UNIVERSITY OF THE WITWATERSRAND

MSC. RESEARCH REPORT

Comparability of the quality of verbal autopsy data from telephonic interviews versus face-to-face interviews in Agincourt, rural northeast South Africa

UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Student number: 2174298
Tinashe Tombe-Nyahuma

Supervisors: Dr Chodziwadziwa Kabudula

A Research Report submitted to the Faculty of Health Sciences in partial fulfilment of the requirements for the degree of Master of Science (MSc) in Epidemiology - Public Health Informatics

August,2024

Tinashe Tombe-Nyahuma ,2024

Comparability of the quality of verbal autopsy data from telephonic interviews versus face to face interviews in Agincourt, rural northeast South Africa.

Supervisors:
Dr Chodziwadziwa Kabudula, University of Witwatersrand

Declaration

I hereby declare that this research report is project work carried out by me under the guidance of Dr. Chodziwadziwa Kabudula. I have taken care in all respects to honour the intellectual property rights of others and have acknowledged the contribution of others. I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Parktown, Johannesburg, August 2024

_____

Tinashe Tombe-Nyahuma

Dedication

To my family, thank you for your love and patience.

Abstract

Background

The Medical Research Council (MRC) and University of the Witwatersrand (WITS) Rural Public Health and Health Transitions Research Unit have relied on face-to-face interviews to collect verbal autopsy data since 1992. In March 2020, all in-person interviews conducted by the Wits Agincourt Research Unit were suspended and replaced with telephonic interviews due to the COVID-19 pandemic. Mortality data has a significant role in measuring the health status of a population as well as informing health policy, programme implementation and evaluation. It is important that quality data is collected when monitoring mortality indicators to ensure accurate and complete information is available for health analysts. Literature argues that the mode of data collection contributes to the quality of data and some studies have shown that data collected using telephonic interviews are inferior to those collected using face-to-face interviews. The effect on the quality of data induced by the change in method in 2020 to collect verbal autopsy data by the MRC/Wits Agincourt Research Unit has not been assessed.

Aim

The purpose of this study was to assess if the quality of verbal autopsy data collected by the MRC/Wits Agincourt Research Unit through telephonic interviews during the COVID-19 pandemic was comparable to the verbal autopsy data collected through face-to-face interviews prior to the pandemic and after. The study also aimed to assess the reliability of verbal autopsy data that was collected telephonically.

Methods

This was a descriptive research study that used data profiling techniques to assess the accuracy, completeness and consistency of verbal autopsy data gathered from the Agincourt Health and Socio-demographic Surveillance System population during the periods 1 April 2020-31 March 2021 using telephonic interviews and 1 April 2019-31 March 2020 and 1 April 2021-31 March 2022 through face-to-face interviews. The Kruskal-Wallis H Test was used to measure if the data quality results from the three datasets were statistically comparable.

Results

The verbal autopsy data that was collected via telephonic interviews during the pandemic and the verbal autopsy data that was collected after the pandemic via face-to-face interviews had

similar overall data quality score of 93.8%. The data that was collected before the pandemic had an overall data quality score of 91.6%. At 95% confidence interval the Kruskal Wallis test result (p-value = 0. 89) showed that there was no difference in the mean rank quality assessment results between data collected face-to-face before and after the COVID-19 pandemic and that collected telephonically during the pandemic.

Conclusion

Our results show that the verbal autopsy data was comparable and there was not much variation in the quality of data collected via telephonic interviews versus that collected via face-to-face interviews in all three data quality dimensions that were measured. Where the results varied slightly, the difference was not statistically significant. Verbal autopsy data collected telephonically is reliable and can be used as a method of data collection by researchers.

Acknowledgements

I wish to acknowledge the following people and organisations:

- I would like to thank my supervisor Dr Chodziwadziwa Kabudula for his support and guidance.
- Thank you to MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) for providing the dataset that was used in this research.
- Thank you to my husband Munyaradzi Nyahuma for the support and taking on the role of father and mother to our children throughout this journey.
- To my parents Mr and Mrs Tombe for always checking up on me and to my siblings Winfrida Tombe, Kudakwashe Tombe and Tawanda Tombe for the love and support.
- To the classmates I started this journey with who have now become family, thank you for the discussions and encouragement.
- Above all, I would like to thank God for his blessings and giving me the strength to complete this journey.

List of Figures

List of Tables

Contents

1 Introduction

1.1 Background

There is a general consensus amongst researchers that in-person interviews are the 'gold standard' for conducting interviews (1). In-person interviews are said to have the advantage of capturing verbal and non-verbal cues like emotions and behaviours by participants and allowing interviewers an opportunity to ask further questions and probe for additional information increasing the chances of collecting high quality data compared to telephonic interviews (2). Restrictions due the COVID-19 pandemic however led to increased usage of telephonic interviews as a method of collecting data in research projects across the globe (3). In the interest of safety and to comply with legal measures put in place, researchers had to adapt and incorporate remote methods of data collection (3) (4).The impact posed by the change in methods of data collection on the quality of data collected is still to be investigated across many studies.

Similar to many research entities in Sub-Saharan Africa, the Medical Research Council (MRC) and University of the Witwatersrand (WITS) Rural Public Health and Health Transitions Research Unit (MRC/Wits-Agincourt Research Unit in short) which has been investigating dynamics of health, population, and social transitions in a rural setting in northeast South Africa since 1992 (5) rely heavily on face-to-face interviews to collect research data. Before the COVID-19 pandemic, primary data consisting of important events like births, deaths, and migration for the Agincourt Health and Socio-demographic Surveillance System (HDSS), which is operated by the MRC/Wits-Agincourt Research Unit, was collected on an annual basis through face-to-face interviews(6). Since 1992, for each death recorded on the annual census, the MRC/Wits-Agincourt Research Unit team used verbal autopsies conducted via face-to-face interviews to collect information for ascertaining causes of death(6). A verbal autopsy is a tool

used to determine cause of death of an individual through a combination of techniques (7). It involves the use of standard questionnaires by trained personnel through an interview to obtain details about the signs, symptoms and demographic information of the deceased from a close caregiver prior to their death(8). The data collected from the questionnaire, thereafter is used by either physicians or a computer-based probabilistic model to determine the cause of death (9). The cause of death is defined as "the condition or injury that started the sequence of events leading directly to death"(10).

Mortality data plays a significant role in measuring the health status of a population as well as informing health policy, programme implementation and evaluation(11). Data on cause of death is monitored to evaluate implemented health interventions together with assessing where interventions maybe required(11). It is important therefore to ensure that all health systems collect data that is of high quality, reliable and accurate (12).Despite the important need for health systems to have accurate , timely and reliable mortality data, the quality of mortality data is poor in low to middle income countries (13) (14).This is mainly due to inefficient civil registration and death certification systems (15). Using 1996 mortality data, the WHO classified South Africa as one of the countries with poor quality mortality data with a death registration completeness score of less than 70% and a score greater than 20% on ill-defined-codes (16).

To improve the quality of mortality data ,South Africa has incorporated other methods of collecting data on mortality, with verbal autopsies being one of the methods(16).Verbal autopsies have played a significant role in improving the quality of mortality data in low and middle income countries(7). WHO recommended verbal autopsy as an efficient tool for collecting mortality data in countries that do not have comprehensive systems for collecting mortality data, particularly mortality data that would have otherwise gone unrecorded due to the death happening outside a healthcare facility (17)(18) (19) (20). To ensure that mortality

data collected through verbal autopsies is of good quality and is consistent, the WHO has gone to great lengths to develop verbal autopsy standardised forms, training instruments on how the verbal autopsies should be administered as well as manuals on the interpretation of cause of deaths (19).

The ISO/IEC 25012 standard defines data quality as "the degree to which a set of characteristics of data fulfils requirements"(21).The key features of quality data include the completeness, the accuracy, relevance and consistency of data(22). It is crucial that public health researchers collect accurate, relevant and complete data on disease outbreaks, mortality rates, cause of death for better decision making (12). Surveillance data of poor quality is detrimental to coming up with better health interventions and health policies, which could lead to dire consequences for the health of the population(12).

In March 2020 all in-person interviews conducted by the MRC/Wits-Agincourt Research Unit were suspended and replaced with telephonic interviews (23). Literature argues that the mode of data collection contributes to the quality of data (24) and that data collected using telephonic interviews may be inferior to those collected using face-to-face interviews (17). However, the effect on the quality of data induced by the change in method of data collection by the MRC/Wits-Agincourt Research Unit had not been assessed. The purpose of this study was to assess if the quality of verbal autopsy data collected by the MRC/Wits-Agincourt Research Unit prior to the pandemic as well as after the pandemic through face-to-face interviews was comparable to the verbal autopsy data collected telephonically during the pandemic. Comparability was measured using accuracy, completeness, and consistency of the data. Findings will assist in making informed decisions for future data collection methodologies in times of pandemics as well as contribute to the literature of whether data collected via telephonic interviews can be compared to data obtained via face-to-face interviews

1.2 Statement of the problem

The MRC/Wits-Agincourt Research Unit has for three decades (1)(6) used face-to-face interviews to collect verbal autopsies for the Agincourt's Health and Socio-demographic Surveillance System (HDSS). However, during the period 1 March 2020 – 28 February 2021, telephonic verbal autopsies were introduced due to the COVID-19 pandemic(23). Literature argues that telephonic interviews are a "methodological compromise" due to the quality of data collected (17, (26). The quality of verbal autopsy data obtained by the MRC/Wits-Agincourt Research Unit using telephonic has not been assessed to give confidence to the data. It is not known if the data is comparable to the data collected via face-to-face verbal autopsies and there are limited studies that have assessed verbal autopsy data collected via alternative methods as face to face interviews have been regarded as the gold standard method for collecting verbal autopsy data (20) (27).

1.3 Justification of the study

Mortality data has a significant role in measuring the health status of a population as well as informing health policy, programme implementation and evaluation (1). Data on cause of death is monitored to identify health problems as well as evaluate implemented health interventions (1,2). It is important that quality data is collected when monitoring mortality data to ensure that accurate and complete information is available for health providers, analysts and policy makers (28). Comparing the quality of data produced by telephonic verbal autopsy interviews with the quality produced by face-to-face verbal autopsy interviews will give confidence in the data that was collected and will inform future research data collection methods. The study will also contribute to the literature of whether data from in-person interviews is comparable to data from telephonic interviews(26).

1.4 Research Question

Is there a difference in the quality of verbal autopsy data collected in Agincourt, from 1 April 2020 to 31 March 2021 by means of telephonic interviews, and that collected through face-to-face interviews between 1 April 2019 - 31 March 2020 and 1 April 2021 – 31 March 2022?

1.5 Study Aim

To assess and compare the quality of verbal autopsy data collected through telephonic and face-to-face interviews.

1.6 Objectives

1. Assess the quality of verbal autopsy data collected through telephonic interviews during the period 1 April 2020-31 March 2021, using three quality indices: syntactic accuracy, completeness, and consistency.
2. Assess the quality of verbal autopsy data collected during the period 1 April 2019-31 March 2020 and 1 April 2021-31 March 2022 via face-to-face interviews using three quality indices: syntactic accuracy, completeness, and consistency.
3. Compare the data quality assessment results of verbal autopsy data collected via telephonic interviews during the period 1 April 2020-31 March 2021 and that collected via face-to-face interviews during the period 1 April 2019-31 March 2020 and 1 April 2021-31 March 2022

1.7 Definition of key terms

Key Term	Definition
Before Covid	The period before the COVID-19 pandemic (1 April 2019 - 31 March 2020). During this period data was collected via face-to-face interviews.
During Covid	The period during the COVID-19 pandemic (1 April 2020 - 31 March 2021). During this period data was collected via telephonic interviews
After Covid	The period after the initial phase of the COVID-19 pandemic (1 April 2021 - 31 March 2022) when data was collected via telephonic interviews

1.8 Outline of Research Report

The research study is outlined as follows**:** Chapter **2** reviews literature on telephonic versus face-to-face interviews. Different data quality dimensions are also explored in this chapter. Chapter **3** focuses on the methods that were used to conduct this study. Chapter **4** presents the results of the data quality assessments. Chapter **5** discusses the findings of the study as well as the limitations of the study. The last chapter, Chapter **6** provides conclusions drawn from the findings and recommendations

2 Literature Review

2.1 Introduction

This chapter reviews the literature on comparability of face-to-face interviews versus telephonic interviews. Literature on data quality dimensions as well as mortality data is also explored.

2.2 Comparability of face-to-face versus telephonic interviews

Literature indicates that method of data collection significantly determines the quality of data collected in research studies (24). There are different methods of collecting data in population and health research, including in-person interviews, telephonic interviews, online questionnaires and focus groups depending on the study design chosen by the researchers (29). While, interviews are the commonly used method of collecting data for population and health research (30), there is ongoing debate about whether telephonic interviews produce the same data quality as face-to-face interviews (2) (31) (26).

In-person interviews are considered to be the "gold standard" when compared to telephonic interviews in population and health research. Some scholars argue that in-person interviews produce good quality data because interviewers have the ability to capture verbal and non-verbal cues like emotions and behaviours of participants during the interview (26). This is based on the premise that, in-person interviews have the ability to allow the interviewer to build great rapport when interviewing the participants and create a conducive environment for the participant to participate fully as well as allow the interviewer to probe for more information if the need arises(26). Lechuga et al. (2012) however argued that an interviewer using the telephone can also build good rapport with respondents as well as probe for further information

if there is anything that needs to be clarified despite not having physical access to the respondent (32) (33).

Telephonic interviews are regarded as a ‘’methodological compromise”(26). Studies have been put forward that test this theory and the results indicate that the data quality produced in telephonic interviews compared to face-to-face interviews was not different (26) (34). Sturges et al. (2004) in a study that aimed to assess the perceptions of correctional officers and jail visitors concluded that “mode of interview did not influence the data to any significant degree” (35). Sweet (2002) in a phenomenological study found that the quality of the data between the two modes of interviews “was not noticeably different”(34).

2.3 Mortality and cause of death

Cause of death data is an essential component of mortality data (36) (37). It provides information on what the population is dying of to guide policy makers and health interventions(36) (38) (39). It is important therefore that data that is accurate and complete on cause of death is collected. The South African Civil Registration System is the main source for cause of death data in South Africa and acquires the data from death notification forms or death report forms(11). The death notification forms are completed by a medical practitioner and in the absence of a medical practitioner the death report form is completed by a traditional healer, headman or police (11) (40). Using data from the death notification forms and death report forms, classification of cause of death is then done by applying the International Classification of Diseases (ICD) codes(11).

Although South Africa’s death registration completion score was estimated to be around 90% by 2007, cause of death data quality remains low (41). A report from Statistics South Africa on 2016 data indicated that 13% of the deaths had ill-defined codes while 13% had an invalid cause of death in the civil registration system(40). Incorrect classification of HIV

deaths to another cause was also identified in several studies (42) (43) (44). The misclassification and ill-defined codes are caused by various factors that include poor completion of death notification forms by medical professionals, legal and social issues like sensitive deaths like suicide that lead to the incorrect information being completed on the death notification form or death report forms as well as lack of skilled personnel to accurately code the cause of death information (43) (45) (27).

Data on cause of death is also collected through verbal autopsies from several research sites across South Africa including seven South African Population Research Infrastructure Network (SAPRIN) research sites that the Agincourt Research site falls under (46). While verbal autopsies have helped improve the quality of mortality data in places where deaths would have gone unrecorded due to lack of civil registration systems, there is ongoing debate on its reliability in determining cause of death (47). There are several factors that contribute to coming up with the final cause of death when using verbal autopsies and some of the limitations mentioned include the reliability of the interviewer recording accurate information given by respondents during the interview especially on the open narratives section of the verbal autopsy tool (48). Recall bias from respondents has also been highlighted as a possible limitation of verbal autopsies as it has an impact on the accuracy of data recorded (49). There is limited literature on the impact of using alternative methods to collect verbal autopsy data as face to face interviews were identified as the standard method to collect verbal autopsy day(20) (27).

2.4 Data Quality Dimensions

While there are studies from multiple disciplines that have assessed the comparability of face-to-face interviews to telephonic interviews using different research questions and addressing

various measures of quality, few focus on the quality dimensions of the data (2) (26) (34) (50) (51) (52) (53). Comparability found in literature mainly focuses on the response patterns found in the data. For example, the response rates to sensitive questions and response patterns based on demographic information. Study participation rate, field performance and suitability of data collection mode to answer research question is also a quality measured in many of the studies (31) (26) (50) (51) (53).

According to Pipino et al. (2003) "the concept of data quality is multidimensional" (54). There is no golden standard framework for data quality(55). Literature on data quality measures is inconsistent and is based on the researcher's perspective(56) (57). Black et al. (2020) identified 60 dimensions of data (58) in their data quality dimensions dictionary. While Stausberg et al (2015) conducted a literature review assessing 36 papers from data quality management registers and cohort studies published between 2005 and 2013.The results indicated 34 data quality indicators (59). Bian et al (2020) came up with 12 common data quality dimensions in their systematic review of Electronic Health Records literature (60).The systematic review included the widely referenced papers of Weiskopf et al (2013) and Chan et al (2010).

Kahn et al (2016) acknowledge the multi dimension definitions that come with data quality and review literature in electronic health records data to come up with a harmonised data quality framework(57). Stausberg et al. (2015) identify accuracy, completeness and correctness as the leading indicators used by researchers in the literature published between 2005 to 2013 to measure data quality in medical research and cohort studies (59). Other researchers identify accuracy, comparability, currency and accessibility as the commonly found data quality indicators in literature (12) (22) (55) (57) (60). Black et al (2013) go on to define completeness, timeliness, validity, accuracy, uniqueness, and consistence as the "core six dimensions of data" (61).

The research and variables under study determine the type of data quality measurement to be used during data quality assessment (62). Not all data quality measurements apply to certain research and variables. Prioritised indicators of the research study have to be established prior to reaching the stage of data quality assessment in order to come up with adequate measures of quality (12) (62) (63). This research study focused on the comparability of telephonic verbal autopsies and face-to-face autopsies using completeness, accuracy and consistency as the key quality measures. The completeness, accuracy and consistency of sections three, four and five of the WHO 2020 VA standard questionnaire (see Table 3.1 in the appendix) as well as certain demographic information in the verbal autopsy questionnaire are essential for coming up with cause of death, it was important therefore that these variables were measured in the assessment (64) (65).Data profiling, was used to conduct the data quality assessments. Jack Olson defines data profiling as "the application of data analysis techniques to existing data stores for the purpose of determining the actual content, structure, and quality of the data"(66).

3 Methods

This section gives a description on the research design, data source and the methods that were followed to conduct the data quality assessment.

3.1 Research Design

This was a descriptive research study that used data profiling techniques to assess the accuracy, completeness and consistency of verbal autopsy data gathered from the Agincourt Health and Socio-demographic Surveillance System population during the periods 1 April 2020-31 March 2021 using telephonic interviews and 1 April 2019-31 March 2020 and 1 April 2021-31 March 2022 through face-to-face interviews (67).

3.2 Data Source

The Agincourt health and socio-demographic surveillance system (HDSS) was established in 1992 by the Medical Research Council (MRC) and the University of Witwatersrand South Africa to investigate dynamics of health, population and social transitions in remote rural South Africa(5). The study site is located in the Bushbuckridge area in Mpumalanga Province and covers 31 villages with a population of about 116 000 people(5). Key aspects of the Agincourt HDSS is the monitoring of demographic events in the population through annual updates that are collected by supervised field workers for each household in the study site through face to face interviews(6). All new events like births, migration as well as deaths are recorded and a longitudinal dataset is maintained and stores all updated information (5).

Mortality data as mentioned above, is one of the vital events collected by the surveillance system using verbal autopsies. A validated questionnaire is used to interview closest caregiver of the deceased and responses to questions related to the signs and symptoms experienced by the deceased prior to their death are recorded. Before the COVID-19 pandemic and after the COVID-19 pandemic data for the surveillance system was collected by field workers using

face to face interviews and during the pandemic that had to change due to the social distancing regulations and data was collected via telephonic interviews

3.3 Dataset

The study was conducted using the three verbal autopsy datasets extracted from the Agincourt health and socio-demographic surveillance system. The first dataset contained data collected using face-to-face interviews during the period 1 April 2019 – 31 March 2020 before the COVID-19 pandemic. The second dataset contained verbal autopsy data collected during the period 1 March 2020 to 31 March 2021 using telephonic interviews. The third dataset contained data collected post the COVID-19 pandemic during the period 1 April 2021 to 31 March 2022.The extracted datasets had different sample sizes with the telephonic dataset being the smallest with 382 observations. In order to increase the power of the data quality assessment, the sample size in the datasets to be compared had to be of similar size. Therefore, 400 observations out of 800 were randomly sampled from the pre-covid dataset. The same principle was applied for the third dataset which had 973 observations.

3.4 Scope of the research

3.4.1 Descriptive Statistics

Prior to conducting the data quality assessment, exploratory data analysis was performed on the three datasets to understand and compare their characteristics. The analysis was performed on the respondent information, deceased information and open narrative sections.

Respondent Data

Respondents are family members/ caregivers of the deceased that were interviewed by field workers either through face-to-face interviews or telephonic interviews to obtain information related to symptoms experienced by the deceased prior to the deceased's death to establish

cause of death. Descriptive statistics on respondents age, sex and relationship to the deceased were compiled and compared among the three datasets

Deceased Data

Descriptive statistics on the deceased were compiled and compared among the three data collection periods. This included the deceased's age, sex, where the deceased died, deceased 's medical history, symptoms experienced prior to the deceased's death, deceased's lifestyle, as well as responses to sensitive questions that were asked about the deceased for example deceased's HIV status.

Narrative Word Count Distribution

The narrative variable contains the information given by the respondents when they were asked to give an open narrative of the symptoms experienced by the deceased prior to their death. For exploratory purposes, the distribution of the number words in the narratives for each dataset was assessed. Stop words like "the"," he"," she"," is", "a" and "are" were removed prior to the word count.

3.4.2 Data Quality Assessment

Data profiling techniques adopted from Abedjan et al illustrated in Fig 3.1 have the ability to assess data quality dimensions comprehensively therefore were used for the data quality assessment(68). The quality assessment measured the syntactic accuracy, completeness and consistency of 60 variables listed in the appendix section in table A.1identified as essential on the verbal autopsy questionnaire. Table 3.1 provides the definitions of the three data quality dimensions that were used in the data profiling process.

Table 3.1 : Definitions of three data quality dimensions measured in the study

Data Quality Dimension	Definition
Syntactic accuracy	"The degree to which a given data is correct and reliable"(69)
Completeness	"The extent to which data is not missing and is of sufficient breadth and depth for the task at hand"(54)
Consistency	"The degree to which data are free of contradiction with themselves "(70)

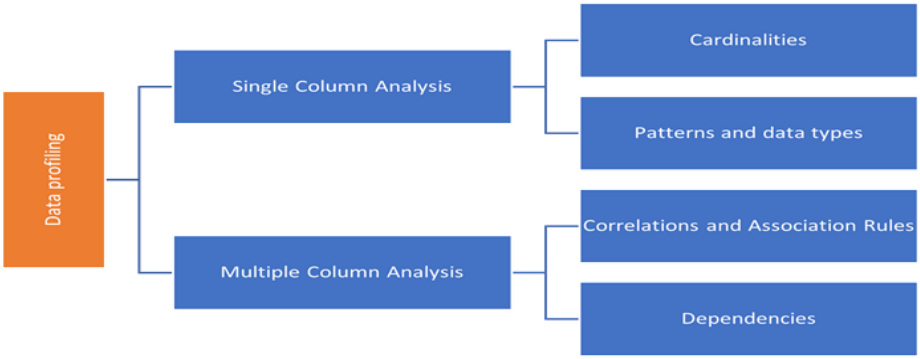

Figure 3.1:Data Profiling Tasks*(68)*

3.4.2.1 Single Column Analysis

Single column analysis involves the analysis of values in a single column of a table/dataset where value ranges, data type patterns and number of non-values are assessed (68) (71) (72). In this study, single column analysis was conducted for each dataset on the 60 selected key variables (Table 3.1 in Appendix 1), looking at the values stored in each column, independent of other columns to measure the three quality dimensions.

Syntactic Accuracy

The accuracy of values in columns was measured using different methods. This included assessing the recorded values versus the defined response options expected to be recorded, for example the "where_died" variable had six defined response options to choose from. All recorded values that did not fall within the classified options or where not missing would be classified as invalid. Range analysis was also used on continuous variables to assess if recorded values were feasible for example the range on the deceased age, if the deceased's age was recorded as 145 years then it would be classified as invalid. Nine variables out of the 60 were not included in the syntactic accuracy analysis because there was no way of measuring their accuracy, the values were only based on the information given by the respondent, for example the deceased's illness duration. The accuracy score was calculated by correct data values/total data values in column. The percentage score of records with correct data values were recorded for each variable in each dataset and compared. Table 3.2 gives examples of the methods that were used to conduct the syntactic accuracy assessment.

Table 3.2 : Examples of variables and methods used during syntactic accuracy assessment

Variable Name	Method of assessment
Deceased's Age	• To measure the accuracy of the recorded deceased's age variable, the accurate age was calculated using the deceased's recorded date of birth and date of death. The newly calculated accurate age was then used to compare to the recorded age. If the recorded age was not similar to the calculated age, then it was classified as incorrect

	• Range analysis was also applied on the variable to measure if the recorded age was within the expected age range of a human being.
Respondent's date of birth	• This was assessed by looking at range to see if it was plausible. For example, if the respondent's year of birth was same as year of interview then that record would be regarded as inaccurate.
Marital status	• The recorded values were assessed by checking if they met the defined responses under marital status for example married, single ,divorced or living with partner
Relationship to deceased	• The recorded values were assessed by checking if they met the defined response options for relationship to deceased variables. Any recorded values that did not fall under the defined response options was classified as inaccurate
Narrative_1	• The recorded values were assessed by checking if they met the defined response options that fell under narrative_1 variable
Family cause	• Accuracy was checked on spelling and repetition of the same cause of death expressed differently

Completeness

The non-null values in each column for each dataset were assessed based on sections. The 60 variables selected for the study were divided and placed under eight sections of the verbal autopsy questionnaire. The completeness assessment was then measured using the average percentage score for each section. Table 3.3 below presents description of the sections and the variables that were classified under them.

Table 3.3 : Verbal Autopsy Questionnaire Sections

A. Respondents Information
Contained demographic information on respondent as well as relationship with deceased
• Respondent's sex • Relationship to deceased • If respondent lived with the deceased • Respondent's date of birth • Relationship other • Interview date
B. Deceased information
Contained demographic information on the deceased
• Study number • Deceased's date of birth • Year deceased died • Deceased's marital status • Deceased's sex • Deceased's date of death • Deceased's age • Deceased's place of death
C. Open narrative from respondent
Contained open narrative descriptive information on the deceased cause of death
• Narrative • Cause of death narrated by the respondent • Classified
D. Deceased Medical History
This section represented data on deceased's medical history. The respondent answered yes or no to whether the deceased had been diagnosed with the following diseases prior to their death
• Tuberculosis • Aids diagnosis • Positive HIV test result • Positive malaria test result

- Negative malaria test result
- Hypertension
- Diabetes
- Asthma
- Cancer
- Sickle cell anemia
- Dengue Fever
- Heart disease
- Epilepsy
- Kidney disease
- Liver disease

E. Deceased Final Symptoms

Represented data on symptoms experienced by the deceased prior to their death

- Fever
- Extreme fatigue
- Breath
- Diarrhea
- Abdominal pain
- Mental confusion
- Night sweats
- Cough
- Chest pain
- Vomiting
- Headache
- Convulsion

F. Deceased Lifestyle

This section contained information on the health lifestyle the deceased lived.

- Alcohol consumption
- Alcohol type drink
- Smoker
- Struggle to be active
- Sedentary lifestyle
- Mobility
- Physically active
- Deceased consumed alcohol
- Alcohol consumed by deceased
- Deceased smoked
- Deceased struggled to be active
- Lived a sedentary lifestyle
- Deceased was mobile
- Deceased was physically active

G. Deceased Illness Duration

Information on how long the deceased was sick for.

- Length of illness before death
- Illness duration in terms of Days/Weeks/Months
- Sudden death
- Duration of illness before death
- Illness duration in terms of days/weeks
- Deceased died a sudden death

H. Civil Registration Information

Contains data on the civil registration of deceased death

- Death form
- Death Form Location
- Immediate cause of death on certificate
- Date of death on certificate
- Death certificate number
- Registered cause of death on certificate

The completeness of responses to sensitive questions e.g HIV status, TB diagnoses and completeness of important variables like narrative, dod, narrative_1, famcause1 were also explored and compared.

Consistency

The consistency of data values and the patterns of data values found in a single column was assessed. All 60 variables were assessed to measure the consistency of data type values for each column. Consistency was also checked on the recorded responses versus the predefined responses the interviewer was supposed to choose from for specific variables. For example, on narrative_1 variable and family cause variable. The following data values represented in table 3.4 were expected in the narrative_1.

Table 3.4 : Defined response options for question related to narrative 1 variable

	Defined Responses	
	• Chronic kidney disease	• Dialysis
	• Fever	• Heart attack
	• Heart problem	• Jaundice
	• Liver failure	• Malaria
	• Pneumonia	• Renal (kidney) failure
	• Suicide	• None of the above words were mentioned
	• Don’t know	• Refused to answer

Consistency on this variable was measured on the percentage of recorded values that were consistent with the defined option responses.

Family Cause

The family cause variable contained the cause of death that was declared by the respondent during the interview. There were no mandated value options for this variable as the respondent

could declare any cause of death based on what they observed prior to the deceased's death. Consistency was therefore measured by checking if cause of death was recorded consistently. For example, old age was one of the causes of death given by the respondents, some records had it as "old age", some "too old" some "she was too old", some "deceased was too old". Despite these values meaning the same thing, in data quality terms they would require cleaning before being classified as one category during analysis stage.

3.4.2.2 Multiple Column Analysis

Multiple column analysis was conducted on each dataset where syntactic accuracy, completeness and consistency was assessed to measure the quality of the relationship of values across the columns.

Syntactic Accuracy

Reference data validation was conducted for columns that had an association. One of the two variables that were assessed was narrative and narrative_1 variable. The narrative variable was a variable where the data collectors recorded the symptoms experienced by the deceased prior to their death as narrated by the respondent. The narrative_1 was a variable that derived its value from the narrative variable. The data collector recorded at least one symptom mentioned in the narrative. The quality assessment measured if the values recorded in narrative_1 were present in the narrative. The data collectors were expected to choose from the predefined options also represented in table 3.4 above if mentioned in the narrative by the individual they were interviewing. Data accuracy was assessed on each allocated symptom per observation.

Key analysis was also conducted to check if there were duplicates in each dataset. Not all variables in the datasets were linked, therefore not all were assessed for accuracy.

Completeness

The overall completeness for all variables per record was measured. Completeness on dependent variables was also assessed.

Consistency

This method of assessment measured the quality of consistency for stipulated association rules amongst values in columns. For example, the variable "relationship other" was associated with the variable "relationship to deceased". The association rule was that, if the value in "relationship to deceased" was mother, father, spouse, child, friend, grandmother, family member or friend then "relationship other" would be null. If the value was "other" then relationship other was expected to have a value. Not all columns were assessed for consistency because not all variables had an association. Percentage score of consistent variables for each dataset was recorded. Table 3.5 below shows the variables that were measured for consistency as well as the method that was used to assess consistency.

Table 3.5 : Variables measured for consistency and method of assessment

Variable Name and Method of Assessment
Relationship to deceased & relationship other
• Percentage score of consistent values found in Relationship to deceased & relationship other
Illness duration and sudden death
• Illness duration was a variable in the dataset that stated how long the deceased was sick or how long the deceased experienced symptoms. The range was based on days, weeks or months. Sudden death was a variable in the dataset that identified if an individual had a sudden death or not. Sudden death is defined as a natural death that occurs instantly within less than 24 hours of the symptoms occurring (73). The assessment compared the value in the illness duration variable to the value in the

sudden death variable. If the illness duration value was 0 and sudden death value recorded as yes, then that record would be classified as having consistent data on those 2 variables. If the illness duration was greater than 0 and sudden death recorded as yes then that record would be classified as inconsistent.

Alcohol usage and alcohol type

- Percentage score of consistent values found in Alcohol usage & alcohol type. If the respondent answered yes to alcohol usage, then alcohol type was expected to have a value. If the respondent answered no, then alcohol type was expected not to have a value

Statistical Analysis

To measure if the data quality results from the three datasets were comparable, a statistical test was used. The data from the results of the data quality assessments was not normally distributed therefore the Kruskal-Wallis H Test was used to compare the quality results of the three datasets. For example, the non-parametric test was used to assess if the average score for completeness for each dataset was significantly different from the others. The Kruskal-Wallis H Test is a non-parametric test that is used to test if data are of the same distribution (74). For the comparisons where the results were significantly different, a further Dunns test was conducted to establish which pair had a difference that was statistically significant. The Dunns test is a non-parametric test that "conduct multiple pairwise comparisons for median difference" (75).

3.5 Experimental Setup & Ethics

3.5.1 Experimental Environment

R programming language using R Studio as the Integrated Development Environment was used to conduct the data quality assessment. The data processing and analysis was performed on a computer running on a windows 10 platform with Core i5 processor, random access memory (RAM) of 16GB and Storage of 512GB.

3.5.2 Ethical Considerations

The protocol was submitted to the Human Research Ethics Committee at the University of the Witwatersrand for ethics approval. Permission was sought from the MRC/Wits Rural Public Health and Health Transitions Health Research Unit to use the verbal autopsy data. All precautions were taken to protect the data. Data was anonymized and personal information that included names and identity number were excluded. The datasets were only used for purposes of this project.

4 Results

4.1 Introduction

This chapter provides results from the exploratory data analysis on the sample characteristics of the three datasets and outlines the results of the data quality assessments that were conducted using data profiling techniques that measured the syntactic accuracy, the completeness and consistency of data in each dataset.

4.2 Descriptive Statistics

Descriptive statistics are presented on respondent data, deceased data and open narrative data to understand the overall characteristics of the datasets.

4.2.1 Respondent Data

Table 4.1 below gives summary statistics on the respondent's sex, age and relationship to the deceased for all data collected during the three time periods.

Table 4.1: Summary of respondent's sex ,age and relationship to deceased

	Before COVID-19 (N=400)		**During COVID-19 (N=382)**		**After COVID-19 (N=400)**	
	n	**%**	**n**	**%**	**n**	**%**
Sex						
Male	76	19	104	27	90	22
Female	312	78	278	73	303	76
Missing	12	3	0	0	7	2
Respondents age						
0	6	2	2	1	3	1
10 – 19	14	4	12	3	19	5
20 – 29	70	18	70	18	91	23
30 – 39	68	17	83	23	65	16
40- 49	76	19	61	16	58	15
50 – 59	68	17	58	15	56	14
60 – 69	50	13	59	15	46	12
70- 79	30	8	26	6	39	10
80+	7	2	10	3	15	4
Missing	11	3	1	0,3	8	2

Relationship to Deceased	Before COVID-19 (N=400)		During COVID-19 (N=382)		After COVID-19 (N=400)	
	n	%	n	%	n	%
Another relationship	46	12	5	1	25	6
Child	66	17	91	24	92	23
Family member	131	33	106	28	187	47
Spouse	43	11	86	23	43	11
Friend	3	1	0	0	5	1
Grand mother	31	8	22	6	9	2
Parent	68	17	72	19	30	8
Refused to answer	2	1	0	0	0	0
Missing	10	3	0	0	7	2

Analysis of the respondent's data shows that females participated in the verbal autopsy interviews the most compared to males in all three datasets, with all data collection periods having 73% and above female participation. The median age for respondents in all three datasets was similar with median age for data collected before covid and after covid having median age of 44 years and the data that was collected during covid had a median age of 45 years. The respondents that were interviewed during the covid period via telephonic interviews had the highest mean age of 46 years while respondents from before and after the covid period had a mean age of 45 years. All three datasets indicated irregularities when it came to the participation of respondents below the age of 10 with some respondents recorded with the age of zero which was not a valid age for participation. In the before covid dataset, 2% of the respondents were recorded to be aged 0 , 1% during covid and 0.1% after covid.

The family member category followed by the child category had highest respondents when it came to the relationship to deceased.

4.2.2 Descriptive Statistics on Deceased

Deceased Age and Sex Distribution

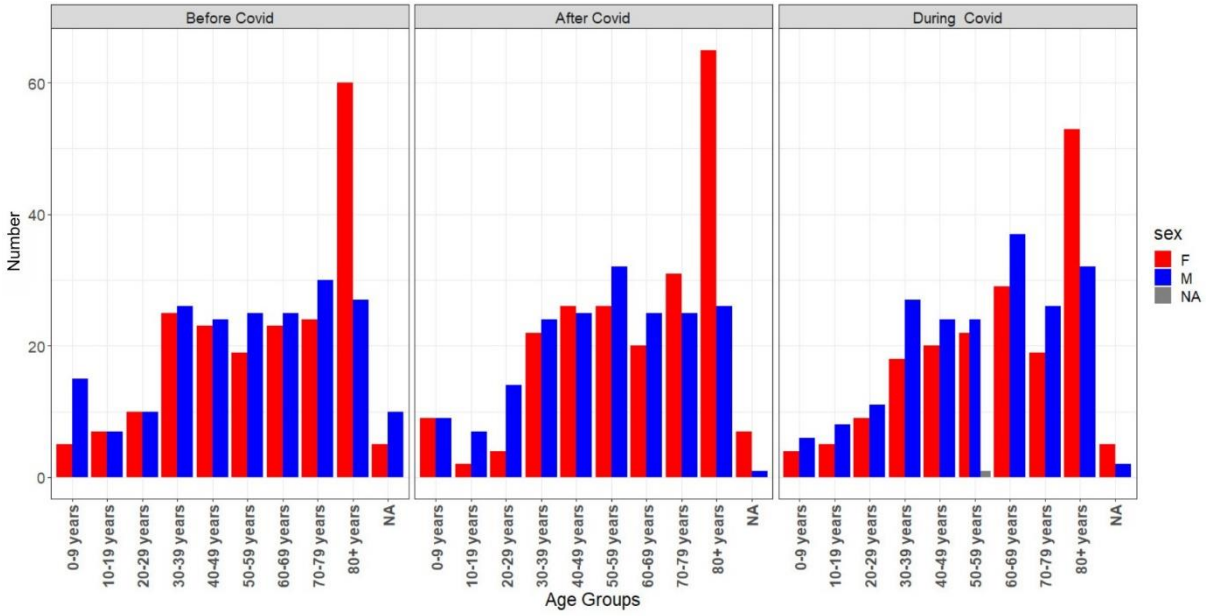

Figure 4.1 : Graph showing deceased's age and sex distribution for the three periods

Figure 4.1 above illustrates the distribution of age and sex amongst the deceased for the three data collection periods. The distribution of the deceased's age and sex was similar in all three datasets with females 80 years and above showing the highest deaths. The average age at death for males was within the 50-56years age range in all datasets with the mean age of males in the data that was collected via face-to-face interviews before covid having a mean age of 53 and 60 for females. The mean age for males at death was 56 years and for females 62 in the data that was collected during covid via telephonic interviews. The deceased males in the data that was collected after covid had a mean age of 53 years and females 63 years. As the age increased

the number of deaths also increased in all three datasets. All three datasets had missing data on the age of the deceased.

The final symptoms that led to death

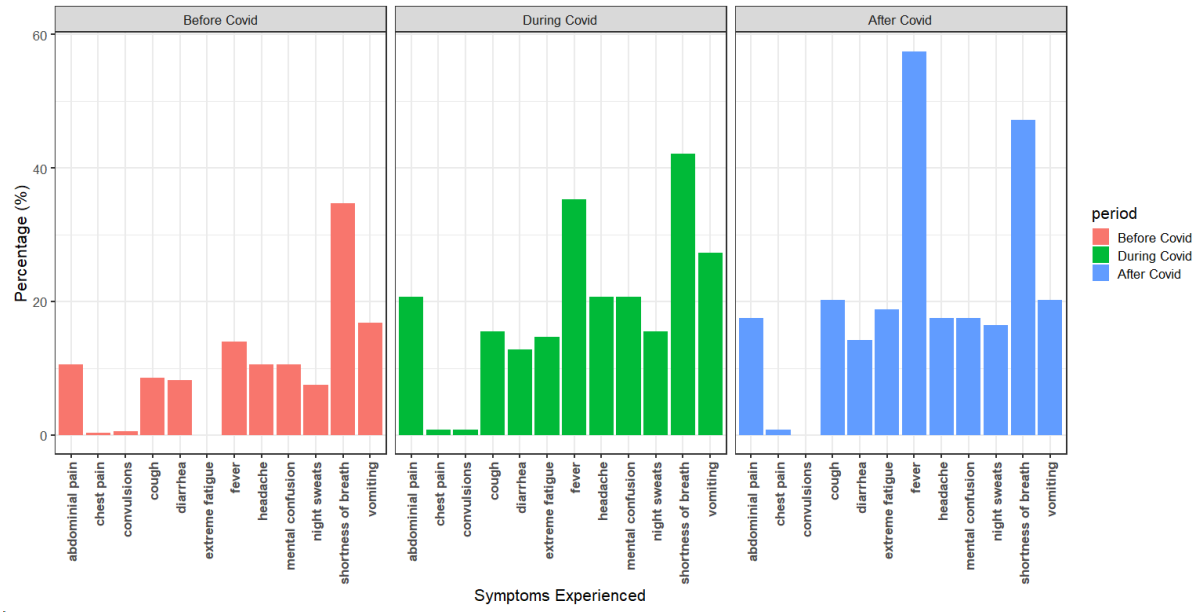

Figure 4.2 : Distribution of deceased's symptoms associated with final death

Figure 4.2 represents the final symptoms exhibited by the deceased for all three periods. The results did not show any noticeable pattern. The most encountered symptoms before covid and during covid was shortness of breath with a frequency of with 35% and 42 % respectively. Fatigue was the common symptom encountered in the after covid dataset with a frequency of 58%. The second most encountered symptom in data that was collected before covid was vomiting with 17%, during covid was fever at 35% and after covid, shortness of breath at 48%. Most of the symptoms experienced were below 20% for all data collection periods with chest pains and convulsions recording the lowest percentage or frequency.

Medical History Associated with final illness

Figure 4.3 represents medical history associated with the deceased's final illness. The distribution of the diagnostic symptoms in all three datasets was similar. The results indicated

that hypertension was the highest diagnosed disease in all three datasets followed by diabetes and HIV. Dengue fever and sickle cell had a zero diagnosis in all three datasets.

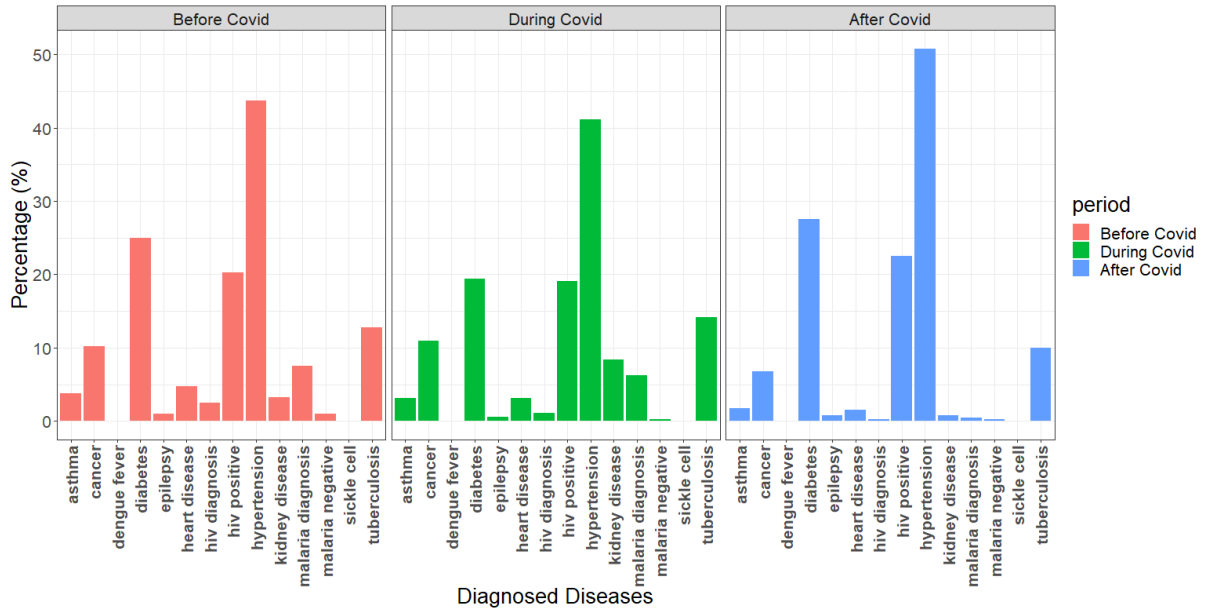

Figure 4.3 : Distribution of deceased's medical history associated with final illness

Deceased Place of Death

The distribution of deceased's place of death in all three datasets is represented in Figure 4.4. The results show that the proportion of people dying at home in all the datasets ranged between 40-48 % with 41% in the data collected before covid, 40.6% during covid and 47.3% after covid. The percentage of people who died at the hospital were high in the data that was collected before and after covid with 40% dying at the hospital in the before covid dataset and 35.8% after covid. No hospital deaths were recorded for data that was collected telephonically during covid. The results also show that the data that was collected during covid was complete for place of death variable with no missing data compared to the before and after covid dataset that had 3% and 2 % respectively of missing data.

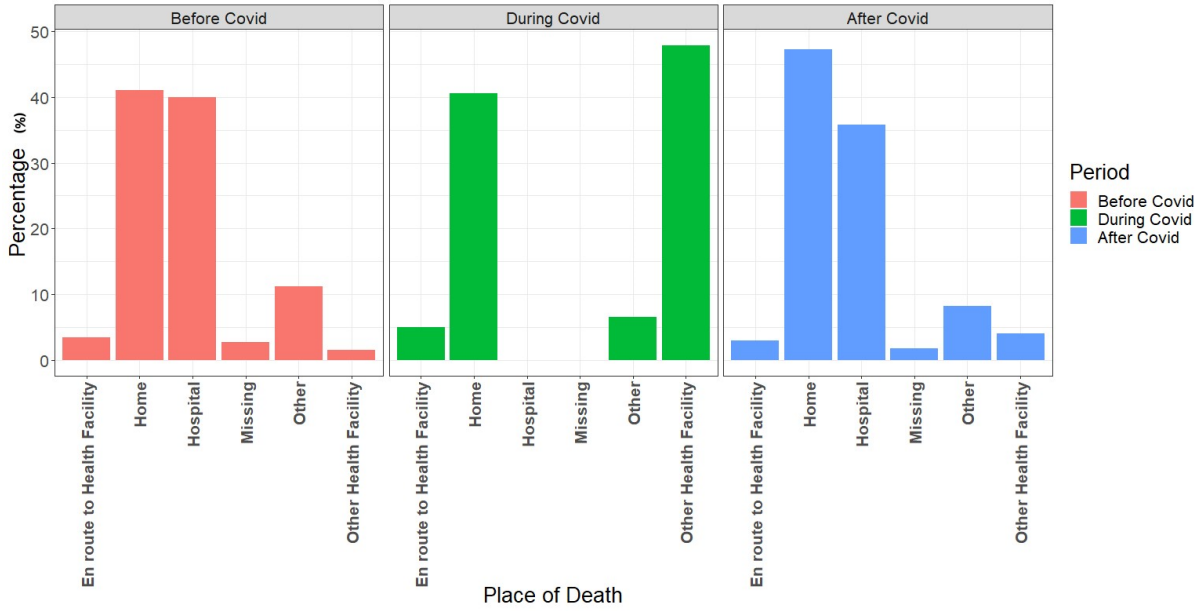

Figure 4.4 : Distribution of deceased's place of death

Narrative Word Count Distribution

Figure 4.5 shows the word count distribution on the narrative variable for each data collection period. The results indicated that the distribution of the number of words in the narrative variable was not similar in the datasets. Data that was collected using telephonic interviews during covid had the highest word count median of 37.5. The data that was collected before covid using face-to-face had a median of 24 and the data that was collected after covid had a median word count of 29. The word count in the before covid dataset ranged from 0 to 118 with a mean of 29 words, 118 being the outlier and 75% of the records had a word count score of 46 and above. The during covid dataset had minimum word count of 0 and maximum of 138 and mean word count of 40 and 75% of the records had a word count of 57 and above, with 2 records having outliers. The after covid dataset had minimum word count of 0 and maximum of 78, mean of 28 words, 75% had a word count of 43 words and above. There were outliers before covid and after covid but there were no outliers after covid. A Kruskal Wallis test was conducted to assess if the word count distribution was the same in

the three different datasets and the results showed that the distribution was not the same at 95% confidence interval and p-value of 0.0, a Dunns test was further conducted to establish which pair had a difference that was statistically significant and the results illustrated in table 4.2 show that word distribution count was not the same in all three data collection periods.

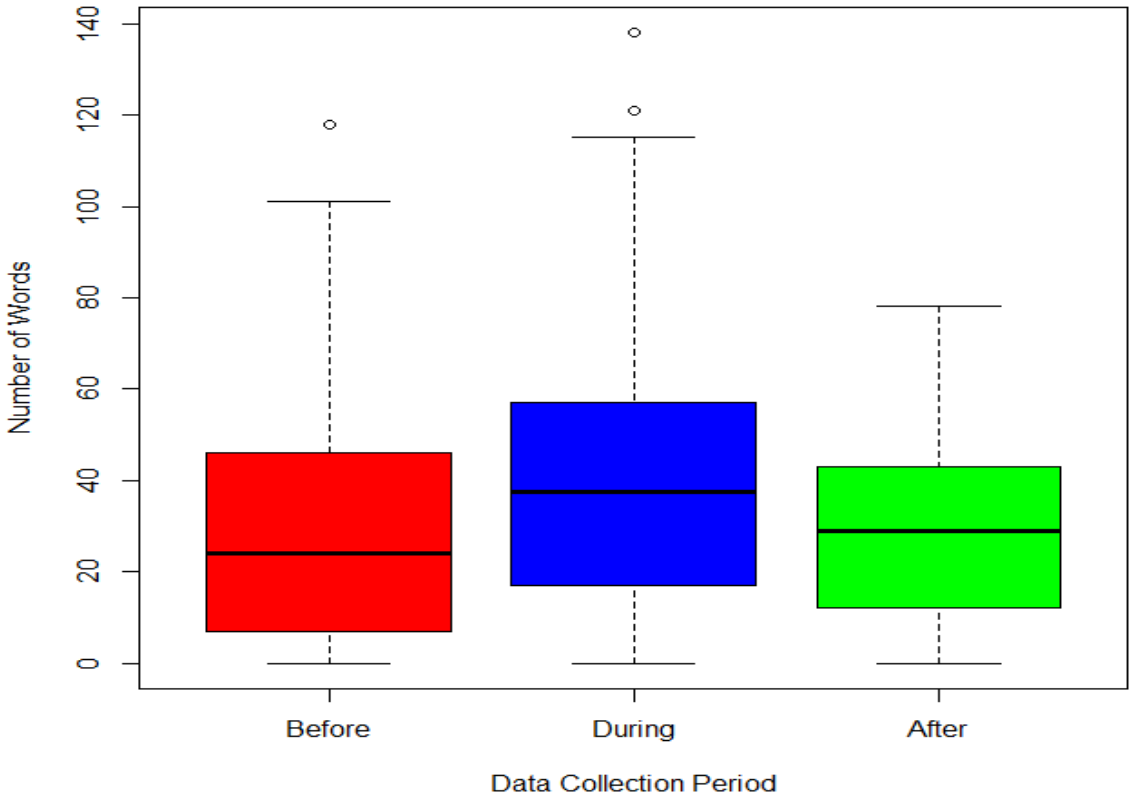

Figure 4.5 : Box plot showing word count distribution for the three data collection periods

Table 4.2 : Dunns test results showing the difference in sum of ranks in word count for the three data collection periods

Group 1	**Group 2**	**No of observations**	**No of observations**	**Statistic**	**p-value**	**adjusted p-value**
After Covid	Before Covid	400	400	-3.5	0.0	0.0
After Covid	During Covid	400	382	5.1	0.0	0.0
Before ovid	During Covid	400	382	8.5	0.0	0.0

4.3 Data Profiling

The results of the single column analysis and multicolumn analysis for all three datasets are presented in this section.

4.3.1 Single Column Analysis

Syntactic Accuracy Assessment

Table 4.3 represents the results of the syntactic accuracy assessment that was conducted on 7 variables for each dataset.

Table 4.3 : Syntactic accuracy assessment results

Variable Name	Before Covid accuracy % score	During Covid accuracy % score	After Covid accuracy % score
Recorded Deceased Age Accuracy	91	44.8	92
Recorded Deceased age range accuracy	100	99.5	100
Respondents dob	98.5	98.7	99.3
Marital Status	100	100	100
Relationship to deceased	100	100	100
Narrative1	98.7	98.1	99.6
Family cause	97.3	98.4	98.5
Average Accuracy	**97.9**	**91.3**	**98.5**

From the variables that were assessed for syntactic accuracy the results indicated that the data that was collected after the pandemic had the highest accuracy rate with a syntactic accuracy rate of 98.5%. It was followed by the dataset which had data collected before the COVID-19 pandemic via face-to-face interviews which had an average accuracy rate of 97.9%. The data that was collected during the COVID-19 pandemic telephonically had the lowest accuracy rate of 91.3% accuracy. The data was not normally distributed therefore a Kruskal-Wallis H Test was conducted to measure if the results in the different data collection periods were statistically significant. At 95% confidence interval the results indicate that there

was no difference in the rank sum of the accuracy scores in all the three datasets as indicated in table 4.4 (p-value of 0.7).

Table 4.4 :Krukskal-Wallis H Test results showing no difference in rank sum of syntactic accuracy scores of the three periods

Chi-squared value	Degrees of freedom	p-value
0.6	2	0.7

Completeness Assessment

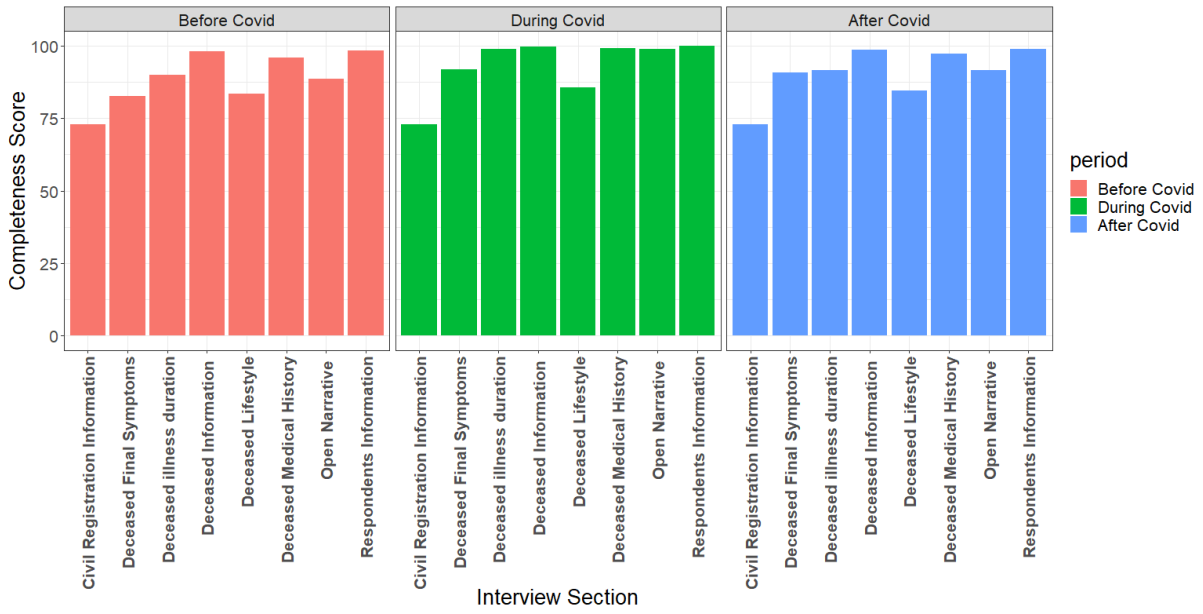

Figure 4.6 : Distribution of completeness scores per interview section for the three data collection periods

Completeness is measured as "the extent to which data is not missing and is of sufficient breadth and depth for the task at hand" (40). In this study completeness was measured on the non-null values in each column per section of the questionnaire for each dataset. Figure 4.6 represents the results from the assessment. The completeness scores per section were similar in the three data collection periods. The civil registration information section recorded the lowest rate of completeness with 72.8% in all three datasets. This was followed by data on

deceased lifestyle, where the completeness rate was 83.5% in the before dataset ,85.6% in the during covid dataset and 84.6% in the after covid dataset. Respondent's information had the highest completeness score of above 98% in all three datasets. The sections in the data that was collected telephonically during covid had the highest average completeness score of 93.3% followed by the data that was collected face-to-face after covid which had a completeness score of 90.75%. The before covid dataset had the lowest completeness score of 88.7%.A Kruskal-Wallis test was conducted to assess if the completeness scores were statistically significant and the results indicated that the difference in scores were not statistically significant with a p-value of 0.3 at 95% confidence interval represented in table 4.5 below.

Table 4.5*:*Krukskal Wallis test results showing no difference in rank sum scores for the completeness of data in the three data collection periods

Chi-squared value	**Degrees of freedom**	**p-value**
0.6	2	0.3

Results on Exploratory Analysis on sensitive Questions

Table 4.6 below,shows the completeness scores of the responses to sensitive questions in all three datasets. The results indicate that the completeness of sensitive questions in the data that was collected before the COVID-19 pandemic via face-to-face interviews had an average completeness rate of 95.7%. The data that was collected during the COVID-19 pandemic telephonically had 99.2% completeness making it the dataset with the highest completeness rate. The data that was collected after the pandemic had a completeness rate of 97% on sensitive questions.

Table 4.6: Completeness scores of responses to sensitive questions

Variable Name	**Before Covid completeness % score**	**During Covid % accuracy score**	**After Covid % accuracy score**

Positive HIV test	95.5	99.2	97.0
Professional HIV diagnosis	95.8	99.2	97.0
TB Diagnosis	95.8	99.2	97.0
Average Completeness	**95.7**	**99.2**	**97.0**

Kruskal-Wallis H Test Results

The data was not normally distributed therefore a Kruskal-Wallis H Test was conducted to assess if the completeness scores in the different data collection periods were significantly different. The results from the Krukskall test indicated that mean rank sum of the completeness of the values in the data collection periods was not the same as shown in table 4.7 with a p-value of 0.02. Further analysis was conducted using the Dunns test and the results in table 4.8 indicate that the completeness rate was not statistically significant different (p-value of 0.5 at 95% confidence interval) between the before covid dataset and after dataset as well as the during covid dataset and the after covid data. The before covid and during covid datasets however had values of completeness (p-value 0.02 at 95% confidence interval) that were statistically significantly different with the during covid dataset having the higher completeness rate.

Table 4.7:Krukskal Wallis test results showing that there is a difference in the rank sum scores of the different data collection periods for sensitive questions scores

Chi-squared value	Degrees of freedom	p-value
7.6	2	0.02

Table 4.8 :Dunns Test results showing the difference in sum ranks of completeness scores for sensitive questions in the three data collection periods

Group 1	Group 2	No of observations	No of observations	Statistic	p-value	p.adjusted
After Covid	Before Covid	3	3	-1.4	0.2	0.5
After Covid	During Covid	3	3	1.4	0.2	0.5
Before Covid	During Covid	3	3	2.8	0.2	0.02

Consistency Assessment

Datatype consistency

The results indicate that all values in columns of the three data collection periods had 100% consistency rate on datatype.

Consistency on narrative_1

The results indicated that the data that was collected before and after the covid_19 pandemic had a consistency rate of 100% when it comes to the data values in that narrative_1 variable. The data that was collected telephonically during the covid 19 pandemic had a consistency rate of 98.2%. The results were not assessed for statistically significant differences due to the very small sample size that did not meet the sample size requirement of the Krukskall test.

Consistency on family cause

The results indicate that 97.9% of the values in the family cause of death variable in the before covid data were consistent. The 2.0% had the same cause of death classified differently or spelt differently. The results also indicate that there is 25.6% coded as 9998 which was meant to be coded as 999, which represents respondent refused to answer the question.

The results also indicated that 98.4% of the values in the family cause variable in the during covid data was recorded consistently. Due to spellings errors values that were of the same category 1.6% were classified as different categories. For example, "Sugar diabetes" and Sugar diates". The results also indicate that 0.3% of records were coded as 99999999999 which was meant to be coded as 999.

The results indicate that 98.5% of the values in the family cause variable in the after covid data was recorded consistently. Due to spellings errors values that were of the same category

were classified as different categories. For example, " She has been witched" and "She has been switched". The results also indicate that there were 15.5% of records coded as 9996 which was meant to be coded as 999.

4.3.2 Multiple Column Analysis

Multiple column analysis was conducted on each dataset to assess the syntactic accuracy, the completeness and consistency of the relationship of values across the columns

Syntactic Accuracy Assessment

Table 4.9 :Multi column analysis syntactic accuracy results

Symptom Allocated	Before Covid accuracy % score	During Covid accuracy % score	After Covid accuracy % score
Kidney	61.5	100	100
Dialysis	0	-	-
Fever	97.6	96.7	98.6
Heart attack	0	0	50
Heart problem	26.7	66	50
Jaundice	0	-	-
Liver	42.9	-	-
Malaria	100	-	0
Pneumonia	100	100	-
Suicide	100	-	100
Non-match	94.2	78.6	92.2
Average accuracy score	**56.6**	**73.6**	**70.1**

NB:- The "-" symbol in the table is an indication that the specific variable was not selected as a response during the interview therefore no assessment was performed on that value.

Table 4.9 represents the results from the syntactic accuracy assessment. The data that was collected via telephonic interviews during covid had the highest accuracy rate for narrative_1 variable with a rate of 70.1%. It was followed by the data that was collected after the COVID-19 pandemic. The data that was collected before the COVID-19 pandemic had the lowest accuracy rate of 56.6% on the narrative_1 variable.

A Kruskal-Wallis test was conducted to assess if the difference in accuracy scores was statistically different. As indicated in table 4.10 at 95% confidence interval and p-value of 0.8 the results indicate that there was no difference in the rank sum of the accuracy scores on narrative_1 in all three datasets.

Table 4.10 : Krukskal-Wallis test results showing no difference in the ranks sum accuracy scores for narrative_1

Chi-squared value	Degrees of freedom	p-value
0.4	2	0.8

Key Analysis

Key analysis was conducted using the unique identifier variable adssid. The results indicated that there were no duplicate observations in all 3 datasets as indicated in table 4.11 with all three data collection periods having a uniqueness score of 100% on the adssid variable.

Table 4.11 : Percentage uniqueness on adssid variable per period

Variable Name	Before Covid uniqueness % score	During Covid % uniqueness score	After Covid % uniqueness score
adssid	100	100	100

Completeness Assessment

Completeness per record

Table 4.12 below presents the summary scores for the completeness of records in the three data collection periods. The results indicate that the during covid dataset had the highest mean score of 93% whilst the before covid dataset had the lowest mean completeness score of 88%. The quartile ranges of all three datasets were almost similar with all ranging above 90% in the first and third quartile. The lowest record completeness score for the before and after covid dataset was 18% which represented 2.5% of the records. The record with the

lowest record completeness score in the during covid data was 48%. The statistical results showed that the record completeness in all three datasets was statistically significantly different with a p-value of 0.00 at 95% confidence interval. Further analysis was conducted using the Dunns test and the results indicated that the completeness rate was statistically significantly different (p-value of 0.00 at 95% confidence interval) for all data collection periods as shown in table 4.13

Table 4.12 :Percentage completeness per record summary

	Minimum %	**First Quartile %**	**Median %**	**Mean %**	**Third Quartile %**	**Maximum %**
Before Covid	18	90	92	88	92	97
During Covid	48	94	94	93	94	98
After Covid	18	92	94	90	94	98

Table 4.13 :Dunns test results showing difference in rank sum scores for completeness per record for the three data collection periods

Group 1	**Group 2**	**No of observations**	**No of observations**	**Statistic**	**p-value**	**adjusted p-value**
After Covid	Before Covid	400	400	-11.8	0.0	0.0
After Covid	During Covid	400	382	5.0	0.0	0.0
Before ovid	During Covid	400	382	16.6	0.0	0.0

Completeness for dependent variables

Table 4.14 :Completeness scores for dependent variables

Variable Name	**Before Covid completeness % score**	**During Covid % completeness score**	**After Covid % completeness score**
Year of interview	100	100	100
Year of death	97.5	100	98.3
Narrative_1	61.8	95.3	70.5
Relationship other	100	100	100

Average completeness score	**89.7**	**98.8**	**92.2**

The results in table 4.14 indicate that the data that was collected via telephonic interviews during covid had the highest completeness score of 98.8% for dependent variables. It was followed by the after covid dataset which had a completeness score of 92.2%. The data that was collected before the COVID-19 pandemic had the lowest completeness score of 89.7%. The data was not normally distributed therefore a Kruskal-Wallis test was used to assess if the differences in completeness scores were statistically significantly different. The results of the test shown in table 4.15 indicated that the rank sum completeness scores of the three data collection periods were not statistically different, with a p-value of 0.7 at 95 % confidence interval.

Table 4.15 :Kruskal-Wallis Test showing no difference in rank sum scores for the completeness of dependent variables for the 3 data collection periods

Chi-squared value	Degrees of freedom	p-value
0.6	2	0.7

Consistency Assessment

The results in table 4.16 below illustrate the results of the consistency assessment that was done on three variables. As shown, the variables relationship other and alcohol type had a 100% consistent score in all three datasets. The sudden death consistency rate varied in all three datasets with the data collected after covid data having the highest consistency rate of 94%. The average consistency rate for all three datasets were almost similar with the data collected after covid having a consistency rate of 98%, 1% higher than the others. No statistical test was conducted due to the small number of records that did not meet the minimum sample size (76).

Table 4.16 :Consistency assessment scores on dependant variables

Variable Name	Before Covid consistency % score	During Covid % consistency score	After Covid % consistency score
Sudden death	92	90	94
Relationship other	100	100	100
Alcohol type	100	100	100
Average Consistency	**97**	**97**	**98**

4.3.3 Overall Data Quality Assessment Results

After performing the data quality assessments for the different variables for each dataset, the overall score for each dimension was then calculated to come up with the overall data quality score for each dataset.

Table 4.17 :Overall data quality assessment for the three data collection periods

Data Profiling Task	Before Covid % score	During Covid % score	After Covid % score
Single Column Analysis			
Syntactic Accuracy	88.1	89.1	91.8
Completeness	88.8	94.6	90.9
Consistency	98.0	97.7	98.6
Average Quality Score	**91.6**	**93.8**	**93.8**

Table 4.17 represents the overall data quality score for the three periods. The results indicate that the data that was collected via telephonic interviews during the pandemic and the data that was collected after the pandemic via face-to-face had the highest quality scores of 93.8%. The data was not normally distributed therefore a Kruskal Wallis test was used to measure if the mean rank of the quality results of the three datasets was significantly significant. At 95% confidence interval and p-value of 0. 9, there was no difference in the mean rank quality assessment results for data collected face-to-face or telephonically as shown by table 4.18.

Table 4.18 : Kruskal Wallis test results showing no difference in rank sum scores for overall data quality in the three data collection periods

Chi-squared value	Degrees of freedom	p-value
0.2	2	0.9

The data collected via face-to-face interviews was combined to get the overall score per data quality dimension for face-to-face interviews. Table 4.19 shows the results.

Table 4.19 :Overall data quality score for face-to-face versus telephonic interviews

Data Profiling Task	Face-to-face interview % Score	Telephonic Interview % Score
Syntactic Accuracy	90.0	89.0
Completeness	90.0	94.6
Consistency	98,4	97.7
Overall data quality score	**92,8**	**93.8**

Table 4.19 shows that the data quality scores for telephonic interviews were slightly higher than the overall face-to-face data quality scores.

5 Discussion

5.1 Introduction Data Profiling

This chapter presents a summary of the key findings of this research study and how the findings relate to existing literature. This chapter also covers the limitations of the study and recommendations for future research.

5.2 Summary of key findings

The aim of the study presented in this report was to assess if the quality of verbal autopsy data collected via telephonic interviews was comparable to the quality of verbal autopsy data collected via face-to-face interviews to assess the reliability of verbal autopsy data collected telephonically. Our results show that overall, the data was comparable and there was not much variation in the quality of data collected via telephonic interviews versus that collected via face-to-face interviews in all three data quality dimensions that were assessed. Where the results varied slightly, the difference was not statistically significant.

The findings in this study support the claim that telephonic interviews produce data of comparable quality to that produced by face-to-face interviews as put forward by other researchers(26) (34) (35). Similar to Nasaruddin et al. (2022) who conducted a study that measured the quality of verbal autopsy data collected telephonically by assessing the accuracy of cause of death data, our study assessed accuracy of several variables including cause of death reported by the family. Verbal autopsy data collected telephonically had an accuracy score 0.89% less than verbal autopsy data collected via face-to-face interviews, affirming that the verbal autopsy data collected telephonically was comparable and reliable (77).

Although not statistically significant, the completeness scores in the verbal autopsy data collected telephonically were higher in all instances measured for completeness compared to the completeness scores for verbal autopsy data collected through face-to-face interviews. Higher scores in completeness for data collected telephonically had not been anticipated based on literature (78) (79).What surprised further ,was the completeness score under the section that had the deceased's civil registration information. We had anticipated the verbal autopsy data collected through face-to-face interviews to have a higher score due to interviewer's proximity to the deceased's place of residence therefore having better access to civil registration documentation compared to the telephonic interviewer. The higher completeness scores on responses to sensitive questions were however in line with other studies. Although the study participants were of a different age group and different questions were asked, the findings on completeness to sensitive questions corroborate with Vogl (2013) and Jäckle et al (31) (78).

Another unanticipated result was the variation in data quality scores between verbal autopsy data collected before the covid pandemic and that collected after covid using the same method of data collection. The study incorporated the before and after data to make the data quality assessment more robust and comprehensive. The verbal autopsy data that was collected before the covid pandemic had the lowest scores in all data quality dimensions assessed when compared to the data collected after the covid pandemic. For example, in the results for multi column analysis the syntactic accuracy score in the verbal autopsy data collected after covid had an accuracy score of 70.1% while the data collected before the covid pandemic had an accuracy score just above 50%. Various factors not covered in this study could have influenced these results and further investigation is needed. We could however speculate that better training of the interviewers during the covid period which extended to the face-to-face period after covid could have influenced the results as it is the same interviewers that were used. It is

also important to note that to a greater extent the data quality scores of verbal autopsy data collected before covid performed lower than the verbal autopsy data collected through telephonic interviews.

Although the main objective of our study was focused on the comparability of data quality scores of the verbal autopsy data on specific data quality dimensions, it was noteworthy to discuss our findings on the descriptive statistics and how they relate to existing literature. The distribution and content of verbal autopsy data that was collected telephonically versus that, collected face-to-face was dependent on the specific variables. Like Chapple (1999), Sweet (2002), Sturges and Hanrahan (2010), our results showed that the depth and content of the data in the responses for deceased age, sex and medical history were comparable. Hypertension and diabetes were found to be the most reported diseases in the verbal autopsy data that was collected telephonically as well as via the face-to-face interviews. While diabetes and hypertension were not the actual causes of death but medical conditions that contributed in coming up with cause of death. It is interesting to note that these results are consistent with existing literature. Hypertension and diabetes are reported to be among the major burden of diseases globally as well as within South Africa. South African 2017 mortality data showed diabetes as the second leading cause of death and hypertension sixth(80) .The distribution on the age and sex of the respondents was also similar in all data collection periods indicating that the sample from where the data was collected was the same. Eliminating the potential of respondent's age and sex impacting on findings in this study as highlighted in other studies (81) .

Whilst the content and depth of data contained in the open narrative variable was comparable in both methods of data collection, it is important to note that contrary to previous studies in literature, the open narrative from verbal autopsy data collected via telephonic interviews had a higher count of words that was statistically significant compared to open narrative data that

was collected through face-to-face(31) (34) .The open narrative from verbal autopsy data collected telephonically had a mean word count of 40 whilst the open narrative data for verbal autopsy data collected before and after through face-to-face interviews had a mean word count of 29. Face-to-face interviews have been known to produce more content and higher word count because of the ability of the interviewer to probe additional information from the respondent as well as observe non-verbal cues (2) (26) but in in this case the data in the open narrative for the data that was collected telephonically was more descriptive and had higher word count compared to the data collected via face-to-face. Further research is needed to understand this outcome however training and expertise of the data collectors could have contributed, an aspect that was not covered in this study.

5.3 Limitations

Some factors that could have influenced our findings were not investigated in this study. They include the level of training the data collectors had prior to collecting the data, the expertise and skills of the data collectors as well as data collection procedures followed in the three data collection points (24)(82)(83) (84). The study was conducted with the assumption that all other factors that could have affected the quality of the data were constant in both the telephonic interviews and face-to-face interviews.

6 Conclusion

Telephonic interviews are regarded as a ‘’methodological compromise” and this study focused on assessing if the quality of verbal autopsy data collected via telephonic interviews during covid was comparable to the quality of verbal autopsy data collected via face-to-face interviews before and after covid to assess the reliability of the data that was collected during covid using telephonic interviews (26). The findings show that the verbal autopsy data collected through telephonic interviews during the COVID-19 pandemic was comparable to data collected through face-to-face interviews before and after the COVID-19 pandemic. Researchers at the Medical Research Council (MRC) and University of the Witwatersrand (WITS) Rural Public Health and Health Transitions Research Unit can have confidence in collecting verbal autopsy data telephonically and can consider it as an alternative method of data collection and not just rely on face-to-face interviews to collect verbal autopsy data.

References

1. Novick G. Is there a bias against telephone interviews in qualitative research? Res Nurs Health. 2008 Aug;31(4):391–8.

2. Irvine A, Drew P, Sainsbury R. Am I not answering your questions properly?' Clarification, adequacy and responsiveness in semi-structured telephone and face-to-face-interviews. Qual Res. 2013 Feb 1;13:87–106.

3. Azad A, Sernbo E, Svärd V, Holmlund previously Bergmark L, Brämberg E. Conducting In-Depth Interviews via Mobile Phone with Persons with Common Mental Disorders and Multimorbidity: The Challenges and Advantages as Experienced by Participants and Researchers. Int J Environ Res Public Health. 2021 Nov 12;18.

4. Self B. Conducting Interviews During the COVID-19 Pandemic and Beyond. Forum Qual Sozialforschung Forum Qual Soc Res [Internet]. 2021 Sep 29 [cited 2022 May 17];22(3). Available from: https://www.qualitative-Research.net/index.php/fqs/article/view/3741

5. Kahn K, Collinson MA, Gómez-Olivé FX, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt health and socio-demographic surveillance system. Int J Epidemiol. 2012 Aug;41(4):988–1001.

6. Kahn K, Tollman SM, Collinson MA, Clark SJ, Twine R, Clark BD, et al. Research into health, population and social transitions in rural South Africa: data and methods of the Agincourt Health and Demographic Surveillance System. Scand J Public Health Suppl. 2007 Aug;69:8–20.

7. Murray CJ, Lozano R, Flaxman AD, Serina P, Phillips D, Stewart A, et al. Using verbal autopsy to measure causes of death: the comparative performance of existing methods. BMC Med. 2014 Jan 9;12(1):5.

8. Thomas LM, D'Ambruoso L, Balabanova D. Verbal autopsy in health policy and systems: a literature review. BMJ Glob Health. 2018 May 1;3(2):e000639.

9. McCormick TH, Li ZR, Calvert C, Crampin AC, Kahn K, Clark SJ. Probabilistic Cause-of-death Assignment using Verbal Autopsies. J Am Stat Assoc. 2016/10/18 ed. 2016;111(515):1036–49.

10. Nojilana Beatrice, Groenewald Pam, Bradshaw Debbie, Reagon Gavin. Quality of cause of death certification at an academic hospital in Cape Town, South Africa : original article. S Afr Med J. 2009 Sep 1;99(9):648–52.

11. Statistics South Africa. Mortality and causes of death in South Africa:Findings from death notification 2017 [Internet]. 2020 Oct. Report No.: STATISTICAL RELEASE P0309.3. Available from: https://www.statssa.gov.za/publications/P03093/P030932017.pdf

12. Chen H, Hailey D, Wang N, Yu P. A review of data quality assessment methods for public health information systems. Int J Environ Res Public Health. 2014 May 14;11(5):5170–207.

13. Menéndez C, Quintó L, Castillo P, Carrilho C, Ismail MR, Lorenzoni C, et al. Limitations to current methods to estimate cause of death: a validation study of a verbal autopsy model. Gates Open Res. 2020;4:55.

14. Yokobori Y, Obara H, Sugiura Y, Kitamura T. Gaps in the civil registration and vital statistics systems of low- and middle-income countries and the health sector's role in improving the situation. Glob Health Med. 2021 Aug 31;3(4):243–5.

15. Mikkelsen L, Phillips DE, AbouZahr C, Setel PW, de Savigny D, Lozano R, et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. The Lancet. 2015 Oct 3;386(10001):1395–406.

16. Joubert J, Rao C, Bradshaw D, Vos T, Lopez AD. Evaluating the Quality of National Mortality Statistics from Civil Registration in South Africa, 1997–2007. PLOS ONE. 2013 May;8(5):1–11.

17. Danso SO, Manu A, Fenty J, Amanga-Etego S, Avan BI, Newton S, et al. Population cause of death estimation using verbal autopsy methods in large-scale field trials of maternal and child health: lessons learned from a 20-year research collaboration in Central Ghana. Emerg Themes Epidemiol. 2023 Feb 16;20(1):1.

18. Byass P, Hussain-Alkhateeb L, D'Ambruoso L, Clark S, Davies J, Fottrell E, et al. An integrated approach to processing WHO-2016 verbal autopsy data: the InterVA-5 model. BMC Med. 2019 May 30;17(1):102.

19. Byass P, Chandramohan D, Clark SJ, D'Ambruoso L, Fottrell E, Graham WJ, et al. Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. Glob Health Action. 2012 Sep 3;5:1–8.

20. Nichols EK, Byass P, Chandramohan D, Clark SJ, Flaxman AD, Jakob R, et al. The WHO 2016 verbal autopsy instrument: An international standard suitable for automated analysis by InterVA, InSilicoVA, and Tariff 2.0. PLoS Med. 2018 Jan;15(1):e1002486.

21. ISO/IEC. Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model. (ISO/IEC 25012, 2008).

22. Laranjeiro N, Soydemir SN, Bernardino J. A Survey on Data Quality: Classifying Poor Data. In: 2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (PRDC). 2015. p. 179–88.

23. Agincourt - MRC/Wits Rural Public Health and Health Transitions Research Unit. Covid-19 Response. 2022. Extending the update to include a Covid-19 Module and Capitalisation on ability to conduct telephonic surveys. Available from: https://www.agincourt.co.za/?page_id=3824

24. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health. 2005 May;27(3):281–91.

25. Farooq M, Villiers C. Telephonic Qualitative Research Interviews: When to consider them and how to do them. Meditari Account Res. 2017 Jun 5;25:291–316.

26. Irvine A. Duration, Dominance and Depth in Telephone and Face-to-Face Interviews: A Comparative Exploration. Int J Qual Methods. 2010 Nov 30;10.

27. Siregar KN, Kurniawan R, Nuridzin DZ, BaharuddinNur RJ, Retnowati, Handayani Y, et al. Strengthening causes of death identification through community-based verbal autopsy during the COVID-19 pandemic. BMC Public Health. 2022 Aug 23;22(1):1607.

28. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep Morb Mortal Wkly Rep Recomm Rep. 2001 Jul 27;50(RR-13):1–35; quiz CE1-7.

29. Gill P, Stewart K, Treasure E, Chadwick B. Methods of data collection in qualitative research: interviews and focus groups. Br Dent J. 2008 Mar 1;204(6):291–5.

30. Safdar N, Abbo LM, Knobloch MJ, Seo SK. Research Methods in Healthcare Epidemiology: Survey and Qualitative Research. Infect Control Hosp Epidemiol. 2016/08/12 ed. 2016 Nov;37(11):1272–7.

31. Vogl S. Telephone Versus Face-to-Face Interviews: Mode Effect on Semistructured Interviews with Children. Sociol Methodol. 2013 Aug 1;43(1):133–77.

32. Cachia M, Millward L. The Telephone Medium and Semi-Structured Interviews: A Complementary Fit. Qual Res Organ Manag Int J. 2011 Jan 1;6(3):265–77.

33. Lechuga V. Exploring culture from a distance: The utility of telephone interviews in qualitative research. Int J Qual Stud Educ. 2012 May 1;25:251–68.

34. Sweet L. Telephone Interviewing: Is It Compatible with Interpretive Phenomenological Research? Contemp Nurse. 2002 Feb;12(1):58–63.

35. Sturges JE, Hanrahan KJ. Comparing Telephone and Face-to-Face Qualitative Interviewing: a Research Note. Qual Res. 2004 Apr 1;4(1):107–18.

36. Jha P. Reliable direct measurement of causes of death in low- and middle-income countries. BMC Med. 2014 Feb 4;12(1):19.

37. Bradshaw, Debbie and Pillay-Van Wyk, Victoria and Laubscher, Ria and Nojilana, Beatrice and Groenewald, Pam and Nannan, Nadine and Metcalf, Carol. Cause of death statistics for South Africa: Challenges and possibilities for improvement. South Afr MRC Burd Dis Res Unit. 2010;

38. Johnson SC, Cunningham M, Dippenaar IN, Sharara F, Wool EE, Agesa KM, et al. Public health utility of cause of death data: applying empirical algorithms to improve data quality. BMC Med Inform Decis Mak. 2021 Jun 2;21(1):175.

39. Hart JD, Sorchik R, Bo KS, Chowdhury HR, Gamage S, Joshi R, et al. Improving medical certification of cause of death: effective strategies and approaches based on experiences from the Data for Health Initiative. BMC Med. 2020 Mar 9;18(1):74.

40. Bradshaw D, Joubert JD, Maqungo M, Nannan N, Funani N, Laubscher R, Cheyip M, Zinyakatira N, Awotiwon O, Nojilana B,, Bezuidenhout F, Martin L, Dempers J, Kahn K, Price J, Lombard C, Morof D, Nichols E, Rao C, Groenewald P. South African National

Cause-of-Death Validation Project: Methodology and Description of a National Sample of Verbal Autopsies [Internet]. Cape Town: South African Medical Research Council; 2020. Available from: http://samrc.ac.za.dedi6.cpt3.host-h.net/sites/default/files/attachments/2022-08/NationalCause-of-deathValidationReport.pdf

41. Nadine Nannan, Debbie Bradshaw, Ria Laubscher, Mikateko Mazinu, Tshifhiwa Nkwenika, Ian Neethling,, Oluwatoyin Awotiwon, Pam Groenewald. Rapid assessment of cause-of-death data collection and public health use in South Africa. South African Medical Research Council; 2023. (South African Medical Research Council).

42. Joubert J, Bradshaw D, Kabudula C, Rao C, Kahn K, Mee P, et al. Record-linkage comparison of verbal autopsy and routine civil registration death certification in rural north-east South Africa: 2006-09. Int J Epidemiol. 2014 Dec;43(6):1945–58.

43. Burger EH, Groenewald P, Bradshaw D, Ward AM, Yudkin PL, Volmink J. Validation study of cause of death statistics in Cape Town, South Africa, found poor agreement. J Clin Epidemiol. 2012;65(3):309–16.

44. Yudkin PL, Burger EH, Bradshaw D, Groenewald P, Ward AM, Volmink J. Deaths caused by HIV disease under-reported in South Africa. AIDS Lond Engl. 2009 Jul 31;23(12):1600–2.

45. Mathers CD, Fat DM, Inoue M, Rao C, Lopez AD. Counting the dead and what they died from: an assessment of the global status of cause of death data. Vol. 83, Bulletin of the World Health Organization. World Health Organization; 2005. p. 171–7.

46. Collinson MA, Mudzana T, Mutevedzi T, Kahn K, Maimela E, Gómez-Olivé FX, et al. Cohort Profile: South African Population Research Infrastructure Network (SAPRIN). Int J Epidemiol. 2022 Aug 1;51(4):e206–16.

47. Fottrell E, Byass P. Verbal Autopsy: Methods in Transition. Epidemiol Rev. 2010 Mar;32(1):38–55.

48. King C, Zamawe C, Banda M, Bar-Zeev N, Beard J, Bird J, et al. The quality and diagnostic value of open narratives in verbal autopsy: a mixed-methods analysis of partnered interviews from Malawi. BMC Med Res Methodol. 2016 Feb 1;16(1):13.

49. Serina P, Riley I, Hernandez B, Flaxman AD, Praveen D, Tallo V, et al. What is the optimal recall period for verbal autopsies? Validation study based on repeat interviews in three populations. Popul Health Metr. 2016;14:40.

50. Jäckle A, Roberts C, Lynn P. Assessing the Effect of Data Collection Mode on Measurement. Int Stat Rev. 2010;78(1):3–20.

51. Rogers TF. Interviews by Telephone and in Person: Quality of Responses and Field Performance*. Public Opin Q. 1976 Jan;40(1):51–65.

52. Midanik L, Greenfield T, Rogers J. Reports of Alcohol-Related Harm: Telephone Versus Face-to-Face interviews. J Stud Alcohol. 2001 Feb 1;62:74–8.

53. Aquilino WS. Telephone Versus Face-to-Face Interviewing for Household Drug Use Surveys. Int J Addict. 1991 Jan 1;27(1):71–91.

54. Pipino L, Lee Y, Wang R. Data Quality Assessment. Commun ACM. 2003 Jul 20;45.

55. Zeleke AA, Naziyok T, Fritz F, Röhrig R. Data Quality and Cost-Effectiveness Analyses of Electronic and Paper-Based Interviewer-Administered Public Health Surveys: Protocol for a Systematic Review. JMIR Res Protoc. 2019 Jan 30;8(1):e10678.

56. Sidi F, Hassany Shariat Panahy P, Affendey L, A. Jabar M, Ibrahim H, Mustapha A. Data quality: A survey of data quality dimensions. 2013 Aug 27;

57. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS Wash DC. 2016;4(1):1244.

58. Andrew Black P vaNederpelt. Dictionary of Dimensions of Data Quality. Dama Nl Foundation; 2020.

59. Stausberg J, Nasseh D, Nonnemacher M. Measuring Data Quality: A Review of the Literature between 2005 and 2013. Stud Health Technol Inform. 2015;210:712–6.

60. Bian J, Lyu T, Loiacono A, Viramontes TM, Lipori G, Guo Y, et al. Assessing the practice of data quality evaluation in a national clinical data research network through a systematic scoping review in the era of real-world data. J Am Med Inform Assoc. 2020 Dec 9;27(12):1999–2010.

61. AskHam NA, Denise Cook, Martin Doyle, Helen Fereday, Mike Gibson, Ulrich Landbeck, et al. The Six Primary Dimensions for Data Quality Assessment. 2013.

62. Kodra Y, Posada de la Paz M, Coi A, Santoro M, Bianchi F, Ahmed F, et al. Data Quality in Rare Diseases Registries. Adv Exp Med Biol. 2017;1031:149–64.

63. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. J Manag Inf Syst. 1996 Mar 1;12(4):5–33.

64. Organization WH. Verbal Autopsy Standards : Ascertaining and Attributing Cause of Death. World Health Organization; 2007. p. 112 p.

65. Ganapathy SS, Yi Yi K, Omar MA, Anuar MFM, Jeevananthan C, Rao C. Validation of verbal autopsy: determination of cause of deaths in Malaysia 2013. BMC Public Health. 2017 Aug 11;17(1):653.

66. Olson JE. Data Quality: The Accuracy Dimension. Morgan Kaufmann; 2003. (The Morgan Kaufmann Series in Data Management Systems).

67. Woodall P, Oberhofer M, Borek A. A Classification of Data Quality Assessment and Improvement Methods. Int J Inf Qual. 2014;3:298–321.

68. Abedjan Z, Golab L, Naumann F. Profiling Relational Data: A Survey. VLDB J. 2015 Aug;24(4):557–81.

69. Miyoshi N, Marques J, Alves D, Azevedo-Marques P. eHealth-interop: an Interoperability Platform developed for Mental Care Integration (Preprint). JMIR Ment Health. 2018 Feb 15;5.

70. Hassenstein MJ, Vanella P. Data Quality—Concepts and Problems. Encyclopedia. 2022;

71. Abedjan Z, Golab L, Naumann F. Data Profiling: A Tutorial. In: Proceedings of the 2017 ACM International Conference on Management of Data [Internet]. New York, NY, USA: Association for Computing Machinery; 2017. p. 1747–51. (SIGMOD '17). Available from: https://doi.org/10.1145/3035918.3054772

72. T. F. Kusumasari, Fitria. Data profiling for data quality improvement with OpenRefine. In: 2016 International Conference on Information Technology Systems and Innovation (ICITSI). 2016. p. 1–6.

73. PAUL O, SCHATZ M. On Sudden Death. Circulation. 1971;43(1):7–10.

74. Ostertagova E, Ostertag O, Kováč J. Methodology and Application of the Kruskal-Wallis Test. Appl Mech Mater. 2014 Aug;611:115–20.

75. Dinno A. Nonparametric Pairwise Multiple Comparisons in Independent Groups using Dunn's Test. Stata J. 2015;15(1):292–300.

76. Kruskal-Wallis Test. In: The Concise Encyclopedia of Statistics [Internet]. New York, NY: Springer New York; 2008. p. 288–90. Available from: https://doi.org/10.1007/978-0-387-32833-1_216

77. Nasaruddin NH, Ganapathy SS, Awaluddin SM, Anuar MFM, Binti Alias N, Mang CY, et al. Conducting verbal autopsy by telephone interview during the pandemic to support mortality surveillance: a feasibility study in Malaysia. West Pac Surveill Response J WPSAR. 2022 Jun;13(2):1–7.

78. Jäckle A, Roberts C, Lynn P. Telephone versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes. 2006 Jan;

79. Rahman RBA. Comparison of telephone and in-person interviews for data collection in qualitative human research. 2015;

80. Kamerman P. Underdiagnosis of hypertension and diabetes mellitus in South Africa. South Afr Med J Suid-Afr Tydskr Vir Geneeskd. 2022 Feb 2;112(1):13519.

81. HERZOG AR, RODGERS WL, KULKA RA. Interviewing Older Adults: A Comparison of Telephone and Face-to-Face Modalities.Public Opin Q. 1983 Jan;47(3):405-1

82. Horsfall M, Eikelenboom M, Draisma S, Smit JH. The Effect of Rapport on Data Quality in Face-to-Face Interviews: Beneficial or Detrimental? Int J Environ Res Public Health. 2021 Oct 15;18(20).

83. Olson K, Smyth JD, Ganshert A. The Effects of Respondent and Question Characteristics on Respondent Answering Behaviors in Telephone Interviews. J Surv Stat Methodol. 2018 Apr;7(2):275–308.

84. Oltmann S. Qualitative Interviews: A Methodological Discussion of the Interviewer and Respondent Contexts. Forum Qual Sozialforschung Forum Qual Soc Res [Internet]. 2016 May;17(2). Available from: https://www.qualitative-research.net/index.php/fqs/article/view/2551

Appendix

Table A.1: List of variables used in the data quality assessment

Data Category	**Data Variables**
Information of respondent	
Sex	resp_sex
Age	resp_dob
Respondent Relationship	Relationship_to_deceased Relationship_other
Living arrangements	LiveWith
Date of interview	Interview_date
Information of the deceased	
Agincourt ADSSID of deceased individual	Adssid
Sex	Sex
Date of birth	Dob
Date of death	Dod
Year of death	yearDeath
Deceased age at time of death	Age
Deceased marital status at time of death	Marital
Where deceased died	where_died
Open Narrative	
Was a death form issued	death_form
Narrative of symptoms experienced by deceased prior to death	Narrative Narrative_1
Cause of death declared by respondent	FamCause1
Medical History Associated with final illness	
Diagnosis of TB	Tuber
HIV test results	hiv_pos
AIDS diagnosis by health professional	Hiv_aids
Positive malaria test result	Malaria

Negative malaria test result	Malarneg
Dengue fever	dengue_fever
Blood pressure	Hypert
Heart disease	Heart_dis
Diabetes	Diabetes
Asthma	Asthma
Epilepsy	Epilepsy
Cancer	Cancer
Kidney disease	Kidney_dis
Sickle cell disease	sickle_cell
Liver disease	Liver_dis
Final Symptoms : Did the deceased experience these symptoms as the final symptoms that led to death	
Fever	Fever
Night sweats	Night_sw
Extreme fatigue	extreme_fatigue
Cough	cough
Difficulty breathing	breath
Chest pain	chest_in
Diarrhoea	diarr
Vomiting	vomiting
Abdominal pain	abdom
Headache	headache
Mental confusion	Men_con
Convulsions	convul
Lifestyle : Questions asked to have a a background of the deceased lifestyle	
Alcohol consumption	alcohol
Alcohol type drink	alcohol_type_drunk
Smoker	smoking
Physically active	active
Struggle to be active	struggle
Sedentary lifestyle	sedet
Mobility	immob
Duration of illness	
Length of illness before death	Ill_duration
Illness duration in terms of Days/Weeks/Months	ill_duration_num_2

sudden death	Sudden
Civil Registration	
Was the death certificate issued from Civil registration	death_form
Can I see the death certificate?	dod_cert
If you have death form, but unavailable, where is it?	DeathFormLocation
Death certificate, death registration number	death_cert_no
Immediate cause of death on death certificate	immediate_cause
Cause of death recorded on death certificate/registration form:	RegCause