

https://dspace.org/

Assessing cervical precancer treatment rate among women utilising the Johannesburg public sector through record linkage of laboratory data

Juggernath, Amilcar M.

2022

https://hdl.handle.net/10539/35204

Downloaded from a DSpace repository. https://dspace.org/



UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

Faculty of Health Sciences, School of Public Health

Department of Community Health

Assessing cervical precancer treatment rate among women utilising the Johannesburg public sector through record linkage of laboratory data

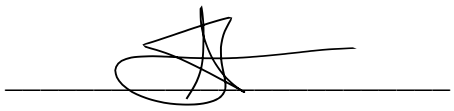
Candidate: Amilcar M. Juggernath
Student no.: 358266
Degree: MMed (Public Health Medicine)
Supervisors: Dr Mary Kawonga, Prof. Pamela Michelow, Dr Chodziwadziwa Kabudula

Johannesburg, August 2022

Declaration

I, Amilcar Munmohan Juggernath, student number 358266, declare that this Research Report is my own, unaided work. It is being submitted as a partial fulfilment of the requirements for the degree of Master of Medicine (Public Health Medicine) in the School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg

This work has not been submitted before for any degree or examination at this or any other University.

Signature of Candidate: 

Name of Candidate: Amilcar Munmohan Juggernath

Signed on the 6th day of August 2022 in Johannesburg

Dedication

This work is dedicated to..

My primary supervisor, Dr Mary Kawonga, for immense patience and mentorship as I journeyed through the research process and the intricate world of public health medicine.

My co-supervisors, Prof. Pamela Michelow and Dr Chodziwadziwa Kabudula, for expanding my public health skills far beyond my comfort zones.

Prof. Leena Thomas for swooping in when needed the most and redefining the concept of "turn-around-time."

My parents, Usha and Spider Juggernath; the Council of Iruy and those who wish to be on it; and the Ramkissoon elders and Gopi twins for the constant support, food, love, nagging, tea, and a place to just relax. And my nephew, Master Elijah Jehhan Juggernath, for making me smile on the dullest of days.

My favourites: Mr Aveel Lalchund, Dr Yashodani Pillay, Ms Serisha Pillay, Ms Dhivanya Moodley, Ms Arushani Govender, Dr Mazvita Mberi, Dr Jesne Kistan, Goten and Broly for putting up with my rants; giving me perspective, ideas, love and space to make sense of things; the cute facial expressions and high-energy barks and greetings; and for just being there always.

Dr Elvira Singh for being the ideal role model and working tirelessly as a pioneer of high quality, usable cancer data in our country, while still having a contagious laugh that I will never forget.

And to all the women who have been affected by the scourge of cervical cancer..

Acknowledgements

I would like to thank:

Prof. Leena Thomas for the assistance through the report-writing process.

Prof. Yasmin Adam (Chris Hani-Baragwanath Hospital) for helping to map out the cervical cancer screening cascade in Johannesburg.

The CDW staff at the NHLS for assisting with data extraction.

Ms Nikita Ramkissoon for the text and language review.

Ms Innocentia Vukeya for her guidance through the submission process.

Abstract

Background

Screening is important to prevent cervical cancer, but is only effective if women who screen positive for precancer undergo precancer treatment. Routine health information systems in South Africa currently do not monitor precancer treatment rates.

Methods

A retrospective cohort of 67 208 women screened in Johannesburg, 2017, used record linkage of laboratory records – cytology with high-grade squamous intraepithelial lesions (HSIL), and histology. The HSIL prevalence; and follow-up, confirmation and precancer treatment rates were calculated; as were the time intervals between these. Variables were compared across age, HIV status, screening facility type and subdistrict through logistic regression, Kruskal-Wallis and Wilcoxon rank-sum tests.

Results

HSIL prevalence was 6.2%. Follow-up occurred in 26.6% of the 4 182 women with HSIL on cytology over a median of 131 days. Treatment rate was 16.1%, performed in a median of 151 days. Screening at hospitals, CHCs and in subdistrict B had the highest follow-up and treatment rates.

Conclusions

Precancer treatment rates are far below the WHO's elimination target. A simpler screening cascade; benchmarking against high-performing subdistricts; and improved information systems may improve precancer treatment rates to reduce cervical cancer incidence.

Table of Contents

Declaration ii
Dedication iii
Acknowledgements iv
Abstract v
Table of Contents vi
List of Figures vii
List of Tables vii
List of Acronyms viii
1. Background 1
1.1. Introduction 1
1.2. Context 1
1.3. Literature review 4
1.4. Problem Statement 8
1.5. Aim and Objectives 9
2. Methods 9
2.1. Study design 9
2.2. Study setting 9
2.3. Study population and sampling 10
2.4. Data extraction and cleaning 11
2.5. Record Linkage 12
2.6. Data management 15
2.7. Data analysis 18
2.8. Ethical considerations 19
3. Results 19
3.1. Participant characteristics 19
3.2. HSIL prevalence and the follow-up, confirmatory test and treatment rates 20
3.3. Time intervals between screening tests, confirmatory tests and treatment 22
3.4. Factors associated with HSIL prevalence and the follow-up, confirmatory test and treatment rates 22
3.5. Factors associated with the time intervals between the screening test, confirmatory test and treatment 26
4. Discussion 27
5. Limitations 34
6. Conclusion and Recommendations 36
6.1. Conclusion 36
6.2. Recommendations 36
7. References 39
8. Appendices 45

List of Figures

Figure 2.1: Summary of developing the final dataset from the cytology and histology datasets
…….. Page 16
Figure 3.1: Summary of the number of women along each step of the cervical cancer screening
…….. Page 21
Figure 8.1: Patient flow through the cervical cancer screening cascade
…….. Page 50

List of Tables

Table 3.1: Summary of the women who had cervical smears, who screened positive with HSIL and who had a first follow-up procedure
…….. Page 20
Table 3.2: Summary of the time intervals between screening tests, confirmatory tests and treatment
…….. Page 23
Table 3.3: Factors associated the follow-up, confirmatory test and treatment rates
…….. Page 25
Table 3.4: Summary of factors associated with time intervals between the screening test, confirmatory test and treatment
…….. Page 26
Table 8.1: Summary of the variables used for components of data processing
…….. Page 51
Table 8.2: Description of summary variables
…….. Page 52
Table 8.3: Description of variables in the final dataset
…….. Page 53

List of Acronyms

aOR	Adjusted odds ratio
ASDR	Age-standardised death rate
CDC	Centers for Disease Control and Prevention
CDW	Corporate Data Warehouse
CHC	Community healthcare centre
CI	Confidence interval
CIN	Cervical intraepithelial neoplasia
cOR	Crude odds ratio
DHB	District Health Barometer
DHIS	District Health Information Software
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HSIL	High-grade intraepithelial lesion
ID	Identity
IQR	Inter-quartile range
LEEP	Loop electrosurgical excision procedure
LLETZ	Large loop excision of the transformation-zone
LSIL	Low-grade intraepithelial lesion
NHLS	National Health Laboratory Service
NIDS	National indicator data set
NILM	Negative for intraepithelial lesion or malignancy
PHC	Primary healthcare clinic
OR	Odds ratio
VIA	Visual inspection with acetic acid
WHO	World Health Organization

1. Background

1.1. Introduction

Cervical cancer is the most common cause of cancer deaths among women in South Africa.[1] Despite a nationwide screening strategy with increasing coverage,[2] the gaps in the screening cascade contribute towards not meeting the goal of reducing the incidence of cervical cancer.[3] Routine monitoring of the screening cascade does not exist which presents an opportunity to use sources outside of the routine health information system to assist in monitoring this programme.

1.2. Context

The World Health Organization (WHO) estimated that there were 10 702 new cases of cervical cancer in South Africa in 2020.[1] While it is the fourth most common cancer among women in the world resulting in an age-standardised death-rate (ASDR) of 7.3 per 100 000 women,[4] it is far more common in South Africa and has worse outcomes with an ASDR of 19.6 per 100 000 women.[1] Persistent human papillomavirus (HPV) infection is the greatest risk factor for developing cervical cancer.[5] HIV co-infection, which is highly prevalent in South Africa,[6] substantially increases this risk .[5,6]

To address the burden of cervical cancer, South Africa revised its Cervical Cancer Prevention and Control Policy[7] in 2017. The policy guides primary prevention strategies such as HPV vaccination, and secondary prevention through screening for early detection and treatment of precancerous cervical lesions. Screening is recommended through three cervical smears, such as a Pap smear, performed at ten-year intervals beginning at age thirty. Women living with HIV, who are at higher risk of developing cervical cancer, are recommended to undergo screening every three years from the time of their HIV diagnosis.[7] The national strategy aligns with the WHO's Comprehensive Cervical Cancer Control[5] guidance document and with the WHO's Global Strategy to Accelerate the Elimination of Cervical Cancer as a Public Health Problem.[8] The elimination strategy includes the "90-70-90" global targets of achieving 90% coverage of girls fully vaccinated against HPV; 70% of eligible women screened for cervical cancer; and a 90% treatment rate of women who screen positive for a precancerous lesion and those who are diagnosed with cervical cancer by 2030.[8]

A reduction in the incidence of cervical cancer through screening may only be achieved if a high proportion of at-risk women are screened and a high proportion of those who screen positive for precancerous lesions are treated.[6,9] Precancerous cervical lesions such as high-grade squamous intraepithelial lesions (HSIL) have distinct cytological changes that may, over a few years, develop into cervical cancer.[5] The national policy recommends a treatment plan for women who have a precancerous lesion. This may entail two steps of a confirmatory diagnostic test and then treatment only if a precancerous lesion is confirmed; or a one-step process of treatment following the screening cervical smear that suggested a precancerous lesion.[7]

In Johannesburg, the health district in the Gauteng province in South Africa where this study was conducted, studies have described a multi-step cascade of visits to multiple healthcare facilities to complete the cervical cancer screening process[10–14] as summarised in Appendix D. A woman begins by having a screening cervical smear performed at a primary-level public sector healthcare facility.[10] Patient information is captured onto a cytology request form, sent with the specimen to an off-site laboratory within the South African National Health Laboratory Services (NHLS) network, and then captured onto the NHLS electronic data system.[15,16] The specimen is then sent to one of the central NHLS cytology laboratories for analysis. The woman returns to the healthcare facility to receive the results (usually four to six weeks after the procedure)[11,14] A woman whose cytology results report HSIL (among other precancerous lesions or suspected cancer) is referred to a colposcopy clinic at a hospital[7,10,17] within the Gauteng province. The appointment for this visit may be made telephonically or may require the woman to present herself to the hospital to schedule the appointment,[10] which is usually booked for two to four months later.[13]

At the colposcopy clinic, the decision to continue with a one- or two-step management process is based on the clinician's findings or the protocol of the facility.[12] The one-step process involves performing a large loop excision of the transformation-zone (LLETZ) at the initial visit.[12,13] This provides both confirmation of the diagnosis of a precancerous lesion and precancer treatment at the same time.[7,12,13] The two-step

process involves a confirmatory biopsy done on the first visit and treatment with a LLETZ at the second visit only if the precancerous lesion is confirmed by the histology result of the biopsy.[7,11] All procedures produce a specimen that is sent for histological analysis. For women following the two-step process, the first set of results that she returns to receive (four to six weeks after the procedure)[11] will confirm the diagnosis and the second would describe the adequacy of the treatment. For those following the one-step process, both these elements are described in the first set of results. A similar procedure of data capturing onto a histology request form, and then into the NHLS electronic data system, is performed.[15]

The NHLS performs cytology and histology analysis for all public sector healthcare facilities in South Africa.[15,16,18] The data collected during procedures includes demographic and clinical details, and details about the procedure and specimen.[15,18] The data for a single procedure is saved as a single record. If a patient has another procedure, a new record is created and the demographic details of the patient is not linked between the records of the two procedures.[15] Therefore, the cytology and histology records of a woman who has gone through the cervical cancer screening cascade exist on the NHLS system separately and are not identified as belonging to the same woman. NHLS data is stored at their Central Data Warehouse (CDW).[14,15,18] Though the data is exclusively pathology-related data and has challenges of not being linked, it may be a rich source of data for research and for programme monitoring and evaluation.[14]

Routine monitoring of the cervical cancer screening programme is currently done by measuring only the cervical smear coverage[19] (the first step of the screening process) and the incidence of cancer[20] (the impact of the screening programme). There are no indicators in the South African national indicator data set (NIDS)[19] that allow for the rest of the screening cascade to be monitored. As a priority cancer affecting women in South Africa,[21] it is important to track indicators that measure how effective the programme is to reduce cervical cancer incidence. These indicators include the smear positivity rate, treatment rate and the waiting periods across the cascade.[7,22]

1.3. Literature review

While indicators along the cervical cancer screening cascade are not routinely measured or monitored, several indicators have been measured in South African research. These indicators include the prevalence of precancerous lesions among women who have had cervical smears, the treatment rate of those with precancerous lesions, and the time periods between the different steps along the screening-to-treatment cascade. Among a list of other indicators, these have been recommended in the national policy,[7] in the national implementation plan[10] and by the WHO.[22]

Studies in South Africa have found mixed results of HSIL prevalence. A national, multi-centre study of 20 000 women in the early 2000s showed an HSIL prevalence of 1.88%.[23] Other smaller studies around the country have since shown a similar[24] or higher[25] prevalence of HSIL. As HIV is a risk factor for developing precancerous lesions,[5] studies focused on samples of women living with HIV[26,27] have found a much greater HSIL prevalence, including a study in Johannesburg which found an HSIL prevalence of 18%.[27]

The precancer treatment rate measures the proportion of women who received treatment among those who had a precancerous lesion detected. In the Overberg district in another South African province, the Western Cape, a study found that 51.3% of women with precancerous lesions had attended their follow-up appointment;[25] in the eThekwini district in the province of KwaZulu Natal, 63.5% received treatment;[26] but in Johannesburg, the treatment rate has been found to be as low as 25.6%.[11] The differences in the follow-up times and the treatment eligibility criteria make it difficult to confidently compare precancer treatment rates between studies. Follow-up times have sometimes not been explicitly documented[25] or have extended to three[11] or even five[26] years after having the screening cervical smear. Treatment eligibility criteria has been influenced by the change in South African guidance between the 2000 national guideline[28] and the 2017 policy[7]. Each document recommended different precancerous lesions for follow-up and the different follow-up plans included repeat cervical smears, confirmatory tests or treatment. Barriers to precancer treatment include transport difficulties due to the multi-step screening-to-treatment cascade[24,29]

and the long time to treatment from the screening cervical smear to the precancer treatment, resulting in loss to follow-up before precancer treatment.[24]

The time to treatment is an important consideration for the success of the cervical cancer screening programme. While no South African standard exists for the time to treatment, the Centers for Disease Control and Prevention (CDC) of the United States of America supports a two-step management process with the standard of three months from the time of screening to the confirmatory test, and then a further two months from confirmatory test to precancer treatment.[30] A recent study in Johannesburg elected to use a timeframe of six to seven months as an appropriate clinical standard of care.[11] In Overberg, a study showed a mean time to treatment of 4.5 months;[25] similarly, in Johannesburg, a median time of 4.4 months has been measured.[31] In eThekwini, however, the median time to treatment among women screened at an HIV clinic attached to a hospital was only 1.3 months.[26] The reduced time to treatment seen in the eThekwini study may be due to the screening test and colposcopy both being performed at the same site and the service being integrated into routine HIV care, making it easier for women to follow up.[32] Hospitals are more likely to have easier access to a laboratory on-site and quicker access to one of the 12 centralised NHLS cytology laboratories[33] therefore reducing the turnaround times of specimen processing and therefore the time at which a woman may follow up for results and subsequent management.

Studies from other countries have also looked at indicators within their cervical cancer screening cascades. A cross-sectional study in Kenya found an HSIL prevalence of 3.96%. with a 3.1 times greater odds of HSIL for women with HIV.[34] A Chinese study found a 4.15% smear abnormality rate, a 70.54% follow-up rate for a colposcopy examination and an 82.40% follow-up rate for their subsequent confirmatory procedure. Loss to follow-up was seen mostly in the elderly, and mild-aged adults who were noted as likely being economic migrants.[35] In Central America, an interventional study across four countries showed that 17.3% of women who were screened had a positive HPV result, 73.7% of these followed up for a confirmatory procedure, and 72.0% of those who required treatment had received it.[36]

Concerns have been expressed about the quality of data available for programme monitoring. For instance, paper-based data that is routinely collected at healthcare facilities may be incomplete and this type of data does not allow for easy aggregation and disaggregation, longitudinal collection, and linkage of patients across different healthcare facilities.[37,38] Many studies of the cervical cancer screening cascade in South Africa have made use of paper-based data sources. This includes primary data through data collection tools used at the point of care,[23,27] and secondary data through extracting data from patient files[17,25] registers at the screening[24,31] and colposcopy clinics,[31] and pathology reports generated by the NHLS.[17,24] Electronic medical records have been used less commonly as a source of secondary data[11,25,26] and has included the exclusive use of NHLS data.[39] Electronic patient records may overcome the challenges associated with paper-based data, but the easy linkage of records over time and across facilities is difficult in the absence of a unique patient identifier, as is often the case in South Africa.[40] Nevertheless, electronic records may have the potential to track indicators along the cervical cancer screening cascade.

One way of overcoming the challenge of unlinked electronic healthcare data is through record linkage. This technique utilises the variables with similar information in different electronic datasets to link records that belong to the same person. Deterministic linkage performs exact matching of records and is most useful when both datasets have a common unique identifier. Probabilistic linkage matches records based on the similarity of information from multiple variables. These similarities generate a composite score that determines whether two records are likely to be a match or if they are likely to be the records of different people. Probabilistic linkage is more suited when unique identifiers are absent and when transcription of variables may be of inconsistent quality.[41] The variables that are often used for the probabilistic record linkage process includes personal identifiers such as the patient's name, sex, date of birth and address.[42]

The public health opportunities of the record linkage technique include its use for broader disease surveillance through using data that is routinely collected from and across different sectors. Using linked data from healthcare facility diagnoses, disease

registries, and death records, it may be possible to determine disease distribution and track healthcare utilisation; while administrative data around socio-economic factors may be used towards understanding the determinants of disease.[42]

In Brazil, the national electronic information system for the cervical cancer screening programme lacks unique identifiers. Research has been conducted to combine the cytology dataset of over 2.4 million records with the histology dataset of 7 477 records through deterministic and probabilistic record linkage. The researchers found that some data elements were of poor quality. The name and surname had additional characters and words and required extensive cleaning and standardisation. Names were compared on a phonetic basis which may have been less accurate because of a large number of Portuguese rather than English names. Addresses were of too poor a quality to use in the linkage process. Nevertheless, high sensitivity and specificity had been achieved at 94.86% and 99.50% respectively.[43]

In South Africa, record linkage has been used in public health research. A pilot study of record linkage at the Agincourt Health and Demographic Surveillance System in the north-eastern part of the country assessed record linkage of records of attendees of healthcare facilities and population surveillance records.[44] The research team found that the highest sensitivity and positive predictive value of the linkage was achieved by using both deterministic and probabilistic linkage on the personal identifiers that were extended to also include the patient's telephone number and a name of a household member; using multiple string comparison techniques such as comparing letters and their order, or comparing strings phonetically; and performing a clerical review on a sample of linked records to adjust the linkage algorithm. The study recognised that monitoring the equitable access of healthcare services across the communities was a potential use for this linked dataset.[44]

A 2018 study used probabilistic record linkage techniques on NHLS data.[18] A combination of identifiers were used to compare different records within an HIV database over time in order to link multiple test records to a single patient. These identifiers included the name, surname, sex, date of birth, facility and province of the

patients. The precision of the record linkage had a sensitivity of 91% and a positive predictive value of 91%. Using the linked data, the study was able to track the various laboratory test results of patients. This was used to monitor the follow-up of people living with HIV in South Africa over six years. The presence of results was used as an indicator that a person followed up for their HIV care at a healthcare facility. The study was able to determine the rate of retention in HIV care over time, the movement of patients across the country, and was able to perform analyses on demographic and disease-specific factors from this database. The researchers recognised that their inferences from monitoring laboratory results may have underestimated retention in care. Any patient who did not undergo laboratory testing did not appear on the NHLS database and so were not considered to have been followed up, irrespective of whether or not appropriate clinical care was received. During the study period, the laboratory monitoring protocols of HIV care had changed, making an assessment of adequate care, through having the correct tests at the correct time intervals, difficult to assess.[18]

1.4. Problem Statement

In South Africa, there is an increasing incidence of cervical cancer despite the growing coverage of cervical smears as part of the cervical cancer screening programme.[2] This suggests a gap between screening test and the treatment of precancerous lesions. There is no robust national or sub-national system to monitor the dropout rate and waiting times of women within the multi-step screening-to-treatment cascade at district level. This study has used record linkage techniques to match cytology and histology records from the NHLS CDW to analyse the proportion of women who were screened for cervical cancer and required treatment who received it. This research should contribute to monitoring the cervical cancer screening programme at district level and, as South Africa is a high HIV prevalence country, add to the body of research around cervical cancer in women living with HIV that is recommend by the WHO.[8]

1.5. Aim and Objectives

The aim of this study was to use record linkage of data from the electronic NHLS database to analyse the follow-up, over two years, of women who had cervical smears done in public sector facilities of Johannesburg in 2017.

The specific objectives were to:

a) Determine the prevalence of HSIL on cervical cytology amongst the cohort of women

who had a screening cervical smear in a public sector facility in Johannesburg, 2017.

b) Determine the proportion of these women with HSIL who had a confirmatory test and/or precancer treatment in the Gauteng province within two years of the initial cervical smear.

c) Determine the time between the screening cervical smear and the i) follow-up procedure, ii) confirmatory test, and iii) treatment; and between the confirmatory test and treatment.

d) Compare these measures across age group, HIV status, and the facility type and subdistrict where screening took place.

2. Methods

2.1. Study design

This was a retrospective cohort study using secondary laboratory data. Record linkage was used to link two laboratory datasets. This linkage was used to indicate if women were utilising healthcare services and following up through the national cervical cancer screening programme.

2.2. Study setting

The Johannesburg district is a metropolitan municipality and the economic hub of South Africa.[45] It is a health district located in the centre of the Gauteng province and is one of the five districts in the province.[45] Johannesburg is divided further into seven geographical subdistricts named *A–G*.[45] Each subdistrict has its own management structure to operationalise the goals of the district.

Primary healthcare services in the public sector of Johannesburg, such as screening cervical smears, are provided across 114 primary healthcare clinics (PHCs), 11 community healthcare centres (CHCs) and two district hospitals.[46] When more specialised services are required, such as colposcopy, patients are referred to one of the two regional or three tertiary hospitals in Johannesburg or may follow up elsewhere in Gauteng where all levels of healthcare services are also available.[46]

In Johannesburg, cervical cancer is ranked as the sixth most common cause of death in women aged 25-64.[2] Prevention of cervical cancer through the screening programme follows the multi-step cascade summarised in Appendix D. A woman has a screening cervical smear at a primary healthcare facility, follows up for the cytology result and is referred for colposcopy if a precancer, such as HSIL, is detected. At the colposcopy clinic within a hospital, either a one-step LLETZ procedure is done to confirm and treat the precancer; or a two-step process of a confirmatory biopsy at the initial visit and a LLETZ at the subsequent visit only if the histology result of the biopsy confirms a precancer.

2.3. Study population and sampling

Johannesburg is the most populous district in the country with approximately 5.5 million people of which around 2.57 million are women.[45] In line with the national policy of three 10-yearly cervical cancer screening tests from the age of 30, the annual target for screening cervical smears is 10% of women over the age of 30 and is estimated at around 157 000.[46]

The study cohort included the cytology records of all women who had a cervical smear at a public sector healthcare facility within the Johannesburg district from 1 January 2017 to 31 December 2017. The histology laboratory records of all these women for a period of two years after the screening test were included in the analysis. No sampling was done.

2.4. Data extraction and cleaning

Data were extracted from the NHLS databases to create two baseline datasets for the purposes of this study – a screening dataset and a histology dataset. To establish the screening dataset, records that included all gynaecological smears performed in Johannesburg in 2017 (69 317) were extracted from the NHLS cytology database. To establish the histology dataset, records that included all gynaecological specimens excised from women in Gauteng from the beginning of 2017 to the end of 2019 (48 420) were extracted from the NHLS histology database. The latter dataset encompassed the whole of Gauteng because women may have a cervical smear in Johannesburg but have follow-up procedures outside Johannesburg. Therefore, to ensure that the follow-up of these screened women was not missed, histology records from all districts in Gauteng were extracted.

To facilitate record linkage between the cytology and histology datasets, the following variables of personal details were extracted from both datasets: name, surname, national identity (ID) number, hospital/clinic number, address, date of birth, age and race. To achieve the study objectives, the following additional variables were extracted for each record: HIV status; specimen collection date, type and result; and facility name, type and subdistrict. These two datasets formed the baseline datasets from which data cleaning and record linkage was performed. The list of data elements extracted from the NHLS datasets is provided in Appendix E.

Data cleaning was performed using STATA[47] and Python.[48] The first step of record linkage is pre-processing through cleaning and standardising the datasets. Non-alphabet characters were removed from the name and surname; second names were removed, and multiple surnames were joined. Ages under 15 and over 100 were recoded based on the date of birth or were recoded as “Missing” if no valid date of birth was available. In the cytology dataset, facility types were recoded based on the facility name using sources from healthcare authorities in Johannesburg.[49,50] Facility type was coded as “Other” for mobile clinics, clinics at workplaces, prisons, universities, and sites providing exclusively research-based services. The cytology dataset was deduplicated – exact duplicate records were removed (7); and for women

who had multiple records with the exact same personal details but with different specimen details, it was inferred that multiple cervical smears were done (904) and only the smear with the most clinically severe result, with a preference for the result of *HSIL*, was retained. An additional 1198 records of procedures that were not cervical smears were removed from the cytology dataset. Each record within the cytology and histology dataset was given a unique identifier. A third dataset ("HSIL dataset") was created to include only those cytology records with the result of HSIL for use in the linkage process.

2.5. Record Linkage

Record linkage was performed between the HSIL dataset and the histology dataset. Deterministic record linkage was not performed because the only unique identifier, the national ID number, had too low a rate of completeness (2.5% of the HSIL dataset records). In the absence of a unique identifier, and to accommodate for variables with slight inaccuracies of the captured values, probabilistic record linkage was performed. This was done using coding in Python.[48]

Performing a comparison between every variable of each record on one dataset with the corresponding variable of each record on a second dataset would require massive computing power and time due to the sheer number of combinations to compare.[51] This potential bottleneck was avoided using the record linkage technique of indexing through blocking. The blocking technique compares only a selected blocking variable between each record of one dataset with each record of the second dataset. If the blocking variable is an exact match, the two records are listed alongside each other as record pairs on a new dataset and only these record pairs are used for the next round of comparisons.[41]

Selecting the blocking variable is restricted by the data quality, uniqueness of the variable and computing power available. A blocking variable should have a high rate of completeness and accuracy. A highly unique blocking variable, such as name, requires high computing power for the blocking process. A blocking variable that is not unique, such as sex or age, will match many records and create a very long list of

record pairs on the new dataset that will subsequently require high computing power for the next round of comparisons. Multiple blocks may be performed using a different blocking variable for each block. The record pairs generated from each block are appended to form a single list. Using multiple blocks will increase the probability of matching as some truly matching record pairs that do not match on one block may match on other blocks. Subsequently, however, too many blocking processes will result in a long list of record pairs that will require high computing power in the next round of comparisons.[52]

Multiple blocks using different combinations of blocking variables were tried between the age, name, surname, and date of birth. The combination of blocks selected was that which resulted in a large number of record pairs on the blocking variable but within the constraints of computing power for the blocking process and subsequent comparisons. Two blocks were performed using the variables of age and then the first two letters of the surname. This resulted in “combined dataset 1” with records from the HSIL dataset alongside paired records with the same age from the histology dataset. The same process was used to create “combined dataset 2” based on the first two letters of the surname. Combined dataset 1 was appended to combined dataset 2 for comparisons to be performed on the other variables between each record pair.

Comparisons were performed on the name, surname and date of birth (day, month and year separately) of each record pair. There were high rates of incompleteness of the address (93.63%) and race (99.15%) variables, and these were therefore omitted from the linkage process. A list of data elements used for record linkage between the two datasets are provided in Appendix E. Each variable of a record that originated from the HSIL dataset was compared with the variable of the record alongside it that originated from the histology dataset. A score was calculated based on how similar each variable was between the records that originated from the HSIL and histology datasets. The string variables of name and surname were compared using the Jaro-Winkler string comparator.[53] This comparator is a metric that compares string variables based on their lengths, number of matching letters in the correct order and the number of transpositions required for the strings to be the same. Furthermore, to

accommodate for spelling errors that often do not affect the first letter of a name or surname, the Jaro-Winkler comparison places a higher weighting on the initial letter of the string variable.[53] The calculated score is between zero and one with the higher score indicating greater similarity.[53] A threshold value of 0.85 (which has been used in other record linkage studies)[54] for the Jaro-Winkler comparison was used to classify names and surnames as a potential matches or non-matches.

As no gold-standard linked dataset was available, an estimation of the *m*-probability (the probability of a true match being identified by the record linkage process as being a match) and the *u*-probability (the probability of a true match being identified by the record linkage process as being a non-match) was generated using an expectation-maximisation algorithm.[41] Thereafter, a second scoring process was performed between potential links. The likelihood ratio of a variable to match and to not match was calculated using the *m*- and *u*-probabilities. These likelihood ratios were then used to calculate a linkage score for each potential link (a real number of which a higher number indicates a higher probability that the two records are a match). A threshold score was set at 15.098 and was then compared with the linkage score of each potential link. Those potential links with scores above the threshold and which had the highest linkage score were then regarded as matched records. These records were deduplicated. This second deduplication was necessary to account for the two blocking processes. As two blocking processes were performed and the two combined datasets were appended, the records that were successfully matched based on both blocking variables and scored above the threshold had appeared twice on the matched dataset. The matched dataset of 1 431 matched records was created.

The matched dataset was manually reviewed through a clerical review of 250 records with the lowest linkage scores. This was to determine at what score the subjective quality of the matches was low and high. The name, surname, age, date of birth, race, address, national ID number, hospital number and facility name between the two matched records were manually compared to each other. It was found that most records with the lowest linkage score (201 records) did not seem to match a single patient. The comparison between the linkage score of each potential link and the

threshold was repeated using the next lowest linkage score as the threshold (15.465). This revised linkage threshold generated a new matched dataset of 1 230 records. A clerical review of the 10% of matched records with the lowest linkage scores showed that most records that were matched through the record linkage process were likely to be true matches. This new threshold was determined to be appropriate to ensure a high probability of true matches.

As some women who had had a cervical smear with the result of HSIL may have had multiple follow-up procedures, the linkage process was repeated to detect the records of these additional procedures. Using the linkage score threshold of 15.465, as determined by the clerical review, the potential links above this threshold which were the second and third most likely matches were added to the matched dataset. This resulted in 801 of the 1 230 records being matched to two histology records and 206 records matched to three histology records.

A successful match of a record between the HSIL and histology datasets was used as an indication that a woman who had had a cervical smear with the result of HSIL had gone on to follow up with a procedure that produced a specimen for histology along the cervical cancer screening cascade. This matched dataset with a single cervical smear and up to three histology procedures was taken forward for further data management.

2.6. Data management

Data management was performed using STATA.[47] The matched records were anonymised by removing the patient's name, surname, date of birth, address, race and hospital and national identity numbers; and the facility names. For the matched record pairs of 1 230 patient records of the matched dataset, the different procedures were re-ordered to ensure that each procedure occurred chronologically, each follow-up procedure was preceded by another procedure (eg. a record should only have a second follow-up procedure if a first follow-up procedure was done), and that the procedures were performed within the study period. Any procedure that preceded the cervical smear, that was performed beyond two years after the cervical smear, or did

not have a date was removed. The development of a final dataset is summarised in Figure 2.1. A list of data elements removed to anonymise the records is provided in Appendix E.

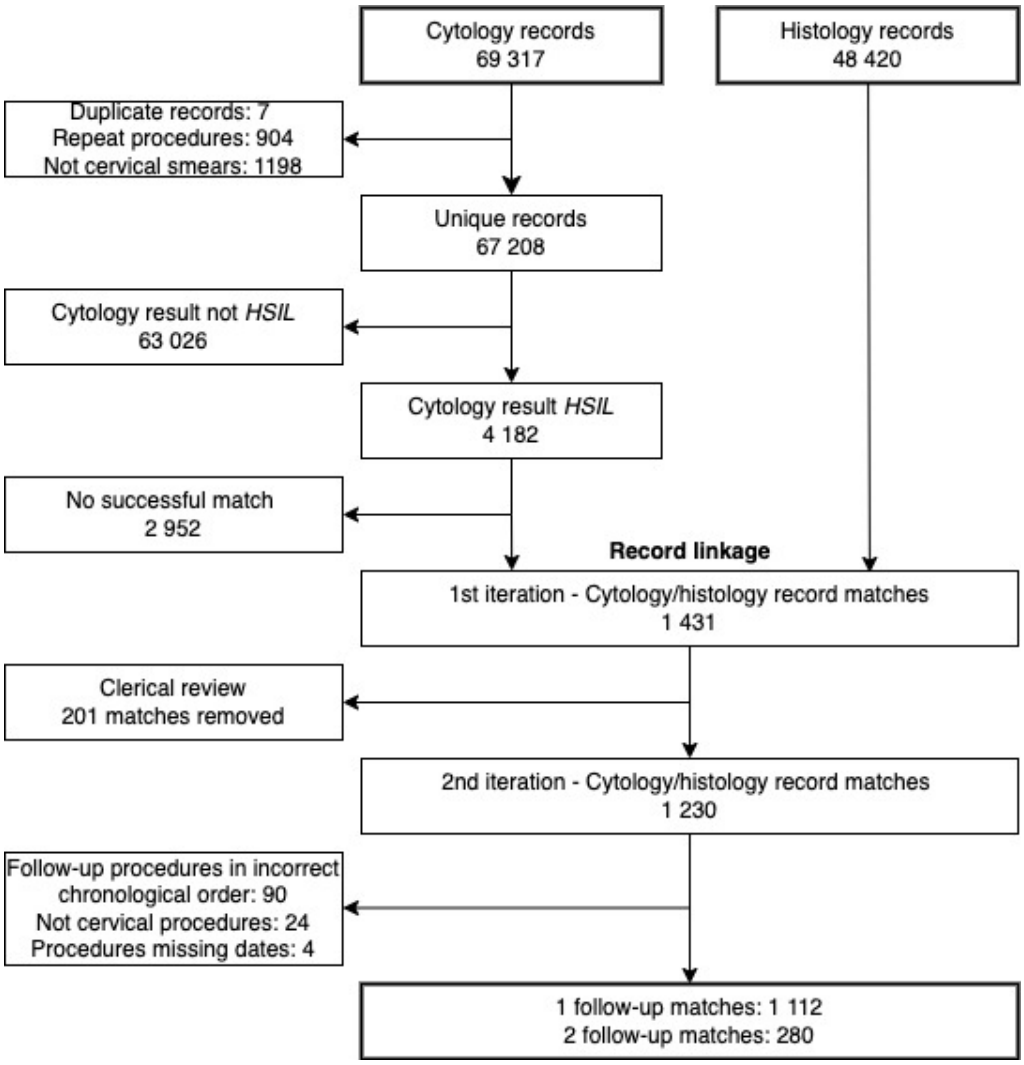


Figure 2.1: Summary of developing the final dataset from the cytology and histology datasets

The histology procedure type and result were already coded using SNOMED CT, an international coding system of clinical terminology,[55] but these codes were frequently missing or incomplete. A free-text variable was available and included a clinical and procedure history from the attending clinician and the histological findings of the pathologist. This variable was manually reviewed, and the procedure type and result were extracted and coded in two new categorical variables for each histology procedure.

Summary variables were developed to indicate the different follow-up steps of the cervical cancer screening cascade. Categorical variables were created and coded (yes/no) to indicate if follow-up after the cervical smear was required (records where the cervical smear result was HSIL and, if a follow-up procedure had taken place, the follow-up result was also HSIL); if a first follow-up had taken place and if it was appropriate for the one-step (treatment) or two-step (confirmatory test) process; if a second follow-up procedure was required (records where the first follow-up was appropriate for the two-step process), if it had taken place and if it was appropriate for precancer treatment. Procedures that indicated treatment were: LLETZ, LEEP (loop electrosurgical excision procedure), hysterectomy (only if the cervix was included in the resection) and cone biopsy. Procedures that indicated a confirmatory test were: punch biopsy and cervical biopsy. A summary of criteria for each summary variable is provided in the Appendix E.

Four numeric time variables were created and calculated between the dates of different procedures. The time to first follow-up was calculated as the interval between the cervical smear and the first histology procedure for women who had at least one follow-up histology procedure. The time to treatment was calculated for women who had a follow-up histology procedure that was appropriate for treatment of a precancerous lesion; and was calculated differently for women who followed up using the one- or two-step management process. For those following the one-step process, the time to treatment was calculated as the interval between the cervical smear and the first histology procedure; and for those following the two-step process, the interval between the cervical smear and the second follow-up procedure. The time to confirmation and time from confirmation to treatment were calculated for women who followed the two-step process and had a histology procedure that was appropriate for treatment of a precancerous lesion. Time to confirmation was calculated as the interval between the cervical smear and the first histology procedure; and time from confirmation to treatment was the interval between the first and second histology procedures.

The final dataset consisted of all women who underwent cervical screening. For those women who had a cytology result of HSIL, the additional variables from the record linkage and further data management processes were merged to the relevant records based on the unique identifier generated for the cytology dataset and preserved through the linkage process. The summary variables were reviewed and only existing or newly created variables with specific information relating to the study objectives were retained and are summarised in Appendix E. An example of a newly created variable in the final dataset is the variable for Management step that is coded only for those women who had a cytology result of HSIL, and was coded *0* if no follow up occurred, *1* if the one-step management process was done, and *2* if the two-step management process was done. This final dataset was then used to perform data analysis towards achieving the objectives of the study.

2.7. Data analysis

Data analysis was performed using STATA.[47] Median and interquartile range (IQR) were used to summarise age (age followed a non-normal distribution), and frequency summaries were calculated across four exposure variables – age group, HIV status, and the screening facility type and subdistrict.

Four binary (yes/no) outcome variables – HSIL on cytology, first follow-up done, confirmatory test done (two step processes begun), and treatment done – were summarised using frequencies across the exposure variables; and crude odds ratios (cOR) and 95% confidence intervals (CI) were calculated for each category within the exposure variables. Multivariate analysis was performed to assess the association between each outcome variable and the exposure variables using a stepwise logistic regression to identify only those exposure variables with a p-value of 0.05 or lower which were subsequently included in a model to calculate the adjusted odds ratios (aOR) and the associated 95% CI.

The confirmatory test rate was calculated as the proportion of women who required a first follow-up who began the two-step process. As recommended by the WHO,[22] the treatment rate calculation excluded women who had a precancerous lesion on

cytology but not on histology as these women were confirmed to not require treatment. The treatment rate was therefore calculated with a numerator of the women who received treatment and a denominator of women who had HSIL on cytology minus those women who had a first histology procedure that was appropriate for the two-step management process but had a histology result that was not HSIL.

The numeric outcome variables were summarised using medians and interquartile ranges (all followed non-normal distributions) across the four exposure variables. The Kruskal-Wallis test (for age group, facility type, subdistrict) and the Wilcoxon rank-sum test (for HIV status) were used to test for significant differences between each outcome variable with a p-value of 0.05. Furthermore, the time to first follow-up and time to treatment were assessed against an additional exposure variable of whether a one- or two-step management process was followed. These two time intervals were summarised and tested for association using the Wilcoxon rank-sum test and a p-value of 0.05.

2.8. Ethical considerations

The Human Research Ethics Committee (Medical) of the University of the Witwatersrand granted ethics approval for this study (reference number: M200648). The Academic Affairs, Research and Quality Assurance Committee of the NHLS authorised the use of laboratory data from the CDW (reference number: PR20298).

3. Results

3.1. Participant characteristics

The analysis included 67 208 women who had a cervical smear in Johannesburg public sector healthcare facilities in 2017. The median age of women was 40 years (IQR: 32 – 50), with 71.77% (48 233) of women within the age group 30-59 years. The HIV status of 26 250 (39.06%) women were missing, and of those with a record of HIV status, 48.84% (20 003) of women had HIV. Most cervical smears were performed in PHCs (44 200; 65.77%) and most were at facilities in subdistricts D (15 597; 23.31%) and F (14 615; 21.75%). The characteristics of the women screened are summarised in Table 3.1.

3.2. HSIL prevalence and the follow-up, confirmatory test and treatment rates

Of the 67 208 women screened, the prevalence of HSIL on cytology was 6.22% (4 182). A summary of women who screened positive with HSIL is summarised in Table 3.1.

Table 3.1: Summary of the women who had cervical smears, who screened positive with HSIL and who had a first follow-up procedure

Characteristic	Cervical smears		Cytology result: HSIL		First follow-up done	
	n	%	n	%	n	%
	67 208	100	4 182	100	1 112	100
Age group						
15-29	11 010	16.38	555	13.27	121	10.88
30-39	20 822	30.98	1 650	39.45	442	28.75
40-49	16 875	25.11	1 210	28.93	284	34.53
50-59	10 536	15.68	476	11.38	116	10.43
60+	6 420	9.55	202	4.83	43	2.87
Missing	1 545	2.30	89	2.13	6	0.54
HIV status						
Negative	20 955	31.18	686	16.40	168	21.76
Positive	20 003	29.76	2 082	49.76	604	78.24
Unknown	26 250	39.06	1 414	33.81	0	0
Screening test facility type						
PHC	44 200	65.77	2 690	64.32	635	57.10
CHC	8 306	12.36	623	14.90	161	14.48
Hospital	8 087	12.03	699	16.71	297	26.71
Other	6 615	9.84	170	4.07	19	1.71
Screening test subdistrict						
A	8 802	13.10	628	15.02	144	12.95
B	8 065	12.00	521	12.46	258	23.20
C	5 645	8.40	345	8.25	87	7.82
D	15 597	23.21	972	23.24	249	22.39
E	6 280	9.34	411	9.83	49	4.41
F	14 615	21.75	818	19.56	224	20.14
G	8 204	12.21	487	11.65	101	9.08

Of the 4 182 women who screened positive with HSIL and required follow-up, 1 112 (26.59%) women had a follow-up procedure, and 766 (68.88%) of these women had a histology result confirming HSIL. Among the 766 women with HSIL confirmed on histology, 62.01% (475) began the one- step process of being treated for a precancerous lesion. The remaining 291 (37.99%) began the two-step process by having a confirmatory histology test at their first follow-up visit, of which 144 (49.48%) followed up a second time to receive treatment of their precancerous lesion. Figure 3.1 summarises the number of women followed up along each stage of the cervical cancer screening cascade.

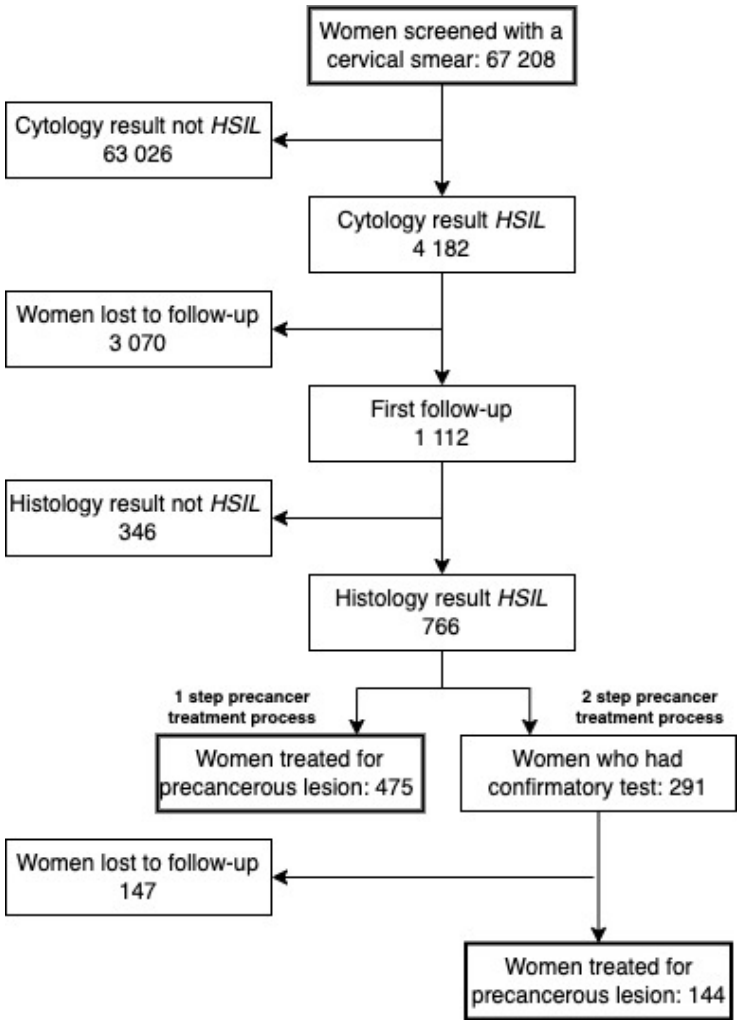


Figure 3.1: Summary of the number of women along each step of the cervical cancer screening

The confirmatory test rate (which also indicates the proportion of women who followed the two-step management process) of the 4 182 women with HSIL on cytology was 11.69% (489). The treatment rate of the 4 182 women was 19.13% (800). However, 346 women with a cytology result of HSIL were shown to not have HSIL on histology of their first follow-up test. These 346 women therefore did not require treatment and were excluded from further analysis along the management processes. The confirmatory test rate and treatment rate among these remaining 3 836 women was found to be 7.59% (291) and 16.14% (619) respectively.

3.3. Time intervals between screening tests, confirmatory tests and treatment

Of the 1 112 women who required and had a first follow-up, the time from their screening cervical smear to follow-up had a median of 131 days (IQR 80-189). For the 291 women with confirmed HSIL on histology who underwent the two-step process, the time from the screening cervical smear to confirmatory test had a median of 123 days (IQR 51-175), and from the confirmatory test to treatment of 125 days (IQR 84-210). The time to treatment had a median of 135 days (IQR 92-194) for women who underwent the one-step management process, 272 days (IQR 186-384) for the two-step management process, and 151 days (IQR 101-246) overall. Table 3.2 summarises the time intervals between procedures across the different characteristics of the participants.

3.4. Factors associated with HSIL prevalence and the follow-up, confirmatory test and treatment rates

The results of the logistic regression analysis assessing the association between the four categorical exposure variables (age group, HIV status, and facility type and subdistrict where the cervical smear was done) and three of the binary outcome variables (first follow-up done, confirmatory test done, and treatment done) are presented in Table 3.3.

Table 3.2: Summary of the time intervals between screening tests, confirmatory tests and treatment

Characteristic	Screening test to first follow-up		Screening test to confirmatory test		Confirmatory test to treatment		Screening test to treatment	
	n	median (IQR)	n	median (IQR)	n	median (IQR)	n	median (IQR)
	1 112	131 (80-189) days	291	123 (51-175) days	144	125 (84-210) days	619	151 (101-246) days
Age group								
15-29	121	127 (69-176)	34	133 (83-177)	18	160 (130-210)	71	128 (78-254)
30-39	442	127 (78-187)	123	122 (41-165)	61	113 (70-211)	257	146 (97-227)
40-49	284	134 (82-194)	107	121 (54-176)	50	118 (98-189)	216	158 (106-262)
50-59	116	135 (92-195)	15	124 (94-155)	9	116 (112-251)	55	168 (119-232)
60+	43	139 (61-188)	10	107 (61-181)	5	156 (111-244)	16	232 (163-362)
HIV status								
Negative	168	131 (82-177)	34	151 (107-224)	16	172 (112-233)	91	141 (99-246)
Positive	604	128 (76-183)	177	115 (49-172)	86	140 (97-243)	337	147 (99-252)
Screening test facility type								
PHC	635	134 (90-189)	175	132 (76-181)	75	119 (98-192)	344	152 (104-244)
CHC	161	127 (90-168)	23	131 (89-187)	13	140 (100-184)	105	132 (100-185)
Hospital	297	125 (52-202)	90	78 (31-141)	56	122 (71-248)	163	168 (102-279)
Other	19	44 (28-161)	3	90 (42-161)	0	-	7	36 (16-202)
Screening test subdistrict								
A	144	56 (29-135)	64	72 (33-176)	17	154 (117-252)	61	78 (34-279)
B	258	126 (52-209)	113	97 (42-162)	65	116 (78-243)	114	230 (124-383)
C	87	144 (110-189)	34	138 (118-176)	20	183 (106-215)	45	220 (136-295)
D	249	137 (107-181)	12	127 (53-178)	4	153 (110-216)	175	138 (112-180)
E	49	150 (100-233)	14	132 (156-350)	5	133 (116-195)	22	185 (126-282)
F	224	132 (86-185)	51	130 (104-166)	33	106 (92-140)	131	154 (93-243)
G	101	146 (116-204)	3	307 (224-337)	0	-	71	139 (116-193)

The odds of a cytology result of HSIL was significantly higher among women aged 30-39 (aOR 1.37; 95% CI 1.21-1.55) and aged 40-49 (aOR 1.23; 95% CI 1.08-1.40) than younger women (referent); living with HIV (aOR 3.02; 95% CI 2.75-3.31) than those without HIV (referent); and women who screened at CHCs (aOR 1.19; 95% CI 1.07-1.33) and hospitals (aOR 1.21; 95% CI 1.08-1.36) compared to PHCs (referent). Most subdistricts had significantly lower odds of HSIL than subdistrict A (referent) on univariate analysis, but these odds were not significant once controlled for the other three exposure variables.

Of the 1 112 women who required and attended a follow-up, after controlling for all exposure variables, facility type and subdistrict were significantly associated with follow-up. Significantly higher odds of follow-up were seen in women who screened at CHCs (aOR 1.28; 95% CI 1.03-1.58) and hospitals (aOR 1.61; 95% CI 1.26-1.90); and in subdistrict B (aOR 2.68; 95% CI 2.02-3.45). Subdistrict E had 0.42 times the odds (aOR 0.42; 95% CI 0.29-0.60) for follow-up than the referent.

Of the 766 women with HSIL confirmed on histology, the odds of having a confirmatory test and entering the two-step management process rather than the one-step process was significantly associated with the screening facility type and subdistrict. Women who screened at a healthcare facility other than the regular PHC, CHC or hospital had an odds of 0.24 (aOR 0.24; 95% CI 0.07-0.78) of having a two-step process than screening at a PHC. Subdistricts D (aOR 0.10; 95% CI 0.05-0.19), E (aOR 0.29; 95% CI 0.16-0.53), F (aOR 0.61; 95% CI 0.40-0.94) and G (aOR 0.05; 95% CI 0.02-0.17) all showed significantly lower odds of following the two-step process; and subdistrict B 2.52 times the odds (aOR 2.52; 95% CI 1.70-3.74) than subdistrict A.

The odds of being treated among those with histology-confirmed HSIL was also significantly associated with facility type and subdistrict. Women who screened at CHCs (aOR 1.34; 95% CI 1.04-1.72) and hospitals (aOR 1.63; 95% CI 1.28-2.09); or screened in subdistricts B (aOR 2.47 95% CI 1.70-3.58), D (aOR 1.80; 95% CI 1.30-2.49), F (aOR 1.78 CI 1.26-2.53), and G (aOR 1.48; 95% CI 1.02-2.15) had significantly higher odds of being treated than their respective reference categories. Subdistrict E had lower odds (aOR 0.47; 95% CI 0.28-0.78) of being treated than subdistrict A.

Table 3.3: Factors associated the follow-up, confirmatory test and treatment rates

	First follow-up done (n=1 112)			Confirmatory test done (n=291)			Treatment done (n=619)		
	n (%)	cOR[a] (95% CI[b])	aOR[c] (95% CI)	n (%)	cOR (95% CI)	aOR (95% CI)	n (%)	cOR (95% CI)	aOR (95% CI)
Age group			removed (p=0.38)			removed (p=0.25)			removed (p=0.50)
15-29	121 (10.88)	ref		34 (11.68)	ref		71 (11.47)	ref	
30-39	442 (38.75)	**1.31 (1.04-1.65)**		123 (42.27)	1.25 (0.85-1.86)		257 (41.52)	1.28 (0.97-1.70)	
40-49	284 (34.53)	**1.67 (1.32-2.11)**		107 (36.77)	**1.54 (1.03-2.31)**		216 (34.89)	**1.55 (1.16-2.07)**	
50-59	116 (10.43)	1.16 (0.86-1.55)		15 (5.15)	**0.53 (0.28-0.99)**		55 (8.89)	0.95 (0.65-1.39)	
60+	43 (3.87)	0.97 (0.66-1.43)		10 (3.44)	0.84 (0.41-1.74)		16 (2.58)	0.62 (0.35-1.09)	
Missing	6 (0.54)			2 (0.69)			4 (0.65)		
HIV status			removed (p=0.08)			removed (p=0.06)			removed (p=0.10)
Negative	168 (21.76)	ref		34 (16.11)	ref		91 (21.26)	ref	
Positive	604 (78.24)	**1.26 (1.03-1.54)**		177 (83.89)	1.56 (0.99-2.44)		337 (78.74)	1.27 (0.98-2.63)	
Unknown									
Facility type of cervical smear									
PHC	635 (57.10)	ref	ref	175 (60.14)	ref	ref	344 (55.57)	ref	ref
CHC	161 (14.48)	1.13 (0.92-1.38)	**1.28 (1.03-1.58)**	23 (7.90)	**0.38 (0.23-0.63)**	1.20 (0.74-1.97)	105 (16.96)	**1.39 (1.10-1.77)**	**1.34 (1.04-1.72)**
Hospital	297 (26.71)	**2.39 (2.01-2.85)**	**1.54 (1.26-1.90)**	90 (30.93)	1.29 (0.92-1.81)	1.41 (0.99-2.00)	163 (26.33)	**2.34 (1.89-2.90)**	**1.63 (1.28-2.09)**
Other	19 (1.71)	**0.41 (0.25-0.66)**	**0.32 (0.19-0.52)**	3 (1.03)	0.66 (0.17-2.58)	**0.24 (0.07-0.78)**	7 (1.13)	**0.28 (0.13-0.61)**	**0.22 (0.10-0.47)**
Subdistrict of facility of cervical smear									
A	144 (12.95)	ref	ref	64 (21.99)	ref	ref	61 (9.85)	ref	ref
B	258 (23.20)	**3.30 (2.56-4.25)**	**2.68 (2.02-3.54)**	113 (38.83)	1.59 (0.95-2.64)	**2.52 (1.70-3.74)**	114 (18.42)	**3.19 (2.27-4.49)**	**2.47 (1.70-3.58)**
C	87 (7.82)	1.13 (0.83-1.54)	1.10 (0.81-1.50)	34 (11.68)	0.94 (0.49-1.78)	0.97 (0.62-1.51)	45 (7.27)	1.44 (0.95-2.17)	1.39 (0.92-2.10)
D	249 (22.39)	1.16 (0.91-1.46)	1.02 (0.80-1.31)	12 (4.12)	**0.05 (0.02-0.10)**	**0.10 (0.05-0.19)**	175 (28.27)	**2.08 (1.53-2.85)**	**1.80 (1.30-2.49)**
E	49 (4.41)	**0.45 (0.32-0.65)**	**0.42 (0.29-0.60)**	14 (4.81)	0.57 (0.25-1.27)	**0.29 (0.16-0.53)**	22 (3.55)	**0.52 (0.31-0.86)**	**0.47 (0.28-0.78)**
F	224 (20.14)	1.27 (1.00-1.61)	1.22 (0.94-1.59)	51 (17.53)	**0.36 (0.21-0.60)**	**0.61 (0.40-0.94)**	131 (21.16)	**1.86 (1.35-2.58)**	**1.78 (1.26-2.53)**
G	101 (9.08)	0.88 (0.66-1.17)	0.83 (0.62-1.11)	3 (1.03)	**0.03 (0.01-0.10)**	**0.05 (0.02-0.17)**	71 (11.47)	1.59 (1.10-2.29)	**1.48 (0.09-0.15)**

a: crude odds ratio. b: confidence interval. c: adjusted odds ratio.

3.5. Factors associated with the time intervals between the screening test, confirmatory test and treatment

Factors associated with the time intervals between the screening test, confirmatory test and treatment, and relevant test results, is summarised in Table 3.4. The time to follow-up was significantly associated with facility type (chi2(3)=8.344; p=0.0394) and subdistrict where the screening test was performed (chi2(6)=70.648; p=0.0001), and whether a one-step or two-step management process was followed (chi2(1)=9.911; p=0.0016).

Table 3.4: Summary of factors associated with time intervals between the screening test, confirmatory test and treatment

Characteristics	Screening test to first follow-up		Screening test to confirmatory test		Confirmatory test to treatment		Screening test to treatment	
	chi2 (df)[a]	p-value	chi2 (df)	p-value	chi2 (df)	p-value	chi2 (df)	p-value
Age group	2.117 (2)	0.7142	1.790 (4)	0.7743	4.224 (4)	0.3765	12.969 (4)	**0.0114**
HIV status	0.085 (1)	0.7709	3.823 (1)	0.0506	0.374 (1)	0.5406	0.001 (1)	0.9750
Screening test facility type	8.344 (3)	**0.0394**	15.892 (3)	**0.0012**	0.650 (2)	0.7225	8.287 (3)	**0.0404**
Screening test subdistrict	70.648 (6)	**0.0001**	22.952 (6)	**0.0008**	7.011 (5)	0.2198	50.643 (6)	**0.0010**
Management steps	9.911 (1)	**0.0016**	-	-	-	-	102.102 (1)	**0.0001**

a: degrees of freedom

The time to treatment was significantly associated with the age group of women (chi2(4)=12.969; p=0.0114), the facility type (chi2(3)=8.287; p=0.0404) and subdistrict (chi2(6)=50.643; p=0.0010), and whether a one- or two-step process was followed (chi2(1)=102.102; p=0.0001). Markedly longer times to treatment were experienced by women aged 61-75 (median 246 days; IQR 160-328), who screened in subdistrict B (median 230; IQR 124-383) and C (median 220; IQR 136-295), and those following the two-step process (median 272; IQR 186-384). Very short times to treatment were found in women who had their screening done at a healthcare delivery platform outside the general public sector facilities (median 36; IQR 16-202) or in subdistrict A (median 78; IQR 34-279).

For the 291 women who began the two-step management process, the first step from their cervical smear to confirmatory test was significantly associated with the facility type (chi2(3)=15.892; p=0.0012) and subdistrict (chi2(6)=22.952; p=0.0008). The

second step of the management process was not significantly associated with any of the exposure variables. The intervals measured were not significantly associated with the HIV status of women.

4. Discussion

Population-level prevention of cervical cancer through screening begins with achieving a high coverage of screening followed by treatment of precancerous lesions that are identified through screening. Key indicators of the cervical cancer programme include the screening and precancerous treatment rates.[7,22] These indicators allow for programme monitoring and to track progress towards the WHO global targets to eliminate cervical cancer as a public health problem.[8] In the absence of routine data on precancer treatment rate,[19] this study set out to use a novel approach of record linkage of laboratory data to develop a dataset that was subsequently used to measure the rates of women following up along the cervical cancer screening cascade. The major findings of this study were that 67 208 cervical smears were performed and there was an HSIL prevalence of 6.22% in the public sector of Johannesburg in 2017; follow-up occurred in around a quarter of the women, of which 38.00% began management with a confirmatory test. Among those who required treatment, 16.14% had received it. The one-step management was more common and had a shorter time to treatment. Overall, the median time to treatment was 151 days. The follow-up rate, confirmatory test rate and treatment rate; and the times from the cervical smear to follow-up, confirmatory test and treatment were significantly associated with facility type and subdistrict. This section discusses the public health implications of these findings and relates the findings to existing literature.

In the 2016/2017 financial year, the District Health Barometer (DHB), which draws data primarily from the District Health Information Software (DHIS), reported that the coverage of cervical cancer screening in the Johannesburg district was 47.40% approximating to 62 000 women.[46] This study found the cervical smear records of over 67 000 women which is greater than the reported district coverage,[46] but more similar to a study of NHLS data where 65 500 cervical smears were found during the 2013/2014 financial year.[39] No study was found that covered the exact study period

and population as this study and the differences in screening coverage across the different sources are likely due to the discordance in dates between the study periods. Furthermore, the higher number of cervical smears found in this study may reflect the DHB's exclusion of the 6 615 screening tests performed at sites listed under "other" which may not be captured on the DHIS; or that exclusively screening cervical smears are included in the DHB study population,[46] while this study could not differentiate between diagnostic and screening cervical smears. The district coverage reported by this study, the 2013/2014 study[39] and the DHB[46] all show that the district annual target of 158 000 screening tests was not reached, and coverage was lower than what was achieved at the national (62%) and provincial (50%) levels.[46] The low screening coverage indicates that most women aged 30-60 years in Johannesburg did not have a cervical smear and could not benefit from early detection and treatment of a precancerous lesion to prevent cervical cancer from developing. The incidence of cervical cancer, therefore, cannot be reduced when screening coverage is low.

This study found an HSIL prevalence of 6.22%, which is similar to studies in other settings in South Africa including the Overberg district (7.7%)[25] and a 2018/2019 study in Johannesburg (4.64%).[17] The high odds of HSIL in women aged 30-49 is consistent with the results of a multicentre, population-based South African survey which found the highest HSIL prevalence among women aged 30-44.[23] This finding supports the policy decision to target this age-group for cervical cancer screening[5,7] to achieve the highest rate of detection of precancerous lesions that may be treated to prevent cervical cancer. This current study also found significantly greater odds of HSIL in women who screened at higher levels of the healthcare system (CHCs and hospitals) than in women who screened at PHCs. This finding is similar to the findings of a previous study in Johannesburg that piloted a decentralised model of care and found higher percentages of precancerous lesions on cytology and of HSIL on histology among women who screened and followed up at the hospital when compared to women who screened and followed up at the primary care level.[32] This current study could not explore the reasons for the differences, but this might be due to higher-risk women, based on medical history and clinical findings, screening at higher levels of care – as was suspected in the previous Johannesburg study.[32] Another possible

reason could be that the quality of cervical smears is greater, and the tests performed at higher levels of care are therefore more sensitive to detecting precancerous lesions. Poor quality cervical smears are less likely to detect precancerous lesions for treatment. Some concern has been expressed by previous studies regarding the quality of cervical smears at lower levels of care;[23,24] however, these studies did not compare the quality between levels of care. Furthermore, it has been noted that with the necessary quality improvement initiatives, comparable quality between cervical smears performed at different levels of the healthcare system would be possible,[24,32] should this be the case in Johannesburg facilities.

The low follow-up rates found in Johannesburg are concerning. This study found that only 26.60% of women with HSIL on cytology had followed up, similar to findings of a 24.86% follow-up rate in a study conducted at an HIV clinic in the district of eThekwini.[26] The loss to follow-up of 3 070 women in this current study was most pronounced in women who screened at PHCs which is particularly concerning given that this facility type was utilised by the majority of women screening for cervical cancer. Of the 766 who followed up and were confirmed through histology to have HSIL, 80.81% had been treated; yet overall, when accounting for the women with unknown histology findings due to loss-to-follow-up, only 16.14% of the women who needed follow-up for treatment had received it. Low follow-up and precancerous treatment rates negate the purpose of a screening cervical smear as women who do not receive follow-up cannot benefit from the early detection and treatment of precancerous lesions. This study has shown a wide gap between current performance of the precancer treatment rate and what is required by the WHO elimination target of treating 90% of women with precancerous lesions,[8] making it unlikely for Johannesburg to meaningfully contribute towards South Africa's goal of meeting this target by 2030.

One barrier to follow-up and precancer treatment is that the screening cascade is complex as it spans many visits to multiple healthcare facilities.[24] Women who screened at a CHC and more so at a hospital rather than a PHC had a significantly increased odds of follow-up and precancer treatment. These favourable findings for

women who screened at higher levels of care may allude to an increased access to follow-up and treatment services when screening and precancer treatment were both performed at a single site. A previous study in Johannesburg illustrated similar favourable results but at a lower level of care as a result of both cervical smear and colposcopy services being provided at the CHC level through a decentralised service delivery model.[32] Studies in other developing countries have assessed less complex cervical cancer screening cascades where women visit the healthcare facility fewer times to receive treatment.[36,38] In African countries, screening has been performed using visual inspection with acetic acid (VIA). Using VIA, in a single consultation, precancerous lesions are detected visually by the clinician and they are able to provide precancer treatment immediately.[38] In a study in Central America, self-sampling HPV screening tests were distributed to women near or at their homes. In one model, the women performed the screening test on themselves and sent the sample to a clinic with a community health worker. The women received their results at home and only needed to visit the healthcare facility if their screening test was positive for HPV. At the healthcare facility, a VIA was conducted and if a precancerous lesion was detected, precancer treatment was provided immediately.[36] Both these examples limited the number of times women need to present to healthcare facilities and therefore removed the opportunity of being lost to follow-up and increased the treatment rate of women.

Differences in follow-up rates across screening facilities in different subdistricts was recognised in an unpublished Johannesburg study a decade ago. Subdistricts B, C and A had the highest follow-up rates, and subdistricts F and E the lowest.[31] Significantly greater odds of following up across subdistricts in this current study was found in subdistrict B, and lower odds in subdistrict E, mirroring this aspect of subdistrict performance from the previous study. While this study could not explore the reasons for high follow-up rates in subdistrict B, in view of the persistently high follow-up rates, and the subdistrict with the highest cervical cancer screening coverage (66%)[46] in Johannesburg, this may allude to the subdistrict overcoming the barriers to uptake of cervical cancer screening services.

The long waiting periods between facility visits for procedures or receiving results has been found repeatedly as a barrier to follow-up along the cervical cancer screening cascade.[24,33,56,57] However, these qualitative and review studies do not quantify the follow-up time intervals that would be acceptable to women. Subdistrict B was found to have the highest absolute number of women who followed up and had among the shortest time to follow-up; and subdistrict E had the lowest absolute number following up and the longest time to follow-up. While causality cannot be determined, this pattern may reflect that women are less likely to follow up in subdistricts with long follow-up intervals; or that subdistricts with greater numbers of women following up are able to streamline their processes in response to this high uptake and improve their service delivery with shorter waiting periods.

Treatment through the one-step management process was shorter than the two-step process. While screening to follow-up was shorter for women following the two-step process, the long confirmatory test to treatment interval resulted in a much longer screening to treatment time for the two-step process. In this study, most women followed the one-step process which may be more acceptable to them. As the one-step process is of comparable quality to the two-step process, particularly in a population with high HIV prevalence,[13] the one-step management process may be appropriate for this setting.

A peculiar pattern was seen in subdistrict B. While the odds of following the two-step management process were greater in this subdistrict, which aligned well with the greater odds of shorter follow-up but longer overall time to treatment; the odds of being treated were markedly higher than other subdistricts despite the two-step process leading to lower treatment rates in other subdistricts. This discordant picture of long times to treatment but high treatment rates supports findings that the factors that facilitate improved follow-up and treatment rates are complex and are not limited to the number of visits required or the waiting periods between procedures. Other healthcare system and patient-related factors play a role in improving follow-up and treatment rates.[58,59]

South Africa, with the highest number of people living with HIV in the world, has approximately 4.8 million women aged 15 years and over who are infected.[60] The South African policy on cervical cancer prevention[7] highlights the importance of screening women with HIV due to their increased risk of both developing precancerous lesions and progressing to develop cervical cancer.[7] This study found that among women who screened for cervical cancer and had a known HIV status, 48.84% had HIV; and of those who had a cytology result of HSIL, 75.22% had HIV. The odds of a screening test having the result of HSIL in women living with HIV was 3.05 times the odds of women without HIV. And while the overall prevalence of HSIL was 6.22%, the prevalence among women with HIV was 10.41%. This finding of a significant association between HSIL prevalence and HIV status supports established literature that describes HIV as a risk factor for cervical intraepithelial lesions.[5,61] This finding is also similar to a recent study in Johannesburg that showed a prevalence of 10.5% in participants with HIV.[17] Among women who required follow-up, the odds of those with HIV following up was 1.26 times that of women without HIV on bivariate analysis, but not significant when controlled for other variables. The HIV status of women did not significantly affect the odds of following the shorter one-step management process nor receiving treatment, suggesting that there may not be any prioritisation for women living with HIV.

With the introduction of HPV vaccination to South Africa as a primary preventative measure for cervical cancer in 2014,[62] the incidence of cervical cancer is expected to decrease. In the interim, a dual approach of vaccinating young women while maintaining screening for older women is recommended to ensure appropriate preventative interventions throughout the life-course. When women who were vaccinated when they were younger reach the age of eligibility for cervical cancer screening, their lower risk for developing cervical cancer should allow the country to reassess the cost-effectiveness of the screening programme and to even consider phasing repeated routine screening out.[63]

The global standard to significantly reduce the burden of cervical cancer is guided by the WHO elimination strategy.[8] South Africa has committed to achieving the 90-70-90

targets for HPV vaccination, screening, and precancer and cancer treatment by 2030.[64] In its year of inception, HPV vaccination coverage reached 65% of eligible young women who received the required two doses; but has since dropped to 56% by 2017.[65] Coverage of screening has increased from 40% in 2016 to 47% in 2020 using the most recent NIDS definition that considers the different screening frequencies of women based on their HIV status and the HIV prevalence of the target population.[2] No robust national monitoring system for precancer- and cancer-treatment rates exists. Studies in different health districts around South Africa have shown treatment rates from around 26%[11] to 64%,[26] and this study found a treatment rate of 16% – all far below the target of 90%. No studies were found that quantify the country's progress towards achieving treatment of women with cervical cancer, leaving a gap in the ability to track this indicator.

Mapping the cascade of the cervical cancer screening programme is often done in South Africa by reviewing medical records and facility registers which are usually paper-based.[17,25,27] Laboratory data has also been used to visualise the cascade across the country,[39] or to add to the data elements of studies using paper-based data;[25,26] and other countries have begun to harness record linkage techniques towards monitoring their cervical cancer screening programme.[43] The technique of record linkage and the use of NHLS data in this study has shown similar results to previous studies using more traditional sources. This method and data source may be a reliable method for routinely monitoring the components of the cervical cancer screening programme.

The COVID-19 pandemic that has affected South Africa from March 2020 has caused a reduction in uptake for non-COVID-19 related services in the country and worldwide[66,67] due to limited mobility, reluctance to visit healthcare facilities and the temporary suspension of non-urgent services such as cancer screening.[66] In addition to catchup programmes to improve cervical cancer screening uptake, innovative methods such as equitable risk stratification to prioritise screening for those at highest risk (eg. women living with HIV), the use of self-screening tests, and telemedicine, rather than in-person visits, to discuss screening results. In-person investigation or

treatment will then be performed on the relatively fewer number of women who screen positive.[68] Despite the negative effects on uptake and adherence along the cervical cancer screening cascade, opportunities have arisen and these may help South Africa reach closer to its screening targets.

5. Limitations

We identified three main limitations of this study. Treatment rate may have been underestimated; the positive impact of measuring treatment rate may have been overestimated; and not all precancerous lesions were reviewed to assess follow-up.

A successful link between records in the cytology dataset and histology dataset was used as a proxy to identify those women who received follow-up. Our study period and setting were limited to a two-year follow-up period, and to public sector facilities within the province of Gauteng. These limits would have excluded any woman who had had appropriate treatment beyond two years, in the private sector, or outside of Gauteng. However, these limits were guided by the CDC time-to-treatment standard of three months between cervical smear and treatment, and South African studies which showed a mean time-to-treatment of 4.5 months[25] and median times of 39 days (IQR 20-95)[26] and 4.4 months (IQR 3.1-7.1) ranging up to 15.4 months[31] in different studies across Johannesburg. This makes it likely that most follow-ups would take place within the two-year timeframe that was chosen for this study. While women may elect to attend private sector facilities or facilities outside of the province, Johannesburg is a district situated in the centre of Gauteng with all referral pathways to higher levels of care, where colposcopy clinics are generally situated, within the province.[49] It is therefore likely that this study will miss very few follow-ups by reviewing facilities beyond the Johannesburg district, but within Gauteng.

Using records from a laboratory service omits procedures that do not produce a specimen for laboratory assessment. This differs from other methods of data collection such as using patient files or facility registers which would indicate that a procedure was performed irrespective of whether a specimen was produced or not. Therefore, despite a woman having appropriate treatment through ablative therapies or excisional

therapies without submitting the specimen to a laboratory, these would not appear on the laboratory record and these women would not be considered to have followed up. This limitation is similar to the study using record linkage on NHLS data to track follow-up visits among people living with HIV which recognised that patients who received adequate clinical care without laboratory investigations could have incorrectly been considered to be lost to follow-up.[18] Regarding cervical cancer screening, the national policy, expresses a preference for excisional techniques and sending specimens to a laboratory for confirmation or assessment of the adequacy of treatment.[7] It would be an anomaly for a colposcopy clinic in Gauteng to omit the step of laboratory assessment, therefore minimising the number of follow-up procedures missed by this study through analysing laboratory records.

Other reasons for possibly underestimating the follow-up rate of women by not successfully linking records between the two procedures include poor data quality – the absence of unique identifiers; incomplete variables of personal information; and inconsistencies such as spelling errors or surname changes – which reduces the ability to link records. To maximise the matching process, the probabilistic record linkage methodology was used. Multiple variables were compared for similarity and therefore data inconsistencies such as spelling errors would still be considered in the matching process. The precision of the record linkage based on a manual review of linked and unlinked records was calculated. The linkage in this study had a sensitivity of 96.64%, a specificity of 95.87%, and a positive predictive value of 95.83% which is similar to other studies using similar techniques.[18,44,69] Of note, as age was used for indexing through the blocking technique, if a change in age or omission of age on one dataset would have resulted in missing a match, underestimating the number of women who had followed up. Age was selected for the blocking process as it was complete in 97.82% of cytology records and 98.06% of histology records; and to further mitigate the risk of under-linking records, an additional block on the first two letters of the surname of each record was used to capture those records that should be matched but have different documented ages.

While a useful measure of service uptake, the treatment rate may overestimate the positive impact of the cervical cancer screening programme in reducing the incidence of cancer. This is because cervical cancer may only be prevented if the therapeutic procedure was of a high quality and if follow-up beyond treatment was adhered to.[12]

A final limitation of this study to provide results towards monitoring the confirmatory and treatment components of the cervical cancer screening cascade is that we assessed the follow-up of only those women with cervical smear results of HSIL. While other cytology findings also require follow-up with colposcopy,[5,7] the South African 2000[28] and 2017[7] cervical cancer policies have consistently recommended colposcopy for women with HSIL. Therefore, analysing the follow-up of these women should give a good indication of the gap between screening tests and the treatment of precancerous lesions.

6. Conclusion and Recommendations

6.1. Conclusion

This study has demonstrated the use of record linkage of laboratory data for use in monitoring the cervical cancer screening programme in Johannesburg. The major findings of this study were very low follow-up and precancer treatment rates. This highlights gaps in the screening programme which is far from meeting the WHO elimination target for precancer treatment, and will not markedly reduce the incidence of cervical cancer. A less complex cascade from screening to treatment with fewer visits to healthcare facilities may reduce loss-to-follow-up; benchmarking the service delivery models of high performing subdistricts may improve the treatment rate; and improving the information system to generate the necessary indicators will enable programme managers to monitor the screening programme towards building a healthcare system that is responsive to the health needs of women.

6.2. Recommendations

The high loss-to-follow-up rate seen in this study is likely due to the complex cascade from screening to precancer treatment that requires multiple visits to multiple healthcare facilities. Simplifying this cascade, possibly though techniques like

screening through self-swabbing, or the VIA that allows both screening and precancer treatment to be performed in one visit should be considered. The WHO recommends that cervical cancer control programmes should promote a single-visit approach. To support this, the WHO further recommends integrating the cervical cancer screening programme into the package of services offered at primary care level; strengthening the laboratory and quality assurance capacity of the healthcare system; and creating an enabling environment by understanding and addressing barriers to access to care.[8]

This study showed that follow-up and treatment rates are higher when screening is performed at a higher level of care. The reasons for this discrepancy should be investigated. At hospitals, where both screening and precancer treatment is performed, the higher treatment rates may be due to easier access to the entire cervical cancer screening cascade at a single healthcare facility. In a Johannesburg study of decentralising precancer treatment, high uptake of services was found at a CHC which was capacitated to perform precancer treatment.[32] Decentralising precancer treatment services to lower levels of the healthcare system (PHCs and CHCs) and capacitating these levels may therefore improve access to cervical precancer treatment services.[9]

This study showed subjectively long waiting periods between procedures in the cervical cancer screening cascade. A long time to follow-up after a screening test, and to precancer treatment is an access barrier that limits the cervical cancer screening programme's ability to successfully undertake secondary prevention of cervical cancer. Further investigation into the programme models used across different subdistricts and levels of care may identify examples of best practice. This will allow for quality improvement projects and benchmarking against those with higher follow-up and treatment rates, and shorter times to treatment.[10] Evidence shows that long waiting times between procedures are often related to the long laboratory turnaround times, delayed reporting of results, long waiting lists for procedures, and inefficiencies of the healthcare system.[10,24] To promote a responsive healthcare system, efforts should be made to capacitate the healthcare system to address these barriers in order to reduce these long time intervals. Recommendations to improve long time intervals

have been covered in local and international literature.[10,56] Healthcare facility managers need to liaise and collaborate with laboratory managers to ensure that quality assurance is being performed to reduce turn-around-times. Electronic or paper-based tracking systems should be developed to promptly inform patients of their results. Referral systems and communication between different levels of care, and facilities and women need to be streamlined and clear.[10,56] Improving waiting periods between procedures will contribute to improving the accessibility of the service by providing a service that is acceptable to women, increasing uptake and improving the chances of success of the cervical cancer screening programme.

A useful way to bring together existing electronic data for use in monitoring of health programmes includes the record linkage technique used in this study. This study used laboratory data to track indicators along the cervical cancers screening cascade and shows some comparability with similar studies that used different data sources. In South Africa, electronic NHLS data is currently accessible, and this source and the techniques of record linkage may be a viable method of programme monitoring at a district, provincial and national level. Further studies to quantify the accuracy, feasibility and efficiency of this methodology may provide insight and guidance on its use. A strong health information system is imperative for a successful cervical cancer screening programme. This should enable the healthcare system to monitor screening coverage, precancer follow-up and treatment rates, and the time intervals between procedures, among other indicators to monitor the cervical cancer screening programme. In the absence of a comprehensive health information system, multiple methods and sources of information may be used to generate the indicators required for programme management and to track progress towards the WHO elimination targets. These methods may include nationally representative surveys or research, or routine data collected at healthcare facilities. Data sources may include laboratory data, as used in this study, data from medical records and files or facility registers.[22,70]

7. References

1. International Agency for Research on Cancer. Cancer profile: South Africa. 2021.
2. Massyn N, Day C, Ndlovu N, Padayachee T. District health barometer 2019/20. Massyn N, Padayachee T, editors. Durban: Health Systems Trust; 2020.
3. Hamdi Y, Abdeljaoued-Tej I, Zatchi AA, Abdelhak S, Boubaker S, Brown JS, et al. Cancer in Africa: the untold story. Front Oncol. 2021;11. doi:10.3389/fonc.2021.650117
4. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49. doi:10.3322/caac.21660
5. World Health Organization. Comprehensive cervical cancer control: A guide to essential practice. 2nd ed. Geneva: World Health Organization; 2014.
6. Denny L, Kuhn L. Cervical cancer prevention and early detection from a South African perspective. In: Padarath A, Barron P, editors. South African Health Review. Durban: Health Systems Trust; 2017. p. 189–95.
7. National Department of Health SA. Cervical cancer prevention and control policy. Pretoria; 2017.
8. World Health Organization. Global strategy to accelerate the elimination of cervical cancer as a public health problem. Geneva: World Health Organization; 2020.
9. Strasser-Weippl K, Chavarri-Guerra Y, Villarreal-Garza C, Bychkovsky BL, Debiasi M, Liedke PER, et al. Progress and remaining challenges for cancer control in Latin America and the Caribbean. Lancet Oncol. 2015;16:1405–38. doi:10.1016/S1470-2045(15)00218-1
10. Cervical Health Implementation Project (CHIP). Implementing cervical screening in South Africa. Volume 1: A guide for programme managers. 2004.
11. Rohner E, Mulongo M, Pasipamire T, Oberlin AM, Goeieman B, Williams S, et al. Mapping the cervical cancer screening cascade among women living with HIV in Johannesburg, South Africa. Int J Gynaecol Obstet. 2021;152:53–9. doi:10.1002/ijgo.13485
12. Adam Y, van Gelderen CJ, de Bruyn G, McIntyre JA, Turton DA, Martinson NA. Predictors of persistent cytologic abnormalities after treatment of cervical intraepithelial neoplasia in Soweto, South Africa: A cohort study in a HIV high prevalence population. BMC Cancer. 2008;8(211):1–8. doi:10.1186/1471-2407-8-211
13. Adam Y, van Gelderen CJ, Newell K. “Look and LLETZ” - A Chris Hani

Baragwanath Hospital experience. South African Med J. 2008;98(2):119–22. doi:10.7196/SAMJ.59

14. National Institute for Communicable Diseases SA. Annual performance plan 2021-2022. 2021.

15. Bassett IV, Huang M, Cloete C, Candy S, Giddy J, Frank SC, et al. Assessing the completeness and accuracy of South African national laboratory CD4 and viral load data: A cross-sectional study. BMJ Open. 2018;8:1–7. doi:10.1136/bmjopen-2018-021506

16. National Health Laboratory Services SA. NHLS diagnostic services [Internet]. [cited 2022 Apr 9]. Available from: https://www.nhls.ac.za/diagnostic-services/

17. Hopkins KL, Jaffer M, Hlongwane KE, Otwombe K, Dietrich J, Cheyip M, et al. Assessing national cervical cancer screening guidelines : Results from an HIV testing clinic also screening for cervical cancer and HPV in Soweto , South Africa. PLoS One. 2021;16(7):1–13. doi:10.1371/journal.pone.0255124

18. Fox MP, Bor J, Brennan AT, MacLeod WB, Maskew M, Stevens WS, et al. Estimating retention in HIV care accounting for patient transfers: A national laboratory cohort study in South Africa. PLoS Med. 2018;15(6):1–21. doi:10.1371/journal.pmed.1002589

19. National Department of Health SA. National indicator data set: data element definitions April 2017 to March 2019. Pretoria; 2017.

20. National Institute for Communicable Diseases SA. National Cancer Registry [Internet]. [cited 2022 Apr 1]. Available from: https://www.nicd.ac.za/centres/national-cancer-registry/

21. National Department of Health SA. National cancer strategic framework for South Africa 2017-2022. Pretoria; 2017.

22. World Health Organization. Improving data for decision-making: a toolkit for cervical cancer prevention and control programmes. Geneva; 2018.

23. Fonn S, Bloch B, Mabina M, Carpenter S, Cronje H, Maise C, et al. Prevalence of pre-cancerous lesions and cervical cancer in South Africa - A multicentre study. South African Med J. 2002;92(2):148–56.

24. Moodley J, Kawonga M, Bradley J, Hoffman M. Challenges in implementing a cervical screening program in South Africa. Cancer Detect Prev. 2006;30(4):361–8. doi:10.1016/j.cdp.2006.07.005

25. Blanckenberg ND, Oettle CA, Conradie HH, Krige FK. Impact of the introduction of a colposcopy service in a rural South African sub-district on uptake of colposcopy. S Afr J Obstet Gynaecol. 2013;19(3):81–5. doi:10.7196/SAJOG.388

26. Katz IT, Butler LM, Crankshaw TL, Wright AA, Bramhill K, Leone DA, et al. Cervical abnormalities in South African women living with HIV with high screening and referral rates. J Glob Oncol. 2016;2(6):375–80. doi:10.1200/JGO.2015.002469

27. Firnhaber C, Van Le H, Pettifor A, Schulze D, Michelow P, Sanne IM, et al. Association between cervical dysplasia and human papillomavirus in HIV seropositive women from Johannesburg South Africa. Cancer Causes Control. 2010;21(3):433–43. doi:10.1007/s10552-009-9475-z

28. National Department of Health SA. National guideline for cervical cancer screening programme. Pretoria; 2000.

29. Sankaranarayanan R, Budukh AM, Rajkumar R. Effective screening programmes for cervical cancer in low- and middle-income developing countries. Bull World Health Organ. 2001;79(10):954–62. doi:10.1590/S0042-96862001001000009

30. Miller JW, Hanson V, Johnson GD, Royalty JE, Richardson LC. From cancer screening to treatment: Service delivery and referral in the National Breast and Cervical Cancer Early Detection Program. Cancer. 2014;120(16):2549–56. doi:10.1002/cncr.28823

31. Jassat W, Kawonga M. An evaluation of the cervical screening programme in Johannesburg Metro district, Gauteng province [Master's thesis]. University of the Witwatersrand; 2010.

32. Maimela G, Nene X, Mvundla N, Sawry S, Smith T, Rees H, et al. The impact of decentralising colposcopy services from tertiary-level to primary-level care in inner-city Johannesburg, South Africa: a before and after study. BMJ Open. 2019;9(e024726). doi:10.1136/bmjopen-2018-024726

33. Jordaan S, Michelow P, Simoens C, Bogers J. Challenges and progress of policies on cervical cancer in South Africa. Heal Care Curr Rev. 2017;5(2). doi:10.4172/2375-4273.1000188

34. Njagi SK, Ngure K, Mwaniki L, Kiptoo M, Mugo NR. Prevalence and correlates of cervical squamous intraepithelial lesions among HIV-infected and uninfected women in central Kenya. Pan Afr Med J. 2021;39(44). doi:10.11604/pamj.2021.39.44.27182

35. Wang SX, Wu JL, Zheng RM, Xiong WY, Chen JY, Ma L, et al. A preliminary cervical cancer screening cascade for eight provinces rural Chinese women: a descriptive analysis of cervical cancer screening cases in a 3-stage framework. Chin Med J (Engl). 2019;132(15):1773–9. doi:10.1097/CM9.0000000000000353

36. Holme F, Jeronimo J, Maldonado F, Camel C, Sandoval M, Martinez-Granera B, et al. Introduction of HPV testing for cervical cancer screening in Central America: The Scale-Up project. Prev Med (Baltim). 2020;135(January). doi:10.1016/j.ypmed.2020.106076

37. Spence AR, Goggin P, Franco EL. Process of care failures in invasive cervical cancer: Systematic review and meta-analysis. Prev Med (Baltim). 2007;45(2–3):93–106. doi:10.1016/j.ypmed.2007.06.007

38. Finocchario-Kessler S, Wexler C, Maloba M, Mabachi N, Ndikum-Moffor F, Bukusi E. Cervical cancer prevention and treatment research in Africa: a

systematic review from a public health perspective. BMC Womens Health. 2016;16(29). doi:10.1186/s12905-016-0306-6

39. Makura CB, Schnippel K, Michelow P, Chibwesha CJ, Goeieman B, Jordaan S, et al. Choropleth mapping of cervical cancer screening in South Africa using healthcare facility-level data from the national laboratory network. AIMS Public Heal. 2016;3(4):849–62. doi:10.3934/publichealth.2016.4.849

40. World Health Organization. Consolidated strategic information guidelines for HIV in the health sector. Geneva: World Health Organization; 2015.

41. Gu L, Baxter R, Vickers D, Rainsford C. Record linkage: current practice and future directions. CSIRO Math Inf Sci Tech. 2003;3.

42. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. Annu Rev Public Health. 2011;32:91–108. doi:10.1146/annurev-publhealth-031210-100700

43. Freire SM, Almeida RT de, Cabral MDB, Bastos E de A, Souza RC, Silva MGP da. A record linkage process of a cervical cancer screening database. Comput Methods Programs Biomed. 2012;108:90–101. doi:10.1016/j.cmpb.2012.01.007

44. Kabudula CW, Clark BD, Gómez-Olivé FX, Tollman S, Menken J, Reniers G. The promise of record linkage for assessing the uptake of health services in resource constrained settings: A pilot study from South Africa. BMC Med Res Methodol. 2014;14(71):1–10. doi:10.1186/1471-2288-14-71

45. Department of Cooperative Governance and Traditional Affairs SA. Profile and analysis: City of Johannesburg Metropolitan. 2020.

46. Massyn N, Padarath A, Peer N, Day C. District health barometer 2016/17. Durban: Health Systems Trust; 2017.

47. Stata Corp. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC; 2021.

48. Python Software Foundation. Python. Beaverton: Python Software Foundation; 2021.

49. Gauteng Department of Health. List of Johannesburg Clinics 2020. 2020.

50. ANOVA Health Institute, Gauteng Department of Health. City of Johannesburg community clinics. 2021.

51. Haque S, Mengersen K. Assessing the accuracy of individual link with varying block sizes and cutoff values using MaCSim approach. Commun Stat - Theory Methods. 2020; doi:10.1080/03610926.2020.1857771

52. Haque S, Mengersen K. Assessing the accuracy of individual link with varying block sizes and cut-off values using MaCSim approach [Internet]. Arxiv. 2020 [cited 2022 Apr 9]. Available from: https://arxiv.org/pdf/2003.05686.pdf

53. Jaro MA. Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. J Am Stat Assoc. 1989;84(406):414–20.

doi:10.1080/01621459.1989.10478785

54. Kabudula CW, Joubert JD, Tuoane-Nkhasi M, Kahn K, Rao C, Gómez-Olivé FX, et al. Evaluation of record linkage of mortality data between a health and demographic surveillance system and national civil registration system in South Africa. Popul Health Metr. 2014;12(23). doi:10.1186/s12963-014-0023-z

55. SNOMED International. SNOMED CT [Internet]. 2022 [cited 2022 Jul 13]. Available from: https://www.snomed.org

56. Rahman R, Clark MD, Collins Z, Traore F, Dioukhane EM, Thiam H, et al. Cervical cancer screening decentralized policy adaptation: an African rural-context-specific systematic literature review. Glob Health Action. 2019;12. doi:10.1080/16549716.2019.1587894

57. Momberg M, Botha MH, Van der Merwe FH, Moodley J. Women's experiences with cervical cancer screening in a colposcopy referral clinic in Cape Town, South Africa: a qualitative analysis. BMJ Open. 2017;7(e013914):1–6. doi:10.1136/bmjopen-2016-013914

58. Khanna N, Phillips MD. Adherence to care plan in women with abnormal papanicolaou smears: A review of barriers and interventions. J Am Board Fam Pract. 2001;14:123–30.

59. Eggleston KS, Coker AL, Das IP, Cordray ST, Luchok KJ. Understanding barriers for adherence to follow-up care for abnormal pap tests. J Women's Heal. 2007;16(3):311–30. doi:10.1089/jwh.2006.0161

60. UNAIDS. HIV estimates with uncertainty bounds 1990-Present [Internet]. 2022 [cited 2022 Mar 25]. Available from: https://www.unaids.org/en/resources/documents/2021/HIV_estimates_with_uncertainty_bounds_1990-present

61. Duerr A, Kieke B, Warren D, Shah K, Burk R, Peipert J, et al. Human papillomavirus-associated cervical cytologic abnormalities among women with or at risk of infection with human immunodeficiency virus. Am J Obstet Gynecololgy. 2001;184(4):584–90. doi:10.1067/mob.2001.111791

62. National Department of Health SA. Minister Aaron Motsoaledi: Launch of Human Papilloma Virus vaccine [Internet]. 2014 [cited 2022 Mar 25]. Available from: https://www.gov.za/national-minister-health-dr-aaron-motsoaledis-speech-launch-human-papilloma-virus-hpv-vaccine

63. Lowy DR, Schiller JT. Reducing HPV-associated cancer globally. Cancer Prev Res. 2012;5(1):18–23. doi:10.1158/1940-6207.CAPR-11-0542

64. World Health Organization. Accelerating the elimination of cervical cancer as a global public health problem: draft decision proposed by Australia, Brazil, Canada, Colombia, Ecuador, India, Kenya, Monaco, Mozambique, New Zealand, Peru, Republic of Korea, South Africa, Sri Lanka, UK. Executive Board, 144. 2019. doi:10.1016/S1470-2045(19)30072-5

65. Amponsah-Dacosta E, Blose N, Nkwinika VV, Chepkurui V. Human

Papillomavirus vaccination in South Africa: Programmatic challenges and opportunities for integration with other adolescent health services? Front Public Heal. 2022;10. doi:10.3389/fpubh.2022.799984

66. Mayo M, Potugari B, Bzeih R, Scheidel C, Carrera C, Shellenberger RA. Cancer screening during the COVID-19 pandemic: a systematic review and meta-analysis. Mayo Clin Proc Innov Qual Outcomes. 2021;5(6):1109–17. doi:10.1016/j.mayocpiqo.2021.10.003

67. Burger R, Day C, Deghaye N, Nkonki L, Rensburg R, Smith A, et al. Examining the unintended consequences of the COVID-19 pandemic on public sector health facility visits: the first 150 days. Cape Town: National Income Dynamics Study–Coronavirus Rapid Mobile Survey (NIDS-CRAM); 2020.

68. Wentzensen N, Clarke MA, Perkins RB. Impact of COVID-19 on cervical cancer screening: challenges and opportunities to improving resilience and reduce disparities. Prev Med (Baltim). 2021;151. doi:10.1016/j.ypmed.2021.106596

69. Silveira DP da, Artmann E. Accuracy of probabilistic record linkage applied to health databases: systematic review. Rev Saude Publica. 2009;43(5).

70. Drummond JL, Were MC, Wools-kaloustian K, Arrossi S. Cervical cancer data and data systems in limited-resource settings: Challenges and opportunities. Int J Gynecol Obs. 2017;138:33–40. doi:10.1002/ijgo.12192

8. Appendices

Appendix A – Plagiarism Declaration

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG | FACULTY OF HEALTH SCIENCES

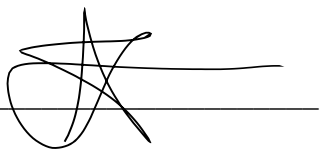
PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

I, Amilcar Juggernath (student number 358266), am a student registered for the degree of MMed (Public Health Medicine) in the academic year 2022.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- I have included a report from *Turnitin* (Appendix B) indicating the level of plagiarism in my research document.

Signature of candidate: 

Name of candidate: Amilcar Juggernath

Signed on the 6th day of August 2022 in Johannesburg

Appendix B – Turnitin report

MMed Research Report_turnitin.pdf

by Amilcar Juggernath

Submission date: 07-Aug-2022 03:11PM (UTC+0200)
Submission ID: 1879724592
File name: MMed_Research_Report_turnitin.pdf (1.07M)
Word count: 17907
Character count: 97783

MMed Research Report_turnitin.pdf

ORIGINALITY REPORT

11%	9%	8%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	hdl.handle.net Internet Source	1%
2	wiredspace.wits.ac.za Internet Source	1%
3	www.science.gov Internet Source	<1%
4	publications.iarc.fr Internet Source	<1%
5	bmcpublichealth.biomedcentral.com Internet Source	<1%
6	www.ncbi.nlm.nih.gov Internet Source	<1%
7	open.uct.ac.za Internet Source	<1%
8	journals.plos.org Internet Source	<1%
9	www.researchgate.net Internet Source	<1%

Appendix C – Ethics Approval Certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG	HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

Office of the Deputy Vice-Chancellor (Research and Postgraduate Affairs)

TO: Drs AM Juggernath & C Kabudula; Prof P Michelow
School of Public Health
Department of Community Health
Medical School
University

E-mail: amilcar.juggernath@gmail.com

CC: Supervisor: Dr M Kawonga
<Mary.Kawonga@wits.ac.za>
and <HREC-Medical Research Office@wits.ac.za>

FROM: Mr Iain Burns
Human Research Ethics Committee (Medical)
Tel: 011 717 1252

E-mail: Iain.Burns@wits.ac.za

DATE: 2021/05/03

REF: R14/49

PROTOCOL NO: **M200648** (This is your ethics application reference number. Please quote it in all enquiries, oral or written, relating to this study.)

PROJECT TITLE: *Assessing cervical precancer treatment rate among women utilising the Johannesburg public sector through record linkage of laboratory data*

Please find attached the Clearance Certificate for the above project. I hope it goes well and that an article in a recognized publication comes out of it. This will reflect well on your professional standing and contribute to Government funding of the University.



MSWorks2000/Iain0007/Clearscan.wps

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R49 Drs AM Juggernath & C Kabudula; Prof P Michelow

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
CLEARANCE CERTIFICATE NO. M200648

NAME: **(Principal Investigator)**	Drs AM Juggernath & C Kabudula; Prof P Michelow
DEPARTMENT:	School of Public Health Department of Community Health Medical School University
PROJECT TITLE:	*Assessing cervical precancer treatment rate among women utilising the Johannesburg public sector through record linkage of laboratory data*
DATE CONSIDERED:	2020/06/26
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr M Kawonga
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	2021/05/03

This Clearance Certificate is valid for 5 years from the date of approval. An extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office secretariat on the 3rd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand, Johannesburg.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated from the research protocol as approved, I/we undertake to submit details to the Committee. **I agree to submit a yearly progress report**. When a funder requires annual re-certification, the application date will be one year after the date when the study was initially reviewed. In this case, the study was initially reviewed in **June** and therefore reports and re-certification will be due in the month of **June** each year. Unreported changes to the study may invalidate the clearance given by the HREC (Medical).

Signature of Principal Investigator

05/05/2021
Date

Appendix D

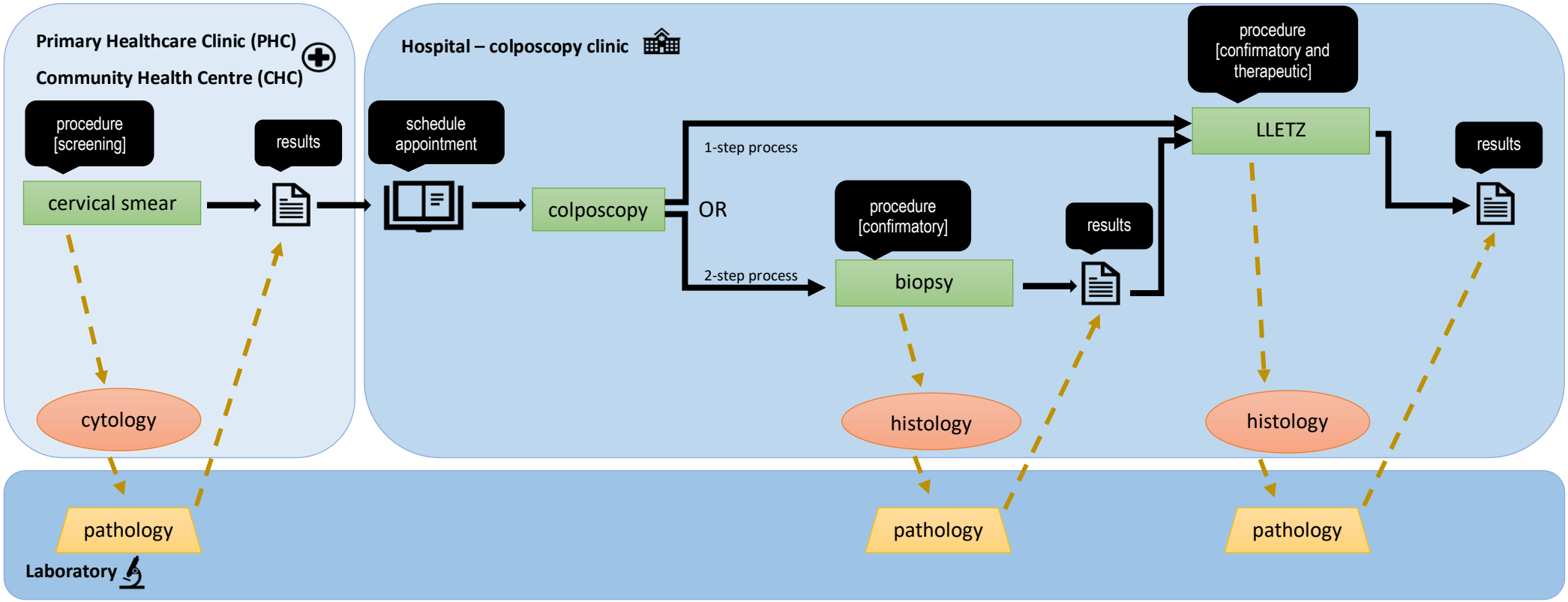


Figure 8.1: Patient flow through the cervical cancer screening cascade

Appendix E

Table 8.1: Summary of the variables used for components of data processing

Data extraction *Variables extracted from the cytology and histology datasets*	**Record linkage** *Variables used to perform record linkage*	**Data anonymisation** *Variables removed to ensure anonymity*
▪ Name ▪ Surname ▪ National ID number ▪ Hospital/clinic number ▪ Address ▪ Date of birth ▪ Age ▪ Race ▪ HIV status ▪ Specimen collection date ▪ Specimen type ▪ Specimen result ▪ Facility name ▪ Facility type ▪ Facility subdistrict	▪ Name ▪ Surname ▪ Date of birth ▪ Age	▪ Name ▪ Surname ▪ National ID number ▪ Hospital/clinic number ▪ Address ▪ Date of birth ▪ Race

Table 8.2: Description of summary variables

Variable name	**Criteria for coding of variable as *Yes***
Cytology result HSIL[a]	Screening test result is *HSIL*
First follow-up done	i) Cytology result HSIL = *Yes* AND ii) At least 1 histology procedure done
First follow-up required	i) Cytology result HSIL = *Yes* AND ii) Histology result of first follow-up = *HSIL* OR First follow-up done = *No*
First follow-up appropriate for the 1-step process	1st Histology procedure = *LEEP*,[b] *LLETZ*,[c] *Hysterectomy* or *Cone biopsy*
First follow-up appropriate for the 2-step process	1st Histology procedure = *Cervical biopsy* or *Punch biopsy*
Second follow-up required	i) First follow-up appropriate for the 2-step process = *Yes* AND ii) 1st Histology result = *HSIL*
Second follow-up done	i) Second follow-up required = *Yes* AND ii) 2 histology procedures done
Second follow-up appropriate for the 2-step process	2nd Histology procedure: *LEEP, LLETZ, Hysterectomy* or *Cone biopsy*
Treatment done	i) First follow-up appropriate for the 1-step process = *Yes* OR ii) Second follow-up appropriate for the 2-step process = *Yes*

a: High-grade intraepithelial lesion. b: Large loop excision of the transformation zone. c: Loop electrosurgical excision procedure.

Table 8.3: Description of variables in the final dataset

Variable name	Variable type		Values	Comment
Unique patient identity	Identifier	Discreet	*1 – 67 208*	Random number generated
Age	Exposure	Discreet	Natural numbers (years)	Source: Cytology dataset. If missing, calculated from Date of birth
Age group	Exposure	Ordinal	*15-29*, *30-39*, *40-49*, *50-59*, *60+*, *Missing*	Age coded into groups
HIV status	Exposure	Nominal	*Positive*, *Negative*, *Unknown*	Source: Cytology dataset
Screening test facility type	Exposure	Nominal	*PHC*[a], *CHC*[b], *Hospital*, *Other*	Source: Cytology dataset. Coded based on Facility name
Screening facility subdistrict	Exposure	Nominal	*A*, *B*, *C*, *D*, *E*, *F*, *G*	Source: Cytology dataset. Coded based on Facility name
Screening result of HSIL	Outcome	Binary	*Yes*, *No*	Source: Cytology dataset.
Treatment required	Outcome	Binary	*Yes*, *No*	Same as the summary variable First follow-up required
First follow-up done	Outcome	Binary	*Yes*, *No*	Same as summary variable described in Table 8.2.
Confirmatory test done	Outcome	Binary	*Yes*, *No*	Same as summary variable First follow-up appropriate for 2-step process = *Yes*
Treatment done	Outcome	Binary	*Yes*, *No*	Same as summary variable described in Table 8.2.
Management step	Outcome	Nominal	*0*, *1*, *2*	Coded based on summary variables. *0* if First follow-up done = *No* *1* if Treatment done = *Yes* AND Confirmatory test done = *No* *2* if Confirmatory test done = *Yes*
Time to follow-up	Outcome	Discreet	Natural numbers (days)	Source: Specimen collection dates from cytology and histology datasets. Calculated if First follow-up done = *Yes* *Date of 1st histology procedure – Date of screening test*
Time to confirmatory test	Outcome	Discreet	Natural numbers (days)	Source: Specimen collection dates from cytology and histology datasets. Calculated if Confirmatory test done = *Yes* *Date of 1st histology procedure – Date of screening test*
Time to treatment	Outcome	Discreet	Natural numbers (days)	Source: Specimen collection dates from cytology and histology datasets. Calculated if Treatment done = *Yes* If Management step = *1*: *Date of 1st histology procedure – Date of screening test* If Management step = *2*: *Date of 2nd histology procedure – Date of screening test*
Time from confirmatory test to treatment	Outcome	Discreet	Natural numbers (days)	Source: Specimen collection dates from cytology and histology datasets. Calculated if Management step = *2* *Date of 2nd histology procedure – Date of 1st histology procedure*

a: Primary healthcare clinic. b: Community healthcare centre