

RESEARCH ARTICLE

# Treatment eligibility and retention in clinical HIV care: A regression discontinuity study in South Africa

Jacob Bor<sup>1,2,3,4\*</sup>, Matthew P. Fox<sup>1,2,4</sup>, Sydney Rosen<sup>1,4</sup>, Atheendar Venkataramani<sup>5</sup>, Frank Tanser<sup>3,6,7,8</sup>, Deenan Pillay<sup>3,9</sup>, Till Bärnighausen<sup>3,8,10,11</sup>

**1** Department of Global Health, Boston University School of Public Health, Boston, Massachusetts, United States of America, **2** Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, United States of America, **3** Africa Health Research Institute, Somkhele, South Africa, **4** Health Economics and Epidemiology Research Office, Department of Internal Medicine, School of Clinical Medicine, Faculty of Health Sciences, University of Witwatersrand, Johannesburg, South Africa, **5** Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **6** School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa, **7** Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, South Africa, **8** Research Department of Infection & Population Health, University College London, London, United Kingdom, **9** Department of Virology, University College London, London, United Kingdom, **10** Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America, **11** Institute of Public Health, University of Heidelberg, Heidelberg, Germany

\* [jbor@bu.edu](mailto:jbor@bu.edu)



**OPEN ACCESS**

**Citation:** Bor J, Fox MP, Rosen S, Venkataramani A, Tanser F, Pillay D, et al. (2017) Treatment eligibility and retention in clinical HIV care: A regression discontinuity study in South Africa. *PLoS Med* 14(11): e1002463. <https://doi.org/10.1371/journal.pmed.1002463>

**Academic Editor:** Noah Kiwanuka, Makerere University College of Health Sciences, School of Public Health, UGANDA

**Received:** September 19, 2016

**Accepted:** October 25, 2017

**Published:** November 28, 2017

**Copyright:** © 2017 Bor et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data are available from the INDEPTH Data Repository: <http://www.indepth-ishare.org/index.php/catalog/140/> (doi: [10.7796/Bor.PLoSMed.2017.Replication.Data](https://doi.org/10.7796/Bor.PLoSMed.2017.Replication.Data)).

**Funding:** JB received support from NIMH of NIH [K01-MH105320] and the Peter T. Paul Career Development Professorship at Boston University. MPF was funded by NIAID of NIH [R01-AI115979] and by the US Agency for International Development [AID-674-A-12-00029]. SR was

## Abstract

### Background

Loss to follow-up is high among HIV patients not yet receiving antiretroviral therapy (ART). Clinical trials have demonstrated the clinical efficacy of early ART; however, these trials may miss an important real-world consequence of providing ART at diagnosis: its impact on retention in care.

### Methods and findings

We examined the effect of immediate (versus deferred) ART on retention in care using a regression discontinuity design. The analysis included all patients ( $N = 11,306$ ) entering clinical HIV care with a first CD4 count between 12 August 2011 and 31 December 2012 in a public-sector HIV care and treatment program in rural South Africa. Patients were assigned to immediate versus deferred ART eligibility, as determined by a CD4 count  $< 350$  cells/ $\mu$ l, per South African national guidelines. Patients referred to pre-ART care were instructed to return every 6 months for CD4 monitoring. Patients initiated on ART were instructed to return at 6 and 12 months post-initiation and annually thereafter for CD4 and viral load monitoring. We assessed retention in HIV care at 12 months, as measured by the presence of a clinic visit, lab test, or ART initiation 6 to 18 months after initial CD4 test. Differences in retention between patients presenting with CD4 counts just above versus just below the 350-cells/ $\mu$ l threshold were estimated using local linear regression models with a data-driven bandwidth and with the algorithm for selecting the bandwidth chosen ex ante. Among patients with CD4

funded by the US Agency for International Development [AID-674-A-12-00029]. AV received support from NIMH of NIH [K23 MH106362]. FT was supported by South African MRC Flagship [MRC-RFA-UFSP-01–2013/UKZN HIVEPI], NICHD of NIH [R01-HD084233], NIAID of NIH [R01-AI124389], and a UK Academy of Medical Sciences Newton Advanced Fellowship [NA150161]. TB was funded by the Alexander von Humboldt Foundation through the Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research. TB is also supported by the Wellcome Trust, the European Commission, the Clinton Health Access Initiative and NICHD of NIH [R01-HD084233], NIAID of NIH [R01-AI124389 and R01-AI12339] and FIC of NIH [D43-TW009775]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: SR is a member of the Editorial Board of PLOS Medicine.

**Abbreviations:** ART, antiretroviral therapy; CACE, complier average causal effect; ITT, intention-to-treat.

counts close to the 350-cells/ $\mu$ l threshold, having an ART-eligible CD4 count (<350 cells/ $\mu$ l) was associated with higher 12-month retention than not having an ART-eligible CD4 count (50% versus 32%), an intention-to-treat risk difference of 18 percentage points (95% CI 11 to 23;  $p < 0.001$ ). The decision to start ART was determined by CD4 count for one in four patients (25%) presenting close to the eligibility threshold (95% CI 20% to 31%;  $p < 0.001$ ). In this subpopulation, having an ART-eligible CD4 count was associated with higher 12-month retention than not having an ART-eligible CD4 count (91% versus 21%), a complier causal risk difference of 70 percentage points (95% CI 42 to 98;  $p < 0.001$ ). The major limitations of the study are the potential for limited generalizability, the potential for outcome misclassification, and the absence of data on longer-term health outcomes.

## Conclusions

Patients who were eligible for immediate ART had dramatically higher retention in HIV care than patients who just missed the CD4-count eligibility cutoff. The clinical and population health benefits of offering immediate ART regardless of CD4 count may be larger than suggested by clinical trials.

## Author summary

### Why was this study done?

- Starting antiretroviral therapy (ART) at HIV diagnosis has health benefits relative to deferring therapy.
- Understanding the magnitude of these benefits is important for countries expanding treatment eligibility.
- Clinical trials may underestimate the real-world benefits of immediate ART, because in such trials, patients assigned to deferred ART are retained in care and carefully monitored to determine future ART eligibility and to assess trial endpoints.
- Although clinical trials have shown the biological benefits of early ART, the behavioral impacts on retention in care are unknown.

### What did the researchers do and find?

- We assessed the association between immediate versus deferred ART eligibility and clinical retention among 11,036 patients presenting for HIV care in rural South Africa, the country with the largest HIV epidemic and largest ART program worldwide.
- Using a regression discontinuity design, we compared 12-month retention in care among patients presenting just above versus just below the 350-cells/ $\mu$ l CD4 threshold used to determine ART eligibility during the period of study.
- We found that, among these patients, immediate eligibility increased initiation of ART by 25 percentage points and 12-month retention by 18 percentage points.

- Among patients starting ART because they had an eligible CD4 count, 12-month retention was 91%; among patients prevented from starting ART because their CD4 count was above 350 cells/ $\mu$ l, 12-month retention was 21%.

### What do these findings mean?

- Offering ART immediately at diagnosis improves retention in care, reducing the number of patients who screen positive for HIV but are then lost to follow-up before they can start ART.
- The benefits of offering ART immediately at diagnosis, regardless of CD4 count, may be greater than previously thought.

## Introduction

Mass provision of HIV treatment has improved life expectancy in southern Africa [1–3], yet HIV remains the leading cause of death and disability [4]. Recent clinical trials show health benefits to antiretroviral therapy (ART) at high CD4 counts [5–7]; WHO now recommends starting HIV patients on ART at diagnosis [8], and many countries have moved to “treat all” policies [9].

Although clinical efficacy has been demonstrated under trial conditions, the effect of immediate versus deferred ART in clinical settings in the “real world” is largely unknown. In addition to the direct health benefit demonstrated in trials [5–7], starting ART immediately also may reduce the burden of disease by retaining in clinical care patients who would otherwise be lost to follow-up. High rates of attrition have been observed among patients who are not yet eligible for ART and who ostensibly are being monitored for disease progression, leading to missed opportunities for counseling and timely initiation of therapy [10–24]. However, the extent to which immediate ART mitigates attrition is unknown. Clinical trials, designed to minimize attrition in both arms, do not observe this phenomenon and may therefore underestimate the benefits of immediate ART. Observational studies have documented lower retention among pre-ART patients compared to patients on ART [11,12], yet these differences could simply reflect the selection of more highly-motivated patients onto ART, rather than a causal effect of ART on retention in care.

In this study, we assessed the association between immediate (versus deferred) ART eligibility and clinical retention in a large public-sector treatment program in rural South Africa. Using a quasi-experimental regression discontinuity design [25–28], we compared retention for patients presenting with CD4 counts just above and below the 350-cells/ $\mu$ l eligibility cutoff used during the study period. Regression discontinuity can be used when a treatment is assigned, at least in part, based on a threshold rule, such as the CD4 eligibility cutoff for HIV treatment [28]. CD4 measurements have high within-individual variation [29] due to laboratory instrument imprecision, sampling variability in blood draws, and random factors such as ambient temperature at the time of the blood draw. Due to random noise in measured CD4 counts, patients just above and below the cutoff are similar on both observed and unobserved factors, but are assigned to different exposures [28]. At the threshold, outcomes are observed in both counterfactual states of the world (eligible/not eligible), and comparisons have a causal

interpretation [30]. Whereas most observational studies rely on strong assumptions about unobserved confounders, regression discontinuity can achieve balance by design, similar to a randomized controlled trial, and therefore enables causal inferences without strong assumptions [28]. This natural experiment provides a unique opportunity to assess the impact of immediate versus deferred eligibility for HIV treatment in a real-world clinical setting.

## Methods

### Ethics

Ethical approval for data collection and analysis was obtained from the University of Kwa-Zulu-Natal Biomedical Research Ethics Committee. The research in this paper consisted of secondary analysis of preexisting de-identified data and was determined to be “not human subjects research” by the Boston University Medical Campus Institutional Review Board (H-35385, “Analysis of the HIV cascade of care in rural South Africa: A secondary data analysis”).

### Study population

The study population for this analysis consisted of all patients in the Hlabisa HIV Treatment and Care Programme (Hlabisa Cohort) [31] whose first CD4 count specimen was collected between 12 August 2011 and 31 December 2012. The Hlabisa HIV Treatment and Care Programme is a collaboration between the Africa Health Research Institute (<https://www.ahri.org>) and the South African Department of Health. The Hlabisa Cohort includes all patients receiving HIV care and treatment services at government facilities (17 clinics and 1 hospital) in Hlabisa sub-district, a poor, largely rural area where 1 in 3 adults is HIV-infected [32]. Patients initiating ART prior to their first CD4 count were excluded from the study.

Data on CD4 counts, viral loads, dates of ART initiation, and routine HIV clinic visits were obtained for all members of the study population. The Hlabisa HIV Treatment and Care Programme collects data on CD4 counts for patients who have not yet initiated ART, including patients who never initiate ART. Patients entered the study on the date when their first CD4 count specimen was collected for lab testing, typically the date of HIV diagnosis. Test results were transferred directly from the laboratory into the Hlabisa Cohort database. All patients were eligible to be followed for at least 12 months. Follow-up was closed on 1 January 2014.

### Outcomes

The primary outcome was 12-month retention in care, which was defined as evidence of any routine clinic visit, lab result (CD4 or viral load), or date of ART initiation within the interval 6 to <18 months after a patient’s first CD4 count, regardless of receipt of ART. By South African national guidelines, all patients would be expected to have at least 2 documented lab tests within this period (Appendix A in [S1 Appendices](#)). Although guidelines delineate semi-annual laboratory monitoring, the wide interval allows for the fact that many patients were late for appointments but still retained in care. Results were robust to narrower intervals.

The primary outcome was designed to capture all clinical contact specified by the national guidelines for pre-ART and ART care. Patients on ART had more opportunities to appear and be classified as retained due to their greater frequency of scheduled clinic visits. In sensitivity analysis, we excluded routine clinic visits and counted patients as retained at 12 months only if they had a CD4 or viral load test or initiated ART during the period 6 to <18 months after their first CD4 count.

As a secondary outcome, we assessed the presence of a CD4 or viral load test or ART start date within 6-month intervals following a patient’s first CD4 count, out to 2 years (0 to <6, 6

to <12, 12 to <18, and 18 to <24 months). Because some patients do not return for lab tests precisely every 6 months, these 6-month intervals will underestimate the proportion of patients retained. However, these analyses may inform how retention evolves over time. Analyses of retention in care out to 18 and 24 months were constrained to the subpopulations observed for that amount of time, i.e., patients presenting before 2 July 2012 (18 months follow-up) and 1 January 2013 (24 months follow-up).

## Exposures

Per South African guidelines during the study period, patients were ART eligible if their CD4 count was <350 cells/ $\mu$ l and/or they had a WHO stage III/IV condition [33]. After blood was drawn for a CD4 count, all patients were instructed to return to the clinic in one week to receive their result. ART-eligible patients were enrolled in several weeks of individual and group counseling and then initiated on ART. At ART initiation, patients were instructed to return for monthly clinic visits to pick up their medication and at 6 and 12 months post-initiation (and annually thereafter) for CD4 count and viral load monitoring. Patients not yet eligible for ART were referred to pre-ART care and were instructed to return every 6 months for CD4 monitoring [31].

Based on these policies, we defined two exposures. First, we defined ART eligibility as having a CD4 count below 350 cells/ $\mu$ l. As our second exposure, we defined ART uptake as initiation of therapy within 6 months of a patient's first CD4 count. Not all patients who had an eligible CD4 count went on to initiate ART: some did not return for their CD4 count results, and others did not complete the counseling sessions even after eligibility was determined. Conversely, some patients with CD4 counts at or above 350 cells/ $\mu$ l initiated ART on account of stage III/IV HIV illness or due to provider discretion. Thus, results for our primary exposure—an ART-eligible CD4 count—have an intention-to-treat (ITT) interpretation. We defined ART uptake at 6 months because, by national guidelines, patients who did not start ART within 6 months had another CD4 count to determine eligibility.

## Study design

To determine the effect of immediate versus deferred ART eligibility on retention, we used a quasi-experimental regression discontinuity design. Regression discontinuity can be implemented when a treatment is assigned based, at least in part, on a threshold rule on a continuous assignment variable [27,28,34–36]. Though commonly used in economics [37–39], regression discontinuity has only recently made inroads in epidemiology and clinical research [26–28,40–42]. Because of random measurement error in the CD4 count laboratory assay [29], assignment to immediate versus deferred treatment is effectively random for those patients with CD4 counts near 350 cells/ $\mu$ l. As such, comparisons of outcomes “just above” and “just below” this threshold have a causal interpretation (see Appendix B in [S1 Appendices](#)).

Our analytic strategy, which was based on a preexisting, single, well-known clinical practice threshold, followed best practices for the conduct and reporting of regression discontinuity designs [37,38,40,43–45]. Our primary analysis tested the null hypothesis, determined a priori, that immediate (rather than deferred) ART eligibility would have no effect on retention in care. We evaluated the relationship between the value of a patient's first CD4 count and retention in care, allowing for a discontinuity at the threshold of 350 cells/ $\mu$ l and different slopes on either side of the threshold. Risk differences at the threshold were estimated using local linear regression with a data-driven Imbens-Kalyanaraman bandwidth and a rectangular kernel. We assessed robustness of the results to a wide range of alternate bandwidths, following the literature (see Appendix B, pp. 3–4, in [S1 Appendices](#)) [45,46]. Results

from these models are ITT effects, i.e., differences in retention for patients assigned to immediate versus deferred treatment eligibility by their CD4 count. The data-driven bandwidth selector chooses the bandwidth that minimizes the mean squared error of the difference in predictions at the threshold (i.e., the ITT effect). The goal is to identify as large a region as possible in which the conditional expectation function (relationship between CD4 count and retention) is approximately linear. The more data included, the less random error in the prediction at the threshold, but also the greater the potential for bias if the relationship is in fact nonlinear [45]. Perhaps the greatest advantage of using a data-driven bandwidth selector is that we eliminate the ability for the investigator to manipulate the results by choosing a “preferred” bandwidth. All models were estimated using a rectangular kernel, i.e., weighting observations within the window of data equally. Additionally, we estimated local logistic regression models and estimated predicted margins at the threshold. Because optimal bandwidth selectors are not currently available for logistic regression, we used the same bandwidth as for the local linear model.

We also assessed the association between immediate versus deferred eligibility and 6-month uptake of ART and used this analysis to estimate the share of patients for whom the decision to initiate ART was based on the eligibility of their CD4 count (so-called compliers) as opposed to disease stage or other factors [47]. Using an instrumental variables approach [37,48], we then estimated the effect of ART uptake on 12-month retention for compliers, using CD4 count < 350 cells/ $\mu$ l as an instrument for ART uptake. Under the assumption that having an eligible CD4 count affected 12-month retention only through uptake of ART, these instrumental variables estimates can be interpreted as the causal effect of ART initiation on retention among compliers (see Appendix B, pp. 5–7, in [S1 Appendices](#)). We estimated complier causal risk differences, also known as complier average causal effects (CACEs) or local average treatment effects (LATEs), using 2-stage least squares regression. We additionally estimated proportions of patients retained among patients who started ART because they were eligible (so-called treated compliers) and among patients who did not start ART because they were ineligible (so-called control compliers), and estimated complier causal relative risks as the ratio of the treated and control complier proportions [49].

The validity of the regression discontinuity design rests on the assumption that other patient characteristics that may influence retention are similar for patients with CD4 counts just above and below 350 cells/ $\mu$ l. To evaluate covariate balance, we assessed whether observed factors (age, sex, date of presentation, and clinic of presentation) were similar on either side of the threshold. Bias can also result if patients or providers manipulate CD4 count values in order to gain access to treatment. To assess systematic manipulation, we tested for heaping of CD4 counts on one side of the threshold [44]. All analyses were conducted using Stata/SE version 14.2. This report has been prepared according to STROBE guidelines, recommended by the Enhancing the QUALity and Transparency Of health Research (EQUATOR) network ([S1 STROBE Checklist](#)). A de-identified analytic dataset ([S1 Data](#)) and replication code ([S1 Code](#)) are available.

## Results

### Sample characteristics

A small proportion of patients (4.1%) initiated ART prior to their first CD4 count and were excluded from the study. The remaining sample included 11,306 patients who entered care with a first CD4 count between 12 August 2011 and 31 December 2012. Of these, 6,225 had CD4 counts < 350 cells/ $\mu$ l and 5,082 had CD4 counts  $\geq$  350 cells/ $\mu$ l. Baseline characteristics were similar just above and below the CD4 count threshold; about 70% of patients were

**Table 1. Balance in baseline characteristics of patients just above and below the 350-cells/ $\mu$ l CD4 count threshold.**

Characteristic	Predicted means for patients with:		Difference at the threshold (95% CI)	<i>p</i> -value	IK bandwidth (cells/ $\mu$ l)	<i>N</i>
	CD4 count just below 350 cells/ $\mu$ l (ART eligible)	CD4 count just above 350 cells/ $\mu$ l (not yet ART eligible)				
Age (years)	30.5	30.4	0.1 year (-1.2, 1.5)	0.856	124.6	4,231
Female	70.4%	72.3%	-1.9% (-6.8, 3.1)	0.453	143.4	4,825
Date of first CD4 count	13 April 2012	24 April 2012	-11.6 days (-26.9, 3.6)	0.135	153.6	5,176
Clinic A	14.1%	13.0%	1.0% (-2.8, 4.8)	0.598	142.9	4,851
Clinic B	12.2%	14.9%	-2.7% (-6.7, 1.3)	0.185	128.2	4,415
Clinic C	17.5%	13.8%	3.8% (-1.2, 8.7)	0.138	83.2	2,955

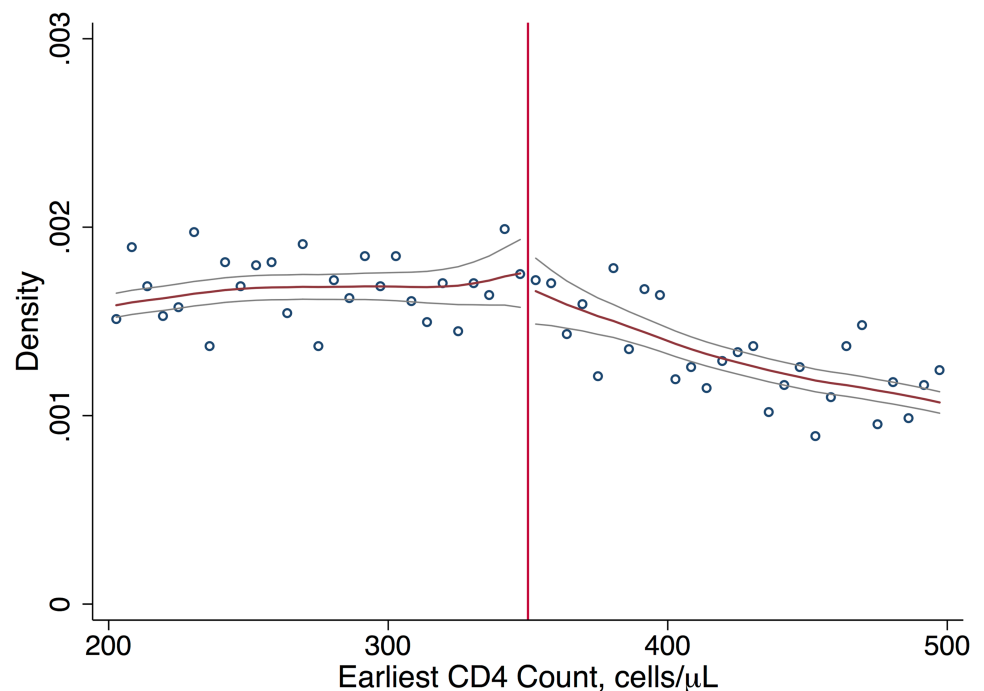
Total *N* = 11,306. Each row displays predicted means just above/below the 350-cells/ $\mu$ l threshold for a different baseline characteristic. The table is analogous to a balance table in a clinical trial: systematic differences between patients just above/below the cutoff would suggest that treatment assignment was not random. Predicted means were based on local linear regression models in which each baseline characteristic was regressed on first CD4 count, with different slopes on either side of the threshold and an intercept shift at the threshold. Each row is based on a separate regression model estimated using data within a window determined by the data-driven Imbens-Kalyanaraman (IK) bandwidth for that characteristic. The IK bandwidth and sample size (*N*) for each regression is reported. Although the total *N* is 11,306, the units included in each model depend on the bandwidth.

<https://doi.org/10.1371/journal.pmed.1002463.t001>

women, and the average age was 30 years (Table 1). We found no evidence of systematic manipulation of CD4 count values around the threshold (Fig 1).

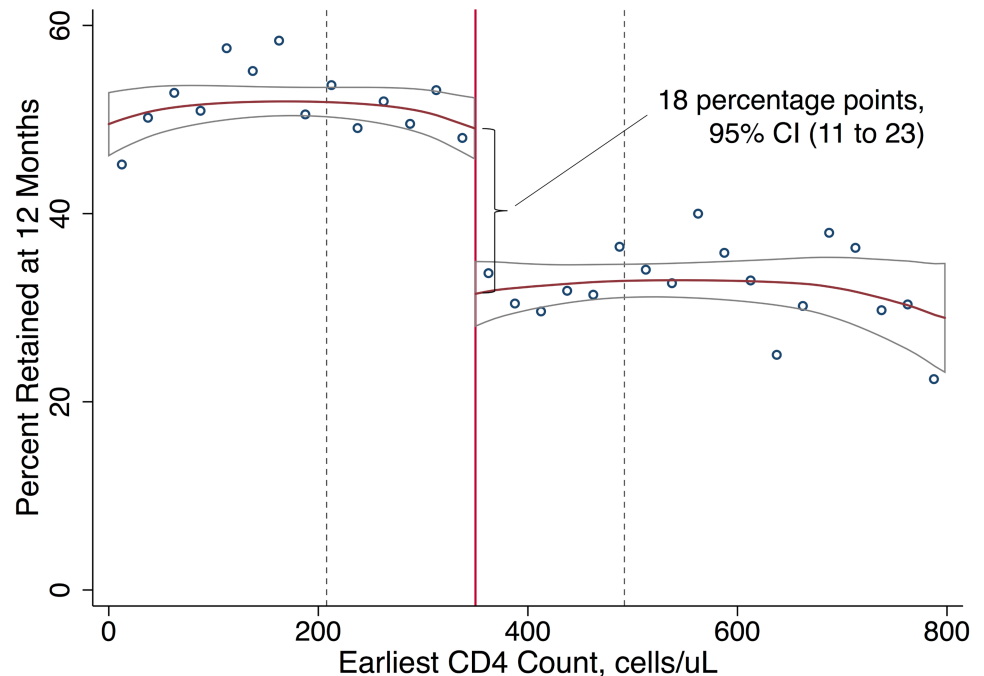
### Immediate versus deferred ART eligibility and retention in clinical care

Among patients presenting with first CD4 counts close to 350 cells/ $\mu$ l, immediate ART eligibility (CD4 count < 350 cells/ $\mu$ l) was associated with higher 12-month retention in care,



**Fig 1. Density of first CD4 counts.** Continuity in the density of first CD4 counts supports our interpretation that patients and providers did not systematically manipulate CD4 counts, e.g., to gain eligibility for ART.

<https://doi.org/10.1371/journal.pmed.1002463.g001>



**Fig 2. Immediate ART eligibility leads to significant gains in 12-month retention.** Twelve-month retention is defined as having a CD4 count or viral load test, initiating ART, or attending a routine clinic visit within the period 6 to <18 months after first CD4 count. The sample excludes patients with <18 months of potential follow-up. Local linear regression estimated with Imbens-Kalyanaraman optimal bandwidth = 142.1 cells. The difference at the threshold was 18 percentage points (95% CI 11 to 23). The effect of interest in regression discontinuity is the difference in the local linear predictions at the threshold, i.e., in the limit, as the threshold is approached from above versus below. The bandwidth defines the region in which the relationship between first CD4 count and the outcome is assumed to be linear in our local linear regression models (Table 2).

<https://doi.org/10.1371/journal.pmed.1002463.g002>

with a difference of 18 percentage points (95% CI 11 to 23), relative to deferred eligibility. Fifty percent of patients with immediate eligibility were retained at 12 months compared to 32% of patients with deferred eligibility, a 56% relative increase with eligibility (Fig 2; Table 2). In sensitivity analyses defining retention in terms of lab results or ART start dates (and excluding clinic visits), estimates were somewhat attenuated, with a difference in 12-month retention of 11 percentage points (95% CI 4 to 18) (Table 2; Fig C1 in S1 Appendices). A gap in retention was observed at all 6-month intervals from first CD4 count to 24 months (Table 2; Fig C2 in S1 Appendices), suggesting that loss to follow-up among patients not yet eligible for ART occurred soon after their initial clinic visit. For all retention outcomes, results were robust to a wide range of alternate bandwidths (Tables D1–D5 in S1 Appendices). Similar results were observed using a logistic rather than linear probability model (Table D6 in S1 Appendices).

### Immediate versus deferred eligibility and ART uptake

Turning to uptake of ART, patients immediately eligible for ART were 25 percentage points (95% CI 20 to 31,  $p < 0.001$ ) more likely to initiate ART within 6 months (Table 2) than those not yet eligible for ART, rising from 18% initiating ART among patients just above the threshold to 43% initiating ART among patients just below the threshold (Fig 3). Even among patients with an eligible CD4 count, a majority (57%) did not initiate ART within 6 months. Fig C3 in S1 Appendices shows that having an eligible CD4 count had no effect on initiation



**Table 2. Intention-to-treat effects of ART eligibility on ART initiation and retention in HIV care.**

Outcome	ART initiation by 6 months	Retained at 12 months (labs, ART, clinic visits)	Retained 0–6 months (labs, ART)	Retained 6–12 months (labs, ART)	Retained 12–18 months (labs, ART)	Retained 18–24 months (labs, ART)	Retained at 12 months (labs, ART)
<b>Risk difference at 350-cells/μl CD4 threshold</b>							
Regression coefficient	25.4	17.9	17.1	8.2	4.6	9.1	11.2
95% CI	(19.7, 31.1)	(11.4, 24.3)	(11.3, 22.9)	(3.8, 12.6)	(−1.0, 10.1)	(2.4, 15.8)	(4.2, 18.1)
<i>p</i> -Value	<0.001	<0.001	<0.001	<0.001	0.108	0.007	0.002
<b>Predicted outcomes at 350-cells/μl CD4 threshold</b>							
Eligible for ART (CD4 just below 350)	43.2	49.7	47.4	28.8	21.7	19.0	41.0
Not eligible for ART (CD4 just above 350)	17.8	31.8	30.3	20.6	17.2	9.9	29.9
IK bandwidth, cells/μl	96.4	142.1	114.2	164.7	125.4	164.2	116.8
<i>N</i>	3,354	3,327	3,937	5,478	2,954	1,734	2,733

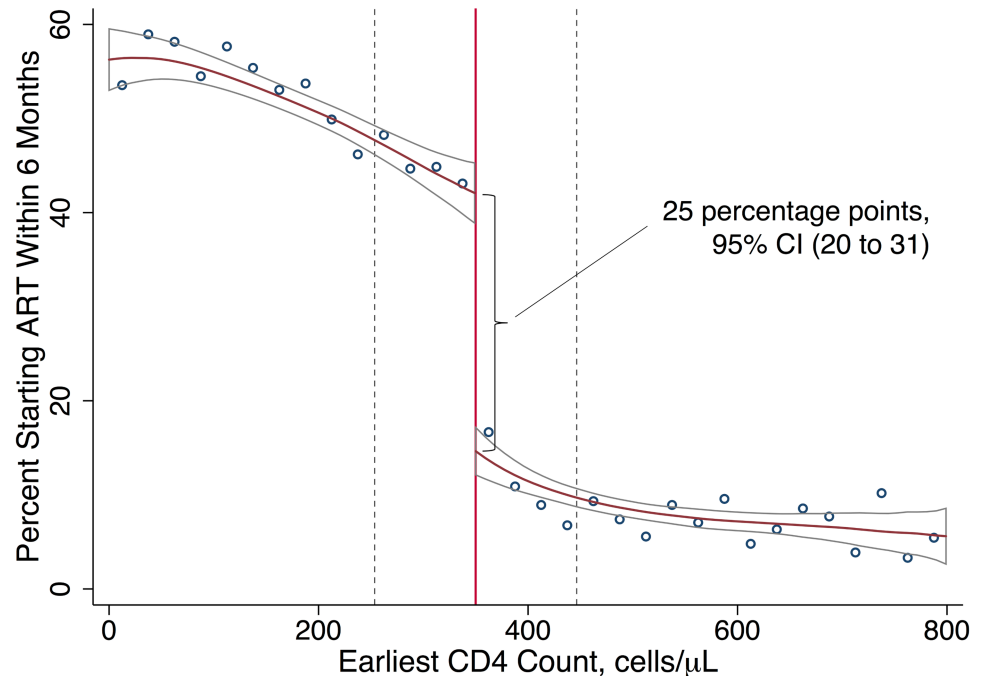
Each column reports the results of a separate linear probability regression discontinuity model, which includes an intercept, an intercept shift at the threshold, and different slopes on either side of the threshold. Results are presented on a percentage point scale (×100). The risk difference estimate shows the regression coefficient, heteroskedasticity-robust 95% CI, and *p*-value for the test that the coefficient is equal to zero. Predicted outcomes at the threshold were estimated by the constant in the regression (prediction if not eligible) and by the sum of the constant and the risk difference (prediction if eligible). Models were estimated for a window of data around the threshold equal to twice the Imbens-Kalyanaraman (IK) optimal bandwidth, which was estimated separately (and is reported separately) for each outcome. All regression coefficients, including coefficients for the slopes of the regression lines on either side of the threshold, are reported in Table D1 in [S1 Appendices](#).

<https://doi.org/10.1371/journal.pmed.1002463.t002>

within the first 2 weeks, consistent with treatment guidelines. The gap in ART uptake apparent at 6 months persisted at 12 months. These results imply that among patients with CD4 counts close to 350 cells/μl, 18% would have initiated ART regardless of CD4 eligibility (so-called always takers), 57% would not have initiated ART regardless of CD4 eligibility (so-called never takers), and 25% of patients would have initiated ART if CD4 count < 350 cells/μl, but not if CD4 count ≥ 350 cells/μl (so-called compliers [48]) (Fig 3). Our study population was followed from the day their first CD4 count was taken, typically the day of diagnosis, regardless of whether patients returned for their results. “Never takers” therefore include patients who tested positive and never came back.

### ART uptake and retention in care among compliers

The results for ART uptake reveal that the ITT effect of CD4 eligibility on retention was substantially diluted by noncompliance—the decision to start ART was based on CD4 count for just 1 in 4 patients presenting close to the threshold. The ITT effect thus underestimates the effect of ART uptake itself on retention in care. Our instrumental variables results for compliers revealed that patients who initiated ART because they had an eligible CD4 count were 70 percentage points (95% CI 42 to 98) more likely to be retained at 12 months than patients who were not initiated because they had an ineligible CD4 count. Retention at 12 months was 91% for compliers assigned to immediate eligibility and 21% for compliers assigned to deferred eligibility. Among compliers, immediate eligibility reduced attrition by 89% (complier causal relative risk = 0.11, 95% CI 0.00–0.32) (Table 3; Table D7 in [S1 Appendices](#)).



**Fig 3. ART uptake increases with an eligible CD4 count.** Local linear regression estimated with Imbens-Kalyanaraman optimal bandwidth = 96.4 cells. The difference at the threshold was 25 percentage points (95% CI 20 to 31). The effect of interest in regression discontinuity is the difference in the local linear predictions at the threshold, i.e., in the limit, as the threshold is approached from above versus below. The data-driven bandwidth refers to the region in which the relationship between first CD4 count and the outcome is assumed to be linear in our local linear regression models (Table 2).

<https://doi.org/10.1371/journal.pmed.1002463.g003>

## Discussion

Understanding the extent to which immediate ART initiation mitigates loss of health and life due to failure to remain in care is important for countries and funding agencies considering WHO recommendations to start patients immediately on therapy regardless of CD4 count [8,9]. Although lower retention has been observed in pre-ART patients compared to ART patients in a wide range of settings, it was hitherto unknown whether this reflected a causal relationship or selection of highly motivated patients onto ART [11,12]. If starting ART causally increased retention in care, then the real-world benefits of immediate ART would be underestimated in clinical trials that actively retain patients not yet eligible for therapy [5–7].

Using a quasi-experimental regression discontinuity design, we found that having an ART-eligible CD4 count at diagnosis significantly improved retention in care for HIV patients in rural South Africa—by 18 percentage points in the ITT analysis. The ITT effect was diluted by the fact that the decision to start ART was based on CD4 count for only a minority of patients (just 1 in 4) presenting for care. (Other patients started ART due to disease stage, while others did not start ART in spite of having an eligible CD4 count.) Among patients whose treatment decision was based on their CD4 count, immediate ART eligibility increased 12-month retention by 70 percentage points relative to deferred eligibility, from 21% to 91% retained.

The retention advantage for ART-eligible patients is perhaps surprising. Clinical guidelines specified that patients who did not start ART should return for CD4 monitoring every 6 months to reassess eligibility. One might expect that patients who are motivated to start life-long ART would tolerate a 6-month delay without exiting care. Further, because the lifetime benefits of ART are greater the earlier a patient starts therapy, there is a strong rationale to stay

**Table 3. Retention in care among patients whose treatment decision was based on the value of their CD4 count.**

	Point estimate	95% CI <sup>a</sup>
<b>Percent retained at 12-months</b>		
Patients who started ART because CD4 count < 350 cells/μl ( <i>compliers assigned to immediate eligibility</i> )	91.0%	(75.8, 100.0]
Patients who did not start ART because CD4 count ≥ 350 cells/μl ( <i>compliers assigned to deferred eligibility</i> )	20.8%	[0.0, 45.8)
Patients who started ART and would have started regardless of CD4 count ( <i>always takers</i> )	86.9%	(77.5, 96.0)
Patients who did not start ART and would not have started regardless of CD4 count ( <i>never takers</i> )	22.1%	(16.5, 28.6)
<b>Absolute and relative effect estimates</b>		
CACE <sup>b</sup> —percentage point difference in retention	70.2	(40.9, 100.0]
CCRR <sup>c</sup> —relative risk of attrition	0.11	[0.00, 0.32)

*N* = 2,366. Estimates calculated based on local linear regression models with a bandwidth of 100 cells/μl. Details on estimation of percent retained among subgroups are provided in Appendix D in [S1 Appendices](#). Percent retained among compliers assigned to deferred eligibility was obtained by subtracting CACE from the percent retained among compliers assigned to immediate eligibility.

<sup>a</sup>All 95% CIs were obtained using the percentile bootstrap (501 resamples). Square brackets are used where the bootstrap CI exceeded logical bounds, e.g., probability less than zero. For CACE, the percentile bootstrap CI was similar to the standard asymptotic 95% CI using heteroskedasticity-robust standard errors: (42.3, 98.1). Both CACE and CCRR were significantly different from the null hypothesis of no effect, *p* < 0.001.

<sup>b</sup>CACE is interpretable as a risk difference and is also known as the local average treatment effect (LATE). CACE was estimated by 2-stage least squares regression.

<sup>c</sup>CCRR is presented in terms of the risk of attrition (1 – retention): among compliers, immediate eligibility reduced attrition by 89%.

CACE, complier causal relative risk; CCRR, complier causal relative risk.

<https://doi.org/10.1371/journal.pmed.1002463.t003>

in pre-ART care in order to initiate as soon as possible. Conceivably, the retention advantage could even favor pre-ART patients as some patients who initiate ART subsequently exit care after experiencing drug side effects and the inconvenience of daily therapy. With respect to the burdens imposed on patients, it might be easier to retain patients on a holding regimen of semi-annual pre-ART appointments than to retain patients on an intensive daily drug therapy. Patients who do not start ART also might be more likely to get sick and need care 12 months later, leading to higher retention in this group.

Contrary to these speculations, immediate eligibility for ART sharply increased retention in care. There are several plausible explanations for the observed results. First, starting ART may shift patients' cognitive reference point for future care-seeking decisions [50]. Patients who have started ART may perceive large costs to defaulting therapy, compared to the more modest costs of delaying initiation among those who have not yet started therapy. These perceptions may be reinforced by clinical guidance to patients that starting ART involves a commitment to take treatment for life [31]. Second, taking daily therapy and returning to the clinic for monthly prescription refills may facilitate habit formation [51], increasing long-term retention in care. Third, auxiliary interventions targeted to ART patients including adherence counseling, support groups, appointment reminders, and outreach to patients who miss appointments may lead to differential retention between patients on ART and

those not yet on ART [52,53]. Fourth, in the context of still-rampant HIV stigma, fear of HIV status disclosure may be a significant barrier to care-seeking among patients not yet on ART [54,55]. Many clinics strongly encourage patients to disclose their HIV status to close family and friends at the time of ART initiation [31]. For ART patients, HIV status disclosure may represent an up-front investment that reduces the costs of future clinic visits. In addition to the behavioral mechanisms above, a fifth possibility is that patient health and quality of life improve due to the antiretroviral drugs themselves [56] and that experiencing these benefits encourages patients to remain in care.

In addition to the benefits of immediate ART eligibility, it is also possible that patients who are told that they are not yet eligible for ART may be inadvertently discouraged from seeking care in the future. The message of deferred eligibility may falsely signal to patients that they do not need or would not benefit from ART. They may also experience anger, hopelessness, or demoralization if they wish to start therapy but are not allowed to, and these experiences may color their attitudes towards the health system and future care-seeking.

Our results highlight the challenges in retaining in care those patients who test positive but are not yet eligible for ART in a resource-limited setting [57]. And yet, the high rates of retention among patients initiated on ART because they were eligible suggest that, in fact, we already have effective techniques to improve retention among patients who have not yet started ART. Further research is needed to identify what specific aspects of initiating ART lead to improved retention in care. These findings may be important for retaining patients who do not wish to start ART on the day of diagnosis, as is now called for in 2017 WHO guidelines [58].

## Causal interpretation of the estimates

Our results were obtained using a regression discontinuity design, a quasi-experimental study design that enables causal inference without the strong assumptions required in most observational studies [26–28,35,40,43,59–61]. So long as values of CD4 measurements are not systematically manipulated by patients or providers, random variability in measured CD4 counts guarantees that patients will be similar (in expectation) in a small range on either side of the 350-cells/ $\mu$ l eligibility threshold [30,35]. We obtained CD4 counts directly from laboratory records (rather than clinical charts) and found no evidence of manipulation, which, if systematic, would lead to heaping in the density of CD4 counts on one side of the threshold [44]. Additionally, there were no systematic differences in observed baseline covariates between patients just above versus below the cutoff. Although we cannot test the assumption that unobserved factors are balanced at the cutoff, our knowledge of the assignment mechanism, the absence of systematic manipulation, and balance on observed characteristics all point to a data-generating process in which quasi-random variation guarantees balance on all factors, similar to a randomized trial.

As with all regression discontinuity designs, causal effects are theoretically identified at the threshold (i.e., in the limit, as the 350-cells/ $\mu$ l cutoff is approached from above and below). In finite samples, however, these causal effects must be estimated using data further from the threshold. We followed current best practice in using local linear regression to estimate the empirical relationship between measured CD4 count and the probability of retention, allowing for an intercept shift at the threshold and different slopes on either side of the threshold [46,62]. The intercept shift at the threshold—i.e., the difference in regression predictions just above and just below 350 cells/ $\mu$ l—estimates the causal effect at the threshold. When using a local linear model to approximate a potentially nonlinear relationship, a key choice is the bandwidth governing the window of data used in the analysis. While a larger bandwidth will

increase the precision of the estimates, this may come at the cost of some bias. We used the data-driven Imbens-Kalyanaraman optimal bandwidth selector, which minimizes the mean squared error (variance plus squared bias) of the difference in predictions at the threshold [45].

In many classical applications of the regression discontinuity design, the assignment variable is associated with the outcome and with the size of the treatment effect. Interestingly, retention was not substantially correlated with first CD4 count in our sample, and slopes were similar on either side of the threshold, evidence that treatment effects may be constant, at least within a range around the 350-cells/ $\mu\text{l}$  cutoff. Because we did not know this *ex ante*, we nevertheless present results from models allowing for different slopes and constrain inferences to the area around the cutoff. By choosing a local effect estimand, our estimates do not rely on extrapolation into unobserved regions nor on assumptions about the functional form of the relationship between CD4 count and retention across the full range of the data [35].

In addition to estimating the ITT effect of ART eligibility on retention, we also estimated the effect of ART initiation itself on retention in care using the eligibility threshold as an instrumental variable. These estimates are interpretable as the effect of starting ART for so-called compliers, *i.e.*, those patients for whom the decision to start (or not to start) ART was based on the value of their CD4 count *vis-à-vis* the 350-cells/ $\mu\text{l}$  threshold. These instrumental variables estimates have a causal interpretation under two additional assumptions. The first assumption, known as the monotonicity or “no defiers” assumption, is that having an eligible CD4 count only increases the chances that a person will start ART. Monotonicity would be violated if there are patients who would start ART if they were ineligible ( $\text{CD4} \geq 350$  cells/ $\mu\text{l}$ ) but would not start ART in a counterfactual world in which they had an ART-eligible CD4 count. It is difficult to conceive of such cases, and this assumption is likely met in our study. The second assumption, known as excludability, is that eligibility differences at the 350-cells/ $\mu\text{l}$  threshold affect retention only through ART uptake. This untestable assumption could be violated if eligibility led to other differences in care apart from ART initiation (*e.g.*, screening for other conditions or pre-ART counseling) that led to increases in later engagement with care. Monotonicity and excludability assumptions are not required for a causal interpretation of the ITT result.

A key strength of regression discontinuity designs (*vis-à-vis* clinical trials) is the ability to assess the causal effects of interventions implemented in real-world settings and in population-representative samples [63]. We studied the complete patient population accessing public-sector HIV care and treatment in one of the poorest and highest HIV-prevalence sub-districts in South Africa. Although our analysis was limited to one sub-district of one country, potentially limiting generalizability, the model of service delivery—decentralized, nurse-led, clinic-based—is common in other HIV-endemic areas of sub-Saharan Africa. Additionally, by including the complete patient population, we avoided the sample selection bias that can result from opt-in participation in clinical trials [64].

Our retrospective analysis of a quasi-experiment avoided many of the potential pitfalls that can lead to bias when investigators knowingly or unknowingly affect outcomes. First, the CD4 count threshold we investigate was set by policy-makers in advance of the study and could not be manipulated by the investigators. Second, patients, providers, and investigators were all blinded to the CD4 count measurement (and hence eligibility status) of the patients at the time when blood was drawn for the patients' first CD4 count. Because we obtained the CD4 results for all blood samples directly from the National Health Laboratory Service, it would have been very difficult for eligibility assignment to be manipulated. Third, the data that we analyzed were collected as part of routine laboratory and clinical monitoring of patients in the Hlabisa HIV Treatment and Care Programme, and thus were not vulnerable to Hawthorne effects or

other investigator biases in collection. Fourth, our analytic approach—local linear regression using a data-driven optimal bandwidth—is a theory-driven and widely used best practice in the conduct of regression discontinuity studies [37,38,40,43–45] and was decided on *a priori*. By using a data-driven bandwidth selector, we eliminated an opportunity for the investigator to manipulate the results by choosing a “preferred” bandwidth. By choosing local linear regression *a priori*, we avoided investigator discretion in the choice of functional form. Finally, following guidelines for the conduct of regression discontinuity studies, we assessed the data for evidence of manipulation of the assignment variable and conducted tests for balance at the threshold on all baseline characteristics observed and available in the dataset. We found no evidence to suggest that patients were dissimilar just above and below the treatment threshold.

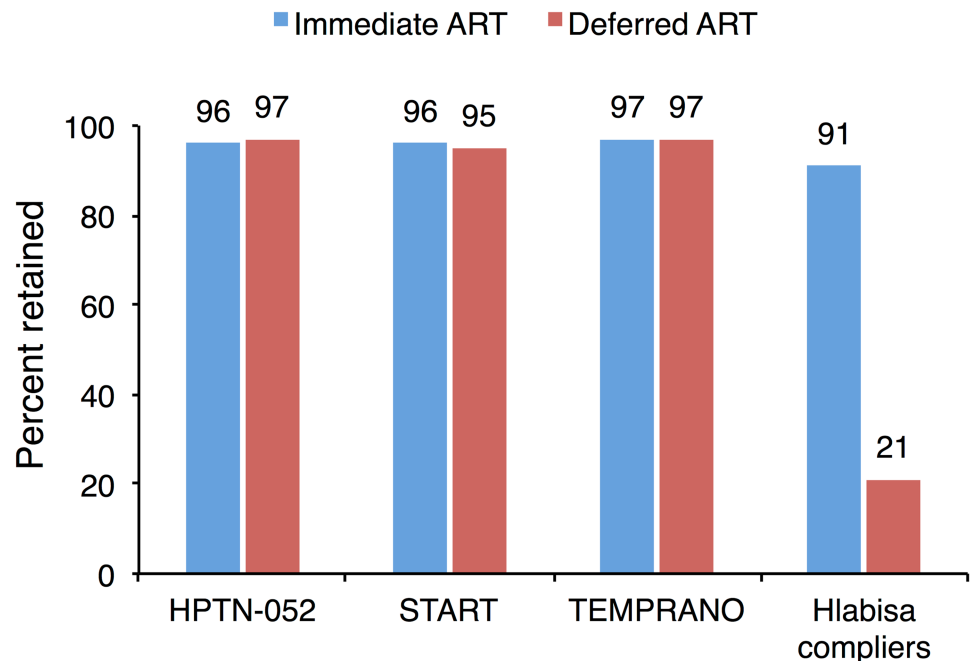
## Limitations

Our study has some limitations. One limitation is that, as discussed above, our local linear regression results may be biased if the relationship between earliest CD4 count and retention in care is nonlinear near the threshold. Nonlinearities were taken into consideration when choosing the bandwidth (the window of data) for the model. Additionally, visual inspection of our figures suggests that in fact the relationships were approximately linear and that any bias resulting from using a linear model would be very small relative to the size of the effect estimates. Our results were also robust to using smaller bandwidths.

A second limitation involves the generalizability of our estimates to patients presenting with different CD4 counts and to different patient populations. As with any regression discontinuity design [26,28,36,41], our results are interpretable as causal effects for patients presenting with CD4 counts close to the 350-cells/ $\mu$ l eligibility threshold. If treatment effects differed across CD4 counts, then our results would not be directly informative of effects at higher (or lower) CD4 counts. Although this is a limitation, it is likely that our estimates are broadly generalizable to other points in the CD4 count distribution. The probability of retention changes little with the value of the patient’s first CD4 count, except at the threshold, and prior analysis showed similar effects at the 200-cells/ $\mu$ l eligibility threshold used prior to August 2011 [42]. Our analysis was also limited to one sub-district of one country, and it is unknown whether the results generalize to other settings.

A third limitation regards the difficulty of measuring retention. There are competing definitions in the literature [65]. In our primary specification, we classified patients as retained if they had any routine contact with the clinic within 6 to <18 months after first CD4 count, including visits to pick up ART medication, laboratory tests (CD4 or viral load), or an ART initiation date, all of which are specified as elements of routine pre-ART or ART care. As a robustness check, we defined an alternate measure of retention based only on laboratory tests and dates of ART initiation. We note that by excluding routine visits, this measure underestimates retention among ART patients and should be interpreted as a lower bound.

A fourth limitation of our analysis is that we were unable to assess longer-term health outcomes that may result from poor retention. In previous analysis, we found large survival benefits of immediate ART eligibility for patients presenting with CD4 counts near the former eligibility threshold of 200 cells/ $\mu$ l [28,42]. However, it is unknown whether these survival benefits extend to patients presenting at higher CD4 counts. We observed significant gaps in retention at 18–24 months. Patients whose CD4 counts are not actively monitored for treatment eligibility may initiate long after their CD4 count falls below the eligibility threshold, or they may not initiate at all. Extended treatment delays have consequences not only for patients themselves but also for population health, with increased potential for onward transmission [66]. Further research will be needed to determine the real-world



**Fig 4. The effect of immediate ART on retention is not observed in clinical trials.** Retention is reported at 2.1 years for the HPTN-052 trial [5], 3 years for the INSIGHT START trial [7], 3.5 years for the TEMPRANO trial [6], and 12 months for Hlabisa. Estimates for Hlabisa are for compliers—those patients whose treatment decision was determined by the value of their CD4 count.

<https://doi.org/10.1371/journal.pmed.1002463.g004>

impacts of deferred ART at higher CD4 counts on long-term engagement with care, health, survival, and onward transmission.

### Implications and next steps

International treatment guidelines are informed (largely) by clinical trials [8,47,58,67], which typically differ from clinical care in the “real world” in important dimensions. One of these dimensions is retention in care. Clinical trials usually seek to retain patients in all treatment arms through systematic monitoring and outreach efforts. For example, in the TEMPRANO trial, 30-month retention was 97% in both arms [6]. The parity of retention across arms in these trials stands in stark contrast to the very large gap—91% versus 21%—we observed in a non-trial setting (Fig 4). Efforts to minimize attrition improve the validity of inferences on clinical endpoints; however, trials cannot then observe the effect of the intervention on retention in care, nor any downstream health impacts that are mediated through retention [5–7,68,69]. In this and other applications, the gap between clinical efficacy (as demonstrated in clinical trials) and real-world effectiveness may turn on the nature of the relationship between the intervention and retention of patients in care—a question of patient behavior, not biology.

Prior observational studies have documented higher retention among ART patients than among pre-ART patients [11,12,23,24]; however, this association is difficult to interpret due to potential selection bias. If highly motivated patients were more likely to initiate ART—leaving a residual of less motivated patients in pre-ART care—then a policy expanding ART eligibility may simply shift patients with low motivation from pre-ART to ART, leading to low retention and poor outcomes for newly eligible patients on treatment. On the other hand, if initiating patients on ART causally increases retention in care, then immediate ART eligibility would improve retention, leading to even larger benefits than observed in clinical trials. Our results

provide evidence, to our knowledge for the first time, to distinguish between these two competing hypotheses. We demonstrate a large difference in retention between pre-ART and ART patients, causally attributable to starting ART itself.

The gap in retention observed in this study would be eliminated if patients were eligible for ART regardless of CD4 count, as under test-and-treat scenarios now being implemented in many countries. We caution that some patients may have little interest in initiating therapy even if eligible [59]. Our results are not informative about the impact of immediate therapy for this group, and increasing demand for ART in such patients may be a challenge. Nor do our estimates generalize to the smaller group of patients who start ART for other reasons (e.g., disease stage) regardless of CD4 count. But for those patients currently barred from initiating due to an ineligible CD4 count, we show that a guideline change allowing immediate initiation could dramatically increase retention in care. Early WHO guidelines for HIV were designed to prioritize the sickest for ART, and there has been concern that expanding eligibility would inappropriately target resources to patients with little incentive to remain on therapy [70]. Our results suggest that expanding eligibility would target patients who are both high need (only 21% would be retained if not eligible) and high performing (91% would be retained if eligible). Countries such as South Africa [9], which have now removed CD4 thresholds, can be encouraged that such a policy will be an efficient step towards expanding HIV treatment coverage.

## Conclusion

Clinical trials have demonstrated the biological efficacy of early ART [5–7]. Effects on patient retention, however, cannot be observed in trials that minimize attrition by design. Our study demonstrates, to our knowledge for the first time, the retention effects of early ART: denying ART eligibility to patients who would be willing to start therapy leads to very large losses from HIV care, losses that would be avoided with immediate ART. Our results thus indicate that the real-world benefits of extending ART eligibility to all patients, regardless of CD4 count, may be larger than previously thought.

## Supporting information

**S1 Appendices. Supporting information on the study methods, supplementary figures, and supplementary tables.** Appendices A and B: supporting information on study methods. Appendix C: supplementary figures. Appendix D: supplementary tables.  
(DOCX)

**S1 Code. Replication code (Stata/SE version 14.2.do file).**  
(DO)

**S1 Data. Description of replication dataset and how to access it.**  
(DOCX)

**S1 STROBE Checklist.**  
(DOC)

## Acknowledgments

We thank the staff of the Africa Health Research Institute and the Hlabisa HIV Treatment and Care Programme for collecting and managing the data. We thank the patients who provided the data analyzed. We thank Anupam Jena and meeting participants at the 2016 Conference on Retroviruses and Opportunistic Infections and the 2016 Epidemiology Congress of the Americas for their feedback on earlier versions of this manuscript.



The contents are the responsibility of the authors and do not necessarily reflect the views of any of the funders or the US government.

## Author Contributions

**Conceptualization:** Jacob Bor, Matthew P. Fox, Sydney Rosen, Till Bärnighausen.

**Data curation:** Frank Tanser, Deenan Pillay, Till Bärnighausen.

**Funding acquisition:** Jacob Bor, Matthew P. Fox, Sydney Rosen, Frank Tanser, Till Bärnighausen.

**Investigation:** Jacob Bor.

**Methodology:** Jacob Bor, Matthew P. Fox, Atheendar Venkataramani, Till Bärnighausen.

**Project administration:** Jacob Bor.

**Resources:** Deenan Pillay.

**Supervision:** Matthew P. Fox, Sydney Rosen, Till Bärnighausen.

**Visualization:** Jacob Bor.

**Writing – original draft:** Jacob Bor.

**Writing – review & editing:** Jacob Bor, Matthew P. Fox, Sydney Rosen, Atheendar Venkataramani, Frank Tanser, Deenan Pillay, Till Bärnighausen.

## References

1. Bor J, Herbst A, Newell M-L, Bärnighausen T. Increases in adult life expectancy in rural South Africa: valuing the scale-up of HIV treatment. *Science*. 2013; 339:961–5. <https://doi.org/10.1126/science.1230413> PMID: 23430655
2. Mills EJ, Bakanda C, Birungi J, Chan K, Ford N, Cooper CL, et al. Life expectancy of persons receiving combination antiretroviral therapy in low-income countries: a cohort analysis from Uganda. *Ann Intern Med*. 2011; 155:209–17. <https://doi.org/10.7326/0003-4819-155-4-201108160-00358> PMID: 21768555
3. Bendavid E, Holmes CB, Bhattacharya J, Miller G. HIV development assistance and adult mortality in Africa. *JAMA*. 2012; 307:2060–7. <https://doi.org/10.1001/jama.2012.2001> PMID: 22665105
4. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380:2095–128. [https://doi.org/10.1016/S0140-6736\(12\)61728-0](https://doi.org/10.1016/S0140-6736(12)61728-0) PMID: 23245604
5. Grinsztejn B, Hosseinipour MC, Ribaudo HJ, Swindells S, Eron J, Chen YQ, et al. Effects of early versus delayed initiation of antiretroviral treatment on clinical outcomes of HIV-1 infection: results from the phase 3 HPTN 052 randomised controlled trial. *Lancet Infect Dis*. 2014; 14:281–90. [https://doi.org/10.1016/S1473-3099\(13\)70692-3](https://doi.org/10.1016/S1473-3099(13)70692-3) PMID: 24602844
6. Danel C, Moh R, Gabillard D, Badje A, Le Carrou J, Ouassa T, et al. A trial of early antiretrovirals and isoniazid preventive therapy in Africa. *N Engl J Med*. 2015; 373:808–22. <https://doi.org/10.1056/NEJMoa1507198> PMID: 26193126
7. Lundgren JD, Babiker AG, Gordin F, Emery S, Grund B, Sharma S, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *N Engl J Med*. 2015; 373:795–807. <https://doi.org/10.1056/NEJMoa1506816> PMID: 26192873
8. World Health Organization. Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV. Geneva: World Health Organization; 2015.
9. Motsoaledi A. Minister Aaron Motsoaledi: Health Dept budget vote 2016/17. Pretoria: South African Department of Health; 2016 [cited 2017 Oct 31]. <http://www.gov.za/speeches/debate-health-budget-vote-national-assembly-10-may-2016-dr-aaron-motsoaledi-minister-health>.
10. Gardner EM, McLees MP, Steiner JF, Del Rio C, Burman WJ. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. *Clin Infect Dis*. 2011; 52:793–800. <https://doi.org/10.1093/cid/ciq243> PMID: 21367734

11. Rosen S, Fox MP. Retention in HIV care between testing and treatment in sub-Saharan Africa: a systematic review. *PLoS Med.* 2011; 8:e1001056. <https://doi.org/10.1371/journal.pmed.1001056> PMID: 21811403
12. Fox MP, Rosen S. Retention of adult patients on antiretroviral therapy in low- and middle-income countries: systematic review and meta-analysis 2008–2013. *J Acquir Immune Defic Syndr.* 2015; 69:98–108. <https://doi.org/10.1097/QAI.0000000000000553> PMID: 25942461
13. Mugglin C, Estill J, Wandeler G, Bender N, Egger M, Gsponer T, et al. Loss to programme between HIV diagnosis and initiation of antiretroviral therapy in sub-Saharan Africa: systematic review and meta-analysis. *Trop Med Int Health* 2012; 17:1509–20. <https://doi.org/10.1111/j.1365-3156.2012.03089.x> PMID: 22994151
14. Kranzer K, Govindasamy D, Ford N, Johnston V, Lawn SD. Quantifying and addressing losses along the continuum of care for people living with HIV infection in sub-Saharan Africa: a systematic review. *J Int AIDS Soc.* 2012; 15:17383. <https://doi.org/10.7448/IAS.15.2.17383> PMID: 23199799
15. Peltzer K, Pengpid S. Socioeconomic factors in adherence to HIV therapy in low- and middle-income countries. *J Health Popul Nutr.* 2013; 31:150–70. PMID: 23930333
16. Hatcher AM, Turan JM, Leslie HH, Kanya LW, Kwena Z, Johnson MO, et al. Predictors of linkage to care following community-based HIV counseling and testing in rural Kenya. *AIDS Behav.* 2012; 16:1295–307. <https://doi.org/10.1007/s10461-011-0065-1> PMID: 22020756
17. Mujugira A, Celum C, Thomas KK, Farquhar C, Mugo N, Katabira E, et al. Delay of antiretroviral therapy initiation is common in east African HIV-infected individuals in serodiscordant partnerships. *J Acquir Immune Defic Syndr.* 2014; 66:436–42. <https://doi.org/10.1097/QAI.0000000000000192> PMID: 24798765
18. Gerdtts SE, Wagenaar BH, Micek MA, Farquhar C, Kariaganis M, Amos J, et al. Linkage to HIV care and antiretroviral therapy by HIV testing service type in Central Mozambique: a retrospective cohort study. *J Acquir Immune Defic Syndr.* 2014; 66:e37–44. <https://doi.org/10.1097/QAI.0000000000000081> PMID: 24326605
19. Hønge BL, Jespersen S, Nordentoft PB, Medina C, da Silva D, da Silva ZJ, et al. Loss to follow-up occurs at all stages in the diagnostic and follow-up period among HIV-infected patients in Guinea-Bissau: a 7-year retrospective cohort study. *BMJ Open.* 2013; 3:e003499. <https://doi.org/10.1136/bmjopen-2013-003499> PMID: 24163204
20. Clouse K, Pettifor AE, Maskew M, Bassett J, Van Rie A, Behets F, et al. Patient retention from HIV diagnosis through one year on antiretroviral therapy at a primary health care clinic in Johannesburg, South Africa. *J Acquir Immune Defic Syndr.* 2013; 62:e39–46. <https://doi.org/10.1097/QAI.0b013e318273ac48> PMID: 23011400
21. Larson BA, Brennan A, McNamara L, Long L, Rosen S, Sanne I, et al. Lost opportunities to complete CD4+ lymphocyte testing among patients who tested positive for HIV in South Africa. *Bull World Health Organ.* 2010; 88:675–80. <https://doi.org/10.2471/BLT.09.068981> PMID: 20865072
22. Hassan AS, Fielding KL, Thuo NM, Nabwera HM, Sanders EJ, Berkley JA. Early loss to follow-up of recently diagnosed HIV-infected adults from routine pre-ART care in a rural district hospital in Kenya: a cohort study. *Trop Med Int Health.* 2012; 17:82–93. <https://doi.org/10.1111/j.1365-3156.2011.02889.x> PMID: 22943164
23. Lessells RJ, Mutevedzi PC, Cooke GS, Newell M-L. Retention in HIV care for individuals not yet eligible for antiretroviral therapy: rural KwaZulu-Natal, South Africa. *J Acquir Immune Defic Syndr.* 2011; 56:e79. <https://doi.org/10.1097/QAI.0b013e3182075ae2> PMID: 21157360
24. Mutevedzi PC, Lessells RJ, Heller T, Bärnighausen T, Cooke G, Newell M-L. Scale-up of a decentralised HIV treatment programme in rural KwaZulu-Natal, South Africa: does rapid expansion affect patient outcomes? *Bull World Health Organ.* 2010; 88:593–600. <https://doi.org/10.2471/BLT.09.069419> PMID: 20680124
25. Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: an alternative to the ex post facto experiment. *J Educ Psychol.* 1960; 51:309–17.
26. O’Keeffe AG, Geneletti S, Baio G, Sharples LD, Nazareth I, Petersen I. Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data. *BMJ.* 2014; 349:g5293. <https://doi.org/10.1136/bmj.g5293> PMID: 25199521
27. Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. *BMJ.* 2016; 352:i1216. <https://doi.org/10.1136/bmj.i1216> PMID: 26977086
28. Bor J, Moscoe E, Mutevedzi P, Newell M-L, Bärnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology* 2014; 25:729–37. <https://doi.org/10.1097/EDE.0000000000000138> PMID: 25061922
29. Hughes MD, Stein DS, Gundacker HM, Valentine FT, Phair JP, Volberding PA. Within-subject variation in CD4 lymphocyte count in asymptomatic human immunodeficiency virus infection: implications for patient monitoring. *J Infect Dis.* 1994; 169:28–36. PMID: 7903975

30. Lee DS. Randomized experiments from non-random selection in U.S. House elections. *J Econom.* 2008; 142:675–97.
31. Houlihan CF, Bland RM, Mutevedzi PC, Lessells RJ, Ndirangu J, Thulare H, et al. Cohort profile: Hlabisa HIV Treatment and Care Programme. *Int J Epidemiol.* 2011; 40:318–26. <https://doi.org/10.1093/ije/dyp402> PMID: 20154009
32. Tanser F, Hosegood V, Bärnighausen T, Herbst K, Nyirenda M, Muhwava W, et al. Cohort profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. *Int J Epidemiol.* 2008; 37:956–62. <https://doi.org/10.1093/ije/dym211> PMID: 17998242
33. Hlabisa Hospital. Memorandum: change to national ART guidelines for adults: CD4 eligibility criteria. Nongoma: Hlabisa Hospital; 2011.
34. Vandenbroucke JP, le Cessie S. Commentary: regression discontinuity design: let's give it a try to evaluate medical and public health interventions. *Epidemiology.* 2014; 25:738–41. <https://doi.org/10.1097/EDE.000000000000145> PMID: 25076150
35. Bor J, Moscoe E, Bärnighausen T. Three approaches to causal inference in regression discontinuity designs. *Epidemiology.* 2015; 26:e28–30. <https://doi.org/10.1097/EDE.0000000000000256> PMID: 25643120
36. Bor J, Moscoe E, Bärnighausen T. The “natural experiment” in regression discontinuity designs: randomization without controlled trials (rapid response). *BMJ.* 2014; 349:g5293. <https://doi.org/10.1136/bmj.g5293>
37. Imbens GW, Lemieux T. Regression discontinuity designs: a guide to practice. *J Econom.* 2008; 142:615–35.
38. Lee DS, Lemieux T. Regression discontinuity designs in economics. *J Econ Lit.* 2010; 48:281–355.
39. Almond D, Doyle JJ, Kowalski AE, Williams H. Estimating marginal returns to medical care: evidence from at-risk newborns. *Q J Econ.* 2010; 125:591–634. PMID: 20634927
40. 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; 68:122–33. <https://doi.org/10.1016/j.jclinepi.2014.06.021> PMID: 25579639
41. Sood N, Bendavid E, Mukherji A, Wagner Z, Nagpal S, Mullen P. Government health insurance for people below poverty line in India: quasi-experimental evaluation of insurance and health outcomes. *BMJ.* 2014; 349:g5114. <https://doi.org/10.1136/bmj.g5114> PMID: 25214509
42. Bor J, Moscoe E, Bärnighausen T. When to start HIV treatment: evidence from a regression discontinuity study in South Africa. Population Association of America 2015 Annual Meeting; 2015 Apr 30–May 2; San Diego, CA. <http://paa2015.princeton.edu/abstracts/151012>
43. Oldenburg CE, Moscoe E, Bärnighausen T. Regression discontinuity for causal effect estimation in epidemiology. *Curr Epidemiol Rep.* 2017; 3(3):233–41. <https://doi.org/10.1007/s40471-016-0080-x> PMID: 284978750
44. McCrary J. Manipulation of the running variable in the regression discontinuity design: a density test. *J Econom.* 2008; 142:698–714.
45. Imbens GW, Kalyanaraman K. Optimal bandwidth choice for the regression discontinuity estimator. *Rev Econ Stud.* 2011; 79:933–59.
46. Hahn J, Todd P, Van der Klaauw W. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica.* 2001; 69:201–9.
47. World Health Organization. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach: 2010 revision. Geneva: World Health Organization; 2010.
48. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica.* 1994; 62:467–75.
49. Imbens GW, Rubin DB. Estimating outcome distributions for compliers in instrumental variables models. *Rev Econ Stud.* 1997; 64:555–74.
50. Tversky A, Kahneman D. Loss aversion in riskless choice: a reference-dependent model. *Q J Econ.* 1991; 106:1039–61.
51. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. *Eur J Soc Psychol.* 2010; 40:998–1009.
52. Bärnighausen T, Chaiyachati K, Chimbindi N, Peoples A, Haberer J, Newell M-L. Interventions to increase antiretroviral adherence in sub-Saharan Africa: a systematic review of evaluation studies. *Lancet Infect Dis.* 2011; 11:942–51. [https://doi.org/10.1016/S1473-3099\(11\)70181-5](https://doi.org/10.1016/S1473-3099(11)70181-5) PMID: 22030332
53. Chaiyachati KH, Ogbuonji O, Price M, Suthar AB, Negussie EK, Bärnighausen T. Interventions to improve adherence to antiretroviral therapy: a rapid systematic review. *AIDS.* 2014; 28(Suppl 2):S187–204. <https://doi.org/10.1097/QAD.0000000000000252> PMID: 24849479

54. Katz IT, Ryu AE, Onuegbu AG, Psaros C, Weiser SD, Bangsberg DR, et al. Impact of HIV-related stigma on treatment adherence: systematic review and meta-synthesis. *J Int AIDS Soc.* 2013; 16(3 Suppl 2):18640. <https://doi.org/10.7448/IAS.16.3.18640> PMID: 24242258
55. Bogart LM, Chetty S, Giddy J, Sypek A, Sticklor L, Walensky RP, et al. Barriers to care among people living with HIV in South Africa: contrasts between patient and healthcare provider perspectives. *AIDS Care.* 2013; 25:843–53. <https://doi.org/10.1080/09540121.2012.729808> PMID: 23061894
56. Lifson AR, Grund B, Gardner EM, Kaplan R, Denning E, Engen N, et al. Improved quality of life with immediate versus deferred initiation of antiretroviral therapy in early asymptomatic HIV infection. *AIDS.* 2017; 31:953–63. <https://doi.org/10.1097/QAD.0000000000001417> PMID: 28121710
57. Govindasamy D, Meghij J, Kebede Negussi E, Clare Baggaley R, Ford N, Kranzer K. Interventions to improve or facilitate linkage to or retention in pre-ART (HIV) care and initiation of ART in low- and middle-income settings—a systematic review. *J Int AIDS Soc.* 2014; 17:19032. <https://doi.org/10.7448/IAS.17.1.19032> PMID: 25095831
58. World Health Organization. Guidelines for managing advanced HIV disease and rapid initiation of anti-retroviral therapy. Geneva: World Health Organization; 2017 [cited 2017 Oct 31]. <http://www.who.int/hiv/pub/guidelines/advanced-HIV-disease/en/>.
59. Bor J, Ahmed S, Fox MP, Rosen S, Meyer-Rath G, Katz IT, et al. Effect of eliminating CD4-count thresholds on HIV treatment initiation in South Africa: an empirical modeling study. *PLoS ONE.* 2017; 12:e0178249. <https://doi.org/10.1371/journal.pone.0178249> PMID: 28617805
60. Bärnighausen T, Tugwell P, Røttingen J-A, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series—paper 4: uses and value. *J Clin Epidemiol.* 2017 Mar 29. <https://doi.org/10.1016/j.jclinepi.2017.03.012> PMID: 28365303
61. Bärnighausen T, Oldenburg C, Tugwell P, Bommer C, Ebert C, Barreto M, et al. Quasi-experimental study designs series—paper 7: assessing the assumptions. *J Clin Epidemiol.* 2017 Mar 29. <https://doi.org/10.1016/j.jclinepi.2017.02.017> PMID: 28365306
62. Gelman A, Imbens G. Why high-order polynomials should not be used in regression discontinuity designs. *J Bus Econ Stat.* 2017 Aug 17. <https://doi.org/10.1080/07350015.2017.1366909>
63. Bor J, Geldsetzer P, Venkataramani A, Bärnighausen T. Quasi-experiments to establish causal effects of HIV care and treatment and to improve the cascade of care. *Curr Opin HIV AIDS.* 2015 Nov; 10(6):495–501. <https://doi.org/10.1097/COH.0000000000000191> PMID: 26476863
64. Westreich D, Pence BW, Turner AN. In populo. *Epidemiology.* 2010; 21:152–3. <https://doi.org/10.1097/EDE.0b013e3181c1e92a> PMID: 20010219
65. Fox MP, Larson B, Rosen S. Defining retention and attrition in pre-antiretroviral HIV care: proposals based on experience in Africa. *Trop Med Int Health.* 2012; 17:1235–44. <https://doi.org/10.1111/j.1365-3156.2012.03055.x> PMID: 22863075
66. Oldenberg C, Bor J, Tanser F, Mutevedzi T, Shahmanesh M, Seage G, et al. Immediate HIV treatment prevents new infections: causal evidence on the real-world impact of immediate versus deferred ART in rural South Africa. 2016 International AIDS Conference; 2016 Jul 18–22; Durban, South Africa.
67. World Health Organization. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach: 2013 revision. Geneva: World Health Organization; 2013.
68. Severe P, Juste MAJ, Ambroise A, Eliacin L, Marchand C, Apollon S, et al. Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. *N Engl J Med.* 2010; 363:257–65. <https://doi.org/10.1056/NEJMoa0910370> PMID: 20647201
69. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med.* 2011; 365:493–505. <https://doi.org/10.1056/NEJMoa1105243> PMID: 21767103
70. World Health Organization. Scaling up antiretroviral therapy in resource-limited settings: treatment guidelines for a public health approach. Geneva: World Health Organization; 2003.