BMJ Open

Cohort Profile: The South African National Health Laboratory Service (NHLS) National HIV Cohort

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-066671
Article Type:	Cohort profile
Date Submitted by the Author:	14-Jul-2022
Complete List of Authors:	MacLeod, William; Boston University School of Public Health, Department of Global Health; University of the Witwatersrand, Health Economics and Epidemiology Research Office Bor, Jacob; Boston University School of Public Health, Departments of Global Health and Epidemiology; Boston University School of Public Health, Department of Global Health Candy, Sue; National Institute for Communicable Diseases Maskew, Mhairi; University of the Witwatersrand, Health Economics and Epidemiology Research Office Fox, Matthew; Boston University School of Public Health, Departments of Global Health and Epidemiology; University of the Witwatersrand, Health Economics and Epidemiology Research Office Bulekova, Katia; Boston University Brennan, Alana; Boston University School of Public Health, Departments of Global Health and Epidemiology; University of the Witwatersrand, Health Economics and Epidemiology Research Office Potter, James; Boston University School of Public Health, Department of Global Health Nattey, Cornelius; University of the Witwatersrand, Health Economics and Epidemiology Research Office Onoya, Dorina; University of the Witwatersrand, Health Economics and Epidemiology Research Office Mlisana, Koleka; National Health Laboratory Service Stevens, Wendy; University of the Witwatersrand, Department of Molecular Medicine and Haematology; National Health Laboratory Service Carmona, Sergio; University of the Witwatersrand, Department of Molecular Medicine and Haematology
Keywords:	HIV & AIDS < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Cohort Profile: The South African National Health Laboratory Service (NHLS) National HIV Cohort

William B. MacLeod[1,2,†,*], Jacob Bor[1,2,3,†], Sue Candy[4], Mhairi Maskew[2], Matthew P. Fox[1,2,3], Katia Bulekova[5], Alana T. Brennan[1,2,3], James Potter,[1] Cornelius Nattey[2], Dorina Onoya[2], Koleka Mlisana[6], Wendy Stevens[6,7], Sergio Carmona[7]

Affiliations:

[1]Department of Global Health, Boston University School of Public Health, Boston, MA, USA

[2]Health Economics and Epidemiology Research Office, University of the Witwatersrand, Johannesburg, South Africa

[3]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

[4]National Institute of Communicable Diseases, Sandringham, South Africa

[5]Research Computing Services, Boston University, USA

[6]National Health Laboratory Service, Sandringham, South Africa

[7]Department of Molecular Medicine and Haematology, University of the Witwatersrand School of Pathology, Johannesburg, South Africa

*correspondence to: wmacleod@bu.edu, 801 Massachusetts Avenue, 3rd Floor, Boston, MA 02118

†equal contributions

Short title: Cohort Profile: South Africa's NHLS National HIV Cohort

Abstract

- **Purpose**: South Africa's National Health Laboratory Service (NHLS) National HIV Cohort was established in 2015 to facilitate monitoring, evaluation, and research on South Africa's National HIV Treatment Program. In South Africa, 84.8% of PLHIV know their HIV status, 70.7% are on ART, and 87.4% are virologically suppressed.
- **Participants**: The NHLS National HIV Cohort includes the laboratory data of nearly all patients receiving HIV care in the public sector since April 2004. Patients are included in the cohort if they have received a CD4 count or HIV RNA viral load test. Using an anonymized unique patient identifier that we developed and validated to linked test results, patients are observed prospectively through their laboratory results as they receive HIV care and treatment. Patients in HIV care are seen for laboratory monitoring every 6 to 12 months. Data collected include age, sex, facility location, and test results for CD4 counts, viral loads, and laboratory tests used to screen for potential treatment complications.
- **Findings to date**: From April 2004 through April 2018, 63 million CD4 count and viral load tests were conducted at 5,483 facilities. 12.6 million unique patients had at least one CD4 count or viral load, indicating they had accessed HIV care, and 7.1 million patients received a viral load test indicating they had started ART. The creation of NHLS National HIV Cohort has enabled longitudinal research on all lab-monitored patients in South Africa's national HIV program, including analyses of: 1) patient health at presentation; 2) care outcomes such as "CD4 recovery", "retention in care", and "viral re-suppression"; 3) patterns of transfer and re-entry into care; 4) facility-level variation in care outcomes; and 5) impacts of policies and guideline changes.
- **Future plans**: Continuous updating of the cohort, integration with available clinical data, and expansion to include tuberculosis and other lab-monitored comorbidities.

STRENGTHS AND LIMITATIONS

Strengths

- Large size and scope—all public sector patients in South Africa.
- We are able to explore outcomes without the problem of silent transfers.
- The cohort contains patient data prior to the initiation of ART: essential for assessing the impact of policy changes on outcomes both before and after treatment.

Limitations

- It is limited to laboratory results with limited demographic information and no clinical visit or pharmacy information.
- The laboratory data doesn't have a unique patient identifier and the matching techniques we use are not perfect and lead to both over- and under-matching of patient records, which potentially could lead to biased findings.

INTRODUCTION

South Africa has the largest HIV treatment program in the world, with over 4.7 million people currently on antiretroviral therapy (ART).[1] It also has the largest number of people living with HIV but not yet on ART, an estimated 3 million people.[1] Based on the fifth South African national HIV sero-behavioural survey conducted in 2017, 84.8% of people living with HIV were aware of their status, 70.7% were on ART, and 87.4% were virologically suppressed.[2] With such high stakes, it is critical to evaluate the impact of the ART program and improve outcomes nationally.

South Africa's National Health Laboratory Service (NHLS) National HIV Cohort was established to facilitate monitoring, evaluation, and research on South Africa's National HIV Care and Treatment Program. The NHLS National HIV Cohort includes the laboratory data of nearly all patients receiving HIV care in the public sector since 2004. Using an anonymized unique patient identifier that we developed and validated, patients can be followed longitudinally through their laboratory results as they progress through the HIV care and treatment cascade. This open, prospective cohort can be used to evaluate changes in HIV treatment guidelines, to assess trends in patient outcomes across space and time, to determine patterns of patient transfer, and to identify areas where outcomes lag behind.

South Africa has conducted a series of five national population-based HIV biomarker surveys since 2002 and a large number of HIV treatment cohorts have been established in South Africa. Analyses of these surveys and cohorts have contributed to our knowledge of HIV treatment outcomes.[2–12] However, the national surveys are only conducted every 4 or 5 years, and the existing cohorts while continuous, also have some drawbacks: they are not nationally

representative, most are in urban areas, and reflect long-standing research collaborations that have resulted in better-than-average patient care. Further, existing cohorts do not systematically track patients lost to follow-up in order to assess re-engagement in care. TIER.net, the country's HIV patient monitoring platform, is national in scope but has limited ability to track patients across sites[13]. While collecting data on all HIV patients nationally would be prohibitively expensive as a research endeavor, routine laboratory data collected in South Africa's HIV program can be leveraged to construct a cohort for monitoring, evaluation, and research.[14,15]

Since the beginning of the national HIV care and treatment program, NHLS provided nearly all laboratory testing for the public-sector program (with the exception of KwaZulu Natal Province which was not fully integrated into NHLS until April 2010). Laboratory testing has been used for CD4 and VL monitoring, for confirmatory HIV diagnostic testing, and to identify potential ART complications and contraindications. Specimens for testing are collected from patients at their care facility and sent to one of 16 laboratories for processing and testing. Results are generated by testing instruments in real time and sent to the NHLS Corporate Data Warehouse (CDW) which manages the distribution of results in real-time. All laboratory test results are maintained within the CDW. All NHLS HIV labs are accredited by the South African National Accreditation System (SANAS). The NHLS supports the labs with training, site visits, and on-site audits. Standard operating procedures have been developed and distributed to ensure standardisation across testing facilities. Laboratory performance on test volumes and turn-around-times are routinely monitored[16].

The NHLS National HIV Cohort has advantages over clinical cohorts. First, the data come directly from the source testing platforms and are less vulnerable to data entry errors that occur when extracting patient charts into databases. Second, the data are lower cost as they are collected for routine patient care. Third, the cohort offers a system-wide perspective on the national program in which transfers, drop-out, and re-entry into care can be observed, enabling evidence generation around national policy decisions. Fourth, the dataset size means that evaluations are robust and can be used to compare facilities, geographic areas, and demographic sub-groups. Fifth, because the data reflect all lab-monitored HIV patients receiving care in the public sector, the cohort reflects public-sector care-seeking patterns.

COHORT DESCRIPTION

Participants in the cohort

The NHLS National HIV Cohort includes all patients presenting for HIV care or treatment and receiving CD4 and/or HIV viral load (VL) monitoring at nearly all government health facilities in South Africa, from April 2004 through April 2018. During this period, 63 million CD4 and VL tests were conducted at 5,483 facilities.

A longitudinal cohort requires a patient identifier in order to follow patients over time. National ID numbers were only available for 2% of the specimens. However, the database contained sufficient information to link laboratory records to individual patients using probabilistic matching techniques. We developed a record linkage algorithm to identify unique patients in

the NHLS database combining elements of probabilistic linkage approaches with concepts from network analysis.[17–19] The methodology is reported in detail elsewhere.[20] In the absence of a true gold standard, we constructed manually-coded datasets for training and validation. We manually reviewed 58,905 candidate matches of 1,000 randomly sampled laboratory records. Relative to these manually-coded data, the algorithm had 93.7% sensitivity, indicating that the algorithm identified all but 6.3% of true matches. The algorithm had 98.6% positive predictive value, indicating that just 1.4% of algorithm-assigned matches were incorrect. Additionally, for those specimens linked to a National ID number, the algorithm correctly identified pairs of records with the same National ID number with 98.5% sensitivity. No linkage is perfect. For example, patients may present to clinics with different names and dates of birth, in which case linkage would be impossible. However, we developed a high-performing linkage approach using the available data in the NHLS database. By creating a validated patient identifier, the linkage enabled analysis of the NHLS database as an HIV Cohort. The cohort was deidentified after linkage, and all analyses are conducted using anonymized data.

Our record linkage of the CDW data was conducted on 152.5 million total laboratory results including 63 million CD4 counts and VL from April 2004 through April 2018. The algorithm assigned these laboratory results to 12,603,979 unique patients with at least one CD4 count or VL, indicating that these individuals had accessed HIV care. **Table 1** shows the demographics of the cohort. The cohort was 64.3% female, 33.7% male, and 2.0% unknown gender. The majority of individuals were 20-49 years old (8.4% had an unknown date of birth), with patients entering the cohort at a median age of 32 years (IQR 25-41 years). In the most recent 12 months of the cohort, HIV patients had laboratory tests at 4,751 facilities within South Africa (**Figure 1)**.

Table 1. Sex and age of patients entering into HIV Care from May 2004 to April 2018.

Parameter (% (n))	May 2004 to April 2008 (n=2,042,315)	May 2008 to April 2013 (n=5,752,084)	May 2013 to April 2018 (n=4,809,580)	May 2004 to April 2018 (n=12,603,979)
Male	31.4% (641,534)	32.4% (1,864,287)	36.2% (1,741,698)	33.7% (4,247,519)
Female	66.3% (1,354,605)	65.9% (3,792,924)	61.6% (2,962,984)	64.3% (8,110,513)
Unknown	2.3% (46,176)	1.6% (94,873)	2.2% (104,898)	2.0% (245,947)
Age Group (years)				
0-4	3.5% (71,011)	2.8% (162,706)	2.3% (111,735)	2.7% (345,452)
5-9	1.4% (28,358)	1.4% (78,910)	1.0% (49,771)	1.2% (157,039)
10-14	0.6% (11,889)	1.0% (60,280)	1.1% (54,329)	1.0% (126,498)
15-19	2.8% (57,167)	3.8% (216,482)	4.4% (211,731)	3.9% (485,380)
20-24	11.2% (229,680)	11.5% (662,334)	11.5% (551,815)	11.5% (1,443,829)
25-29	17.7% (360,484)	16.8% (965,872)	15.6% (751,784)	16.5% (2,078,140)
30-34	18.3% (374,633)	15.9% (917,420)	15.7% (755,664)	16.2% (2,047,717)
35-39	14.1% (287,302)	13.3% (763,369)	12.4% (597,389)	13.1% (1,648,060)
40-44	9.9% (202,827)	9.2% (530,075)	9.3% (448,095)	9.4% (1,180,997)
45-49	6.4% (131,450)	6.5% (375,617)	6.4% (310,034)	6.5% (817,101)
50-54	3.9% (79,764)	4.2% (242,140)	4.6% (221,419)	4.3% (543,323)
55-59	2.0% (41,344)	2.5% (144,966)	3.0% (144,236)	2.6% (330,546)
60-64	0.9% (17,757)	1.3% (73,998)	1.8% (85,267)	1.4% (177,022)
65+	0.4% (8,772)	0.6% (33,536)	0.8% (40,675)	0.7% (82,983)

Age is age at first entry into HIV Care--at first CD4 or first VL test.

Of the 12.6 million patients in the cohort, 7.1 million ever initiated ART. To estimate the number of patients currently receiving ART, we limited the cohort to the last 18 months from

November 2016 through April 2018. We estimate there were 4.4 million patients receiving ART as evidenced by VL monitoring during that time period, similar to the 4.7 million on ART UNAIDS estimated for 2018.[21] The number of patients entering HIV care, initiating HIV treatment, and actively on HIV treatment by year is shown in **Table 2.** Since 2004-05 the number with a first VL (a proxy for initiating HIV treatment, in South Africa, only patients on ART are monitored virologically) has grown from 27,937 to 913,604 per year.

We found 4.2 million CD4 and 592,261 VL tests that were unlinked to other tests. Patients in the cohort that had just a single CD4 and no VL include the large number of patients who present clinically with HIV, have blood drawn for a CD4, but do not seek follow-up care. Still, some of these singleton lab results may be "stray" lab results that should have been linked to a patient. Singleton VL represent 8.3% of all VL tests.

Participant Follow-Up

In South Africa, as in many countries, HIV care and treatment decisions are informed by routine laboratory monitoring, as specified in standardized national guidelines.[22] While the cohort lacks information on clinical visits, pharmacy data, or ART regimen, it is nevertheless possible to follow important care and treatment decisions through the labs.

Within the cohort, follow-up is determined by the dates of CD4 and VL monitoring, as well as screening for potential treatment complications, known as the ART work-up. Laboratory-based monitoring of CD4 and VL has been a standard component of national HIV care and treatment

Table 2. Number of patients entering into HIV Care, receiving first HIV VL test, on ART, and HIV VL suppressed by 12-month period from May 2004 to April 2018.

Year (12-month period May 1 to April 30)	Number of patients entering into HIV Care[a]	Number of patients receiving first HIV VL test[b]	Number of patients on ART[c]	Percent HIV VL suppressed[d]
2004-05	118,829	27,937	24,208	31.1
2005-06	219,942	111,865	120,796	39.4
2006-07	302,510	176,979	272,078	43.4
2007-08	344,935	223,188	420,542	50.3
2008-09	410,563	283,929	609,729	55.4
2009-10	479,406	249,971	716,555	65.7
2010-11	915,898	525,457	1,066,016	72.3
2011-12	713,101	582,175	1,465,428	72.2
2012-13	626,296	672,543	1,875,968	74.0
2013-14	619,396	715,786	2,316,959	74.6
2014-15	629,246	785,503	2,801,889	77.1
2015-16	666,492	952,256	3,429,148	79.6
2016-17	629,754	919,229	3,951,406	79.4
2017-18	461,076	913,604	4,419,017	83.2

[a]Numbers of patients into HIV Care is the number of patients with a first CD4 or first VL.

[b]Number of patients receiving first HIV VL test is the number of first of VL tests

[c]Number of patients on ART is the number of patients who had a VL test in the 18-month period from November 1, (year – 1) to April 30, (year + 1).

[d]Percent of HIV VL suppressed are 100*(the number of VL < 400 cp/ml/Total number of VL tests) during the same 18-month period. Patients who did not receive a viral load test are not included.

guidelines during the period of study and a routine part of clinical care. During the period of observation, patients testing positive for HIV at a health facility have blood drawn for a CD4 on the same day or shortly thereafter. Thus, a patient's first CD4 test is a reasonable proxy for the date when a patient was diagnosed as HIV-infected within the health system and presented to a health facility. At the start of the HIV care program, VL tests were conducted at treatment initiation, and then every six-months. Since 2010, VL monitoring has been limited to 6- and 12-months after treatment initiation and annually thereafter. **Table 3** lists the evolving criteria for ART initiation and ART laboratory monitoring in South Africa from 2004 to 2016.

Within the cohort, for patients on treatment, a total of 26.7 million follow-up VL have been conducted which corresponds to a median (IQR) of 2 (1-5) VL per person. This translates to 23,381,315 total person years of follow-up on HIV treatment for a median (IQR) of 2.1 (1.0-5.0) years per person.

Variables measured

The variables in the NHLS National HIV Cohort are described in **Table 4** along with percent completeness. The cohort includes the type of test, tests results, test date, and geographic information along with patient's date of birth and sex. **Table 5** is a frequency listing of test type by three time periods. The tests included in the cohort includes all CD4, VL, HIV confirmatory tests (PCR/Elisa), and ART-workup labs for patients receiving HIV care.

Despite the lack of clinical and pharmacy data, the cohort can be used to generate a wealth of information about the national HIV program. Information on facility can be linked to facility geocodes, which can be mapped (**Figure 1**), aggregated to the local municipality, district, and

Table 3. Evolving criteria for ART initiation and ART monitoring in South Africa from 2004 to 2016.

Year	Guidelines	ART Initiation Criteria/CD4 and VL monitoring
2004	National ARV Treatment Guidelines (2004)[23]	*Criteria for ART Initiation* CD4 <200 cells/mm3 irrespective of stage or WHO Stage IV AIDS-defining illness, irrespective of CD4 count and Patient expresses willingness and readiness to take ART adherently *ART Monitoring* CD4: Staging and 6 monthly VL: Staging and 6 monthly
2010	The South African Antiretroviral Treatment Guidelines 2010[24]	*Criteria for ART Initiation* CD4 count <200cells/mm3 irrespective of clinical stage or CD4 count <350cells/mm3 in patients with TB/HIV or Pregnant women or WHO stage IV irrespective of CD4 count or MDR/XDR irrespective of CD4 *Require fast track (i.e. ART initiation within 2 weeks of being eligible)* Pregnant women eligible for lifelong ART; Patients with very low CD4 (<100); Stage 4, CD4 count not yet available, or; MDR/XDR TB *ART Monitoring* CD4: Staging, 6 and 12 months and annually thereafter VL: 6 and 12 months and annually thereafter
2011	Circular on New Criteria for Initiation of Adults on ART at CD4 Count of 350 cells/mm3 and below. August 26, 2011	Baseline CD4 for initiation at CD4 < 350 cells/mm3
2012	Accelerating Access to ART Services and Uptake (circular) April 17, 2012	*Criteria for ART Initiation* CD4 count <350 cells/mm3 irrespective of clinical stage or patients with TB/HIV irrespective of CD4 count or Pregnant women or WHO stage IV irrespective of CD4 count or MDR/XDR irrespective of CD4 *Require fast track (i.e. Same day ART initiation)* Pregnant women eligible for lifelong ART; Patients with low CD4 (<200); Stage 4, CD4 count not yet available, or; MDR/XDR TB Screening and treatment of patients with very low CD4 Counts (<100) for cryptococcal infection *ART Monitoring* CD4: Staging, 12 months and annually thereafter VL: 6 and 12 months and annually thereafter
2013	The South African Antiretroviral Treatment Guidelines 2013[25]	*Criteria for ART Initiation* CD4 count <350 cells/mm3 irrespective of WHO clinical stage or Irrespective of CD4 count and All types of TB (In patients with TB drug resistant or sensitive, including extra pulmonary TB) and WHO stage 3 or 4 irrespective of CD4 count *Require fast track (i.e. ART initiation within 7 days of being eligible)* HIV positive women who are pregnant or breast feeding or Patients with low CD4 <200 or Patients with Stage 4, irrespective of CD4 count or Patients with TB/HIV co morbidity with CD4 count < 50 (Patients

		with Cryptococcus meningitis or TB meningitis (defer ART for 4-6 weeks) *ART Monitoring* CD4: Staging and 1 year VL: 6 and 12 months and annually thereafter
2015	National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults 2015[26]	*Criteria for ART Initiation* CD4 count <500 cells/mm3 irrespective of WHO clinical stage or All types of TB or WHO stage 3 or 4 or HBV coinfection irrespective of CD4 count *Immediate Initiation* All HIV-positive pregnant or breastfeeding women, as long as no active TB *Require fast track (i.e. ART initiation within 7 days of being eligible)* Patients with low CD4 <200 or Patients with Stage 4, irrespective of CD4 count *ART Monitoring* CD4: Staging and 1 year VL: 6 and 12 months and annually thereafter
2016	Circular on Implementation of the Universal Test and Treat Strategy for HIV Positive Patients and Differentiated Care for Stable Patients August 22, 2016[27]	*Criteria for ART Initiation* All HIV positive adolescents and adults regardless of CD4 count *ART Monitoring* CD4 Staging and 1 year. VL monitoring at 6 and 12 months and annually thereafter.

Table 4. Variables in the NHLS National HIV Cohort

Variable	Description	Percent completeness of variable
BU_uniq_ID	Unique patient identifier	100%
Episode_no	NHLS Episode identifier	100%
Sex	Sex	98.0%
Age	Age in years at testing date	91.6%
Province	Province of health facility where specimen was taken	99.8%
District	District of health facility	99.8%
Local_Municipality	Local municipality of health facility	99.8%
Facility	NHLS Facility Code of health facility	99.8%
Ward	NHLS Ward Code of ward at health facility	99.8%
Test_name	Type of test (e.g. CD4, VL)	100%
Test_date	Date of test	100%
Test_result	Test result	100%

Table 5. Frequency of all tests included in the NHLS National HIV Cohort in three time periods, May 2004-April 2018

Name of Test	Use in HIV care	2004-08	2008-13	2013-18	All time periods
Alanine Aminotransferase	Measure of liver injury and determining choice of ART	1,970,138	9,301,723	9,176,996	20,448,857
CD4 Count	ART eligibility and disease progression	3,599,594	15,483,376	17,578,415	36,661,385
Serum Cryptococcal Antigen	To detect and prevent cryptococcal meningitis	17	71,374	1,017,903	1,089,294
Creatinine Clearance	Measure of kidney function and determining choice of ART	2,880	6,113,182	25,422,404	31,538,466
HIV Elisa Confirmatory	HIV diagnostic test primarily in infants and young children	663,518	820,322	370,465	1,854,305
HIV Elisa Screening	HIV diagnostic test primarily in	1,129,249	1,448,212	639,528	3,216,989

	infants and young children				
Hemoglobin	Measure of overall health	3,841,090	12,318,399	14,260,748	30,420,237
HIV PCR	HIV diagnostic test primarily used for HIV diagnosis in infants	39,706	109,079	108,212	256,997
HIV RNA Viral Load	Monitoring efficacy of ART therapy	1,120,821	6,716,430	18,922,454	26,759,705
Total		**12,367,013**	**52,382,097**	**87,497,125**	**152,246,235**

provincial levels, or linked to external population-level data such as HIV prevalence, poverty levels, or the population age distribution. Information on test dates can be used to assess longitudinal outcomes such as retention in care (18 months without a monitoring lab) at the patient level, and to assess trends in patient outcomes over time.

Participant attrition

As in other passive surveillance cohorts[5], attrition from care (and its inverse, "retention in care") is a primary outcome of interest. Data are generated through lab monitoring as part of routine clinical care and no data are collected beyond what is clinically indicated. No efforts are made by the research team to retain patients in care nor to actively follow-up patients who have left care. However, accuracy of retention estimates are enhanced by the national perspective of the cohort which is robust to silent transfers. As the study population includes all people who have sought public sector care for HIV in South Africa, attrition occurs only if a person emigrates from South Africa and can no longer seek HIV care.

FINDINGS TO DATE

The ability to link laboratory results to create records of individual patients has enabled 1) longitudinal patient-level epidemiologic research for the complete national HIV care and treatment program including patients not yet on ART; 2) assessment of concepts such as "CD4 recovery", "retention in care", and "viral re-suppression" that require individual-level longitudinal data and monitoring of these concepts at all public-sector facilities nationally; 3) tracking of patients as they seek care at different clinics within the health system and assessing patterns of transfer; 4) evaluation of policies and guideline changes. Finally, linkage with facility

geocodes has enabled integration of the cohort with publicly available data on outcomes and exposures at the facility or district level, including programmatic data (e.g. clinic staffing, size) and population-based data (e.g. population density, poverty, HIV prevalence, and mortality).

Key findings from the NHLS National HIV Cohort include the following:

- *Retention in HIV care is underestimated by not accounting for within system patient transfer* – Estimated retention in HIV care from both the initiating clinic and a national perspective. At the clinic level retention in care was 29.1% by 6 years. However, when accounting for transfers to other clinics, retention in care was 63.3% by 6 years.[28]

- *There is large spatial heterogeneity in the HIV care cascade* – Estimated rates of VL testing and suppression from April 2014 through March 2015 across public facilities. We identified wide spatial variation in VL suppression, ranging from 69 to 82% at the provincial level.[29] **Figure 2** shows a map of district level estimates of VL suppression. The cohort was also used to develop a summary measure of quality of care at the facility level based on patients' longitudinal outcomes in different facilities. Year-to-year, quality was found to be highly correlated within facilities, but varied widely across facilities.[30]

- *A high proportion of patients present with advanced disease* – Documented that a consistently large share (~33%) of patients entered into care with a CD4 <200 cells/uL from 2011 to 2016 despite increased CD4 treatment eligibility standards.[31] Late presentation persisted into the UTT era.[32]

- *A wave of adolescents will require HIV treatment* – Identified adolescents who entered care in childhood (likely perinatally-infected) and adolescents who entered care in their later teenage years (likely infected via sexual transmission). A 10- to 20-fold increase in the numbers of adolescents on ART from 2004-2007 to 2012-2014 was found and resulting in a "wave" of 15-19 year-olds who will require HIV treatment over the next decade due to both the aging of perinatally-infected children into adolescence and increased numbers of adolescent girls seeking HIV care for the first time.[33]

- *Viral monitoring for treatment failure* – VL monitoring is conducted to identify patients at risk for treatment failure, to target these patients with adherence counseling, and to switch them to second-line therapy if needed. Comparing outcomes among patients with VL results just above vs. just below the 1000 copies/mL[3] threshold, we found that while having a VL>1000 increases the probability of follow-up VL monitoring, most patients with elevated VL do not receive monitoring within recommended timelines.[34,35]

STRENGTHS AND LIMITATIONS

The main strength of the NHLS National HIV Cohort is its size and scope. With millions of patients over many years, the cohort can be used to conduct robust evaluations of policy change in South Africa's public-sector treatment programme. It can also be used to take advantage of variation in program implementation (e.g. increases in HIV testing in some areas before others, or implementation of the National Adherence Strategy) to evaluate the impacts of these interventions[36].

Second, the cohort is unique in its ability to explore outcomes without the problem of silent transfers.[37] Silent transfers, patients who move from one facility to another without informing their sending clinic, leads to overestimates of attrition from HIV care and misclassification of outcomes in programmatic evaluations. Because the cohort contains information on the clinic where the lab investigation was conducted, we are able to identify movement between clinics and not misclassify these movements as lost to follow-up.

Third, the cohort contains patient data prior to the initiation of ART: essential for assessing the impact of policy changes on outcomes both before and after treatment—something few clinical cohorts in South Africa capture[38].

A weakness of the cohort is that it is limited to laboratory results with limited demographic information and no clinical visit or pharmacy information. While we have been able to overcome some of these limitations through imputation of ART start dates[39], we do not have data on medication or visit adherence, or clinical diagnoses that would be useful for describing a national program.

Second, the matching techniques we use are not perfect and lead to both over- and under-matching of patient records, which potentially could lead to biased findings. Because the dataset is so large, random error is typically approaching zero in any analysis. This makes systematic errors (like the over- and under-matching) the main source of error in studies using this cohort, a problem that can be explicitly modelled using quantitative bias analysis.[40] We have also assessed the sensitivity of our results to matching parameters and found the results to be quite robust.[28]

Third, as with other clinical databases, data collection is part of routine clinical care. This means that if patients do not present for care, we cannot observe their CD4 or VL. Further, adherence to laboratory monitoring guidelines may vary across facilities, which may contribute to differences in outcomes across facilities.

Finally, the cohort does not have data on death and is unable to link to the National Vital Registration System to obtain mortality data due to the paucity of national ID numbers collected. This means that we are currently not able to describe the impact of interventions and policy changes on mortality. Instead we use other indicators of poor outcomes such as unsuppressed VL, failure to gain CD4 cell count, and attrition from care.

Future Plans

Our future plans for the cohort fall into three categories all with the aim of enhancing the research and clinical value of the cohort: 1) continuous updating of the cohort; 2) integrate the cohort with clinical databases, and 3) expand the scope of cohort to include tuberculosis and other lab-monitored comorbidities.

Collaboration: Access to primary data is subject to restrictions owing to privacy and ethics policies set by the South African Government. Requests for access to the data can be made to the National Health Laboratory Services directly (http://www.aarms.nhls.ac.za) and require a full protocol submission. Inquiries can be made via the Office of Academic Affairs and Research at the National Health Laboratory Service. To find out more about the cohort, contact academic.research@nhls.ac.za.

Ethics clearance: We received ethics clearance from the Wits HREC (M150429) and Boston University IRB (H-31968, H-33442).

Patient and Public Involvement: Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Acknowledgements: Jaco Grobler was instrumental in programming the first iteration of the NHLS unique patient ID which was the genesis for this work. The authors are also grateful for the advice and technical assistance of Naseem Cassim concerning details of the CDW architecture. The authors would like to acknowledge the contributions of the dedicated staff at the National Health Laboratory Service Corporate Data Warehouse for their support and assistance with data extracts and staff of the National Priority Programme for maintaining a service that ensures the generation of good quality assay results.

Funding Acknowledgments: This work was supported by grants R01AI115979, R01-AI152149, K01MH105320, K01DK116929-01, and R01-HD084233 from the National Institutes of Health. WBM and SC were supported by USAID through Cooperative Agreement AID- 674-A-12-0020. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Contributorship: WM, JB, MM, MF, and CN wrote the manuscript. WM, JB, SC, KM, WS and SC conceptualized the creation of the laboratory cohort. WM, JB, SC, MM, KB, AB, JP, and DO

conducted analyses contributing to the creation of the cohort. All authors read and approved the final manuscript.

References

1. UNAIDS. *UNAIDS Data 2019 Annual Report*. Geneva; 2019.

2. Marinda E, Simbayi L, Zuma K, et al. Towards achieving the 90–90–90 HIV targets: Results from the south African 2017 national HIV survey. *BMC Public Health*. 2020;20(1):1-12. doi:10.1186/s12889-020-09457-z

3. Fox MP, Maskew M, Macphail a P, et al. Cohort Profile: The Themba Lethu Clinical Cohort, Johannesburg, South Africa. *Int J Epidemiol*. March 2012:6-8. doi:10.1093/ije/dys029

4. Haas AD, Zaniewski E, Anderegg N, et al. Retention and mortality on antiretroviral therapy in sub-Saharan Africa: Collaborative analyses of HIV treatment programmes: Collaborative. *J Int AIDS Soc*. 2018;21(2):1-7. doi:10.1002/jia2.25084

5. Egger M, Ekouevi DK, Williams C, et al. Cohort Profile: The international epidemiological databases to evaluate AIDS (IeDEA) in sub-Saharan Africa. *Int J Epidemiol*. May 2011. doi:10.1093/ije/dyr080

6. Houlihan CF, Bland RM, Mutevedzi PC, et al. Cohort profile: Hlabisa HIV treatment and care programme. *Int J Epidemiol*. 2011;40(2):318-326. doi:10.1093/ije/dyp402

7. Kluberg SA, Fox MP, LaValley M, Pillay D, Bärnighausen T, Bor J. Do HIV treatment eligibility expansions crowd out the sickest? Evidence from rural South Africa. *Trop Med Int Heal*. 2018;23(9):968-979. doi:10.1111/tmi.13122

8. Bor J, Chiu C, Ahmed S, et al. Failure to initiate HIV treatment in patients with high CD4 counts: evidence from demographic surveillance in rural South Africa. *Trop Med Int Heal*. 2018;23(2):206-220. doi:10.1111/tmi.13013

9. Macleod WB, Maskew M, Jaffray I, Macphail AP, Ive P, Fox MP. The Feasibility of Using Screening Criteria to Reduce Clinic Visits for Stable Patients on Antiretroviral Therapy in South Africa. *J Acquir Immune Defic Syndr*. October 2012. doi:10.1097/QAI.0b013e318278e976

10. Fox MP, Shearer K, Maskew M, et al. Treatment outcomes after 7 years of public-sector HIV treatment. *AIDS*. 2012;26(14):1823-1828. doi:10.1097/QAD.0b013e328357058a

11. Brennan AT, Maskew M, Ive P, et al. Increases in regimen durability associated with the introduction of tenofovir at a large public-sector clinic in Johannesburg, South Africa. *J Int AIDS Soc*. 2013;16:1-12. doi:10.7448/IAS.16.1.18794

12. Haas AD, Johnson LF, Grimsrud A, et al. Extending Visit Intervals for Clinically Stable Patients on Antiretroviral Therapy: Multicohort Analysis of HIV Programs in Southern Africa. *J Acquir Immune Defic Syndr*. 2019;81(4):439-447. doi:10.1097/QAI.0000000000002060

13. Osler M, Hilderbrand K, Hennessey C, et al. A three-tier framework for monitoring antiretroviral therapy in high HIV burden settings. *J Int AIDS Soc*. 2014;17. doi:10.7448/IAS.17.1.18908

14. Virnig BA, McBean M. Administrative Data for Public Health Surveillance and Planning.

Annu Rev Public Health. 2001;22(1):213-230. doi:10.1146/annurev.publhealth.22.1.213

15. Lee CH, Yoon HJ. Medical big data: Promise and challenges. *Kidney Res Clin Pract*. 2017;36(1):3-11. doi:10.23876/j.krcp.2017.36.1.3

16. NHLS. National CD4 Testing Programme. https://www.nhls.ac.za/priority-programmes/cd4/. Accessed December 23, 2021.

17. Fellegi I, Sunter A. A Theory for Record Linkage. *J Am Stat Assoc*. 1969;40:1183–1210.

18. Jaro MA. Probabilistic linkage of large public health data files. *Stat Med*. 1995. doi:10.1002/sim.4780140510

19. Winkler WE. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. *Proc Sect Surv Res Am Stat Assoc*. 1990. doi:10.1007/978-1-4612-2856-1_101

20. Bor J, MacLeod W, Oleinik K, et al. Building a National HIV Cohort from Routine Laboratory Data: Probabilistic Record-Linkage with Graphs. *bioRxiv*. 2018:450304. doi:10.1101/450304

21. UNAIDS. *Country Factsheets South Africa | 2018*. Geneva; 2018.

22. South African National Department of Health. 2019 ART Clinical Guidelines for the Management of HIV in Adults, Pregnancy, Adolescents, Children, Infants and Neonates. 2019;(March):26.

23. South African National Department of Health. *National Antiretroviral Treatment*

Guidelines. Vol 328.; 2004. doi:10.1136/bmj.328.7434.280

24. South African National Department of Health. *The South African Antiretroviral Treatment Guidelines 2010*.; 2010.

25. South African National Department of Health. *The South African Antiretroviral Treatment Guidelines 2013*.; 2013.

26. South African National Department of Health. National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults. 2015;(January).

27. South African National Department of Health. *Circular on the Implementation of the Universal Test and Treat Strategy for HIV Positive Patients and Differentiated Care for Stable Patients*. Pretoria, South Africa; 2016.

28. Fox MP, Bor J, Brennan AT, et al. Estimating retention in HIV care accounting for patient transfers: A national laboratory cohort study in South Africa. *PLoS Med*. 2018;15(6):e1002589. doi:10.1371/journal.pmed.1002589

29. MacLeod W, Bor J, Crawford K, Carmona S. *Analysis of Big Data for Better Targeting of ART Adherence Strategies: Spatial Clustering Analysis of Viral Load Suppression by South African Province, District, Sub-District and Facility (April 2014-March 2015)*. Pretoria, South Africa; 2015.

30. Bor J, Gage A, Onoya D, Maskew M, Tripodis Y, Fox MP, Puren A, Carmona S, Mlisana K MW. Quality of Care in South African Facilities Providing ART Services. *PLoS Med*.

2021;18(3).

31. Carmona S, Bor J, Nattey C, et al. Persistent High Burden of Advanced HIV Disease Among Patients Seeking Care in South Africa's National HIV Program: Data From a Nationwide Laboratory Cohort. *Clin Infect Dis*. 2018;66(suppl_2):S111-S117. doi:10.1093/cid/ciy045

32. J B, MP F, C N, et al. Late presentation persists under UTT in South Africa: a national cohort study. In: *Conference on Retroviruses and Opportunistic Infections*. Boston, Massachusetts; 2020.

33. Maskew M, Bor J, MacLeod W, Carmona S, Sherman GG, Fox MP. Adolescent HIV treatment in South Africa's national HIV programme: a retrospective cohort study. *lancet HIV*. 2019;6(11):e760-e768. doi:10.1016/S2352-3018(19)30234-6

34. Harlow AF, Bor J, Brennan AT, et al. Impact of Viral Load Monitoring on Retention and Viral Suppression: A Regression Discontinuity Analysis of South Africa's National Laboratory Cohort. *Am J Epidemiol*. 2020;27708(February):1-19. doi:10.1093/aje/kwaa140

35. Fox MP, Brennan AT, Nattey C, et al. Delays in repeat HIV viral load testing for those with elevated viral loads: a national perspective from South Africa. *J Int AIDS Soc*. 2020;23(7):e25542. doi:10.1002/jia2.25542

36. Fox MP, Nattey C, Maskew M, MacLeod W, Brennan A BJ. *Impact of South Africa's National HIV Testing Campaign on Numbers Enrolling in Care: HE2RO Policy Brief Number 35*. Health Economics and Epidemiology Research Office, Johannesburg, South Africa;

2020.

37. Hallett TB, Eaton JW. A side door into care cascade for HIV-infected patients? *J Acquir Immune Defic Syndr*. 2013;63 Suppl 2:S228-32. doi:10.1097/QAI.0b013e318298721b

38. Bor J, Fox MP, Rosen S, et al. Treatment eligibility and retention in clinical HIV care: A regression discontinuity study in South Africa. *PLoS Med*. 2017;14(11):1-20. doi:10.1371/journal.pmed.1002463

39. Maskew M, Bor J, Hendrickson C, et al. Imputing HIV treatment start dates from routine laboratory data in South Africa: a validation study. *BMC Health Serv Res*. 2017;17(1):41. doi:10.1186/s12913-016-1940-2

40. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Dordrecht, Heidelberg, London, New York: Springer; 2009.

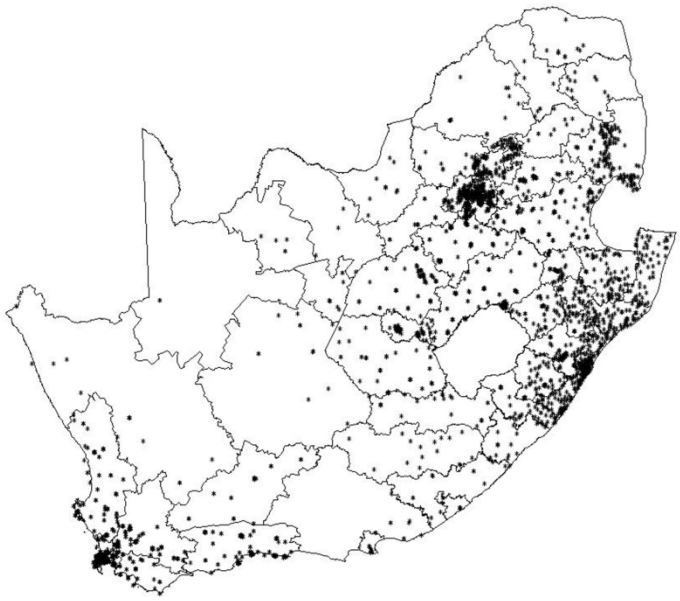

Figure 1. Location of health facilities providing HIV VL tests from May 1, 2017 to April 30, 2018. A total of 4,839 facilities in the NHLS database that provided HIV VL tests during the time period May 1, 2017 to April 30, 2018. 4,751 had valid longitude and latitude values and are represented in this map. Each asterisk represents one facility.

226x151mm (300 x 300 DPI)

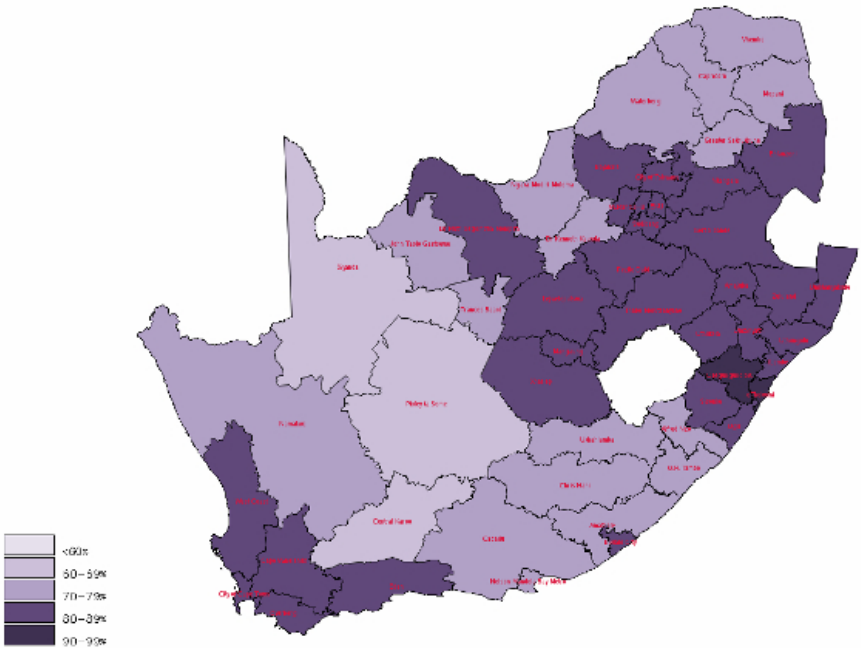

Figure 2. Viral load suppression by district South Africa, May 1, 2017 to April 30, 2018.

228x152mm (72 x 72 DPI)

BMJ Open

Cohort Profile: The South African National Health Laboratory Service (NHLS) National HIV Cohort

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2022-066671.R1
Article Type:	Cohort profile
Date Submitted by the Author:	18-Aug-2022
Complete List of Authors:	MacLeod, William; Boston University School of Public Health, Department of Global Health; University of the Witwatersrand, Health Economics and Epidemiology Research Office Bor, Jacob; Boston University School of Public Health, Departments of Global Health and Epidemiology; Boston University School of Public Health, Department of Global Health Candy, Sue; National Institute for Communicable Diseases Maskew, Mhairi; University of the Witwatersrand, Health Economics and Epidemiology Research Office Fox, Matthew; Boston University School of Public Health, Departments of Global Health and Epidemiology; University of the Witwatersrand, Health Economics and Epidemiology Research Office Bulekova, Katia; Boston University Brennan, Alana; Boston University School of Public Health, Departments of Global Health and Epidemiology; University of the Witwatersrand, Health Economics and Epidemiology Research Office Potter, James; Boston University School of Public Health, Department of Global Health Nattey, Cornelius; University of the Witwatersrand, Health Economics and Epidemiology Research Office Onoya, Dorina; University of the Witwatersrand, Health Economics and Epidemiology Research Office Mlisana, Koleka; National Health Laboratory Service Stevens, Wendy; University of the Witwatersrand, Department of Molecular Medicine and Haematology; National Health Laboratory Service Carmona, Sergio; University of the Witwatersrand, Department of Molecular Medicine and Haematology
Primary Subject Heading:	HIV/AIDS
Secondary Subject Heading:	Epidemiology
Keywords:	HIV & AIDS < INFECTIOUS DISEASES, Public health < INFECTIOUS DISEASES, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Cohort Profile: The South African National Health Laboratory Service (NHLS) National HIV Cohort

William B. MacLeod[1,2,†,*], Jacob Bor[1,2,3,†], Sue Candy[4], Mhairi Maskew[2], Matthew P. Fox[1,2,3], Katia Bulekova[5], Alana T. Brennan[1,2,3], James Potter,[1] Cornelius Nattey[2], Dorina Onoya[2], Koleka Mlisana[6], Wendy Stevens[6,7], Sergio Carmona[7]

Affiliations:

[1]Department of Global Health, Boston University School of Public Health, Boston, MA, USA

[2]Health Economics and Epidemiology Research Office, University of the Witwatersrand, Johannesburg, South Africa

[3]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

[4]National Institute of Communicable Diseases, Sandringham, South Africa

[5]Research Computing Services, Boston University, USA

[6]National Health Laboratory Service, Sandringham, South Africa

[7]Department of Molecular Medicine and Haematology, University of the Witwatersrand School of Pathology, Johannesburg, South Africa

*correspondence to: wmacleod@bu.edu, 801 Massachusetts Avenue, 3rd Floor, Boston, MA 02118

†equal contributions

Short title: Cohort Profile: South Africa's NHLS National HIV Cohort

Abstract

- **Purpose**: South Africa's National Health Laboratory Service (NHLS) National HIV Cohort was established in 2015 to facilitate monitoring, evaluation, and research on South Africa's National HIV Treatment Program. In South Africa, 84.8% of PLHIV know their HIV status, 70.7% are on ART, and 87.4% are virologically suppressed.
- **Participants**: The NHLS National HIV Cohort includes the laboratory data of nearly all patients receiving HIV care in the public sector since April 2004. Patients are included in the cohort if they have received a CD4 count or HIV RNA viral load test. Using an anonymized unique patient identifier that we developed and validated to linked test results, patients are observed prospectively through their laboratory results as they receive HIV care and treatment. Patients in HIV care are seen for laboratory monitoring every 6 to 12 months. Data collected include age, sex, facility location, and test results for CD4 counts, viral loads, and laboratory tests used to screen for potential treatment complications.
- **Findings to date**: From April 2004 through April 2018, 63 million CD4 count and viral load tests were conducted at 5,483 facilities. 12.6 million unique patients had at least one CD4 count or viral load, indicating they had accessed HIV care, and 7.1 million patients received a viral load test indicating they had started ART. The creation of NHLS National HIV Cohort has enabled longitudinal research on all lab-monitored patients in South Africa's national HIV program, including analyses of: 1) patient health at presentation; 2) care outcomes such as "CD4 recovery", "retention in care", and "viral re-suppression"; 3) patterns of transfer and re-entry into care; 4) facility-level variation in care outcomes; and 5) impacts of policies and guideline changes.
- **Future plans**: Continuous updating of the cohort, integration with available clinical data, and expansion to include tuberculosis and other lab-monitored comorbidities.

STRENGTHS AND LIMITATIONS

Strengths

- Large size and scope—all public sector patients in South Africa.
- We are able to explore outcomes without the problem of silent transfers.
- The cohort contains patient data prior to the initiation of ART: essential for assessing the impact of policy changes on outcomes both before and after treatment.

Limitations

- It is limited to laboratory results with limited demographic information and no clinical visit or pharmacy information.
- The laboratory data doesn't have a unique patient identifier and the matching techniques we use are not perfect and lead to both over- and under-matching of patient records, which potentially could lead to biased findings.

INTRODUCTION

South Africa has the largest HIV treatment program in the world, with over 4.7 million people currently on antiretroviral therapy (ART).[1] It also has the largest number of people living with HIV but not yet on ART, an estimated 3 million people.[1] Based on the fifth South African national HIV sero-behavioural survey conducted in 2017, 84.8% of people living with HIV were aware of their status, 70.7% of those who know their HIV status were on ART, and 87.4% of those on ART were virologically suppressed.[2] With such high stakes, it is critical to evaluate the impact of the ART program and improve outcomes nationally.

South Africa's National Health Laboratory Service (NHLS) National HIV Cohort was established to facilitate monitoring, evaluation, and research on South Africa's National HIV Care and Treatment Program. The NHLS National HIV Cohort includes the laboratory data of nearly all patients receiving HIV care in the public sector since 2004. Using an anonymized unique patient identifier that we developed and validated, patients can be followed longitudinally through their laboratory results as they progress through the HIV care and treatment cascade. This open, prospective cohort can be used to evaluate changes in HIV treatment guidelines, to assess trends in patient outcomes across space and time, to determine patterns of patient transfer, and to identify areas where outcomes lag behind.

South Africa has conducted a series of five national population-based HIV biomarker surveys since 2002 and a large number of HIV treatment cohorts have been established in South Africa. Analyses of these surveys and cohorts have contributed to our knowledge of HIV treatment outcomes.[2–12] However, the national surveys are only conducted every 4 or 5 years, and the existing cohorts while continuous, also have some drawbacks: they are not nationally

representative, most are in urban areas, and reflect long-standing research collaborations that have resulted in better-than-average patient care. Further, existing cohorts do not systematically track patients lost to follow-up in order to assess re-engagement in care. TIER.net, the country's HIV patient monitoring platform, is national in scope but has limited ability to track patients across sites[13]. While collecting data on all HIV patients nationally would be prohibitively expensive as a research endeavor, routine laboratory data collected in South Africa's HIV program can be leveraged to construct a cohort for monitoring, evaluation, and research.[14,15]

Since the beginning of the national HIV care and treatment program, NHLS provided nearly all laboratory testing for the public-sector program (with the exception of KwaZulu Natal Province which was not fully integrated into NHLS until April 2010). Laboratory testing has been used for CD4 and VL monitoring, for confirmatory HIV diagnostic testing, and to identify potential ART complications and contraindications. Specimens for testing are collected from patients at their care facility and sent to one of 16 laboratories for processing and testing. Results are generated by testing instruments in real time and sent to the NHLS Corporate Data Warehouse (CDW) which manages the distribution of results in real-time. All laboratory test results are maintained within the CDW. All NHLS HIV labs are accredited by the South African National Accreditation System (SANAS). The NHLS supports the labs with training, site visits, and on-site audits. Standard operating procedures have been developed and distributed to ensure standardisation across testing facilities. Laboratory performance on test volumes and turn-around-times are routinely monitored[16].

The NHLS National HIV Cohort has advantages over clinical cohorts. First, the data come directly from the source testing platforms and are less vulnerable to data entry errors that occur when extracting patient charts into databases. Second, the data are lower cost as they are collected for routine patient care. Third, the cohort offers a system-wide perspective on the national program in which transfers, drop-out, and re-entry into care can be observed, enabling evidence generation around national policy decisions. Fourth, the dataset size means that evaluations are robust and can be used to compare facilities, geographic areas, and demographic sub-groups. Fifth, because the data reflect all lab-monitored HIV patients receiving care in the public sector, the cohort reflects public-sector care-seeking patterns.

COHORT DESCRIPTION

Participants in the cohort

The NHLS National HIV Cohort includes all patients presenting for HIV care or treatment and receiving CD4 and/or HIV viral load (VL) monitoring at nearly all government health facilities in South Africa, from April 2004 through April 2018. During this period, 63 million CD4 and VL tests were conducted at 5,483 facilities.

A longitudinal cohort requires a patient identifier in order to follow patients over time. National ID numbers were only available for 2% of the specimens. However, the database contained sufficient information to link laboratory records to individual patients using probabilistic matching techniques. We developed a record linkage algorithm to identify unique patients in

the NHLS database combining elements of probabilistic linkage approaches with concepts from network analysis.[17–19] The methodology is reported in detail elsewhere.[20] In the absence of a true gold standard, we constructed manually-coded datasets for training and validation. We manually reviewed 58,905 candidate matches of 1,000 randomly sampled laboratory records. Relative to these manually-coded data, the algorithm had 93.7% sensitivity, indicating that the algorithm identified all but 6.3% of true matches. The algorithm had 98.6% positive predictive value, indicating that just 1.4% of algorithm-assigned matches were incorrect. Additionally, for those specimens linked to a National ID number, the algorithm correctly identified pairs of records with the same National ID number with 98.5% sensitivity. No linkage is perfect. For example, patients may present to clinics with different names and dates of birth, in which case linkage would be impossible. However, we developed a high-performing linkage approach using the available data in the NHLS database. By creating a validated patient identifier, the linkage enabled analysis of the NHLS database as an HIV Cohort. The cohort was deidentified after linkage, and all analyses are conducted using anonymized data.

Our record linkage of the CDW data was conducted on 152.5 million total laboratory results including 63 million CD4 counts and VL from April 2004 through April 2018. The algorithm assigned these laboratory results to 12,603,979 unique patients with at least one CD4 count or VL, indicating that these individuals had accessed HIV care. **Table 1** shows the demographics of the cohort. The cohort was 64.3% female, 33.7% male, and 2.0% unknown gender. The majority of individuals were 20-49 years old (8.4% had an unknown date of birth), with patients entering the cohort at a median age of 32 years (IQR 25-41 years). In the most recent 12 months of the cohort, HIV patients had laboratory tests at 4,751 facilities within South Africa (**Figure 1)**.

Table 1. Sex and age of patients entering into HIV Care from May 2004 to April 2018.

Parameter (% (n))	May 2004 to April 2008 (n=2,042,315)	May 2008 to April 2013 (n=5,752,084)	May 2013 to April 2018 (n=4,809,580)	May 2004 to April 2018 (n=12,603,979)
Male	31.4% (641,534)	32.4% (1,864,287)	36.2% (1,741,698)	33.7% (4,247,519)
Female	66.3% (1,354,605)	65.9% (3,792,924)	61.6% (2,962,984)	64.3% (8,110,513)
Unknown	2.3% (46,176)	1.6% (94,873)	2.2% (104,898)	2.0% (245,947)
Age Group (years)				
0-4	3.5% (71,011)	2.8% (162,706)	2.3% (111,735)	2.7% (345,452)
5-9	1.4% (28,358)	1.4% (78,910)	1.0% (49,771)	1.2% (157,039)
10-14	0.6% (11,889)	1.0% (60,280)	1.1% (54,329)	1.0% (126,498)
15-19	2.8% (57,167)	3.8% (216,482)	4.4% (211,731)	3.9% (485,380)
20-24	11.2% (229,680)	11.5% (662,334)	11.5% (551,815)	11.5% (1,443,829)
25-29	17.7% (360,484)	16.8% (965,872)	15.6% (751,784)	16.5% (2,078,140)
30-34	18.3% (374,633)	15.9% (917,420)	15.7% (755,664)	16.2% (2,047,717)
35-39	14.1% (287,302)	13.3% (763,369)	12.4% (597,389)	13.1% (1,648,060)
40-44	9.9% (202,827)	9.2% (530,075)	9.3% (448,095)	9.4% (1,180,997)
45-49	6.4% (131,450)	6.5% (375,617)	6.4% (310,034)	6.5% (817,101)
50-54	3.9% (79,764)	4.2% (242,140)	4.6% (221,419)	4.3% (543,323)
55-59	2.0% (41,344)	2.5% (144,966)	3.0% (144,236)	2.6% (330,546)
60-64	0.9% (17,757)	1.3% (73,998)	1.8% (85,267)	1.4% (177,022)
65+	0.4% (8,772)	0.6% (33,536)	0.8% (40,675)	0.7% (82,983)

Age is age at first entry into HIV Care--at first CD4 or first VL test.

Of the 12.6 million patients in the cohort, 7.1 million ever initiated ART. To estimate the number of patients currently receiving ART, we limited the cohort to the last 18 months from

November 2016 through April 2018. We estimate there were 4.4 million patients receiving ART as evidenced by VL monitoring during that time period, similar to the 4.7 million on ART UNAIDS estimated for 2018.[21] The number of patients entering HIV care, initiating HIV treatment, and actively on HIV treatment by year is shown in **Table 2.** Since 2004-05 the number with a first VL (a proxy for initiating HIV treatment, in South Africa, only patients on ART are monitored virologically) has grown from 27,937 to 913,604 per year.

We found 4.2 million CD4 and 592,261 VL tests that were unlinked to other tests. Patients in the cohort that had just a single CD4 and no VL include the large number of patients who present clinically with HIV, have blood drawn for a CD4, but do not seek follow-up care. Still, some of these singleton lab results may be "stray" lab results that should have been linked to a patient. Singleton VL represent 8.3% of all VL tests.

Participant Follow-Up

In South Africa, as in many countries, HIV care and treatment decisions are informed by routine laboratory monitoring, as specified in standardized national guidelines.[22] While the cohort lacks information on clinical visits, pharmacy data, or ART regimen, it is nevertheless possible to follow important care and treatment decisions through the labs.

Within the cohort, follow-up is determined by the dates of CD4 and VL monitoring, as well as screening for potential treatment complications, known as the ART work-up. Laboratory-based monitoring of CD4 and VL has been a standard component of national HIV care and treatment

Table 2. Number of patients entering into HIV Care, receiving first HIV VL test, on ART, and HIV VL suppressed by 12-month period from May 2004 to April 2018.

Year (12-month period May 1 to April 30)	Number of patients entering into HIV Care[a]	Number of patients receiving first HIV VL test[b]	Number of patients on ART[c]	Percent HIV VL suppressed out of all VL tests[d]
2004-05	118,829	27,937	24,208	31.1
2005-06	219,942	111,865	120,796	39.4
2006-07	302,510	176,979	272,078	43.4
2007-08	344,935	223,188	420,542	50.3
2008-09	410,563	283,929	609,729	55.4
2009-10	479,406	249,971	716,555	65.7
2010-11	915,898	525,457	1,066,016	72.3
2011-12	713,101	582,175	1,465,428	72.2
2012-13	626,296	672,543	1,875,968	74.0
2013-14	619,396	715,786	2,316,959	74.6
2014-15	629,246	785,503	2,801,889	77.1
2015-16	666,492	952,256	3,429,148	79.6
2016-17	629,754	919,229	3,951,406	79.4
2017-18	461,076	913,604	4,419,017	83.2

[a]Numbers of patients into HIV Care is the number of patients with a first CD4 or first VL.

[b]Number of patients receiving first HIV VL test is the number of first of VL tests

[c]Number of patients on ART is the number of patients who had a VL test in the 18-month period from November 1, (year – 1) to April 30, (year + 1).

[d]Percent of HIV VL suppressed are 100*(the number of VL < 400 cp/ml/Total number of VL tests) during the same 18-month period. Patients who did not receive a viral load test are not included.

guidelines during the period of study and a routine part of clinical care. During the period of observation, patients testing positive for HIV at a health facility have blood drawn for a CD4 on the same day or shortly thereafter. Thus, a patient's first CD4 test is a reasonable proxy for the date when a patient was diagnosed as HIV-infected within the health system and presented to a health facility. At the start of the HIV care program, VL tests were conducted at treatment initiation, and then every six-months. Since 2010, VL monitoring has been limited to 6- and 12-months after treatment initiation and annually thereafter. **Table 3** lists the evolving criteria for ART initiation and ART laboratory monitoring in South Africa from 2004 to 2016.

Within the cohort, for patients on treatment, a total of 26.7 million follow-up VL have been conducted which corresponds to a median (IQR) of 2 (1-5) VL per person. This translates to 23,381,315 total person years of follow-up on HIV treatment for a median (IQR) of 2.1 (1.0-5.0) years per person.

Variables measured

The variables in the NHLS National HIV Cohort are described in **Table 4** along with percent completeness. The cohort includes the type of test, tests results, test date, and geographic information along with patient's date of birth and sex. **Table 5** is a frequency listing of test type by three time periods. The tests included in the cohort includes all CD4, VL, HIV confirmatory tests (PCR/Elisa), and ART-workup labs for patients receiving HIV care.

Despite the lack of clinical and pharmacy data, the cohort can be used to generate a wealth of information about the national HIV program. Information on facility can be linked to facility geocodes, which can be mapped (**Figure 1**), aggregated to the local municipality, district, and

Table 3. Evolving criteria for ART initiation and ART monitoring in South Africa from 2004 to 2016.

Year	Guidelines	ART Initiation Criteria/CD4 and VL monitoring
2004	National ARV Treatment Guidelines (2004)[23]	*Criteria for ART Initiation* CD4 <200 cells/mm3 irrespective of stage or WHO Stage IV AIDS-defining illness, irrespective of CD4 count and Patient expresses willingness and readiness to take ART adherently *ART Monitoring* CD4: Staging and 6 monthly VL: Staging and 6 monthly
2010	The South African Antiretroviral Treatment Guidelines 2010[24]	*Criteria for ART Initiation* CD4 count <200cells/mm3 irrespective of clinical stage or CD4 count <350cells/mm3 in patients with TB/HIV or Pregnant women or WHO stage IV irrespective of CD4 count or MDR/XDR irrespective of CD4 *Require fast track (i.e. ART initiation within 2 weeks of being eligible)* Pregnant women eligible for lifelong ART; Patients with very low CD4 (<100); Stage 4, CD4 count not yet available, or; MDR/XDR TB *ART Monitoring* CD4: Staging, 6 and 12 months and annually thereafter VL: 6 and 12 months and annually thereafter
2011	Circular on New Criteria for Initiation of Adults on ART at CD4 Count of 350 cells/mm3 and below. August 26, 2011	Baseline CD4 for initiation at CD4 < 350 cells/mm3
2012	Accelerating Access to ART Services and Uptake (circular) April 17, 2012	*Criteria for ART Initiation* CD4 count <350 cells/mm3 irrespective of clinical stage or patients with TB/HIV irrespective of CD4 count or Pregnant women or WHO stage IV irrespective of CD4 count or MDR/XDR irrespective of CD4 *Require fast track (i.e. Same day ART initiation)* Pregnant women eligible for lifelong ART; Patients with low CD4 (<200); Stage 4, CD4 count not yet available, or; MDR/XDR TB Screening and treatment of patients with very low CD4 Counts (<100) for cryptococcal infection *ART Monitoring* CD4: Staging, 12 months and annually thereafter VL: 6 and 12 months and annually thereafter
2013	The South African Antiretroviral Treatment Guidelines 2013[25]	*Criteria for ART Initiation* CD4 count <350 cells/mm3 irrespective of WHO clinical stage or Irrespective of CD4 count and All types of TB (In patients with TB drug resistant or sensitive, including extra pulmonary TB) and WHO stage 3 or 4 irrespective of CD4 count *Require fast track (i.e. ART initiation within 7 days of being eligible)* HIV positive women who are pregnant or breast feeding or Patients with low CD4 <200 or Patients with Stage 4, irrespective of CD4 count or Patients with TB/HIV co morbidity with CD4 count < 50 (Patients

		with Cryptococcus meningitis or TB meningitis (defer ART for 4-6 weeks) *ART Monitoring* CD4: Staging and 1 year VL: 6 and 12 months and annually thereafter
2015	National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults 2015[26]	*Criteria for ART Initiation* CD4 count <500 cells/mm3 irrespective of WHO clinical stage or All types of TB or WHO stage 3 or 4 or HBV coinfection irrespective of CD4 count *Immediate Initiation* All HIV-positive pregnant or breastfeeding women, as long as no active TB *Require fast track (i.e. ART initiation within 7 days of being eligible)* Patients with low CD4 <200 or Patients with Stage 4, irrespective of CD4 count *ART Monitoring* CD4: Staging and 1 year VL: 6 and 12 months and annually thereafter
2016	Circular on Implementation of the Universal Test and Treat Strategy for HIV Positive Patients and Differentiated Care for Stable Patients August 22, 2016[27]	*Criteria for ART Initiation* All HIV positive adolescents and adults regardless of CD4 count *ART Monitoring* CD4 Staging and 1 year. VL monitoring at 6 and 12 months and annually thereafter.

Table 4. Variables in the NHLS National HIV Cohort

Variable	Description	Percent completeness of variable
BU_uniq_ID	Unique patient identifier	100%
Episode_no	NHLS Episode identifier	100%
Sex	Sex	98.0%
Age	Age in years at testing date	91.6%
Province	Province of health facility where specimen was taken	99.8%
District	District of health facility	99.8%
Local_Municipality	Local municipality of health facility	99.8%
Facility	NHLS Facility Code of health facility	99.8%
Ward	NHLS Ward Code of ward at health facility	99.8%
Test_name	Type of test (e.g. CD4, VL)	100%
Test_date	Date of test	100%
Test_result	Test result	100%

Table 5. Frequency of all tests included in the NHLS National HIV Cohort in three time periods, May 2004-April 2018

Name of Test	Use in HIV care	2004-08	2008-13	2013-18	All time periods
Alanine Aminotransferase	Measure of liver injury and determining choice of ART	1,970,138	9,301,723	9,176,996	20,448,857
CD4 Count	ART eligibility and disease progression	3,599,594	15,483,376	17,578,415	36,661,385
Serum Cryptococcal Antigen	To detect and prevent cryptococcal meningitis	17	71,374	1,017,903	1,089,294
Creatinine Clearance	Measure of kidney function and determining choice of ART	2,880	6,113,182	25,422,404	31,538,466
HIV Elisa Confirmatory	HIV diagnostic test primarily in infants and young children	663,518	820,322	370,465	1,854,305
HIV Elisa Screening	HIV diagnostic test primarily in	1,129,249	1,448,212	639,528	3,216,989

	infants and young children				
Hemoglobin	Measure of overall health	3,841,090	12,318,399	14,260,748	30,420,237
HIV PCR	HIV diagnostic test primarily used for HIV diagnosis in infants	39,706	109,079	108,212	256,997
HIV RNA Viral Load	Monitoring efficacy of ART therapy	1,120,821	6,716,430	18,922,454	26,759,705
Total		**12,367,013**	**52,382,097**	**87,497,125**	**152,246,235**

provincial levels, or linked to external population-level data such as HIV prevalence, poverty levels, or the population age distribution. Information on test dates can be used to assess longitudinal outcomes such as retention in care (18 months without a monitoring lab) at the patient level, and to assess trends in patient outcomes over time.

Participant attrition

As in other passive surveillance cohorts[5], attrition from care (and its inverse, "retention in care") is a primary outcome of interest. Data are generated through lab monitoring as part of routine clinical care and no data are collected beyond what is clinically indicated. No efforts are made by the research team to retain patients in care nor to actively follow-up patients who have left care. However, accuracy of retention estimates are enhanced by the national perspective of the cohort which is robust to silent transfers. As the study population includes all people who have sought public sector care for HIV in South Africa, attrition occurs only if a person emigrates from South Africa and can no longer seek HIV care.

Patient and Public Involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

FINDINGS TO DATE

The ability to link laboratory results to create records of individual patients has enabled 1) longitudinal patient-level epidemiologic research for the complete national HIV care and treatment program including patients not yet on ART; 2) assessment of concepts such as "CD4

recovery", "retention in care", and "viral re-suppression" that require individual-level longitudinal data and monitoring of these concepts at all public-sector facilities nationally; 3) tracking of patients as they seek care at different clinics within the health system and assessing patterns of transfer; 4) evaluation of policies and guideline changes. Finally, linkage with facility geocodes has enabled integration of the cohort with publicly available data on outcomes and exposures at the facility or district level, including programmatic data (e.g. clinic staffing, size) and population-based data (e.g. population density, poverty, HIV prevalence, and mortality).

Key findings from the NHLS National HIV Cohort include the following:

- *Retention in HIV care is underestimated by not accounting for within system patient transfer* – Estimated retention in HIV care from both the initiating clinic and a national perspective. At the clinic level retention in care was 29.1% by 6 years. However, when accounting for transfers to other clinics, retention in care was 63.3% by 6 years.[28]

- *There is large spatial heterogeneity in the HIV care cascade* – Estimated rates of VL testing and suppression from April 2014 through March 2015 across public facilities. We identified wide spatial variation in VL suppression, ranging from 69 to 82% at the provincial level.[29] **Figure 2** shows a map of district level estimates of VL suppression. The cohort was also used to develop a summary measure of quality of care at the facility level based on patients' longitudinal outcomes in different facilities. Year-to-year, quality was found to be highly correlated within facilities, but varied widely across facilities.[30]

- *A high proportion of patients present with advanced disease* – Documented that a consistently large share (~33%) of patients entered into care with a CD4 <200 cells/uL from

2011 to 2016 despite increased CD4 treatment eligibility standards.[31] Late presentation persisted into the UTT era.[32]

- *A wave of adolescents will require HIV treatment* – Identified adolescents who entered care in childhood (likely perinatally-infected) and adolescents who entered care in their later teenage years (likely infected via sexual transmission). A 10- to 20-fold increase in the numbers of adolescents on ART from 2004-2007 to 2012-2014 was found and resulting in a "wave" of 15-19 year-olds who will require HIV treatment over the next decade due to both the aging of perinatally-infected children into adolescence and increased numbers of adolescent girls seeking HIV care for the first time.[33]
- *Viral monitoring for treatment failure* – VL monitoring is conducted to identify patients at risk for treatment failure, to target these patients with adherence counseling, and to switch them to second-line therapy if needed. Comparing outcomes among patients with VL results just above vs. just below the 1000 copies/mL[3] threshold, we found that while having a VL>1000 increases the probability of follow-up VL monitoring, most patients with elevated VL do not receive monitoring within recommended timelines.[34,35]

STRENGTHS AND LIMITATIONS

The main strength of the NHLS National HIV Cohort is its size and scope. With millions of patients over many years, the cohort can be used to conduct robust evaluations of policy change in South Africa's public-sector treatment programme. It can also be used to take advantage of variation in program implementation (e.g. increases in HIV testing in some areas

before others, or implementation of the National Adherence Strategy) to evaluate the impacts of these interventions[36].

Second, the cohort is unique in its ability to explore outcomes without the problem of silent transfers.[37] Silent transfers, patients who move from one facility to another without informing their sending clinic, leads to overestimates of attrition from HIV care and misclassification of outcomes in programmatic evaluations. Because the cohort contains information on the clinic where the lab investigation was conducted, we are able to identify movement between clinics and not misclassify these movements as lost to follow-up.

Third, the cohort contains patient data prior to the initiation of ART: essential for assessing the impact of policy changes on outcomes both before and after treatment—something few clinical cohorts in South Africa capture[38].

A weakness of the cohort is that it is limited to laboratory results with limited demographic information and no clinical visit or pharmacy information. While we have been able to overcome some of these limitations through imputation of ART start dates[39], we do not have data on medication or visit adherence, or clinical diagnoses that would be useful for describing a national program.

Second, the matching techniques we use are not perfect and lead to both over- and under-matching of patient records, which potentially could lead to biased findings. Because the dataset is so large, random error is typically approaching zero in any analysis. This makes systematic errors (like the over- and under-matching) the main source of error in studies using this cohort, a problem that can be explicitly modelled using quantitative bias analysis.[40] We

have also assessed the sensitivity of our results to matching parameters and found the results to be quite robust.[28]

Third, as with other clinical databases, data collection is part of routine clinical care. This means that if patients do not present for care, we cannot observe their CD4 or VL. Further, adherence to laboratory monitoring guidelines may vary across facilities, which may contribute to differences in outcomes across facilities.

Finally, the cohort does not have data on death and is unable to link to the National Vital Registration System to obtain mortality data due to the paucity of national ID numbers collected. This means that we are currently not able to describe the impact of interventions and policy changes on mortality. Instead we use other indicators of poor outcomes such as unsuppressed VL, failure to gain CD4 cell count, and attrition from care.

Future Plans

Our future plans for the cohort fall into three categories all with the aim of enhancing the research and clinical value of the cohort: 1) continuous updating of the cohort; 2) integrate the cohort with clinical databases, and 3) expand the scope of cohort to include tuberculosis and other lab-monitored comorbidities.

Collaboration: Access to primary data is subject to restrictions owing to privacy and ethics policies set by the South African Government. Requests for access to the data can be made to the National Health Laboratory Services directly (http://www.aarms.nhls.ac.za) and require a full protocol submission. Inquiries can be made via the Office of Academic Affairs and Research

at the National Health Laboratory Service. To find out more about the cohort, contact academic.research@nhls.ac.za.

Ethics clearance: We received ethics clearance from the Wits HREC (M150429) and Boston University IRB (H-31968, H-33442).

Acknowledgements: Jaco Grobler was instrumental in programming the first iteration of the NHLS unique patient ID which was the genesis for this work. The authors are also grateful for the advice and technical assistance of Naseem Cassim concerning details of the CDW architecture. The authors would like to acknowledge the contributions of the dedicated staff at the National Health Laboratory Service Corporate Data Warehouse for their support and assistance with data extracts and staff of the National Priority Programme for maintaining a service that ensures the generation of good quality assay results.

Funding Acknowledgments: This work was supported by grants R01AI115979, R01AI152149, K01MH105320, K01DK116929-01, and R01HD084233 from the National Institutes of Health. WBM and SC were supported by USAID through Cooperative Agreement AID- 674-A-12-0020. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Contributorship: William B. MacLeod, Jacob Bor, Sue Candy, Mhairi Maskew, Matthew P. Fox, Katia Bulekova, Alana T. Brennan, James Potter, Wendy Stevens and Sergio Carmona made substantial contributions to the conception or design of the work.

William B. MacLeod, Jacob Bor, Sue Candy, Katia Bulekova, James Potter, Cornelius Nattey, Dorina Onoya, and Koleka Mlisana made substantial contributions to the acquisition, analysis or interpretation of data for the work.

William B MacLeod, Jacob Bor, Mhari Maskew, Matthew P. Fox, and Cornelius Nattey drafted the work.

Sue Candy, Katia Bulekova, Alana T. Brennan, James Potter, Wendy Stevens, and Sergio Carmona critically revised the work for important intellectual content.

All authors gave final approval of the version to be published, and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Figure Legends

Figure 1. Location of health facilities providing HIV VL tests from May 1, 2017 to April 30, 2018. A total of 4,839 facilities in the NHLS database that provided HIV VL tests during the time period May 1, 2017 to April 30, 2018. 4,751 had valid longitude and latitude values and are represented in this map. Each asterisk represents one facility.

Figure 2. Viral load suppression by district South Africa, May 1, 2017 to April 30, 2018.

References

1. UNAIDS. *UNAIDS Data 2019 Annual Report*. Geneva; 2019.

2. Marinda E, Simbayi L, Zuma K, et al. Towards achieving the 90–90–90 HIV targets: Results from the south African 2017 national HIV survey. *BMC Public Health*. 2020;20(1):1-12. doi:10.1186/s12889-020-09457-z

3. Fox MP, Maskew M, Macphail a P, et al. Cohort Profile: The Themba Lethu Clinical Cohort, Johannesburg, South Africa. *Int J Epidemiol*. March 2012:6-8. doi:10.1093/ije/dys029

4. Haas AD, Zaniewski E, Anderegg N, et al. Retention and mortality on antiretroviral therapy in sub-Saharan Africa: Collaborative analyses of HIV treatment programmes: Collaborative. *J Int AIDS Soc*. 2018;21(2):1-7. doi:10.1002/jia2.25084

5. Egger M, Ekouevi DK, Williams C, et al. Cohort Profile: The international epidemiological databases to evaluate AIDS (IeDEA) in sub-Saharan Africa. *Int J Epidemiol*. May 2011. doi:10.1093/ije/dyr080

6. Houlihan CF, Bland RM, Mutevedzi PC, et al. Cohort profile: Hlabisa HIV treatment and care programme. *Int J Epidemiol*. 2011;40(2):318-326. doi:10.1093/ije/dyp402

7. Kluberg SA, Fox MP, LaValley M, Pillay D, Bärnighausen T, Bor J. Do HIV treatment eligibility expansions crowd out the sickest? Evidence from rural South Africa. *Trop Med Int Heal*. 2018;23(9):968-979. doi:10.1111/tmi.13122

8. Bor J, Chiu C, Ahmed S, et al. Failure to initiate HIV treatment in patients with high CD4 counts: evidence from demographic surveillance in rural South Africa. *Trop Med Int Heal*. 2018;23(2):206-220. doi:10.1111/tmi.13013

9. Macleod WB, Maskew M, Jaffray I, Macphail AP, Ive P, Fox MP. The Feasibility of Using Screening Criteria to Reduce Clinic Visits for Stable Patients on Antiretroviral Therapy in South Africa. *J Acquir Immune Defic Syndr*. October 2012. doi:10.1097/QAI.0b013e318278e976

10. Fox MP, Shearer K, Maskew M, et al. Treatment outcomes after 7 years of public-sector HIV treatment. *AIDS*. 2012;26(14):1823-1828. doi:10.1097/QAD.0b013e328357058a

11. Brennan AT, Maskew M, Ive P, et al. Increases in regimen durability associated with the introduction of tenofovir at a large public-sector clinic in Johannesburg, South Africa. *J Int AIDS Soc*. 2013;16:1-12. doi:10.7448/IAS.16.1.18794

12. Haas AD, Johnson LF, Grimsrud A, et al. Extending Visit Intervals for Clinically Stable Patients on Antiretroviral Therapy: Multicohort Analysis of HIV Programs in Southern Africa. *J Acquir Immune Defic Syndr*. 2019;81(4):439-447. doi:10.1097/QAI.0000000000002060

13. Osler M, Hilderbrand K, Hennessey C, et al. A three-tier framework for monitoring antiretroviral therapy in high HIV burden settings. *J Int AIDS Soc*. 2014;17. doi:10.7448/IAS.17.1.18908

14. Virnig BA, McBean M. Administrative Data for Public Health Surveillance and Planning.

Annu Rev Public Health. 2001;22(1):213-230. doi:10.1146/annurev.publhealth.22.1.213

15. Lee CH, Yoon HJ. Medical big data: Promise and challenges. *Kidney Res Clin Pract*. 2017;36(1):3-11. doi:10.23876/j.krcp.2017.36.1.3

16. NHLS. National CD4 Testing Programme. https://www.nhls.ac.za/priority-programmes/cd4/. Accessed December 23, 2021.

17. Fellegi I, Sunter A. A Theory for Record Linkage. *J Am Stat Assoc*. 1969;40:1183–1210.

18. Jaro MA. Probabilistic linkage of large public health data files. *Stat Med*. 1995. doi:10.1002/sim.4780140510

19. Winkler WE. String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. *Proc Sect Surv Res Am Stat Assoc*. 1990. doi:10.1007/978-1-4612-2856-1_101

20. Bor J, MacLeod W, Oleinik K, et al. Building a National HIV Cohort from Routine Laboratory Data: Probabilistic Record-Linkage with Graphs. *bioRxiv*. 2018:450304. doi:10.1101/450304

21. UNAIDS. *Country Factsheets South Africa | 2018*. Geneva; 2018.

22. South African National Department of Health. 2019 ART Clinical Guidelines for the Management of HIV in Adults, Pregnancy, Adolescents, Children, Infants and Neonates. 2019;(March):26.

23. South African National Department of Health. *National Antiretroviral Treatment*

Guidelines. Vol 328.; 2004. doi:10.1136/bmj.328.7434.280

24. South African National Department of Health. *The South African Antiretroviral Treatment Guidelines 2010*.; 2010.

25. South African National Department of Health. *The South African Antiretroviral Treatment Guidelines 2013*.; 2013.

26. South African National Department of Health. National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults. 2015;(January).

27. South African National Department of Health. *Circular on the Implementation of the Universal Test and Treat Strategy for HIV Positive Patients and Differentiated Care for Stable Patients*. Pretoria, South Africa; 2016.

28. Fox MP, Bor J, Brennan AT, et al. Estimating retention in HIV care accounting for patient transfers: A national laboratory cohort study in South Africa. *PLoS Med*. 2018;15(6):e1002589. doi:10.1371/journal.pmed.1002589

29. MacLeod W, Bor J, Crawford K, Carmona S. *Analysis of Big Data for Better Targeting of ART Adherence Strategies: Spatial Clustering Analysis of Viral Load Suppression by South African Province, District, Sub-District and Facility (April 2014-March 2015)*. Pretoria, South Africa; 2015.

30. Bor J, Gage A, Onoya D, Maskew M, Tripodis Y, Fox MP, Puren A, Carmona S, Mlisana K MW. Quality of Care in South African Facilities Providing ART Services. *PLoS Med*.

2021;18(3).

31. Carmona S, Bor J, Nattey C, et al. Persistent High Burden of Advanced HIV Disease Among Patients Seeking Care in South Africa's National HIV Program: Data From a Nationwide Laboratory Cohort. *Clin Infect Dis*. 2018;66(suppl_2):S111-S117. doi:10.1093/cid/ciy045

32. J B, MP F, C N, et al. Late presentation persists under UTT in South Africa: a national cohort study. In: *Conference on Retroviruses and Opportunistic Infections*. Boston, Massachusetts; 2020.

33. Maskew M, Bor J, MacLeod W, Carmona S, Sherman GG, Fox MP. Adolescent HIV treatment in South Africa's national HIV programme: a retrospective cohort study. *lancet HIV*. 2019;6(11):e760-e768. doi:10.1016/S2352-3018(19)30234-6

34. Harlow AF, Bor J, Brennan AT, et al. Impact of Viral Load Monitoring on Retention and Viral Suppression: A Regression Discontinuity Analysis of South Africa's National Laboratory Cohort. *Am J Epidemiol*. 2020;27708(February):1-19. doi:10.1093/aje/kwaa140

35. Fox MP, Brennan AT, Nattey C, et al. Delays in repeat HIV viral load testing for those with elevated viral loads: a national perspective from South Africa. *J Int AIDS Soc*. 2020;23(7):e25542. doi:10.1002/jia2.25542

36. Fox MP, Nattey C, Maskew M, MacLeod W, Brennan A BJ. *Impact of South Africa's National HIV Testing Campaign on Numbers Enrolling in Care: HE2RO Policy Brief Number 35*. Health Economics and Epidemiology Research Office, Johannesburg, South Africa;

2020.

37. Hallett TB, Eaton JW. A side door into care cascade for HIV-infected patients? *J Acquir Immune Defic Syndr*. 2013;63 Suppl 2:S228-32. doi:10.1097/QAI.0b013e318298721b

38. Bor J, Fox MP, Rosen S, et al. Treatment eligibility and retention in clinical HIV care: A regression discontinuity study in South Africa. *PLoS Med*. 2017;14(11):1-20. doi:10.1371/journal.pmed.1002463

39. Maskew M, Bor J, Hendrickson C, et al. Imputing HIV treatment start dates from routine laboratory data in South Africa: a validation study. *BMC Health Serv Res*. 2017;17(1):41. doi:10.1186/s12913-016-1940-2

40. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. Dordrecht, Heidelberg, London, New York: Springer; 2009.

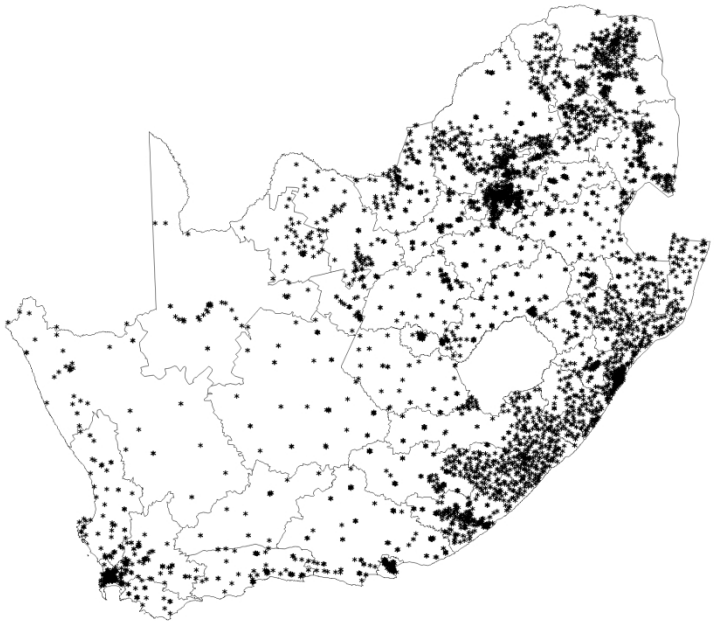

Figure 1. Location of health facilities providing HIV VL tests from May 1, 2017 to April 30, 2018. A total of 4,839 facilities in the NHLS database that provided HIV VL tests during the time period May 1, 2017 to April 30, 2018. 4,751 had valid longitude and latitude values and are represented in this map. Each asterisk represents one facility.

203x135mm (300 x 300 DPI)

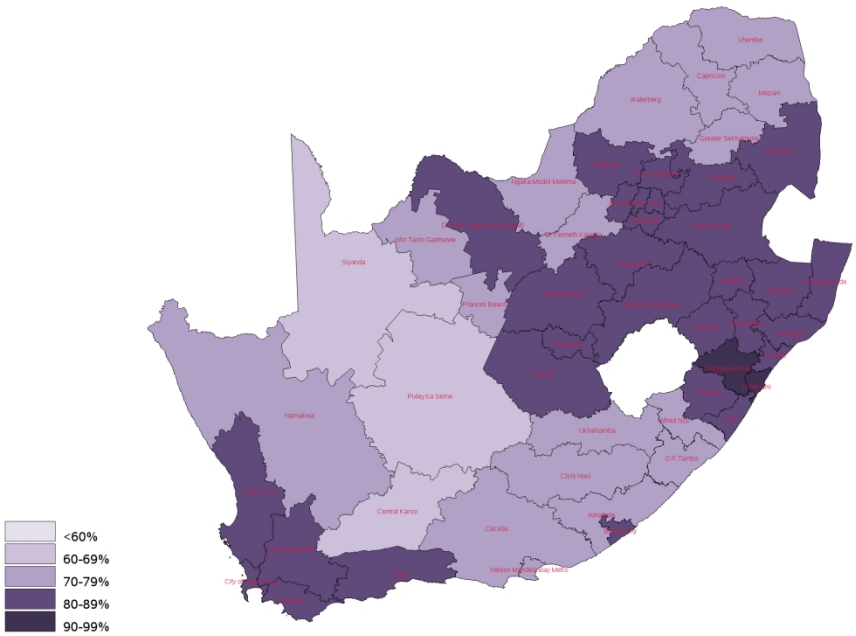

Figure 2. Viral load suppression by district South Africa, May 1, 2017 to April 30, 2018.

516x344mm (118 x 118 DPI)