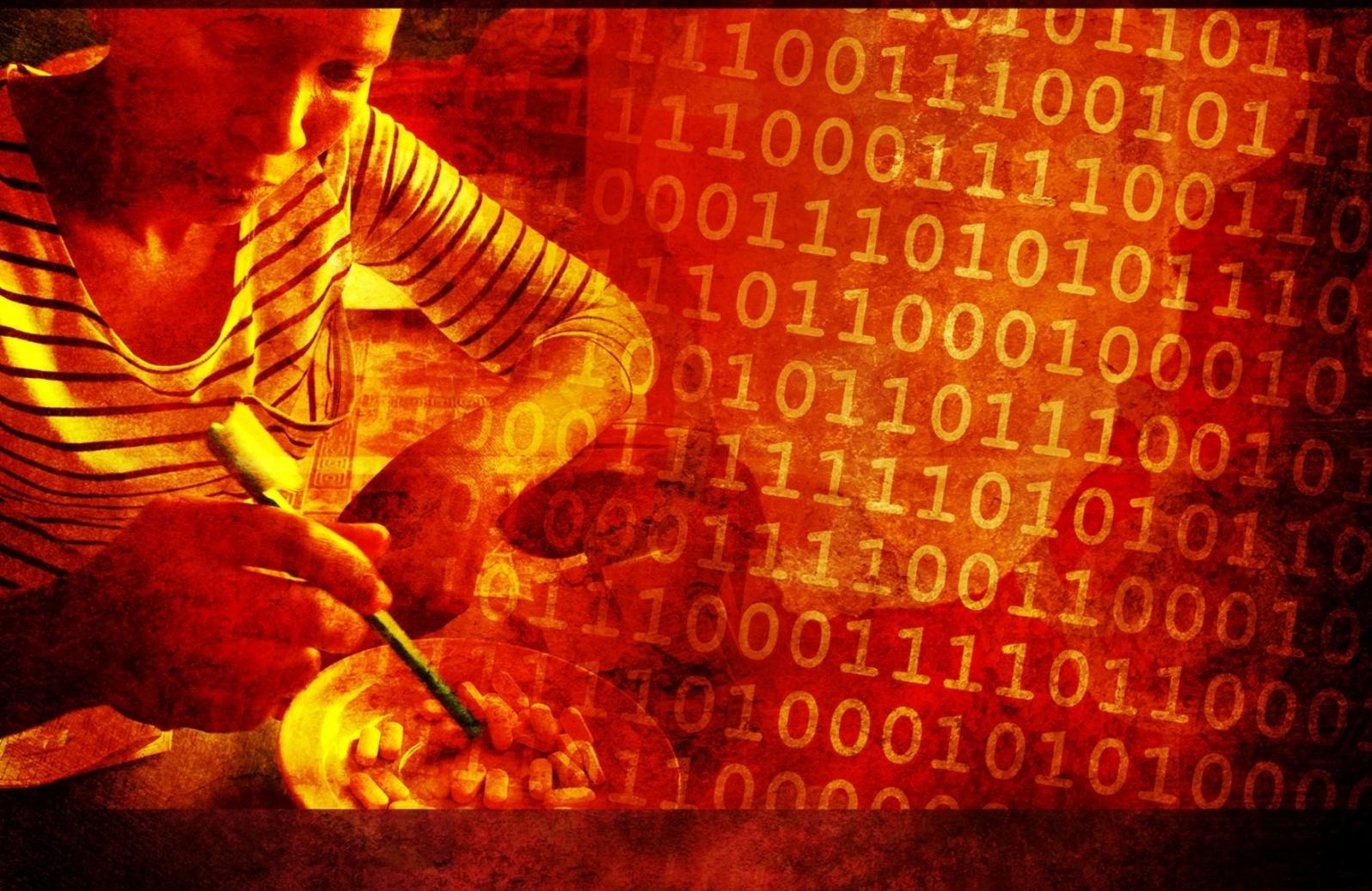


Analysis of Big Data for better targeting of ART Adherence Strategies

Spatial clustering analysis of viral load suppression by South African province, district, sub-district and facility (April 2014 – March 2015)

November 2015



health

Department:
Health
REPUBLIC OF SOUTH AFRICA

This study is made possible by (a) financing and technical support from the World Bank and (b) the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the co-authors of the report and do not necessarily reflect the views of USAID or the United States Government. In addition, the findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent.

Analysis of Big Data for better targeting of ART Adherence Strategies

01001011110101110110110110100011101001111011010101110010100

Spatial clustering analysis of viral load
suppression by South African province,
district, sub-district and facility
(April 2014–March 2015)

November 2015

This page is intentionally left blank

Table of Contents

| | |
|--|-----------|
| Abbreviations | iii |
| Acknowledgements | v |
| EXECUTIVE SUMMARY | 1 |
| Background | 1 |
| Methods | 1 |
| Results | 2 |
| Discussion | 5 |
| INTRODUCTION | 7 |
| METHODS | 8 |
| Patient-level Linking | 8 |
| Record Linkage Procedure | 9 |
| Validation of Record Linkage | 11 |
| Record Linkage Results | 11 |
| Clinic Level Linking | 12 |
| Steps in the Construction of the Linking Files | 13 |
| NHLS CDW Laboratory Dataset | 13 |
| RESULTS | 14 |
| Viral Suppression at National and Province Level | 14 |
| Viral Suppression at District And Sub-district Level | 16 |
| Viral Load Suppression at Health Facility Level | 26 |
| Viral Suppression by Age and Gender | 29 |
| DISCUSSION | 30 |
| References | 35 |

Figures

| | |
|---|----|
| ES Figure 1 Population with high viral load (VL >1,000 cp/mL) by health sub-district, April 2014–March 2015, South Africa | 3 |
| Figure 1 Flow diagramme for the record linkage procedures | 12 |
| Figure 2-A Proportion viral load suppression (VL <400 cp/mL) by district, April 2014–March 2015, South Africa | 21 |
| Figure 2-B Proportion viral load suppression (VL <400 cp/mL) by health sub-district, April 2014–March 2015, South Africa | 21 |
| Figure 3-A Population with viral load suppression (VL <400 cp/mL) by district, April 2014–March 2015, South Africa | 22 |
| Figure 3-B Population with viral load suppression (VL <400 cp/mL) by health sub-district, April 2014–March 2015, South Africa | 22 |

| | | |
|------------|--|----|
| Figure 4-A | Proportion high viral load (VL >1,000 cp/mL) by district, April 2014–March 2015, South Africa | 23 |
| Figure 4-B | Proportion high viral load (VL >1,000 cp/mL) by health sub-district, April 2014–March 2015, South Africa | 23 |
| Figure 5-A | Population with high viral load (VL >1,000 cp/mL) by district, April 2014–March 2015, South Africa | 24 |
| Figure 5-B | Population with high viral load (VL >1,000 cp/mL) by health sub-district, April 2014–March 2015, South Africa | 24 |
| Figure 6-A | Proportion of HIV patients with known suppression (had VL test <400 cp/mL) by district, April 2014–March 2015, South Africa | 25 |
| Figure 6-B | Proportion of HIV patients with known suppression (had VL test <400 cp/mL) by health sub-district, April 2014–March 2015, South Africa | 25 |
| Figure 7-A | Average HIV clinic size by district, April 2014–March 2015, South Africa | 27 |
| Figure 7-B | Average HIV clinic size by health sub-district, April 2014–March 2015, South Africa | 27 |
| Figure 8 | Scatterplot of standardised viral load suppression versus local weighted average with fitted line showing correlation among health facilities, April 2014–March 2015, South Africa | 28 |

Tables

| | | |
|------------|--|----|
| ES Table 1 | Summary table of viral load results by level..... | 2 |
| ES Table 2 | Summary table of viral load results in the impact evaluation’s health facilities..... | 4 |
| Table 1 | Validity of unique patient identifiers | 12 |
| Table 2 | Viral load (VL) suppression for people in care and on ART results by province, April 2014–March 2015 | 15 |
| Table 3 | Comparison of proportions of viral load (VL) tests done and VL suppression for most recent test and all tests by province, April 2014–March 2015 | 17 |
| Table 4 | Viral load (VL) suppression for people in care and on ART results by district, April 2014–March 2015 | 18 |
| Table 5 | Viral load suppression for people in care and on ART results by facility, April 2014–March 2015 | 26 |
| Table 6 | Mean proportion viral load suppression by facility size | 26 |
| Table 7 | Global autocorrelation statistics for viral load suppression by facility | 28 |
| Table 8 | Viral load (VL) suppression for people in care and on ART results by age group, April 2014–March 2015 | 29 |
| Table 9 | Viral load (VL) suppression for people in care and on ART results by gender, April 2014–March 2015 | 29 |

Abbreviations

| | |
|-----------------|---|
| AIDS | Acquired immunodeficiency syndrome |
| ART | Antiretroviral therapy |
| CCMT | Comprehensive Care, Management, and Treatment |
| CDW | Corporate Data Warehouse |
| CI | Confidence Interval |
| cp/mL | copies per millilitre |
| DHIS | District Health Information System |
| DOB | Date of Birth |
| HIV | Human Immunodeficiency Virus |
| M&E | Monitoring & Evaluation |
| NDoH | National Department of Health |
| NHLS | National Health Laboratory Service |
| RA | Research Assistant |
| SCC | Boston University Shared Computing Cluster |
| TB | Tuberculosis |
| TIER.Net | Tiered ART Monitoring Strategy |
| TROA | Total Remaining on ART |
| VL | HIV Viral Load |

This page is intentionally left blank

Acknowledgements

This report was authored by:

William MacLeod^{1,2}, Jacob Bor^{1,2}, Kathryn Crawford³, and Sergio Carmona^{4,5}

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

² Health Economics and Epidemiology Research Office (HE2RO), Faculty of Medical Sciences, University of Witwatersrand, Johannesburg, South Africa

³ Department of Environmental Health, Boston University School of Public Health, Boston, USA

⁴ National Health Laboratory Service, Johannesburg, South Africa

⁵ Department of Haematology and Molecular Medicine, Faculty of Medical Sciences, University of Witwatersrand, Johannesburg, South Africa

Other contributors

National Department of Health: Yogan Pillay, Mokgadi Phokojoe, and Tshepo Molapo

World Bank: Nicole Fraser, Zara Shubber and Marelize Görgens

National Health Laboratory Service/Wits: Wendy Stevens

Boston University School of Public Health: Matthew Fox

Health Economics and Epidemiology Research Office: Mhairi Maskew

Right to Care: Ian Sanne

Support for developing master list of health facilities

Catherine White (Clinton Health Access Initiative), Zoe McLaren (University of Michigan), Nassim Caseem (NHLS), Louzanne Oosthuizen (University of Stellenbosch) and Calle Hedberg (HISP)

Support for patient-level linking

Sue Candy (NHLS CDW) and Jaco Grobler (BiTanium) and Katia Oleinik (Boston University)

Peer review of the report

Andrew Phillips, Nathan Ford, Thomas Finkbeiner, Joy de Beyer, Paolo Belli, Zlatan Sabic

Funding sources

Right to Care (USAID), PEPFAR, DFID, World Bank, NDOH, NHLS



This page is intentionally left blank

EXECUTIVE SUMMARY

Background

South Africa has more persons living with HIV and patients on antiretroviral therapy (ART) than any other country in the world (World Health Organisation 2014). Most ART patients are treated in public sector facilities, but there is a lack of information at national, district and health facility level on the outcomes of the treatment programme. In particular, there is a dearth of information about the proportion of ART patients who are virally suppressed (the ultimate outcome of an HIV treatment programme for maximum individual clinical benefit and population-level HIV prevention benefits in terms of reduced HIV transmission). This is in part because of fragmented health information systems, the lack of unique identifiers captured for each ART patient, and other historical reasons. The country's District Health Information System (DHIS) reports numbers of patients on ART (Total Remaining on ART—TROA), the proportion of patients with viral load (VL) tests captured in that system, and the proportion of these tests with a result indicating viral suppression. The most recently reported ART cohort data from the DHIS showed 46% of ART patients with a VL test and 83% VL suppression among these (Overmeyer, 2015).

Laboratory monitoring test results are stored in the National Health Laboratory Services' (NHLS) database. This database has the potential to provide important strategic information on the ART programme reach and quality if the testing data were to be systematically linked to patient data. However, linking the data is difficult because the NHLS and Department of Health (DOH) use separate lists of health facility names and identifications, and the NHLS test data are not automatically linked to the patient data stored in the DHIS. The NHLS database is a database of tests, not persons, and—given that a person can test more than once and via different health facilities—there is a disconnect between data systems. This makes individual and facility linkages and determinations of the percentage of ART patients with viral load tests done, and viral suppression by ART patient, difficult and limits spatial analysis of laboratory data for targeting ART adherence investments across South Africa.

Methods

1. In order to analyse the NHLS data we developed an algorithm for patient level linking of CD4 count and VL test data using probabilistic record linkage methods.
2. Incorporating the work of other institutions, we merged DHIS and NHLS facility lists. This allowed us to link DHIS reports of TROA data with the NHLS database of

public sector VL tests at the facility level to measure the proportion of patients receiving a VL test in a 12-month period and to group the tests results in four categories (<400, 400–1000, >1000, and >10,000 copies/mL). We used the unique patient identifier in the NHLS lab database to select only the most recent VL test in a 12-month period, for each patient.

3. We then determined the proportion of viral load tests done (VLD) and proportion of ART patients virally suppressed (VLS) by province, district, sub-district and health facility, and provided tables and maps. Next, we used the VLS results from the health facilities to match intervention and control facilities in the NDOH/World Bank impact evaluation of the adherence guidelines. We also assessed the relationship between facilities' TROA size and viral suppression levels.
4. In order to test the hypothesis that viral suppression levels at one health facility are independent of values at neighbouring health facilities, two indices of spatial autocorrelation were calculated (Moran's I and Geary's c) and a scatter plot developed.

Results

From April 2014 through March 2015, 3,775 public facilities reported 2,993,125 patients on ART. During this period, 2,199,890 unique patients received 2,995,133 VL tests. At the national level, 75% of patients had received a VL test in the previous 12 months and 78% of those tested had a suppressed VL (below 400 copies/mL), 19% had VL results above 1,000 copies/mL, and 12% were above 10,000 copies/mL (see ES Table 1). The proportion of viral load suppression increased with age from a low of 51% in children under 5 years to 83% in patients 50 years and older. One in three ART patients aged under 25 years was not virally suppressed. Approximately 5% more female patients had viral suppression than to males. Sixteen percent of male and 11% of female ART patients had a VL test result above 10,000 copies/mL.

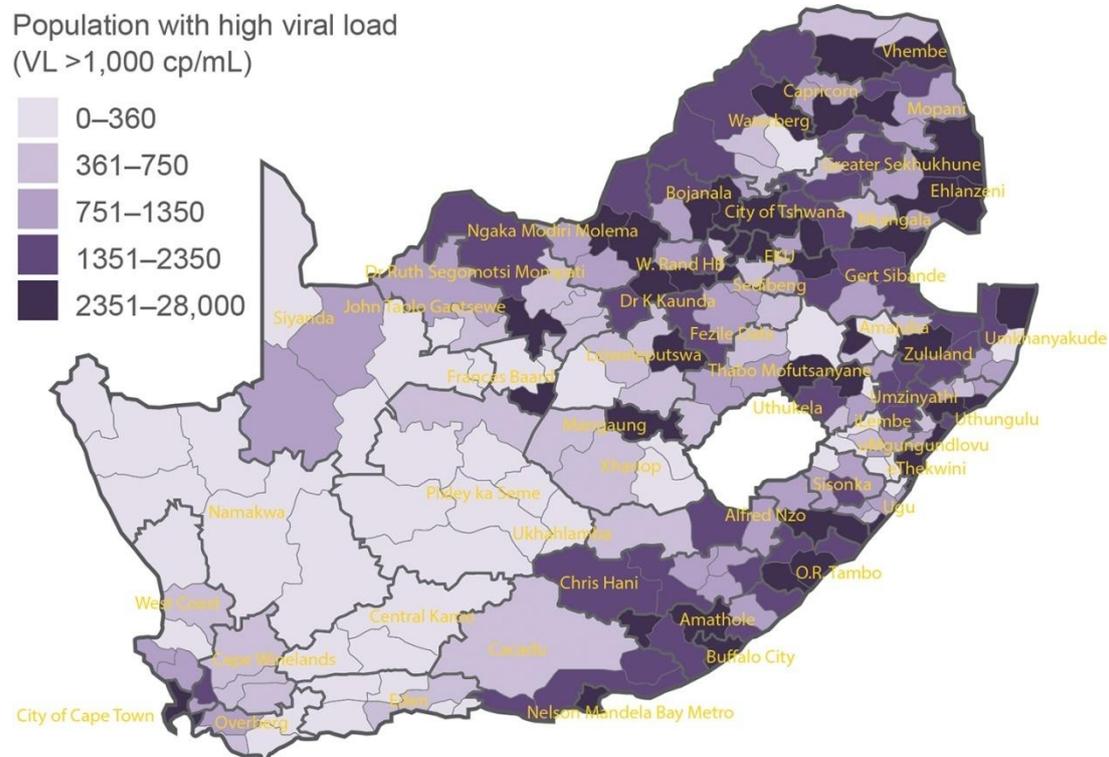
ES Table 1 Summary table of viral load results by level

| | VL Test in 12 month period, VLD | | VL<400 cp/mL, VLS | | VL>10000 cp/mL | | Known to be suppressed (VLD x VLS) | |
|-----------------------|---------------------------------|---------|----------------------|---------|----------------|---------|------------------------------------|---------|
| | Lowest | Highest | Lowest | Highest | Lowest | Highest | Lowest | Highest |
| National ¹ | 75% | | 78% | | 12% | | 58% | |
| Province ¹ | 71% | 82% | 69% | 82% | 10% | 20% | 52% | 65% |
| District ² | 54% | 99% | 47% | 86% | 8% | 35% | 34% | 73% |
| Facility ³ | n/a | n/a | 20% ^{&} | 96% | 1% | 67% | n/a | n/a |

Sources: 1=see table 2 for details; 2=see Table 4 for details; 3=see Table 5 for details; &=the few facilities with lower percentages had sample sizes too small to take into account.

At the provincial level, among patients who had had a VL test in the last 12 months, the proportion with suppressed VL ranged from 69 to 82% (ES Table 1). Three provinces had 25% or more patients with VL test results above 1,000 copies/mL, and one province had 20% of VL tests above 10,000 copies/mL. Comparison across districts showed much more heterogeneity in terms of VL test coverage (54% - 99%), VL suppression (47% - 86%) and high viral load above 10,000 copies/mL (8% - 35%). These differentials were even more pronounced in the comparison across sub-districts (see ES Figure 1) and health facilities (see ES Table 1). Only 3.4% of all ART facilities met the 90% target for viral load suppression; 200 clinics had VL suppression levels below 50%.

ES Figure 1 Population with high viral load (VL >1,000 cp/mL) by health sub-district, April 2014–March 2015, South Africa



Source: NHLS Corporate Data Warehouse, May 2015 data extract.

We also calculated the overall proportion of patients with *known* suppression (product of VL done and VL suppressed—see summary table ES Table 1). The proportion of patients known to be suppressed was 58% nationally; Western Cape (65%) and Free State (64%) provinces had the highest proportion of patients known to be suppressed.

Median monthly TROA per facility ranged from 1 to 18,786. TROA was associated with VL suppression at the facility level, controlling for proportion of patients tested and province. Two-thirds of all ART patients are seen in the 25% largest facilities and had VL suppression levels 14.5% (95% CI: 13.1–15.9) higher than the 25% smallest facilities. Spatial correlation analysis of VL data indicated that neighbouring ART facilities tend to have similar VL suppression rates.

The VL suppression results for individual health facilities were used to pair health facilities by similar levels of suppression for the NDOH/World Bank impact evaluation of the adherence guidelines (see ES Table 2). The 24 health facilities in four selected health districts were also matched for similar TROA between the pairs of intervention and control facilities.

ES Table 2 Summary table of viral load results in the impact evaluation's health facilities (the matched controls are shown below the respective intervention facilities in *italics*)

| District and facilities | Midpoint TROA | Patient received VL (%) | % VL<400 cp/ml |
|---|---------------|-------------------------|----------------|
| Gauteng: Ekurhuleni MM | | (76) | 77 |
| Khumalo Clinic | 1,898 | 1,174 (62) | 84 |
| <i>Zonkizizwe 1 Clinic</i> | <i>1,056</i> | <i>820 (78)</i> | <i>85</i> |
| Phola Park CHC | 2,688 | 2,640 (98) | 88 |
| <i>Ramokonopi CHC</i> | <i>2,731</i> | <i>2,219 (81)</i> | <i>79</i> |
| Motsamai Clinic | 1,084 | 782 (72) | 72 |
| <i>Tamaho Clinic</i> | <i>1,055</i> | <i>903 (86)</i> | <i>73</i> |
| KwaZulu-Natal: uThungulu DM | | (70) | 71 |
| Buchanana Clinic | 1,110 | 952 (86) | 83 |
| <i>Ntambanana Clinic</i> | <i>1,183</i> | <i>996 (84)</i> | <i>85</i> |
| King Dinizulu Clinic | 1,889 | 1,720 (91) | 71 |
| <i>Nkwalini Clinic</i> | <i>1,049</i> | <i>942 (90)</i> | <i>68</i> |
| Thokozani Clinic | 4,981 | 2 848 (57) | 84 |
| <i>Nseleni CHC</i> | <i>7,265</i> | <i>3 596 (49)</i> | <i>84</i> |
| Limpopo: Mopani DM | | (68) | 69 |
| Giyani CHC | 1,682 | 900 (54) | 73 |
| <i>Dzumeri CHC</i> | <i>1 437</i> | <i>951 (66)</i> | <i>75</i> |
| Mugodeni Grace CHC | 1,722 | 1,597 (93) | 74 |
| <i>Motupa Clinic</i> | <i>972</i> | <i>648 (67)</i> | <i>69</i> |
| Tzaneen Clinic | 1,350 | 916 (68) | 77 |
| <i>Nkowankowa CHC</i> | <i>1,013</i> | <i>235 (23)</i> | <i>77</i> |
| North West: Bojanala Platinum DM | | (63) | 76 |
| Hebron Clinic | 1,128 | 1,057 (94) | 83 |
| <i>Bafokeng CHC</i> | <i>3,309</i> | <i>2,209 (67)</i> | <i>79</i> |
| Letlhabile CHC | 4,155 | 2,732 (66) | 83 |
| <i>Majakaneng Clinic</i> | <i>1,205</i> | <i>1,088 (90)</i> | <i>80</i> |
| Tlhabane CHC | 4,024 | 2,198 (55) | 81 |
| <i>Wonderkop Clinic</i> | <i>4,172</i> | <i>1,077 (26)</i> | <i>84</i> |

Discussion

Geographic differentials in VL test coverage and VL suppression: This analysis demonstrates, in South Africa, the diversity in the proportion of patients getting a VL test annually and patient virally suppressed. The variation shows by geographic location and facility, as well as by age, gender and demography. The two provinces with best VL suppression had suppression levels 13% higher than the three provinces with lowest levels, and when comparing across districts the differential was even greater at 39%. Viral suppression achieved 96% in one health facility, but was as low as 20% at the other end of this vast spectrum of viral suppression results.

A pattern of VL suppression by size of patient population: The districts and health facilities with larger patient populations had better viral suppression levels than districts and facilities with low patient numbers. This analysis suggests that even if facilities have high patient loads, they can still achieve good viral suppression results, whereas facilities with small ART populations seem to have more difficulties in achieving viral suppression in their patients.

VL suppression varied with the age and gender of the ART patients: Viral load suppression also was associated with age, with higher levels of viral suppression in the older age groups. The high rate of unsuppressed viral loads in the 0-4-year-old age group may be attributed in part to the lack of understanding of treatment and monitoring procedures by the parents or guardians of this population. These results indicate that high VL is a critical problem affecting young ART patients and particularly children under 5 years. Females received approximately two-thirds of VL tests, about the same as their 67% share of all HIV positive persons on treatment nationally. Our finding that one in six male and one in nine female ART patients has a high risk of HIV transmission due to VL levels above 10,000 copies/mL is cause for concern and requires urgent targeting of treatment adherence support to these patients.

Consideration on the linked NHLS data and the routine data: The comparison of the NHLS to the DHIS results showed very similar proportions of VL suppressed but large differences in proportions of patients receiving a VL test. We hypothesize three possible reasons for these differences: incomplete coverage of facilities by the Tiered ART Monitoring Strategy (TIER.Net); if early TIER.Net adopters are better managed facilities they might have higher VL suppression than later TIER.Net adopters; and differences in completeness of data capturing.

Utility of this analysis: This analysis of the patient-level laboratory cohort demonstrates its ability to provide detailed information on viral load suppression in South Africa. The information can identify facilities with low VL suppression and low

proportions of patients tested in the previous 12 months, enabling failures in ART adherence and viral suppression to be identified and corrected.

The South African government target is 90% VL suppression among ART patients. This analysis shows that while some facilities are achieving this goal, it is not being met by any entire district or province. There is great potential to learn about successful promotion of ART adherence from the one-in-30 facilities with 90% or more of ART patients virally suppressed.

By any measure, the health system needs to improve compliance with the ART monitoring guidelines, the processing of laboratory information, and communication of VL test results at clinic level. More complete VL monitoring and reporting in the DHIS is crucial to enable satisfactory monitoring of the health of patients and of the treatment programme's performance and impact. This would enable managers routinely to combine the DHIS and NHLS data, providing a powerful tool for better clinical care and programmatic monitoring. In the short term, we recommended using a combination of NHLS laboratory data and DHIS data for programmatic and facility monitoring of the ARV programme. VL data, disaggregated to the level of decision-making, should be used to guide the allocation of resources for programme improvements.

This work shows that Big Data—in this case a secondary analysis of large sets of routine data and innovative methodology of record linkage—can inform programme improvements. Specifically, the present analysis highlights patient demographics, health facilities and districts that need enhanced adherence support, and it identifies success stories where viral suppression is achieved and ART patients regularly monitored to ascertain treatment effectiveness. ■

01 INTRODUCTION 01110110110

South Africa has the largest population of HIV-infected people in the world and the largest population of people on antiretroviral therapy (ART), but lacks important strategic information on the effectiveness of the treatment management programme. In 2012, an estimated 6.4 million people were living with HIV and over 2.0 million people were on ART in South Africa (Shisana, O et al., 2014). In 2013, the South African government and international donors spent 2.3 billion USD on HIV/AIDS and tuberculosis (Meyer-Rath et al., 2015). However, implementation of comprehensive monitoring and evaluation (M&E) has lagged behind the rollout of the Comprehensive Care, Management and Treatment (CCMT) programme. More information is needed to refine treatment strategies to improve their impact and direct scarce resources where they are most needed.

With the aim of providing useful data for this purpose and establishing a standardised M&E system at CCMT clinics nationally, The National Department of Health (NDoH) began implementing the Tiered ART Monitoring Strategy in September 2011 (TIER.Net system). This monitors various clinical and laboratory indicators (National Department of Health, 2011). The three tiers are Tier 1: paper registers; Tier 2: non-networked electronic system; and Tier 3: a networked electronic medical record system. All facilities with more than 500 patients remaining on ART (TROA or Total Remaining On ART) have been prioritised to implement the TIER.Net system. As of November 2013, 967 of 1,803 eligible facilities had completed all stages of TIER.Net implementation. In the September 2013 report on ART data, NDoH reported that, as of the second quarter of 2013, “approximately 36% of patients had a viral load done and of the viral loads done nearly 80% of those patients on ART have suppressed viral loads” (National Department of Health, 2013). Data reported through TIER.Net is published by the District Health Information System (DHIS).

With regards to bridging the information gap caused by the inadequate M&E coverage, the National Health Laboratory Service (NHLS) has an existing database that is a potential secondary source of laboratory M&E indicators, such as proportion of patients with suppressed viral load, or CD4 count at initiation. The NHLS database contains all public sector viral load and CD4 count tests since 2010. The use of this database is currently limited because it lacks unique patient identifiers. In a sample of over 43 million lab tests, only 3.8% of the tests were associated with a valid national ID number.

HIV viral load suppression is the most important indicator of successful antiretroviral therapy (Ford et al., 2014; WHO, 2014). For ART patients with susceptible virus, good adherence to their medications typically leads to viral suppression (defined in South Africa as a viral load of less than 400 copies per millilitre, National Department of

Health, 2015). Suppression has benefits for the patient in that CD4 counts recover allowing the patient to fight off disease, and for the community as patients with low circulating viral loads are less infectious (Miller, Powers, Smith, & Cohen, 2013). Viral load results of 400 or more copies per mL (cp/mL) suggest that either the patient is non-adherent to treatment or that they are infected with drug-resistant virus. In either case, intervention is necessary—for adherence counselling, or to change the drug regimen (WHO, 2013).

In line with the World Health Organization’s 90-90-90 targets, the NDoH has set a viral suppression target of 90% of patients on ART in South Africa (UNAIDS, 2014). Current South African national treatment guidelines call for routine viral load tests 6 and 12 months after treatment initiation and yearly tests thereafter. Additional tests may be ordered if a clinician needs additional diagnostic information (National Department of Health, 2015).

In order to assess viral suppression rates in South Africa’s national treatment programme, we created a patient-level cohort using viral load and CD4 count test results stored in the NHLS’s Corporate Data Warehouse (CDW). This report presents an analysis of viral load suppression in this cohort at the provincial, district, sub-district and facility level in South Africa for a 12 month period from April 2014 to March 2015. ■

02 METHODS 0110110110100011

This section describes the analytical methods applied, and the creation of the patient and clinic level cohort from the NHLS CDW laboratory data.

Patient-level Linking

Health facilities use test requisition forms that state the patient name, date of birth, gender, and facility name. Patients may move between facilities, and their personal information is often recorded differently on different requisitions. To be able to track changes in VL/CD4, a patient-linked cohort of viral load test results stored in the NHLS CDW was created by linking patient-level information associated with each test result, to create a data base in which all test results for each individual patient were linked together. Recall that only a small proportion of the laboratory tests are associated with a national patient identification number. A patient-level cohort was created by linking viral load test results for each individual using names and date of birth. The creation of the patient level cohort was a multi-step procedure that used non-probabilistic and probabilistic (“fuzzy”) linking procedures using surname, first name and dates of birth associated with test results. The central task in fuzzy record linkage is to create a

unique identifier that simultaneously minimises over-matching (falsely combining records that should remain separate) and under-matching (falsely separating records that should be combined) (Christen, 2006).

The NHLS CDW has had a series of unique patient identifiers. The first matching algorithm was based upon exact matching of names, dates of birth and street address. This also used some hospital identification numbers. In mid-2014, the CDW implemented a fuzzy linking algorithm, but did not have the time and resources to evaluate the success of the unique patient identifier in regards to over and under-matching. The project team was asked to evaluate the CDW algorithm and make recommendations to improve it. The evaluation process identified a number of steps that improved the matching algorithm.

Record Linkage Procedure

The record linkage method consisted of five steps: data cleaning, exact-linking, pre-processing, fuzzy linking, and consolidation via network analysis. The methods were implemented using Boston University's Shared Computing Cluster (SCC), a network of high speed, multiprocessor computers that enabled development of a comprehensive algorithm that could compare millions of records in less than 12 hours per programme run.

The procedures were as follows.

1. Data cleaning

- a. Non-alphabetical characters, single initials, etc. were omitted.
- b. Words that were not names, e.g., "Mr.", "Dr.", "Mother", "No Name", etc. were omitted.
- c. Laboratory tests identified as being conducted as part of research studies and outside the national CCMT programme were omitted.

2. Pre-processing

- a. When there were multiple first or last names, matches were searched on all name components.
- b. First name / last name inversions were searched for and corrected. The version with the highest number of lab tests conducted was regarded as correct.
- c. A statistically informed "nicknames and alternate spellings" file (more than 16,000 names) was created, and matches on those nicknames/alternate spellings were searched for.

3. Fuzzy (probabilistic) linking

- a. A categorical variable for DOB similarity, distinguishing between exact matches, probable matches, plausible matches, and non-matches, was generated.
- b. The Jaro-Winkler algorithm was used to measure the similarity of first name pairs and last name pairs (Winkler, 2006). To compare names, all 20 million cleaned, exact-matched, pre-processed records were compared against each other. No blocking was conducted by name, allowing for detection of similar names even when the initial letters differed (e.g., Carl ~ Karl). To compare 20 million records against each other meant 400 trillion potential comparisons. To reduce the size of the problem, probabilistic matching (allowing for differences in names) was only conducted when the two dates of birth were +/- 11 years apart.
- c. The first name similarity score, last name similarity score, and DOB similarity score were combined into a single summary score equal to $\text{tot_sim} = \text{first_sim} * w + \text{last_sim} * (1-w) - \text{penalty}(\text{DOB_sim})$, where specific penalties were assigned for lack of similar DOBs. The weights and penalties were chosen empirically in an initial analysis, which sought simultaneously to minimise the number of unique IDs and the number of patient IDs combining a large number of exact matches. These penalties were chosen such that a cut-off of 0.90 would be a plausible initial threshold. The top 10 potential matches were identified for each of the 20 million exact-matched pre-processed records.
- d. Records were determined to belong to the same patient if the total similarity score was above a threshold value (see below).

4. Identifying unique patients via network analysis

- a. Using network concepts, the threshold for a match depended on the density of a region in the full network of names/DOBs. Specifically, in areas where names were commonly matched to others, a higher threshold was used; in sparser areas of the network, a lower threshold could be tolerated. This concept was operationalised with the following decision rule: a cluster would only be accepted as “a patient” if the diameter of the cluster (i.e., the shortest distance between the furthest nodes) was less than or equal to 3, and there were no more than 5 connections emanating from any given node. Together, these rules imply that no cluster greater than size 10 would be accepted as a patient.
- b. To implement this approach, a threshold of 0.90 was initially used, and all clusters were identified. The decision rule was then applied and the threshold was raised for clusters that did not satisfy the rule, eliminating low probability

links, to re-identify clusters. This procedure was iterated in .01 increments in the threshold until all clusters satisfied the rule.

Validation of Record Linkage

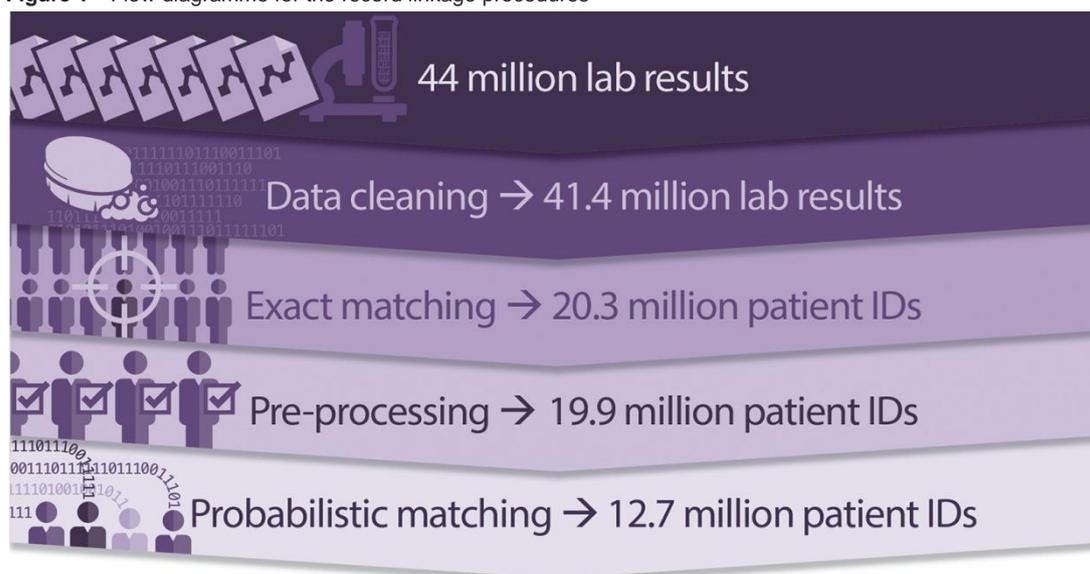
The final stage in the record linkage exercise was to validate it against a gold standard. In the case of the NHLS database, no gold standard exists that could capture the potential flow of patients across different sites within the national health system. Therefore, to validate the approach, a gold standard was manually constructed from NHLS data. Specifically, the project team:

- Drew a random sample of 1,000 lab results from the full database of 40 million VL and CD4 count results
- Started with a very liberal (low threshold) version of the matching algorithm and generated candidate matches for each of the 1,000 lab results. There was an average of 59 candidates per lab result (range 0, 838)
- Graded the match quality. For each of the candidate matches for each of the 1,000 lab results, four trained research assistants (RAs) graded the match quality on a 4-pt scale: 1 = almost certainly not a match, 2 = plausible match, 3 = probable match, 4 = almost certain match. A total of 59,000 candidate matches were graded twice each by separate RAs. After all matches were graded, refresher training was conducted and a third RA evaluated all candidate matches in which there was disagreement between the first two RAs in order to determine a final match quality
- Considered all 3s and 4s as “matches” and all 1s and 2s as “not matches”

Record Linkage Results

Figure 1 displays a flow diagram showing the numbers of observations at each stage of the linkage process. At the start, there were over 44.0 million lab results. After exact matching and pre-processing, there were 19,856,964 patient IDs. The probabilistic matching resulted in 12,684,248 unique patient IDs.

Table 1 displays total numbers of unique patients and Type 1 and Type 2 error probabilities for three available identifiers: Exact Match ID based on first name, last name, and DOB; the initial CDW unique patient ID that was developed in 2014; and the new Boston University unique patient ID.

Figure 1 Flow diagramme for the record linkage procedures**Table 1** Validity of unique patient identifiers

| | Exact match ID | CDW unique patient ID | New unique patient ID |
|--|----------------|-----------------------|-----------------------|
| Total number of patients identified | 20,258,414 | 18,124,393 | 12,684,248 |
| Number of patients with >100 results | 157 | 194,342 | 157 |
| Type 1 error (over-matching) ¹ | 0% | 17% | 15% |
| Positive Predictive Value (1 – Type 1) | 100% | 83% | 85% |
| Type 2 error (under-matching) ² | 36% | 27% | 11% |
| Sensitivity (1–Type 2 error) | 64% | 73% | 89% |

Notes: 1=Type 1 error is the probability that a lab result that was linked for a given patient identifier was not matched in the gold standard dataset. Type 1 error is over-estimated in all cases (probably by ~5 percentage points) because the cut-off date for our gold standard dataset was prior to the cut-off date for the fuzzy-matched dataset. We will fix this in future reporting. 2=Type 2 error is the probability that a lab result matched in the gold standard dataset was not matched by a given patient identifier.

Clinic Level Linking

The NDoH and NHLS lists of facilities use different names/identifiers for many facilities. In order to combine DHIS information on the patients on ART at the clinic level (TROA) with the test result data, the project undertook to link the DHIS and NHLS facilities lists. Analysis and utilization of NHLS data to inform policy has been hampered by the lack of a single, unified clinic-level linking file matching NHLS facilities to NDoH facilities.

The project team set out to construct a uniform NHLS-NDoH linking file, for use by NHLS and its partners. Several other efforts to construct clinic-level linking files were discovered. In October-November 2014, the project coordinated a call among several of these partners, which laid the groundwork for the process to create a unified list of matching NHLS-NDoH facilities.

Steps in the Construction of the Linking Files

Linking files were constructed by following these steps:

1. Obtaining official lists of facilities from NHLS and NDoH and compiling a master facility linking file using files from collaborators
2. Assessing modal matches and initial agreement scores from the merged linking files
3. Manually reviewing to identify “best” matches and updating agreement scores
4. Assessing NDoH geo-coordinates among matches

91% of all health facilities in the NHLS data base and that had any CD4/VL tests were linked to a NDoH facility with a medium to high level of agreement. Results from linking the NHLS list to the NDoH clinics reporting persons on ART care resulted in linking 91.8% of the facilities accounting for 95.8% of the patients on ART in care in NDoH facilities.

NHLS CDW Laboratory Dataset

The NHLS dataset contained 2,995,133 viral load results from April 1, 2014 to March 31, 2015 from public health facilities in South Africa. Viral load results from each patient in the cohort were restricted to the most recent result in the time period in order to limit the number of tests to no more than one per patient. The total number of patients having a viral load test in the time period was calculated, as well as the proportion of those patients with tests less than 400 cp/mL (virologically suppressed), between 400 and 1,000 cp/mL (considered at risk for poor adherence), and greater than 1,000 (non-adherent to treatment or drug-resistant) and 10,000 cp/mL (high risk for HIV transmission). The proportion of all viral load tests less than 400 cp/mL and greater than 1,000 cp/mL was calculated for comparison purposes.

These results were compared to data from the NDoH’s DHIS. The DHIS data were extracted from an Excel spreadsheet: DHIS_\$ZA_NDOH5_NIDS_OU5_11_13_ART.xlsm. This spreadsheet contained the monthly values for TROA from January 2014 to March 2015 for 3,775 clinics. The median TROA value of October and November 2014 for each clinic was used as the TROA value. October and November 2014 were chosen because they were the midpoint of the analysis period from April 2014 to March 2015. If a clinic did

not have TROA values for October and November, the median TROA value for the 12-month period April 2014 to March 2015 was used. One hundred and two facilities with 12 or fewer viral load test results were excluded.

Combining the number of unique patient viral load tests with the TROA, it was possible to calculate the proportion of patients that had a viral load test within the previous 12 months. The known proportion of virally suppressed patients was calculated by taking the product of the proportion of patients having a test in the time period and the proportion of virally suppressed tests.

Associations between clinic patient size (TROA) and viral load suppression were also calculated. For this analysis clinic size was divided into quartiles, and tested for differences in viral load suppression using mixed methods analysis of variance adjusting for multiple comparisons using Scheffe's test (Kleinbaum, Kupper, Muller, & Nizam, 1997).

In order to test the hypothesis that viral load suppression at one health facility is independent of values at neighbouring health facilities two indices of spatial autocorrelation were calculated (Moran's I and Geary's c) (Jacquez, 2008). A scatterplot of viral load suppression by the average suppression of neighbouring facilities was created. The scatterplot demonstrates standardised viral load suppression with a mean of zero and a standard deviation of one. The mean standardised viral load suppression was weighted for neighbouring facilities to account for the distance from the index facility. ■

03 RESULTS 0110110110110101110

Viral Suppression at National and Province Level

Table 2 shows the viral load suppression results nationally and by province for the time period from April 1, 2014 to March 31, 2015, obtained through linkage of patient and test result data and using the newly created health facilities lists. Nationally, there were slightly more than 2.9 million persons on ART. 78% of all persons who had a viral load test were virologically suppressed (VL less than 400 cp/mL), 19% had a viral load greater than 1,000 cp/mL, and 12% had a viral load greater than 10,000 cp/mL. The proportion of patients *known* to be suppressed (having had a VL test and found suppressed) was 58% nationally. Considering only those patients with a VL test, KwaZulu-Natal and Free State were the two provinces with the highest levels of virologic suppression at 82%, while Northern Cape, Limpopo and North West Provinces had the lowest levels of virologic suppression among those tested at 69–70%.

Table 2 Viral load (VL) suppression for people in care and on ART results by province, April 2014–March 2015

| Province | Total remaining on ART (TROA) (Oct–Nov 2014) DHIS (1) | Patients with VL test in 12 month period % (N) NHLS/DHIS (2) | VL<400 cp/mL % (N) NHLS (3) | Proportion known to be suppressed NHLS/DHIS (4) | VL 400–1000 cp/mL % (N) NHLS (5) | VL>1000 cp/mL % (N) NHLS (6) | VL>10000 cp/mL % (N) NHLS (7) |
|-----------------------|---|--|-----------------------------|---|----------------------------------|------------------------------|-------------------------------|
| KwaZulu-Natal | 906,783 | 72% (655,937) | 82% (540,553) | 59% | 3% (18,122) | 15% (97,262) | 10% (63,042) |
| Free State | 163,073 | 78% (127,318) | 82% (104,938) | 64% | 2% (2,382) | 16% (19,998) | 11% (13,579) |
| Gauteng | 660,849 | 73% (484,233) | 80% (385,673) | 59% | 4% (21,502) | 16% (77,058) | 10% (49,145) |
| Western Cape | 174,008 | 82% (142,156) | 80% (113,498) | 65% | 3% (3,835) | 17% (24,823) | 11% (16,191) |
| Entire country | 2,951,159 | 75% (2,199,890) | 78% (1,709,867) | 58% | 4% (80,873) | 19% (409,150) | 12% (272,836) |
| Eastern Cape | 307,288 | 80% (246,452) | 73% (180,293) | 59% | 3% (8,518) | 23% (57,642) | 16% (40,400) |
| Mpumalanga | 282,750 | 71% (200,903) | 73% (146,749) | 52% | 6% (11,550) | 21% (42,604) | 14% (27,445) |
| Limpopo | 214,066 | 76% (161,926) | 70% (113,193) | 53% | 5% (8,357) | 25% (40,376) | 17% (28,241) |
| North West | 200,833 | 74% (148,091) | 70% (103,415) | 52% | 4% (5,450) | 26% (39,226) | 18% (27,395) |
| Northern Cape | 41,511 | 79% (32,874) | 69% (22,695) | 55% | 3% (867) | 28% (9,312) | 20% (6,655) |

Notes: DHIS data comes from Excel Spreadsheet: DHIS_\$ZA_NIDS13_NDOH5_13_14_ART_Monthly.xlsm; NHLS data comes from May 2015 Corporate Data Warehouse extract

The proportion of patients with viral load values greater than 1,000 and 10,000 copies per mL were also highest in Northern Cape, Limpopo, and North West Provinces. The Western Cape (65%) and Free State (64%) had the highest proportion of patients *known* to be suppressed.

Table 3 compares suppression results from the most recent test done per patient, with the results looking at all viral load tests done in the study period. A total of 2,786,676 viral load tests were performed during the study period. 74% of those tests were virologically suppressed and 22% were greater than 1,000 cp/mL, compared to 78% suppressed and 19% greater than 1,000 cp/mL of only the most recent test per patient. At the provincial level, difference in viral suppression between all tests and the most recent tests ranged from 2%–4%—except for the Western Cape where there was a 7% difference. The total number of viral load tests was 94% of the TROA nationally. At the provincial level, all provinces except the Western Cape had fewer viral load tests than TROA. Clinics in the Western Cape averaged 1.1 tests per patient in the study period.

Viral Suppression at District and Sub-district Level

Table 4 shows viral load suppression in the 52 districts. Sixteen districts had viral load suppression 80% or greater and 16 districts had viral load suppression 70% or lower. We found a positive relationship between district TROA size and proportion suppressed. The 16 districts with the highest viral load suppression had a mean population on treatment of 87,970 patients, while the 16 districts with lowest viral load suppression had a mean population of only 27,138. The proportion of patients that had a viral load test in the previous 12 months ranged from a low of 54% to 99%. The proportion of patients with *known* viral load suppression ranged from 34% to 73%.

Figures 2a and 2b show the distribution of viral load suppression by district and sub-district. The lowest levels of viral load suppression were found in the Northern, Eastern and Western Cape sub-districts, and a few North West sub-districts. Figures 3a and 3b show the same distribution by absolute numbers. These figures demonstrate that the areas with lower levels of virologic suppression and lower population density have fewer numbers of suppressed patients. In 17 districts, between 25% and 47% of patients had viral loads >1,000 cp/mL. Figures 4a and 4b show the proportion of unsuppressed viral loads distributed by district and sub-district. Figure 5a and 5b show the number of patients with an unsuppressed viral load distributed by district and sub-district. The contrast between these two and the other figures is very telling: while viral load suppression is much greater in KwaZulu-Natal and Gauteng provinces, they have much larger populations of patients with unsuppressed viral loads. Figures 6a and 6b show the proportion of HIV patients with known suppression by district and sub-district. This shows a concentration of these patients in some metropolitan areas and in Free State, Eastern Cape, and KwaZulu Natal.

Table 3 Comparison of proportions of Viral Load (VL) tests done and VL suppression for most recent test and all tests by province, April 2014–March 2015

| Province | Total remaining on ART (Oct/Nov 2014) DHIS (1) | Most recent VL load test in 12 month period April 2014–March 2015 | | | All VL load tests in 12 month period April 2014–March 2015 | | |
|-----------------------|--|---|-----------------------------|------------------------------|--|-----------------------------|------------------------------|
| | | Number of patients with VL in past 12 months NHLS (2) | VL<400 cp/mL % (N) NHLS (3) | VL>1000 cp/mL % (N) NHLS (4) | Number of VL tests in past 12 months NHLS (5) | VL<400 cp/mL % (N) NHLS (6) | VL>1000 cp/mL % (N) NHLS (7) |
| KwaZulu-Natal | 906,783 | 655,937 | 82% (540,553) | 15% (97,262) | 866,033 | 79% (680,789) | 18% (157,777) |
| Free State | 163,073 | 127,318 | 82% (104,938) | 16% (19,998) | 151,279 | 80% (120,571) | 18% (27,594) |
| Gauteng | 660,849 | 484,233 | 80% (385,673) | 16% (77,058) | 614,951 | 76% (466,577) | 19% (118,240) |
| Western Cape | 174,008 | 142,156 | 80% (113,498) | 17% (24,823) | 191,066 | 73% (140,194) | 23% (44,169) |
| Entire country | 2,951,159 | 2,199,890 | 78% (1,709,867) | 19% (409,150) | 2,786,676 | 74% (2,069,416) | 22% (606,247) |
| Eastern Cape | 307,288 | 246,452 | 73% (180,293) | 23% (57,642) | 304,629 | 69% (210,817) | 27% (82,664) |
| Mpumalanga | 282,750 | 200,903 | 73% (146,749) | 21% (42,604) | 245,961 | 70% (172,988) | 24% (58,509) |
| Limpopo | 214,066 | 161,926 | 70% (113,193) | 25% (40,376) | 192,813 | 68% (130,370) | 27% (52,391) |
| North West | 200,833 | 148,091 | 70% (103,415) | 26% (39,226) | 179,263 | 67% (119,968) | 29% (52,578) |
| Northern Cape | 41,511 | 32,874 | 69% (22,695) | 28% (9,312) | 40,681 | 66% (26,904) | 31% (12,622) |

Notes: DHIS data comes from Excel Spreadsheet: DHIS_\$ZA_NIDS13_NDOH5_13_14_ART_Monthly.xlsm; NHLS data comes from May 2015 Corporate Data Warehouse extract

Table 4 Viral load (VL) suppression for people in care and on ART results by district, April 2014–March 2015

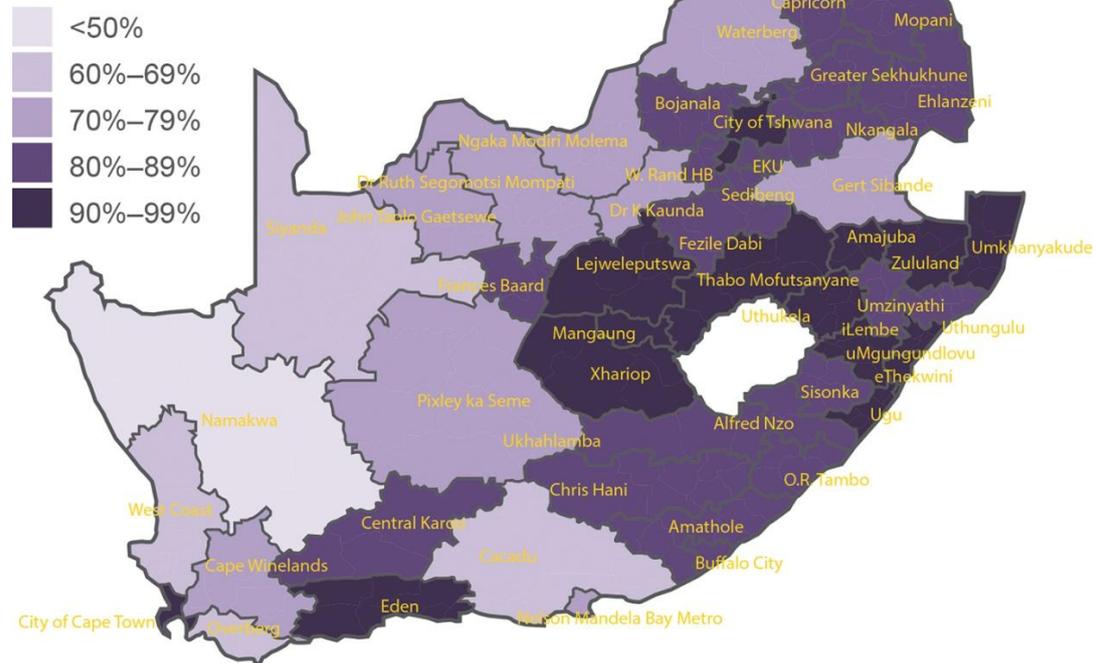
| Province | District | Total remaining on ART (TROA) (Oct–Nov 2014) DHIS (1) | Patients with VL in 12 month period % (N) NHLS/DHIS (2) | Proportion known to be suppressed NHLS/ DHIS (3) | VL<400 cp/mL % (N) NHLS (4) | VL400–1000 cp/mL % (N) NHLS (5) | VL>1000 cp/mL % (N) NHLS (6) | VL>10000 cp/mL % (N) NHLS (7) |
|----------|-------------------|---|---|--|-----------------------------|---------------------------------|------------------------------|-------------------------------|
| KZN | Amajuba DM | 38,397 | 62% (23,780) | 53% | 86% (20,337) | 1% (326) | 13% (3,117) | 9% (2,125) |
| KZN | eThekwini MM | 296,354 | 69% (204,524) | 59% | 86% (176,625) | 2% (3,600) | 12% (24,299) | 8% (15,857) |
| KZN | Ugu DM | 66,101 | 78% (51,472) | 65% | 84% (43,234) | 3% (1,306) | 13% (6,932) | 9% (4,462) |
| KZN | Umkhanyakude DM | 60,591 | 83% (50,265) | 70% | 84% (42,131) | 2% (884) | 14% (7,250) | 10% (4,884) |
| WC | Cape Town MM | 124,371 | 85% (106,244) | 71% | 83% (87,855) | 2% (2,543) | 15% (15,846) | 10% (10,280) |
| GP | Tshwane MM | 135,795 | 71% (95,880) | 59% | 83% (79,246) | 2% (1,928) | 15% (14,706) | 10% (9,793) |
| KZN | uMgungundlovu DM | 106,778 | 66% (70,745) | 55% | 83% (58,961) | 4% (2,685) | 13% (9,099) | 8% (5,857) |
| FS | T Mofutsanyane DM | 47,337 | 77% (36,302) | 64% | 83% (30,231) | 2% (645) | 15% (5,426) | 10% (3,657) |
| FS | Mangaung MM | 39,015 | 78% (30,442) | 65% | 83% (25,294) | 2% (562) | 15% (4,586) | 10% (3,109) |
| FS | Lejweleputswa DM | 41,063 | 73% (30,126) | 61% | 83% (25,098) | 2% (542) | 15% (4,486) | 10% (3,060) |
| KZN | Uthukela DM | 48,941 | 88% (43,226) | 72% | 82% (35,401) | 2% (857) | 16% (6,968) | 10% (4,425) |
| KZN | iLembe DM | 46,826 | 83% (38,939) | 68% | 82% (32,028) | 2% (968) | 15% (5,943) | 10% (3,778) |
| WC | Eden DM | 14,023 | 68% (9,564) | 55% | 81% (7,745) | 2% (222) | 17% (1,597) | 11% (1,038) |
| KZN | Zululand DM | 73,179 | 68% (49,884) | 55% | 81% (40,187) | 2% (855) | 18% (8,842) | 12% (6,012) |
| GP | Johannesburg MM | 244,286 | 73% (177,847) | 59% | 81% (143,527) | 4% (7,182) | 15% (27,138) | 10% (17,623) |
| FS | Xhariep DM | 9,251 | 91% (8,396) | 73% | 80% (6,732) | 2% (204) | 17% (1,460) | 11% (917) |
| FS | Fezile Dabi DM | 26,408 | 84% (22,052) | 67% | 80% (17,583) | 2% (429) | 18% (4,040) | 13% (2,836) |
| GP | Sedibeng DM | 50,060 | 85% (42,625) | 67% | 79% (33,493) | 6% (2,388) | 16% (6,743) | 10% (4,389) |
| EC | Buffalo City MM | 39,230 | 86% (33,819) | 68% | 79% (26,753) | 2% (814) | 18% (6,252) | 13% (4,484) |

| Province | District | Total remaining on ART (TROA) (Oct–Nov 2014) DHIS (1) | Patients with VL in 12 month period % (N) NHLS/DHIS (2) | Proportion known to be suppressed NHLS/ DHIS (3) | VL<400 cp/mL % (N) NHLS (4) | VL400–1000 cp/mL % (N) NHLS (5) | VL>1000 cp/mL % (N) NHLS (6) | VL>10000 cp/mL % (N) NHLS (7) |
|----------|----------------------|---|---|--|-----------------------------|---------------------------------|------------------------------|-------------------------------|
| NW | Bojanala Platinum DM | 86,753 | 63% (54,280) | 49% | 78% (42,333) | 2% (1,291) | 20% (10,656) | 13% (7,178) |
| GP | Ekurhuleni MM | 172,576 | 76% (130,649) | 59% | 78% (101,513) | 6% (7,675) | 16% (21,461) | 10% (13,043) |
| All | Entire Country | 2,951,159 | 75% (2,199,890) | 58% | 78% (1,709,867) | 4% (80,873) | 19% (409,150) | 12% (272,836) |
| EC | A Nzo DM | 39,997 | 67% (26,648) | 51% | 77% (20,601) | 3% (746) | 20% (5,301) | 15% (3,906) |
| EC | O Tambo DM | 74,052 | 65% (48,330) | 50% | 76% (36,668) | 2% (1,170) | 22% (10,491) | 16% (7,501) |
| KZN | Umzinyathi DM | 42,114 | 73% (30,943) | 56% | 76% (23,541) | 4% (1,192) | 20% (6,210) | 14% (4,219) |
| NC | Frances Baard DM | 17,810 | 89% (15,816) | 67% | 76% (12,006) | 2% (383) | 22% (3,427) | 15% (2,384) |
| MP | Ehlanzeni DM | 151,573 | 68% (102,338) | 51% | 75% (77,244) | 5% (4,629) | 20% (20,465) | 13% (13,696) |
| GP | West Rand DM | 58,133 | 64% (37,232) | 48% | 75% (28,049) | 6% (2,276) | 19% (6,907) | 11% (4,256) |
| KZN | Uthungulu DM | 88,622 | 70% (61,790) | 52% | 74% (45,861) | 5% (3,351) | 20% (12,578) | 13% (7,831) |
| EC | Amathole DM | 38,050 | 91% (34,487) | 67% | 74% (25,610) | 4% (1,387) | 22% (7,490) | 14% (4,948) |
| LP | Sekhukhune DM | 39,014 | 75% (29,112) | 55% | 74% (21,648) | 5% (1,376) | 21% (6,088) | 14% (4,153) |
| KZN | Harry Gwala DM | 38,882 | 78% (30,369) | 57% | 73% (22,248) | 7% (2,098) | 20% (6,023) | 12% (3,592) |
| MP | Nkangala DM | 60,392 | 68% (41,307) | 49% | 72% (29,871) | 5% (2,234) | 22% (9,202) | 14% (5,838) |
| EC | Joe Gqabi DM | 17,456 | 81% (14,052) | 58% | 72% (10,120) | 4% (550) | 24% (3,382) | 17% (2,432) |
| LP | Capricorn DM | 41,347 | 86% (35,490) | 61% | 71% (25,276) | 4% (1,397) | 25% (8,817) | 18% (6,302) |
| EC | C Hani DM | 36,978 | 95% (35,232) | 68% | 71% (25,106) | 4% (1,393) | 25% (8,733) | 17% (6,025) |
| WC | Central Karoo DM | 2,175 | 66% (1,437) | 47% | 71% (1,020) | 3% (43) | 26% (374) | 18% (259) |

| Province | District | Total remaining on ART (TROA) (Oct–Nov 2014) DHIS (1) | Patients with VL in 12 month period % (N) NHLS/DHIS (2) | Proportion known to be suppressed NHLS/ DHIS (3) | VL<400 cp/mL % (N) NHLS (4) | VL400–1000 cp/mL % (N) NHLS (5) | VL>1000 cp/mL % (N) NHLS (6) | VL>10000 cp/mL % (N) NHLS (7) |
|----------|---------------------------|---|---|--|-----------------------------|---------------------------------|------------------------------|-------------------------------|
| LP | Mopani DM | 58,722 | 68% (39,906) | 48% | 70% (28,109) | 6% (2,336) | 24% (9,461) | 16% (6,472) |
| EC | N Mandela Bay MM | 43,667 | 88% (38,395) | 62% | 70% (26,692) | 3% (1,107) | 28% (10,596) | 20% (7,709) |
| MP | G Sibande DM | 70,786 | 81% (57,258) | 56% | 69% (39,634) | 8% (4,687) | 23% (12,937) | 14% (7,911) |
| NC | J T Gaetsewe DM | 10,530 | 54% (5,690) | 37% | 69% (3,953) | 2% (135) | 28% (1,602) | 20% (1,138) |
| LP | Vhembe DM | 43,677 | 72% (31,585) | 49% | 68% (21,583) | 6% (1,943) | 26% (8,059) | 17% (5,500) |
| NW | Ruth Segomotsi Mompoti DM | 28,334 | 75% (21,130) | 51% | 68% (14,353) | 3% (576) | 29% (6,201) | 21% (4,468) |
| NW | Dr K Kaunda DM | 43,054 | 92% (39,405) | 61% | 67% (26,277) | 3% (1,204) | 30% (11,924) | 23% (8,921) |
| LP | Waterberg DM | 31,307 | 83% (25,833) | 53% | 64% (16,577) | 5% (1,305) | 31% (7,951) | 23% (5,814) |
| NC | Pixley ka Seme DM | 5,787 | 89% (5,179) | 55% | 62% (3,214) | 3% (155) | 35% (1,810) | 25% (1,303) |
| WC | Cape Winelands DM | 18,426 | 73% (13,393) | 44% | 61% (8,170) | 6% (750) | 33% (4,474) | 21% (2,826) |
| NW | Ngaka Modiri Molema DM | 42,692 | 78% (33,276) | 48% | 61% (20,428) | 7% (2,374) | 31% (10,474) | 21% (6,859) |
| NC | ZF Mgcau DM | 5,834 | 85% (4,966) | 49% | 57% (2,852) | 3% (140) | 40% (1,974) | 29% (1,463) |
| EC | Sarah Baartman DM | 17,859 | 87% (15,489) | 49% | 56% (8,744) | 9% (1,350) | 35% (5,395) | 22% (3,394) |
| WC | West Coast DM | 6,227 | 99% (6,179) | 56% | 56% (3,460) | 5% (328) | 39% (2,391) | 26% (1,624) |
| WC | Overberg DM | 8,788 | 61% (5,339) | 34% | 56% (3,005) | 4% (239) | 39% (2,095) | 27% (1,453) |
| NC | Namakwa DM | 1,550 | 79% (1,223) | 37% | 47% (576) | 5% (66) | 47% (581) | 35% (431) |

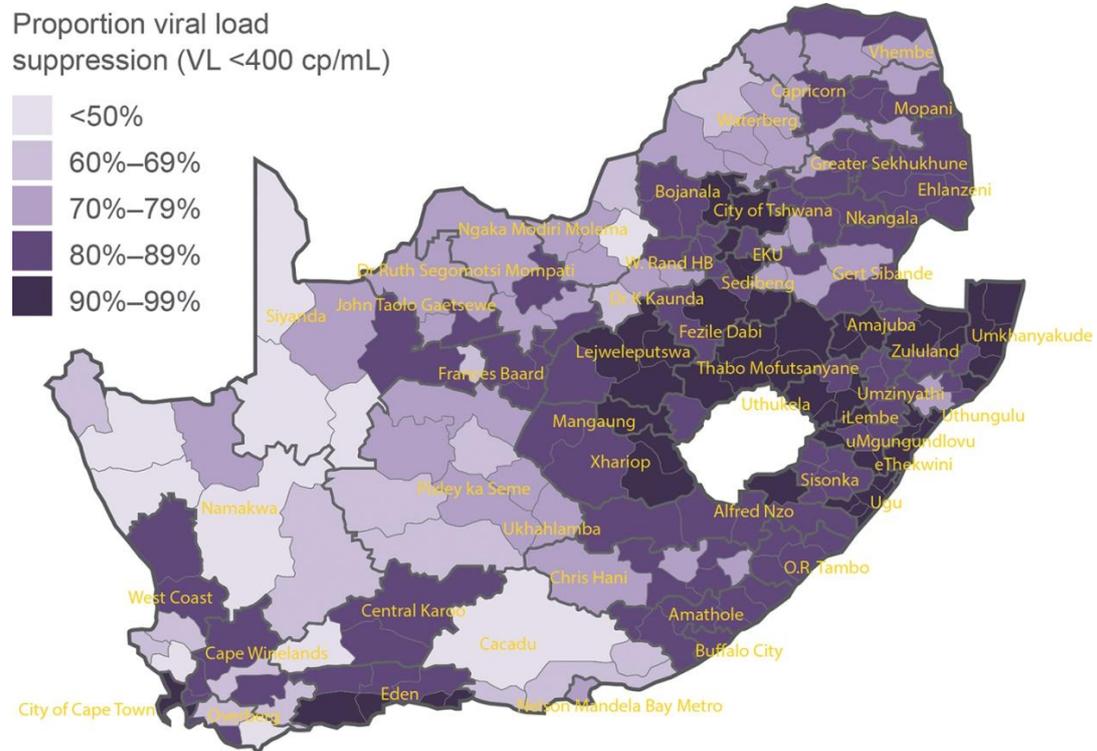
Notes: DHIS data come from Excel Spreadsheet: DHIS_\$ZA_NIDS13_NDOH5_13_14_ART_Monthly.xlsx; NHLS data come from May 2015 Corporate Data Warehouse extract; EC=Eastern Cape, FS=Free State, GP=Gauteng, KZN=KwaZulu-Natal, LP=Limpopo, MP=Mpumalanga, NC=Northern Cape, NW=North West, WC=Western Cape

Figure 2-A Proportion viral load suppression (VL <400 cp/mL) by district, April 2014–March 2015, South Africa
Proportion viral load suppression (VL <400 cp/mL)



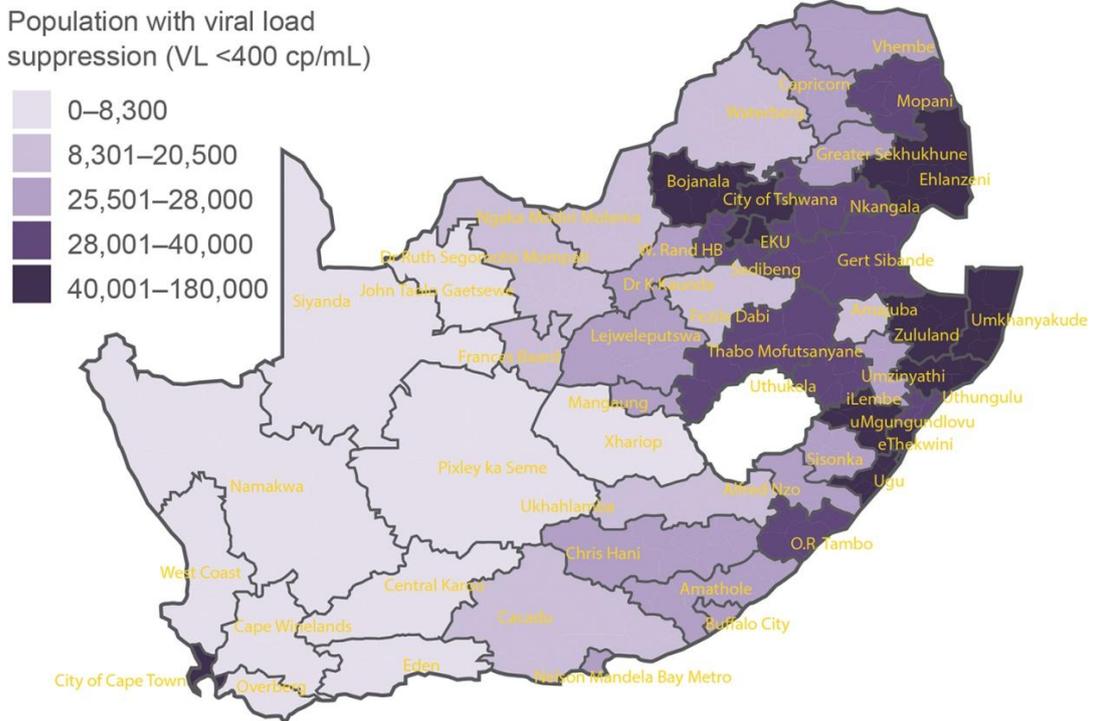
Source: NHLS Corporate Data Warehouse, May 2015 data extract.

Figure 2-B Proportion viral load suppression (VL <400 cp/mL) by health sub-district, April 2014–March 2015, South Africa
Proportion viral load suppression (VL <400 cp/mL)



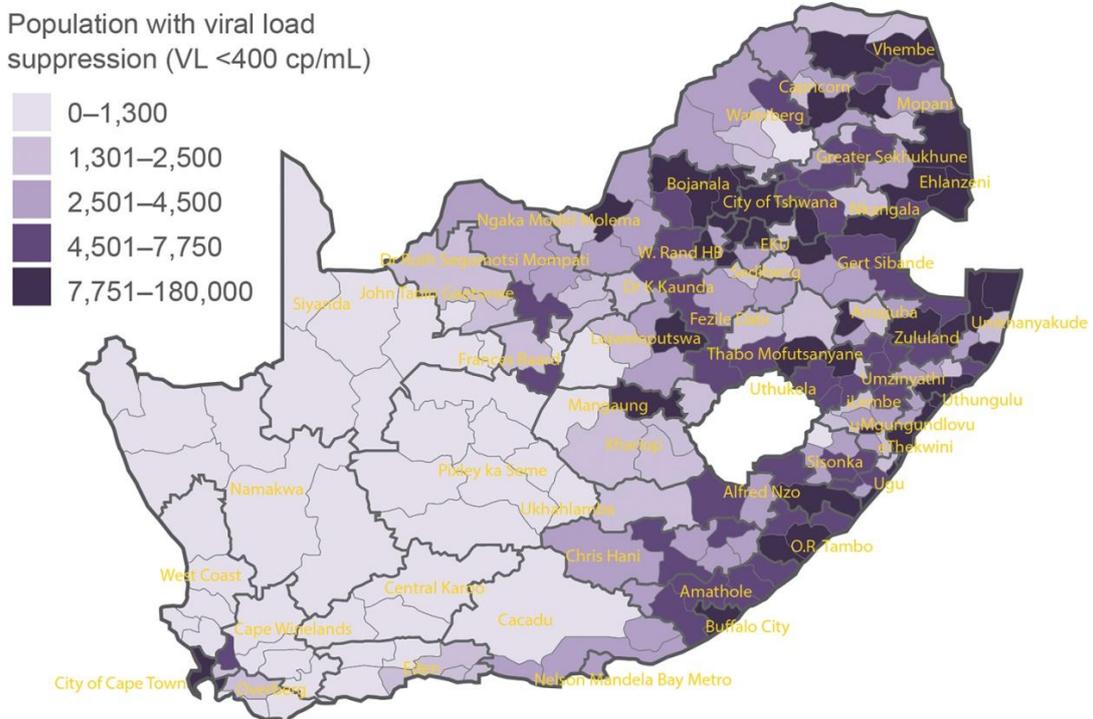
Source: NHLS Corporate Data Warehouse, May 2015 data extract.

Figure 3-A Population with viral load suppression (VL <400 cp/mL) by district, April 2014–March 2015, South Africa



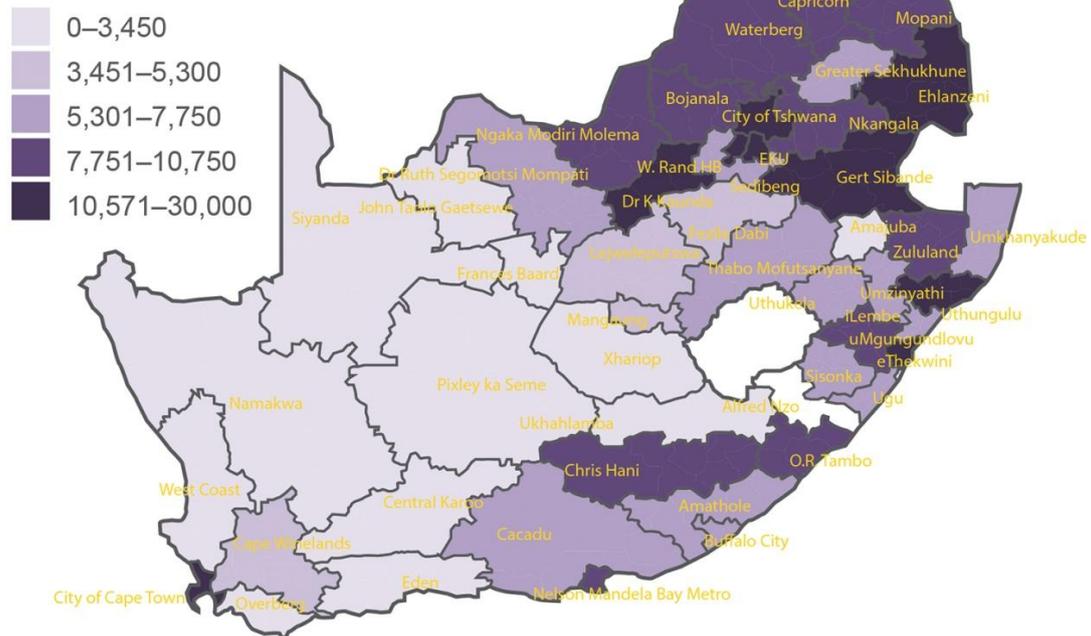
Source: NHLS Corporate Data Warehouse, May 2015 data extract.

Figure 3-B Population with viral load suppression (VL <400 cp/mL) by health sub-district, April 2014–March 2015, South Africa



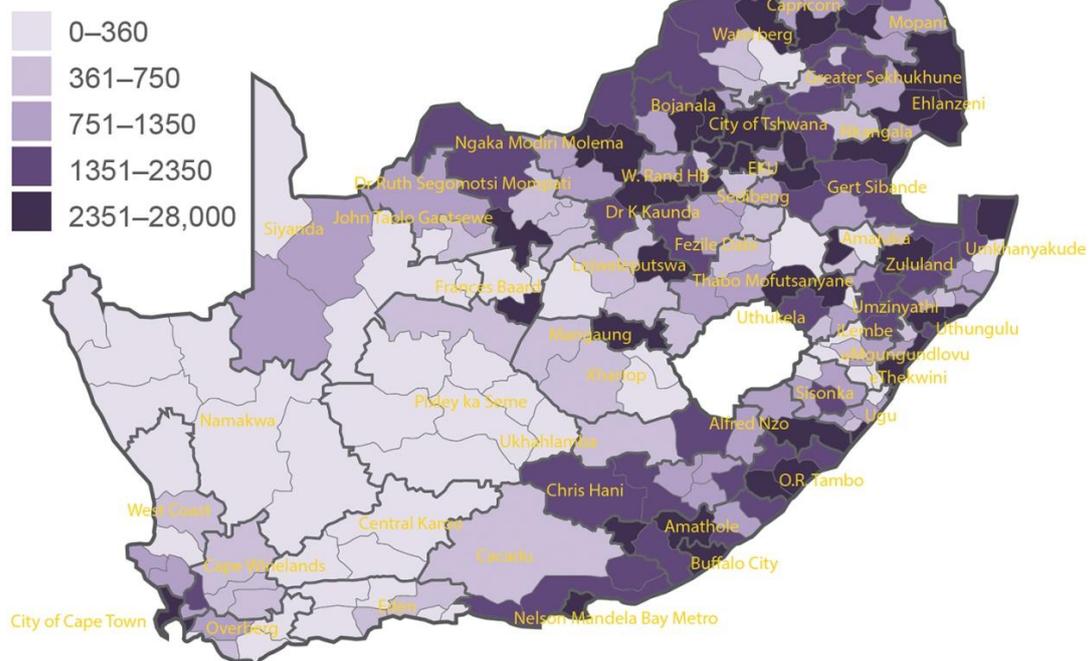
Source: NHLS Corporate Data Warehouse, May 2015 data extract.

Figure 5-A Population with high viral load (VL >1,000 cp/mL) by district, April 2014-March 2015, South Africa
 Population with high viral load (VL >1,000 cp/mL)



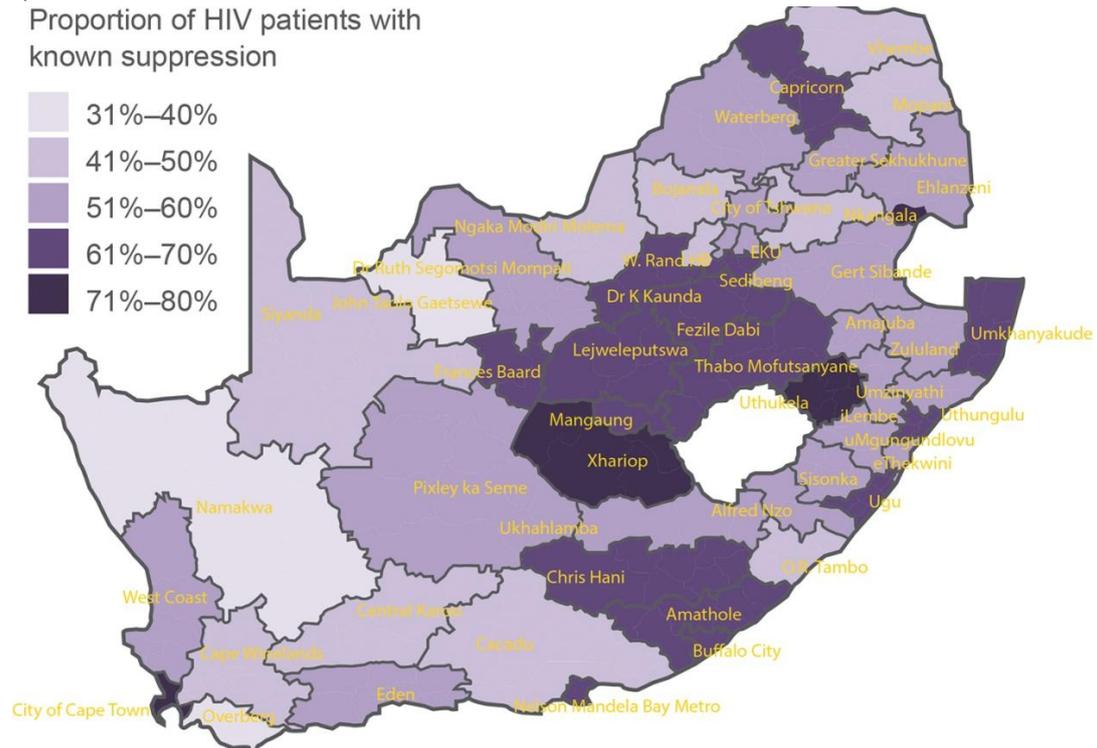
Source: NHLS Corporate Data Warehouse, May 2015 data extract.

Figure 5-B Population with high viral load (VL >1,000 cp/mL) by health sub-district, April 2014–March 2015, South Africa
 Population with high viral load (VL >1,000 cp/mL)



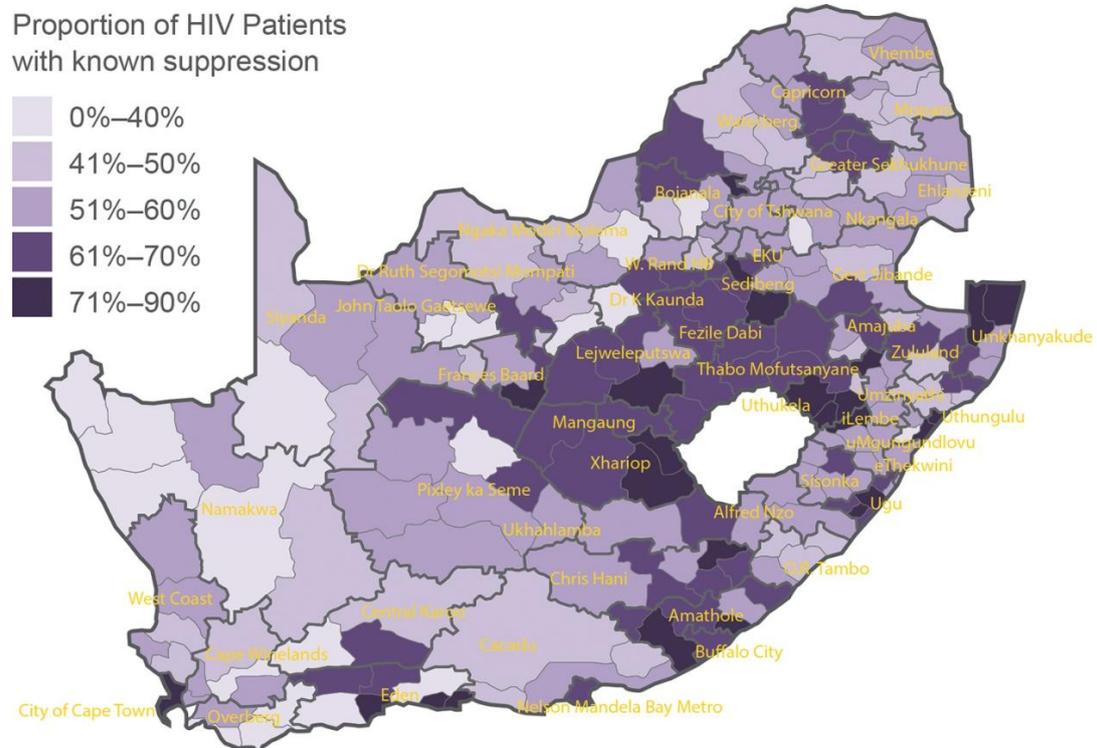
Source: NHLS Corporate Data Warehouse, May 2015 data extract.

Figure 6-A Proportion of HIV patients with known suppression (had VL test <400 cp/mL) by district, April 2014–March 2015, South Africa



Source: NHLS Corporate Data Warehouse, May 2015 data extract.

Figure 6-B Proportion of HIV patients with known suppression (had VL test <400 cp/mL) by health sub-district, April 2014–March 2015, South Africa



Source: NHLS Corporate Data Warehouse, May 2015 data extract.

Viral Load Suppression at Health Facility Level

Table 5 shows viral load suppression for 3,674 facilities that report a TROA of 12 or greater in the DHIS. Viral load suppression varied widely by facility: 132 facilities had 90% or better viral suppression, 218 had less than 50% viral suppression.

Table 5 Viral load suppression for people in care and on ART results by facility, April 2014–March 2015

This table contains data for 3,674 public sector health facilities the following variables included: Province, district, sub-district, facility name, TROA, number of patients with VL in past 12 months, percent and number per VL category (<400, 400–1000, >1000, >10000 cp/mL), and VL category. The table can be viewed in a separate window by clicking on this link.

The VL suppression results for individual health facilities informed the design of the NDOH/World Bank impact evaluation of the adherence guidelines. This evaluation compares intervention and control facilities, with the intervention (a standardised package of adherence strategies for patients on chronic treatments) randomly allocated to 12 health facilities. Facilities with similar VL suppression levels (and in the same TROA band) were matched to form pairs. The pairs and their VLS and TROA are shown in the Executive Summary Table 2.

Table 6 shows viral load suppression among facilities broken down into quartiles by number of patients on treatment. Facilities with more patients on treatment had higher proportions of patients virologically suppressed. The 25% of facilities with at least 922 patients on treatment had a mean virologic suppression rate almost 15% higher than the 25% of facilities with 175 or fewer patients on treatment. The observed differences in viral suppression were all statistically significant and this was a linear relationship. Facility type was not a confounder of the relationship between virologic suppression and facility size: over 83% of all the facilities were classified as clinics (rather than hospitals or other type of facility). When facility type was included in the mixed model regression, it had minimal effect on the predicted coefficients.

Table 6 Mean proportion viral load suppression by facility size

| Number of patients in facility DHIS (1) | Number of facilities DHIS (2) | Total number of patients on treatment DHIS (3) | Mean proportion viral load suppression NHLS (4) |
|---|-------------------------------------|--|---|
| More than 922 | 919 | 2,020,925 | 79.0% |
| 415–922 | 918 | 579,486 | 75.2% |
| 176–414 | 918 | 262,049 | 71.6% |
| 13–175 | 919 | 88,699 | 64.5% |
| All Facilities | 3,674 | 2,951,159 | 71.3% |

Sources: DHIS data comes from Excel Spreadsheet: DHIS_\$ZA_NIDS13_NDOH5_13_14_ART_Monthly.xlsm; NHLS data comes from May 2015 Corporate Data Warehouse extract.

Table 7 shows Moran’s I and Geary’s c statistics for viral load suppression for all the facilities. Moran’s I has possible values from -1 (perfect dispersion) to 1 (perfect correlation) and a value of 0 indicates a random spatial pattern. The value of Moran’s I in this data set is 0.246; 95% CI (0.241, 0.251). The range of Geary’s c is from 0 (perfect correlation) to 2 (perfect dispersion). The value of Geary’s c in this data set was 0.54; 95% CI (0.48, 0.60). These values indicate that neighbouring facilities are likely to have similar viral load suppression rates.

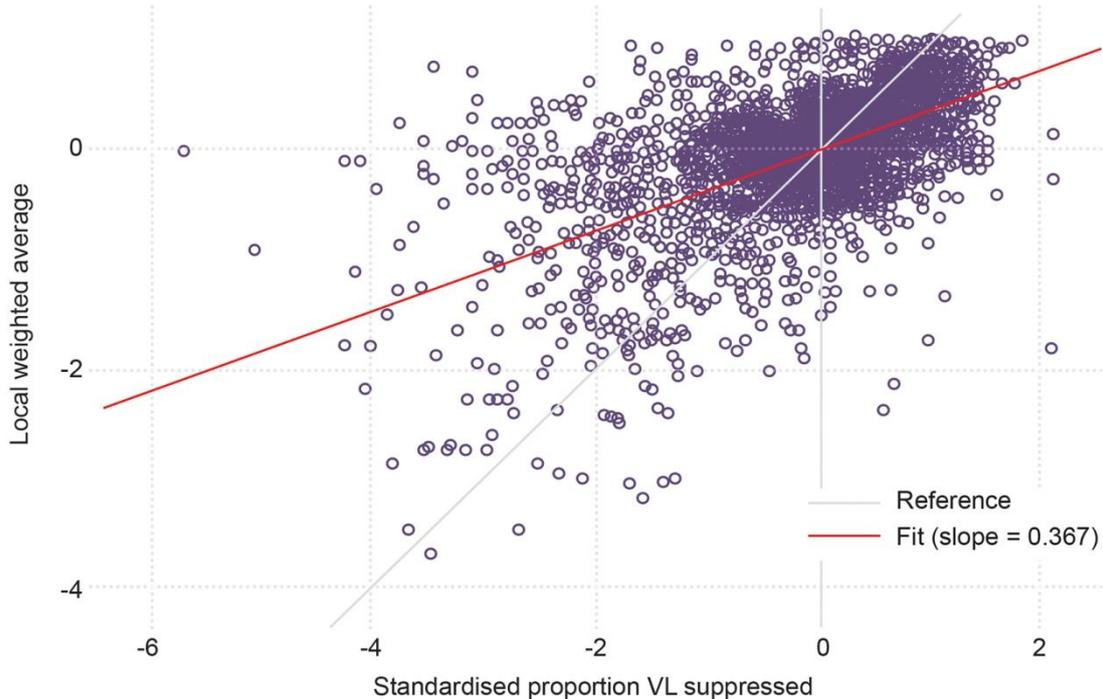
Table 7 Global autocorrelation statistics for viral load suppression by facility

| Coefficient | Statistic (95% CI) |
|-------------|----------------------|
| Moran’s I | 0.246 (0.241, 0.251) |
| Geary’s c | 0.54 (0.48, 0.60) |

Source: NHLS data from May 2015 Corporate Data Warehouse extract.

Figure 8 is a scatterplot of the standardised proportion of viral load suppression versus the weighted average of neighbouring facilities (Local Weighted Average), which shows the same correlation graphically. The reference line has a slope of 1 and indicates perfect correlation. The fitted line has a slope of 0.367, showing a positive correlation between the proportion of suppressed viral load at a facility and its neighbouring facilities.

Figure 8 Scatterplot of standardised viral load suppression versus local weighted average with fitted line showing correlation among health facilities, April 2014–March 2015, South Africa



Viral Suppression by Age and Gender

Tables 8 and 9 show proportion suppressed by age group and gender. The proportion of viral load suppression increased with age from a low of 51% in the youngest age group (children under 5 years) to 83% in the population 50 years and older. One in three ART patients under 25 years of age was not virally suppressed. The differences between age groups is statistically significant at the $p < 0.0001$ level.

Females had a higher proportion of viral suppression than men and this difference was significant at the $p < 0.0001$ level. One in six male and one in nine female ART patients had a viral load $> 10,000$ copies/mL. ■

Table 8 Viral load (VL) suppression for people in care and on ART results by age group, April 2014–March 2015

| Age Group | Number of Patients with VL in past 12 months | VL < 400 cp/mL % (N) | VL 400–1000 cp/mL | | VL > 10,000 cp/mL |
|-----------|--|----------------------|-------------------|---------------------|-------------------|
| | | | % (N) | VL 1000 cp/mL % (N) | % (N) |
| 0–4 | 33,338 | 51% (16,848) | 5% (1,537) | 45% (14,953) | 36% (11,854) |
| 5–14 | 108,573 | 67% (72,690) | 5% (5,116) | 28% (30,768) | 18% (19,895) |
| 15–24 | 144,605 | 65% (93,974) | 4% (5,799) | 31% (44,832) | 21% (30,167) |
| 25–49 | 1,642,191 | 79% (1,290,862) | 4% (58,839) | 18% (292,490) | 12% (194,870) |
| 50+ | 370,602 | 83% (306,888) | 4% (13,544) | 14% (50,170) | 9% (32,772) |

Source: NHLS data from May 2015 Corporate Data Warehouse extract.

Table 9 Viral load (VL) suppression for people in care and on ART results by gender, April 2014–March 2015

| Gender | Number of Patients with VL in past 12 months | VL < 400 cp/mL % (N) | VL 400–1000 cp/mL | | VL > 10000 cp/mL |
|---------|--|----------------------|-------------------|-----------------------|------------------|
| | | | % (N) | VL > 1000 cp/mL % (N) | % (N) |
| Female | 1,546,590 | 79% (1,221,812) | 4% (56,005) | 17% (268,773) | 11% (171,883) |
| Male | 713,956 | 74% (528,359) | 4% (27,694) | 22% (157,903) | 16% (113,238) |
| Unknown | 38,763 | 79% (30,813) | 3% (1,163) | 18% (6,788) | 12% (4,613) |

Source: NHLS data from May 2015 Corporate Data Warehouse extract.

04 DISCUSSION 011010101010110

This paper presents a novel analysis of data from the DHIS and the NHLS CDW to provide a snapshot of ARV treatment outcomes in South Africa. The novel elements are:

- 1 The probabilistic patient level linking which allows the creation of a patient-level cohort from individual laboratory results; and
- 2 The clinic linking which allows the combination of clinic-level reporting through the DHIS and viral load laboratory results from the patient cohort to be compared at the facility, health sub-district, district, provincial and national levels

The analysis shows that nationally, 75% of patients on ART received a viral load test in the 12-month period, 78% of those tested were virologically suppressed, and 58% of the ART patient population is *known* to be suppressed. This is in contrast to the DHIS data from the adult cohort which estimated that 46% of patients on ART had a viral load test in the past 12 months, 83% were virologically suppressed (Overmeyer, 2015) and therefore only 38% of the ART patients were *known* to be suppressed.

The robustness of the findings is dependent in part upon the accuracy of the patient level linking algorithm. We have estimated that this linking algorithm has an under-linking rate of 11% and an over-linking rate of 15%. Under-linking results in tests not being joined to patients, which results in patients contributing multiple viral load tests (i.e., patients being counted more than once), while over-linking results in tests being incorrectly linked to a different patient resulting in some patients' viral load tests not being counted, and a misleading picture of the viral load test frequency and results for the patient to whom the results are incorrectly linked. Under-linking and over-linking has an effect on the overall estimate of the number of patients who had a viral load test, and a smaller effect on the pattern of viral load suppression (the level of distortion depends on the differences in viral load among those whose records are incorrectly linked).

The effects of over-linking and under-linking are dependent upon the time period being measured. A short time period reduces the chances that a patient will have multiple viral load tests which reduces the chance of over or under-linking. Thus, the analysis was limited to a 12-month period, chosen because it is expected that all patients would have at least one viral load test in this time period and that a low proportion would have had more than one.

Results from all viral load tests were compared to results limited to the patient's last test in the study period. At the national level and most provinces, the difference in viral load suppression was 2–4% lower when looking at all test results. Viral load suppression rates from all test results provides a lower boundary for viral load suppression as it includes patients who have had more than one test in a 12-month period; the most likely contributor to this difference would be from patients with unsuppressed viral loads in whom a repeat viral load test was requested following adherence counselling.

Regarding the number of patients who were tested, there is a strong consistency between the TROA and the number of patients who have been tested at least once in the 12-month period. At the national, provincial, and district levels, TROA always exceeds the number of patients tested, which is logically consistent and expected. This same finding doesn't always hold true at the facility level, but we believe that this is due to some minor issues with facility matching between the DHIS and NHLS datasets.

The analysis of the DHIS and NHLS CDW datasets also required facility level linking. We were able to link 95% of the 3,674 clinics reporting a TROA in the DHIS with the NHLS laboratory viral load test results. The matched facilities account for 96% of the TROA in the DHIS. While the matching was not entirely complete, the results of an unmatched analysis of NHLS data at provincial level show negligible differences in viral load suppression. We excluded 102 facilities with a TROA of 12 or less from the analysis. These 102 facilities accounted for a total of 523 TROA (0.02% of the total TROA), so we do not believe that their inclusion would make any difference in our results.

This analysis assumes that clinics are following viral load monitoring guidelines which call for the routine use of viral loads at yearly intervals except for the first test at six months after commencing ART. It is possible that some facilities test people who are yet to start ART or who are known to be off ART, or that frequent repeat VL tests are being done in those identified as having a VL greater than 1,000 cp/mL. While this might explain some of the diversity we see in facility-level virologic suppression, it is unlikely to explain the differences in virologic suppression by facility TROA size.

This analysis demonstrates the diversity in the proportion of patients virally suppressed by geographic location, facility and demography in South Africa. Nationally, viral load suppression was 78%, but some provinces, districts and facilities were able to achieve higher suppression rates. Among provinces, the two that performed best had suppression rates 11–12% higher than the three worst-performing provinces. At the district level, the best-performing districts had suppression rates more than 40% higher than the worst-performing districts. The gap between the best and poorest performing facilities was even greater. We found that the best-performing districts had larger patient populations than poorer-performing districts.

We found a positive linear correlation between facility TROA size and viral load suppression. The 25% of facilities with the largest TROA had viral load suppression rates 3.8% higher than the second quartile, 7.4% higher than the third quartile, and almost 15% higher than the 25% of facilities with the smallest TROA. Implementing HIV treatment for large numbers of patients can be challenging, especially in regards to patient waiting times. However, this analysis suggests that facilities with high patient loads have been able to provide high-quality care as evidenced by the high rates of viral load suppression found in their patient populations. When examining possible explanations for this, we ruled out the type of site as the vast majority of sites were clinics and type of site had no influence on the results. Many other factors could contribute to the differences in viral suppression across sites: staffing level, staffing type, time since ART provision started at the site, and rural/urban location, however this data could not be taken into consideration within the scope of this analysis. Moreover, we cannot rule out data issues—results at facilities with small patient populations are more sensitive to a small number of errors, whereas for big facilities, errors get absorbed in the size of the population.

We demonstrated correlations in the proportions of viral load suppression between neighbouring facilities using two different methods of correlation. This can be interpreted as evidence that neighbouring facilities have more similar levels of virologic suppression than facilities that are further away. Possible explanations could include shared patient populations with similar socio-economic profiles; or shared provincial and district governance with similar policies, programmatic support and health systems factors. However, due to the limited information in our data set we cannot test any of these explanations.

Viral load suppression also was associated with age, with older age groups having higher levels of viral suppression. The two oldest age groups contributed 88% of the viral load tests; 79% of the tests given to 25–49 year olds and 83% of tests given to patients 50 years and older were suppressed. Around two-thirds of the tests given to the 5–14 year olds and 15–24 year olds were suppressed, while only 51% of the tests given to the 0–4 year olds were suppressed. The high rate of unsuppressed viral loads in the 0–4 year old age group may be attributed, in part, to the lack of understanding of treatment and monitoring procedures in this population. Nevertheless, these results indicate that high viral load while on treatment is a critical problem affecting adolescents and particularly children under 5 enrolled in the ART programme. Nationally, patients 25-49 years of age make up 73% of persons on ART and the 50 and older age group another 13%, which is similar to the age distribution of test results in the NHLS dataset (Shisana, O et al., 2014).

Viral testing by gender showed that females received approximately two-thirds of tests and approximately 5% more females had viral suppression. Nationally, females are estimated to make up 67% of all HIV positive persons on treatment (Shisana, O. et al., 2014), so their testing rate is about what would be expected.

The comparison of the NHLS to the DHIS results showed close results for the proportion of viral load suppressed but large differences in the proportion of patients receiving a viral load test. The national difference in proportion viral load suppressed was 2% (78 vs. 83%) while the difference in viral load tests done was 29% (75 vs. 46%). There are a number of possible explanations for the large differences between DHIS and our findings. Firstly, the TIER.Net data collection is incomplete: the results come from a national cohort of over 400,000 patients at a time when the same system was reporting over 2 million patients on ART. The facilities adopting TIER.Net early are likely to be better managed facilities and might have higher VLS than later TIER.Net adopters. Secondly, the data are captured two different ways. The NHLS data are captured from the laboratories using automated methods and housed in a centralised data warehouse, while the TIER.Net data are captured at the clinic level. For the test to be captured at the clinic level, test results must be received by the clinic from the laboratory; results must be printed and filed in the patient's folder; and the results entered into TIER.Net. An evaluation of the clinic laboratory interface noted that process compliance of a clinic receiving lab results was such that only 35% of lab results were filed in patients' folders (Young & Galloway, 2010). Such low compliance could explain a large proportion of the observed differences.

While the range of viral load suppression at the provincial level is narrow, it is much greater at the district and facility level. If all facilities followed the national viral load monitoring guidelines, then the results could be accepted at face value and lower viral load suppression rates could be interpreted as patients not doing as well in one facility compared to another. However, it is possible that some facilities might have adopted alternative strategies for viral load testing such as selectively testing ART patients who they think are not doing well on treatment. Evidence for this also exists in the variability of proportions of patients estimated to have been tested in the past 12 months across districts. Some districts tested less than 70% of their patients in the past 12 months, while others tested more than 90%. If facilities are using different monitoring strategies then a straight comparison is not possible. It would be important to understand why facilities are not fully monitoring all of their patients.

In addition to providing patient-level results of viral suppression in South Africa, this analysis also debuts the patient-level laboratory cohort, demonstrating its ability to provide detailed information on viral load suppression in South Africa. This information can be used to identify facilities that have low viral load suppression, a low proportion of

patients tested in the previous 12 months, and a low proportion of patients known to be virologically suppressed, so that problems can be identified and corrected. It can also help us learn from the facilities that achieve near-universal viral suppression among their ART patients. This report has demonstrated the utility of using NHLS test results in combination with DHIS information. If this combination of data are routinely used to track progress on ART treatment effectiveness in the future, it could inform the allocation of resources for programme improvements.

The South African government has set a target of 90% viral load suppression among HIV patients on treatment. This analysis shows that while some facilities are achieving this goal, it is not being achieved by any entire district or province. By any measure, the health system needs to improve the monitoring of patients' viral loads, the processing of information and communication of viral suppression results at clinic level. More complete viral load monitoring is crucial to enable satisfactory monitoring of the health of patients who are not having regular viral load tests. The adoption and routine use of the National Health ID number will make the combination of DHIS and NHLS laboratory data a powerful tool for better clinical care and programmatic monitoring. In the short term, the use of a combination of both NHLS laboratory and DHIS data is recommended for programmatic and facility monitoring of the ARV programme.

This work shows that Big Data—in this case a secondary analysis of large sets of routine data and innovative methodology of record linkage—can inform programme improvements. Specifically, the present analysis highlights health facilities and districts which need enhanced adherence support, and it also identifies success stories where viral suppression is achieved and ART patients regularly monitored to ascertain treatment effectiveness. ■

References

- Christen, P. (2006). A Comparison of Personal Name Matching: Techniques and Practical Issues. Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), (September), 290–294. <http://doi.org/10.1109/ICDMW.2006.2>
- Ford, N., Stinson, K., Davies, M.-A., et al. (2014). Is it safe to drop CD4+ monitoring among virologically suppressed patients: a cohort evaluation from Khayelitsha, South Africa. *Aids*, 28(July), Epub. <http://doi.org/10.1097/QAD.0000000000000406>
- Jacquez, G. M. (2008). Spatial Cluster Analysis. In S. Fotheringham & J. Wilson (Eds.), *The Handbook of Geographic Information Science* (pp. 395–416). Blackwell.
- Kleinbaum, D., Kupper, L., Muller, K., & Nizam, A. (1997). *Applied Regression Analysis and Multivariable Methods* (3rd ed.). Duxbury Press.
- Meyer-Rath, G., Chiu, C., Johnson, L., et al. (2015). South Africa's Investment Case – What are the country's "best buys" for HIV and TB? In Proceedings from the South African AIDS Conference. Durban.
- Miller, W. C., Powers, K. A., Smith, M. K., & Cohen, M. S. Community viral load as a measure for assessment of HIV treatment as prevention. *The Lancet Infectious Diseases*, 13(5), 459–64. [http://doi.org/10.1016/S1473-3099\(12\)70314-6](http://doi.org/10.1016/S1473-3099(12)70314-6). 2013.
- National Department of Health. (2011). Tier T1 AND T2 ART Monitoring and Evaluation Implementation Plan.
- National Department of Health. (2013). *Antiretroviral Health Indicators Update-- Directorate: Monitoring and Evaluation Issue III*. Pretoria.
- National Department of Health. (2015). National Consolidated Guidelines for the Prevention of Mother-to-Child Transmission of HIV and the Management of HIV in Children, Adolescents and Adults. Pretoria, South Africa.
- Overmeyer R (2015). Alignment between MTSF core indicators and DIP tracer indicators. NDOH/DIP SI TWG. 22 October 2015.
- Shisana, O, Rehle, T., LC, S., Zuma, K., et al. *South African National HIV Prevalence, Incidence and Behaviour Survey, 2012*. Cape Town: HSRC Press. 2014.
- UNAIDS.. 90-90-90: An ambitious treatment target to help end the AIDS epidemic. Geneva. 2014.
- WHO. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a public health approach June 2013. *Who 2013 Consolidated Guideline*. Geneva.
- WHO. Technical and Operational Considerations for Implementing HIV Viral Load Testing. 2014.
- WHO. Global update on the health sector response, 2014. October 2014, ISBN 978 92 4 150758 5.
- Winkler, W.. Overview of record linkage and current research directions. *Technical Report RR2006/02*. Washington, DC. 2006.
- Young, T., & Galloway, M.. Integrated Systems Analysis of Clinic-Laboratory Interface: Understanding the nature and limitation of the pre-analytical and post-analytical phases of specialist laboratory tests at South African primary health care clinics. 2010.