PHDC engagement evaluation PROTOCOL

Title

Evaluating engagement with HIV care for people on ART in a high-burden, low-resource setting in South Africa.

Core study team

Dr Claire Keene (co-principal investigator), Dr Tamsin Phillips, Dr Aaloke Mody, Jonathan Euvrard, Dr Jacob McKnight, Prof Mike English and Prof Catherine Orrell (co-principal investigator)

Background and rationale

Engagement with HIV care is an important concept

HIV is now considered a chronic disease, which changes the dynamics of patient engagement with treatment[1,2]. Engagement is a complex, multi-dimensional, dynamic process comprising interaction with the health system and treatment-taking behaviour[3,4], comprising the dimensions of retention, adherence and active self-management. Engagement is increasingly recognised as critical to achieving and maintaining treatment success, making its measurement a priority to support individual patient management, strategic decisions about services and programmatic strategies to control the epidemic[5,6]. However, evaluations are not comprehensive, and measures are not standardised, undermining the health system's efforts to respond to the evolving epidemic[7–9].

Longitudinal and latent class methods of engagement analysis

There is a trend towards valuing the dynamics of engagement more than single time-point measures, which reduce this complex concept to a cross-sectional summary and obscure the heterogeneity between patients and within individuals over time[10,11], introduce systematic errors and bias, and as a proxy for consistent engagement fail to identify the pattern of care or gaps where the health service could intervene to improve engagement overall[12]. The dynamic nature of engagement over time is important to understand how engagement facilitates sustained treatment success for individual management and the understanding of intervention success or failure[10]. Patterns of engagement may be more important than engagement status at a specified point: treatment interruptions are associated with poor outcomes[13,14] and have proved more predictive of outcomes than average or point adherence[15].

A more complex understanding of engagement calls for more sophisticated measures[15]. Considering the dimensions of engagement as time-varying covariates in a survival analysis allows a more nuanced exploration of the impact of engagement on treatment outcomes, which reflects the realities of people's changing engagement over time[16,17]. However, survival analysis only evaluates time to one (Kaplan-Meier) or the first of multiple (competing risks) outcomes[18]. Multistate analysis can help characterise the longitudinal course of engagement with HIV care, accounting for people 'churning' in and out of the system[16–18]. It captures the shifts in the proportion of the ART population in different states over time and the timing of these transitions[18]. This could help programmes plan service delivery, as multistate approaches can reveal the care gaps at distinct time points that undermine treatment success, and the factors associated with them in order to identify targets for intervention.

Longitudinal trajectories of single or multiple dimensions of engagement can also be explored using group-based trajectory analysis to uncover different subgroups of engagement trajectories[19]. Latent class methods are applicable when "there is a desire for an individualised approach, whilst at the same time a need to understand something general about a patient population"[20]. Traditional variable-centred regression draws conclusions based on the relationships between variables, so does not capture the complex interactions between biological, psychological and social factors[20], which are relevant when evaluating engagement behaviour. Person-centred techniques like group-based trajectory modelling (a longitudinal application of latent class analysis[19]) could reveal underlying, qualitatively different, categorical subgroups (the unobserved heterogeneity[21]) of ways individuals engage with HIV care, based on their observable patterns of retention and adherence over time[22].

While the subgroups are not literal distinct entities, they help to approximate the more complex reality of HIV care engagement, with each group trajectory providing a summary of the cluster of individuals grouped together because they have similar trajectories[19]. Distinct, categorical subgroups of engagement trajectories would be useful in practice to support the development of services and interventions targeted at different subgroups of patients and their specific behavioural patterns. This offers a novel opportunity to direct interventions to "behavioural phenotypes" distinct from the demographic categories we traditionally use to target differentiated services, often with poor success[10].

Segmentation of the patient population is helpful to tailor support programmes and provide a basis for resource allocation and the development of interventions[23]. It also shifts service delivery to an approach between a 'one size fits all' and a fully individualised programme. This supports the notion of 'precision public health', which is characterised by optimising care strategies for well-characterised population strata[24,25]. Differentiated service delivery has already begun to respond to the differing needs of people who are stable on treatment, however this does not encompass the whole population of people on ART[26]. The engagement subgroups identified with group-based trajectory modelling represent patient typologies that may help to understand how people engage with care in a cyclical manner, 'churning' in and out of care over time. This could help understand 'patient profiles' of different patterns of engagement, and so develop interventions tailored to behavioural patterns as well as the traditional demographic categories[27,28], and ensure that routine services are relevant to respond to each subgroup's needs.

Using multiple methods provides different viewpoints on the longitudinal dynamics of engagement behaviour, and together could help to broaden the understanding of engagement behaviour over time and the factors that influence it. Better understanding of the complex concept of engagement coupled with more nuanced routine monitoring could support more robust strategic decision-making.

Routine data in health system monitoring and research

Longitudinal cohorts are more complex as they require adjustments and planning to enable follow up, such as unique patient identifiers, tracking of patients over time and linking of data from different sources to these individuals[29]. Routine data may provide a ready cohort of patients for such analyses, and is in great demand for health systems to answer questions with large, person-level datasets and to support services to become 'learning systems'[30]. Advantages over primary data include reduced burden of data collection, wider coverage, longitudinal data points and relatively low levels of self-report bias[29,30]. However, routine health information systems are made up of a patchwork of multiple databases[31] and are susceptible to surveillance bias and other confounding[30]. Routine data's primary purpose is to support service delivery, thus it can be biased in answering research questions as routine data leaves out aspects that are significant to patients or that influence their outcomes[30]. Longitudinal analysis can be difficult as routine data is vulnerable to missing information[32] and following individual patients over time is one of the main challenges to

this approach[12]. In addition, many health systems do not have the capacity to generate and use the information from routine data to change practice[33].

The limitations of routine data may be mitigated by using more sophisticated analysis methodology, linking databases (healthcare and other governmental, social and commercial agencies) or supplementing the 'big' dataset with 'small data' purposively collected in order to create a more comprehensive picture[30].

Evaluation of engagement in care in Khayelitsha and Gugulethu

Sub-Saharan Africa has the highest rates of HIV in the world[34], contributing to a double burden of infectious and non-communicable diseases that places pressure on governments and worsens dependency on donor funding[35]. Despite routinely better rates of retention and adherence than patients in high income settings[36,37], PLWH in sub-Saharan Africa face particular challenges engaging with care[38], and health systems are restricted in their ability to monitor engagement[39].

Using group-based trajectory modelling, Mody et al found six distinct trajectories that described how PLWH engaged with ART in Zambia: 1. Consistently high adherence and retention (28.5% of the sample), 2. Early non-adherence with consistent retention (22.2%), 3. Gradually decreasing adherence and retention (21.6%), 4. Early loss to follow-up and reengagement (8.6%), 5. Early loss to follow-up (8.7%), and 6. Late loss to follow-up (10.4%)[10]. There subgroups were found to be associated with mortality and to have a stronger association with treatment outcomes than patient demographics[10].

Khayelitsha and Gugulethu are high HIV-prevalence settlements in South Africa where communities face many challenges common in urban sub-Saharan Africa[40,41]. However, quality routine data are available[42], so provide a unique opportunity to gain insight into the dynamics of engagement and its measurement in a lower-resourced, high burden setting. This analysis will test the theory that the groups identified by Mody et al are generalisable and explore engagement and its associations in this setting using multiple methods of longitudinal analyses.

Combining longitudinal analysis of multiple dimensions and identifying the latent subgroups of engagement from these indicator variables, could be a relevant application of routine data to provide clearer insight into engagement behaviour, its dynamics and the factors associated with optimal or poor engagement. This study will evaluate the utility of routine data in longitudinal evaluation of engagement with HIV care at individual and programmatic levels.

Research aims and objectives

The aim is to explore engagement behaviour through the patterns of measurable, longitudinal indicators, in order to gain insight into

- Engagement behaviour and its dynamics in this population
- Patient phenotypes based on behavioural patterns as well as demographics
- The impact of longitudinal measurement and latent class methods on inferences from routine data

The objectives of the study are:

1. To examine the association of retention and adherence with HIV-specific treatment outcomes in Khayelitsha and Gugulethu using multiple methods, including survival analysis, group-based trajectory modelling and multi-state analysis.

   a. Explore the impact of the duration of trajectories, pivotal events and past trajectories on the future patterns of behaviour and risk of poor treatment outcomes.
   b. Evaluate the difference in outputs using different measures of adherence and retention and identify an optimal model based on statistical and theoretical criteria
2. To identify subgroups of patients that engage with the system in different ways in Khayelitsha and Gugulethu based on their patterns of retention and adherence over time, and profile the subgroups to evaluate their association with patient factors (demographics, clinical and health system factors).
3. To evaluate the validity of a model of engagement with care developed by Mody et al in a Zambian cohort [10] by replicating the analysis in another setting
4. To develop a predictive model for virologic failure based on patterns of behaviour, health system, clinical and patient characteristics

Methodology

Study design

This is a retrospective, quantitative, longitudinal analysis of routine data.

Study setting

This analysis will focus on the lower-resource settings found in Khayelitsha and Gugulethu in South Africa. Both are characterised by high HIV prevalence against a backdrop of high rates of non-communicable disease, poverty, unemployment, violence and high mobility for work, which makes engagement particularly challenging[40,41]. Both areas have large, mature cohorts on antiretroviral therapy[43].

This setting will be compared to the whole Cape Town Metro (of which Khayelitsha and Gugulethu are a part) in order to understand the nuance of engagement in a lower resource setting.

Data

Provincial Health Data Centre (PHDC)

The PHDC is a Western Cape Department of Health routine health-data repository that consolidates and curates patient-level information from routine data systems into one centralised database intended to improve patient care at the clinician and health-system level (including planning and management), as well as support epidemiological understanding and evaluation through exploration of research questions using an exported dataset. The PHDC data consists of linked, de-identified, individual patient information rather than aggregate totals, making patient level, longitudinal and cross-sectional analyses possible with continuity across services, facilities and time[42].

The database covers the public health system (healthcare in South Africa is provided in two systems: the well-resourced private sector and the public system, which functions as a single service provider where health services are provided free of charge depending on income). Over five million people (75% of the population) in the Western Cape access public healthcare, and most patients accessing care for conditions such as HIV and tuberculosis use the public sector[42]. Electronic patient records are not used, but data is regularly captured into an electronic database by data clerks at all facilities. Across the 52 hospitals and 354 primary care clinics in the Western Cape, the PHDC estimated 15 million encounters with patients per year, and eight million patients active between 2009 and 2019 (of 17 million ever registered on the public system)[42]. The request for data will also include a request to link the eligible cohort to the national mortality data.

Limitations include that the data is only collated from Western Cape systems, and particularly for Khayelitsha patients who tend to travel to and from the neighbouring Eastern Cape province, the PHDC data misses data when they seek care in other provinces.

Sample and eligibility

The sample size for the routine data will include all eligible PLWH on ART in Khayelitsha and Gugulethu. This is expected to be approximately 50 000 patients on ART in Khayelitsha and 35 000 in Gugulethu. The eligibility for inclusion in the analysis is outlined in Table 1.

Table 1. Eligibility for inclusion in the analysis

Eligibility criterion	Justification
Relationship to ART: people living with HIV who have ever initiated ART	Focus on the engagement behaviour dynamics with ART
Age: ≥15 years old when they were initiated	Identify a cohort exposed to adult regimens and treatment engagement dynamics (e.g. no reliance on a parent or other caregiver for access to care)
Period of evaluation: patients who initiated ART between 1st January 2013 and 31st December 2020 will be included. The analysis will include patient time up to the 30th September 2021 to allow at least nine months of follow up for each individual, and the database will be closed on 31st December 2021 to allow for capturing and evaluation of outcomes in the dataset.	By the 1st January 2013 the services were largely similar to the current offering: all first line patients were on single dose regimens (introduced in March 2012[44]), ART was initiated and managed by nurses (NIMART was in policy by 2013[45]), and clubs were an option (rolled out in the Western Cape by the end of 2011[46] with 16 000 stable ART patients in over 600 clubs receiving care through adherence clubs in the Cape Metro by December 2012[2]). System changes during this period will be considered in the interpretation of the results, increasing numbers on ART for the same number of facilities, increasing number of patients cared for in differentiated service delivery options, the introduction of the more tolerable ARV dolutegravir for first and second-line patients, an increasing focus on providing better services for those struggling to engage and a change in the eligibility criteria to initiate ART.

Variables

The variables for analysis include variables that reflect engagement, treatment outcome variables and influential factors (Table 2).

Table 2. List of variables from the Provincial Health Data Centre's routine dataset

	ENGAGEMENT INDICATORS	**TREATMENT OUTCOMES**	**INFLUENTIAL FACTORS**
Variable explanation	The retention and adherence variables are the indicator variables that will be used to identify the subgroups of the latent engagement variable.	The outcomes include virologic failure (two consecutive VL ≥1000 copies/mL), sustained virologic suppression (two consecutive VL <400 copies/mL and <50 copies/mL), death, immunological failure (CD4 after six months on treatment ≤CD4 at initiation) and loss to follow up (no interaction with the health system for >90 and >180 days).	The influential factors are demographic, clinical and service factors that will be evaluated for association with subgroup membership. Pivotal events are events that are identifiable in the data, and which could potentially change the person's engagement with care, or the way that the health system interacts with the individual. For example, events like pregnancy and a diagnosis of tuberculosis both have the potential to change a person's motivation to engage, but also trigger more intensive follow up from the health system. Other potentially pivotal events include hospitalisation, a change in ART regimen, an episode of tuberculosis or chronic disease diagnosis.
PHDC routine data	• Dates of interaction with the health system: visit constancy	• Viral load (dates and values) • Death outcome and date • CD4 (dates and values) • Dates of interaction with the health system	**Demographics** • Date of birth • Sex **Clinical characteristics** • CD4 at initiation (value and date) **Health service factors**

	and visit gaps
	• Pharmacy refill dates and number pills dispensed: medication possession ratio since last visit
		• Date and location of diagnosis
		• Regimen at initiation and subsequent switches
		• Duration on ART and year of initiation (date of initiation)
		• Family planning prescribed
		• Chronic medication prescribed (hypertension, diabetes)
		• Pivotal events
		  ○ Pregnancy episode and date
		  ○ Tuberculosis diagnosis and date
		  ○ Mental health diagnosis and date
		  ○ Hospitalisation (ward and date)
		• Dates of allied health interactions (including dietician, psychology, counsellor, social worker, physiotherapist, community care workers and mental health)

Analysis

The analysis will be conducted in R software (R Foundation for Statistical Computing, Vienna, Austria) and may make use of the MPlus software for more complicated analyses. Categorical variables will be described using counts and proportions. Continuous variables will be described using means and standard deviation (normally distributed) or medians and interquartile ranges (non-normally distributed).

Longitudinal analysis will be conducted using

1. Survival analysis (survival to HIV specific treatment outcomes including virologic failure, mortality and loss to follow-up) with a Cox regression model or non-parametric survival methods as appropriate, as well as a competing risk approach. Retention and adherence will be included as time-varying covariates
2. Multi-state analysis to describe and quantify the transition of patients through mutually exclusive and collectively exhaustive states of retention and adherence over time.
3. Group-based trajectory modelling to identify subgroups that follow distinct longitudinal patterns of engagement over time, using retention and adherence measures as the indicator variables. These will be modelled both individually and with multi-trajectory analysis simultaneously to identify subgroups that follow joint trajectories[10,19].

Different combinations of retention, adherence and treatment outcome measures (Table 3) will be evaluated to identify an optimal model to describe engagement behaviour over time, based on statistical and theoretical criteria.

Table 3. Measures of retention and adherence for the analyses

Adherence	Retention	Summary of engagement process
Medication possession ratio *= proportion of time that a patient has medication on hand as a proxy for the maximum possible adherence*[47] Refill gaps *= medication-free days where patient could not have medication in hand and it would be impossible for a patient to have taken medication*[48] Maintenance of adherence	Visit gaps *= 90 days without a visit*[32] *(commonly used and clinically relevant) where the patient returns to care after the gap*[50]. Visit gaps *= 180 days without a visit*[32] *(strong association with outcomes) where the patient returns to care after the gap*[50].	Viral load suppression <50 copies/mL *= a summary measure of engagement demonstrating a combination of retention (has a viral load conducted) and adherence (viral load is suppressed <50 copies/mL)* Viral load suppression <400 copies/mL

= presence or absence of '≥2 consecutive individual refills with <95% adherence'[49] Pharmacy composite *= measure of attendance (on time or late pick up) and coverage (MPR)*	Fixed point retention *= interaction with an HIV service at a specified point in time, with a window of 90 days (WHO definition)*[51] Visit constancy *= reflects the presence of ≥1 HIV visit in each 90 day*[36,52] *and 180 day*[32,53] *interval* Late for visit *= interaction with an HIV service >14 days after the expected date*[52] *(calculated from the last pharmacy refill date and quantity dispensed)* Combination of fixed point and timing of attendance *= point retention and ≥75% appointments met on time (± 14 days) during follow up)*[54]	*= a summary measure of engagement demonstrating a combination of retention (has a viral load conducted) and adherence (viral load is suppressed <400 copies/mL)* Viral load suppression <1000 copies/mL *= a summary measure of engagement demonstrating a combination of retention (has a viral load conducted) and adherence (viral load is suppressed <1000 copies/mL)*

The proposed methods for each objective are described in Supplementary material Table 1.

Ethical considerations

Confidentiality and data safety

The data management will comply with both the United Kingdom General Data Protection Regulation (UK GDPR)[55], the South African Health Act[56] and the South African Protection of Personal Information (POPI) Act[57]. The primary author has undertaken a research integrity course and an ethics workshop (including data management concerns) provided by the University of Oxford and a good clinical practice course conducted by Clinical Research Education and Development (CREDE) in Cape Town.

Once ethics approval has been obtained, a formal application to the PHDC will be completed for a de-identified dataset of populations that meet the described eligibility criteria. Only the minimum data to answer the research question will be applied for and no primary data will be collected. This is routinely collected data and the dataset will be de-identified by the PHDC before the core study team has access to it. This application will constitute a data sharing agreement between the PHDC and the research team named in this protocol.

The data will be transferred to the primary researcher, located in Oxford, which is governed by the UK GDPR. Other researchers in the core analysis team are located at the University of Cape Town, which is governed by the South African POPI Act and Washington University, which is governed by a number of US Federal data protection laws. Data will be shared using One Drive file sharing, restricting access to the core research team. This follows the cross-border data sharing and information flow regulations in the POPI Act[58].

During analysis, the dataset will be accessible only via a password protected laptop with whole disk encryption, and only accessible by the core study team. The dataset will be encrypted with a password and stored in a folder in the primary author's university-provided One-Drive account, accessible only with a single sign on through the University of Oxford. The laptop used to access the dataset through the One-Drive account for analysis will have the University of Oxford Sophos Endpoint Detection and Response (anti-virus) product installed, complete regular back-ups through

university-supplied Hierarchical File Server (HFS) Hub Backup and Archive Service, and meet a minimum level of IT hygiene as prescribed by the University of Oxford.

Post-analysis, data will be stored for five years after the publication of the work in an encrypted folder in a One-Drive account in order to support corrections. After this time it will be deleted. Any requests for data access in the interest of open access research will be referred to the PHDC to make a formal application for the data.

Potential benefits to participants

There are no immediate, direct benefits to participants, but the findings of this study could inform better service delivery options and facilitate the development of interventions responsive to their needs.

Potential risks to participants

The study poses minimal risk to participants. There is very little chance that participants could be identified from triangulation of information, considering the minimal identifying information (date of birth, sex and access to care in Khayelitsha, Gugulethu or the Cape Town Metro) and the large sample size. The link to identifiers is held by the PHDC and will not be released to the research team for this study. The results will not place any individual, family or community at social, psychological or economic risk.

Consent

As this is a secondary analysis of de-identified routine data, no individual patient consent will be sought. However once ethical approval is received, provincial Department of Health approval will be obtained and PHDC approval for use of the dataset will be applied for.

Dissemination of results

The study team will make use of a range of forums to disseminate results, including manuscripts submitted to peer-reviewed publications and local and international scientific conferences. The outcomes of the study will be shared with the Department of Health and local health authorities to inform policy and practice.

Resources

Collaboration

The project will draw on expertise from the research team including researchers from the University of Oxford, University of Cape Town and Washington University.

Funding

The time of the primary author is funded through the Clarendon Foundation as part of a PhD in the Nuffield Department of Medicine at the University of Oxford.

Timeline

Table 4 gives an overview of the expected milestones

Table 4. Gantt chart with expected timeline

	2022												2023											
	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
Ethics and provincial approval																								
Application for data																								
Data analysis																								
Write-up																								
Dissemination																								

References

1. Angwenyi V, Aantjes C, Kajumi M, De Man J, Criel B, Bunders-Aelen J. Patients experiences of self-management and strategies for dealing with chronic conditions in rural Malawi. *PloS one*. 2018;13(7):e0199977. doi:10.1371/journal.pone.0199977

2. Swendeman D, Ingram BL, Rotheram-Borus MJ. Common elements in self-management of HIV and other chronic illnesses: an integrative framework. *AIDS Care*. 2009;21(10):1321-1334. doi:10.1080/09540120902803158.Common

3. Mallinson RK, Relf M V., Dekker D, Dolan K, Darcy A, Ford A. Maintaining normalcy: A grounded theory of engaging in HIV-oriented primary medical care. *Advances in Nursing Science*. 2005;28(3):265-277. doi:10.1097/00012272-200507000-00008

4. Osler M, Hilderbrand K, Goemaere E, et al. The Continuing Burden of Advanced HIV Disease over 10 Years of Increasing Antiretroviral Therapy Coverage in South Africa. *Clinical Infectious Diseases*. 2018;66(Figure 1):S118-S125. doi:10.1093/cid/cix1140

5. Mugavero MJ, Davila JA, Nevin CR, Giordano TP. From access to engagement: Measuring retention in outpatient HIV clinical care. *AIDS Patient Care and STDs*. 2010;24(10):607-613. doi:10.1089/apc.2010.0086

6. Wachira J, Middlestadt S, Reece M, Peng CYJ, Braitstein P. Physician communication behaviors from the perspective of adult HIV patients in Kenya. *International journal for quality in health care : journal of the International Society for Quality in Health Care*. 2014;26(2):190-197. doi:10.1093/intqhc/mzu004

7. Mugavero MJ, Westfall AO, Zinski A, et al. Measuring Retention in HIV Care: The Elusive Gold Standard. *J Acquir Immune Defic Syndr*. 2012;61(5):574-580. doi:10.1097/QAI.0b013e318273762f.Measuring

8. Craker L, Tarantino N, Whiteley L, Larry Brown. Measuring Antiretroviral Adherence among Young People Living with HIV: Observations from a Real-Time Monitoring Device versus Self-Report. *AIDS Behaviour*. 2019;23(8):2138-2145. doi:10.1007/s10461-019-02448-1.Measuring

9. Bright FAS, Kayes NM, Worrall L, McPherson KM. A conceptual review of engagement in healthcare and rehabilitation. *Disability and Rehabilitation*. 2015;37(8):643-654. doi:10.3109/09638288.2014.933899

10. Mody A, Eshun-Wilson I, Sikombe K, et al. Longitudinal engagement trajectories and risk of death among new ART starters in Zambia: A group-based multi-trajectory analysis. *PLoS Medicine*. 2019;16(10):1-25. doi:10.1371/journal.pmed.1002959

11. Gachara G, Mavhandu LG, Rogawski ET, Manhaeve C, Bessong PO. Evaluating Adherence to Antiretroviral Therapy Using Pharmacy Refill Records in a Rural Treatment Site in South Africa. *AIDS research and treatment*. 2017;2017:5456219. doi:10.1155/2017/5456219

12. Rollins NC, Becquet R, Orne-Gliemann J, et al. Defining and analyzing retention-in-care among pregnant and breastfeeding HIV-infected women: Unpacking the data to interpret and improve PMTCT outcomes. *Journal of Acquired Immune Deficiency Syndromes*. 2014;67:S150-S156. doi:10.1097/QAI.0000000000000355

13. Meresse M, March L, Kouanfack C, et al. Patterns of adherence to antiretroviral therapy and HIV drug resistance over time in the Stratall ANRS 12110/ESTHER trial in Cameroon. *HIV medicine*. 2014;15(8):478-487. doi:10.1111/hiv.12140

14. Parienti JJ, Massari V, Descamps D, et al. Predictors of virologic failure and resistance in HIV-infected patients treated with nevirapine- or efavirenz-based antiretroviral therapy. *Clinical Infectious Diseases*. 2004;38(9):1311-1316. doi:10.1086/383572

15. Eby J, Chapman J, Marukutira T, et al. The adherence-outcome relationship is not altered by diary-driven adjustments of microelectronic monitor data. *Pharmacoepidemiology and drug safety*. 2015;24(12):1313-1320. doi:10.1002/pds.3887

16. Ehrenkranz P, Rosen S, Boulle A, et al. The revolving door of HIV care: Revising the service delivery cascade to achieve the UNAIDS 95-95-95 goals. *PLoS Medicine*. 2021;18(5):1-10. doi:10.1371/journal.pmed.1003651

17. Kaplan S. Contemporary disengagement from antiretroviral therapy in Khayelitsha, South Africa. *Dissertation Abstracts International: Section B: The Sciences and Engineering*. 2018;78(11-B(E)):No-Specified. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc15&NEWS=N&AN=2017-36665-066

18. Mody A, Tram KH, Glidden D v., et al. Novel Longitudinal Methods for Assessing Retention in Care: a Synthetic Review. *Current HIV/AIDS Reports*. 2021;18(4):299-308. doi:10.1007/s11904-021-00561-2

19. Nagin DS, Odgers CL. Group-Based Trajectory Modeling in Clinical Research. *Annual Review of Clinical Psychology*. 2010;6:109-138. doi:10.1146/annurev.clinpsy.121208.131413

20. Kongsted A, Nielsen AM. An introduction to Latent Class Analysis in health research. *Journal of Physiotherapy*. 2016;63(1):55-58. doi:10.1016/j.jphys.2016.05.018

21. Weller BE, Bowen NK, Faubert SJ. Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology*. 2020;46(4):287-311. doi:10.1177/0095798420930932

22. Law EH, Harrington R. A Primer on Latent Class Analysis. *Values and Outcomes Spotlight*. 2016;(December):18-19.

23. Bloem S, Stalpers J, Groenland EAG, Van Montfort K, Van Raaij WF, De Rooij K. Segmentation of health-care consumers: Psychological determinants of subjective health and other person-related variables. *BMC Health Services Research*. 2020;20(1):1-12. doi:10.1186/s12913-020-05560-4

24. Elkind Velmovitsky P, Bevilacqua T, Alencar P, Cowne D, Pelegrini Moria P. Convergence of Precision Medicine and Public Health Into Precision Public Health: Towards a Big Data Perspective. *Frontiers in Public Health*. 2021;9:561873.

25. Dolley S. Big Data's Role in Precision Public Health. *Frontiers in Public Health*. 2018;6(68).

26. Barker C, Dutta A, Klein K. Can differentiated care models solve the crisis in HIV treatment financing? Analysis of prospects for 38 countries in sub-Saharan Africa. *Journal of the International AIDS Society*. 2017;20(Supplement 4):A4. doi:http://dx.doi.org/10.7448/IAS.20.5.21648

27. Nosetti L, Paglietti MG, Brunetti L, et al. Application of latent class analysis in assessing the awareness , attitude , practice and satisfaction of paediatricians on sleep disorder management in children in Italy. *PlosONE*. 2020;1:1-17.

28. Liu Z, Zhang Y, Tian L, Sun B, Chang Q, Zhao Y. Application of latent class analysis in assessing the competency of physicians in China. *BMC Medical Education*. 2017;17(208):1-8. doi:10.1186/s12909-017-1039-4

29. World Health Organization. *Consolidated HIV Strategic Information Guidelines: Driving Impact through Programme Monitoring and Management*.; 2020.

30. Deeny SR, Steventon A. Making sense of the shadows : priorities for creating a learning healthcare system based on routinely collected data. *BMJ Quality and Safety*. 2015;24:505-515. doi:10.1136/bmjqs-2015-004278

31. Gitelman L. *" Raw Data " Is an Oxymoron*. The MIT Press; 2013.

32. Phillips TK, Orrell C, Brittain K, Zerbe A, Abrams EJ, Myer L. Measuring retention in HIV care: the impact of data sources and definitions using routine data. *AIDS (London, England)*. 2020;34(5):749-759. doi:10.1097/QAD.0000000000002478

33. Sheikh K, Abimbola S. *Learning Health Systems: Pathways to Progress*.; 2021.

34. The World Health Organization. WHO Africa: HIV/AIDS. Published 2020. Accessed February 12, 2021. https://www.afro.who.int/health-topics/hivaids

35. Angwenyi V, Bunders-Aelen J, Criel B, Lazarus J V, Aantjes C. An evaluation of self-management outcomes among chronic care patients in community home-based care programmes in rural Malawi: A 12-month follow-up study. *Health & social care in the community*. Published online July 2020. doi:10.1111/hsc.13094

36. Poles G, Li M, Siril H, et al. Factors associated with different patterns of nonadherence to HIV care in Dar es Salaam, Tanzania. *Journal of the International Association of Providers of AIDS Care*. 2014;13(1):78-84. doi:10.1177/1545109712467068

37. Mills EJ, Nachega JB, Buchan I, et al. Adherence to antiretroviral therapy in sub-Saharan Africa and North America: A meta-analysis. *Journal of the American Medical Association*. 2006;296(6):679-690. doi:10.1001/jama.296.6.679

38. Haas AD, Zaniewski E, Anderegg N, et al. Retention and mortality on antiretroviral therapy in sub-Saharan Africa: collaborative analyses of HIV treatment programmes. *Journal of the International AIDS Society*. 2018;21(2). doi:10.1002/jia2.25084

39. Tique JA, Howard LM, Gaveta S, et al. Measuring Health Literacy Among Adults with HIV Infection in Mozambique: Development and Validation of the HIV Literacy Test. *AIDS and Behavior*. 2017;21(3):822-832. doi:10.1007/s10461-016-1348-3

40. Groenewald P, Bradshaw D, Daniels J, et al. Local-level mortality surveillance in resource-limited settings: a case study of Cape Town highlights disparities in health. *Bulletin of the World Health Organisation*. 2010;88(6):444–451.

41. Strategic Development Information and GIS Department (SDI&GIS). 2011 Census Suburb Profiles. Published online 2013.

42. Boulle A, Heekes A, Tiffin N, et al. Data Centre Profile : The Provincial Health Data Centre of the Western Cape. *International Journal of Population Data Science*. 2019;4(2):06. doi:https://doi.org/10.23889/ijpds.v4i2.1143

43. Stinson K, Goemaere E, Coetzee D, et al. Cohort Profile: The Khayelitsha antiretroviral programme, Cape Town, South Africa. *International Journal of Epidemiology*. 2017;21:e21. doi:10.1093/ije/dyw057

44. Mbengashe TD. Circular: Changes in regime for HIV positive pregnant women and note on those with a psychiatric illness, 19 March 2012. Published online 2012:2.

45. Mouton J, Africa F. Circular 169/2013: NIMART Guidelines. Published online 2013.

46. Wilkinson LS. ART adherence clubs: A long-term retention strategy for clinically stable patients receiving antiretroviral therapy. *Southern African Journal of HIV Medicine*. 2013;14(2):48-50. doi:10.7196/SAJHIVMED.924

47. Mekuria LA, Prins JM, Yalew AW, Sprangers MAG, Nieuwkerk PT. Which adherence measure - self-report, clinician recorded or pharmacy refill - is best able to predict detectable viral load in a public ART programme without routine plasma viral load monitoring? *Tropical medicine & international health : TM & IH*. 2016;21(7):856-869. doi:10.1111/tmi.12709

48. Orrell C, Cohen K, Leisegang R, Bangsberg DR, Wood R, Maartens G. Comparison of six methods to estimate adherence in an ART-naïve cohort in a resource-poor setting: which best predicts virological and resistance outcomes? *AIDS Research & Therapy*. 2017;14(1):20. doi:10.1186/s12981-017-0138-y

49. Genn L, Chapman J, Okatch H, et al. Pharmacy Refill Data are Poor Predictors of Virologic Treatment Outcomes in Adolescents with HIV in Botswana. *AIDS and Behavior*. 2019;23(8):2130-2137. doi:http://dx.doi.org/10.1007/s10461-018-2325-9

50. Nawar E. Longitudinal analysis of interruptions in HIV care and treatment among HIV-positive pregnant women engaged in clinical care in Kigali, Rwanda. *Dissertation Abstracts International Section A: Humanities and Social Sciences*. 2020;81(7-A):No-Specified.

51. World Health Organization. *HIV Test–Treat–Retain Cascade Analaysis: Guide and Tools*.; 2017.

52. Font H, Rollins N, Essajee S, et al. Retention-in-care in the PMTCT cascade: definitions matter! Analyses from the INSPIRE projects in Malawi, Nigeria and Zimbabwe. *Journal of the International AIDS Society*. 2020;23(10):e25609. doi:10.1002/jia2.25609

53. Agaba PA, Genberg BL, Sagay AS, et al. Retention in Differentiated Care: Multiple Measures Analysis for a Decentralized HIV Care and Treatment Program in North Central Nigeria. *Journal of AIDS & clinical research*. 2018;9(2). doi:10.4172/2155-6113.1000756

54. Ahoua L, Arikawa S, Tiendrebeogo T, et al. Measuring retention in care for HIV-positive pregnant women in Prevention of Mother-to-Child Transmission of HIV (PMTCT) option B+ programs: the Mozambique experience. *BMC public health*. 2020;20(1):322. doi:10.1186/s12889-020-8406-5

55. Information Commissioner's Office. *Guide to the General Data Protection Regulation (GDPR)*.; 2021.

56. South African Parliament. *South African National Health Act*.; 2003.

57. South African Parliament. *Protection of Personal Information Act 4*.; 2013.

58. Adams R, Adeleke F, Anderson D, et al. *POPIA Code of Conduct for Research*. Vol 117.; 2021.

59. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation*. 2016;133(6):601-609. doi:10.1161/CIRCULATIONAHA.115.017719

60. Dignam JJ, Zhang Q, Kocherginsky M. The use and interpretation of competing risks regression models. *Clinical Cancer Research*. 2012;18(8):2301-2308. doi:10.1158/1078-0432.CCR-11-2097

61. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *American Journal of Epidemiology*. 2009;170(2):244-256. doi:10.1093/aje/kwp107

62. Nylund KL, Asparouhov T, Muthén BO, et al. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling : A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal*. 2007;14(4):535-569. doi:10.1080/10705510701575396

63. Mody A, Lyons PG, Guillamet CV, et al. The Clinical Course of Coronavirus Disease 2019 in a US Hospital System : A Multistate Analysis. *American Journal of Epidemiology*. 2021;00(00):1-14. doi:10.1093/aje/kwaa286

64. Titman AC, Sharples LD. Model diagnostics for multi-state models. *Statistical Methods in Medical Research*. 2010;19:621-651.

Supplementary material

Table 1. Analytic methods for each objective

Objective	*Analytic methods*		
	Survival analysis	Group-based trajectory modelling	Multi-state analysis
To examine the association of retention and adherence with HIV-specific treatment outcomes in Khayelitsha and Gugulethu using multiple methods, including survival analysis, group-based trajectory modelling and multi-state analysis. • *Explore the impact of the duration of trajectories, pivotal events and past trajectories on the future patterns of behaviour and risk of poor treatment outcomes.* • *Evaluate the difference in outputs using different measures of adherence and retention and identify an optimal model based on statistical and theoretical criteria*	Survival analysis will be conducted from ART initiation as time zero, to HIV specific treatment outcomes including virologic failure, mortality and loss to follow-up. This will be done using a Cox regression model or non-parametric survival methods as appropriate, as well as a competing risk approach. Each patient will contribute a duration of follow-up from their ART initiation until the closure of the database. Individuals will be right censored for all analyses if they are transferred out of the province and at the closure of the database. Competing events include negative treatment outcomes: - Virologic failure - Mortality - Loss to follow-up Competing events will be right-censored in the traditional survival analysis, creating a cause-specific hazard model for each event (while censoring the other events). They will be treated as competing risks for the competing risks approach. Retention and adherence will be included as time-varying covariates to evaluate their impact on survival time to negative treatment outcomes. This will be explored using a cause-specific hazards and a subdistribution hazards model (to adjust for competing events), adjusting for other	Group-based trajectory modelling will be used to identify subgroups of individuals that follow distinct longitudinal patterns of engagement over time, using retention and adherence measures as the indicator variables. The observed indicators will be used to estimate (with maximum likelihood estimation[19]) both the trajectories for the subgroups (modelled as a function of time using flexible polynomials[10]) and the subgroup distribution in the population that produces the best fit for the data[19]. The indicators will be modelled individually and with multi-trajectory analysis simultaneously to identify subgroups that follow joint trajectories[10,19] Measures of engagement will be repeated at 30-day, 90-day and 180 day intervals from time zero (ART initiation) to the point of right censorship (death, transfer to a new facility out of the province or the end of the observation period). Individuals will not be censored if they are missing data or 'lost to follow up': missing engagement indicator values will be considered to indicate non-retention, with the acknowledged limitations that there may be some missing at random or missing due to health system error (such as healthcare worker omission of a scheduled VL or lack of capturing). Individuals will remain under observation until they meet one of the censoring criteria. Those who are censored in the first 180 days will be excluded from analysis to ensure a minimum observation period to develop meaningful engagement trajectories.	A multi-state model will be used to describe and quantify the transition of patients through expected states of retention and adherence over time. The analyses will be conducted with initiation of ART as time zero, as well as by calendar year (with 1st January 2013 as time zero). Individuals will be categorised into mutually exclusive and collectively exhaustive states of engagement[63]. Care states will include: - Currently retained and adherent - Currently retained and adherent with a history of suboptimal engagement - Retained but not fully adherent - Adherent but not fully retained - Not retained and not adherent <90 days and <180 days (intermittent engagement) - Not retained, not adherent >90 days and >180 days - Death Non-parametric multistate analytical techniques (based on the Aalen Johnson method) will be applied to probability of being in each state at a given moment since time zero, and the movements between states[63]. Left censoring will be applied at initiation of ART. Individuals will be right censored if they are transferred out of the province and at the closure of the database. Each patient will contribute a duration of follow-up from their ART initiation until the closure of the database. Local scores tests

patient characteristics in a multivariate analysis.

Cause-specific hazard ratios are the marginal rate of occurrence [59] in a hypothetical world where no competing events exist [60], so describes pure aetiological associations [61]. Subdistribution hazards reflect the real-world incidence of the event, taking into account the impact of competing events [60], and predict expected individual risk [61]. Subdistribution hazards are valuable for the planning of service delivery, the evaluation of treatment policy and for clinical or policy decision-making in a population where competing outcomes must be accounted for [59].

Each participant will contribute a duration of follow up from their ART initiation until the closure of the database. Tracing studies have been conducted that provide weights that can be applied to calculate the proportion 'out of care' who are likely to have died or actually be in care after a 'silent' transfer. However, these studies trace patients from unlinked datasets, whereas the PHDC has already partially followed up patients out of care by triangulating different sources of interaction with the system that account somewhat for transfer between public facilities, and include mortality data that (while not perfect) accounts for those who have died. PLWH who have transferred out of the province or into private healthcare may be missed, but using weights from the literature would overestimate death and transfer and bias the analysis. If no tracing studies are published by the time of analysis that account for the partial tracing of the PHDC, no weights will be used.

The probabilities of individuals belonging to a specific trajectory group given their observed patterns of retention and adherence will be estimated. These posterior probabilities will be based on the observed data up until the time of being censored (to account for varying durations of follow up) and used to assign individuals to the group they most likely belong to[10].
The aim of the model is to "summarise the distinctive features of the data in the most parsimonious – and useful – fashion possible"[19].
Models with a sequentially larger number of subgroups (neighbouring class models[62]), will be evaluated[19]. As the order of the trajectory polynomial is not known a priori but must be prespecified when estimating a model, a series of models with sequentially smaller orders will be tested[10].

test the assumption of time homogeneity for a particular transition intensity near to a particular time point[64]. Pearson-type goodness-of-fit tests will also be conducted[64].

The association of treatment outcomes with each state will be assessed.
Separate analyses stratifying by various states (time zero being the entry date into that state, e.g. entry into 'engaged with a history of poor engagement'), pivotal events (time zero being the date of the pivotal event with comparison of the analysis before and after), composite states (e.g. engaged and engaged with a history of poor engagement) will be conducted.

Each model will be individually well identified, and the final latent class model (polynomial order and number of subgroups) will be selected using:

1. Theoretical considerations to test the model adequacy to answer the research questions[19,21]: evaluate whether the groups are interpretable and make substantive sense, with comparison to groups found previously[10] and the InCARE framework
2. Relative model fit statistics: Bayesian Information Criterion and the bootstrapped likelihood ratio test [BLRT][19,62]
3. Described with classification diagnostics: latent class posterior probabilities (with comparison of the proportion assigned to each group to the estimated group membership, estimation of the average posterior probability of individuals assigned to each group using the maximum probability assignment rule and the odds ratio of being assigned to the correct trajectory group using posterior probabilities compared to population-level group distribution[10]), entropy (separation between groups), number of members in subgroups and face validity of qualitative subgroup differences[20,21].

Multinomial logistic regression will be used to evaluate the association of group membership with treatment outcomes. Pivotal events will be included as time-varying covariates to evaluate their impact on trajectories within a group.

In a subset of patients with at least three years of follow-up, engagement trajectories from the first two years from ART initiation (as time zero) will be compared with longer trajectories to evaluate the association of earlier behaviour on later

		engagement patterns and treatment outcomes. For PLWH who have five years of follow up, the association of cumulative trajectories (one year, two years, three years, four years and five years before an assessment of outcomes) with treatment outcomes will be compared to explore the impact of duration of follow-up on the predictive power of the model.	
To identify subgroups of patients that engage with the system in different ways in Khayelitsha and Gugulethu based on their patterns of retention and adherence over time, and profile the subgroups to evaluate their association with patient factors (demographics, clinical and health system factors).	N/A	Based on the posterior probability group assignments, the baseline demographics, clinical characteristics and health system factors will be described for each identified engagement subgroup and stratified analyses will be conducted across these factors to evaluate the generalisability of the behaviour patterns[10]. Potential confounders, colliders and interactions between the variables available in the PHDC data will be identified through the process of a literature review and the development of directed acyclic graphs of causal relationships between the variables, following which the associations of the identified subgroups with predictive variables will be evaluated using multinomial logistic regression The Bolck–Croon–Hagenaars (BCH) approach for associations with predictor variables will also be used.	Alluvial[63] and Sankey diagrams will be used to visualise the transitions between states over time. Median duration of stay within each state will be estimated. The probability of transition from each state to another will be calculated and compared[63]. The instantaneous rate of entry into each state and transition intensities (instantaneous rate of transition to the next immediate state) will be calculated, and peaks of transfer to particular states identified[63]. The proportion in each state at some point of their treatment journey and at particular time points since ART initiation (6 months, 1 year, 2 years, 5 years and 9 years) will be presented. All analyses will be stratified by patient and health system characteristics, with independent exposure variables and confounders identified through drawing directed acyclic graphs of a priori hypotheses of causal relationships. The cumulative incidence of each state over time and Cox proportional hazards analysis will be used to identify influential factors that are associated with transition to each

			state (proportional hazards assumption to be checked using Schoenfeld residuals).
To evaluate the validity of a model of engagement with care developed by Mody et al in a Zambian cohort [10] *by replicating the analysis in another setting*	N/A	The routine data model will be developed using the Khayelitsha data and replicated using the Gugulethu data. The models from these two populations will be compared to those identified by Mody et al in Zambia[10] and the rest of the Cape Town Metro to evaluate the generalisability of identified subgroups. The subgroups identified by the Zambian study will also be applied to this dataset. Posterior probabilities and associations with treatment outcomes will be calculated and compared to those calculated from developing the model anew	N/A
To develop a predictive model for virologic failure based on patterns of behaviour, health system, clinical and patient characteristics	Using the outputs from the previous two objectives, a model will be selected to describe the association of engagement with treatment outcomes. A combination of pattern of behaviour/group membership and patient characteristics will be used to develop a model to predict virologic failure, using logistic regression or other appropriate techniques. This will be developed in a training set and validated in the remaining data. Model performance will be assessed using the area under the curve. Goodness of fit will also be assessed.		