





BMJ Open Cohort profile: the South African National Health Laboratory Service (NHLS) National HIV Cohort

William B MacLeod ^{1,2} Jacob Bor,^{2,3} Sue Candy,⁴ Mhairi Maskew,² Matthew P Fox ^{2,3} Katia Bulekova,⁵ Alana T Brennan ^{2,3} James Potter,¹ Cornelius Nattey,² Dorina Onoya ² Koleka Mlisana,⁶ Wendy Stevens,^{7,8} Sergio Carmona⁷

To cite: MacLeod WB, Bor J, Candy S, *et al.* Cohort profile: the South African National Health Laboratory Service (NHLS) National HIV Cohort. *BMJ Open* 2022;**12**:e066671. doi:10.1136/bmjopen-2022-066671

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-066671>).

WBM and JB contributed equally.

Received 14 July 2022
Accepted 07 September 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr William B MacLeod;
wmacleod@bu.edu

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 Programme. In South Africa, 84.8% of people living with HIV know their HIV status; 70.7% who know their status are on ART; and 87.4% on ART 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 (VL) test. Using an anonymised unique patient identifier that we have developed and validated to linked test results, we observe patients prospectively through their laboratory results as they receive HIV care and treatment. Patients in HIV care are seen for laboratory monitoring every 6–12 months. Data collected include age, sex, facility location and test results for CD4 counts, VLs and laboratory tests used to screen for potential treatment complications.

Findings to date From April 2004 to April 2018, 63 million CD4 count and VL tests were conducted at 5483 facilities. 12.6 million unique patients had at least one CD4 count or VL, indicating they had accessed HIV care, and 7.1 million patients had a VL test indicating they had started antiretroviral therapy. The creation of NHLS National HIV Cohort has enabled longitudinal research on all lab-monitored patients in South Africa's national HIV programme, including analyses of (1) patient health at presentation; (2) care outcomes such as 'CD4 recovery', 'retention in care' and 'viral resuppression'; (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.

INTRODUCTION

South Africa has the largest HIV treatment programme in the world, with over 4.7 million people currently on antiretroviral therapy (ART).¹ It also has the largest number of

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ 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 antiretroviral therapy: essential for assessing the impact of policy changes on outcomes both before and after treatment.
- ⇒ It is limited to laboratory results with limited demographic information and no clinical visit or pharmacy information.
- ⇒ The laboratory data does not have a unique patient identifier, and the matching techniques we use are not perfect and lead to both overmatching and undermatching of patient records, which potentially could lead to biased findings.

people living with HIV (PLHIV) but not yet on ART, an estimated 3 million people.¹ Based on the fifth South African national HIV serobehavioural survey conducted in 2017, 84.8% of PLHIV 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 programme 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 programme. The NHLS National HIV Cohort includes the laboratory data of nearly all patients receiving HIV care in the public sector since 2004. Using an anonymised unique patient identifier that we have developed and validated, we can follow up patients 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, and 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 patients with HIV nationally would be prohibitively expensive as a research endeavour, routine laboratory data collected in South Africa's HIV programme can be leveraged to construct a cohort for monitoring, evaluation and research.^{14 15}

Since the beginning of the national HIV care and treatment programme, NHLS provided nearly all laboratory testing for the public-sector programme (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 viral load (VL) monitoring, for confirmatory HIV diagnostic testing and for identifying potential ART complications and contraindications. Specimens for testing are collected from patients at their care facility and sent to 1 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. 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 programme 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

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

Parameter (% (n))	May 2004–April 2008 (n=2 042 315)	May 2008–April 2013 (n=5 752 084)	May 2013–April 2018 (n=4 809 580)	May 2004–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 (8772)	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 viral load test.

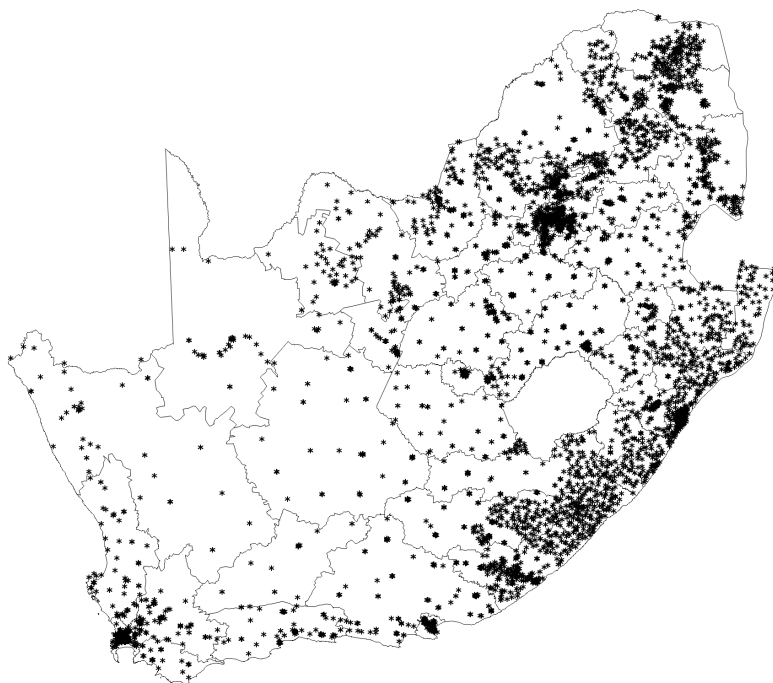


Figure 1 Location of health facilities providing HIV VL tests from 1 May 2017 to 30 April 2018. A total of 4839 facilities in the National Health Laboratory Service database provided HIV VL tests during the time period of 1 May 2017–30 April 2018. A total of 4751 had valid longitude and latitude values and are represented in this map. Each asterisk represents one facility. VL, viral load.

evaluations are robust and can be used to compare facilities, geographical areas and demographic subgroups. Fifth, because the data reflect all lab-monitored patients with HIV 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 VL monitoring at nearly all government health facilities in South Africa, from April 2004 to April 2018. During this period, 63 million CD4 and VL tests were conducted at 5483 facilities.

A longitudinal cohort requires a patient identifier in order to follow up 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 1000 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 anonymised 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, patients with HIV had laboratory tests at 4751 facilities within South Africa ([figure 1](#)).

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 from May 1 to April 30)	Patients entering into HIV care (n)*	Patients receiving first HIV VL test (n)†	Patients on ART (n)‡	Percent HIV VL suppressed out of all VL tests§
2004–2005	118 829	27 937	24 208	31.1
2005–2006	219 942	111 865	120 796	39.4
2006–2007	302 510	176 979	272 078	43.4
2007–2008	344 935	223 188	420 542	50.3
2008–2009	410 563	283 929	609 729	55.4
2009–2010	479 406	249 971	716 555	65.7
2010–2011	915 898	525 457	1 066 016	72.3
2011–2012	713 101	582 175	1 465 428	72.2
2012–2013	626 296	672 543	1 875 968	74.0
2013–2014	619 396	715 786	2 316 959	74.6
2014–2015	629 246	785 503	2 801 889	77.1
2015–2016	666 492	952 256	3 429 148	79.6
2016–2017	629 754	919 229	3 951 406	79.4
2017–2018	461 076	913 604	4 419 017	83.2

*Numbers of patients into HIV care are the number of patients with a first CD4 or first VL.

†Number of patients receiving first HIV VL test is the number of first of VL tests.

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

§Per cent of HIV VL suppressed are $100 \times (\text{the number of VL} < 400 \text{ cp/mL} / \text{total number of VL tests})$ during the same 18-month period. Patients who did not receive a VL test are not included.

ART, antiretroviral therapy; VL, viral load.

Of the 12.6 million patients in the cohort, 7.1 million ever initiated ART. To estimate the number of patients currently undergoing ART, we limited the cohort to the last 18 months from November 2016 to 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 the Joint United Nations Programme on HIV/AIDS (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 2004–2005, 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 VLs 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 standardised 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 workup. Laboratory-based monitoring of CD4 and VL has been a standard component of national HIV care and treatment

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 programme, VL tests were conducted at treatment initiation, and then every 6 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 VLs have been conducted, which corresponds to a median of 2 (IQR 1–5) VLs per person. This translates to 23 381 315 total person years of follow-up

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) ³⁶	<p>Criteria for ART initiation</p> <p>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</p> <p>ART monitoring</p> <p>CD4: staging and every 6 months</p> <p>VL: staging and every 6 months</p>
2010	The South African Antiretroviral Treatment Guidelines 2010 ³⁷	<p>Criteria for ART initiation</p> <p>CD4 count <200 cells/mm³ irrespective of clinical stage or CD4 count <350 cells/mm³ in patients with TB/HIV or pregnant women or WHO stage IV irrespective of CD4 count or MDR/XDR irrespective of CD4</p> <p>Require fast track (ie, ART initiation within 2 weeks of being eligible)</p> <p>Pregnant women eligible for lifelong ART, patients with very low CD4 (<100)</p> <p>Stage 4, CD4 count not yet available</p> <p>MDR/XDR TB</p> <p>ART monitoring</p> <p>CD4: staging, 6 and 12 months and annually thereafter</p> <p>VL: 6 and 12 months and annually thereafter</p>
2011	Circular on New Criteria for Initiation of Adults on ART at CD4 Count of 350 cells/mm ³ and below, 26 August 2011	Baseline CD4 for initiation at CD4 <350 cells/mm ³
2012	Accelerating Access to ART Services and Uptake (circular), 17 April 2012	<p>Criteria for ART initiation</p> <p>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</p> <p>Require fast track (ie, same day ART initiation)</p> <p>Pregnant women eligible for lifelong ART, patients with low CD4 (<200)</p> <p>Stage 4, CD4 count not yet available</p> <p>MDR/XDR TB</p> <p>Screening and treatment of patients with very low CD4 counts (<100) for cryptococcal infection</p> <p>ART monitoring</p> <p>CD4: staging, 12 months and annually thereafter</p> <p>VL: 6 and 12 months and annually thereafter</p>
2013	The South African Antiretroviral Treatment Guidelines 2013 ³⁸	<p>Criteria for ART initiation</p> <p>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 resistance or sensitivity, including extrapulmonary TB) and WHO stage 3 or 4 irrespective of CD4 count</p> <p>Require fast track (ie, ART initiation within 7 days of being eligible)</p> <p>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 comorbidity with CD4 count <50 (patients with <i>Cryptococcus meningitis</i> or TB meningitis (defer ART for 4–6 weeks)</p> <p>ART monitoring</p> <p>CD4: staging and 1 year</p> <p>VL: 6 and 12 months and annually thereafter</p>

Continued

Table 3 Continued

Year	Guidelines	ART initiation criteria/CD4 and VL monitoring
2015	National Consolidated Guidelines for the Prevention of Mother-to-Child Transmission of HIV and the Management of HIV in Children, Adolescents and Adults 2015 ³⁹	Criteria for ART initiation CD4 count < 500 cells/mm ³ 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 (ie, 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, 22 August 2016 ⁴⁰	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

ART, antiretroviral therapy; TB, tuberculosis; VL, viral load.

on HIV treatment for a median of 2.1 (IQR 1.0–5.0) years per person.

Variables measured

The variables in the NHLS National HIV Cohort are described in [table 4](#) along with per cent completeness. The cohort includes the type of test, tests results, test date and geographical information along with the 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 programme. Information on facility can be linked to facility geocodes, which can be mapped ([figure 1](#)), aggregated to the local municipality, district and provincial levels, or linked to external population-level data such as HIV prevalence, poverty levels or the population age distribution. Information

Table 4 Variables in the NHLS National HIV Cohort

Variable	Description	Per cent completeness of variable (%)
BU_uniq_ID	Unique patient identifier	100
Episode_no	NHLS Episode identifier	100
Sex	Sex	98.0
Age	Age at testing date (years)	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 (eg, CD4, VL)	100
Test_date	Date of test	100
Test_result	Test result	100

NHLS, National Health Laboratory Service; VL, viral load.

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	2004–2008	2008–2013	2013–2018	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 infants and young children	1 129 249	1 448 212	639 528	3 216 989
Haemoglobin	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

ART, antiretroviral therapy; NHLS, National Health Laboratory Service.

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 is 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, conduct, 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 epidemiological research for the complete national HIV care and treatment programme including patients not yet on ART; (2) assessment of concepts such as ‘CD4 recovery’, retention in care and ‘viral resuppression’ 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; and (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 (eg, clinic staffing and size) and population-based data (eg, 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.²³
- ▶ *There is large spatial heterogeneity in the HIV care cascade*—estimated rates of VL testing and suppression from April 2014 to March 2015 across public facilities. We identified wide spatial variation in VL suppression, ranging from 69% to 82% at the provincial level.²⁴ [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.²⁵
- ▶ *A high proportion of patients present with advanced disease*—documented that a consistently large share (~33%) of patients entered into care with a CD4 of <200 cells/ μ L from 2011 to 2016 despite increased CD4 treatment eligibility standards.²⁶ Late presentation persisted into the Universal Test and Treat (UTT) era.²⁷

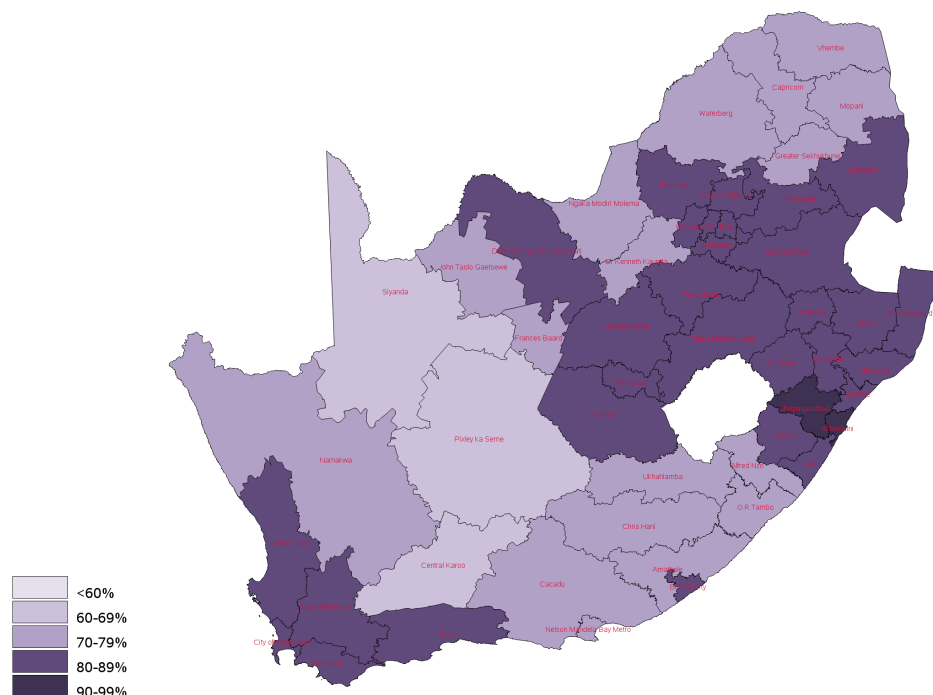


Figure 2 Viral load suppression by district in South Africa, 1 May 2017–30 April 2018.

- ▶ *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-fold to 20-fold increase in the numbers of adolescents on ART from 2004 to 2007 to 2012–2014 was found and resulting in a ‘wave’ of adolescents aged 15–19 years old who will require HIV treatment over the next decade due to both the ageing of perinatally infected children into adolescence and increased numbers of adolescent girls seeking HIV care for the first time.²⁸
- ▶ *Viral monitoring for treatment failure.* VL monitoring is conducted to identify patients at risk of treatment failure, to target these patients with adherence counselling and to switch them to second-line therapy if needed. Comparing outcomes among patients with VL results just above versus just below the 1000 copies/mL³ threshold, we found that while having a VL of >1000 increases the probability of follow-up VL monitoring, most patients with elevated VL do not receive monitoring within recommended timelines.^{29 30}

Strengths and limitations

The main strength of the NHLS National HIV Cohort is, first, 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 programme implementation (eg, increases in HIV testing in some areas before others,

or implementation of the National Adherence Strategy) to evaluate the impacts of these interventions.³¹

Second, the cohort is unique in its ability to explore outcomes without the problem of silent transfers.³² Silent transfers, patients who move from one facility to another without informing their sending clinic, lead 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.³³

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,³⁴ we do not have data on medication or visit adherence, or clinical diagnoses that would be useful for describing a national programme.

Second, the matching techniques we use are not perfect and lead to both overmatching and undermatching 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 overmatching and undermatching) the main source of error in studies using this cohort, a problem that can be explicitly modelled

using quantitative bias analysis.³⁵ We have also assessed the sensitivity of our results to matching parameters and found the results to be quite robust.²³

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) integrating the cohort with clinical databases, and (3) expanding the scope of the cohort to include tuberculosis and other lab-monitored comorbidities.

Author affiliations

¹Department of Global Health, Boston University School of Public Health, Boston, Massachusetts, USA

²Health Economics and Epidemiology Research Office, University of the Witwatersrand, Johannesburg-Braamfontein, South Africa

³Departments of Global Health and Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA

⁴Centre for HIV and STIs, National Institute for Communicable Diseases, Sandringham, South Africa

⁵Research Computing Services, IS&T, Boston University, Boston, Massachusetts, USA

⁶Academic Affairs, Research & Quality Assurance, National Health Laboratory Service, Johannesburg, South Africa

⁷Department of Molecular Medicine and Haematology, University of the Witwatersrand, Johannesburg-Braamfontein, South Africa

⁸National Priority Programmes, National Health Laboratory Service, Sandringham, South Africa

Twitter Cornelius Nattey @corneliusnattey

Acknowledgements Jaco Grobler was instrumental in programming the first iteration of the National Health Laboratory Service (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 acknowledge the contributions of the dedicated staff at the NHLS 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.

Contributors WBM, JB, SC, MM, MPF, KB, ATB, JP, WS and SC made substantial contributions to the conception or design of the work. WBM, JB, SC, KB, JP, CN, DO and KM made substantial contributions to the acquisition, analysis or interpretation of data for the work. WBM, JB, MM, MPF and CN drafted the work. SC, KB, ATB, JP, WS and SC critically revised the work for important intellectual content. All authors gave final approval of the version to be published and agreed 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. WBM is responsible for the overall content as guarantor.

Funding 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.

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein) or of any geographical or locational reference does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by University of the Witwatersrand HREC (M150429) and Boston University institutional review board (H-31968 and H-33442). We received a waiver of the need for a consent form as this was essentially a retrospective record review of millions of records, and the construction of the cohort would not have been practical without the waiver. The final cohort does not have any personal identifiers.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. 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 Service (NHLS) 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 NHLS. To find out more about the cohort, contact academic.research@nhls.ac.za.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

William B MacLeod <http://orcid.org/0000-0001-8003-8874>

Matthew P Fox <http://orcid.org/0000-0001-9887-0634>

Alana T Brennan <http://orcid.org/0000-0002-7746-0668>

Dorina Onoya <http://orcid.org/0000-0002-2664-7342>

REFERENCES

- UNAIDS. UNAIDS data 2019 annual report. Geneva; 2019.
- 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;20:1–12.
- Fox MP, Maskew M, MacPhail AP, *et al*. Cohort profile: the Themba Lethu clinical cohort, Johannesburg, South Africa. *Int J Epidemiol* 2013;42:430–9.
- Haas AD, Zaniewski E, Anderegg N, *et al*. Retention and mortality on antiretroviral therapy in sub-Saharan Africa: collaborative analyses of HIV treatment programmes. *J Int AIDS Soc* 2018;21:1–7.
- 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* 2012;41:1256–64.
- Houlihan CF, Bland RM, Mutevedzi PC, *et al*. Cohort profile: Hlabisa HIV treatment and care programme. *Int J Epidemiol* 2011;40:318–26.
- Kluber SA, Fox MP, LaValley M, *et al*. Do HIV treatment eligibility expansions crowd out the sickest? Evidence from rural South Africa. *Trop Med Int Health* 2018;23:968–79.
- 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 Health* 2018;23:206–20.
- MacLeod WB, Maskew M, A Jaffray I, *et al*. The feasibility of using screening criteria to reduce clinic visits for stable patients on antiretroviral therapy in South Africa. *J Acquir Immune Defic Syndr* 2013;62:e82–6.

- 10 Fox MP, Shearer K, Maskew M, *et al.* Treatment outcomes after 7 years of public-sector HIV treatment. *AIDS* 2012;26:1823–8.
- 11 Brennan AT, Maskew M, Iwe 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;16:18794–12.
- 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;81:439–47.
- 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;17:18908.
- 14 Virnig BA, McBean M. Administrative data for public health surveillance and planning. *Annu Rev Public Health* 2001;22:213–30.
- 15 Lee CH, Yoon H-J. Medical big data: promise and challenges. *Kidney Res Clin Pract* 2017;36:3–11.
- 16 NHLS. National CD4 testing programme. Available: <https://www.nhls.ac.za/priority-programmes/cd4/> [Accessed 23 Dec 2021].
- 17 Fellegi IP, Sunter AB. A theory for record linkage. *J Am Stat Assoc* 1969;64:1183–210.
- 18 Jaro MA. Probabilistic linkage of large public health data files. *Stat Med* 1995;14:491–8.
- 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.
- 20 Bor J, MacLeod W, Oleinik K. Building a national HIV cohort from routine laboratory data: probabilistic record-linkage with graphs. *bioRxiv* 2018;450304.
- 21 UNAIDS. Country Factsheets South Africa. 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.
- 23 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:e1002589.
- 24 MacLeod W, Bor J, Crawford K. 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.
- 25 Bor J, Gage A, Onoya D. Quality of care in South African facilities providing art services. *PLoS Med* 2021;18.
- 26 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:S111–7.
- 27 Bor J, Fox MP, Nattey C. Late presentation persists under UTT in South Africa: a national cohort study. *Conference on Retroviruses and Opportunistic Infections*, Boston, Massachusetts, 2020.
- 28 Maskew M, Bor J, MacLeod W, *et al.* Adolescent HIV treatment in South Africa's national HIV programme: a retrospective cohort study. *Lancet HIV* 2019;6:e760–8.
- 29 Harlow AF, Bor J, Brennan AT. 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;27708:1–19.
- 30 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:e25542.
- 31 Fox MP, Nattey C, Maskew M, *et al.* *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.
- 32 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.
- 33 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;14:1–20.
- 34 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;17:41.
- 35 Lash TL, Fox MP, Fink AK. *Applying quantitative bias analysis to epidemiologic data.* Dordrecht, Heidelberg, London, New York: Springer, 2009.
- 36 South African National Department of Health. National antiretroviral treatment guidelines; 2004: 328.
- 37 South African National Department of Health. The South African antiretroviral treatment guidelines 2010; 2010.
- 38 South African National Department of Health. The South African antiretroviral treatment guidelines 2013; 2013.
- 39 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.
- 40 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.