for the one TB data element that could be measured.

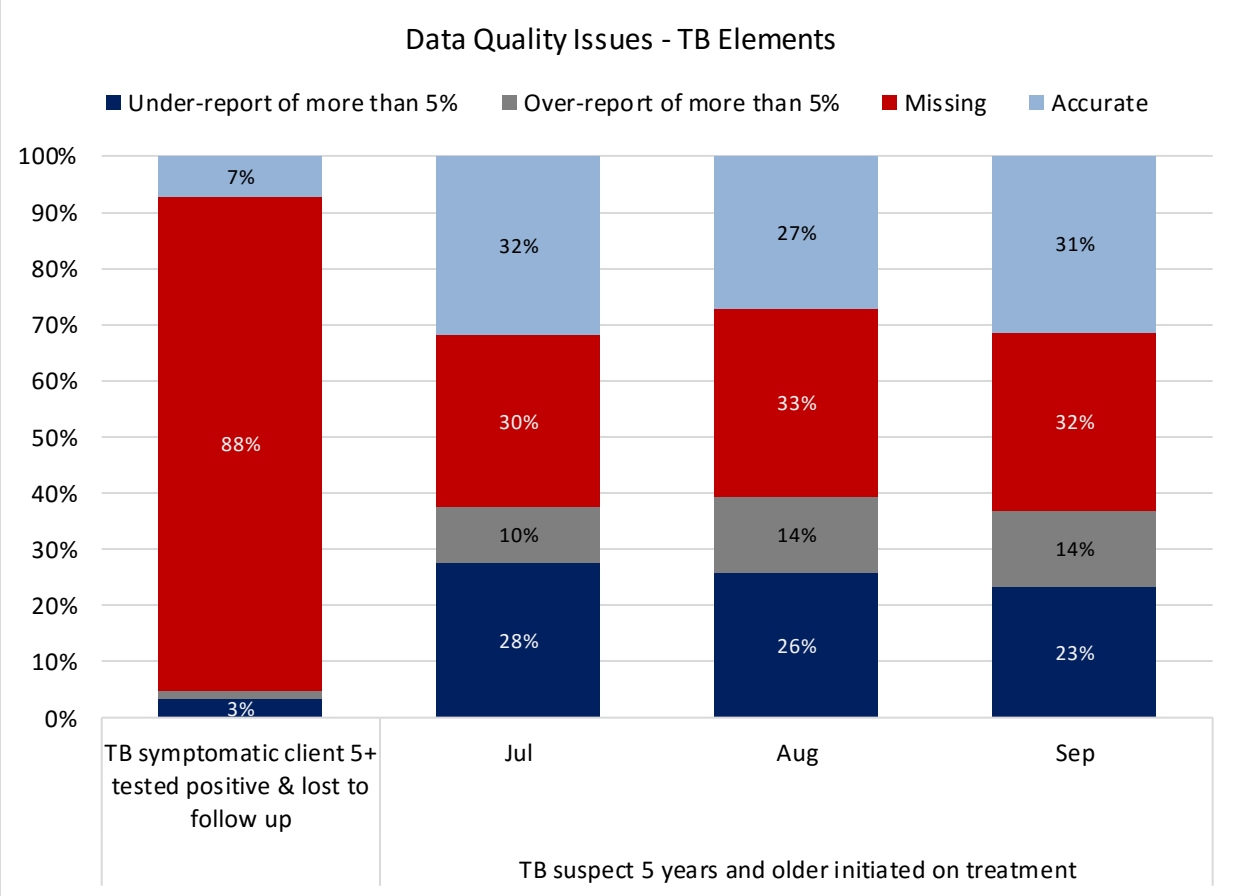


Figure 7: Data Quality for TB elements (Percentage of facilities)

C) HIV & AIDS ELEMENTS

The accuracy of HIV and AIDS related data at PHC facilities is troubling. Figures 8 and 9 reveal that only one-third of facilities produced accurate data, and that was for three out of seven elements alone. For the other elements, fewer than 20 percent of facilities had accurate data. Accuracy was best for records relating to infant PCR tests, and antenatal HIV care. However there was a significant amount of missing data. Problems with extreme over or under reporting, and missing data showed up mostly in data related to HIV testing. Missing data for "Infant 1st PCR test positive around 6 weeks" is to be expected because, although the tests are routinely conducted, the results can only be obtained from a laboratory test and a clinical follow-up. Clinical laboratory interface is a known issue in PHC facilities. Initiatives by PEPFAR partners and the Department of Health are underway to improve this.

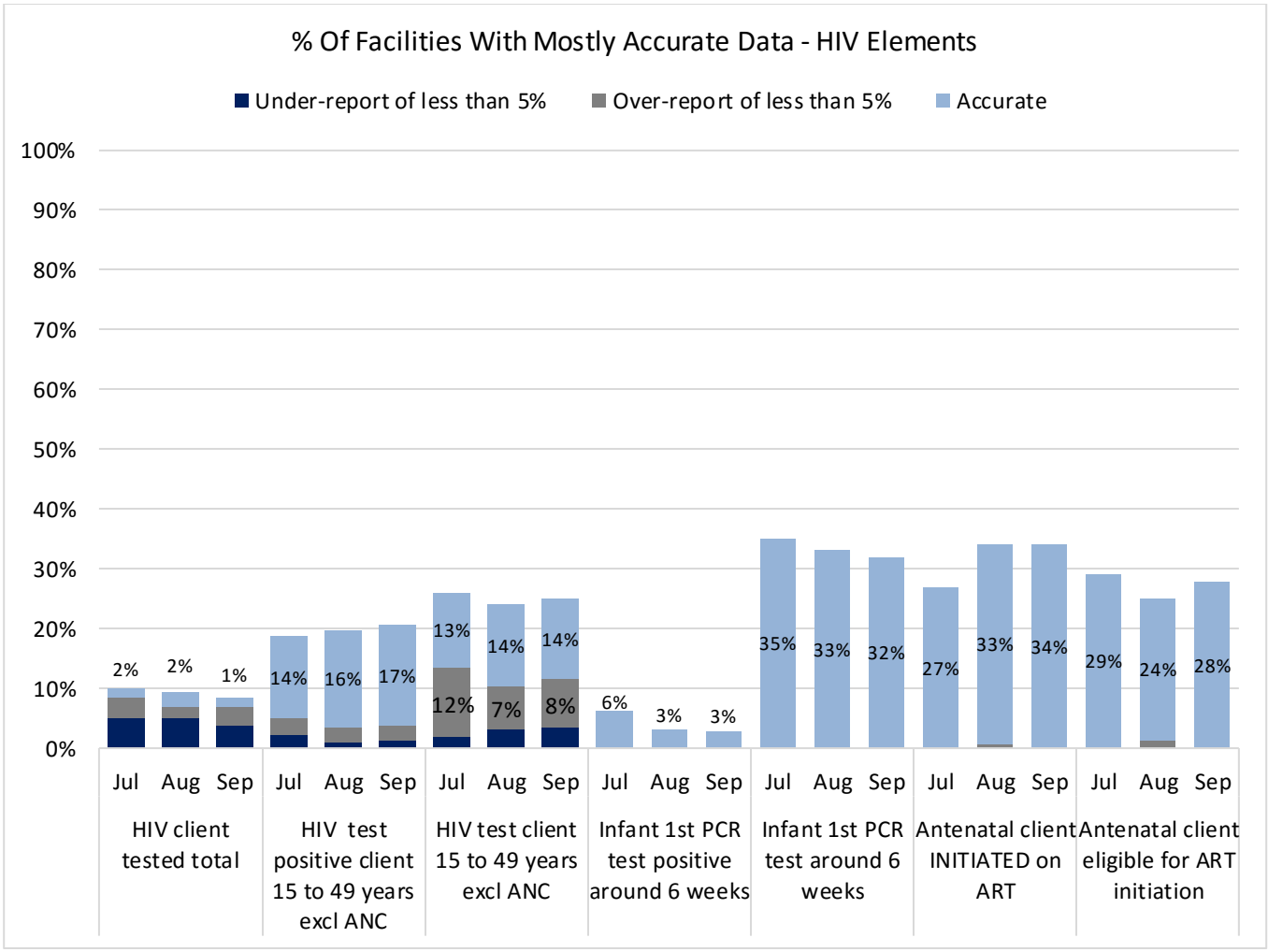


Figure 8: Percentage of facilities with mostly accurate data for HIV elements

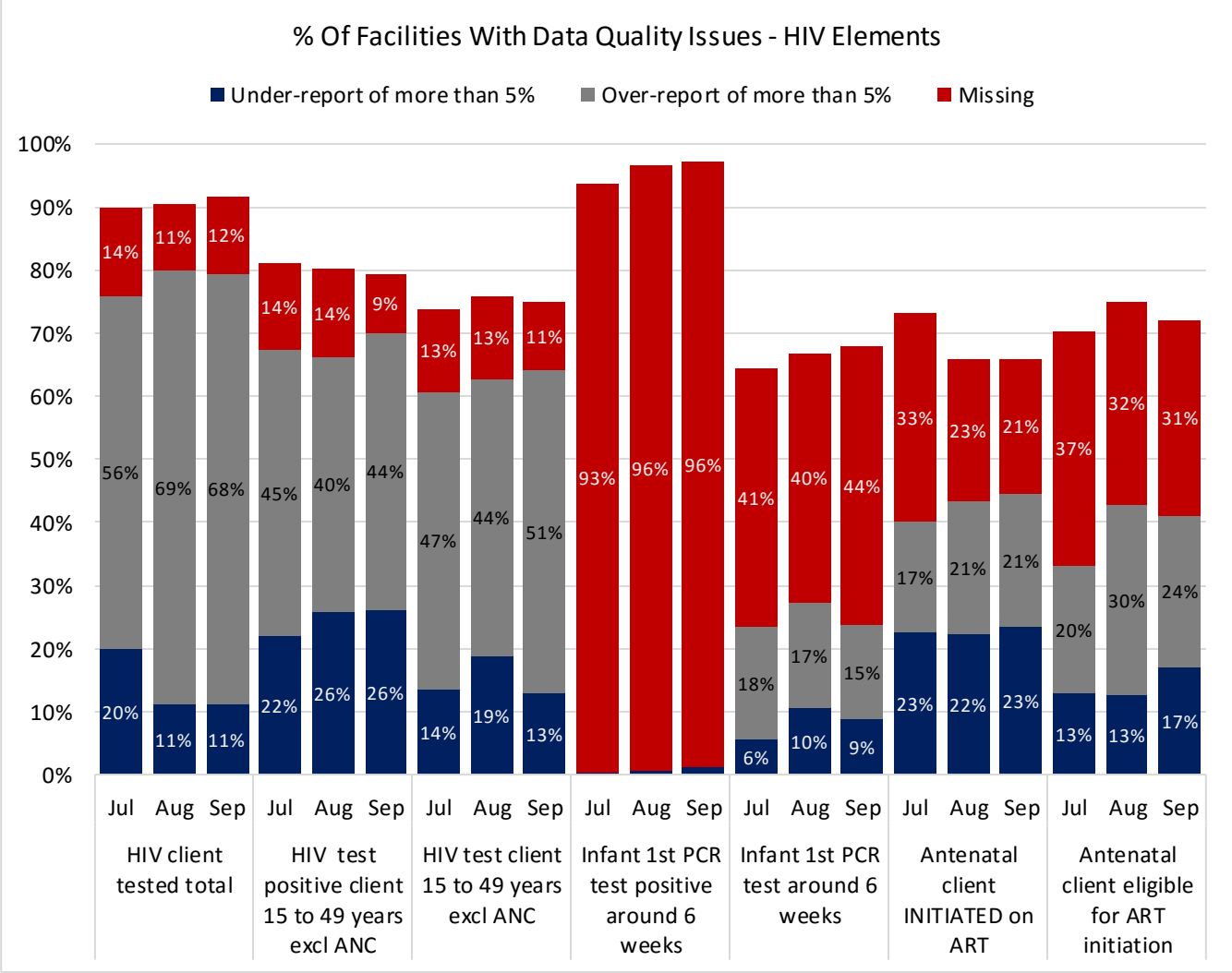


Figure 9: Percentage of facilities with data quality issues for HIV elements

D) DIABETES AND HYPERTENSION ELEMENTS

Two Non-Communicable Disease (NCD) data elements were recounted— clients who started hypertension and clients who started diabetes treatment. About 33% of facilities showed accurate data for patients on hypertension treatment, but almost the same amount are over reporting this data. This means that facility managers and senior personnel responsible for data recording are not adequately checking to ensure that the correct hypertension information is being recorded, or they are purposively inflating results. With regards to diabetes, at almost 80 percent of facilities, the diabetes treatment data for July was missing (i.e. either the data at source was not available during recount or the data in the DHIS was not available). This improved slightly during August and September, but overall, fewer than 15 percent of facilities had accurate diabetes treatment data. Figure 10 shows the data quality for these elements.

Finally, missing data were pronounced on data elements related to procedures that do not occur frequently at PHC facilities (e.g. deliveries and medical male circumcision). This is not unexpected, but also limited the ability to conduct adequate recount statistics. Data was also

missing on health indicators that require long-term tracking or waiting for test feedback (e.g. some of the TB elements, diabetes, and Infant 1st PCR test positive around 6 weeks). This is troubling because improving health outcomes of infectious and chronic diseases is especially vital for maintaining a healthy population, and accurate tracking of patient treatment is essential for preventing the development of drug resistance diseases.

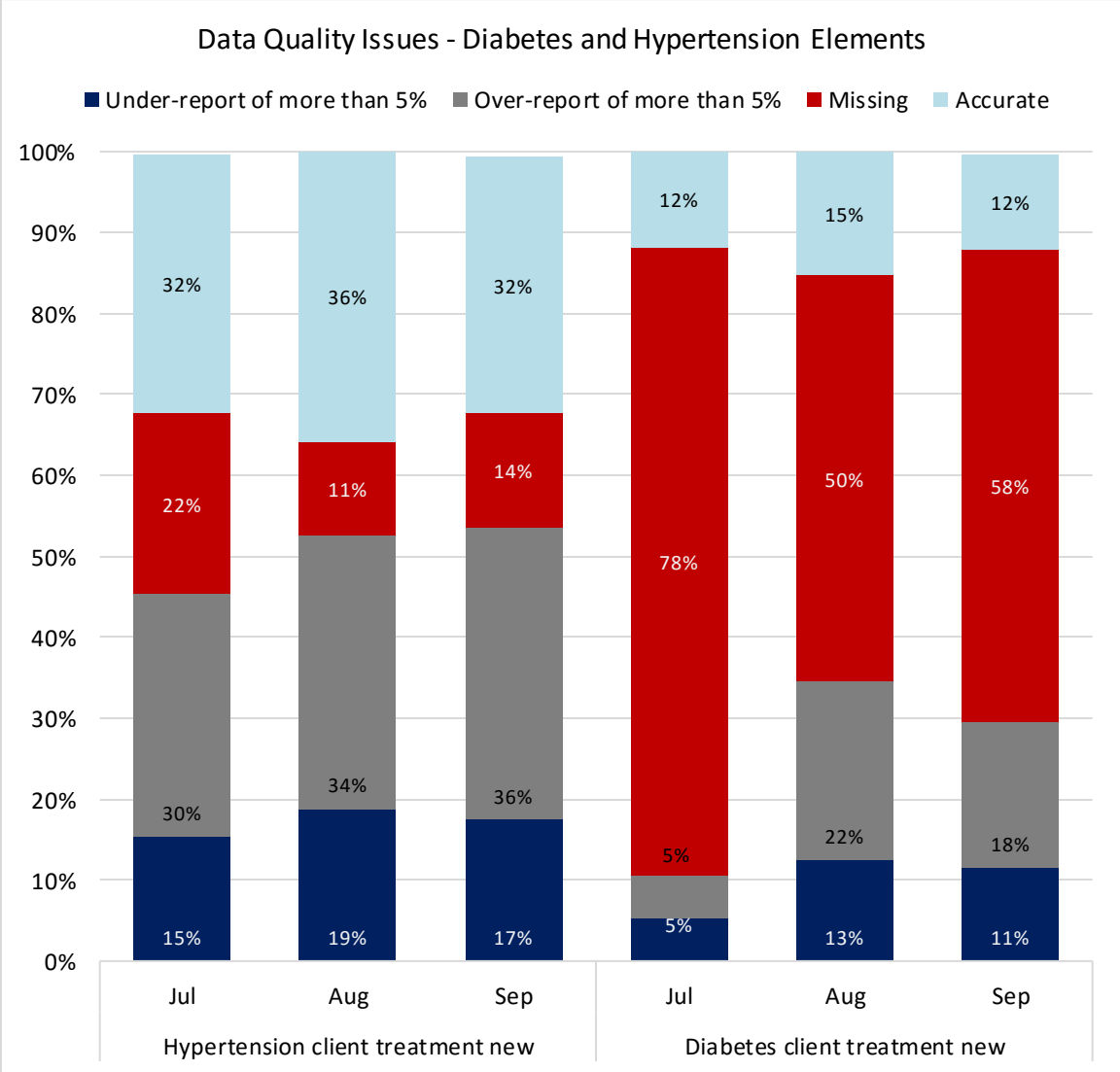


Figure 10: Data Quality for NCD and Headcount elements (Percentage of facilities)

- **Finding 3: Data quality varies by indicator and by district. City of Johannesburg PHC facilities, more than others, struggled with missing data. Sedibeng PHC facilities, tended to underreport more than other facilities. West Rand PHC facilities tended to perform better on a few of the assessed data elements.**

A selection of the best performing data elements per programme is shown below, and depicted by district. This includes headcount, immunisation, TB treatment initiation, antenatal ART initiation, and hypertension data.

Headcount data, by district

On the PHC headcount indicator for September, more of the facilities in City of Johannesburg and Ekurhuleni had data quality issues. Approximately 10% of the Johannesburg PHC facilities had missing data (i.e. either the headcount register was unavailable at site, or the DHIS data was not available). Significant over-reporting was a problem at approximately 55% of the City of Johannesburg facilities, and significant under-report was a problem at about 10% of the City of Johannesburg facilities. On PHC headcount, more of the Tshwane and Sedibeng facilities did well (i.e. either 100% accurate, or an over or undercount of less than 5%).

Although City of Johannesburg consistently had the most PHC facilities with missing data (possibly because some of the data is captured centrally and not on-site), they did not consistently have the poorest data quality across the different programmes. In fact, for the data element checking the number of children younger than one year of age that are fully immunized, Tshwane Metro facilities performed poorest—despite having the best data quality for the headcount indicator reported above. Over-report tended to be a significant problem for the PHC headcount indicator—in Tshwane Metro, 62% of the assessed PHC facilities over reported on this indicator, in Ekurhuleni Metro 50% over-reported, and in City of Johannesburg 49% of the assessed sites over reported on headcount. Financial incentives in the health budget allocation system may be partly responsible for the tendency to over-report on this indicator.

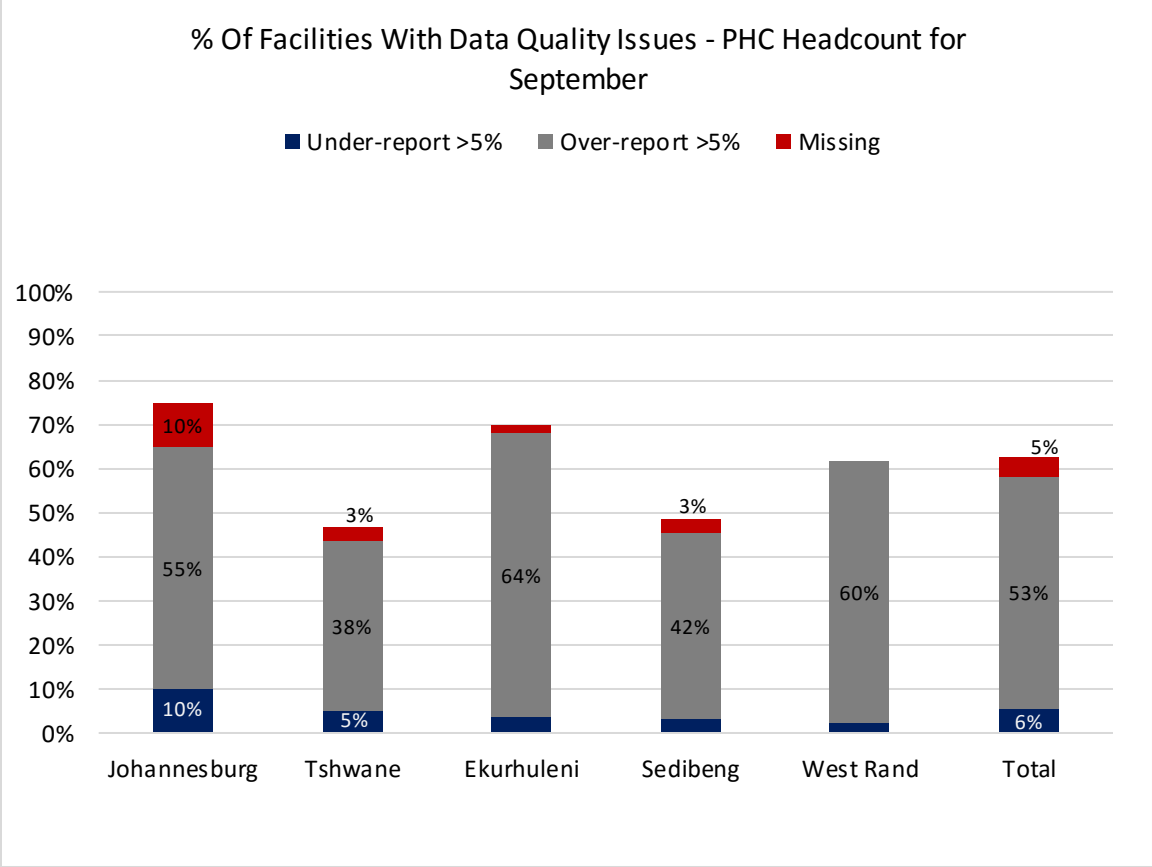


Figure 11: Percentage of facilities with mostly accurate data by district - PHC headcount

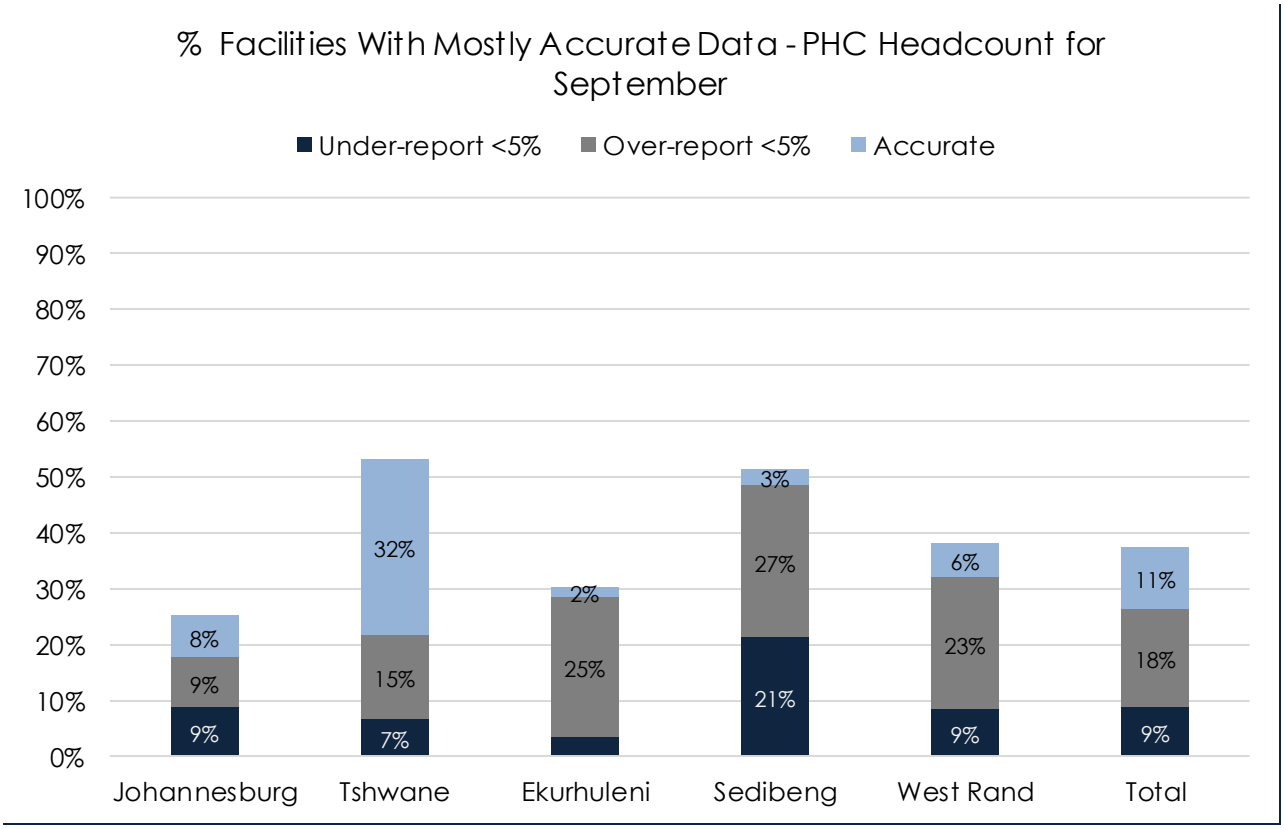


Figure 12: Percentage of facilities with data quality issues by district – PHC Headcount

Infant immunisation, by district

Immunisation data were well maintained in the West Rand (43% accurate) and Sedibeng (36% accurate) districts, but was lacking in quality on the other districts, especially in Tshwane. Over reporting was the biggest problem for this data element, although the West Rand had a tendency to over repot their results. The quality of data in PHC facilities is depicted in the two figures below.

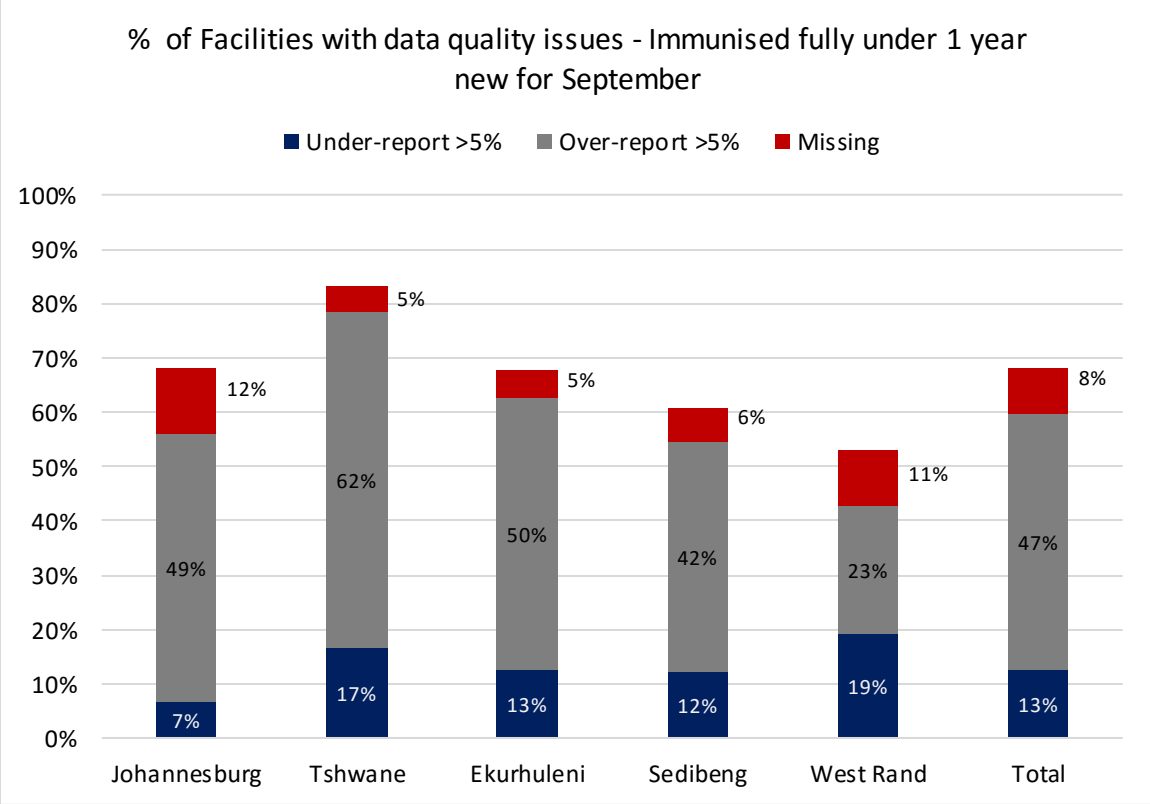


Figure 13: Percentage of facilities with data quality issues by district – Immunised

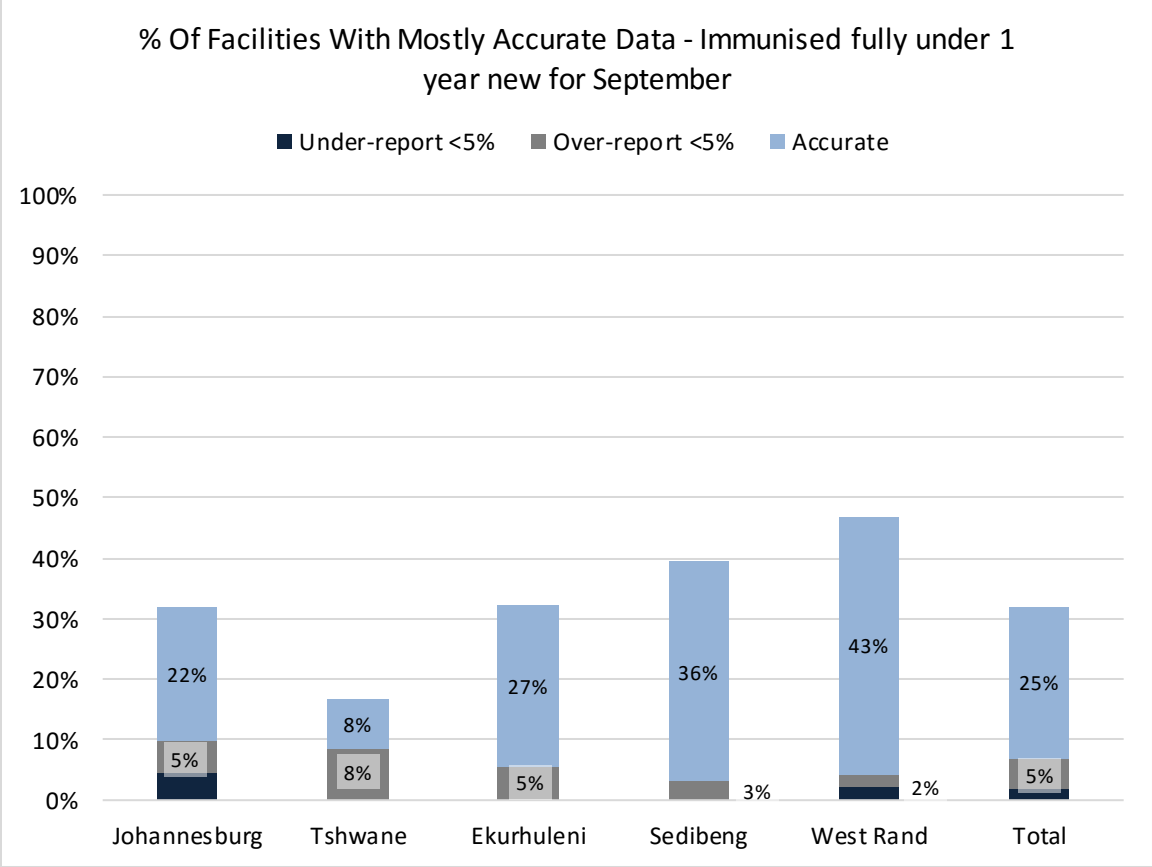


Figure 14: Percentage of facilities with mostly accurate data by district – Immunised

TB treatment initiated, by district

More sites in Ekurhuleni (36%) and West Rand (49%) had accurate data on the TB element related to the initiation of TB treatment for TB suspects 5 years and older. More than a third of the Johannesburg facilities (35%), had missing data for this element, and more than a third of the Sedibeng sites (36%), had significant problems with under reporting this data. Missing data on this element seemed to be a problem for approximately 32% of the PHC facilities across Gauteng—possibly because TB registers, more than others, were not accessible at the assessed PHC facilities.

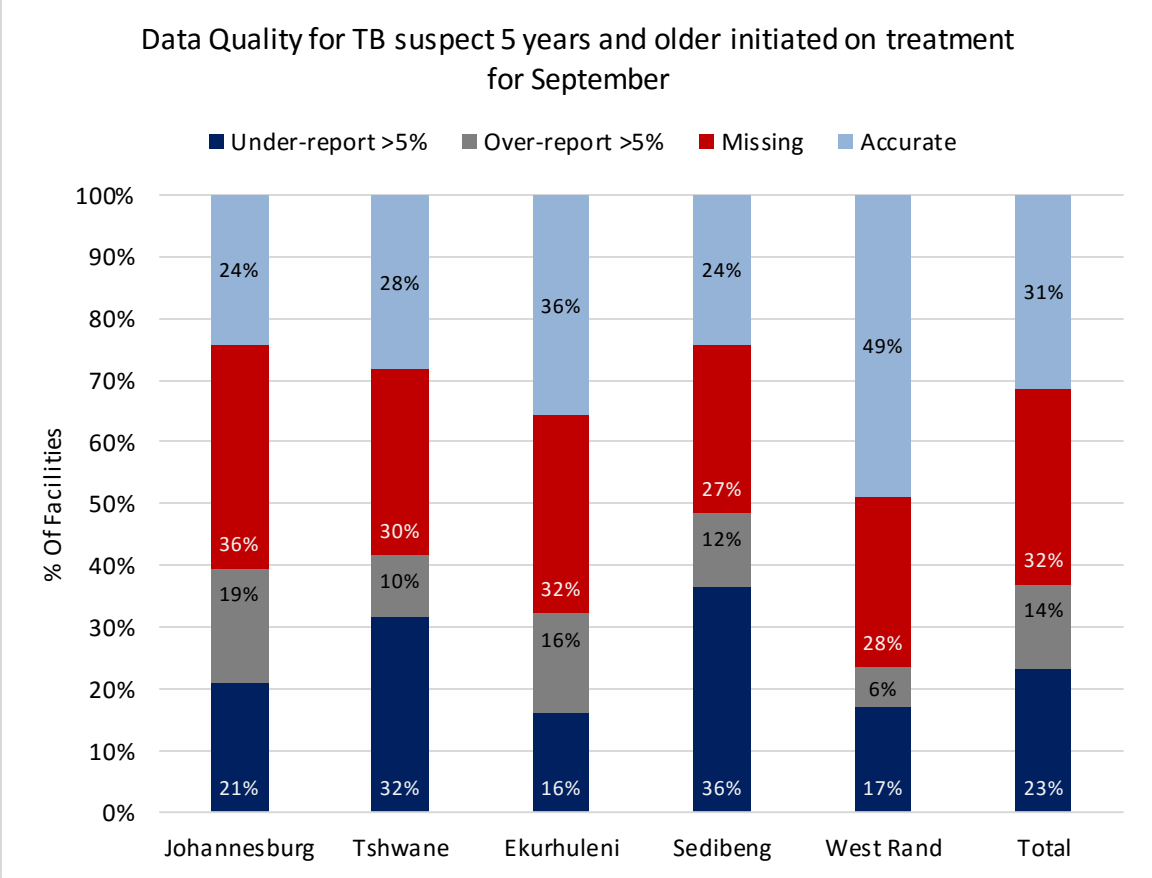


Figure 15: Data Quality for TB suspect 5 years and older initiated on treatment for September

Antenatal ART initiation, by district

More of the Gauteng PHC facilities (34%) had accurate data on the data element recording the number of Antenatal clients initiated on ART, than on any of the other assessed data elements. More of the Ekurhuleni and West Rand PHC facilities particularly, had accurate data for this data element. Missing data was a problem at 9% of the assessed PHC facilities in Ekurhuleni, and at 13% in the West Rand. Once again, City of Johannesburg had a significant percentage of assessed PHC facilities with missing data, but this is because of the large percentage of clinics that fall under the local authority and does not deliver a comprehensive package of services. The figure below depicts the data quality at different facilities in the five districts.

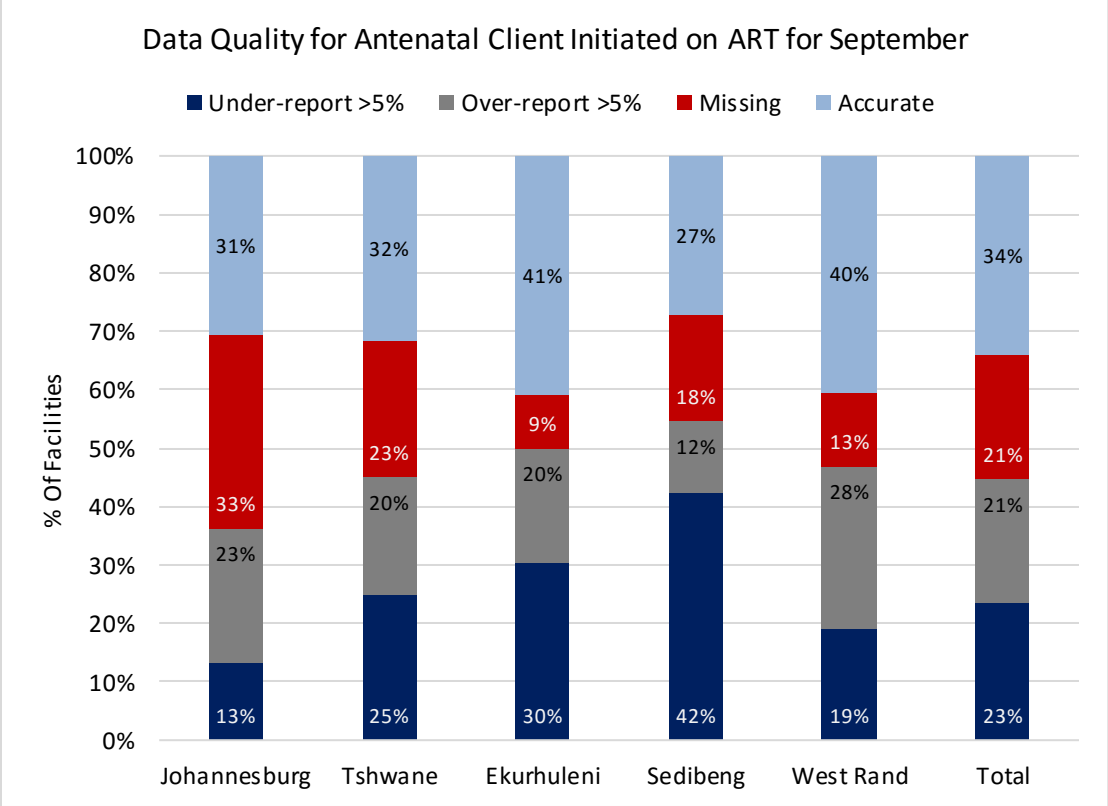


Figure 16: Data Quality for Antenatal Client Initiated on ART for September

Hypertension data, by district

Ekurhuleni facilities did comparatively well on the indicator related to the number of new hypertension clients. Half (50%) of the assessed Ekurhuleni facilities had accurate data. Sedibeng facilities performed worse than those in other health districts—48% had significant over reports, and 27% had significant under reports.

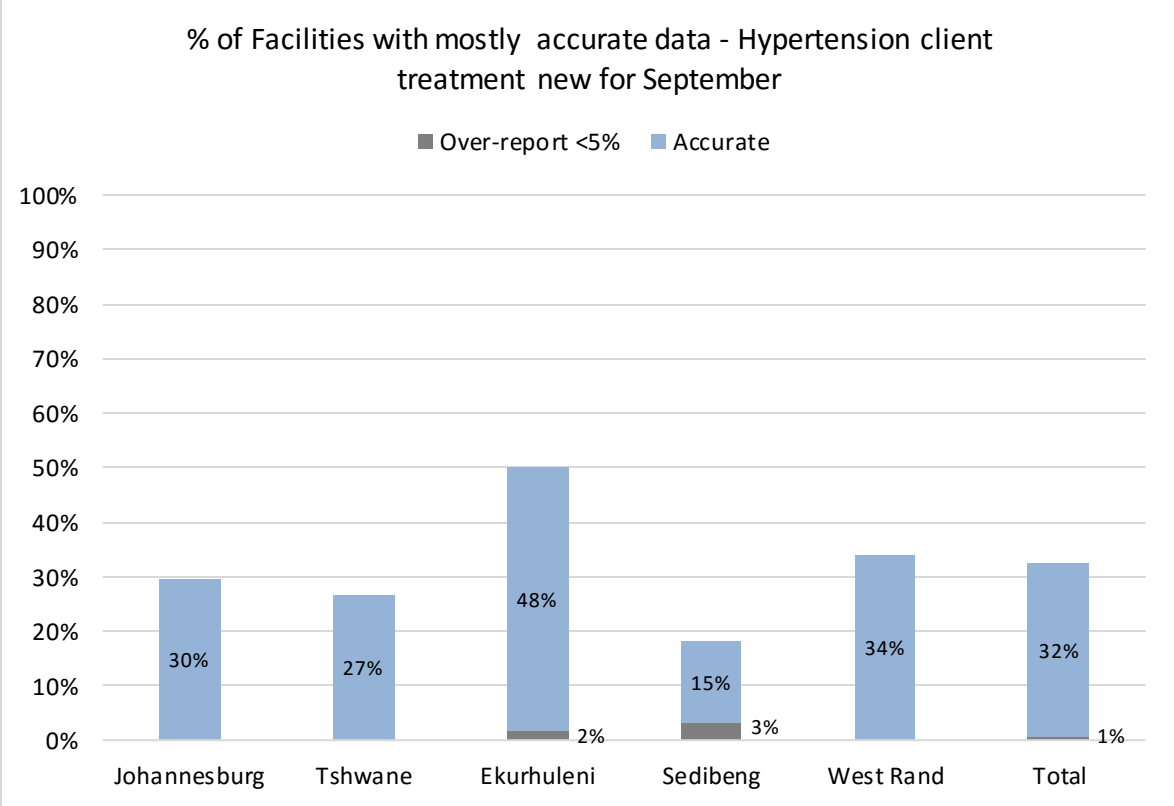


Figure 17: Percentage of facilities with mostly accurate data by district – Hypertension

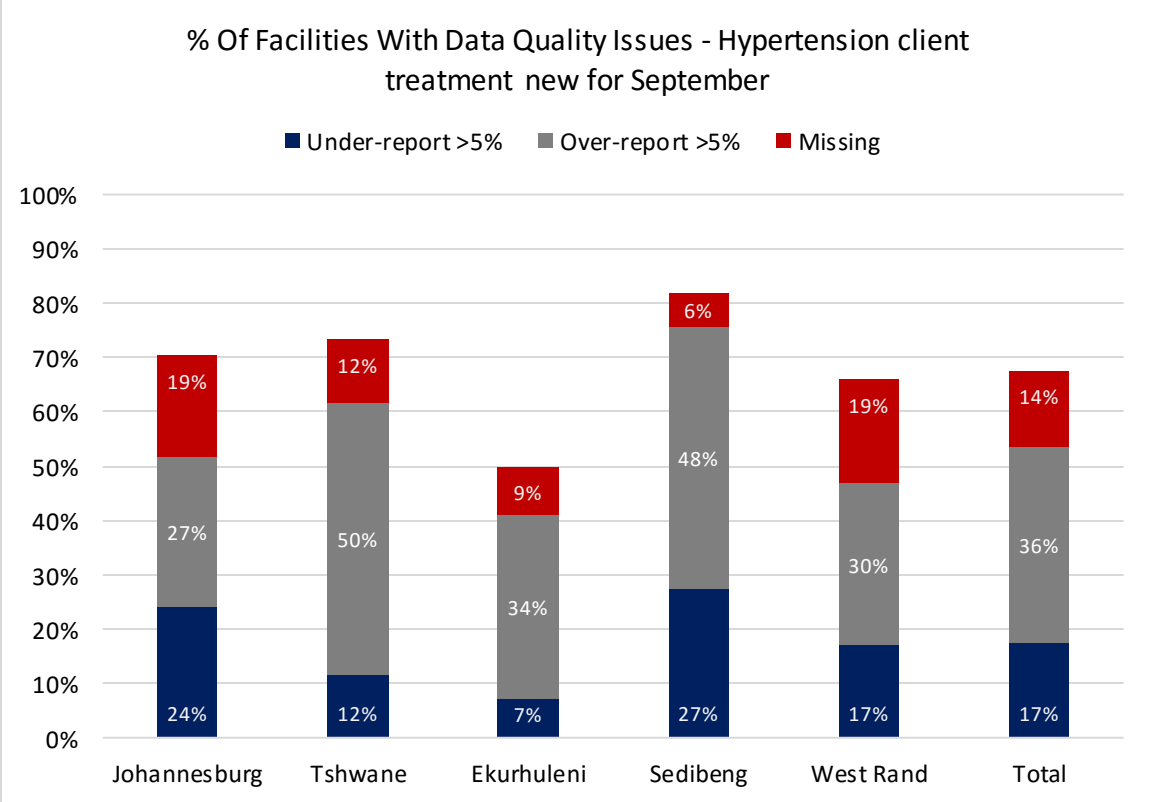


Figure 18: Percentage of facilities with data quality issues by district – Hypertension

- **Finding 4: Missing data are most prevalent in small clinics, while over-reporting is most prevalent in large clinics.**

The type of data quality issues varied by clinic size. Small clinics are those with a monthly headcount of less than 1 800 clients. Large clinics are those with a monthly headcount of more than 11 000 clients. The number of different sized clinics sampled in this assessment is depicted in Table 4: Number of assessed PHC facilities in each size or headcount category below.

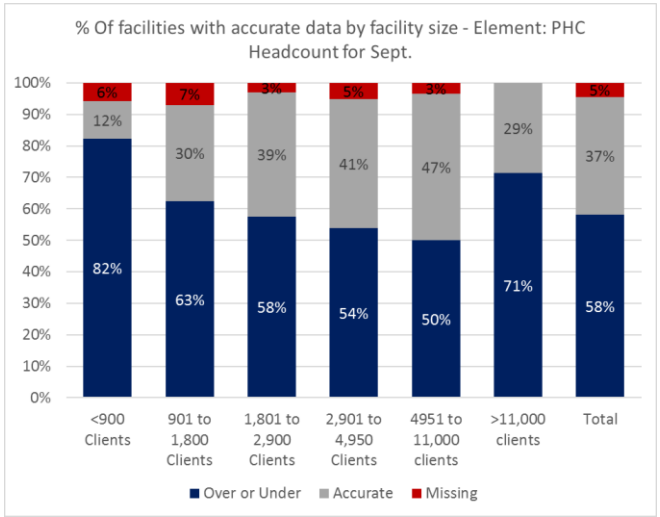
Table 4: Number of assessed PHC facilities in each size or headcount category[6]

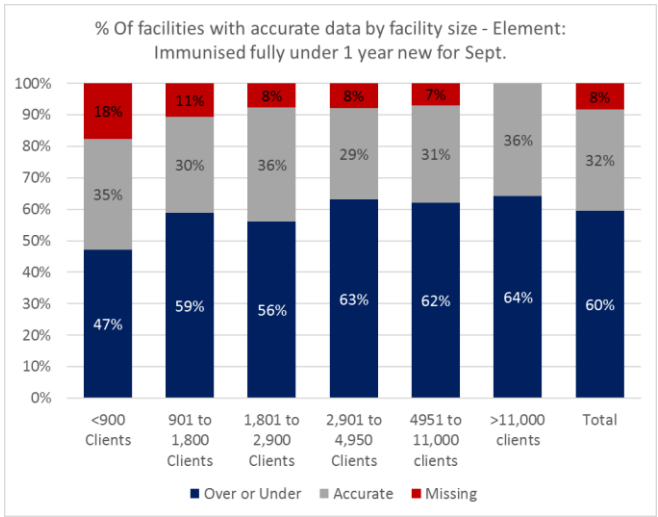
Value Label	Value	Frequency	Percent
<900 Clients	1	17	5.92
901 to 1,800 Clients	2	56	19.51
1,801 to 2,900 Clients	3	66	23
2,901 to 4,950 Clients	4	76	26.48
4951 to 11,000 clients	5	58	20.21
>11,000 clients	6	14	4.88
Total		287	100

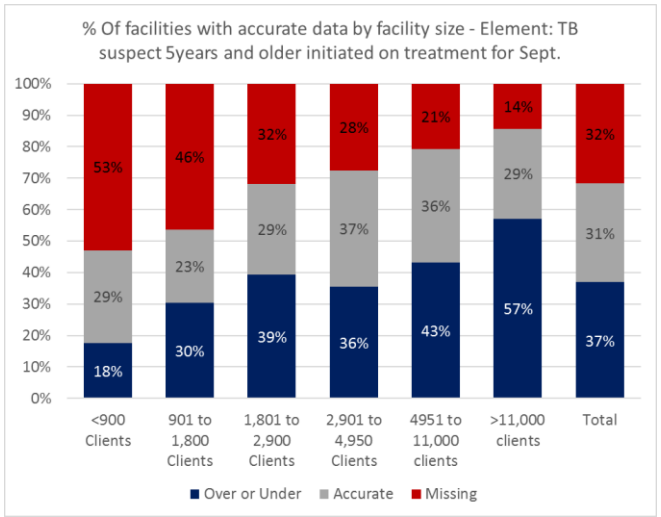
Approximately seventeen of the assessed PHC facilities had monthly headcounts smaller than 900, and a further fifty-six facilities had monthly headcounts between 900 and 1 800. Approximately fourteen of the facilities had monthly headcounts in excess of 11 000.

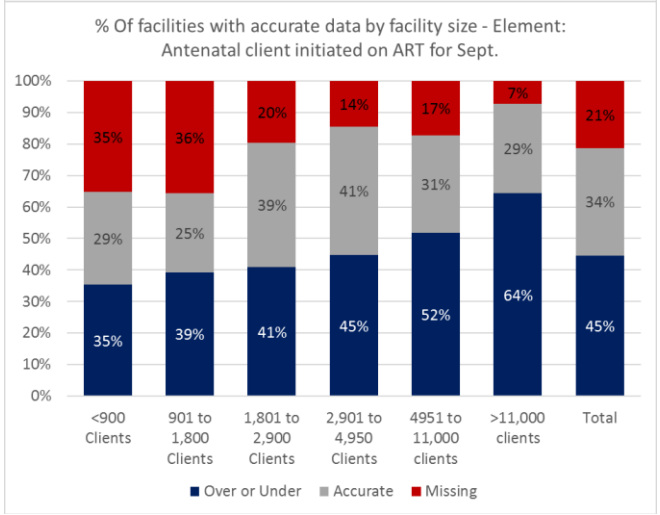
The five figures on the next page indicate that the small clinics tended to have more missing data—possibly because space for storing the registers may be limited on site or because staff responsible for data-management at these small facilities had to split their time between many different tasks and could not always keep track of the placement of registers. These figures also indicate that a disproportionately large percentage of the large PHC facilities tended to significantly over- or under-report on the various recounted data elements.

6 For the purposes of categorizing PHC facilities into size categories, the average recounted headcount for the period July, August, and September 2015 was taken. If recount statistics on headcount was not available for any of these months, the average headcount as per the DHIS for the same three months was used. Cut-off values around the 5th, 25th, 50th, 75th and 95th percentiles were taken to create six size categories.









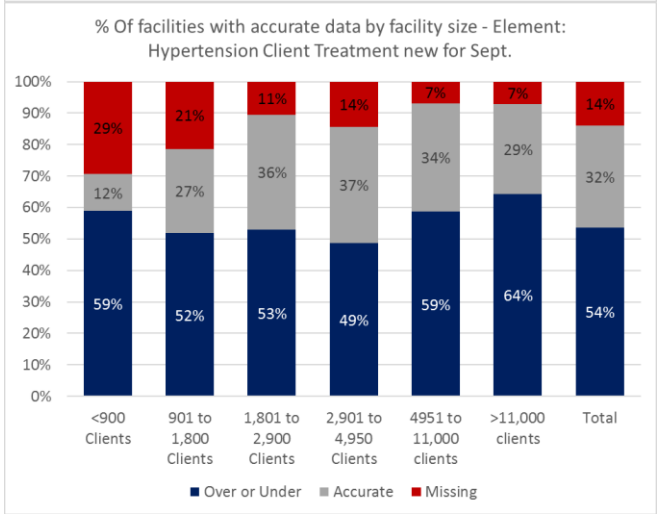


Figure 19: Percentage of facilities with mostly accurate data by facility size

A deeper investigation into the discrepancies at these large facilities indicates that they tended to over-report. This suggests that a significant improvement in the province level data quality can be achieved if these few large clinics are targeted for an intervention that aims to reduce over-reporting. The table below shows how over reporting takes place for certain health indicators in September at the large facilities only. As an example, ten facilities over-reported the PHC headcount by at least 5% in September. Eight large facilities over-reported the infant immunisation data by at least 5%. Similarly, six facilities over-reported the TB suspect data by at least 5% in September.

Table 5: Under- and over-reporting on select health indicators at large PHC facilities

	PHC headcount total—Sept	Immunised fully under 1 year new—Sept	TB suspect 5 years and older initiated on treatment—Sept	Antenatal client initiated on ART—Sept	Hypertension client treatment new—Sept
Under-report >5%	0	1	2	1	1
Under-report <5%	1	1			
Accurate	0	2	4	4	4
Over-report <5%	3	2			0
Over-report >5%	10	8	6	8	8
Missing	0	0	2	1	1
Valid Total	14	14	12	13	13
Total	**14**	**14**	**14**	**14**	**14**

- **Finding 5: Data quality issues such as missing data are problematic at up to 40% of PHC facilities . Human error contributes to these data quality problems.**

Summary of discrepancies between recount and reported data

Two questions were asked to determine the reasons for the possible discrepancies between the recount data and summary data. These two questions were:

1. Enter reason for registries / all documentation not being provided
2. Provide a reason from the data capturer why the summary and the count do not match

This summary includes the results for these two questions across all the PHC facilities for the three months under review (July, August and September 2015) with the detailed type and number of reasons provided for the PHC headcount indicator only. Since reasons for discrepancies on the other indicators tended to be similar and less voluminous, they are only reported when particularly illuminating.

Reasons why registers or all documentation were not provided

Interviewees provided four reasons for why registers or all documents could not be provided to the Assessment team. These are listed below from the most common, to the least common, response.

1. All or some of the registers or documentation were missing[7] i.e. could not be obtained or found at the facility, or the person did not know what happened to the registers
2. Registers or documentation were at another site or facility (e.g. the sub-district, or being audited)
3. Data capturers or clerks not available to provide the requested registers and documents
4. Loose pages used as registers not available (Ran out of register or late delivery of registers to PHC)

[7] Missing (refers to those files that could not be found at the time of the audit by the relevant personnel at the PHC)

Table 6: Types of reasons provided at PHC facilities for registries or all documentation not being provided for PHC headcount

Types of Reasons	PHC headcount				
	July	**Aug**	**Sep**	**Total (No.)**	**Total %**
All registers missing and some missing, Don't know what happened to registers, Unattainable	39	24	23	86	91%
Data capturer/derks not available	2			2	2%
Loose pages used as registers not available (Ran out of register or late delivery of registers to PHC)	2			2	2%
Registers taken off site (sub-district, audit)	3	1		4	4%
Number of PHCs that provided a reason	**46**	**25**	**23**	**94**	**100%**

Reasons provided by the data-capturer for why the summary and the count data do not match.

Fieldworkers were asked to obtain reasons from the data-capturer for why the summary and the count data did not match at the time of the quality audit. A total of fourteen reasons were most commonly provided and are listed below in order or frequency, from most to least frequently cited.

Most frequent	1. Inaccurate filling in of registers (ticks not in line, no dates, no age, using cross and numbers) 2. Some registers or summaries missing/lost/technical issue 3. Human error in counting totals
Less Frequent	4. Using sources other than the registers (e.g. notebooks, consulting rooms, loose papers) 5. All registers or summaries not attainable 6. Under reported 7. Over reported 8. Do not know
Least Frequent	9. Include other statistics (mobile clinics, consulting rooms, dentists) 10. Data Capturer not available at time of recount (on leave, working night shift) 11. Registers off-site 12. New registers 13. Newly appointed Data Capturer 14. Use of both manual and electronic system

The most commonly cited reason at these facilities provided by the data-capturer for data discrepancies have to do with inaccurately filling in the registers. These inaccuracies include not ticking the relevant boxes, no dates on the registers, no ages provided for patients, using crosses instead of ticks which were hard to read, ticks crossing over two blocks and a number entered instead of ticks Another reason that was most cited was that some registers or summaries were either missing or lost because renovations were being done; they were somewhere in the storage facilities, or they were in a doctor's office, or there were technical issues with the computers at the facility. Human error in counting the ticks and totals was also a frequently mentioned reason

for why discrepancies occurred when comparing the summaries for the month and the recount totals.

Reasons that were less frequently cited include; Using data sources other than the prescribed registers (for e.g. notebooks, consulting room books, loose papers); all registers or summaries not attainable; Under reporting of the counts in the monthly summaries; over reporting of counts in the monthly summaries; and the data-capturer not knowing the reason for why there is a discrepancy.

Reasons that were the least frequently cited include; adding other statistics from mobile clinics, consulting rooms and dentists into the monthly counts that were not counted at the time of the recount. Some data capturers were not available at the time of the recount to also provide reasons for the discrepancies as they were either on leave or work the night shift. Another reason was that registers were off-site and could not be counted, and that in August new registers were employed through the rationalisation process that may have led to challenges in counting the data. The last reason provided was that the data capturer at the facility was newly appointed and s/he could provide no explanation.

Table 7: Types of reasons provided by the data-capturer why the summary and the count data do not match PHC headcount

		PHC Headcount Indicator				
	Reason Types	**July**	**August**	**Sep**	**Total**	**%**
1	Inaccurate filling in of registers (ticks not in line, no dates, no age, using cross and numbers)	60	86	70	216	37%
2	Some registers or summaries missing/lost/technical issue	57	44	43	144	25%
3	Human error in counting totals	32	39	46	117	20%
4	All registers or summaries not attainable	4	7	8	19	3%
5	Using sources other than the registers (e.g. notebooks, consulting rooms, loose papers)	7	6	6	19	3%
6	Under reported	5	4	5	14	2%
7	Over reported	4	5	5	14	2%
8	Do not know	6	2	4	12	2%
9	Indude other statistics (mobile dinics, consulting rooms, dentists)	3	5	1	9	2%
10	Registers off-site	4	1	1	6	1%
11	Data capturer not available at time of recount (on leave, working night shift)	3	2	1	6	1%
12	New registers		2	1	3	1%
13	Newly appointed data capturer		1		1	0%
14	Use of both manual and electronic system		1		1	0%
Total number of PHCs that have provided reasons		**185**	**205**	**191**	**581**	**100%**

Additional reasons provided by the data capturers for the discrepancies in the summary data and the re-count data are provided below:

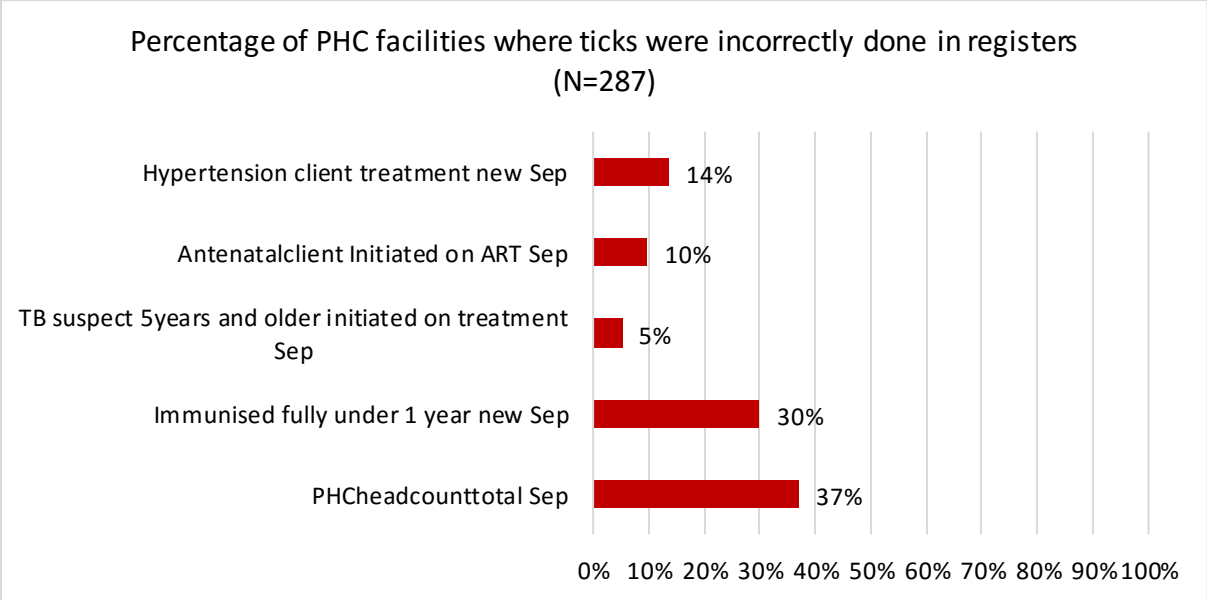
Indicators Reviewed	Period of review (July, August and September 2015)
HIV indicators	
Clients tested for HIV including ANC	HCT and ANC registers were counted in July and for August and September the counts were to be obtained from the comprehensive PHC register. For this indicator the reasons also covered some registers and summaries being missing, human error in counting and calculating as well as inaccurate completing of the registers. It was also reported for this indicator at a few PHC's that the correct register was not being used and this is element was captured on pieced of paper other than the register. The reason for the inaccurate completing of register was due to ticking both the negative and positive columns, which spoilt the entry. At sites were there was a monthly summary it was noted that a few of these sites that this data element did not appear on the monthly summary across the three months under review. At one clinic it was noted that this indicator for September did not appear on the system (TIER.net). At times statistics from other sources for example, from campaigning were included in the summary forms and no hardcopy record of this was available.
TB Indicators	
TB lost to follow up	Additional reasons provided for this indicator included that the summary report did not include this indicator. This was seen for some of the PHC facilities across the three months under review, with this being more frequent in the months of August and September than in July. In addition, discrepancies were also noted due to TB registers not being circulated within the PHC hence it had not been used.
TB symptom 5 years and older screen rate	At many of the PHC facilities for this indicator the TB symptom-screening book was not provided. In the few instances where the tool was not provided, the PHC registers were used for this indicator. In cases were the tools were provided it was inaccurately completed or not completed or used at all. It was noted at a few PHC facilities that The nurses do not use the screening materials as they have too much work to do. At one facility it was reported that the Adult Patient form for TB screening was used instead of the TB symptoms screening book. It was also noted that notebooks were either used or the only records of this indicator were kept in patient files. At some PHC facilities this indicator was not provided in monthly summary tool. For this indicator it was also noted that patients were being screened by students off-site, which may have resulted in data discrepancies.
Maternal Health Indicators	
Mother postnatal Visit	No additional reasons—there were a significant number of facilities that did not have maternity wards and hence this indicator was not counted
Cervical Cancer Screening	For this indicator the addition reason was that this indicator was not included and could not be found in the summary forms

- **Finding 6: Fewer than 5% of PHC facilities do not always use official registers**

Official registers are not always used, but this is only a problem in fewer than 5% of PHC facilities. The exception seems to be data elements that record Infant PCR tests and PHC headcount—these were recorded on unofficial registers in 8% and 10% of health facilities respectively. A reason for this good, consistent use of the registers may be due to the efforts to rationalise registers, which started in August and began showing positive effects in September.

- **Finding 7: Ticks are not properly done in registers. Headcount and MCWH indicators are vulnerable to this issue**

Basic errors, such as irregular ticks in the registers are common, especially in data elements relating to headcounts and Mother, Child and Women's Health (MCWH). No difference in the tick accuracy was observed when July, August and September data was compared, indicating that if the users of the registers have poor ticking habits, these tend to be maintained over time. Fieldworkers observed that inaccuracies include not ticking in the relevant boxes, no dates on the registers, no ages provided, using crosses instead of ticks, ticks crossing over two blocks and numbers entered instead of ticks. Improving the accuracy of ticks, may require a special awareness campaign by the DOH, and could lead to a quick gain in terms of this aspect of data quality assurance.



- **Finding 8: Recount at hospitals were not as useful to identify data quality issues**

The methodology and tools used to conduct the recount at PHC facilities, were adapted for use in hospitals, but proved to be less useful in determining data quality. This is because the

comparison of data that is directly captured onto the Medicom system (used in a few hospitals) is unlikely to differ from that which is captured in the DHIS records. Discrepancies were found on some indicators, but it was often because the recount was incomplete due to the volume of data that had to be processed at hospitals. Annex C contains a comparison of the DHIS data to the recounted data, together with an indication of the reasons why discrepancies were found.

- **Summary of the Recount Findings**

The evidence provided in this section substantiates the finding that, at most Gauteng PHC facilities the quality of data needs to be improved. While performance across programmes and data elements varied, only about 10 percent of facilities had accurate headcount data, and about one third had accurate data on other important health-related elements. The accuracy of headcount data improved over time with the rationalization of registers but no other data improved in this time period. The biggest concern for data quality is over reporting, missing data, and poor recording in the registers.

Data quality across different districts varied greatly across the different indicators, but City of Johannesburg facilities, more than others, struggled with missing data. Sedibeng facilities, tended to underreport more than other facilities. West Rand facilities tended to perform better on a few of the assessed data elements.

At large clinics (i.e. those with monthly headcounts more than 11,000 clients) over-reporting was a problem on most of the assessed data elements. At small clinics (i.e. those with monthly headcounts smaller than 1,800 clients) missing data was a problem. Missing data was pronounced on data elements related to procedures that do not occur frequently at PHC facilities (e.g. deliveries and medical male circumcision), and on those requiring long-term tracking or waiting for test feedback (e.g. some of the TB elements, diabetes, and Infant 1st PCR test positive around 6 weeks). Improving the clinical laboratory interface is a possible quick gain that may yield benefits both in terms of data quality and clinical management of cases.

The quality of record keeping at times leaves much to be desired. Ticks are not properly done in registers, and especially headcount and MCWH indicators are vulnerable to this issue. Official registers are not always used, and although this is only a problem in a small percentage (<5%) of PHC facilities, it compromises the overall quality of health data, and subsequently of health-related policies based on this flawed information.

Explanations for data discrepancies provided by data quality officers suggest that data quality problems are mainly the result of human error. To address this issue, data quality officers need

to better understand *why* human errors are occurring and *how* they can take steps to minimize *these* errors.

3.2 Level of Technical Resourcing

- **Finding 9: Most facilities have the necessary IT equipment to capture data, but the majority of computers are used by more than one person for the purpose of capturing more than one dataset.**

Most facilities have access to functioning computers to capture data but usually more than one person can use any given computer, often for the purpose of capturing more than one dataset. Although this might be an efficient way to use technical resources it could negatively influence data quality since the volume of data to be captured requires dedicated computer terminals to be available. Issues that could be improved are; regular backups of data at facilities, rollout of up-to date DHIS versions, and connectivity to the Internet.

Background

The health sector's Negotiated Service Delivery Agreement (NSDA) 2010–2014, produced in 2010 observed: Although large sums of money have been used to procure health ICT and HIS in South Africa in the past, the ICT and HIS within the health system, are not meeting the requirements to support the business processes of the health system thus rendering the healthcare system incapable of adequately producing data and information for management and for monitoring and evaluating the performance of the national health system. (Source: SAHR 2013/2014)

The South African Health Review (SAHR) of 2013/2014, in its chapter on eHealth Strategy refers to the findings of a landscape analysis completed by CSIR and its partners, which identified a total of 42 health information systems being used in the public sector[8]. The SAHR provides information on health information of systems used in Gauteng province. Some of which are: District Health Information System (DHIS), national Electronic TB Register, the national three-TIER monitoring and evaluation system for antiretroviral therapy, government-owned Patient Administration and Billing System, Gauteng Medicom; Soarian MedSuite and PharmAssist.

[8] Council for Scientific and Industrial Research (CSIR). Health Information Systems in Public Healthcare Facilities in South Africa: A review of the current status and trends—Version 1.0. February 2013: unpublished report

In 2012 the eHealth Strategy for South Africa 2012–2016 was produced in an effort to coordinate the implementation of eHealth projects and programmes, and guides efforts to strengthen information management. One of the strategic priorities of the eHealth Strategy is eHealth Foundations which aims to deploy eHealth capability in a step-wise manner. With the four areas identified as providing the foundations for all other eHealth activities: infrastructure, connectivity, registration of patients, facilities and providers, and a basic national electronic health record. [9] On 23 April 2014, the Ministry issued the National Health Normative Standards Framework for Inter-operability in eHealth. This framework aims to address the inability of the ICT systems to exchange or share information or data between heterogeneous systems.

Data collection

The facility manager and senior data person in each facility responded to the level of technical (ICT) resourcing available to capture data. The following questions were asked:

- How many computers do the data personnel use to capture data? What is the functionality status of each computer?

Each computer was named using the location of the computer the user or primary use of the computer. Using this as a basis the following status indicators were collected for each computer:

- Functionality: Fully functional, Some issues, Not functional
- Multiple uses: Yes (the computer is used for capturing multiple datasets or for other uses such as administrative tasks), No
- Multiple users: Yes indicates more than one primary user
- Datasets: TB, HIV, PHC, Other
- DHIS Version
- Antivirus Updated: Yes, No
- Last Backup: Today, Yesterday, This week, This month, More than a month ago
- Connectivity: Not connected, Basic connectivity (email capacity), Full connectivity (email plus other web access)

Table 8: Technology resources table (included in systems tool)

Computer Station	Function-ality	Multiple Uses	Multiple Users	Data Sets	DHIS Version	Antivirus Updated	Last Backup	Connectivity
(Identify computer as reception, TIER, PHC, facility manager i.e. where it is OR who it is used by OR primary use)	Fully functional	Yes	Yes	TB		Yes	Today	Not connected
	Some issues	No	No	HIV		No	Yesterday	Basic connectivity
	Not functional			PHC			This week	Full connectivity
				Other			This month	
							More than a month	

9 Masilela T.C, Foster R, Chetty M. The eHealth Strategy for South Africa 2012-2016: how far are we? In: Padarath A, English R, editors. South African Health Review 2013/14. Durban: Health Systems Trust; 2014. URL: http://www.hst.org.za/publications/south-african-health-review-2013/14

Computers used for data capture

All but three facilities visited had access to at least one computer to use for data capturing activities. The three clinics reported to not have any computers are:

Table 9: Facilities with no computers

District Name	Sub-district name	Site Name
Sedibeng	Sedibeng—Emfuleni	**Zone 14 Clinic**
Johannesburg	Johannesburg C	**Cosmo City Clinic**
West Rand	West Rand—Merafong City	**Kokosi Clinic**

Two facilities in Tshwane, namely Sokhulumi Clinic and Soshanguve 2 Clinic, had access to 8 computers. Many facilities (110 or 38%) had access to two computers, while 79 facilities (28%) made use of only one computer. Overall, a total of 664 computers were reportedly used in the 286 assessed facilities. The percentage of facilities per district with no or up to 8 computers for data capture is illustrated below:

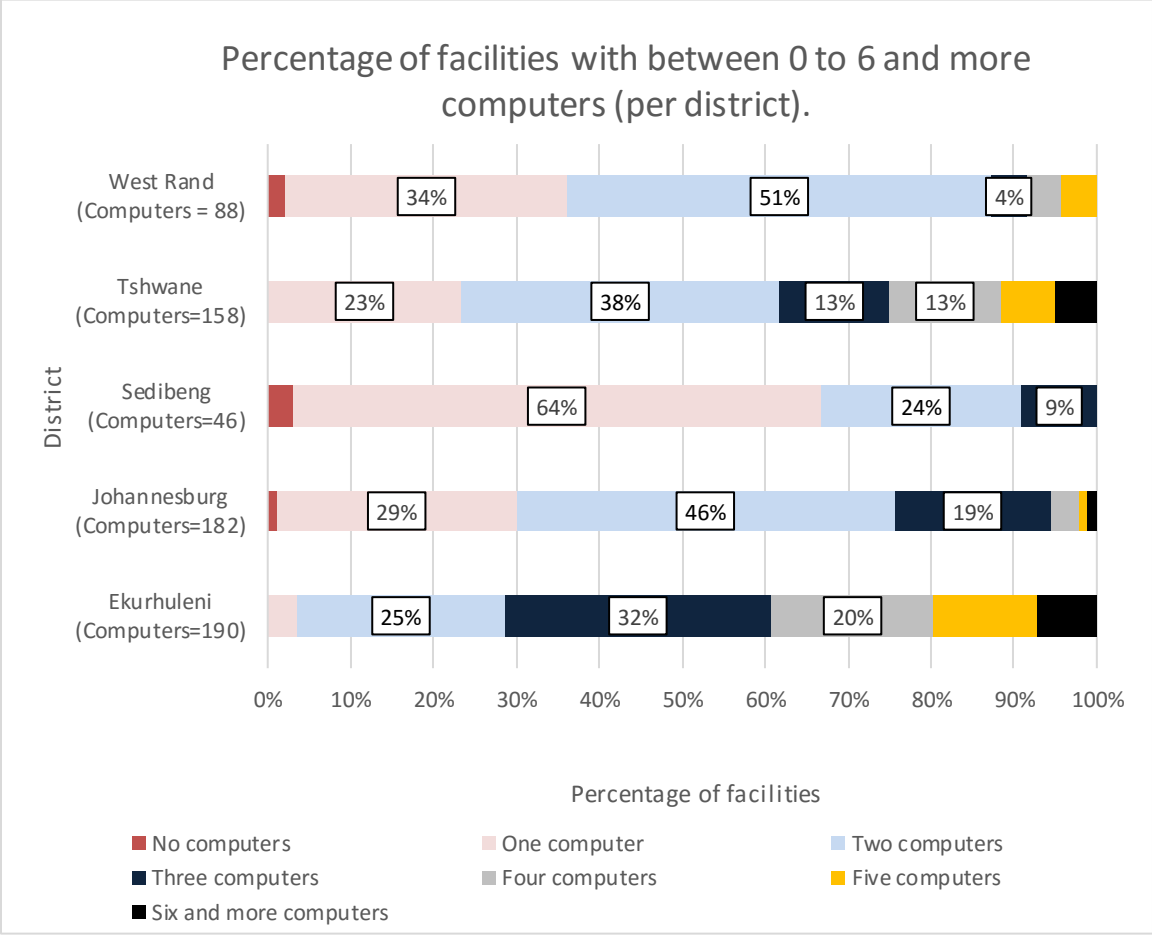


Figure 20: Number of computers per facility

Ekurhuleni district had the highest number of computers per facility with an average of 3.4, while Sedibeng had the lowest number on average namely 1.4.

Functionality of computers

A good percentage of computers (87%) were reported to be fully functional and staff members were able to utilise the technology at the facilities to capture data. Sedibeng district had the lowest number of computers. It also had 16 computers (35%) that did not function optimally. In the Tshwane district only 7 computers (4%) were reportedly as not fully functional.

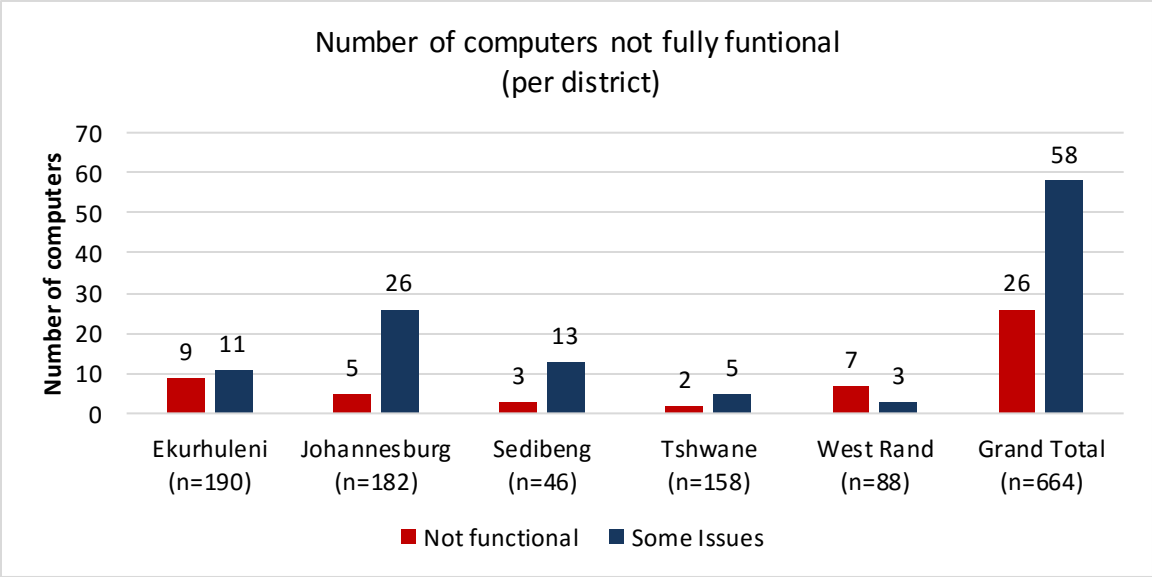


Figure 21: Number of computers not fully functional

In addition, sixteen facilities had only one computer that was not even fully functional for data capturing. This likely has a negative influence on accurate and timely data entries. Of these facilities, half had a data score of less than 75% or no data.

Table 10: List of clinics with only one computer that is not fully functional

District Name	Sub-district name	Site Name	Data Quality Score	Functionality
Ekurhuleni	Ekurhuleni East 1	Daveyton East Clinic	91.48%	Some issues
Johannesburg	Johannesburg C	Rex Street Clinic	69.05%	Some issues
Johannesburg	Johannesburg C	Davidsonville Clinic	73.01%	Some issues
Johannesburg	Johannesburg C	Tshepisong Clinic	1.53%	Some issues
Johannesburg	Johannesburg E	Thoko Mngoma Clinic	No data	Some issues
Johannesburg	Johannesburg G	Protea South Clinic	85.37%	Some issues
Johannesburg	Johannesburg G	Lenasia Ext 10 Clinic	96.08%	Some issues
Johannesburg	Johannesburg G	Wildebeesfontein Clinic	23.42%	Some issues
Sedibeng	Sedibeng—Emfuleni	Driehoek Clinic	77.55%	Not functional
Sedibeng	Sedibeng—Emfuleni	Retswelapele Clinic	91.1%	Not functional
Sedibeng	Sedibeng—Emfuleni	Zone 13 Clinic	98.97%	Some issues

Sedibeng	Sedibeng—Emfuleni	Dr Helga Kuhn Clinic	0%	Some issues
Sedibeng	Sedibeng—Lesedi	Ratanda Ext 7 Clinic	35.22%	Some issues
Tshwane	Tshwane 1	Ga-Rankuwa View Clinic	92.59%	Not functional
Tshwane	Tshwane 3	Saulsville Clinic	43.86%	Some issues
Tshwane	Tshwane 6	Stanza Bopape CHC	93.34%	Some issues

Analysing the number of fully functional computers per 5 000 PHC headcount, shows that 43% of facilities had between one and three fully functional computers (this is the DHIS system average for three months from July to September 2015). Only 20% of facilities that had available data reported having one or less fully functional computer per 5 000 PHC headcount. The Sedibeng district, where 39% of facilities had access to one or fewer fully functional computers, had the biggest ICT challenges. In Johannesburg, a quarter of facilities faced the same challenges.

Table 11: Number of facilities with fully functional computers per 5000 PHC headcount

Sub-district	0	More than 0 up to 1	More than 1 to 3	More than 3 up to 6	More than 6 up to 10	More than 10	No data	Total	Average PHC headcount (DHIS)
Ekurhuleni East 1	1	1	4	2				8	7145
Ekurhuleni East 2			2	1	2	1		6	3969
Ekurhuleni North 1			4		1		1	6	5534
Ekurhuleni North 2	1	1	4	2		1		9	5751
Ekurhuleni South 1			7	2	2			11	6233
Ekurhuleni South 2		2	8	3	3			16	7337
Johannesburg A			4	1		1	1	7	3548
Johannesburg B			7	2	1			10	2532
Johannesburg C	3	2	4	1			1	11	5277
Johannesburg D	1	7	9	2	1		1	21	9130
Johannesburg E	1		2	3			1	7	4055
Johannesburg F	1	2	5	5				13	4402
Johannesburg G	6	1	7	4	3			21	4327
Sedibeng—Emfuleni	7	4	11	3			1	26	4931
Sedibeng—Lesedi	1			3				4	1771
Sedibeng—Midvaal	1		2					3	3921
Tshwane 1	1	4	5	8	1			19	7609
Tshwane 2		1	3	5				9	4383
Tshwane 3	1	3	3	4				11	5934
Tshwane 4		1	2	2				5	3746
Tshwane 5			3			1		4	5432
Tshwane 6	1		5	1		1		8	6768
Tshwane 7			1	1	1	1		4	1923
West Rand—Merafong City	1		8	3	2			14	2819

Sub-district	0	More than 0 up to 1	More than 1 to 3	More than 3 up to 6	More than 6 up to 10	More than 10	No data	Total	Average PHC headcount (DHIS)
West Rand—Mogale City		2	9	4	2	1		18	3055
West Rand—Randfontein			2	4	1	1		8	3088
West Rand—Westonaria			3	2		2		7	3548
Grand Total	**27**	**31**	**124**	**68**	**20**	**10**	**6**	**286**	**5137**
Percentage	**9%**	**11%**	**43%**	**24%**	**7%**	**3%**	**2%**	**100%**	

Table 12: Percentage of facilities with fully functional computers per PHC headcount

District	0	More than 0 up to 1	More than 1 to 3	More than 3 up to 6	More than 6 up to 10	More than 10	No data
Ekurhuleni	4%	7%	52%	18%	14%	4%	2%
Johannesburg	13%	13%	42%	20%	6%	1%	4%
Sedibeng	27%	12%	39%	18%	0%	0%	3%
Tshwane	5%	15%	37%	35%	3%	5%	0%
West Rand	2%	4%	47%	28%	11%	9%	0%
Total Percentage	**9%**	**11%**	**43%**	**24%**	**7%**	**3%**	**2%**

Although the technical challenges for the Tshwane districts seemed to be at an average level specific sub-districts namely Tshwane 1 and 3 experienced challenges. The table below lists the sub-districts where access to ICT might be a constraining factor in performing duties.

Table 13: List of sub-districts with a high percentage of facilities with 0 to 1 fully functional computers per 5,000 PHC headcount

Sub-district	No fully functional computers	More than 0 up to 1 fully functional computers	Total percentage No or up to one fully functional computers
Ekurhuleni East 1	13%	13%	25%
Ekurhuleni North 2	11%	11%	22%
Johannesburg C	27%	18%	45%
Johannesburg D	5%	33%	38%
Johannesburg F	8%	15%	23%
Johannesburg G	29%	5%	33%
Sedibeng—Emfuleni	27%	15%	42%
Sedibeng—Lesedi	25%	0%	25%
Sedibeng—Midvaal	33%	0%	33%
Tshwane 1	5%	21%	26%
Tshwane 3	9%	27%	36%

Datasets captured on computers

The facility manager and senior data person provided information pertaining to the purpose of each computer. This helped determine whether computers were used exclusively to capture a single dataset. Computers were classified as having multiple uses if they were used to capture more than one dataset, or if they were used for other tasks such as administrative functions.

A total of 491 computers (74%) were used either to capture multiple datasets, or for other activities. Thus, only 173 computers (26%) were exclusively dedicated to capture only one dataset. Sedibeng had only 1.4 computers on average per facility (7%), therefore it follows that the percentage of computers dedicated to capture a specific dataset would be low.

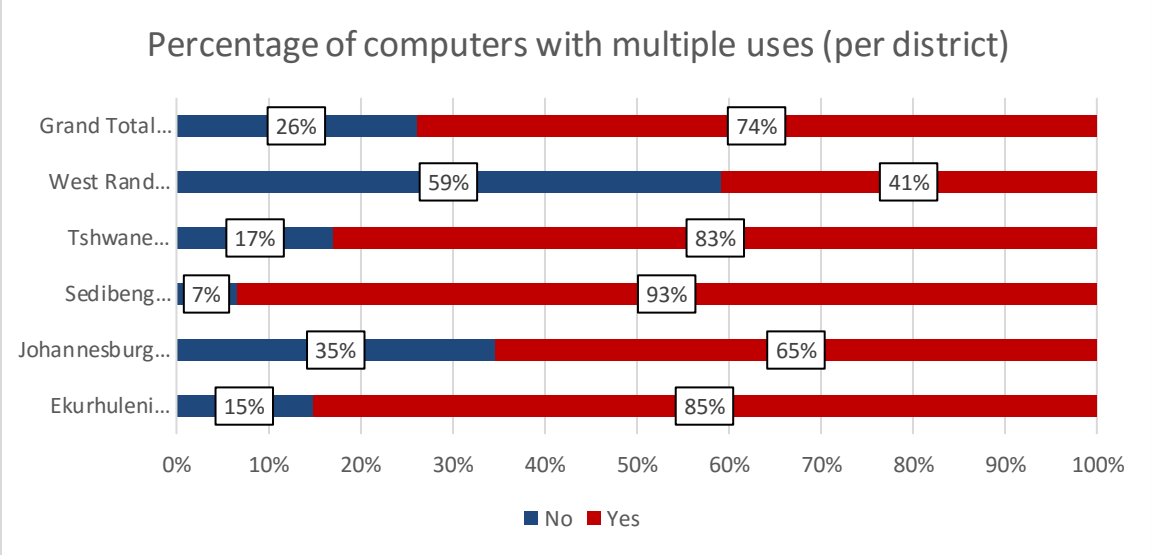


Figure 22: Percentage of computers used for multiple purposes

Computers dedicate to a single user

The graph below depicts that most computers were used by more than one person. Since the Sedibeng district had the lowest number of computers on average per facility it follows that the computers (91%) at these facilities would be used by more than one person. And yet, Ekurhuleni had many computers per facility but still most computers (84%) were used by more than one person. This might be an indication of the efficient use of resources but could lead to competition amongst users impeding availability.

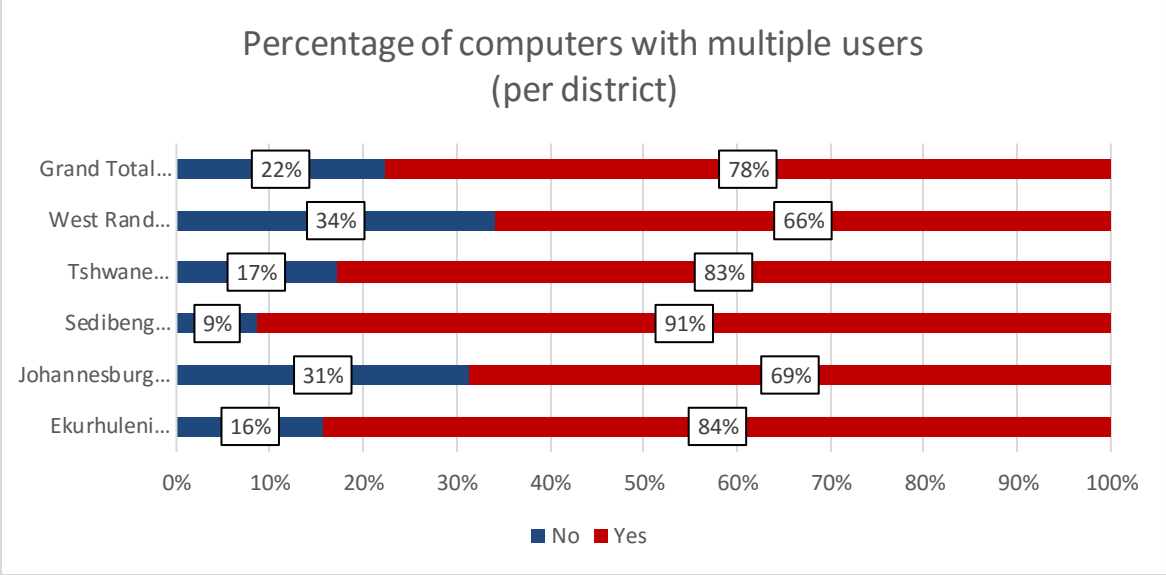


Figure 23: Percentage of computers used by more than one person

Datasets captured on the computers where multiple datasets are stored

As illustrated above, most of the computers at facilities were used for more than one purpose. The data below confirms this fact and shows that 43% of the 664 computers include the TB dataset, 64% include the HIV and 51% include the PHC headcount dataset. About 30% (197) of the computers host other datasets.

Table 14: Number of computers with specific dataset

	DATASETS			
	TB	**HIV**	**PHC**	**Other**
Number of computers	287	422	339	197
Total number of computers = 664				
Percentage computers	**43%**	**64%**	**51%**	**30%**

DHIS updates running on facility computers?

Users reported a diverse set of DHIS updates at facilities. For example, a few variants of DHIS update 1 were running on some computers. This could compromise the integrity of data quality if data are shared across platforms and software version, and it can also confuse data capturers if they forget which update they have most recently used.

For this analysis of the DHIS software, all the different variants of a specific update are grouped together such that update 1.8, 1.8.0, 1.803, 1.8.3, 1.8.300.0, 1.8.400.00 is reported as update 1.8 of the DHIS. The Results of the analysis concerning DHIS updates, which was conducted on 341

computers, are illustrated in the next table. The data for the remainder of the computers was not relevant or of a low quality. The facility personnel reported that about one third of computers (34%), for which usable data was available, had DHIS update 1 installed, roughly 24% of computers had update 1.8 installed, and about 20% had update 2 of the DHIS system installed.

Table 15: DHIS versions at facilities

District	Hard copy / *Pads	TIER.net	Update 1	Update 1.3	Update 1.8	Update 2	Reported updates higher than 2	Total
Ekurhuleni	2*	27	35	0	10	23	0	97
Johannesburg	2	12	8	3	30	14	5	74
Sedibeng	1	1	1	0	2	5	1	11
Tshwane	0	5	70	0	7	25	2	109
West Rand	0	2	2	4	34	0	8	50
Total computers	**5**	**47**	**116**	**7**	**83**	**67**	**16**	**341**
Percentage	**1%**	**14%**	**34%**	**20%**	**2%**	**5%**	**24%**	**100%**

Status of Antivirus software

About 80% of computers were reported to run updated antivirus software. However, the Sedibeng district had only 46% of its computers safeguarded against viruses. This poses a risk to the data at facility level and should be addressed as a matter of urgency.

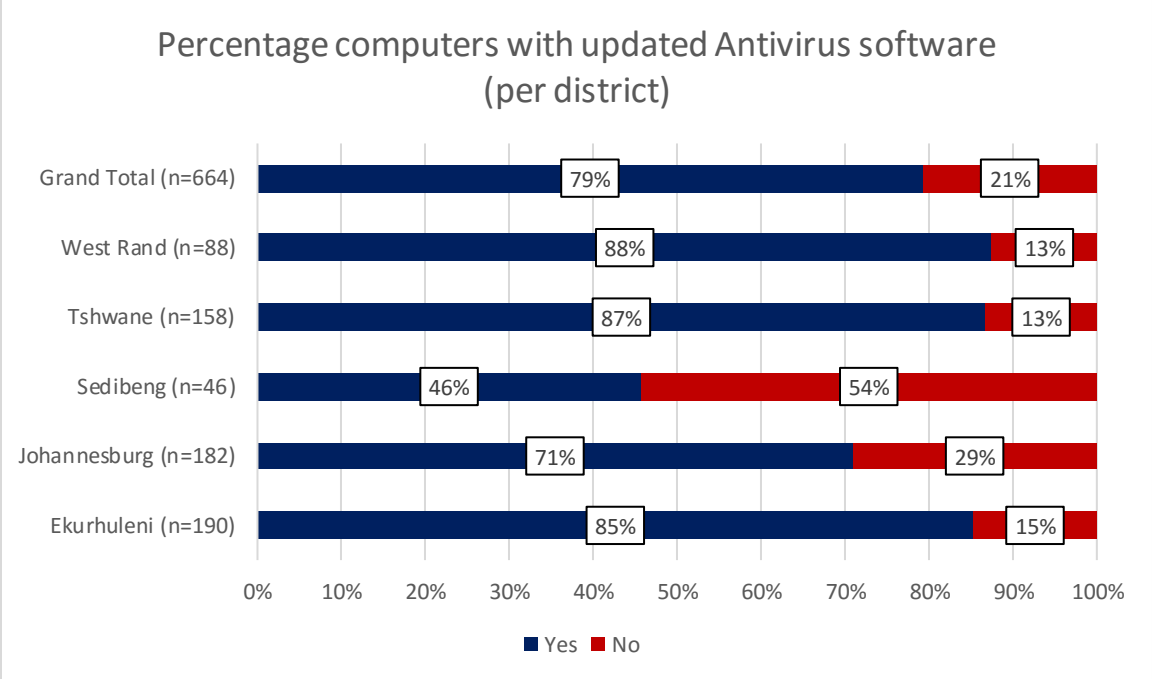


Figure 24: Status of Antivirus software

Status of backup

With 304 facilities (46%) having backups older than one week, facilities run the risk of losing more than one week's data in the event of an equipment failure. More concerning is the fact that

nearly 200 facilities (30%) reported to have only done a backup more than a month ago. Sedibeng district was the least exposed with 76% facilities reporting that a backup was done on the day or the previous day of assessment.

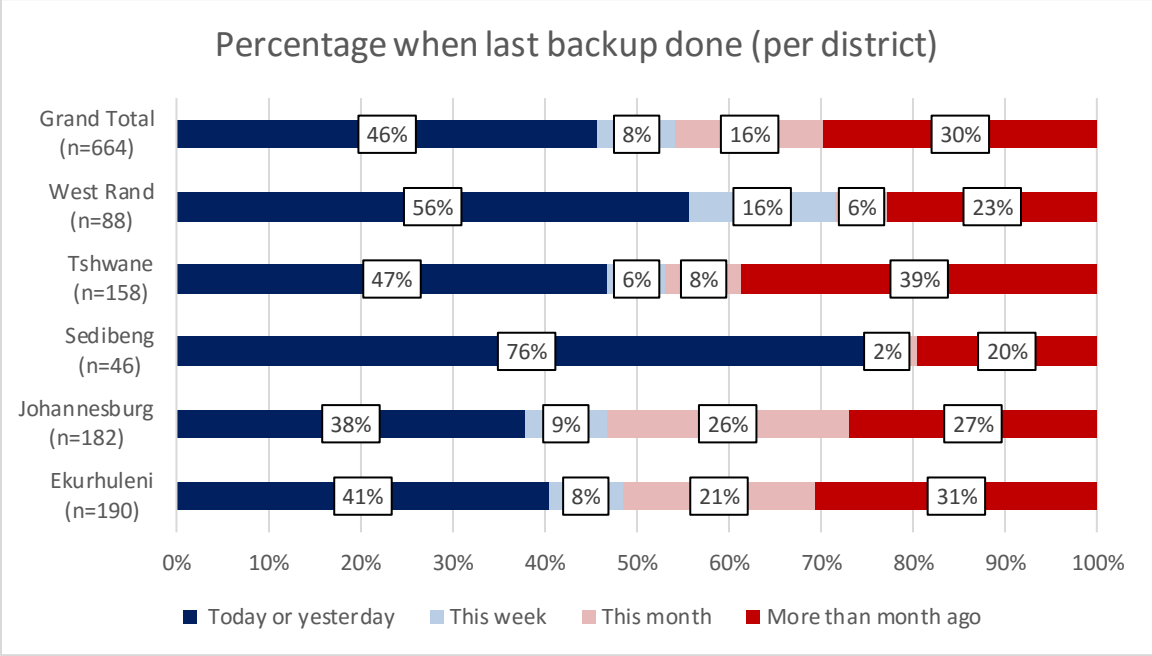


Figure 25: Computer backup status

Are computers connected to the Internet?

Facility representatives responded to the following questions regarding connectivity of each computer:

- Is the computer fully connected? Thus email plus other web access?
- Does the computer have basic connectivity (email capacity)? or
- Is the computer not connected?

The following responses to question about Internet connectivity represent the responses of the facility manager or senior data person at each facility. Fieldworkers could not confirm these responses because they did not test the actual abilities of each computer to connect to the Internet.

The findings reveal that 39% of computers are reportedly not connected to the Internet. Sedibeng and West Rand districts reported the lowest connectively with only 11% and 15% of computers respectively having any form of connectivity. The Ekurhuleni (67%) and Tshwane (62%) districts reported higher percentage of facilities with full connectivity.

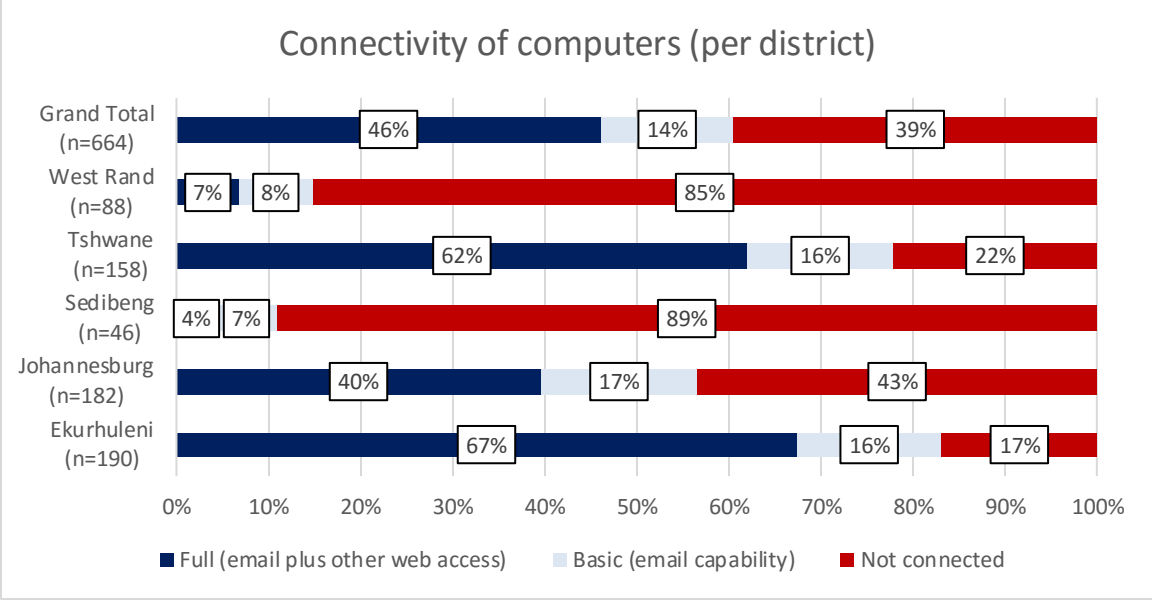


Figure 26: Internet connectivity of computers

- **Summary of the Technical Resourcing**

The findings from this section reveal that the PHC facilities are reasonably equipped with the technical resources to capture health-related data. However, the resources are not necessarily sufficient to ensure excellent quality of the data that are captured.

Computers

Data indicate that, throughout the province, all but three facilities had access to computers to use for capturing data, and a total of 669 computers were in use during the assessment period. Ekurhuleni district had the highest number of computers per facility with an average of 3.4, while Sedibeng had the lowest number on average at 1.4 per facility. On the downside, thirteen facilities had only one, partially functioning computer. On the upside, almost 90% of computers were reportedly fully functional.

Most computers were used for capturing more than just one dataset or for other tasks. Almost 80% of computers at facilities were shared and had more than one primary user. Although both these aspects might indicate efficient use of technical resources, sharing resources could also have potential negative impacts on data quality, and on access to a computer for data capturing.

Software

Data indicate that many different versions of the DHIS software are installed and used on computers. This will likely impede data quality and consistency of datasets. While antivirus software is generally updated—except in the Sedibeng district.

3.3 Human Resource Capacity at Health Facilities to Carry out Data Management

- **Finding 10: Most clinics have the necessary staff to carry out data management, however this varies by district**

Clinic staff

Clinical staff members at health care facilities include permanent or contracted doctors, professional nurses, other practitioners (e.g. dentists), or support personnel. Table 15 and 16 show that according to the facility managers in the surveyed clinics, some facilities were well supplied with appropriate clinical employees, while the supply of doctors at PHC facilities was limited. The median scores reveal that on average many clinics in Johannesburg, Ekurhuleni and Sedibeng district do not have one permanent doctor employed. Instead they relied on nurses and contracted doctors. Facility managers reported that the number of permanent and contracted doctors employed ranged anywhere from between zero to nine (median=1); professional nurses employed ranged from one to seventy-one (median=7); "other professional practitioners" employed ranged from zero to thirty-six (median=0); and support staff ranged from zero to one hundred and twenty-three (median=6).

According to the facility managers, there were four clinics that employed only professional nurses for clinical duties—i.e. they had no doctors, support staff or "other practitioners" engaged. These four clinics with low clinic staff levels are:

- Zone 3 Clinic in the Emfuleni Local Municipality (7 nurses only)
- Glenanda Clinic in the Johannesburg F Health sub-district (4 nurses only)
- Venterspos Clinic in the Westonaria Local Municipality (2 nurses only)
- Mountain View Clinic in the Johannesburg G Health sub-district (4 nurses only)

Administrative staff

In terms of staffing levels, the fieldworkers also asked facility managers about the different people who *may be responsible* for data capturing and processing, including permanent or contracted data capturers, or other administrative staff members. Out of the 286 clinics surveyed, at least 175 clinics were reported to have employed only one type of administrative personnel. There were 20 clinics that had no data capturers or dedicated administrative staff and most of these were located in the Tshwane Metro Municipality. The names of these clinics with reportedly low administrative staff levels are listed below.

Johannesburg F Health sub-District	Bellavista Clinic
Johannesburg G Health sub-District	Ennerdale Ext 9 Clinic
	Lawley Clinic
	Thulamntwana Clinic

Tshwane 1 Health sub-District	Boikhutsong Clinic
	Ga-Rankuwa View Clinic
	Mercy Winterveldt NGO Clinic
	Phedisong 4 CHC
	Rosslyn Clinic
Tshwane 2 Health sub-District	Kekana Gardens Clinic
	Ramotse Clinic
	Refentse Clinic
Tshwane 3 Health sub-District	Bophelong (Region C) Clinic
	Saulsville Clinic
Tshwane 4 Health sub-District	Eldoraigne Clinic
	Laudium Clinic
	Lyttelton Clinic
	Olievenhoutbosch Ext 13 Clinic
Tshwane 6 Health sub-District	Silverton Clinic
	Eersterust CHC

Table 16: Total number of clinic and administrative employees across 286 sampled Gauteng PHC facilities

	All Districts	Johannesburg	Tshwane	Ekurhuleni	Sedibeng	West Rand
Clinical Staff	**Total**	**Total**	**Total**	**Total**	**Total**	**Total**
Doctors (Permanent)	281	84	66	43	19	69
Doctors (Contracted)	240	52	63	55	23	47
Professional Nurses	3082	904	853	685	301	339
Other (e.g. Dentists)	288	100	52	25	94	17
Support Staff	2985	791	626	738	333	497
Administrative Staff						
Data Capturers (Permanent)	354	85	57	78	56	78
Data Capturers (Contracted)	293	131	33	90	23	16
Other	252	76	44	58	19	55

Table 17: Median number of administrative employees in Gauteng PHC facilities

	All districts		Johannesburg		Tshwane		Ekurhuleni Metro		Sedibeng		West Rand	
	Median	**Range**	**Median**	**Range**	**Median**	**Range**	**Median**	**Range**	**Median**	**Range**	**Median**	**Range**
Clinical Staff												
Doctors (Permanent)	1	0 – 9	0	0 – 9	1	0 – 5	0	0 – 4	0	0 – 5	1	0 – 6
Doctors (Contracted)	1	0 – 9	0	0 – 3	1	0 – 4	1	0 – 9	1	0 – 6	0	0 – 6
Professional Nurses	7	1 – 71	5	2 – 71	11	1 – 59	7	3 – 59	7	2 – 39	7	2 – 19
Other (e.g. Dentists)	0	0 – 36	0	0 – 36	0	0 – 8	0	0 – 6	1	0 – 25	0	0 – 2
Support Staff	6	0 – 123	5	0 – 73	6	0 – 73	6	0 – 123	8	0 – 59	6	0 – 58
Administrative Staff												
Data Capturers (Permanent)	1	0 – 10	1	0 – 3	1	0 – 10	1	0 – 4	2	0 – 4	1	0 – 5
Data Capturers (Contracted)	0	0 – 21	1	0 – 21	0	0 – 5	1	0 – 6	0	0 – 6	0	0 – 9
Other	0	0 – 12	0	0 – 11	0	0 – 8	0	0 – 12	0	0 – 8	0	0 – 7

Capacity of employees: Level of data-related training

In addition to the facility managers, the fieldworkers spoke to a total of 636 other persons who *self-identify as being responsible* for data management at their facility. Of these people, 605 reported their official role as: 104 administrative clerks, 341 data capturers, 130 facility managers and 30 "others" (e.g. volunteers or intern clerks).

These personnel attend to data related tasks, meaning that they should all have received some training on the data collection software or data use policy. Findings show that in terms of data-related training, between 72% and 73% of respondents had reportedly received DHIS and TIER training, 36% had ETR training, while 55% or less had received any formal instruction on DHMIS policy or NIDS and PIDS (See Figure below).

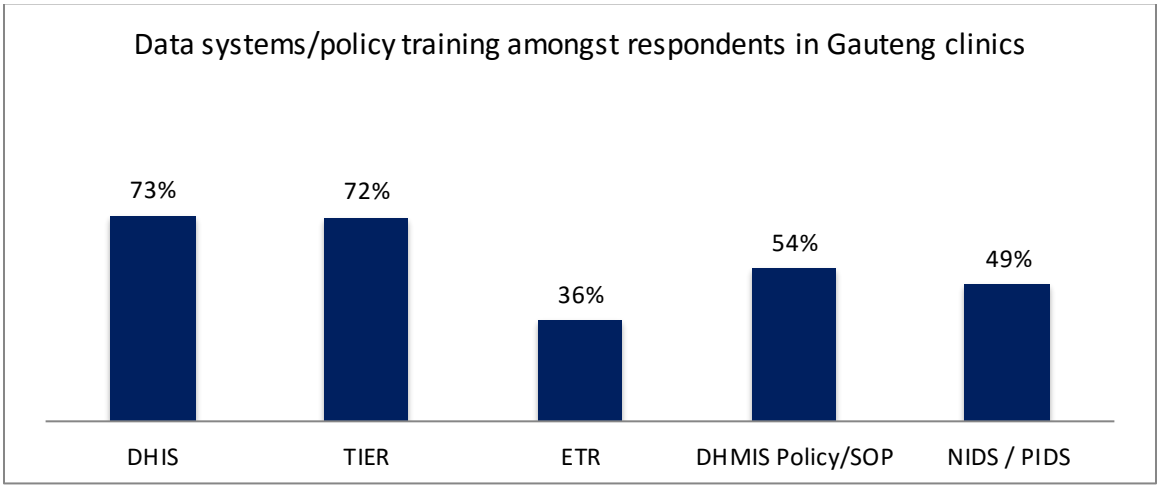


Figure 27: Level of data-related training amongst 636 employees at 286 Gauteng PHC facilities

There were 34 (12%) facilities where none of the respondents stated that they had received DHIS or TIER training, 140 (48%) were none said they had received ETR training, 74 (26%) were none said they had received DHMIS Policy/SOP training, and 84 (29%) facilities where none of the respondents had reportedly received any training in NIDS or PIDS (Table 18).

Table 18: Level of data-related training amongst 636 employees at 286 Gauteng PHC facilities

Level of training	DHIS	TIER	ETR	DHMIS Policy/SOB	NIDS / PIDS
High level *(>=60% of respondents trained)*	209 (72%) facilities	198 (69%) facilities	92 (32%) facilities	146 (50%) facilities	136 (47%) facilities
Low-moderate level *(0< mod<60%)*	46 (16%) facilities	57 (20%) facilities	57 (20%) facilities	69 (24%) facilities	69 (24%) facilities
No training *(0 % of respondents trained)*	**34 (12%) facilities**	**34 (11%) facilities**	**140 (48%) facilities**	**74 (26%) facilities**	**84 (29%) facilities**

Most facilities showed a high-level of database-related training, with 60% or more of the personnel interviewed saying that they were trained in at least one of the five listed database trainings. Of concern are the twenty-two facilities where none of the respondents said they had received any data-related training. These facilities, listed below, are not concentrated in any particular district or sub-district.

Johannesburg G Health sub-District	Lenasia Ext 10 Clinic
Johannesburg A Health sub-District	Midrand West Clinic
Johannesburg B Health sub-District	Bosmont Clinic
Johannesburg G Health sub-District	Wildebeesfontein Clinic
	Protea South Clinic
Tshwane 2 Health sub-District	Refentse Clinic
Tshwane 1 Health sub-District	Sedilega Clinic
	Soshanguve Block JJ Clinic
Tshwane 3 Health sub-District	Laudium CHC
	Atteridgeville Clinic
Ekurhuleni North 2 Health sub-District	Northmead Clinic
	Dresser Clinic
Ekurhuleni South 1 Health sub-District	Jabulane Dumane CHC
Ekurhuleni South 2 Health sub-District	Sunrise View Clinic
Lesedi Local Municipality	Rensburg Clinic
	Ratanda Ext 23 Clinic
Emfuleni Local Municipality	Tshepiso Clinic
	Rietspruit Clinic
Randfontein Local Municipality	Badirile Clinic
Mogale City Local Municipality	Rietvallei 2 and 3 Clinic
	Fanyana Nhlapo Clinic
Merafong City Local Municipality	Khutsong West Clinic

Capacity of employees: education level

The interviewed employees were fairly well-educated. The Figure 28 below shows the average education level of those who provided education information. Most employees (98%) interviewed had at least a grade twelve level education and almost one fifth had completed a university level degree. Of these, 13% had a post-graduate degree. This indicates that the people responsible for data capturing and processing at the time of the assessment were a well-educated set of employees who were most likely appropriately capable of handling the data.

Nevertheless, the distribution of highly skilled employees was not even throughout the district clinics. Figure 30 reveals that fieldworkers in the Johannesburg Metro area where easily able to interview at least one person who had a post graduate degree in 43 different clinics. However, in Tshwane there were only 15 clinics where one or more persons could be interviewed who were highly educated. Such people were also not easily found for interviewing in the Sedibeng and the West Rand clinics. This could be a simple consequence of fieldworkers struggling to find skilled employees to interview—perhaps because these employees were elsewhere occupied—or it could indicate that these clinics struggle to attract highly skilled employees.

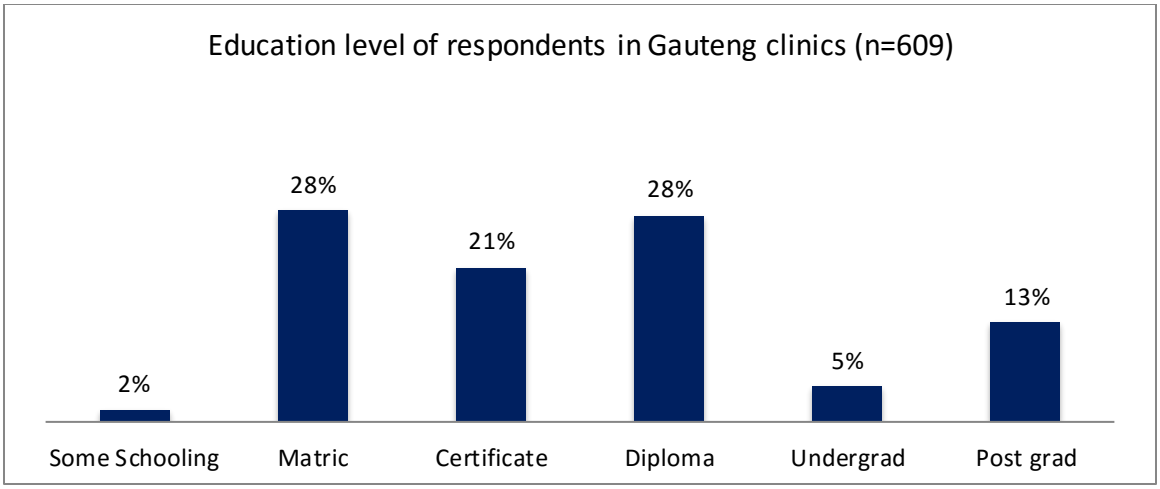


Figure 28: Education level of respondents in Gauteng clinics

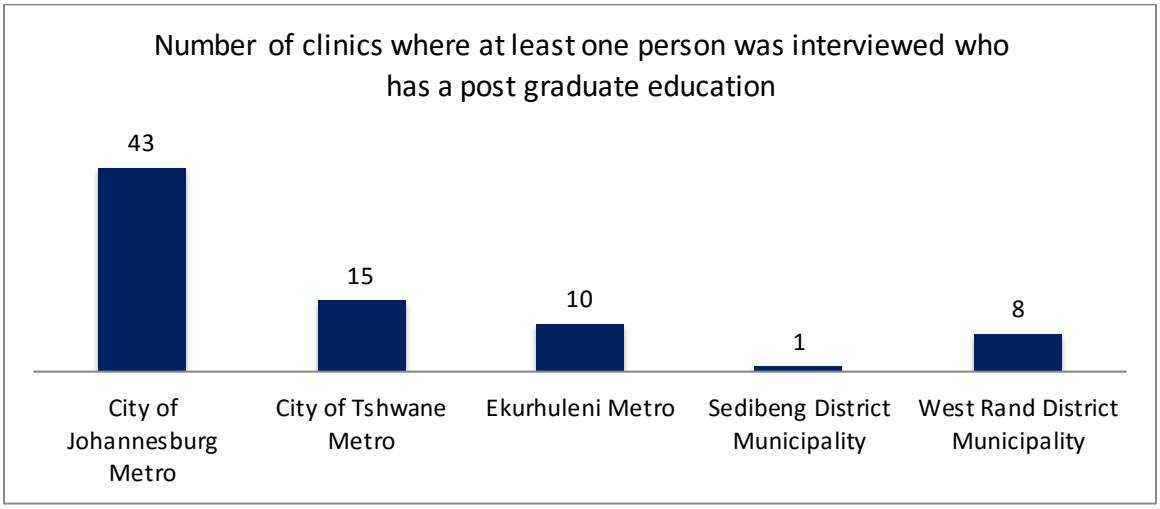


Figure 29: Number of clinics where at least one interviewed person has a post graduate degree

Employee tenure

Of the clinic representatives interviewed (621 people), on average they had been employed for slightly more than six years. Twenty-five (4%) of these people were long time employees of over twenty-five years, 9% had been employed between ten and twenty-five years, while at least 51%

had only been employed four years or less. Only 17% of the respondents said they had been employed at the clinic for only about one year.

Capacity of employees: Employment level

The employment type amongst these respondents also varied (Figure 30). The majority of employees that the fieldworkers spoke to worked fulltime in the clinic (79%), while the rest were either engaged on a part time (3%) or volunteer (1%) bases. Roughly 1% had been seconded or were PWPS employees (8%).

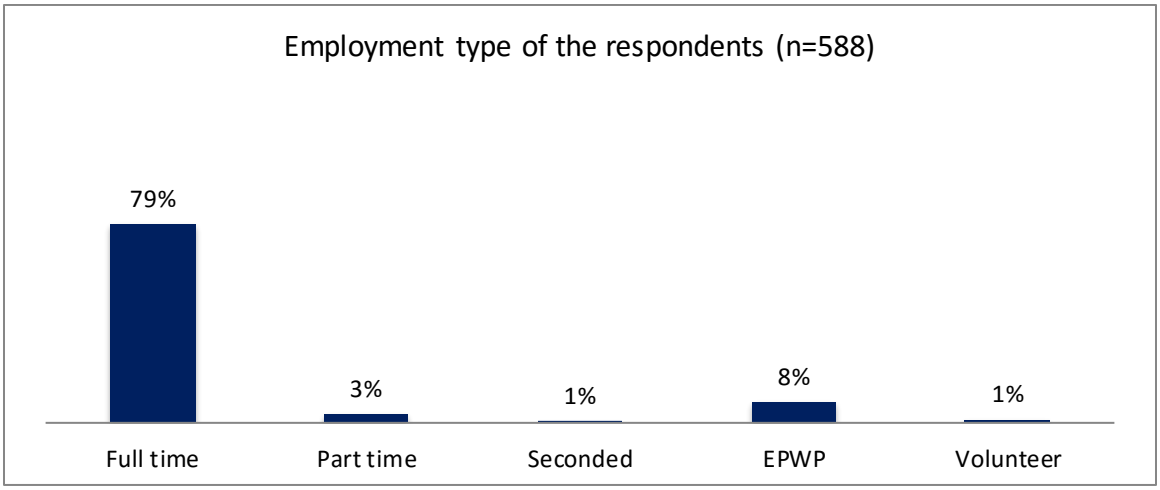


Figure 30: Types of employment

Consulting rooms

Fieldworkers asked the facility managers to talk about the physical facilities at the clinic. The managers reported that the amount of consulting rooms available for clinical consults varied by clinic type, and can range anywhere from two to as many as fifty-eight rooms (see

Table 16). Facilities in Gauteng had on average eight rooms for consultations. Tshwane tops the list with facilities that accommodated a median of twelve consulting rooms, whereas facilities in Johannesburg, Sedibeng and West Rand Districts had around seven rooms for consultations.

Table 19: Total and median number of consulting rooms across all sampled facilities

	Total rooms	Median in clinics	Range in clinics
Johannesburg	888	7	2 – 58
Tshwane	782	12	3 – 34
Ekurhuleni	597	9	2 – 39
Sedibeng	303	7	2 – 36
West Rand	344	7	2 – 23
Total	**2914**	**8**	**2 – 58**

- **Finding 11: Facility managers and senior data persons are generally able to manage the volume of data in the time available, but where shortages of administrative staff was found, this was not the case**

Burden of time: facility managers

Table 20 shows the per cent of time that facility managers and senior data persons reportedly devoted to particular data capturing, processing and use-related tasks during the assessment period. On average, facility managers said they spend roughly 16% of their work time on clinical tasks, around the same amount on supervision and on administration respectively, and about 12% of time on data management. This indicates an even spread of work-time-allocation for facility managers. There was little variation in this regard across the districts.

While most managers balance their time well, an exceptional few reportedly spent too much time on one task alone. Facility managers at twenty different facilities said they spent more than 55% of their time on clinical tasks alone, which points to shortages with clinical staff. This was especially evident at Jabulane, Dumane CHC, and Zone 13 Clinics where the managers were said to have spent more than 75% of their time on clinic tasks. However, these appear to be isolated cases, and there is no discernible pattern of difference between the districts or sub-districts. Moreover, this is self-reported information and the managers could have misjudged their time.

Burden of time: Senior data persons

The findings suggest that the senior persons who are responsible for data processes at the PHC facilities are also good at balancing their time between data tasks. Like facility managers they reportedly distribute their time rather evenly, spending on average no more than 10% of their time on a given task, with the exception of compiling monthly reports, which they say consumes about 12% of their time.

There were a few isolated cases where senior data persons could balance their time more evenly. At the Muldersdrift and Randgate Clinics, the senior data persons report that they spent more than 55% of their time preparing the monthly reports. At Moagle and Khutsong South clinics

they reportedly spend over 60% of their time collating and summarising PHC data, perhaps indicating limited support or a lack of skill in this task. In total there were ten individuals who found that one data-related task occupied at least 40% of their workload—five of whom were located in West Rand District clinics, and three in City of Johannesburg clinics.

Table 20: Percentage of work time devoted to the following tasks

	All Districts	Johannesburg	Tshwane	Ekurhuleni	Sedibeng	West Rand
Facility manager						
	% of time	**% of time**	**% of time**	**% of time**	**% of time**	**% of time**
Clinical tasks (consulting patients)	16	27	15	15	10	14
Supervision/oversight/mentoring	15	21	18	15	9	11
Administration	14	22	15	15	8	12
Data management (capturing, verifying, reviewing or using)	12	20	12	11	7	12
Senior data person						
	% of time	**% of time**	**% of time**	**% of time**	**% of time**	**% of time**
Collating and summarising PHC data	9	14	9	8	4	8
Capturing TIER data	9	12	9	9	6	7
Verifying data, doing spot checks	7	11	10	7	5	5
Supporting staff to capture and manage data	7	11	8	6	3	5
Preparing the monthly summary report	12	20	10	12	6	12
Preparing data/reports on request	7	11	8	7	4	6
Responding to inputs from sub-district, district, province and national	7	11	7	7	4	6

- **Finding 12: Data-related training seems to help facility managers balance-out their distribution of tasks.**

The systems and staffing data reveal that higher numbers of data-trained personnel at a facility were associated with a decrease in the amount of time that a facility manager spent on clinical tasks (r= -0.13), and an increase in time that s/he spent on supervision, oversight or mentoring (r= 0.12). These are small but statistically significant correlations. In contrast, the more data-trained personnel at a facility, the less time a facility manager spent on administrative and data management tasks (although not statistically significant). In principle this could indicate that as personnel gain the knowledge and capacity to deal with data themselves they can handle the workload better, which allows managers more time to supervise.

Amongst senior data personnel, levels of training at the facility had little effect on how their time was spent between tasks, with two exceptions: 1) an increase in the amount of personnel who were trained in TIER and ETR was associated with an increase in the amount of time that a senior data person spent on capturing TIER data (r= 0.20 and r= 0.12). Not an unsurprising finding. However, more trained personnel can be associated with a reduction in the amount of time that the senior data person spent on verifying data and doing spot checks. This is troubling and may indicate that while training in data software and data policy leads to awareness about data capturing and data systems, it does not promote a practical emphasis on data quality. Put differently, personnel consume their time with data capturing and reporting, possibly at the expense of conducting data quality control. The focus on data capturing after training on TIER may, however, also relate to a temporary focus on back-capturing of TIER data.

- **Finding 13: Data use at the facility level appears limited, and likely negatively impacts data quality**

In each clinic, the facility managers and senior data persons were asked if they had requested or pulled a report from the system other than the monthly report. The pie graph (Figure 27) below shows that PHC information was reported as most popular (34%) source of information, followed by TIER data (31%). Facility managers and senior data persons seemed to have had less interest in, or need for, raw TB data, only requesting or pulling this information from the system 10% of the time. However, they relied on information from "other" source at least 25% of the time.

There appears to be some differences between districts in what interests facility managers and senior data persons. For example, TIER data was the most used in Johannesburg and Sedibeng, while PHC was more popular in Tshwane and Ekurhuleni, and "other" sources were more used in the West Rand. On average facility managers and senior data persons in Ekurhuleni and Sedibeng used TB information more than those from any other district (18% of the time).

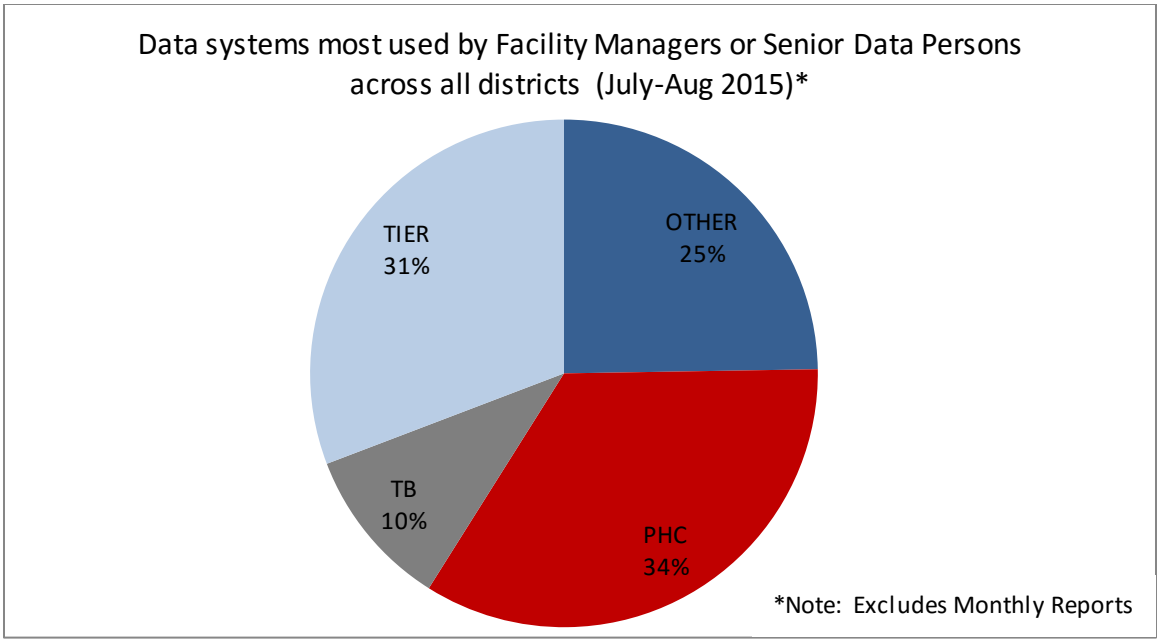
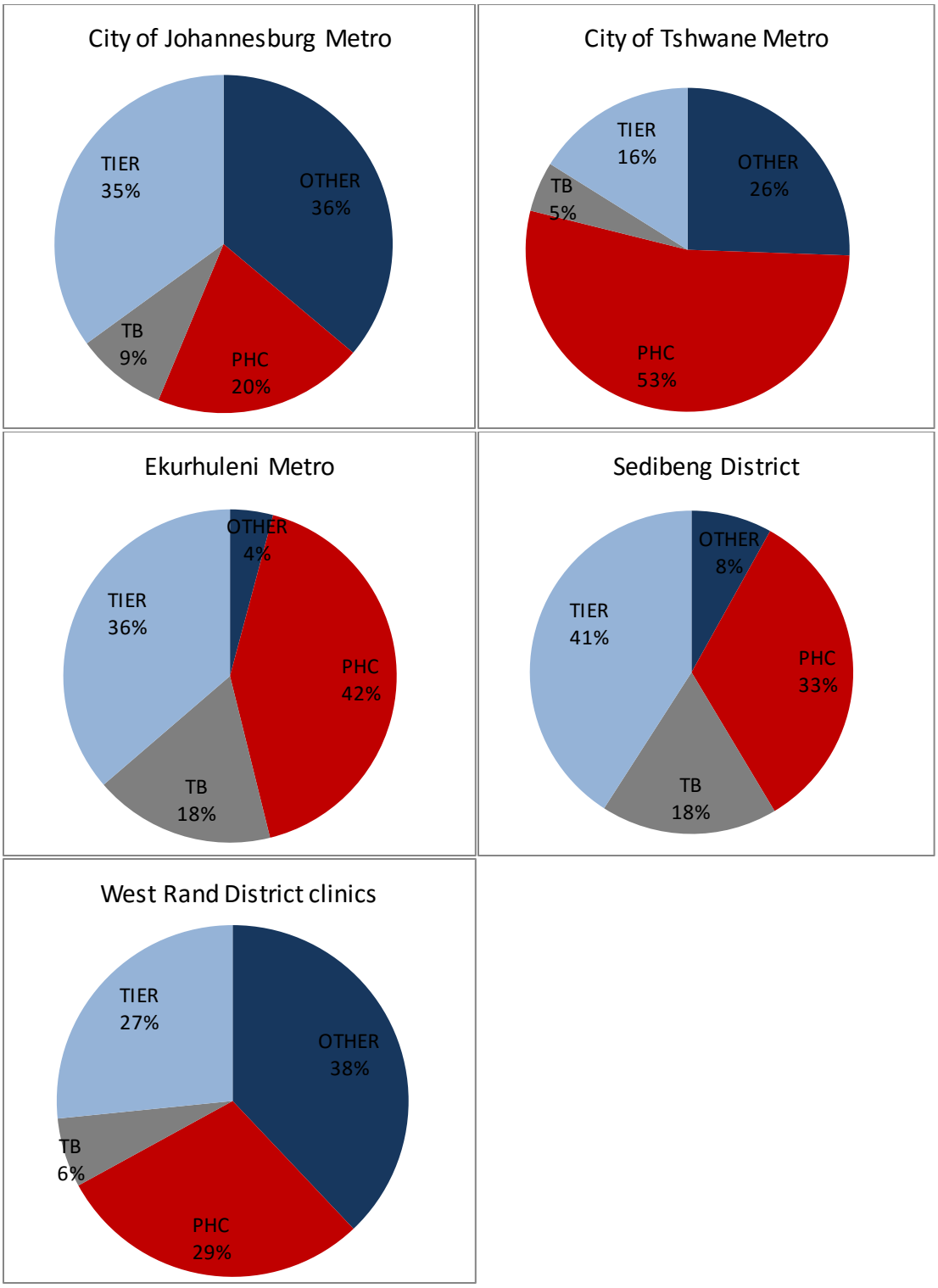


Figure 31: Data systems most used

Data systems most used by facility managers or senior data persons in each district



Despite limited data use by facility managers and senior data persons, there was evidence of routine discussion on data quality, timeliness, completeness and accuracy (see Figure 33 below). The degree to which these discussions focused on problem identification and recommendations for remedial actions were not assessed in this study.

Fieldworkers found evidence of data discussions in the meeting agendas, minutes, and other facility notes. While most facility managers and data persons kept record of the different discussions about data quality in their meeting notes, in Ekurhuleni and Sedibeng clinics this evidence was also located in other, less-structured, sources, perhaps indicating some lack of organised filing and record keeping. Evidence of less routine discussions about data quality could be found in other documents and sources, such as:

- Monthly and weekly hash reports
- Daily summary sheets
- Quarterly feedback graphs or reports
- Quality improvement plans
- Office notebooks or files
- Dedicated statistics files or books
- Manager's personal diary, calendar, and/or loose sheets of paper

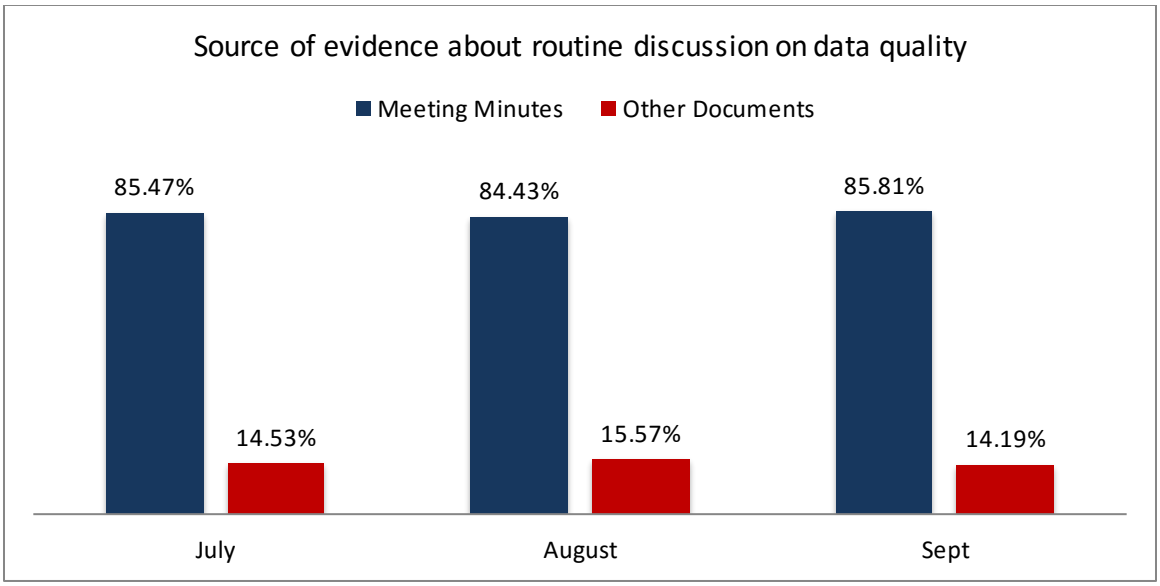


Figure 32: Source of evidence about routine discussion on data quality per month

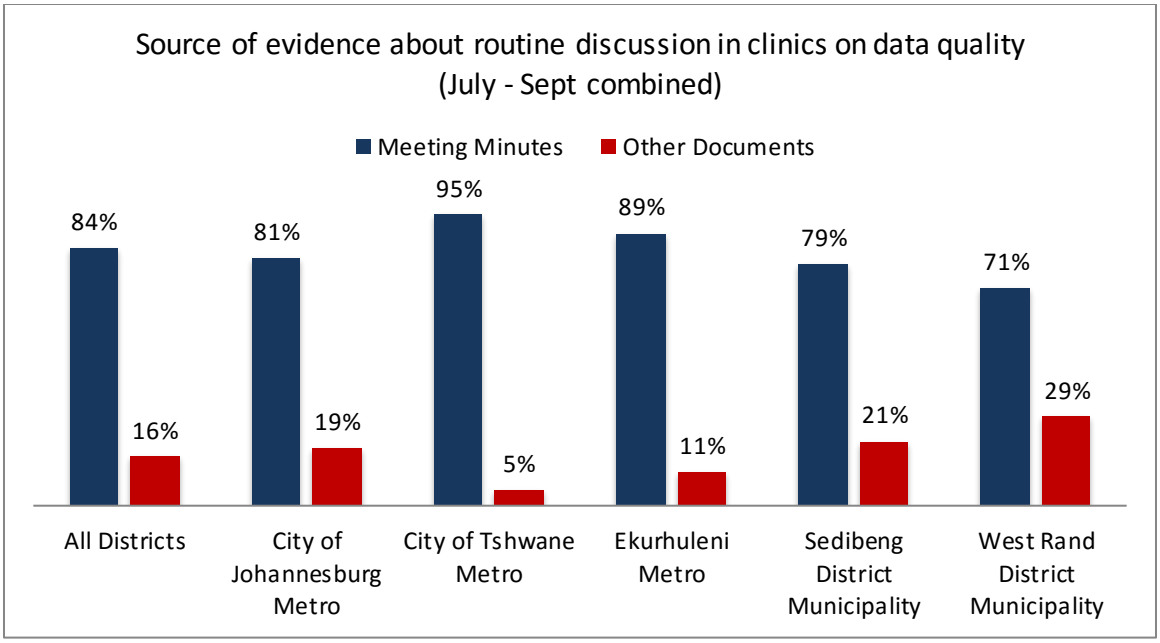


Figure 33: Source of evidence about routine discussion on data quality by district

Table 21: Time burden: Percentage of work time devoted to the following tasks

Percentage of work time devoted to the following tasks	All Districts			City of Johannesburg Metro	City of Tshwane Metro	Ekurhuleni Metro	Sedibeng District Municipality	West Rand District Municipality
Time burden for facility manager								
	% of time	Std. Dev.	Max %	% of time	% of time	% of time	% of time	% of time
Clinical tasks (consulting patients)	28%	0.16	75%	30%	25%	26%	29%	29%
Supervision/oversight/mentoring	26%	0.14	75%	23%	30%	27%	27%	22%
Administration	25%	0.13	100%	24%	25%	26%	23%	25%
Data management (capturing, verifying, reviewing or using)	22%	0.12	75%	22%	21%	20%	21%	24%
Time burden for senior data person								
	% of time	Std. Dev.	Max %	% of time	% of time	% of time	% of time	% of time
Collating and summarising PHC data	15%	0.09	77%	15%	15%	14%	12%	17%
Capturing TIER data	15%	0.09	57%	14%	14%	17%	20%	14%
Verifying data, doing spot checks	13%	0.07	50%	12%	16%	12%	14%	11%
Supporting staff to capture and manage data	12%	0.07	44%	13%	13%	11%	10%	10%
Preparing the monthly summary report	21%	0.14	77%	22%	17%	22%	19%	24%
Preparing data/reports on request	12%	0.07	62%	12%	13%	12%	13%	11%
Responding to inputs from sub-district, district, province and national	12%	0.09	80%	12%	12%	12%	12%	13%

- **Summary of the Human Resource Capacity at PHC Facilities**

The results of the systems and staffing survey reveal that most clinics had the necessary staff to carry out data management. They employed data capturers, administrative clerks, facility managers, and volunteers or interns for routine data capturing and processing, and in some instance nurses were also doing this work. These individuals were usually well-educated with at least a grade twelve matriculation, and very often a post-high-school certificate of some kind.

The burden of data work is heavy, and often it falls to primarily one person. In at least 61% of clinics, only one administrative person was employed to deal with data-related tasks and processing. At the other facilities, this task fell to clinical or support personnel who were not necessarily trained in data processing, or who had other competing duties. The Tshwane Metro Municipality had twenty clinics with no data capturers or dedicated administrative staff employed. Thus, in terms of human resources, PHC clinics had appointed employees dedicated to data-related tasks, but the number of employees dedicated to this task was less than adequate.

Other results show that the personnel engaged in data-related tasks do not necessarily have the capacity to ensure a high standard of data *quality*. Although they were well educated, appointed at the appropriate level, and usually full-time employees, there were instances where volunteers were being asked to do many of the data-related tasks. In addition, about 17% of these employees were recent hires who had worked at the clinic for less than a year. The findings also reveal that not all the employees interviewed had been trained on the various data-systems, possibly because many were new to the clinic. Only about 73% had received DHIS or TIER training, 36% had received ETR training, and less than 55% had received no formal instruction on DHMIS policy, or NIDS and PIDS. If high data quality standards are to be met, new employees must receive the appropriate data training.

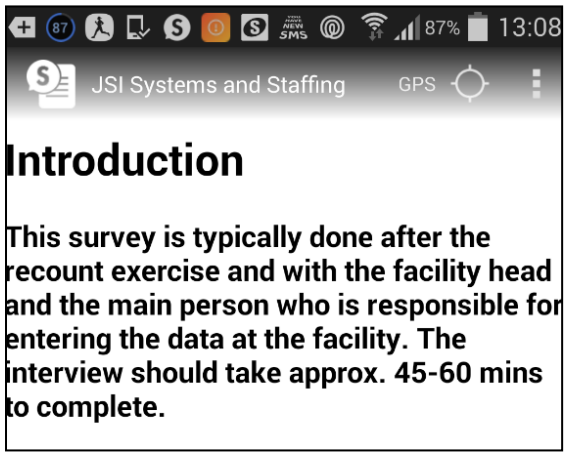
Further findings demonstrate that facility managers and senior data persons were generally able to manage the volume of data in the time available, but that they could pay more attention to supervising data quality, not simply data collection and capturing. Having more individuals trained in data-related tasks helped the facility managers at the clinics balance out their tasks, but it also reduced the amount of time that personnel spent on data quality checks and verifications. In other words, awareness about data collection—that comes from the training—is potentially encouraging a culture of "quantity" and not "quality" in the data collection process. Personnel consume their time with data capturing and reporting, possibly at the expense of conducting data quality control.

Finally, data use at the facility level seemed to be limited, and this in turn is likely to have a negative impact on data quality. Meeting minutes and other office documents show evidence of routine discussion about data processing and data quality at many PHC facilities, indicating that employees are beginning to be concerned with the integrity of information. Nevertheless,

managers and senior data persons at the facilities do not rely much on the data being collected at their facilities. Reasons for this warrant further exploration, as it is likely to have a negative impact on data quality.

3.4 Implementation of DHMIS Policy Status

- **Finding 14: The DHMIS policy is well implemented at about two thirds of the assessed PHC facilities**



In an effort to assess whether the DHMIS policy was implemented well across the health facilities, the facility managers were asked about the "Tools and resources used to guide data collection," "Data verification," "The availability of tools, registers and forms," and the "Maintenance of data collection tools in line with Standard Operating Procedures (SOP)." Their responses to these questions were recorded in two ways. First, the fieldworkers noted the spontaneous response, and next they prompted the manager with further hints and then also recorded these answers. This strategy helps ensure a more unbiased reflection of how facility managers truly perceive their implementation of DHMIS policy. The results for this section are therefore presented as either a "spontaneous" or a "prompted" mention.

Tools and resources used to guide data collection

Spontaneous indications by facility managers and senior data persons of using data management tools and resources

The facility manager and the senior data person were asked to verify which tools and resources they used to guide their data collection and management at the facility. The use of these tools and resources were also observed by the fieldworker to verify this claim and is reported in the following section.

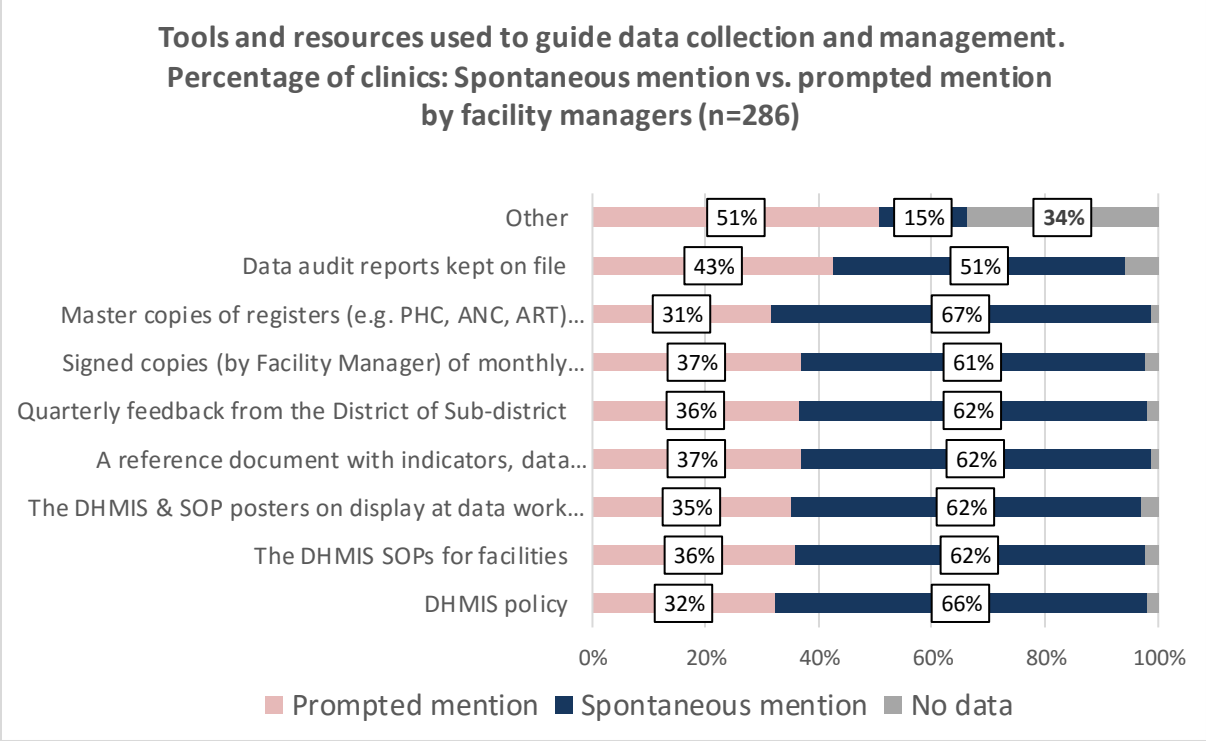


Figure 34: Percentage clinics: Spontaneous vs. prompted mention by facility managers of tools and resources used

Table 22: Number of clinics: Spontaneous vs. prompted mention by facility managers of tools and resources used

	Response from facility manager		
Tools and resources	**Prompted mention**	**Spontaneous mention**	**Blank**
DHMIS policy	92	188	6
The DHMIS standard operating procedures for facilities	102	177	7
The DHMIS and SOP posters on display at data work station	100	177	9
A reference document with indicators, data elements, definitions and formulas	105	177	4
Quarterly feedback from the district of sub-district	104	176	6
Signed copies (by facility manager) of monthly reports	105	174	7
Master copies of registers (e.g. PHC, ANC, ART) kept on file	90	192	4
Data audit reports kept on file	122	147	17
Other:	145	44	97

At the majority of the facilities the facility manager spontaneously mentioned that seven of the eight tools and resources were used to guide data collection. The data audit reports kept on file were mentioned spontaneously at just over 50% of the PHC facilities. Facility managers at 192 (67%) of the facilities spontaneously mentioned that master copies of registers were kept on file, and at 66% of the facilities the DHMIS policy were reportedly used to guide data collection and management of the facility. At 62% of the facilities the facility manager spontaneously mentioned

that the following were used to guide data collection: DHMIS SOPs for facilities; the DHMIS and SOP posters on display at data work stations; a reference document with indicators, data elements, definitions and formulas, and quarterly feedback from the district or sub-district. Only at 44 (15%) facilities did the facility manager spontaneously mention other tools being used. However, at about half of the facilities (147 or 51%), managers acknowledged that other tools and resources were used for guiding data collecting and processing, but this occurred only after being prompted by the fieldworkers. Specific tools and resources mentioned include:

- TIER.net (7 facility managers)
- Note book, daily journals (5 facility managers)
- Computers (3 facility managers)
- Loose pages (2 facility managers)

Overall at majority of the facilities, facility managers spontaneously mentioned that the respective tools and resources were used to guide data collection at the facilities. The responses provided by senior data managers were similar to that of facility managers in response to this question. This is illustrated in the figure below.

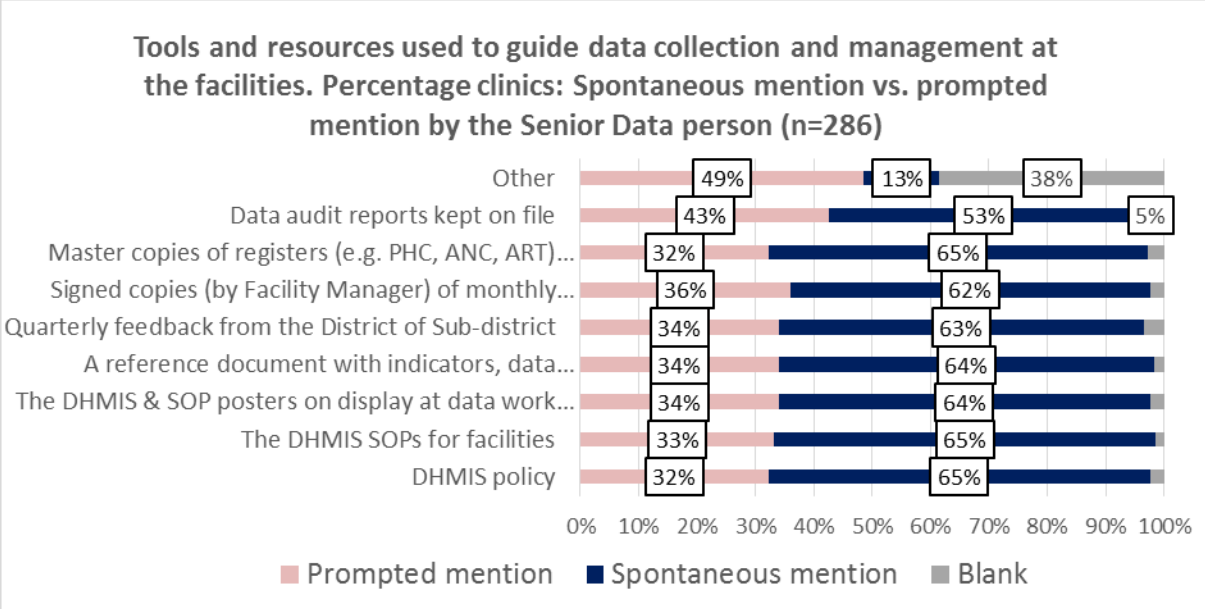


Figure 35: Tools and resources used to guide data collection and management at the facilities: Percentage of facilities: Spontaneous mention vs. prompted mention by senior data person

Observed use of tools and resources at facilities

The fieldworkers also verified that these tools and resources were used in practice—and not just reportedly used—by observing their use to guide data collection and management at each facility.

The figure below illustrates the percentage of facilities where the use of the tools and resources was observed.

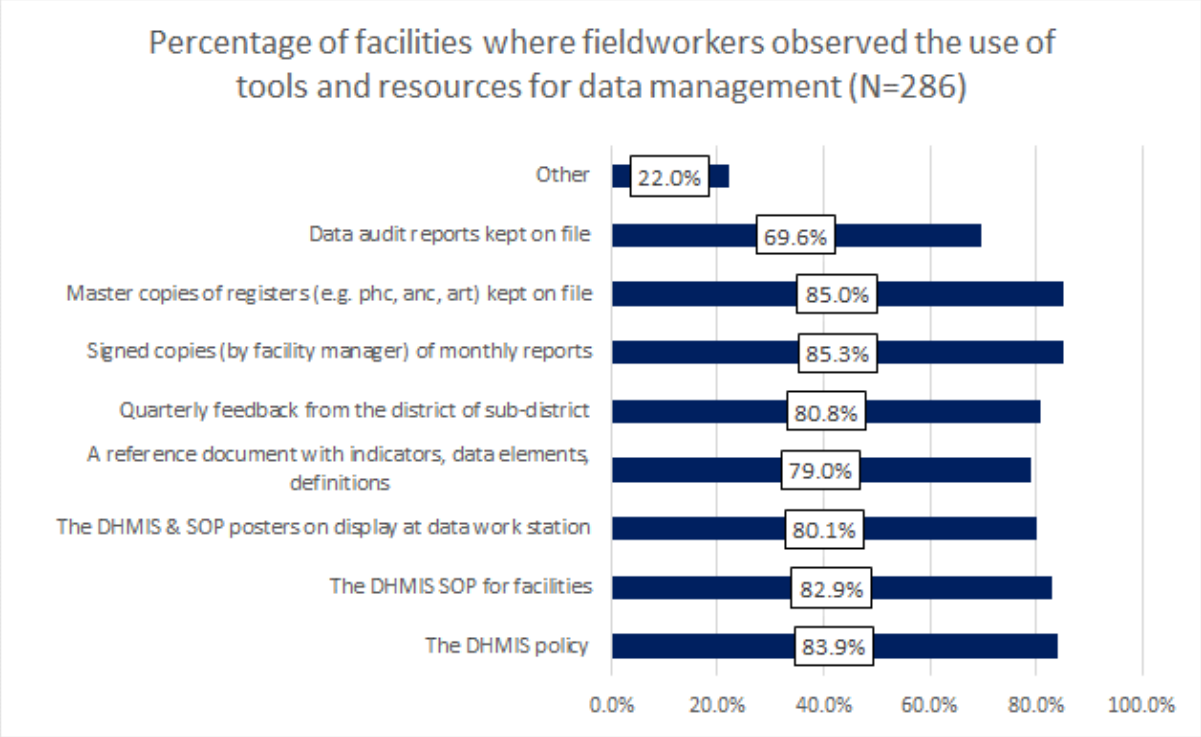
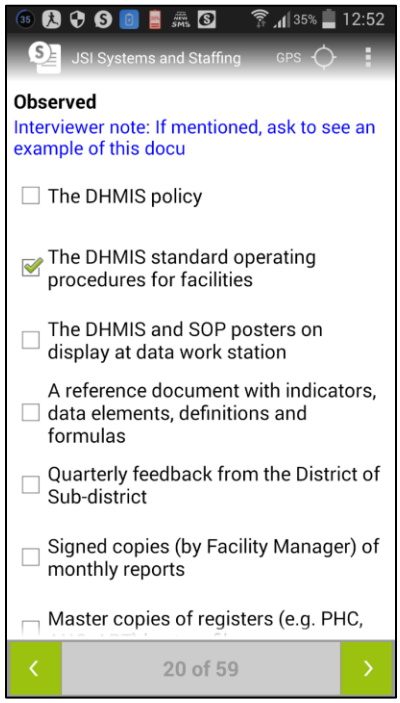


Figure 36: Percentage of facilities where fieldworkers observed the use of data management tools and resources

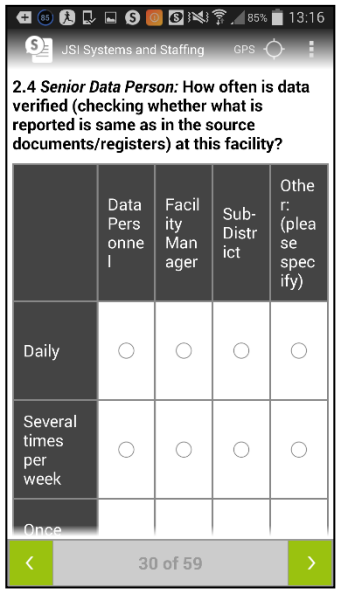
Fieldworkers observed the use of tools and resources in at least 70% of the facilities. They also observed that the master copies of registers were kept on file at 243 (85.0%) facilities, and the signed copies of monthly reports were kept at 244 (85.3%) of the facilities. This is the case as these were the main data sources that were used by the fieldworkers for the recount process. The DHMIS policy was also observed at 240 (83.9%) of the facilities indicating that they were in use followed by the DHMIS SOP being observed at 237 (82.9%) of the facilities. According to fieldworkers, the DHMIS and SOP posters were present at the workstations in 229 (80%) facilities. However, data audit reports were only observed at 199 (69.6%) facilities.



This verifies that at the time of the data recount, fieldworkers could visually locate the eight data tools and resources at more than 70% of the health facilities, which concurs with the facility managers and senior data persons' accounts discussed earlier.

Verification of Data

The main person responsible for data entry at each facility (i.e. a senior data person) was asked to provide information about the process and the frequency of data verification. They were asked a series of questions and the following answer options were provided: "Daily," "Several times per week," "Once per week," "Less often," and "Don't do data verification." The senior person at the clinic reported the frequency of verification conducted by the data personnel, facility manager, sub-district representative and any other personnel at the facility.



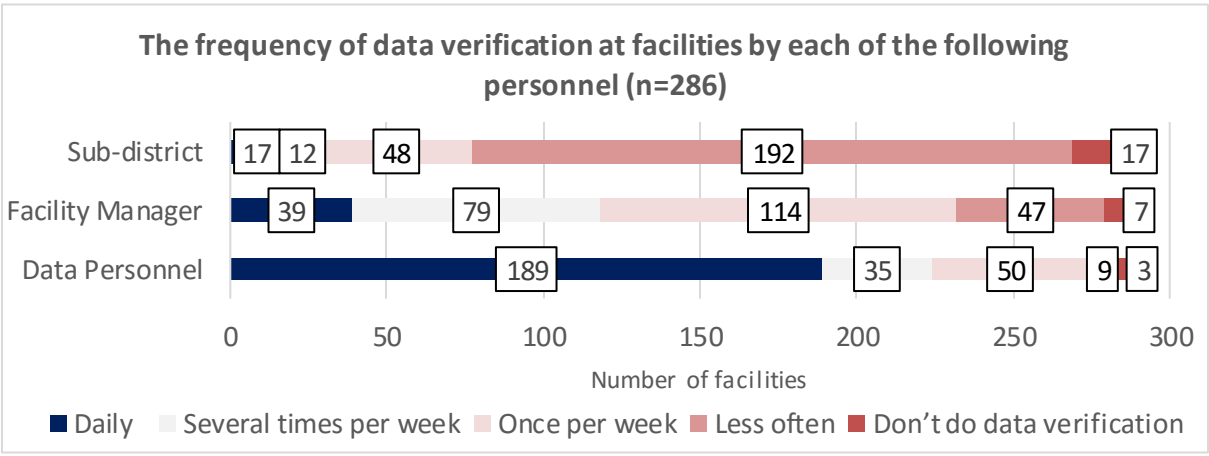


Figure 37: The frequency of data verification across facilities surveyed by the data personnel, facility manager, sub-district and other personnel

Overall, the frequency of data varies for the different roles of facility personnel responsible for aspects data management. Senior data persons reported that only 189 (66%) of 286 data personnel verify data on a daily basis. The majority of the facility managers (232 or 82%) reportedly verify data at least once per week. At 79 (28%) facilities the facility manager verifies data several times a week while thirty-nine (14%) facility managers verify data on a daily basis.

Furthermore sub-district representatives reportedly verify data at all but 17 (6%) of the assessed facilities. At 58 (20%) of the 286 facilities senior data personnel reported that data verification

was conducted by employees other than the data personnel, facility manager or sub-district representative. Other people that conduct data verification checks include:

- District Personnel (8 facilities)
- Clinicians (7 facilities)
- Data capturers (5 facilities)
- Clerks (4 facilities)
- Data companies (2 facilities)
- ANOVA Health Institute personnel (2 facilities)
- TIER representatives (2 facilities) and;
- Assistant facility managers (2 facilities)

Availability of tools, registers and forms

This sub-section discusses the availability of registers and data collection tools at PHC facilities. The senior data person was asked whether the facility ran out of registers over the three-month period, July, August and September 2015. The availability of eleven registers were reviewed. The registers included three PHC registers, four TB Registers, two HIV registers and two Maternal Health registers. The majority of facilities claimed registers were available (they had not run out of registers) for the three-month period.

Due to the rationalisation process the PHC daily register was discontinued at the end of July 2015 and two new registers, the PHC headcount register and the PHC comprehensive register were initiated from August 2015 onwards. Hence, the PHC daily register was used for the month of July and the PHC headcount register and the PHC comprehensive register were used for August and September 2015.

In terms of the PHC daily register the figure below illustrates that 91 (32%) of the 286 facilities claimed to have ran out of the PHC daily register in July 2015. This might be due to the change over to the new registers were the PHC daily register may have no longer been issued. Thirty-one (11%) facilities ran out of the headcount register for one month and at 35 (12%) facilities for the two months of August and September 2015. This trend was similar for the PHC comprehensive register. Again this might have been a result of the initial implementation of the rationalisation of registers.

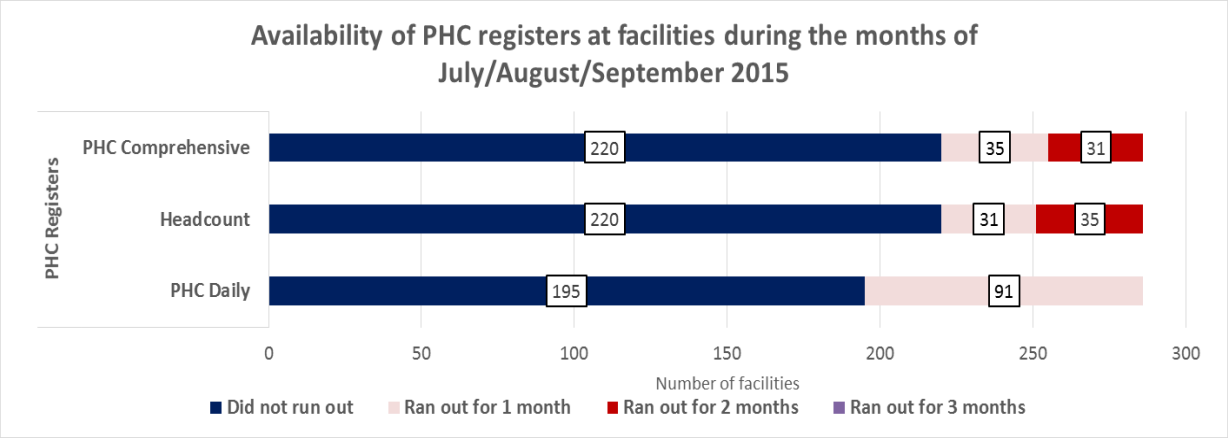


Figure 38: Availability of PHC Registers at facilities during the months of July, August and September 2015

In terms of the remaining registers, a higher number of facilities ran out of TB screening registers when compared to the other TB registers, with one-third of the facilities running out of this register for one month and more. Managers at thirty-five (12%) of the 286 facilities mentioned that they had run out of this register in all three months, with 21 (7%) mentioning that the facilities ran out of TB identification and follow up registers and TB registers in those months. Overall, two-thirds 195 (68%) of the facilities reported to have a constant supply of the TB screening tool within this period.

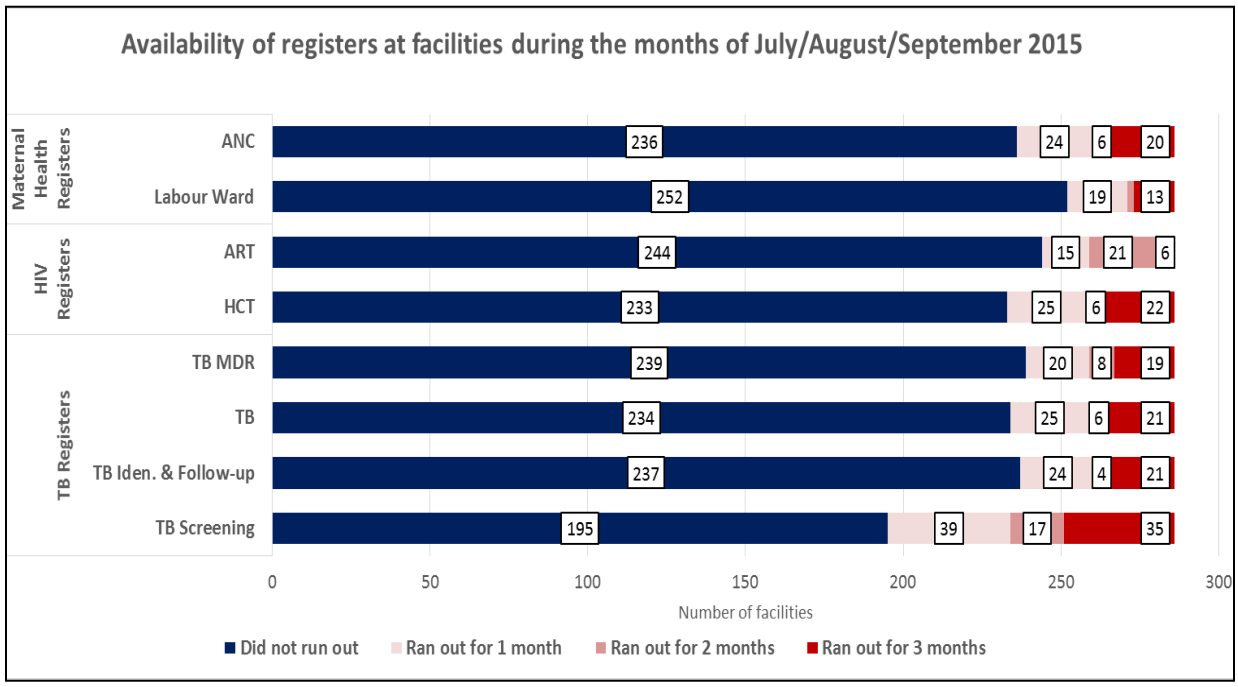


Figure 39: Availability of registers at facilities during the months of July, August and September

The figure below, presents the availability of the TB screening registers for the three months across the five Gauteng district municipalities. It is clear from the graph that in Sedibeng only a

few facilities ran out of registers, 6 (18%) out of 33 PHC clinics. Just under 50% of the facilities in the West Rand had their TB Screening registers available, with the remaining facilities either running out of registers for one month (11), two months (5) and all three months (9). This trend was similar across the Ekurhuleni, Tshwane and Johannesburg districts, with over 50% of the facilities having the TB screening registers available over the three-month period.

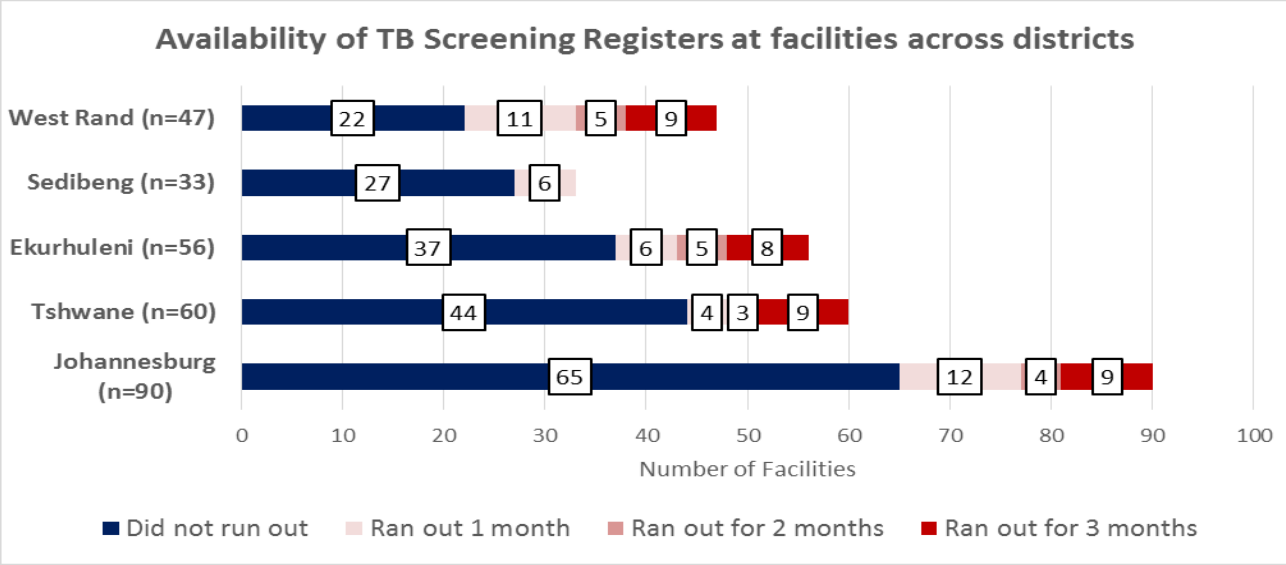


Figure 40: Availability of TB screening Registers at facilities across districts

Maintenance of data collection tools in line with Standard Operating Procedures (SOP)

At each facility, the fieldworker requested to view one register and checked whether it was maintained in accordance with eight of the SOP requirements, including: pages numbered correctly, entries completed correctly, lines drawn, entries legible, dated, signed off by data person, signed. The fieldworker indicated how well the SOP requirement was met by checking one of the following four options: "yes," "mostly yes," "mostly no," and "no." The SOP requirements were observed for the three PHC registers (PHC headcount register, PHC comprehensive register and PHC daily register) at a majority (248 or 86%) of the 286 sites as depicted in the table below.

Table 23: Data collection tools that were observed at PHC facilities at the time of the audit

Registers		Facilities	
		No.	%
HIV registers	ANC	3	1%
	ART	4	1%
	HTC	16	6%
Maternity	Labour ward register	1	0%
PHC	Headcount	54	19%
	PHC comprehensive	47	16%
	PHC daily	147	51%
TB tools	TB iden. and follow-up register	2	1%
	TB	4	1%
	TB screening	7	2%
	TB MDR	1	0%
Total facilities		**286**	**100%**

The following observed responses were provided for the PHC health facilities where data collection tools where checked.

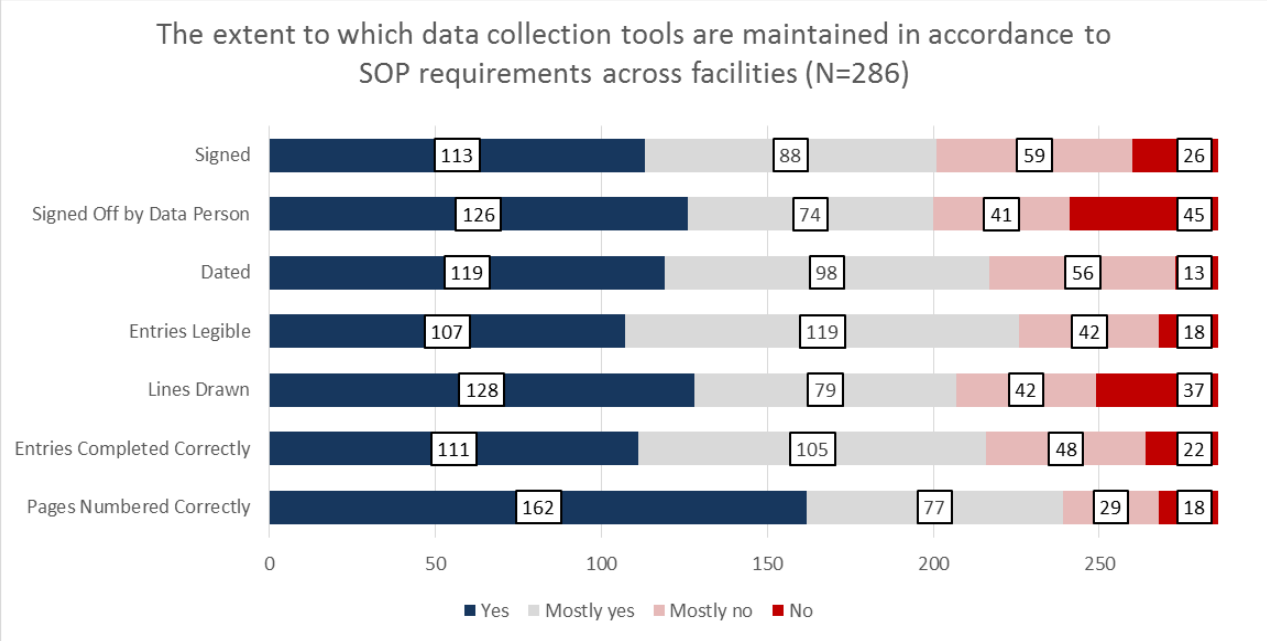


Figure 41: The extent to which data collection tools are maintained in accordance to SOP requirements across facilities

As can be seen from the figure above, over two-thirds of the facilities maintain the tools in line with SOP requirements. Only two SOP requirements were not well implemented and they both have to do with getting final approval from an authority. First, the appointed data person at only

86 (30%) facilities had signed off on the data when required to do so. Second, the data collectors at only 85 (29%) facilities appropriately signed the data collection tool. It was observed that at 45 (15%) facilities the data person had not signed off on the at all. In addition at 37 (12%) facilities lines were not drawn on the respective tools according to the SOP requirements.

With regard to the extent to which calculations on the data collection tools were correct, 95 (33%) of the 286 facilities had correct calculations with a further 140 (49%) facilities making a few errors in their calculations. Very few facilities (only 9) had incorrect calculations; however 42 (14%) facilities had many errors observed in their tools. This can be linked with the frequency of data verification at the sites, as these errors can be prevented with more frequent and accurate data verification.

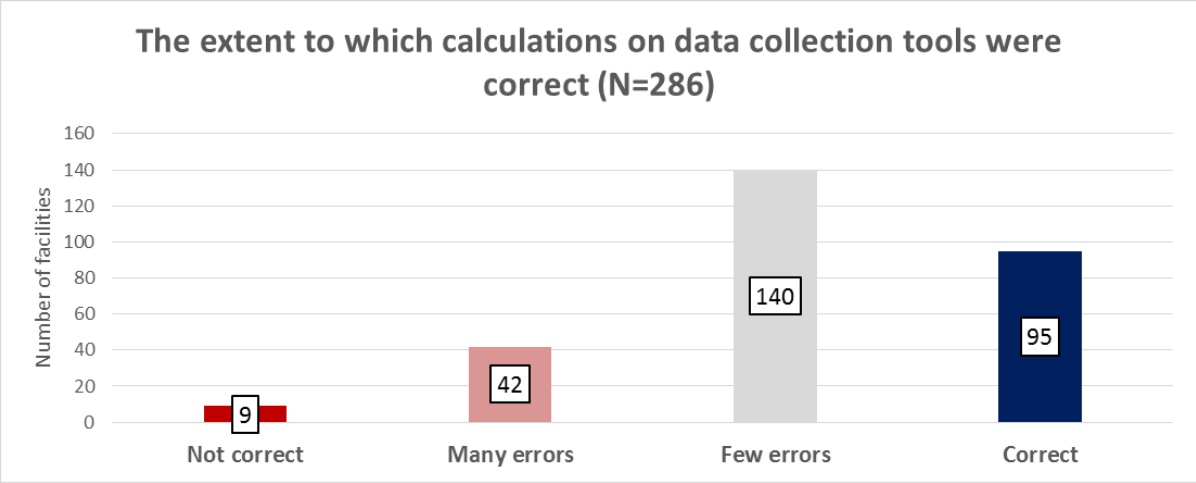


Figure 42: The extent to which calculations on data collection tools were correct

- **Summary—DHMIS Policy Implementation**

The evidence from the system and staffing survey completed with relevant facility staff shows that overall the DHMIS Policy is well implemented at a large number of PHC facilities. Evidence of this successful implementation is seen in terms of the use of tools and resources to guide data collection; the frequency of data verification; the availability of tools, registers and forms at health facilities; and the maintenance of data collection tools in line with SOPs.

Almost all of the facility managers and senior data persons at the 286 facilities indicated using the relevant tools and resources to guide data collection. Fieldworkers also observed that the tools and resources were actually used—not just reportedly so—at more than 70% of the health facilities. In addition, they observed that master copies of registers were in fact kept on file, and that signed copies of monthly reports were available at 85% of the health facilities.

Overall, the frequency of data verification fluctuates as per the role of facility personnel (i.e. data personnel, facility manager or sub-district representative). Data verification should take place more frequently and in particular, the senior data persons and the facility managers need to pay more attention to this task at least once a week. It is concerning that sub-district representatives reportedly only verify data at 6% of the health facilities. Instead it is other employees such as district personnel, clinicians, data capturers, facility clerks, data companies, TIER representatives, and assistant facility managers who generally take up this responsibility.

In terms of the availability of tools, registers and forms it was noted that two-thirds of the health facilities ran out of the PHC daily register in July 2015. This might be due to the change over to the new registers were the PHC daily register may have no longer been issued. About 12% of facilities ran out of the PHC headcount register and the PHC comprehensive register in August and September, which could also be due to the implementation of the rationalisation of registers. A higher number of facilities ran out of TB screening registers when compared to the other TB registers, with one-third of the facilities running out of the TB screening register for one month and more. In terms of the availability of the TB screening registers across districts, it was seen in the Sedibeng district that 82% of the facilities had registers available over the three-month period. Fewer than 50% of the facilities in the West Rand had their TB Screening registers available and at over 50% of the facilities in the Ekurhuleni, Tshwane and Johannesburg districts.

Overall, over two-thirds of the health facilities maintain the tools, especially the PHC tools, in line with SOP requirements. The two SOP requirements that were not implemented well includes the sign-off of the tool by a data person and the tool signed by data collectors. Almost 50% of the health facilities had few calculation errors noted on their data collection tools, with 18% of the facilities having many errors to no correct calculations in their PHC tools. From the

above the facilities have shown signs of adhering to the SOP requirements, however there are aspects that need further attention to ensure consistent adherence to these requirements.

4 Summary of Baseline Findings

Charles Babbage is often quoted as saying: "Errors using inadequate data are much less than those using no data at all". Indeed, this report indicates that while there are various data inadequacies that need to be addressed in the Gauteng Department of Health's strategic information system, the move towards digitised, accessible data is good overall for the country's goal to improve health outcomes.

- **Data quality needs to be improved**

This assessment found that data quality needs to be improved at most facilities. The biggest concern for data quality is over reporting, missing data, long-term data capturing fatigue, and poor recording in the registers. Performance across programmes and data elements varied greatly, but TB, HIV and Diabetes data were consistently the poorest on record. Data for facility headcount, MWCH, antenatal HIV care, immunization and hypertension is better in terms of quality, but still only about one third of PHC facilities are within the 5% acceptable error threshold set by the Auditor General.

Data quality across different districts varied greatly across the different indicators, but City of Johannesburg facilities, more than others, struggled with missing data. Sedibeng facilities, tended to underreport more than other facilities. West Rand facilities tended to perform better on a few of the assessed data elements.

At large clinics (i.e. those with monthly headcounts more than 11,000 clients) there is over-reporting on most of the assessed data elements. At small clinics (i.e. those with monthly headcounts smaller than 1,800 clients) there is missing data. Missing data is pronounced on data elements related to procedures that do not occur frequently at PHC facilities (e.g. deliveries and medical male circumcision), and on those requiring long-term tracking or waiting for test feedback (e.g. some of the TB elements, diabetes, and Infant 1st PCR test positive around 6 weeks)

- **Record keeping needs to be improved**

Record keeping at times was being done where ticks are not properly done in registers, especially for headcount and MCWH indicators. Official registers are not always used, and although this is only a problem in a small percentage (<5%) of PHC facilities, it compromises the overall quality of health data, and subsequently of health-related policies based on this flawed information.

- **Human error needs to be minimised in order to improve data quality**

Explanations for data discrepancies provided by data quality officers suggest that they view data quality problems as simply the result of human error in recording and summarising data. While this perception may not be entirely inaccurate, it absolves officers of taking ownership for ensuring that staff are adequately trained and supervised, and for responsibly implementing and following SOP protocols. The data quality officers need to better understand not only *why* such human errors are occurring but also *how* they can take steps to minimize such error in the future at their facility. Hopefully this report will provide a starting point in that direction.

- **Technical resources are available to support data quality in most facilities**

The findings reveal that the PHC facilities are reasonably equipped with the technical resources to capture health-related data. However, the resources are not necessarily sufficient to ensure excellent quality of the data that is captured.

From what was ascertained, throughout the province, all but three facilities had access to computers to use for capturing data, and a total of 669 computers were in use during the assessment period. Ekurhuleni district had the highest number of computers per facility with an average of 3.4, while Sedibeng had the lowest number on average at 1.4 per facility. On the downside, thirteen facilities had only one, partially functioning computer. On the upside, almost 90% of computers were reportedly fully functional.

Most computers were used for capturing more than just one dataset or for other tasks. Almost 80% of computers at facilities were shared and had more than one primary user. Although both these aspects might indicate efficient use of technical resources, sharing resources could also have potential negative impacts on data quality, and on access to a computer for data capturing.

In terms of the software used for data capturing, the results show that many different versions of the DHIS software are installed and used on computers. This might impede data quality and consistency of datasets. Although the system was not interrogated in detail, it is generally recommended that the number of software versions operating at the same computer be kept to a minimum in order to ensure consistency and lower costs.

While antivirus software is generally updated—except in the Sedibeng district—backing up computer data is not done regularly enough, which leaves the facilities' data vulnerable. The DOH should urgently address this issue. Moreover, only about 60% of computers are connected to the Internet, making it difficult in these instances to forward reports and data summaries to the appropriate stakeholders and district officials. Issues that could be improved are: regular backups of data at facilities, roll-out of DHIS versions, and connectivity to the Internet.

The results of the systems and staffing survey reveal that most clinics had the necessary staff to carry out data management. They employed data capturers, administrative clerks, facility managers, and volunteers or interns for routine data capturing and processing, and in some instance nurses were also doing this work. These individuals were usually well educated with at least a Grade twelve matriculation, and very often a post-high-school certificate of some kind. The burden of data work is heavy, and often it falls to primarily one person. In at least 61% of clinics, only one administrative person was employed to deal with data-related tasks and processing. At the other facilities, this task fell to clinical or support personnel who were not necessarily trained in data processing, or who had other competing duties. The Tshwane Metro Municipality had twenty clinics with no data capturers or dedicated administrative staff employed. Thus, in terms of human resources, PHC clinics had appointed employees dedicated to data-related tasks, but the number of employees dedicated to this task was less than adequate.

Other results show that the personnel engaged in data-related tasks do not necessarily have the capacity to ensure a high standard of data *quality*. Although they were well educated, appointed at the appropriate level, and usually full-time employees, there were instances where volunteers were being asked to do many of the data-related tasks. In addition, about 17% of these employees were recent hires who had worked at the clinic for less than a year. The findings also reveal that not all the employees interviewed had been trained on the various data-systems, possibly because many were new to the clinic. Only about 73% had received DHIS or TIER training, 36% had received ETR training, and less than 55% had received no formal instruction on DHMIS policy, or NIDS and PIDS. If high data quality standards are to be met, new employees must receive the appropriate data training.

Further findings demonstrate that facility managers and senior data persons were generally able to manage the volume of data in the time available, but that they could pay more attention to supervising data quality, not simply data collection and capturing. Having more individuals trained in data-related tasks helped the facility managers at the clinics balance out their tasks, but it also reduced the amount of time that personnel spent on data quality checks and verifications. In other words, awareness about data collection—that comes from the training—is potentially encouraging a culture of "quantity" and not "quality" in the data collection process. Personnel consume their time with data capturing and reporting, possibly at the expense of conducting data quality control.

Furthermore, data use at the facility level seemed to be limited, and this in turn is likely to have a negative impact on data quality. Meeting minutes and other office documents show evidence of routine discussion about data processing and data quality at many PHC facilities, indicating that employees are beginning to be concerned with the integrity of information. Nevertheless,

managers and senior data persons at the facilities do not rely much on the data being collected at their facilities. Reasons for this warrant further exploration, as it is likely to have a negative impact on data quality.

The evidence from the system and staffing survey completed with relevant facility staff shows that overall the DHMIS Policy is well implemented at a large number of PHC facilities. Evidence of this successful implementation is seen in terms of the use of tools and resources to guide data collection; the frequency of data verification; the availability of tools, registers and forms at health facilities; and the maintenance of data collection tools in line with SOPs.

Almost all of the facility managers and senior data personnel at the 286 facilities said they used the relevant tools and resources to guide data collection. Fieldworkers observed this use at more than 70% of the health facilities. In addition, they observed that master copies of registers were in fact kept on file, and that signed copies of monthly reports were available at 85% of the health facilities.

The frequency of data verification fluctuates as per the role of facility personnel (i.e. data personnel, facility manager or sub-district representative). Data verification should take place more frequently and in particular, the senior data persons and the facility managers need to pay more attention to this task at least once a week. It is concerning that sub-district representatives reportedly only verify data at 6% of the health facilities. Instead it is other employees such as district personnel, clinicians, data capturers, facility clerks, data companies, TIER representatives, and assistant facility managers who generally take up this responsibility.

In terms of the availability of tools, registers and forms it was noted that two-thirds of the health facilities ran out of the PHC daily register in July 2015. This might be due to the change over to the new registers were the PHC daily register may have no longer been issued. About 12% of facilities ran out of the PHC headcount register and the PHC comprehensive register in August and September, which could also be due to the implementation of the rationalisation of registers. A higher number of facilities ran out of TB screening registers when compared to the other TB registers, with one-third of the facilities running out of the TB screening register for one month and more. In terms of the availability of the TB screening registers across districts, it was seen in the Sedibeng district that 82% of the facilities had registers available over the three-month period. Fewer than 50% of the facilities in the West Rand had their TB Screening registers available and at over 50% of the facilities in the Ekurhuleni, Tshwane and Johannesburg districts.

Finally, over two-thirds of the health facilities maintain the tools, especially the PHC tools, in line with SOP requirements. The two SOP requirements that were not implemented well include the sign-off of the tool by a data person and the tool signed by data collectors. Almost 50% of the health facilities had few calculation errors noted on their data collection tools, with 18% of the facilities having many errors to no correct calculations in their PHC tools. From the above the

facilities have shown signs of adhering to the SOP requirements, however there are aspects that need further attention to ensure consistent adherence to these requirements.

Gauteng Health, therefore, still has some ways to go before realising the priceless value of good data. As this data assessment demonstrates, however, steps towards improving the data quality are underway.

5 Recommendations

The findings of the Data Quality Assessment were shared with managers and other stakeholders in the Gauteng Department of Health, during a symposium held on the 18th and 19th of April 2016. The symposium was entitled: "Integrated Performance Information Planning Symposium". Based on work in smaller commissions, a range of recommendations were made to improve the data quality in the Gauteng Department of Health. These recommendations are reflected here, in as far as they relate to the findings in this report.

Quick gains

In order to quickly improve the facilities' performance on audits conducted by the Auditor General, a campaign to ensure that every clinical and administrative staff member understands how to tick the various registers may be necessary. Work to help small clinics with document management may immediately yield an improvement in terms of "missing" data, and a focus on reducing the overcount at PHC clinics with headcounts of 11 000 and higher, may also significantly improve the data quality for Gauteng Province.

Strengthening use as a strategy to also improve Data Quality should be pursued as a primary objective

In the short term, it is recommended that the management accountability tool be used to report on health outcomes. The performance of facilities in clusters and districts in relation to targets could be reviewed quarterly, and further used to strengthen target setting.

The analysis of sub-facility level data across clusters or districts may also help to identify positive and negative outliers, that could be used as input into cluster meetings and the development of action plans. The possibility of checking on the functioning of referral pathways and levels of service may be possible with such analysis.

Such additional data analysis may require technology work to do significant integration across data sets and software systems (NHLS, SANBS, EMS, DHIS, BAS, PAAB, MEDICOM, PERSAL Rx Solution) and also some rework of the reporting functionality available in the health information systems. Ideally, data dashboards and early warning systems that draw on data as it is captured into the DHIS, would enhance the usefulness. In the meantime, a reporting tool that would help managers at facility level pull relevant information from the various systems into one report for discussion in meetings may be useful.

It may even require additional human resources to do the analysis at District and Cluster level. Public health specialists may be drawn upon to provide the analysis, but also to input into meetings together with the DCST. Training on analysis and interpretation of data may also be necessary for all facility heads. The investment in this kind of capacity is likely to drive data quality far more than an investment at the data capture side.

In the short term is it is recommended that management accountability tool be used and the school of public health should be mobilised to give input on analysis of data. Targets should be reviewed and the tool should be implemented to report on trends using the current indicators. The public health specialists could assist to analyse trends. Public health specialists, DCST are part of the cluster meetings. It is recommended that training in analysis and interpretation (to read data, write a story on data) is considered in the medium term and in the long term appropriate technology solutions.

Internal oversight and feedback from checking activities according to SOPs for Data Management

Since errors in data capturing can be expected to be found only if internal verification and checking of data takes place, it is recommended that daily, weekly and monthly verification as per the DHMIS policy and SOPs be championed at every facility. It was suggested that compliance with this is likely to increase, if the performance management of each manager also pertinently includes an item around data verification and checking.

A data quality assurance activity, conducted monthly at sub-district level by either district staff, or a peer from another facility, may ensure that a focus on data quality is maintained. The possibility of extending internal-audit functions to also include a sample of non-financial data may be investigated. A rigorous and routine quality assurance system will have the best chance of success if it is linked to appropriate systemic incentives and disincentives and consequence management.

A special data quality task team may be established that focuses on identifying areas for quick gains. For example if a few clinics are found to consistently over-report figures, the task team may intervene to support, monitor and track the facilities with specific data quality improvement measures.

Strengthening the Staffing Component responsible for Data Management

Where relevant, facility managers may lobby for long-term employment of data capturers and continued staff training. Whilst this is being pursued, however, the DOH may start to develop norms and standards for data management units in facilities. Such standards

could outline how many staff members can be expected per 5 000 patient headcount, what kind of skills and training these staff members require, and what kind of support may be necessary.

Since staff turnover among data capturers is a real issue, the DOH may look into the development of a standard orientation manual supported by appropriate audio-visual and interactive learning materials for data operators. This will reduce the down-time between the recruitment of new data staff members, and their efficient deployment at facility level. It may also help to standardise the understanding of the various health information systems.

Options for strengthening the HIM facility champions initiative may be further explored.

Strengthening of the Technical Capacity for Data Management

Beyond systems integration, and the development of reporting tools, attention should also be given to the issue of ensuring that adequate ICT resources are available at each facility. The number of functioning computers exclusively in use for data management is important to consider, but measures should also be taken to ensure that updated versions of software are deployed as soon as it is available, and that backups and antivirus software are maintained.

Continue alignment and rationalization of indicators and data collection tools, and investigate the introduction of some record management and data management tools

The Department of Health should prioritize a process for reviewing the data elements and indicators every two years. Changes to NIDS and PIDS should be made within the collaboratively set timeframes for such review. Retraining or re-orientation of all facility managers and their data teams is necessary to ensure that the changes are immediately made also at facility level.

A re-emphasis of SOPS at various levels may be necessary, and tools such as standardised checklists to ensure that data verification has taken place could be made available electronically or manually.

Further rationalization of registers may be possible in order to simplify the collection of information, remove the duplication of data collection, ensure that appropriate information is available at each level of care and decrease the number of registers. Such a rationalization may only be effective if record management can be improved. Assigning

unique patient identifiers, specifying uniform patient folders and introducing the scanning of patient records at facility level may assist in this regard.

6 References

Adelekan, T., N. Makhubele and A. Ngcobo. 2015. *Rapid Assessment Report of the Rationalised Registers*. Pretoria, South Africa.

Auditor General. 2010. *General Report on the National Audit Outcomes 2009-10*. Pretoria: Auditor General of South Africa.

Barron, P., C. Day and F. Monticelli. 2007. *The District Health Barometer 2006/07*. Durban: Health Systems Trust.

Council for Scientific and Industrial Research (CSIR). *Health Information Systems in Public Healthcare Facilities in South Africa: A review of the current status and trends*—Version 1.0. February 2013: unpublished report

CSIR and NDOH. 2013. *National Health Normative Standards Framework for Interoperability in eHealth in South Africa*. Pretoria: CSIR, National Department of Health.

English, R., T. Masilela, P. Barron and A. Schönfeldt. 2011. "Health Information Systems in South Africa." Pp. 81-89 in *South African Health Review* 2011. Edited by P. Padarath and R. English. Durban: Health Systems Trust.

Garrib, A., K. Herbst, L. Dlamini, A. McKenzie, N. Stoops, T. Govender and J. Rohde. 2008. "An Evaluation of the District Health Information System in Rural South Africa." *South African Medical Journal*, 98(7):549-52.

Gauteng DOH. 2013a. *Annual Performance Report for 2012/2013*. Pretoria: Gauteng Provincial Department of Health.

Gauteng DOH. 2013b. *Gauteng Department of Health: Strategic Plan 2015/16-2019/20*. Pretoria: Gauteng Provincial Department of Health.

Gauteng DOH. 2015. *Department of Health: Province of Gauteng Annual Report 2014/2015 (Pr 219/2015)*. Pretoria: Gauteng Provincial Department of Health.

JSI, ESI and NDOH. 2011. *District Health Information System (DHIS): Rapid Information Needs Assessment Draft Report*. Pretoria: John Snow Inc. (JSI), ESI, and National Department of Health.

Kareithi, R., I. Brodsky, S. Podges and T. Nutley. 2015. "Using the Data Demand, Quality, and Use Facilitated Group Self-Assessment Tool." *Technical Brief Series*. Chapel Hill, NC: Measure Evaluation. Available at http://www.cpc.unc.edu/measure/sifsa/ technical-briefs, accessed 24 March 2016.

Loveday, M., J. Smith and F. Monticelli. 2006. *Health Information Audit Report, South Africa*. Durban: Health Systems Trust.

Masilela T.C, Foster R, Chetty M. *The eHealth Strategy for South Africa 2012-2016: how far are we?* In: Padarath A, English R, editors. South African Health Review 2013/14. Durban: Health Systems Trust; 2014.

Massyn, N., N. Peer, A. Padarath, P. Barron and C. Day. 2015. *The District Health Barometer 2014/15*. Durban: Health Systems Trust.

Mate, K. S., B. Bennett, W. Mphatswe, P. Barker and N. Rollins. 2009. "Challenges for Routine Health System Data Management in a Large Public Programme to Prevent Mother-to-Child HIV Transmission in South Africa." *Plos One* 4(5):5483.

MEval and SIFSA. 2014. *Strengthening Evidence Based Health Management in Uthungulu District, Kwazulu Natal, South Africa*. Chapel Hill, NC: Measure Evaluation (MEval), Strategic Information for South Africa (SIFSA).

MEval and SIFSA. 2015. *Status of Data Demand, Quality and Use: Baseline Results of Assessed Health Districts in South Africa*. Chapel Hill, NC: Measure Evaluation (MEval) Strategic Information for South Africa (SIFSA).

Mphatswe, W., B. Mate, H. Ngidi, J. Reddy, P. M. Barker and N. Rollins. 2012. "Improving Public Health Information: A Data Quality Intervention in Kwazulu-Natal, South Africa." *Bulletin of the World Health Organisation*, 90(3):176-82.

NDOH. 2010. *Negotiated Service Delivery Agreement. A Long and Healthy Life for All South Africans*. Pretoria: National Department of Health.

NDOH. 2011a. *District Health Management Information Systems (DHMIS) Policy*. Pretoria: National Department of Health.

NDOH. 2011b. *Health Data Advisory and Co-Ordination Committee Report*. Pretoria: National Department of Health.
NDOH. 2012. *National eHealth Strategy, South Africa 2012/13-2016/17*. Pretoria: National Department of Health.
NMOH. 2014. *Annual Performance Plan for 2014/2015*. Pretoria: National Ministry of Health.
Nombembe, T. 2013. "Auditing Government Performance Information—Statement by the Auditor General of South Africa." Technical Article for the *International Journal of Government Auditing*. Available at http://www.intosaijournal.org/highlights/auditing_gov_performance_apr2013.html, accessed 24 March 2016.
Nutley, T. and H. W. Reynolds. 2013. "Improving the Use of Health Data for Health System Strengthening." *Global Health Action* 6(3). Available at http://www.globalhealthaction.net/index.php/gha/article/view/20001, accessed 24 March 2016.
Republic of South Africa. 1992. *Births and Deaths Registrations Act (Act 51 of 1992)*. Pretoria: South African Government.
Republic of South Africa. 1999a. *Statistics Act (Act 6 of 1999)*. Pretoria: South African Government.
Republic of South Africa. 1999b. *The Public Audit Act (Act 1 of 1999)*. Pretoria: South African Government.
Republic of South Africa. 2001. *Public Service Regulation: Government Notice No. R. 1 of 5, January 2001*. Pretoria: South African Government.
Republic of South Africa. 2002. *Electronic Communications and Transactions Act [No. 25 of 2002]*. Pretoria: South African Government.
Republic of South Africa. 2003. *National Health Act (Act 61 of 2003)*. Pretoria: South African Government.
Republic of South Africa. 2007. *Policy Framework for the Government-Wide Monitoring and Evaluation System*. Pretoria: The Presidency.
Republic of South Africa. 2012. *National Development Plan for 2030*. Pretoria: South African Government. Available at http://www.gov.za/issues/national-development-plan-2030, accessed 9 April 2016.
Schönfeldt, A. 2014. *Verifying the Current Number of Health Facilities in South Africa* (Interview date: 29 April 2014). Interviewers: R. N. M. Kareithi and S. Podges.
Sibuyi, I. N. 2014. *The District Health Information System (DHIS) as a Support Mechanism for Data Quality Improvement in Waterberg District, Limpopo: An Exploration of Staff Experiences*. Masters in Public Health, Department of Public Health, University of South Africa, Pretoria.
Vital Wave Consulting. 2009. "Health Information Systems in Developing Countries: A Landscape Analysis." *Research Paper and Strategic Briefing Series*. Palo Alto, CA: Vital Wave Consulting.
WHO. 2005. *International Health Regulations: Areas of Work for Implementation*. Geneva: World Health Organization (WHO).
WHO. 2007. *Everybody's Business—Strengthening Health Systems To Improve Health Outcomes: WHO's Framework For Action*. Geneva: World Health Organization. Available at http://apps.who.int/iris/bitstream/10665/43918/1/ 9789241596077_eng.pdf, accessed 24 March 2016.

7 Annexure A: Assessment Tools

8 Annexure B: PHC Recount Tables per District and Sub-District

9 Annexure C: PHC Systems and Staffing Tables per District and Sub-District

10 Annexure D: Hospital Recount Tables

11 Annexure E: Hospital Systems and Staffing Findings