American Journal of Epidemiology
© The Author(s) 2020. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
Vol. 189, No. 12
DOI: 10.1093/aje/kwaa140
Advance Access publication:
July 10, 2020

Original Contribution

Impact of Viral Load Monitoring on Retention and Viral Suppression: A Regression Discontinuity Analysis of South Africa's National Laboratory Cohort

Alyssa F. Harlow*, Jacob Bor, Alana T. Brennan, Mhairi Maskew, William MacLeod, Sergio Carmona, Koleka Mlisana, and Matthew P. Fox

* Correspondence to Alyssa F. Harlow, Department of Epidemiology, Boston University School of Public Health, Boston, MA 02118 (e-mail: afharlow@bu.edu).

Initially submitted November 18, 2019; accepted for publication July 7, 2020.

South African guidelines recommend repeat viral load testing within 6 months when human immunodeficiency virus (HIV) viral loads exceed 1,000 copies/mL. We assessed whether South African facilities follow viral load monitoring guidelines and whether guidelines improve HIV-related outcomes, using a regression discontinuity design in a national HIV cohort of 174,574 patients (2013–2015). We assessed whether patients with viral loads just above versus just below 1,000 copies/mL were more likely to receive repeat testing in 6 months, and we compared differences in clinic transfers, retention, and viral suppression. The majority (67%) of patients with viral loads of >1,000 copies/mL did not receive repeat testing within 6 months, and these patients were 8.0% (95% confidence interval (CI): 6.2, 9.7) more likely to receive repeat testing compared with ≤1,000 copies/mL. Eligibility for repeat testing (>1,000 copies/mL) was associated with greater 12-month retention (risk difference = 2.9%, 95% CI: 0.6, 5.2) and combined suppression and retention (risk difference = 5.8%, 95% CI: 3.0, 8.6). Patients with viral loads of >1,000 copies/mL who actually received repeat testing were 85.2% more likely to be both retained and virally suppressed at 12 months (95% CI: 35.9, 100.0). Viral load monitoring might improve patient outcomes, but most patients with elevated viral loads do not receive monitoring within recommended timelines.

HIV; regression discontinuity design; viral load monitoring

Abbreviations: ART, antiretroviral therapy; CACE, complier average causal effect; CI, confidence interval; HIV, human immunodeficiency virus; ITT, intention-to-treat; RD, risk difference.

World Health Organization human immunodeficiency virus (HIV) treatment guidelines call for routine viral load testing as a key component of antiretroviral therapy (ART) monitoring (1). Daily ART suppresses HIV viral load in most patients. Persistent viremia typically signals gaps in adherence or resistance to first-line medications. Guidelines call for viral load testing at 6 and 12 months after ART initiation, and annually thereafter. If a viral load is >1,000 copies/mL, the World Health Organization recommends addressing adherence concerns and repeating viral load testing after 3–6 months (1). If the repeat value is >1,000 copies/mL, switching to second-line ART is recommended. South Africa has adopted the World Health Organization viral load monitoring guidelines as national policy (2).

Treatment guidelines indicate action should be taken for patients with viral loads of >1000 copies/mL, while patients with viral loads of ≤1000 copies/mL do not require immediate intervention. Coupled with adherence counseling, repeat viral load testing at 3–6 months reveals whether a patient has resuppressed or has a resistant strain requiring second-line ART. Although viral load is a continuous, numeric biomarker, the threshold rule assists clinicians in identifying patients at high risk for treatment failure. Patients with viral loads of ≤1,000 copies/mL are at reduced risk of disease progression and HIV transmission (3). Randomized controlled trials comparing routine laboratory monitoring with clinical-only monitoring of ART-initiated patients find reduced mortality and improved clinical outcomes

for immunologically and virologically monitored patients (4–6). Existing studies do not capture implications of an eligibility threshold for repeat viral load testing.

Viral load monitoring is challenging, and issues like specimen transportation, limited laboratory equipment and staff, provider-patient communication, and clinic transfers can impede meeting repeat testing goals (7). In prior research, our group and others found that average time to repeat testing after an initial viral load test was 3 months (8–10), but others find the process can take >6 months (11, 12). It is unknown whether South African health systems are meeting the goal of repeat testing after an elevated viral load within 6 months, and whether being eligible for repeat testing leads to improved patient outcomes.

We evaluated the effect of being eligible for repeat viral load testing on HIV-related outcomes using a national patient cohort developed from South Africa's national laboratory database. To estimate the causal impact of the policy, we used a regression discontinuity design (13–15), comparing patients with viral load values just above and just below 1,000 copies/mL. Because viral load is a noisy, continuous biomarker, patients with viral loads just above and below the threshold are expected to be similar on observed and unobserved factors, enabling robust causal inference without strong assumptions about absence of residual confounding. We assessed the extent to which South African facilities follow guidelines and estimated the effect of this policy on retention, viral suppression, and clinic transfer for patients with viral loads just above and just below 1,000 copies/mL.

METHODS

Sample

We analyzed data from South Africa's National HIV Laboratory Service (NHLS), the sole provider of diagnostic pathology services for public-sector HIV treatment programs. NHLS data includes date and results of each laboratory test, with the date representing date of blood draw. After blood draw, results are processed and sent back to facilities or, in some cases, processed by a computer system. Laboratory results are transferred from the laboratory information system to be stored in the NHLS Corporate Data Warehouse. Previously, we used probabilistic data linkage methods (16) to turn >50 million CD4 count and viral load results into a national HIV cohort. The methods have been described in detail elsewhere and used for national evaluations (17). Briefly, we employed a novel matching method to link records to each other, built on Fellegi-Sunter probabilistic record linkage methods (18), Jaro-Winkler string comparisons (19), and graph-based entity resolution. The method was validated by comparison with a manually matched set of 59,000 potential matches with 93.7% sensitivity and 98.6% positive predictive value (20).

For this analysis, we included deidentified data for all patients with a viral load of 400–2,500 copies/mL in South Africa's public sector HIV program, from January 2013 until July 2015. Patients entered the study on the date of their viral load test and were followed longitudinally through December 2016—at least 18 months for all patients. Patients with multiple viral load results in this period were allowed to enter the study multiple times at each viral load test (viral load test results are the unit of observation). We excluded viral load results if the same patient had an elevated test in the prior 12 months because these could be "repeat" tests. For example, if a patient had 2 elevated viral load results (>1,000) within 12 months, only the first test was included as an observation. The second test was still used to assess outcomes but not considered its own unit of observation. We restricted the sample to laboratory tests after 2012 to retain the most up-to-date data that still allowed for longitudinal outcomes. While the database does not have date of treatment initiation, during the study period guidelines called for viral load testing 6 months after patients were initiated on treatment (21). We included viral loads of 400–2,500 copies/mL to focus on the area around the threshold of 1,000 copies/mL (400 and 2,500 are equidistant from 1,000 in log-units). Viral load assays changed over time to be more sensitive, but all assays detect viral loads of <400 copies/mL, and any change would be balanced above and below the 1,000 copies/mL threshold.

Exposure

Our primary exposure was eligibility for repeat viral load testing, defined as a viral load of >1,000 copies/mL. Patients with a viral load of 400–1,000 copies/mL follow normal monitoring protocols and repeat testing after 1 year (6 months if it was their first test), while those with a viral load of >1,000 copies/mL are eligible for adherence counseling and repeat testing within 3–6 months (Figure 1) (1). We additionally created a variable for receiving a repeat test, defined as ≥1 additional viral load test within 6 months of the test of interest. We did not require a minimum time between the 2 viral load tests.

Outcomes

We defined outcomes by having a qualifying laboratory test (described in each section below) within 12 months (defined as at least 6 up to <18 months, or 183–547 days), 18 months (defined as at least 12 up to <24 months, or 365–729 days) and 24 months (defined as at least 18 up to <30 months, or 548–911 days) after eligibility (Figure 2). All patients had ≥18 months of follow-up and were included in analyses of 12-month outcomes. For 18- and 24- month outcomes, we included only patients with the full range of potential follow-up (24 months for 18-month outcomes; 30 months for 24-month outcomes).

Retention. South African guidelines call for ≥1 laboratory test within a 12-month period, and ≥2 laboratory tests during a patient's first year on treatment. We estimated retention at 12, 18, and 24 months after eligibility as having any CD4 or viral load testing within the respective time intervals (described above).

Retained and virally suppressed. Viral suppression is defined as <400 copies/mL per treatment guidelines (1, 2).

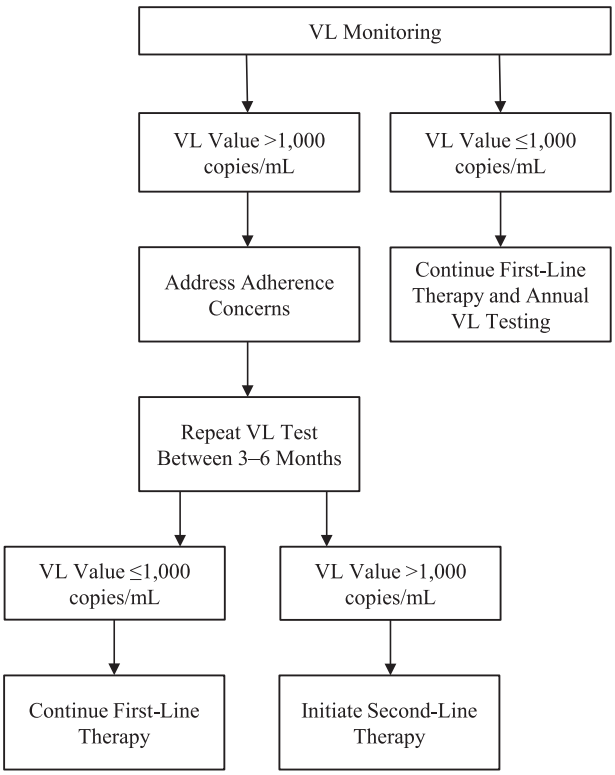

Figure 1. Repeat viral load (VL) monitoring policy, South Africa. Patients with viral load values greater than 1,000 copies/mL are recommended to receive repeat viral load testing within 3–6 months, whereas viral load values up to 1,000 copies/mL are given annual testing.

We defined being retained and virally suppressed at 12, 18, and 24 months after eligibility as having a viral load of <400 copies/mL within the intervals defined above. The negative outcome was having a nonsuppressed viral load (i.e., retained but not suppressed) or having no viral load within the same interval (i.e., not retained, unknown if suppressed).

Clinic transfer. We defined clinic transfer as having ≥1 HIV laboratory test at a different clinic from the viral load result of interest at any point over follow-up.

Study design and statistical methods

Regression discontinuity design. We estimated the effect of eligibility for repeat viral load testing within 6 months (vs. eligibility for annual testing) by comparing outcomes of patients with viral loads just below and above the threshold for action (1,000 copies/mL). A regression discontinuity design is appropriate when the treatment is fully or partially assigned based on a threshold rule for a continuous assignment variable (13–15). Here, "treatment" is intensive monitoring for possible treatment failure featuring repeat viral load testing within 6 months of a viral load of >1,000 copies/mL, which we observe, and adherence counseling, which we do not observe. The continuous assignment variable is the patient's viral load value. Due to random measurement variability in viral load assays, treatment eligibility is as good as randomly assigned for patients near 1,000 copies/mL; therefore, comparing patients with viral loads just above and below the threshold is akin to a locally randomized trial. Because not all patients with viral loads of >1,000 copies/mL will have repeat testing within 6 months, our effect estimates have an intention-to-treat (ITT) interpretation.

Statistical approach. We estimated the relationship between viral load and outcomes using local linear regression models, allowing for a discontinuity at the threshold of 1,000 copies/mL by fitting separate slopes on either side of the threshold. The model is represented by $E[Y|Z] = \beta_0 + \beta_1(Z - c) + \beta_2 Ti + \beta_3((Z - c) \times Ti)$, where β_1 represents the assignment variable (viral load value, Z) centered on the threshold of 1,000 copies/mL (c), β_2 is an exposure indicator

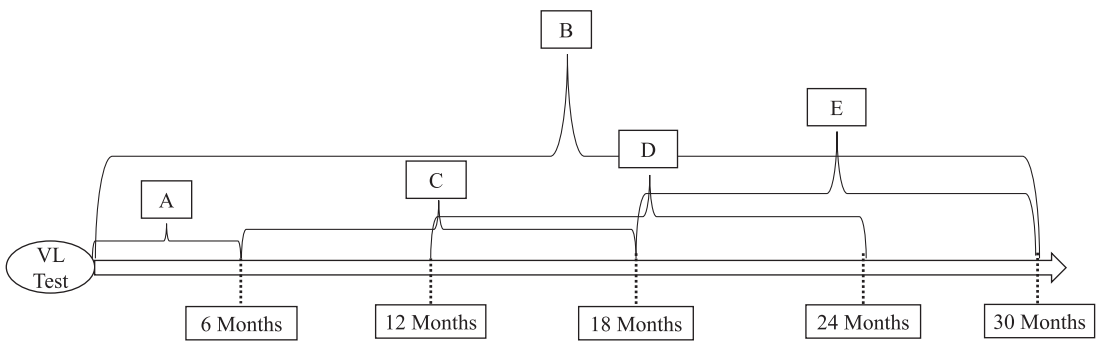

Figure 2. Visual representation of treatment and outcomes for regression discontinuity analysis of patients with an elevated viral load (VL) in South Africa, 2013–2015 A) Treatment includes having a repeat viral load test within 6 months of viral load test. B) Clinic transfer outcome: transferring clinics at any point after the VL test. C) 12-month outcomes are those that occur within the interval of at least 6 up to <18 months (183–547 days) from the viral load test. D) 18-month outcomes are those that occur within the interval of at least 12 up to <24 months (365–729 days) from the viral load test. E) 24-month outcomes are those that occur within the interval of at least 18 up to <30 months (548–911 days) from the viral load test.

Table 1. Demographic and Clinical Factors for Patients With an Elevated Viral Load, Comparing Viral Loads Just Above and Below the Eligibility Threshold for Repeated Testing of 1,000 copies/mL (n = 191,764), South Africa, 2013–2015

Characteristic	Viral Load Test Result		Comparison of Viral Load Test Result of >1,000 vs. ≤1,000 copies/mL	
	Just Below 1,000 copies/mL (n = 81,271)	Just Above 1,000 copies/mL (n = 110,493)	RD[a]	95% CI
Percent of total viral load sample	42.3	57.7		
Age, years[b]	36.2	36.2	0.04	−0.4, 0.5
Female	66.4	66.9	0.5	−1.4, 2.6
Earliest CD4 count[b]	282.7	282.3	−0.39	−8.9, 8.2
No. of days since first CD4 or viral load test[b]	1,180.9	1,193.5	12.6	−22.8, 48.1
Viral load test year				
2013	39.7	39.9	0.02	−2.0, 2.3
2014	39.1	40.0	0.09	−1.1, 2.9
2015	21.2	20.1	−1.1	−2.7, 0.6
Province				
EC	13.0	12.7	−0.3	−1.6, 1.0
FS	4.1	3.7	−0.4	−1.3, 0.6
GP	33.4	34.0	0.6	−1.3, 2.6
KZ	18.5	19.0	0.5	−1.1, 2.2
LP	7.7	7.5	−0.3	−1.3, 0.7
MP	9.4	9.8	0.4	−0.9,1.6
NC	2.0	1.8	−0.2	−0.8, 0.4
NW	6.3	6.2	−0.1	−0.9, 0.8
WC	5.5	5.1	−0.4	−1.5, 0.7
Clinic size, patients/year[b,c]	2,365.9	2,371.6	5.7	−106.2, 117.6

Abbreviations: CI, confidence interval; EC, Eastern Cape; FS, Free State; GP, Gauteng; KZ, KwaZulu-Natal; LP, Limpopo; MP, Mpumalanga; NC, Northern Cape; NW, North West; RD, risk difference; WC, Western Cape.

[a] Estimates derived from separate local linear regression models for each covariate using data-driven optimal bandwidths and rectangular kernel.

[b] Values are expressed as mean and mean difference.

[c] Represents the mean number of patients seen per calendar year among the 4,264 clinics where eligible patients received care (1,911 clinics for viral loads of ≤1,000 copies/mL, and 2,353 clinics for viral loads of >1,000 copies/mL).

for being eligible for a repeat test (viral load of >1,000 copies/mL, T), and β_3 is a term for interaction between the assignment variable ($Z - c$) and exposure indicator (T). ITT effects were estimated on the risk difference scale using separate models for each outcome. β_2 is the risk difference of the outcome for those with viral loads just below and above the 1,000-copies/mL threshold, interpretable as the ITT effect of the policy.

We selected data-driven optimal bandwidths around the threshold using the Calonico, Cattaneo, and Titiunik method (21) and the STATA packages rdrobust and rdbwselect (STATA/MP, version 16.1; StataCorp LLC, College Station, Texas) (22, 23). We used heteroskedasticity-robust standard errors and a rectangular kernel, which weights data equally within the selected bandwidth.

To assess policy adherence, we plotted the predicted probability of receiving a repeat viral load test within 6 months against binned viral load values. We estimated the risk difference of receiving a repeat viral load test at the threshold by fitting a local linear regression model regressing exposure (having a repeat viral load test within 6 months) by model parameters described above.

We tested for systematic manipulation of the assignment variable by plotting a histogram of first viral load results and visually assessing bunching at the threshold of 1,000 copies/mL, and conducting a McCrary density test, which assesses for discontinuity in the density of the assignment variable at the threshold (24, 25). We compared available covariates—age, sex, first CD4 count, time since entry to care, viral load test year, province, and clinic size—for

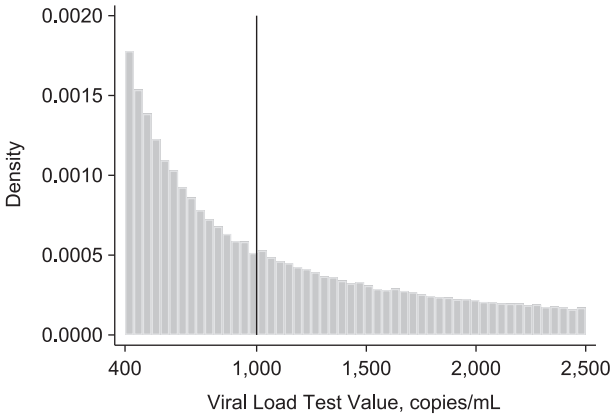

Figure 3. Histogram of viral load values among patients with an elevated viral load in South Africa, 2013–2015. Plotted histogram of viral load laboratory test values to check for bunching on either side of the 1,000 copies/mL threshold. Lack of bunching around the threshold suggests that patients and providers did not systematically manipulate the assignment variable (viral load values) ($n = 191{,}764$).

patients with viral loads just above and below the threshold using local linear regression with Calonico, Cattaneo, and Titiunik bandwidths. Akin to a balance table in a randomized trial, if characteristics are balanced above and below the threshold, this lends confidence that the treatment was as good as randomly assigned. Although balance cannot be assessed on unobserved factors, balance is generated by the treatment assignment mechanism. Due to the nature of the study design, crude regression results are expected to be unconfounded and other covariates are not included in regression models.

Complier average causal effect. Not all patients with viral loads of >1,000 copies/mL received intensive viral monitoring; hence, differences at the threshold are interpreted as the ITT effect of eligibility for intensive monitoring. To understand the effect of intensive viral monitoring itself—rather than eligibility—we used an instrumental variables approach to scale the ITT effect by the share of patients receiving intensive monitoring because of a viral load of >1,000 copies/mL, estimating a complier average causal effect (CACE) (14). CACE risk differences and 95% confidence intervals were estimated for each outcome using 2-stage least squares regression. CACE estimates can be interpreted causally under assumptions of monotonicity and excludability (14). The monotonicity assumption states that the threshold rule can affect a patient's actual treatment status in only one direction. Relative to the counterfactual of viral load of ≤1000, having a viral load of >1,000 copies/mL might increase (or have no effect on) chances of a repeat viral load but cannot decrease chances of a repeat viral load. This is plausible because it is unlikely that indication for repeat testing causally reduces chances of having a repeat test. The excludability assumption requires that eligibility at the threshold affect outcomes only though intensive monitoring, measured by the presence of a repeat viral load.

Sensitivity analyses. In sensitivity analyses (Web Appendixes 1 and 2), we repeated our analysis using a 400-copies/mL threshold (Web Figures 1–2, Web Tables 1–2), excluded patients' first viral load tests (Web Table 3), tested for effect measure modification by CD4 count (Web Tables 4–6) and clinic size (Web Tables 7–10), replicated our analysis allowing patients to enter the study only once (Web Tables 11–12), tested multiple bandwidths (Web Tables 13–14), and repeated our analysis using Calonico, Cattaneo, and Titiunik bias–corrected estimates and confidence intervals (Web Table 15) (22).

Ethical approval

We received ethical approval for the use of deidentified data with a waiver of consent from the Boston University Institutional Review Board and the Human Research Ethics Committee of the University of Witwatersrand.

RESULTS

Our analytical sample included 174,574 patients from 4,264 clinics from 9 South African provinces. The 174,574 patients contributed 191,764 separate first viral load results between January 2013 and July 2015 (81,271 viral loads of 1,001–2,500 copies/mL and 110,493 of 400–1,000 copies/mL). Through December 2016, eligible patients contributed 284,150 CD4 counts and 533,676 viral load results. The 533,676 viral load results include the separate "first" tests (units of observation), and the follow-up tests of interest. See Web Figure 3 for the exclusions flowchart. Characteristics were balanced around the threshold (Table 1). Median age was 35 years (interquartile range, 28–42), 66% of the sample were women, and median first CD4 count was 240 cells (interquartile range, 133–380). Viral load density was smooth across the threshold (difference in log-density at the threshold = −0.02, 95% confidence interval (CI): −0.04, 0.01), indicating no evidence of systematic manipulation of the assignment variable (Figure 3, Web Figure 4).

Adherence to repeat viral load policy

Repeat viral load testing was completed for 72% of patients with a viral loads of >1,000 copies/mL. Median time until repeat testing for viral loads of >1,000 copies/mL was 6.6 months. Approximately one-third of observations had a repeat test within the strict 6-month interval (data not shown). Patients with viral loads eligible for repeat testing (>1,000 copies/mL) were 8.0% (95% CI: 6.2, 9.7) more likely to receive repeat testing within 6 months than those not eligible (32.8% vs. 24.8%) (Table 2 and Figure 4). A majority of patients (67%) with viral loads of >1,000 copies/mL did not receive repeat testing within the recommend time frame. We explored flexibility of the 6-month cutoff by allowing for +3-month range (e.g., repeat test within 9 months). With this date range, repeat testing increased by 12% for all patients, and difference in repeat

Table 2. HIV-Related Outcomes (%) for Patients With an Elevated Viral Load, Comparing Those Above and Below Threshold of 1,000 copies/mL ($n = 191{,}764$), South Africa, 2013–2015

Characteristic	Viral Load Test Result		Comparison of Viral Load Test Result of >1,000 vs. ≤1,000 copies/mL				
	Just Below 1,000 copies/mL	Just Above 1,000 copies/mL	ITT RD[a]	ITT 95% CI	Bandwidth, copies/mL	CACE RD[b] (RD/First-Stage RD)	CACE 95% CI
Repeat viral load testing within 6 months (first-stage RD)	24.8	32.8	8.0	6.2, 9.7	197.6		
Retained at 12 months	62.1	65.0	2.9	0.6, 5.2	136.0	38.1	7.7, 68.4
Retained at 18 months[c]	58.5	60.0	1.6	−0.8, 4.0	161.9	19.9	−10.0, 49.8
Retained at 24 months[d]	54.7	54.6	−0.1	−2.8, 2.5	183.4	1.5	−28.4, 31.4
Retained and suppressed at 12 months[e]	35.8	41.6	5.8	3.0, 8.6	93.7	85.2	35.9, 100.0
Retained and suppressed at 18 months[c,e]	36.3	38.5	2.2	−0.2, 4.6	150.6	31.1	−2.6, 64.8
Retained and suppressed at 24 months[d,e]	36.6	36.7	0.1	−2.3, 2.5	202.2	1.5	−28.3, 31.4
Transferred clinics at least once	18.6	18.9	0.3	−1.4, 2.1	152.4	4.3	−19.2, 27.8

Abbreviations: CACE, complier average causal effect; CI, confidence interval; ITT, intention-to-treat; RD, risk difference.

[a] Difference between viral load tests just above 1,000 cells/mL and viral load tests just below 1,000 cells/mL. Data-driven Calonico, Cattaneo, and Titiunik (21) bandwidth obtained using rdrobust package in STATA.

[b] CACE estimates use outcome optimal bandwidth (not first-stage bandwidth). CACE RDs and 95% CIs estimated using 2-stage least squares regression. The CACE is estimated as a linear probability model, which has no constraints on generating CIs outside the bounds of 0 and 1.0. Square brackets are used where the CACE 95% CI exceeds the logical bounds of 1.0. First-stage RD represents the RD in repeat testing within 6 months for viral loads just above and just below 1,000 copies/mL.

[c] 151,703 viral load values had at least 24 months of follow-up time and were included in 18-month outcomes.

[d] 113,617 viral load values had at least 30 months of follow-up time and were included in 24-month outcomes.

[e] Negative outcome is having a viral load test in that specified time interval that is ≥400 copies/mL. Excludes patients with no viral load test in specified time interval.

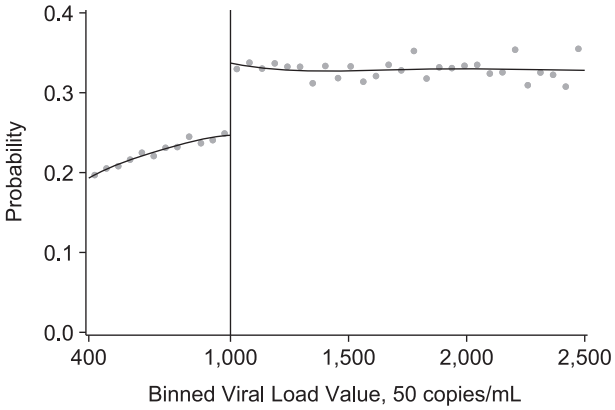

Figure 4. Probability of repeat viral load testing within 6 months among patients with an elevated viral load in South Africa, 2013–2015. Probability of receiving a repeat viral load within 6 months based on viral load values. Risk difference estimated using local linear regression model with a data-driven optimal bandwidth of 197.6 copies/mL. The risk difference at the threshold of 1,000 copies/mL is 8.0% (95% confidence interval: 6.2, 9.7). Gray dots represent sample averages within binned viral load values. Black lines represent polynomial fit of 4.

testing between patients >1,000 copies/mL versus ≤1,000 copies/mL remained approximately 8% (44.9% vs. 37.3%; risk difference (RD) = 7.6%, 95% CI: 5.6, 9.7).

Eligibility for repeat viral load testing and HIV-related outcomes

Eligibility for repeat testing (>1,000 copies/mL) was associated with a 2.9% (95% CI: 0.6, 5.2) increase in retention at 12 months (Table 2 and Figure 5). No difference was found for retention at 18 months (RD = 1.6% ,95% CI: −0.8, 4.0) or 24 months (RD = –0.1%, 95% CI: −2.8, 2.5). Eligible patients were 5.8% (95% CI: 3.0, 8.6) more likely to be both retained and virally suppressed at 12 months (Figure 5). A small difference was found for combined retention and suppression at 18 months (RD = 2.2%, 95% CI: −0.2, 4.6) but no difference at 24 months (RD = 0.1%, 95% CI: −2.3, 2.5) or for clinic transfers (RD = 0.3%, 95% CI: −1.4, 2.1) (Table 2). Sensitivity analyses did not meaningfully alter our conclusions (see Web Tables 1–16).

Eligibility for repeat viral load testing and outcomes among compliers

Given low adherence to the policy, ITT effects on retention and suppression substantially underestimate the effect on patients who actually received the intensive monitoring indicated by an elevated viral load. CACE estimates indicate intensive viral monitoring induced by having a viral load of >1,000 copies/mL increased retention at 12 months by 38.1% (95% CI: 7.7, 68.4) and shared retention and viral suppression at 12 months by 85.2% (95% CI: 35.9, 100.0) (Table 2).

DISCUSSION

This study is one of the first to evaluate adherence to South Africa's HIV viral load monitoring policy on a national level. The results demonstrate high rates of repeat testing after an elevated viral load among HIV patients receiving care in South Africa's public-sector health system but low rates of repeat viral load testing within the stated 6-month guideline. Having an elevated viral load increased the probability of a repeat viral load by just 8%. Two-thirds of patients with elevated viral loads (>1,000 copies/mL) did not receive repeat testing within 6 months, suggesting that South African national monitoring guidelines are not being followed in many cases. Using presence of a repeat viral load as an indicator of intensive monitoring, we estimated the impact for patients of actually receiving monitoring under assumptions of excludability and monotonicity. We found that intensive monitoring increased the probability of being retained and virally suppressed 12 months later by 85%. Therefore, when carried out, guideline-recommended intensive monitoring is effective in improving patient outcomes.

There are a number of patient-, provider-, and institutional-level barriers that might inhibit timely viral load monitoring for eligible HIV patients. Even if a patient is indicated to return for additional monitoring, HIV-related stigma, transportation issues, migrant labor, or poor patient-provider communication might prevent patients from returning for testing. In addition, staffing shortages, power outages, and lack of provider awareness of the importance of routine monitoring might also inhibit consistent monitoring (26, 27).

Our findings suggest that patients who are monitored are more likely to be retained and have resuppression, consistent with previous evidence from sub-Saharan Africa on viral load monitoring and clinical outcomes. A randomized trial from Uganda comparing routine viral load monitoring with CD4 count monitoring and clinical monitoring found that patients in the viral load monitoring arm had 70% fewer AIDS-defining events than those in the clinical monitoring arm (28). Other work shows that patients enrolled in ART programs with viral load monitoring are switched to second-line therapy faster and at higher CD4 counts compared with those using CD4 count monitoring (29). Our study demonstrates that "real-world" viral monitoring, if adhered to, can have positive impacts on viral suppression. While our findings are identified for patients with viral loads close to the threshold of 1,000 copies/mL, there is reason to believe that the behavioral impacts of intensive monitoring (e.g., better adherence, switching to second-line therapy) might be similar for patients at higher viral loads, and benefits of resuppression might be even greater. Policy guidelines are only as good as their implementation, and lack of adherence to the monitoring policy substantially dilutes its impact, as demonstrated by the discrepancy between the results of our ITT and CACE analysis. It is likely that increased adherence to the monitoring guidelines would improve patient outcomes.

Approximately one-quarter of patients with viral loads of ≤1,000 copies/mL received repeat testing within 6 months. One driver of this finding might be that all patients are indicated to have a repeat viral load test within 6 months

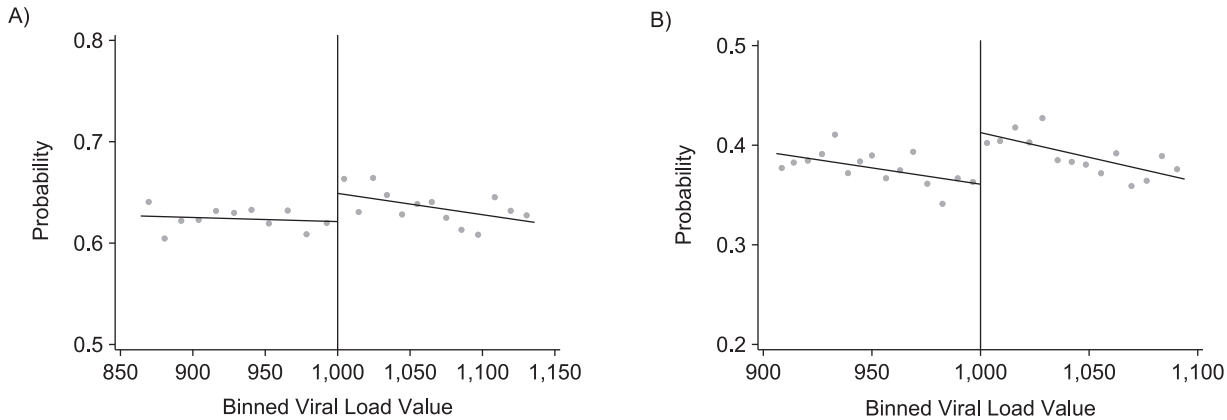

Figure 5. Intention-to-treat results of eligibility for repeat viral load (VL) testing in South Africa, 2013–2015. Results from regression discontinuity analysis of the effect of VL monitoring within 6 months on retention and viral suppression at 12 months. Gray dots represent sample averages within binned VL values. Black lines represent linear polynomial fits. Plots are generated using data-driven optimal bin lengths and bandwidths from corresponding regression estimates. A) Probability of retention at 12 months. The risk difference at the threshold of 1,000 copies/mL is 2.9% (95% confidence interval: 0.6%, 5.2%). Mean bin lengths are 16.5 to the left of the threshold, and 14.1 to the right of the threshold. Bandwidth is +/− 136 copies/mL. B) Probability of retention and suppression at 12 months. The risk difference at the threshold of 1,000 copies/mL is 5.8% (95% confidence interval: 3.0%, 8.6%). Mean bin lengths are 6.3 to the left of the threshold, and 6.7 to the right of the threshold. Bandwidth is +/− 96 copies/mL.

after their first. However, results remained even in sensitivity analyses restricting to patients' second viral load test (e.g., only allowing patients to enter the study once at their viral load was tested a second time) (Web Table 12). The most plausible explanation for this finding is that viral loads might be tested but not always used to inform patient care. For example, providers might have concerns related to self-reported adherence, symptoms, or treatment resistance, prompting repeat testing.

Our study has limitations. First, results could be biased if patients just above and below the threshold of 1,000 copies/mL differed by unobserved covariates that were also associated with outcomes. However, given random measurement variability in testing assays, it is unlikely that patients would systematically differ around the threshold. Balance in observed covariates at the threshold supports our interpretation that the treatment assignment mechanism is "as good as random" for patients close to the threshold. Second, the probabilistic matching used to generate the data set is vulnerable to over- or undermatching, potentially resulting in over- or underestimation of outcomes. Given the study design, linkage errors are unlikely to be correlated with treatment assignment. Third, we relied on laboratory data for all measures, including retention outcomes. Patients might receive care without laboratory testing, and so estimates of retention are likely underestimated. That said, our primary outcome of "retained and virally suppressed" is measured similarly to other cohorts. In addition, turnaround time from blood draw to return of viral load test to the clinic might result in misclassification of time to repeat testing. This misclassification is likely nondifferential in that there is little reason to believe the error would be different for patients just above and below 1,000 copies/mL. Fourth, it is possible some patients could receive adherence counseling without a repeat viral load test. If adherence counseling is the primary mechanism through which repeat testing affects retention and suppression, and not all patients that receive adherence counseling also have a repeat viral load test, then our group of compliers would be greater than indicated by our data, and our CACE findings would be overestimated. We explored this possibility in sensitivity analyses and found that even if 50% of patients receiving adherence counseling do not have a repeat viral load test, repeat testing increased the probability of being retained and virally suppressed at 12 months by 42.6% (Web Table 16).

In conclusion, despite evidence that viral load monitoring improves outcomes, this national policy was not being followed in South Africa's public-sector care system for the majority of patients receiving HIV care between 2013 and 2015. These patients might miss out on adherence interventions and experience delays in diagnosis of treatment failure and initiation of second-line therapy. With the establishment of the World Health Organization 90-90-90 strategy, which aims to see that 90% of all people receiving ART are virally suppressed by 2020 (30), we might see improvements in intensive viral load monitoring in the near future.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts (Alyssa F. Harlow, Jacob Bor, Alana T. Brennan, Matthew P. Fox); Department of Global Health, Boston University School of Public Health, Boston, Massachusetts (Jacob Bor, Alana T. Brennan, William MacLeod, Matthew P. Fox); Health Economics and

Epidemiology Research Office, Department of Internal Medicine, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa (Jacob Bor, Alana T. Brennan, Mhairi Maskew, William MacLeod, Matthew P. Fox); National Health Laboratory Service, Johannesburg, South Africa (Sergio Carmona, Koleka Mlisana); Department of Molecular Medicine and Haematology, University of the Witwatersrand, Johannesburg, South Africa (Sergio Carmona); and Department of Medical Microbiology, School of Laboratory Medicine and Medical Sciences, University of KwaZulu Natal, Durban, South Africa (Koleka Mlisana).

This work was supported by the National Institutes of Health (grants 1R01AI115979-01, 1K01MH105320-01A1, and 1K01DK116929-01A1). W.M. and S.C. were supported by the US Agency for International Development (cooperative agreement AID-674-A-12-0020).

The National Institutes of Health had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The US Agency for International Development played no role in the manuscript, did not review the manuscript, and played no role in the decision to submit.

Conflict of interest: none declared.

REFERENCES

1. World Health Organization. *Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a Public Health Approach*. Geneva, Switzerland: World Health Organization; 2013.
2. South Africa 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*. Pretoria, South Africa: National Department of Health; 2015. https://sahivsoc.org/Files/ART%20Guidelines%2015052015.pdf. Accessed December 6, 2018.
3. Ellman TM, Alemayehu B, Abrams EJ, et al. Selecting a viral load threshold for routine monitoring in resource-limited settings: optimizing individual health and population impact. *J Int AIDS Soc*. 2017;20(S7):e25007.
4. DART Trial Team DT, Mugyenyi P, Walker AS, et al. Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomised non-inferiority trial. *Lancet*. 2010;375(9709):123–131.
5. Apondi R, Bunnell R, Ekwaru JP, et al. Sexual behavior and HIV transmission risk of Ugandan adults taking antiretroviral therapy: 3 year follow-up. *AIDS*. 2011;25(10):1317–1327.
6. Chang LW, Harris J, Humphreys EH. Optimal monitoring strategies for guiding when to switch first-line antiretroviral therapy regimens for treatment failure in adults and adolescents living with HIV in low-resource settings. *Cochrane Database Syst Rev*. 2010;(4):CD008494.
7. El-Sadr WM, Rabkin M, Nkengasong J, et al. Realizing the potential of routine viral load testing in sub-Saharan Africa. *J Int AIDS Soc*. 2017;20(suppl 7):e25010.
8. Rohr JK, Ive P, Horsburgh CR, et al. Marginal structural models to assess delays in second-line HIV treatment initiation in South Africa. *PLoS One*. 2016;11(8):e0161469.
9. Fox MPMP, Van Cutsem GVG, Giddy J, et al. Rates and predictors of failure of first-line antiretroviral therapy and switch to second-line ART in South Africa. *J Acquir Immune Defic Syndr*. 2012;60(4):428–437.
10. Petersen ML, Tran L, Geng EH, et al. Delayed switch of antiretroviral therapy after virologic failure associated with elevated mortality among HIV-infected adults in Africa. *AIDS*. 2014;28(14):2097–2107.
11. Narainsamy D, Mahomed S. Delays in switching patients onto second-line antiretroviral treatment at a public hospital in eThekwini, KwaZulu-Natal. *South Afr J HIV Med*. 2017; 18(1):696.
12. Murphy RA, Court R, Maartens G, et al. Second-line antiretroviral therapy in sub-Saharan Africa: it is time to mind the gaps. *AIDS Res Hum Retroviruses*. 2017;33(12): 1181–1184.
13. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *J Clin Epidemiol*. 2015;68(2):122–133.
14. Bor J, Moscoe E, Mutevedzi P, et al. Regression discontinuity designs in epidemiology. *Epidemiology*. 2014;25(5):729–737.
15. Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: an alternative to the ex post facto experiment. *J Educ Psychol*. 1960;51(6):309–317.
16. Bor J, MacLeod W, Oleinik K, et al. Building a national HIV cohort from routine laboratory data: probabilistic record-linkage with graphs. *bioRxiv*. 2018. (doi: 10.1101/450304). Accessed December 6, 2018.
17. Maskew M, Bor J, Macleod W, et al.. The youth treatment bulge in South Africa: increasing numbers, inferior outcomes among adolescents on ART [abstract]. *Presented at the 21st International AIDS Conference*, Durban, South Africa, July 19, 2005.
18. Fellegi I, Sunter A. A theory for record linkage. *J Am Stat Assoc*. 1969;64(328):1183–1210.
19. Winkler W. *String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage*. Washington, DC: American Statistical Association; 1990:354–359. https://files.eric.ed.gov/fulltext/ED325505.pdf. Accessed July 1, 2020.
20. Maskew M, Bor J, Hendrickson C, et al. Imputing HIV treatment start dates from routine laboratory data in South Africa: a validation study. *BMC Health Serv Res*. 2017;17(1): 1–9.
21. Calonico S, Cattaneo MD, Titiunik R. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*. 2014;82(6):2295–2326.
22. Calonico S, Cattaneo MD, Arbor A, et al. Rdrobust: software for regression-discontinuity designs. *The Stata Journal*. 2017; 17(2):372–404.
23. Calonico S, Cattaneo MD, Arbor A, et al. Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*. 2014;14(4):909–946.
24. McCrary J. Manipulation of the running variable in the regression discontinuity design: a density test. *J Econom*. 2008;142(2):698–714.
25. Smith LM, Lévesque LE, Kaufman JS, et al. Strategies for evaluating the assumptions of the regression discontinuity design: a case study using a human papillomavirus vaccination programme. *Int J Epidemiol*. 2017;46(3): 939–949.
26. Lecher S, Williams J, Fonjungo PN, et al. Progress with scale-up of HIV viral load monitoring—seven sub-Saharan African countries, January 2015–June 2016. *MMWR Morb Mortal Wkly Rep*. 2016;65(47):1332–1335.

27. Roberts T, Cohn J, Bonner K, et al. Scale-up of routine viral load testing in resource-poor settings: current and future implementation challenges. *Clin Infect Dis*. 2016;62(8): 1043–1048.
28. Mermin J, Ekwaru JP, Were W, et al. Utility of routine viral load, CD4 cell count, and clinical monitoring among adults with HIV receiving antiretroviral therapy in Uganda: randomised trial. *BMJ*. 2011;343:d6792.
29. ART-LINC of IeDEA Study Group. Switching to second-line antiretroviral therapy in resource-limited settings: comparison of programmes with and without viral load monitoring. *AIDS*. 2009;23(14):1867–1874.
30. Bor J, Brennan A, Carmona A, et al. *Towards 90–90-90 How Close Is South Africa to Reaching the UNAIDS HIV Treatment Targets?* Johannesburg, South Africa: Health Economics and Epidemiology Research Office, University of the Witwatersrand; 2016. [Policy Brief]. http://www.heroza.org/wp-content/uploads/2016/12/NHLS-AIDS-2016-HE2RO-Policy-Brief-FINALII.pdf. Accessed January 19, 2018.