Longitudinal latent class and joint modelling of antiretroviral adherence

Author:
Campbell McDuling

Student Number:
MCDCAM001

Supervisors:
Francesca Little

Catherine Orrell

May, 2023

Plagiarism Declaration:

1. I know that plagiarism is a serious form of academic dishonesty.

2. I have read the document about avoiding plagiarism, am familiar with its contents and have avoided all forms of plagiarism mentioned there.

3. Where I have used the words of others, I have indicated this by the use of quotation marks.

4. I have referenced all quotations and other ideas borrowed from others.

5. I have not and shall not allow others to plagiarise my work.

Signed by candidate

Signature: _____

Contents

1 Abstract

This dissertation investigates the relationship between antiretroviral therapy (ART) adherence and viral outcomes in people living with HIV in South Africa using advanced statistical modeling techniques. Utilizing data from the ADD-ART study, a prospective cohort of 238 adults on ART in Cape Town, the research employs survival analysis, joint modeling, and longitudinal latent class analysis to compare different adherence monitoring tools and examine heterogeneity in adherence behaviors.

Key findings include: Electronic Adherence Monitoring (EAM) and tenofovir diphosphate levels in dried blood spots were more strongly associated with viral non-suppression than self-reported adherence; joint modeling revealed a stronger association between EAM adherence and viral outcomes compared to traditional survival models, with each additional missed dose in the preceding 30 days associated with an 81% increase in the hazard of viral non-suppression; longitudinal latent class analysis identified five distinct adherence trajectory groups, with poorer or declining adherence groups experiencing significantly higher rates of viral non-suppression; baseline viral load and prior tuberculosis exposure were significant predictors of subsequent viral non-suppression, even after accounting for adherence.

The results highlight the importance of using objective adherence measures, the value of advanced statistical techniques in HIV research, and the need for personalized adherence support strategies. Limitations include potential violations of model assumptions and generalizability constraints.

2 Introduction

Sub-optimal adherence to antiretroviral treatment (ART) is a major barrier in the global fight against the HIV/AIDS epidemic. In South Africa, a high disease burden and resource limited country, improving adherence is a pressing research imperative. The relationship between adherence, viral outcomes and viral resistance is well-known - with better adherence resulting in enhanced viral suppression and a lower risk of developing resistance for the individual, and lower rates of transmission between individuals [90] [69]. Optimal monitoring of adherence is thus a crucial step in identifying poorly adherent individuals and providing them with targeted intervention. Understanding adherence behavioural patterns may provide further knowledge for the design and implementation of these interventions to help improve sub-optimal adherence.

While direct monitoring of adherence is not feasible at scale, different proxy measures to estimate adherence have been proposed. Widely-used candidate proxies for adherence include self-reporting (SR), pharmacy refill monitoring (PRM), pill counts, electronic adherence monitoring (EAM), and drug concentrations in blood plasma or dry blood spots (DBS). However, these tools are associated with viral outcomes to different degrees and through different mechanisms. Analysis of these associations can be approached with varying degrees of complexity. In some cases, taking a simpler approach - such as smoothing over longitudinal data or analysing longitudinal and survival outcomes separately - may be preferred. However, more advanced methods which make full use of the available information can lead to more accurate estimates which reduce the chances of: (a) incorrectly rejecting null hypotheses (Type I error); and (b) failing to detect truly significant effects (Type II error). Estimating these associations in an efficient way that preserves Type I and Type II error can help to more reliably uncover how each measurement type relates to true adherence. Once validated as a reliable proxy, adherence monitoring tools facilitate the analysis of adherence behaviour.

This study proposes leveraging survival analysis and joint modelling to evaluate the association between the time to viral non-suppression and longitudinal adherence data from select monitoring tools. In this dissertation, the joint modeling techniques used involve the harmonized modeling of longitudinal profiles and survival outcomes. Longitudinal latent class analysis methods can then be applied to explore heterogeneous patterns within a sample of longitudinal adherence profiles. Moreover, characterizing subgroups exhibiting distinct adherence behaviors could yield valuable insights into context-specific determinants and consequences of adherence dynamics.

2.1 Background in HIV/AIDS and antiretroviral therapy adherence

HIV/AIDS remains a major public health challenge in South Africa, which has one of the highest HIV prevalence rates globally. By the end of 2018, of an estimated 37.9 million people living with HIV (PLWH) globally, around 7.7 million were living in South Africa [26].

ART, the primary treatment for HIV/AIDS, works by suppressing the replication of the virus in the body. Effective ART allows the immune system to recover, and prevents disease progression to AIDS and the risk of transmission [7]. South Africa initiated a public sector ART program in 2004 to provide treatment through government health facilities, initially limiting eligibility to those with CD4 counts below 200 cells/μL [7]. In September 2016, South Africa adopted the 'Universal Test and Treat' (UTT) policy, making ART accessible to all HIV-positive individuals [7]. This in an attempt to align with the UNAIDS 90-90-90 targets of having 90% of people living with HIV knowing their status, 90% of those diagnosed on ART, and 90% of those on ART virally suppressed by 2020 [26]. As of 2019, the data suggest that these targets had not been fully realized: an estimated 90% of the population knew their status, 68% of which were on treatment, with 87% of those on treatment achieving viral suppression [26].

The standard public sector first-line ART regimen in South Africa has evolved significantly in response to technological advancement, and research into therapeutic efficacy, safety, medication adherence, and viral resistance. Initially consisting of stavudine or zidovudine along with lamivudine and efavirenz or nevirapine, it transitioned in 2010 to tenofovir as the preferred nucleoside reverse transcriptase inhibitor backbone [13]. Most recently in 2019, dolutegravir, an integrase inhibitor, replaced efavirenz due to its higher genetic barrier to resistance, lower toxicity, and longer half-life allowing less frequent and more flexible dosing [7] [95]. While

there is ongoing research into long-acting injectable ART [9], in practice options are currently still limited to oral dosing - usually administered on a once-daily basis. The necessity of frequent oral dosing places more emphasis on the importance of good adherence and more opportunity for poor adherence to be observed.

ART inhibits the reverse transcription, integration and assembly of the virus' replication cycle - repressing an infected persons viral load. This mechanism can be disrupted, possibly through the complex interaction of individual-specific pharmacokinetics/pharmacodynamics, or if the virus evolves resistance against the treatment. Figure 1 provides an illustration of the means through which viral non-suppression can arise. The role of pharmacokinetics/dynamics is not very well understood, since treatment outcomes are usually positive with good adherence and no resistance. It is suggested, however, that treatment efficacy in this respect can be impacted by the host's genetics [92] [88] [96] [77] and/or the concurrent use of certain medications [32] [60]. Drug resistance, on the other hand, is understood to allow the virus to replicate more easily within the host. Selection for drug resistance can occur prior to treatment or be acquired during treatment [16]. Pre-treatment resistance can arise if an individual has been exposed to the medication previously, for example through recreational purposes, or has been infected by a strain of the virus that has already gained resistance in the population. Acquired resistance, on the other hand, is a likely eventuality of poor adherence [16].

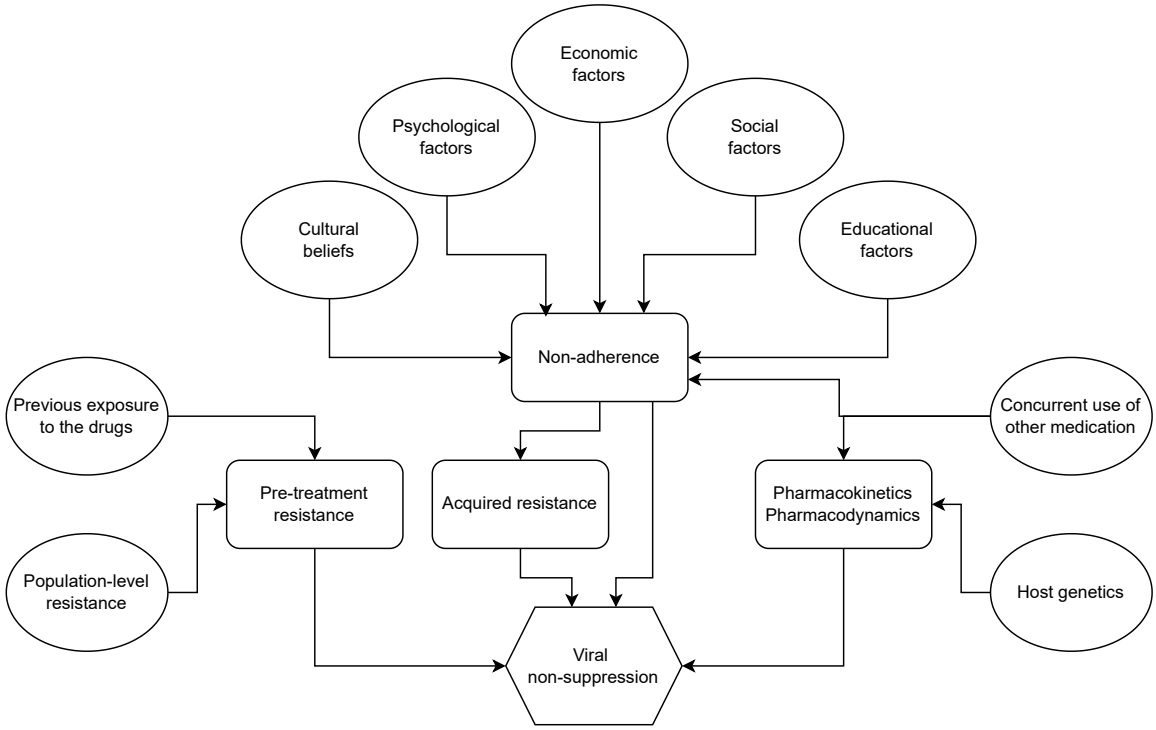


Figure 1: The mechanisms through which viral non-suppression can arise (adapted from [16])

Despite medication adherence being a nuanced and highly context-specific issue, several factors have been identified as capable of contributing to poor ART adherence in South Africa. These include: cultural and belief factors such as beliefs about the efficacy of ART; psychological factors such as depression, substance abuse, stress and forgetfulness; economic factors like financial and transportation constraints; social factors such as HIV stigma and fear of disclosure; and educational factors such as knowledge of the interplay between the virus, ART and adherence; among other aspects like drug toxicity and side-effects, and higher level health system policies [7]. Under conditions of good medication adherence and no viral resistance, ART is thus highly effective at suppressing the viral load below detectable levels. Suboptimal adherence therefore remains

a major barrier to successful treatment outcomes for both the individual and the collective.

While South Africa has made much progress in its public ART rollout, suboptimal adherence remains a significant obstacle, driven by various behavioral, socioeconomic, and cultural factors. Continuing to break down these barriers through enhanced adherence monitoring and counseling programs, education, and tailored strategies for high-risk groups is crucial to fully realizing the potential benefits of widespread ART access. To inform these strategies effectively, it is essential that research on ART adherence is underpinned by reliable measurement techniques. Only through reliably measured adherence can researchers deconstruct the nuances behind adherence-related factors and provide insights into their contribution to South Africa's ongoing HIV/AIDS burden.

2.2 ART adherence monitoring

Several methods have been used to measure adherence to ART, each with it's own limitations in capturing true adherence behavior. Possibly the simplest of these - self-report (SR) surveys - ask patients to recall their medication-taking behavior over a specified time period [107]. For example, a quantitative self-report tool may ask individuals to report how many doses they missed in the preceding 30 days. While inexpensive and easy to implement, self-reports tend to overestimate adherence due to recall and social desirability biases [90].

Another commonly used monitoring mechanism is pharmacy refill monitoring (PRM). This method tracks timely pharmacy refills as a proxy for adherence, assuming collected medications are taken as prescribed [33]. A few metrics can be obtained from PRM data - including medication-free days and average number of pills received per day monitored. PRM is very useful in observational studies, especially when implemented at scale, since the data is relatively accessible and inexpensive to implement. It provides an objective measure of medication possession but does not capture actual dosing behaviour [33].

Another method, electronic adherence monitoring (EM/EAM) has gained traction in recent times. Electronic monitoring devices, such as pill bottles with microchips, record the date and time of bottle openings over an observation window [29]. This provides granular longitudinal data on medication access patterns but cannot confirm pill ingestion. EAM may also cause bias: due to curiosity openings - a term used to describe situations where many 'openings' are recorded that do not correspond to medication access (eg: a child may get hold of the device and play with the lid); if individual's medication access does not correspond to ingestion (eg: an individual opens the device once and removes a week's worth of medication); or by altering adherence behavior due to patient awareness of monitoring [55].

A more accurate but more costly and invasive technique is to measure drug levels from biological samples, providing objective evidence of recent medication ingestion. The Urine Tenofovir Rapid Assay (UTRA) is used to detect the presence of drug metabolites in urine samples. These have a very short half-life and thus the test only indicates recent medication intake on the order of a handful of days. Drug levels can also be measured through blood plasma samples, or dried blood spots. Drug levels measured in plasma samples reflects adherence in the last 4-5 days [36]. That measured through dried blood spots (TFV-DP in DBS) has a much longer half-life, measuring cumulative drug levels in red blood cells, and reflects long-term adherence on the order of 6-8 weeks [10]. When compared to the former two, the latter method is less susceptible to 'white-coat dosing' - a phenomenon where individuals tend to increase their adherence shortly before their clinic visits [33]. Drug concentration measurements thus provide objective evidence of recent drug intake but can be biased by individual differences in pharmacokinetics [69]. It is thus ideal to adjust analyses by relevant potential confounders, which depend on the measurement method, such as body mass index (BMI) or creatinine levels [69]. There is however also a lead time involved in processing biological samples. Therefore, while providing an objective measure of adherence, these may not be preferred over methods like EAM which enable continuous and immediate monitoring.

Each of these methods offers a different perspective on adherence behavior. Combining multiple measures can provide a more comprehensive assessment of true adherence patterns, informing both clinical care and research efforts to improve ART outcomes [33].

2.3 Research Objectives and Roadmap

The precise and accurate monitoring of adherence is a valuable tool for healthcare workers in South Africa, since it allows the rapid identification of non-adherent individuals which - depending on the monitoring method - may allow intervention before viral outcomes deteriorate. This dissertation uses data from an observational clinic-based adherence study conducted in the Western Cape province of South Africa. This study, which is described in detail in Section 5, collected data on longitudinal virologic and adherence outcomes, including different adherence monitoring mechanisms. Descriptive statistics are presented in Section 6.2 which detail the socio-demographic composition at study entry of this cohort of 250 virally-suppressed people living with HIV.

This dissertation thus primarily investigates the use of survival analysis and joint modelling to estimate the association between viral outcomes and adherence as measured by different monitoring tools - self-report surveys; monthly EAM data; and TFV-DP from DBS. This is achieved by conducting a thorough exploratory analysis in Section 6; and implementing survival and longitudinal modelling techniques in Sections 7.1 - 7.3 to build valid sub-models which can be integrated into the joint modelling framework in Section 7.4. To demonstrate the authors command of this suit of statistical tools, a thorough study of the statistical methods underlying survival and longitudinal analyses, and joint modelling is presented in Section 4.

Secondly, the dissertation explores the use of longitudinal latent class modelling to examine unexplained heterogeneity in longitudinal EAM data. This is achieved by using longitudinal finite mixture models to cluster individuals into groups with similar longitudinal adherence profiles. These methods are detailed in Section 4.4, and implemented in Section 7.5.

Section 8 then provides a comprehensive discussion on the key results and limitations resulting from the analyses, before the conclusion is presented in Section 9.

3 Literature review

The literature review that follows focuses on the statistical components of the research. It is organized into two sections that relate to the two main research questions of this dissertation and the methodology used to address them. The first section explores existing literature involving statistical modelling of the association between monitoring mechanisms for antiretroviral adherence and viral outcomes. The second section aims to outline published research regarding statistical approaches to modelling heterogeneous longitudinal adherence behaviours. This will ultimately help to contextualize the relevance of the research and analysis presented in this dissertation.

The quality of the primary studies was evaluated through the relevance of their research question, statistical methodology, population, sampling, intervention, data collection and/or data analysis. Evidence was organized together by grouping primary studies based on their relevance to the two research questions, the study population and the statistical methods used. The time of publication was restricted to an eleven year window from the beginning of 2013 to the end of 2023, except in the case of very notable earlier work. An exhaustive search was conducted via UCT's Primo database, Google Scholar, and PubMed. Two tools powered by artificial intelligence algorithms - Semantic Scholar and Connected Papers - were then used to search for peripherally related publications.

3.1 Modelling the association between adherence and viral outcomes

It is no new notion that adherence to antiretroviral medication is crucial in achieving and maintaining low viral loads. Direct monitoring of adherence is not feasible at scale and thus several studies have explored the relationship between a single proxy for adherence and viral outcomes, or clinical outcomes related to viral outcomes such as CD4 counts.

One early prospective, observational study conducted by Paterson et. al. [70] revealed a statistically significant association between increased adherence, measured by a microelectronic monitoring system, and both improved viral outcomes and CD4 counts. The full text was not recoverable, however, and thus the statistical methods used to assess this association could not be determined.

A study in 2001 by Bangsberg et. al. [4] assessed adherence by monitoring pill counts and categorizing the mean proportion of pills taken over the study window into three bins ($< 50\%$; $50\% - 90\%$; $> 90\%$). The study then used Kaplan-Meier analysis with a log-rank test to investigate differences between the adherence groups' time to progression-free-survival (as defined by the event of a CD4 count below 200 cells per microliter or development of opportunistic infection). This showed a significant difference between the adherence groups and the progression to AIDS. The researchers also assessed the relative risk of developing AIDS by using the same survival outcome in Cox Proportional Hazards models, regressed against the mean proportion of pills taken and independently adjusted for a number of hypothesized baseline covariates. This revealed a significant reduction in the risk of progression to AIDS with a 10% increase in mean adherence, along with associations with baseline CD4 count, viral load and proportion of months virally suppressed at baseline. These methods however did not make use of the full available information since the longitudinal pill counts were collapsed by taking the mean, and both of the aforementioned studies were conducted among a very different population in more developed countries.

In a two-arm RCT, Kiwuwa-Muyingo et. al. [47] used pill counts to measure adherence among participants in Uganda and Zimbabwe, and employed a discrete time Cox proportional hazards model to assess the association between adherence and long-term mortality. The results indicated that poor adherence was significantly associated with increased odds of mortality in both study arms when compared to those with optimal adherence.

Nachega et. al. [62] used a similar statistical approach as in [4] - the baseline-covariate adjusted Cox Proportional Hazards model - to measure the association between average adherence over the study period and all-cause mortality in a cohort of HIV-infected South Africans. In this case, adherence was measured by the proportion of pharmacy claims collected over the study window (pharmacy refill data), and categorized into 20% quantiles. The study found a significant increase in the relative hazard of mortality between each

successive adherence quantile, but again did not make full use of the longitudinal information provided by the adherence measures.

A study conducted in Ethiopia in 2017 by Seyoum et. al. [87] made a step in this direction. The authors showed that the use of joint modelling approaches can provide more powerful estimation of association between CD4 count and adherence (as measured by pill counts), while better preserving the Type I error. The authors however considered both outcomes as longitudinal, specifying the joint distribution appropriately. Although there was no survival component in their analysis, the authors' approach still allowed them to measure a significant association between the longitudinal pill count adherence and change in CD4 cell counts. Another paper was published using the same data which implements a joint model of the association between CD4 counts and time-to-death of HIV/TB co-infected patients, but this analysis did not include any direct measures of adherence [93].

Mchunu et. al. [56] also used joint modelling to estimate the association between CD4 count and risk of death among 642 HIV/TB co-infected individuals in a RCT in South Africa - but without any adherence component. Luvanda et. al. [52] similarly did not incorporate adherence into an joint model that measured the association between viral load and survival time, using secondary data from a retrospective cohort of Tanzanian PLWH. A study that is peripherally related to this paper by context, but closely related by statistical methods is that of Spreafico and Ieva [91]. This used pharmacy refill data as a longitudinal adherence outcome and specified a joint model to measure the association between this adherence proxy and time-to-death among patients who had experienced heart failure between 2000 and 2012.

The publications mentioned above considered relatively imprecise measures of adherence, namely pill counts and pharmacy refill data. There is evidence that other proxies for adherence may provide a better estimation of true adherence, and thus be more closely associated with viral outcomes [90] [54]. Electronic adherence monitoring (EAM) devices have shown noteworthy promise in this respect. A 2016 publication by Evans et. al. [19], measured the association between adherence from EAM data and elevated viral load over a cohort of 49 PLWH in Johannesburg who were failing second-line ART. This study used a log-binomial regression model, with viral suppression as a binary outcome variable, and categorized mean adherence as a covariate. The findings suggested a significantly higher relative risk of viral non-suppression for poorly adherent individuals (adherence $\leq 80\%$) compared to adherent individuals. This statistical analysis again did not utilize the longitudinal nature of the adherence data. Petersen et. al. [71] used ensemble machine learning methods to detect viremea using EAM, clinical and demographic data. While these methods, by design, do not explicitly measure any conditional association, and are purely predictive in nature, the study did find that the model's predictive accuracy was significantly improved when incorporating real-time EAM data in addition to clinical and demographic data.

Most recently, tenofovir diphosphate (TFV-DP) levels in dried blood spots (DBS) has been proposed as a reliable measure of cumulative ART adherence. Castillo-Mancilla et. al. [10] used a simple one-way ANOVA model to show that adherence, as approximated by TFV-DP levels, was significantly higher among subjects who did not develop viral failure than those who developed viral failure without resistance among a cohort of 288 PLWH in Kwa-Zulu Natal, South Africa. Similarly, Jennings et. al. [37] estimated the odds of viral failure at the next clinic visit for PLWH in the Western Cape with TFV-DP < 400 fmol/punch when compared to subjects with > 800 fmol/punch. This revealed significantly higher relative odds of viral failure for those with lower adherence, however this analysis did not make use of the viral failure survival time nor full use of the longitudinal nature of this adherence measure.

Of the above studies, none compared the adherence-viral outcomes association between multiple adherence proxies in the same population. Gill et. al. [25] in 2010 used a logistic regression to estimate the association between undetectable viral load (a binary outcome indicated by ¡200 *copies/ml*) and categorized mean adherence over a 6 month period. To measure adherence, the study team considered self-report along a visual analog scale, pill counts, EAM and EAM with dose-timing - which additionally required a dose event to occur within a 1hr window of a pre-specified time. They found that EAM was more strongly associated with viral suppression than self-report or pill counts, and additionally found that EAM with dose-timing was more strongly associated than without dose-timing.

A 2017 study by Orrell et. al. [69] compared the ability of six adherence monitoring mechanisms (SR; clinic-based pill count; two measures using PR data; efavirenz drug concentrations in blood plasma; and EAM) to predict viral failure and viral resistance at around 4 months and 12 months following the participants' entry into care. For the analysis, a simple logistic regression was used to estimate the probability of viral failure/resistance, independently at the two time points, as a function of average adherence. Results indicated that electronic adherence and pharmacy refill based methods were most predictive of viral outcomes and viral resistance around the 12 month time point, while self-reported adherence and efavirenz drug monitoring were both not associated with the outcomes. These methods did not make effective use of the longitudinal nature of the adherence and viral outcomes.

Jennings et. al. [38] conducted a similar study called ADD-ART which concluded in 2019. In this study, viral outcomes and adherence (through EAM and TFV-DBS) were observed among a cohort of 250 virally suppressed PLWH in four primary healthcare clinics in Cape Town at regular intervals over the course of 24 months. The analysis used multiple logistic regression models with viral breakthrough (VB) as a binary outcome variable, adjusted for adherence and known confounders. As measures of adherence, the team independently considered: EAM counts at 1, 2 and 3 months prior to VB; and TFV-DP levels at 0, 1 and 2 months prior to VB. Therefore, only cross-sectional adherence data was used - neglecting the longitudinal information. The analysis revealed strong associations with VB for all adherence estimates, with the associations being marginally stronger for TFV-DP than EAM at all comparable time points.

In a different approach, Smith et. al. [89] conducted a systematic review and meta-analysis to evaluate the accuracy of different adherence monitoring tools at detecting viral non-suppression. The results showed significant variability in both sensitivity and specificity of the different adherence measures, with very low-certainty associated with all methods. The authors' analyses did however show that the sensitivity and specificity for the studies that used EAM to detect viral non-suppression was higher than for any other method.

3.2 Modeling longitudinal adherence behaviour

This section explores relatively complex approaches to modeling heterogeneous adherence behaviors, as well as studies that have modeled mean adherence trajectories over time using methods such as generalised linear mixed effects models and generalized estimating equations. These approaches provide valuable insights into overall adherence trends while accounting for the longitudinal nature of the data. In a study of HIV-positive youth in the United States, Naar-King et al. [61] applied repeated measures ANOVA model - essentially a random effects model - to analyze temporal adherence differences in terms of self-efficacy following a motivational intervention [61]. The results did not show any significant improvement to self-efficacy over the 24-month period. Though this study modelled longitudinal adherence (as self-efficacy), it did not attempt to find any variables which could explain heterogeneity in the adherence outcome.

Orrell et. al. [68] used a generalised linear model to investigate predictors of average cumulative EAM data and treatment interruptions among a South African cohort of PLWH. The results indicated that daily text message reminders did not have any significant effect on overall adherence but did reduce the frequency of treatment interruptions longer than 72 hours, and no significant associations between adherence and any of the following were found: gender, non-disclosure, anxiety, depression [68].

Vrijens et al. [98] used a time-dependent continuation model to capture longitudinal adherence data in a cohort of 35 patients followed up over 12 months, and associate these to viral outcomes. The methodology enabled the authors to capture distinct categories of longitudinal adherence which were predictive of different probable viral outcomes. At a high level, this is a similar approach as that implemented in this dissertation in terms of identifying latent heterogeneity in longitudinal adherence and evaluating the association between these latent profiles and viral outcomes.

Voisin et. al [97] collected data on self-report adherence, psychological distress, HIV stigma, substance use, social support, family acceptance, self-efficacy, and demographics from a cohort of 92 young African-American PLWH in the USA, at three time points over a 24 month period. The authors used an unspecified longitudinal latent class model to identify two latent trajectories of self-reported adherence - representing

consistently high adherence ($N = 62$) and consistently low adherence ($N = 30$). Using logistic regression, the authors found that those with high self-reported adherence: were less likely to have daily/weekly alcohol or cannabis use over the previous 3 months; reported higher family acceptance; and reported higher self-efficacy. The authors found no differences between the groups' experiences of HIV stigma, social support, or psychological distress. This study relied on self-reported adherence, which is often a poor measure of true adherence, and only had data at three time points spread over 24 months. There is thus potential for bias in the analysis through the unreliable measure of adherence, and the lack of longitudinal granularity means that there were likely patterns of heterogeneous adherence which could not be revealed. The population from which this study sampled was also very different from that in this research and thus there is likely little ground to generalise the results to the South Africa context.

Zhang et. al. [108] conducted a similar analysis in a 2021 publication, using Latent Class Profile Analysis to uncover heterogeneity in self-reported self-management scores in a cohort of 868 PLWH in Sichuan Province, China. The authors then used a multinomial logistic regression to explore the association between latent class membership, and socio-demographic and disease-related predictors. In this case, the study was cross-sectional: self-reported adherence management was assessed using a 20 item, 3 factor HIV Self-Management Scale [100] at a single time-point - and thus no longitudinal components were considered. The population considered in this study also likely differs from the South African context.

A 2019 study by Mody et. al. [59] used a multivariate group-based trajectory model (GBTM) to cluster longitudinal adherence and retention profiles, and multinomial logistic regression to assess associations with baseline characteristics and adherence/retention profiles of a cohort of 38 879 PLHW in Zambia. Pharmacy refill data (as medication possession ratio [MPR]) was used as a measure of adherence, with loss to follow-up [LTFU] as a binary variable to indicate retention in care. The model was thus specified with two longitudinal outcomes: the MPR was modelled assuming a censored normal distribution; LTFU was modelled as a binary outcome with a logit link function. The latent class analysis revealed six distinct latent longitudinal groups representing: consistently high adherence and retention; early non-adherence but consistent retention; gradually decreasing adherence and retention; early LTFU with later re-engagement; early LTFU without re-engagement; and late LTFU without re-engagement. The multinomial logistic regression did not find any significant associations between the latent class membership for any group and any of the socio-demographic or clinical variables that they collected. The authors then implemented an adjusted Poisson regression model with a time offset to assess associations between latent group membership and incidence rate ratios of mortality. This revealed significantly higher mortality incidence rates for both early LFTU groups and the late LFTU group when compared to the group with persistently strong adherence and good retention.

In a 2020 study, researchers also utilised the GBTM to cluster longitudinal adherence profiles among a cohort of 1206 PLWH in Brazil - with the aim of identifying whether adherence was improved when switching from multiple to single tablet regimens [22]. This study used proportion of days covered from pharmacy refill monitoring (PRM) data as a measure of adherence. The authors implemented the GBTM with four groups for the periods before and after the regimen switch, and compared the cluster assignment for individuals. The findings indicated that the majority of individuals either migrated towards a more adherent group or did not change cluster assignment. The authors did not investigate the virologic outcomes in the sample and used a relatively naive measure of adherence in PR data.

Keiser et. al. [42] took a different approach to the topic by conducting a latent class analysis on socio-demographic and behavioural characteristics of 4483 PLWH at enrollment into the Swiss HIV Cohort Study. They then separately regressed virologic, adherence, and retention outcomes against these latent classes. The analysis revealed some significant differences between the latent classes' probability of viral suppression and adherence to ART, but the Swiss cohort is so different from that considered here that these particular results may not be relevant to this research.

Khorashadizadeh [45] presented an analysis which is applicable to this research in many ways. The study was a retrospective cohort study which observed CD4 counts, as well as time to HIV-related death or censoring, over 213 PLWH in Iran between 1989 and 2014. The analysis included: (a) a shared-random effects joint model of longitudinal CD4 counts and time-to-death; and (b) a latent class joint model of the same outcomes. Results related to the former model indicated that increasing deviations from the population average CD4

counts were associated with increasing risk of death. The latter analysis suggested that there were three subgroups within the sample, corresponding to 'High', 'Moderate' and 'Low' risks of death based on their different longitudinal CD4 profiles. While this analysis did not directly consider either adherence or viral outcomes, it showed the potential of coupling latent class analysis with joint modelling to reveal unaccounted for heterogeneity within a sample.

4 Statistical methods

4.1 Longitudinal models

Many experimental and observational studies feature design characteristics that lead to the formation of grouped data. For example, consider a longitudinal multi-center trial. This exhibits a nesting structure, where repeated observations are nested within participants, who are further nested within study sites. Of particular interest is the presence of repeated measures data, where many observations are recorded over time for each participant. This induces an inherent correlation between observations from the same individual, rendering conventional single-class generalized linear models unsuitable for analysis [72].

In the case of this dissertation, ignoring the correlation between adherence measures collected from the same individual over time will tend to bias the statistical analysis. Thus, alternative statistical methodologies must be explored to appropriately address the correlated nature of the data and enable valid inferences while accounting for the inherent groupings.

The first of these considered here are mixed-effects models. Mixed-effects models are a class of regression models that are commonly used to analyse grouped data. These express an outcome as a linear combination of fixed effects (population-level parameters) and random effects (grouping-level parameters). Mixed effects models are thus often used to analyse longitudinal data, where the random effects result in a covariance structure that effectively captures the inherent correlation between observations that come from the same group [72]. Generalized linear mixed effects models (GLMMs) are defined for the case where the outcome variable is distributed according to some parametric distribution in the exponential family. These, along with many linear regression models, assume that the random errors are normally distributed with mean zero and constant variance. In addition, the random effects are assumed to be homoskedastic and normally distributed around zero, and independent of the random errors.

Consider the repeated sampling of a random variable Y - such as the number of missed doses recorded by an EAM device between each clinic visit - sampled at each of $j \in (1, ..., n_i)$ occasions, such that $\boldsymbol{y_i}$ represents the vector of observed responses for subject $i \in (1, ..., N)$. The generalized mixed-effects model is specified as [72]:

$$g(\boldsymbol{y_i}) = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\epsilon_i}, \tag{1}$$

where $g(.)$ is an appropriate link function - determined by the distributional form of Y; $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ are design matrices for the fixed and random effects, respectively; $\boldsymbol{\beta}$ is a vector of fixed effect coefficients; $\boldsymbol{b_i} \sim N(\boldsymbol{0}, \boldsymbol{R})$ is a vector of subject-specific random effect coefficients; and $\boldsymbol{\epsilon}_i \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ represent within-subject residual errors associated with each of the sampling occasions.

The inclusion of random effects decomposes the covariance of $\boldsymbol{Y_i}$ into two sources - that inherited from the heterogeneous random effects and that from the the residual variance [20]:

$$Cov(\boldsymbol{Y_i}) = Cov(\boldsymbol{Z_i}\boldsymbol{b_i}) + Cov(\boldsymbol{\epsilon_i}) = \boldsymbol{Z_i}\boldsymbol{R}\boldsymbol{Z_i}^T + \sigma^2\boldsymbol{I}. \tag{2}$$

Where necessary, this last assumption of uncorrelated residual errors with constant variance can be relaxed by explicitly modeling the error covariance structure. These models can also be extended with further random effects to incorporate further levels of nesting - such as in the case of multi-center longitudinal studies.

Mixed-effects models are conditional models, thus very useful when the objective of analysis is centered around capturing grouping-level variability over the repeated measurements (such as estimating individual trajectories over time). An alternative approach is marginal models, which emphasize the population-average effects while regarding within-subject correlations as nuisance parameters by directly specifying a working correlation structure [20]. Suppose that the marginal mean response is linearly related to some covariates, X_{ij}, through a known link function $g(.)$:

$$g(\mu_{ij}) = \boldsymbol{X}_{ij}^T\boldsymbol{\beta}. \tag{3}$$

The marginal model is specified if the variance of the response, conditional on the X_{ij}, can be written as a function of the mean, scaled by some constant ϕ:

$$Var(Y_{ij}|X_{ij}) = \phi v(\mu_{ij}).$$

The pairwise within-subject correlation, conditional on the X_{ij}, can also be written as a function of the mean and a set of unknown parameters $\boldsymbol{\alpha}$, such that the covariance matrix can be expressed as:

$$\boldsymbol{V}_i = \boldsymbol{A}_i^{\frac{1}{2}} Corr(\boldsymbol{Y}_i) \boldsymbol{A}_i^{\frac{1}{2}} \approx Cov(\boldsymbol{Y}_i),$$

where $\boldsymbol{A_i}$ is a diagonal matrix with $Var(Y_{ij})$ on the diagonals; and $Corr(\boldsymbol{Y}_i) = h(\boldsymbol{\mu_i}, \boldsymbol{\alpha})$ [20].

4.1.1 Estimation in longitudinal models

A common method of estimating the parameters in a GLMM is the Expectation-Maximization (EM) algorithm [72]. The EM algorithm is an iterative method for estimation which involves iterating between an Expectation (E) step, where the random effects are estimated given the current parameter values, and a Maximization (M) step, where the model parameters are updated given the estimated random effects [72]. This process continues until some convergence criteria are met. The EM algorithm is attractive as it is relatively simple to implement and can handle models with complex random effects structures [20].

Another estimation approach, called adaptive Gaussian quadrature, bases its solution on numerical integration [73]. This method approximates the integrals involved in the likelihood function using a weighted sum of function evaluations at carefully chosen quadrature points [73]. These weights and quadrature points are adapted based on the random effects distribution, improving the accuracy of the approximation [73]. Adaptive Gaussian quadrature can provide more accurate parameter estimates than the EM algorithm, particularly for models with non-normal random effects [20].

4.1.2 Mixed effects models for count data

When the response variable is a continuous outcome, with a conditional distribution that is approximately normal, the identity function is used as the link function in equation 1. However, count outcomes - such as the number of missed ARV doses - are discrete and thus would break the assumption of normally distributed (and continuous) within-subject random errors. An alternative is to assume that the conditional distribution for the outcomes follow a Poisson distribution [72]. Following the notation introduced previously:

$$\boldsymbol{y}_i|\boldsymbol{b}_i \sim Poisson(\boldsymbol{\lambda}_i), \tag{4}$$

where $\boldsymbol{\lambda}_i$ is the vector of expected counts for the i-th participant.

One can then specify a log link function to allow the conditional expectation of $\boldsymbol{y}_i$ to be modeled as a linear function of the fixed and random effects [72]:

$$log(\boldsymbol{\lambda}_i) = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i}, \tag{5}$$

where $\boldsymbol{b_i} \sim N(\boldsymbol{0}, \boldsymbol{R})$.

An important property of the Poisson distribution is the mean-variance relationship, where $Var(\boldsymbol{y}_i|\boldsymbol{b}_i) = \boldsymbol{\lambda}_i$. To elaborate, the Poisson GLMM fundamentally assumes that the variance of the count response, given the random effects, is equal to the conditional mean. This implies that the random effects explain all additional variability in the response. Many data generating processes will not satisfy this assumption, in which case the data may exhibit underdispersion or overdispersion.

If there is additional variability in the response, one option is to turn to the negative binomial GLMM [72]. This follows a very similar structure to the Poisson GLMM. The material difference is that the conditional distribution of $\boldsymbol{y}_i$ is assumed to follow a negative binomial distribution with a shape parameter θ which

makes it much more flexible than the Poisson distribution. In particular, the same formulation as equation 5 applies but with:

$$\boldsymbol{y}_i|\boldsymbol{b}_i \sim negbin(\boldsymbol{\lambda}_i, \theta\boldsymbol{I}). \tag{6}$$

In some cases, the observed counts may be biased by some varying but known exposure. For example, the number of possible missed doses for an individual is influenced by the total number of possible missed doses in a certain time period. This may vary between individuals and also between observation points. In this case, an offset term can be introduced into the model to adjust the expected counts. In this case, the model would be specified as:

$$log(\boldsymbol{\lambda}_i) = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + log(\boldsymbol{o}_i), \tag{7}$$

where $log(\boldsymbol{o}_i)$ represents the logarithm of the vector of offset terms for individual i, and thus o_{ij} represents the number of possible counts for the i-th individual's j-th observation.

This model can be re-arranged to reflect the effect of the offset's inclusion in adjusting the estimated expected counts:

$$log(\frac{\lambda_{ij}}{o_{ij}}) = \boldsymbol{X}_i^T\boldsymbol{\beta} + \boldsymbol{Z}_i^T\boldsymbol{b_i}.$$

4.1.3 Mixed effects model diagnostics

The assumptions underlying mixed effects models can be assessed with the use of DHARMa residuals. Conventional residual diagnostics use raw, Pearson or Deviance residuals - which are not easily interpretable in the case of mixed-effects models due to the deconstruction of the covariance induced by the random effects (see equation 2). This is especially complicated by models which are specified with multiple levels of random effects [23]. DHARMa residuals were thus introduced to provide a reliable and interpretable method for diagnosing misspecification with this class of models [31].

DHARMa residuals are generated via a simulation procedure and inverse transform sampling. For each observation, many responses are simulated generating a simulated probability distribution under the fitted model from which an empirical cumulative density function (eCDF) is obtained. The residual is then defined through the inverse transformation of the eCDF at the observed value of the response. This provides residuals which are constrained between 0 and 1, expected to follow a uniform distribution, and roughly symmetrical when plotted against predictors [18] [31]. These residuals can thus be plotted against their expected distribution and against the predictors to assess the overall model fit, over/underdispersion, homoskedasticity, and to identify outliers [31].

4.1.4 A note on missingness in longitudinal data

Longitudinal studies can produce data that is incomplete for various reasons. There are three main categories for the mechanisms through which missingness can arise: missing completely at random (MCAR); missing at random (MAR); and missing not at random (MNAR) [50]. These differ based on dependence between missingness, observed data and relevant unobserved data. Understanding the missing data mechanism is crucial since it can potentially have very material implications for an analysis.

MCAR occurs when the probability of missingness is purely stochastic - independent of both the observed and unobserved data [50]. MAR occurs when the probability of missingness is related to observed variables but not to unobserved variables. An example may be when drop-out is related to observed baseline health status. MNAR occurs when the probability of missingness is related to both observed and unobserved variables. This means that the missingness is non-random and at least some of this dependence cannot be captured with the measured data. MCAR and MAR are ignorable in the case of the GLMM model, however MNAR presents a non-ignorable problem which introduces bias into the model [35]. In the case of this research, it is assumed that any missing longitudinal adherence data arises independently of unobserved variables and is at least MAR.

4.2 Survival models

Survival analysis covers a suit of techniques which aim to model survival experiences with time-to-event data, such as the time to viral non-suppression as the response variable [14] [48]. Most broadly, survival analysis provides a means of estimating the probability of survival or failure over time. Framing this as a regression problem can allow researchers to quantify the association between covariates and survival experiences; adjust for potential confounding; and to thus estimate the probability of survival conditional on the values of these covariates with an estimated uncertainty. However, conventional statistical methods such as linear regression, contain assumptions unsuited to analysing this type of data. This is due to the inherent censoring of survival times - the time of the event is usually not observed for all individuals during the observation window(s). This is a specific type of missing data which induces bias into estimates and any resulting inference when obtained from analytical methods which ignore the censoring [80].

The choice of appropriate analytical methods in this respect depends on the censoring mechanism. Censoring mechanisms can be classed as left, right, or interval censored; as well as either informative or non-informative [48] [80]. Right censored data is common to see in many experimental and observational studies. This usually occurs when the event times for a proportion of the individuals under observation are not recorded before the end of the study (or study withdrawal). There are two possibilities in this situation: either an individual never experiences the event or they experience the event after the observation period. Thus the event times are considered to be censored. Left censoring is less common and occurs when event times are not recorded because they occurred before the commencement of the observation period. Lastly, interval censoring occurs when events occur sometime between the discrete observation times. For example, an individual may develop a disease sometime between clinic visits $(n-1)$ and n, but the event is only recorded at visit n. In some cases, especially when the gaps between observation times are relatively close together, interval censoring is ignored and the event is assumed to occur when it is first observed.

Censoring mechanisms are also classed according to whether the event under interest is independent of the censoring mechanism. This is the case under non-informative censoring, while under informative censoring the probability of experiencing an event is dependent on whether an individual is censored or not [14] [80]. The methods presented here assume that the probability of viral non-suppression is independent of withdrawal from the study, and thus non-informative. The methods further assume that the event times are right-censored - thus that viral non-suppression occurs sufficiently near to the recorded observation time and that unobserved events may occur after the period which an individual is under observation.

Kaplan-Meier (KM) survival analysis is a non-parametric method for estimating and visualizing survival probabilities over time [39]. Let T be a random variable denoting the survival times (such as time from study entry to viral non-suppression), and let the probability of failure at $T = t$ be represented by the probability density function: $f(t) = P(T = t)$ [39]. The cumulative density function, $F(t) = P(T \leq t)$, thus represents the probability of failure at or before t. The survival function is defined as the complement of $F(t)$, and represents the probability of survival to at least time t: $S(t) = P(T > t)$ [39].

The KM estimator for the probability of survival at time t is defined as [39]:

$$\hat{S}(t) = \prod_{t_i \leq t} (1 - \frac{d_i}{n_i}), \tag{8}$$

where t_i are the ordered event times, d_i is the number of events at time t_i, and n_i is the number of subjects at risk just before t_i (the "risk set") [39].

KM curves plot these estimated survival probabilities against time, providing a visual representation of the survival experience. To compare survival between groups, the sample can be stratified by a categorical variable, and separate KM curves constructed for each stratum. The log-rank hypothesis test can then be used to formally assess differences in survival distributions between strata [53]. While KM analysis does not provide effect size estimates like regression models, it is very valuable for exploratory analysis.

The foundations of modern survival analysis lie in the formulation of the hazard function, proposed in Cox's seminal 1972 publication [15]. The hazard function represents the probability of failure in the instant

following time t, given survival up to time t [15]. This can be interpreted as the instantaneous risk of failure at t and is derived from the above probability density function and consequent survival function definitions [15]:

$$\begin{aligned} h(t) &= P(t \leq T < (t + \Delta t) | T > t) \\ h(t) &= \lim_{\Delta t \to 0} \frac{P(t \leq T < (t + \Delta t) | T > t)}{\Delta t} \\ h(t) &= \frac{f(t)}{S(t)}. \end{aligned} \tag{9}$$

The Cox proportional hazards (PH) model expresses the hazard of failure for individual i as a function of fixed effects ($\mathbf{w}_i$) [15]:

$$h_i(t|\mathbf{w}_i) = h_0(t) \exp^{\eta_i}, \tag{10}$$

where $\eta_i = \mathbf{w}_i^T \boldsymbol{\gamma}$ is a linear combination of fixed effects; and $h_0(t)$ is the baseline hazard function, representing the hazard of failure for an individual with $\eta_i = 0$. The baseline hazard has no assumed distributional form and carries no assumptions in the Cox-PH model, hence it's designation as a semi-parametric model. The parameters $\boldsymbol{\gamma}$ are estimated independently and then used to estimate the baseline hazard [14].

This model is easily linearized by expressing the logarithm of the hazard ratio as a linear function of the fixed effects [14]:

$$log(\frac{h_i(t|\mathbf{w}_i)}{h_0(t)}) = \mathbf{w}_i^T \boldsymbol{\gamma}, \tag{11}$$

and thus the coefficients $\boldsymbol{\gamma}$ estimate the marginal association between each of the fixed effects and the logarithm of the hazard ratio *at any time t*.

For example, γ_p represents the magnitude of effect of a one unit increase in the continuous independent variable w_p on the logarithm of the hazard at any t, holding all else constant. $\exp^{\gamma_p}$ thus represents the factor change in the risk of failure associated with a unit increase in w_p at any time t, ceteris paribus [14].

Given a set of n unique, ordered failure times, maximum likelihood estimates of the model parameters can be estimated through numerical optimization methods such as the Newton-Raphson procedure by optimizing the following (partial) log-likelihood function [14]:

$$log(L(\boldsymbol{\gamma})) = \sum_{i=1}^{n} \delta_i [\boldsymbol{\gamma}^T \mathbf{w}_i - log(\sum_{l \epsilon R(t_i)} \exp^{\boldsymbol{\gamma}^T \mathbf{w}_l})], \tag{12}$$

where δ_i indicates whether the i-th event t_i is censored ($\delta_i = 0$) or not; and $R(t_i)$ represents the risk-set at time t_i [15].

4.2.1 Extension for recurrent events

The basic Kaplan Meier analysis and standard Cox-PH model assumes a single event, with conditional independence assumed between event times. In many applications, however, it is possible for each experimental unit to experience multiple events of the same type. For example, when monitoring viral outcomes, a participant may exhibit viral non-suppression many times during the observation period - with the events thus clustered within participant. If this unobserved heterogeneity is ignored, the assumption of conditional independence will be broken, and the estimates from both the Kaplan-Meier estimator and the standard Cox-PH model would be biased.

One naive method for dealing with this could be extending the Kaplan-Meier analysis by changing the definition of the risk set. At each time t, those at risk could be considered as those who had not yet experienced their terminal event or censoring. This is implemented by structuring data in a counting process format, where each subject-event is entered as a (start, stop) interval and event indicator observation [21]. At each event time t_i, the risk set - n_i in equation 8 - is then defined as the number of intervals which intersect t_i. The same KM estimator as presented in equation 8 is thus used naively by adjusting the risk set

[94] [12]. A naive estimator which assumes independence, this approach is insufficient for many applications involving inference but may provide a useful exploratory tool.

Overall, this presents a similar problem to that encountered when modeling repeated measures data. In a similar fashion, one can deal with this dependence between recurrent events by extending Cox models either with: (a) marginal models, which account for the dependence through the estimation of standard errors [101]; or (b) with frailty models.

An example of a marginal model is that proposed by Andersen and Gill [1]. This estimates population level effects under a counting process format and accounts for heterogeneity due to clustering of events within subjects through robust standard error estimation. The hazard function for the i-th individual at the j-th event time is given by [1]:

$$h_{ij}(t|\mathbf{w}_i(t)) = h_0(t)\exp(\boldsymbol{\gamma}^T\mathbf{w}_i). \tag{13}$$

The regression coefficients $\boldsymbol{\gamma}$ are estimated within a likelihood-based methodology, as with the standard Cox-PH model. To account for within-subject correlation, robust standard errors are used. The robust estimate is estimated by [1]:

$$\text{Var}(\boldsymbol{\gamma}) = I(\boldsymbol{\gamma})^{-1}\left(\sum_{i=1}^{n} U_i(\boldsymbol{\gamma})U_i(\boldsymbol{\gamma})^T\right)I(\boldsymbol{\gamma})^{-1}, \tag{14}$$

where $U_i(\boldsymbol{\gamma})$ is the score vector for the i-th individual and $I(\boldsymbol{\gamma})$ is the observed information matrix.

The Andersen-Gill model above assumes conditionally proportional hazards and non-informative censoring over time [1]. Additionally, recurrence of events within individuals are assumed to be conditionally independent - i.e. fully explained by the covariates.

Frailty models adjust for clustering of recurrent events in a manner analogous to the random effects in a linear mixed effects model. The Cox-PH model is adjusted to include a frailty term, ω_i, specific to cluster i. The model is formulated as follows: let T_{ij} represent the event time of the j-th non-suppression event for the i-th individual, $i \in (1,\ldots,n)$, $j \in (1,\ldots,n_i)$, where n_i is the number of events observed for individual i. The hazard function for the j-th event for individual i is given by [48]:

$$h_{ij}(t|\mathbf{w}_i) = h_0(t)\omega_i\exp(\mathbf{w}_i^T\boldsymbol{\gamma}), \tag{15}$$

where the shared frailty term ω_i is assumed to follow a known distribution, often the gamma or log-normal.

If one lets $v_i = log(\omega_i)$, then the model can be rewritten as:

$$log(\frac{h_{ij}(t|\mathbf{w}_i)}{h_0(t)}) = \mathbf{w}_i^T\boldsymbol{\gamma} + v_i, \tag{16}$$

which closely resembles a generalized linear mixed-effects model. Here the failure times are assumed to be independent conditional on $\mathbf{w}_i$.

Given that the frailties are unobserved, the parameter estimation of the standard Cox-PH model cannot be followed. Instead, the model parameters must be estimated via the EM algorithm [48].

4.2.2 Extension for time-varying covariates

The survival analysis methods presented above assume that the risk of failure can be modelled as a function of independent variables which are constant over time. These could be factors measured at baseline such as sex and time since ART initiation.

There are, however, other factors which are not constant over time and influence the risk of failure. These can be internal or external/ancillary covariates: internal when variation over time is generated as a result of the actions/behaviour of the individual under observations; and external when variation is completely

independent of the individual's action [48]. Adherence is an example of an internal time-varying covariate, where an individual's adherence varies over the course of time and is related to their own behaviour.

The extended Cox model incorporates time-varying covariates $\boldsymbol{y}_i(t)$ in the following formulation [48]:

$$h_i(t|\mathbf{w}_i, \boldsymbol{y}_i(t)) = h_0(t) \exp^{\mathbf{w}_i^T \boldsymbol{\gamma} + \boldsymbol{y}_i(t)^T \boldsymbol{\alpha}}, \tag{17}$$

where α_p would represent the average association between the log hazard of failure at t and the value of y_{ip} at t, ceteris paribus.

In some cases, the covariate may vary over time but it's association with the survival outcome is constant. In other cases, this association may change over time and induce time-dependency into the data generating process. The latter would cause *time-dependency* in the covariate's coefficient - which results in a violation of the PH assumption [5] [109]. It is thus possible to assess time-dependency by fitting a model which ignores it and evaluating the PH assumption. If the assumption is violated, then the model can be adjusted by introducing an appropriately specified time-based step function such that a separate hazard ratio can be estimated for each time interval [5] [109].

The extended Cox model introduced above assumes that the time-varying covariates evolve in a step-wise fashion, changing value at each observation time and remaining constant between measurements [80]. Furthermore, the values of $y_i(t)$ are assumed to be observed with no uncertainty - yet another assumption that may present the above methodology as untenable in many real world applications. Section 4.3 outlines a statistical approach which does not suffer from these limiting assumptions.

4.2.3 Survival model diagnostics

One major caveat of the standard Cox-PH model is the Proportional Hazards (PH) assumption. This states that the marginal effects (γ_p) of each of the covariates on the hazard of failure are constant over time, and thus the hazard of failure at t is proportional to the baseline hazard for all t [15]. If w_{age} represents age, then the PH model assumes that the marginal effect of age (γ_{age}) on the survival experiences remains constant - hence the emphasis on *at any time t* in the interpretation of the model parameters. There are many situations where this may not be the case and thus this assumption must always be validated under a standard Cox-PH model. A formal test for non-proportionality using a form of weighted partial (Schoenfeld) residuals is provided by Grambsch and Therneau [27]. This estimates $\hat{\gamma}_p(t)$ for each p using the weighted residuals and tests for a non-zero slope of $\hat{\gamma}_p(t)$ against t. This can also be assessed via subjective interpretation of a plot of these scaled Schoenfeld residuals against the observation time.

The Cox model also assumes that the relative hazard can be expressed as a linear combination of the independent variables. The functional form of the covariates should thus be assessed for linearity. Martingale residuals can be used as a diagnostic tool to this extent. Furthermore, observations which are outlying with respect to the model fit should be investigated using deviance residuals, and score residuals can be used to investigate observations which disproportionately influence the estimation of parameters. Lastly, the overall fit of the model to the data can be investigated through the distribution of Cox-Snell residuals. More detail is given to these residuals and their uses in the Appendix (see A.1.1).

4.3 Joint models

In the previous two sections, it was shown: (a) how linear mixed-effects models can be used to model repeated measures data - such as longitudinal adherence outcomes; and (b) how time-to-event data can be analysed in a regression framework while accounting for recurrent events and/or exogenous time-varying factors. This section presents a harmonised approach for the modelling of repeated measures and time-to-event data through the application of joint models. Joint models of this type are generally used with two main aims in mind: firstly, these methods provide a efficient and accurate way of estimating the association between survival outcomes and endogenous time-dependent covariates; secondly, these models can be used to primarily model the longitudinal response while accounting for missing data, for example due to dropout, by leveraging the available survival information [80]. In this section, the methods are presented with focus on

the former aim - since the primary goal is to model the association between the risk of viral non-suppression and longitudinal adherence outcomes.

The basic joint model is comprised of two sub-models with a unified likelihood function: one model that specifies the form of the underlying survival model and a second that specifies the form of the longitudinal model. For a single time-dependent covariate $y_i(t)$ and a basic survival outcome with no recurrent events, the survival sub-model can be specified as:

$$h_i(t|\mathbf{w}_i, M_i(t)) = h_0(t)\exp(\mathbf{w}_i^T\boldsymbol{\gamma} + \alpha m_i(t)), \tag{18}$$

where $m_i(t)$ is the *true and unobserved value* of Y for individual i at t, and thus α measures the association between the true value of the longitudinal outcome and the logarithm of the hazard of failure; and $M_i(t)$ represents the entire history of the true longitudinal process up to t [80].

The longitudinal sub-model's role pertains to the estimation of the true value of the longitudinal variable at time t. This is not observable because the longitudinal outcome is measured with error, and usually at a set of discrete time points [80].

To account for this, a (generalized) linear mixed-effects model can be used to estimate $m_i(t)$ for any t (and not just the times at which the variable is observed). This takes the same form as that presented in equation 1:

$$m_i(t) = E(y_i(t)|b_i) = g^{-1}({x_i}^T(t)\beta + z_i^T(t)b_i), \tag{19}$$

where $g^{-1}()$ specifies the inverse of the link function. This depends on the nature of Y (eg: continuous vs categorical). The same assumptions regarding homoskedastic and normally distributed random errors and random effects, centered around zero, apply to this sub-model as they do to the model when used alone.

While there are other methods to analyze longitudinal data, such as the marginal models touched on in Section 4.1, the choice of a mixed-effects model here is not arbitrary. The inclusion of subject-specific random effects in the model allows the reconstruction of not just the true population-level longitudinal trajectory ($x_i^T\beta$), but the entire history of the true trajectory *for each individual* ($M_i(t) = m_i(s), 0 \leq s < t$) [80]. Furthermore, the joint model assumes that the random effects account for any dependence between the repeated longitudinal measures [80]:

$$p(\boldsymbol{y}_i|\boldsymbol{b}_i;\boldsymbol{\theta}) = \prod_j p(y_{ij}|\boldsymbol{b}_i;\boldsymbol{\theta}),$$

and between these and the survival outcome:

$$p(\boldsymbol{y}_i, T_i, \delta_i|\boldsymbol{b}_i;\boldsymbol{\theta}) = p(\boldsymbol{y}_i|\boldsymbol{b}_i;\boldsymbol{\theta})p(T_i, \delta_i|\boldsymbol{b}_i;\boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ subsumes the parameters for the longitudinal and survival sub-models [80].

The joint likelihood for the joint model can thus be specified as [80]:

$$p(\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{y};\boldsymbol{\theta}) = \prod_i p(\boldsymbol{y}_i|\boldsymbol{b}_i;\boldsymbol{\theta})p(T_i, \delta_i|\boldsymbol{b}_i;\boldsymbol{\theta})p(\boldsymbol{b}_i;\boldsymbol{\theta}). \tag{20}$$

Estimation of the joint model parameters $(\boldsymbol{\theta};\boldsymbol{b})$ can be done via two different frameworks: EM algorithm for maximum likelihood estimation or Markov Chain Monte Carlo (MCMC) sampling of the Bayesian joint posterior probability distribution [80]. The Bayesian approach estimates the parameters by combining the likelihood in equation 20 and an (often uninformative) prior distribution to arrive at a posterior probability distribution, specified by [80]:

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{b}|\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{y}) &= \frac{\prod_i p(\boldsymbol{y}_i|\boldsymbol{b}_i;\boldsymbol{\theta})p(T_i, \delta_i|\boldsymbol{b}_i;\boldsymbol{\theta})p(\boldsymbol{b}_i;\boldsymbol{\theta})p(\boldsymbol{\theta})}{\prod_i p(T_i, \delta_i, \boldsymbol{y}_i)} \\ &\propto \prod_i p(\boldsymbol{y}_i|\boldsymbol{b}_i;\boldsymbol{\theta})p(T_i, \delta_i|\boldsymbol{b}_i;\boldsymbol{\theta})p(\boldsymbol{b}_i;\boldsymbol{\theta})p(\boldsymbol{\theta}). \end{aligned} \tag{21}$$

The *JMbayes2* package uses a combination of Gibbs sampling and Metropolis-Hastings algorithm to efficiently estimate properties of the posterior distribution in equation 21 [81].

4.3.1 Parameterisations of longitudinal covariate

As defined above, the joint model can be used to estimate the association between the risk of failure at time t with the value of a time-dependent variable Y *at the same time point* t. This is referred to as the current value parameterization, which is not necessarily always the ideal case.

In some applications it may be of more interest to investigate the association with the risk of failure and the rate of change of Y at t, or the accumulation of Y up to t, or perhaps the value of Y at $t-c$, where c specifies some lag of time. All of these examples and more are possible through a more generalized expression for the parameterization of $y_i(t)$ in the joint model. The generalized formulation can be written as [80]:

$$h_i(t) = h_0(t)exp(\mathbf{w}_i^T\boldsymbol{\gamma} + \alpha f(m_i(t-c))). \tag{22}$$

The function $f(.)$ transforms the longitudinal variable in an appropriate way. For example, if the accumulation of Y were of interest, one could use $f(m_i(t)) = \int_{s=0}^{t} m_i(s)ds$ [80]. Alternatively, if the rate of change of Y were of interest, then $f(m_i(t)) = \frac{d}{dt}m_i(t)$ would be suitable [80]. Ultimately, there are many ways which the longitudinal covariate could thus be parameterized under the generalization provided by equation 22 which allows much more flexibility in the model.

4.3.2 Extension for recurrent events

The joint model can be adjusted to handle survival data with recurrent events [80]. One can simply integrate the frailty model specified in equation 16 into the joint modelling approach. The joint model (with the current value parameterization and a continuous Y) would thus be fully specified by:

$$\begin{aligned} h_i(t) = h_0(t)exp(\mathbf{w}_i^T\boldsymbol{\gamma} + \alpha m_i(t) + v_i) \\ y_i(t) = m_i(t) + e_i(t), \end{aligned} \tag{23}$$

where $m_i(t) = {x_i}^T(t)\beta + z_i^T(t)b_i$; $v_i \sim N(0, \sigma_v^2)$; $b_i \sim N(0, \sigma_b^2)$; and $e_i \sim N(0, \sigma_e^2)$.

4.3.3 Extension for multiple longitudinal covariates

In the case that the risk of failure is thought to depend on more than one time-dependent variable, the joint model can be extended by using a multivariate generalized longitudinal sub-model. Consider q longitudinal covariates, one can thus define [80]:

$$g_q(y_{iq}(t)) = x_{iq}^T(t)\beta + z_{iq}^T(t)b_{iq} + e_{iq}(t) = m_{iq}(t) + e_{iq}(t), \tag{24}$$

where the random effects and random errors are assumed to follow their own multivariate normal distributions centered around zero; and $g_q()$ defines an appropriate link function for each of the q variables.

The multivariate outcomes can then be incorporated into the relative risk sub-model as [80]:

$$h_i(t) = h_0(t)exp(\mathbf{w}_i^T\boldsymbol{\gamma} + \sum_q \alpha_q m_{iq}(t)). \tag{25}$$

4.3.4 Joint model diagnostics

There are three main aspects to validating a fitted joint model: (1) evaluating overall model fit; (2) validating the assumptions of each of the sub-models; and (3) evaluating convergence of the Bayesian estimation procedure.

The overall model fit uses comparative tools to assess the relative fits of different model specifications to the data by estimating the expected predictive accuracy of each. For this purpose, model fit statistics based on the Akaike Information Criteria (AIC) are used - where a model which fits the data well would yield

a statistic relatively lower in magnitude than a more poorly fitted model. The AIC statistic uses the log-likelihood evaluated at the MLE of the parameters with a penalty for the number of parameters in the model [24]:

$$AIC = -2log(p(\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{y}|\hat{\boldsymbol{\theta}}_{MLE})) + 2k,$$

where k is the number of parameters.

Overall model fit for models with hierarchical structures or informative priors are not suitably assessed by AIC, mostly due to the effective number of parameters not being as straightforward as in simpler models [24]. Therefore, it is more appropriate to evaluate the fit of a joint model using Deviance Information Criteria (DIC) or Watanabe-Akaike Information Criteria (WAIC).

DIC is computed by estimating the likelihood at the Maximum A Posteriori (MAP) estimates (i.e. the estimates obtained from maximizing the Bayesian posterior distribution) and penalizing by the effective number of parameters:

$$DIC = -2log(p(\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{y}|\hat{\boldsymbol{\theta}}_{MAP})) + 2k_{DIC},$$

where k_{DIC} estimates the number of effective parameters [24].

WAIC is similar to the DIC in that it takes a Bayesian approach to estimates the log pointwise predictive density and applies a different penalty that also estimates the effective number of parameters. WAIC can be defined as:

$$WAIC = -2log(\prod_{i=1}^{n} p(\boldsymbol{T}, \boldsymbol{\delta}, \boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{b})) + 2k_{WAIC},$$

where k_{WAIC} estimates the number of effective parameters and the log pointwise predictive density is evaluated by drawing from posterior simulations [24].

Another useful statistic for overall model fit, often useful in Bayesian settings, is the Logarithm of the Pseudo-Marginal Likelihood (LPML). This is related to how well a model predicts the observed data, with less negative values being associated with a better overall fit [2].

The second aspect of model diagnostics is the validation of the assumptions underlying the longitudinal and survival sub-models. This is usually done by assessing residuals of the sub-model. With respect to the Cox model, the same residual diagnostic procedures outlined in Section 4.2.3 and expanded on in Appendix A.1 should be followed. With respect to the longitudinal model, plots should be used to assess the assumptions of homoskedasticity and normality for the error terms (Dharma residuals) and the subject-specific random effects (as outlined in Section 4.1.3).

A third procedure is required to check the validity of a joint model fit. This has to do with the stability of the MCMC estimation procedure. Graphical methods can be used to assess convergence of the estimation algorithm and the resulting validity of the parameter estimates. Typically this includes traceplots, which plot the value of a parameter estimate against iterations, and kernal density plots of the parameter estimates over all iterations [74]. A stable estimation procedure would be reflected by: (a) a roughly stationary process in the traceplots (i.e: no sudden and prolonged deviations from a stable mean, no global trend); and (b) density plots that are unimodal.

4.4 Finite mixture models

Finite mixture models (FMM) are a powerful and flexible model-based method for capturing heterogeneity in data. In general, FMMs assume that the population consists of a finite number of distinct sub-populations, each with a different underlying data-generating mechanism. In this case, it is reasonably assumed that a mixture of distributions can approximate the observed distribution of the response(s). It should thus be evident why FMMs are extremely useful in clustering and latent class analyses, however these methods extend to any analysis where the modelling of an unknown distributional shape is of interest ([57]). As an example, FMMs could be used to probabilistically cluster individuals with similar adherence and/or viral outcomes by deconstructing the overall distributions of these responses at baseline into a finite number of sub-distributions.

In general, if $\boldsymbol{Y} = (\boldsymbol{Y_1}, ..., \boldsymbol{Y_n})^T$ represents a random sample of a p-dimensional vector (i.e. n observations taken over p variables) with a probability density function $f(\boldsymbol{y_j})$, and $\boldsymbol{y}$ represents a realisation of $\boldsymbol{Y}$, we can represent the general parametric FMM as [57]:

$$f(\boldsymbol{y_j}, \boldsymbol{\Psi}) = \sum_{g=1}^{G} \pi_g f_g(\boldsymbol{y_j}, \boldsymbol{\theta_g}), \tag{26}$$

where $f_g(.)$ are parametric 'component densities' associated with each of the G sub-populations; and π_g are 'mixing proportions' - probabilities that are hence constrained by:

$$0 \leq \pi_g \leq 1 \quad \forall \quad g \in 1, \ldots, G,$$

and

$$\sum_{g=1}^{G} \pi_g = 1.$$

Group assignment can be based on maximising the posterior probability of an individual belonging to a specific cluster, which is computed as:

$$P(\boldsymbol{y_j} | j \in I_g, \boldsymbol{\theta}) = \frac{\pi_g f_g(\boldsymbol{y_j}|\boldsymbol{\theta_g})}{\sum_{g'=1}^{G} \pi_{g'} f(\boldsymbol{y_j}|\boldsymbol{\theta_{g'}})}, \tag{27}$$

where I_g indicates whether $\boldsymbol{y_j}$ belongs to the g-th group.

Each of the component densities are defined by an unknown parameter set, $\boldsymbol{\theta_g}$, while all of the unknown parameters in the mixing model in equation (26) are subsumed in the vector $\boldsymbol{\Psi}$:

$$\boldsymbol{\Psi} = (\boldsymbol{\pi}^T, \boldsymbol{\theta}^T).$$

4.4.1 Longitudinal finite mixture models

The concept of finite mixture models is easily extended to the longitudinal, repeated measures context.

Assume that $\boldsymbol{y_i}$ is a vector of measurements for subject $i \in 1, ..., n$ repeated on each of $t \in 1, ..., T$ occasions. Furthermore, assume that among the population there are G subclasses - each with a distinct longitudinal trajectory which can be modelled by some specification of a longitudinal model (such as the GLMM or GEE models introduced earlier). The marginal probability of a randomly chosen trajectory for $\boldsymbol{y_i}$ is defined as [63]:

$$P(\boldsymbol{y_i}) = \sum_{g=1}^{G} \pi_g f_g(\boldsymbol{y_i}), \tag{28}$$

where $f_g(\boldsymbol{y_i})$ is the distribution of $\boldsymbol{y_i}$, conditional on individual i belonging to subclass g; and π_g are the mixing proportions.

For a continuously distributed response, the multivariate normal density could approximate the conditional distribution $f_g(\boldsymbol{y_i})$:

$$\boldsymbol{y_i}^g \sim N(\boldsymbol{\mu}^g, \boldsymbol{\Omega}^g),$$

where $\boldsymbol{\mu}^g$ represents the expected vector of responses over the T time-points for individuals in subclass g; and $\boldsymbol{\Omega}^g$ represents the covariance matrix associated with these repeated responses.

The specification of $\boldsymbol{\mu}^g$ and $\boldsymbol{\Omega}^g$ determines the formulation of different longitudinal FMMs [17].

4.4.2 Group-based trajectory models

Consider the case where the relationship between the response and time is defined by only fixed effects, such as an order k polynomial of time. In this case, all individuals in class g are assumed to follow the same mean trajectory but this differs between classes [65]. Therefore, the parameters of the longitudinal model are distinct to group g - such that the mean response for group g is defined as:

$$\boldsymbol{\mu}^g = \sum_{j=0}^{k} \beta_j^g \boldsymbol{t}^j,$$

where $\boldsymbol{t}$ is a vector of time points at which the response is recorded.

Furthermore, consider restricting $\boldsymbol{\Omega}^g$ to be diagonal - implying the assumption of conditional independence, between the measured variables among individuals in the same group, over time [17]. In other words, individual deviations from $\boldsymbol{\mu}^g$ are assumed to be uncorrelated but serial dependence is still assumed at the population level due to the group-specific specification of $f(\boldsymbol{y_i})$ [64]. This case represents the *latent class growth analysis* (LCGA) group of models.

A subset of these models, the group-based trajectory model (GBTM) is defined when one further places the restriction of constant residual variances across time and between groups [17]:

$$\boldsymbol{\Omega}^1 = \boldsymbol{\Omega}^2 = ... = \boldsymbol{\Omega}^G = \omega^2 \boldsymbol{I}.$$

This assumption greatly simplifies the model, resulting in significantly fewer unknown parameters to be estimated, and is thus very useful when the response's temporal behaviour is complex or the sample size is relatively small [65].

It is important to note that assuming conditional independence over time (diagonal $\boldsymbol{\Omega}$) may render the LGCA class of models inadequate in many cases - while the further assumptions of constant residual variance over time ($\boldsymbol{\Omega}^g = \omega_g^2 \boldsymbol{I}$) and between groups ($\omega_1^2 = \omega_2^2 = ... = \omega_G^2$) may deem the GBTM as a gross oversimplification of the true data generating mechanism. For these reasons, the GBTM may be a useful starting point to identify the number of components, G, and the functional form of the longitudinal response - serving as an informative foundation upon which to build more complex models.

4.4.3 Growth mixture models

As was introduced in Section 4.1, the assumption of conditional independence over time is often not appropriate for repeated measures data, and a common way of amending this is to model the repeated measurements as a function of fixed and random effects - as in equation 1. The growth mixture model (GMM) extends the GLMM by assuming that the sample is drawn from a mixture of sub-populations, with each following a distinct mean longitudinal trajectory described by the set of unique parameters $\boldsymbol{\beta}^g$. The generalized GMM can thus be expressed as [65]:

$$g(\boldsymbol{y_i}^g) = \boldsymbol{X_i}\boldsymbol{\beta}^g + \boldsymbol{Z_i}\boldsymbol{b_i}^g + \boldsymbol{\epsilon_i}^g. \tag{29}$$

The same assumptions for GLMMs, as defined in Section 4.1, apply to this model - but are now defined separately per class ($\boldsymbol{b_i}^g \sim N(\boldsymbol{0}, \boldsymbol{R}^g)$; $\boldsymbol{\epsilon_i}^g \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}^g)$). Moreover, the covariance decomposition defined in equation 2 now applies per class:

$$Cov(\boldsymbol{y_i^g}) = Cov(\boldsymbol{Z_i}\boldsymbol{b_i}^g) + Cov(\boldsymbol{\epsilon_i}^g) = \boldsymbol{Z_i}\boldsymbol{R^g}\boldsymbol{Z_i}^T + \boldsymbol{\Sigma}^g.$$

It is clear that the inclusion of random effects allows the GMM to relax the assumption of conditional independence over time between individuals in the same class, and thus provides a much more robust model when compared to the GBTM. As with the GBTM, the GMM model can be simplified by assuming that the residual variances: are time-constant ($\boldsymbol{\Sigma^g} = \sigma_g^2 \boldsymbol{I}$); are constant across groups ($\boldsymbol{\Sigma}^1 = \boldsymbol{\Sigma}^2 = ... = \boldsymbol{\Sigma}^G$); and are constant across both time and groups ($\boldsymbol{\Sigma}^1 = \boldsymbol{\Sigma}^2 = ... = \boldsymbol{\Sigma}^G = \sigma^2 \boldsymbol{I}$) [17]. An additional simplification can be made by assuming that the covariance of the random effects is constant across groups ($\boldsymbol{R}^1 = \boldsymbol{R}^2 = ... = \boldsymbol{R}^G$) [17].

4.4.4 GBTM: Extension to count responses

The concepts introduced in Sections 4.4 - 4.4.3 can be easily extended to non-continuously distributed outcomes.

Where the response is a count variable, the conditional distribution of the outcome for individual i in class g at time t, $f_g(y_{it})$, can be approximated by the Poisson distribution with rate λ_t^g:

$$y_{it}^g \sim Poisson(\lambda_t^g).$$

In this scenario, the average rate of event occurrences for individuals in a given class at a given time, λ_t^g, can be modeled in either of the longitudinal frameworks presented in Sections 4.4.2 and 4.4.3, employing a log-link function [65].

Under Poisson assumptions, the expected conditional probability of a zero-count response is defined as $e^{-\lambda}$ - an evident result from the Poisson distribution. In some cases, additional zero counts may be generated outside of the Poisson process resulting in an zero-inflated distribution of counts. The zero-inflated Poisson distribution can thus be defined by modeling the zero-part of the distribution separately. Furthermore, varying exposure times can be accounted for by incorporated a time-at-risk offset variable κ_{it} [66]. Incorporating this theory results in the following conditional distribution [66]:

$$y_{it}^g \sim (1 - q_t^g) Poisson(\lambda_t^g \kappa_{it}) + q_t^g I(y_{it}^g = 0), \tag{30}$$

where $I(y_{it}^g = 0)$ is an indicator function for zero values [66].

In the above, λ_t^g can be modelled as a low-order polynomial of time, using a log-link function, and the zero-part proportion (q_t) can be modelled proportionately to $log(\lambda_t^g)$ for the excess probability of zero [49]:

$$logit(q_t^g) \propto log(\lambda_t^g) = \sum_{k=0}^{K} \beta_k^g t^k.$$

A joint likelihood is defined for this model as [66]:

$$L_G(\boldsymbol{\theta}) = \prod_{i=1}^{N} \sum_{g=1}^{G} \pi_g \prod_{t=L}^{U} \left[(1 - q_t^g) e^{-\kappa_{it} \lambda_t^g} \frac{(\kappa_{it} \lambda_t^g)^{y_{it}}}{y_{it}!} + q_t^g I(y_{it} = 0) \right],$$

where $\boldsymbol{\theta}$ subsumes all unknown parameters and can be estimated via maximum likelihood. The posterior probabilities of group membership at the maximum likelihood estimates can then be estimated using the same Bayesian framework defined in equation 27.

4.4.5 Longitudinal mixture model diagnostics

A couple of sources of subjective uncertainty are encountered with longitudinal mixture models. Firstly, the functional form for the underlying longitudinal model must be specified. Secondly, mixture models have an unsupervised learning type nuance in that the number of groups or clusters is imposed. Both of these sources of uncertainty can be minimized by using model fit statistics to choose a model specification that provides the best fit to the data. Traditionally, these include AIC and BIC. Nielsen et. al. also provide

a cross-validation error (CVE) based solution [66]. This involves recomputing the MLE and corresponding predicted value under leave-one-out cross-validation rules, and estimating an average cross-validation error (CVE). The minimum CVE thus corresponds to the model with the best predictive accuracy [66].

The global uncertainty of group assignment can also be assessed by looking at the distribution of posterior probabilities. If most of the mass of these distributions is concentrated at zero and one, it indicates that the groups are relatively well separated. If the posterior probabilities are relatively uniformly distributed, this implies that the mean longitudinal profiles and the assignment of members to these is pretty arbitrary.

Another step in assessing longitudinal mixture models is to investigate how well the fitted model captures the observed heterogeneous trends. This can be achieved by plotting the fitted model trajectories for each group against the observed (mean) trajectories for the individuals assigned to each group (based off their maximum posterior mixture probability). Group assignment is probabilistic, however, so perhaps a better reflection of the model's performance would be to weight the observed means by the mixture probabilities and use this to compare to the fitted trajectories [102]. These two approaches are of course practically indistinguishable if most of the posterior probabilities are near zero or one.

4.5 Software

All analyses implemented in this dissertation were achieved using the R language for statistical programming [75]. The suit of packages provided by the R *tidyverse* were utilised for data manipulation and visualisation, and convenient programming [105]. In addition, the packages *lubridate*, *foreign*, *psych*, *readxl*, and *gridExtra* were all used for specific programming and data manipulation tasks [28] [76] [106] [104] [3].

A selection of additional R packages were used for their uses for particular data visualisation tasks in the exploratory analysis. These include *ggpubr*, *GGally*, and *survminer* [40] [85] [41].

A number of R packages were used to apply the statistical models implemented in this research. The survival models were fit using *survival* [94], while *ldatools* and *survminer* were utilised for some of the corresponding residual diagnostic plots [6] [41]. For the longitudinal models, the package *GLMMadaptive* was used to fit the models [82] and *ggeffects* was used to illustrate the fitted longitudinal curves [51]. The package *DHARMa* was used to generate and visualise the residual diagnostic plots for these longitudinal models [30]. To fit the joint models, the package *JMBayes2* was utilised [83]. Lastly, the Group-Based Trajectory Models were fit to the data using the package *crimCV* [67].

All code used for data manipulation, data visualisation and data analysis can be found at the GitHub repository supplied in the Appendix (A.3).

5 Data description

5.1 ADD-ART

The data used in this research project was obtained from a longitudinal, clinic-based ART adherence study run by the Desmond Tutu Health Foundation's Center for Adherence and Therapeutics (D-CAT). The study, named 'ADD-ART', was a prospective cohort study designed to assess the ability of 2 adherence monitoring techniques (EAM and TFV-DP DBS) in predicting viral outcomes. Conducted between 2017 and 2019, this observational study recruited 250 virally-suppressed PLWH who were receiving ART from four primary healthcare clinics in Cape Town, Western Cape - the Hannah Crusaid Treatment Centre; Nyanga Clinic; Gugulethu Clinic; Vuyani Clinic. The population was restricted to adults who had been on an ART regimen containing tenofovir for at least 4 months and no more than 24 months. Each participant was enrolled in the study for 12 months. All data was obtained with permission from the D-CAT, as well as their partners in this study.

5.1.1 Outcomes

The main outcome of interest in ADD-ART was anti-retroviral adherence, for which there were three measures in this study. At monthly intervals, participants' self-reported adherence was reported and blood samples were obtained to measure TFV-DP levels in DBS. Furthermore, each participant was given an EAM Wisepill device at baseline - which records pill box openings by sending signals to a secure website when a cellular connection is available. This signal is usually sent at the time of opening, but the device can store the events and time of events for up to 30 days in the case of connection or other issues. As a result, each participant had daily, binary measurements for pill-box openings along with the time of each opening. These daily measures can be conveniently summarised into monthly counts, or a proportion of days/doses covered by device signal.

Two survival outcomes are considered and defined. The primary survival outcome is time in days from study entry until viral breakthrough (VB) - with VB defined as a monthly viral load > 1000 *copies/ml* (the cutoff used in public sector clinics in the Western Cape of South Africa). The secondary survival outcome is defined as time in days from study entry until loss-to-care - where loss-to-care is defined as having no clinic contact for 90 consecutive days.

5.1.2 Covariates

At baseline and at the final visit, comprehensive data on psychological and socio-demographic factors, and medical history were collected. The socio-demographic characteristics include age; sex; education; income; employment status; dwelling type; access to running water; electricity; and toilet; and food insecurity (missed meals in the last month). To record medical history, each patient self-reported time since HIV+ diagnosis; time on ART; type of ART regime; last CD4+ count; last VL result; and comorbid conditions. Psychological and other contextual factors were measured by administering the following: Life Events Checklist [84]; Patient-Clinic Relationship Scale (PCR) [86] [11]; Medication Specific Social Support Scale (MSSS) [84]; Berger HIV Stigma Scale [58]; HIV Treatment Knowledge Inventory (HTK) [99] [78]; Beliefs About Medication Questionnaire (BAM)[34]; HIV Disclosure Status Questionnaire [84]; Kessler Psychological Distress Scale (K10) [43] [44]; Substance Abuse and Mental Illness Symptoms Screener (SAMISS) [8] [103]. Body Mass Index (BMI) was later collected through an amendment to the protocol, and resulting in incomplete data taken at inconsistent times.

6 Exploratory analysis

6.1 Data management

6.1.1 General data cleaning and processing

The data used in these analyses were obtained from two different sources - the clinic data and the EAM data. The EAM data contained at least one observation per participant per day. To get 30-day counts of the observations, the EAM data had to be merged with the clinic data on participant ID and visit date - and observations tallied in the 30 days preceding the visit date. Prior to doing this, the EAM data was cleaned to mitigate known sources of bias (elaborated on in Section 6.1.2).

An important note here is that participants only started using the EAM device at the baseline visit. There is thus no EAM data at baseline. The baseline visits were thus removed from the longitudinal data, and the timeline was shifted one unit backwards such that the first visit was defined as time zero. The analyses thus refer to time from visit 1, although for mathematical purposes visit 1 is treated as time zero. Baseline information, such as viral loads and demographic information, were still retained.

Missingness was also investigated in the clinic data and some participants were excluded from the analysis on these grounds. This process is summarised in Figure 2. 11 participants attended less than 3 out of the expected 12 visits, and were thus excluded on the basis of containing very little information on the longitudinal adherence and viral outcomes. These participants were investigated in more detail and were all early drop-out with little EAM data. A further 1 participant was excluded on the basis of not recording observations for tuberculosis exposure at baseline - a covariate used in the modelling procedure. This left a final sample of 238 participants included in the analysis, excluding a total of 12 individuals.

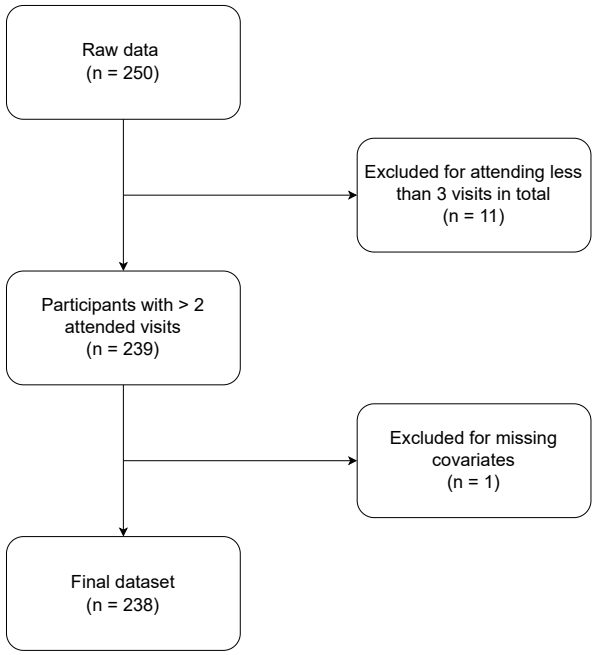


Figure 2: Flow diagram of ADD-ART analysis sample

Additional notes on missing data across these 238 participants which did not result in a further loss of sample size are as follows: (i) body-mass index (BMI) was collected through a protocol amendment mid-way through the study and only on 195 participants (only 85 of these were collected by visit 1); (ii) participants intermittently missed visits, resulting in incomplete longitudinal data. The former point resulted in BMI being excluded from formal analyses and the second point is handled via the estimation procedure for all models.

Another necessary data processing step was the formatting of the time-to-viral non-suppression data in counting process format. This is necessary for modelling recurrent events of this type. This involved representing each participant's event as a row with an event indicator, and a start and stop time. The start time represents the time of the last event for this participant, and the stop time is the time at which each event was recorded (or the last observation time for censored points).

All scripts used to execute the data processing procedure can be found in the GitHub repository linked in the Appendix (see A.3).

6.1.2 EAM bias correction

The EAM data was received in long format, with observations consisting of a participant ID, a date and time, and a categorical definition of the type of signal. These are defined as *intake*, *heartbeat*, or *none*. *Intakes* signal pill box openings, *heartbeats* are device signals (pings) that are sent when an opening hasn't been recorded in a certain amount of time, and *none* refers to days when neither of the aforementioned are recorded (no signal).

One potential source of bias with EAM device data is that pill box openings may not always signal dosages. Some patterns of openings can reveal this and be corrected in the data management phase, while some are not so easily corrected. Curiosity openings [55], for example, can sometimes be detected when many openings are recorded in a short space of time. Furthermore, as stated in the study protocol [79], the majority of participants were on a single-pill, single daily dose ART regimen. To guard against biasing adherence towards overestimation, it is thus reasonable to assume that only one opening can be considered as a dosage on any given day. The EAM data was thus filtered so that only one intake per participant per day could be recorded.

Technical bias could also be introduced if many *heartbeat* signals are recorded on one day, or if a a *heartbeat* signal was recorded on the same day as in *intake*. A few more adjustments were thus possible to guard against bias in the EAM adherence: only one heartbeat per day was allowed (ignoring this would lead to an overestimation of the number of missed doses and thus an underestimation of adherence); and, if an opening and a heartbeat were recorded on the same day, then the heartbeat was removed (ignoring this would mean that the opening would be effectively 'cancelled out' by the heartbeat, underestimating adherence).

While the EAM device data was recorded more or less continuously, the clinic data could only be recorded if and when a participant attended their monthly clinic visit. As a result, there are situations where a participant missed their clinic visits, for example for two months, but their EAM device was still in use. In these cases, longitudinal clinic data such as viral load and drug concentration measurements were not available but there were still two months of EAM data. The participant's adherence at the missed visit(s) could thus be estimated by imputing a visit date 30 days after the last attended visit, and then aggregating the EAM data over the 30 days preceding this date. This strategy allowed the retention of more longitudinal EAM data than would have been possible if this was ignored.

Lastly, EAM data is further complicated by the fact that, on the occasional day, no signal is received from the EAM device due to technical problems such as battery death (the previously mentioned *none* type observation). This leads to varying exposure times between individuals and between visits. An offset was thus defined as the number of signals (*heartbeats* + *intakes*) received in the previous 30 days, which accounts for the varying time-at-risk for non-adherence between individuals and visits.

6.2 Descriptive statistics

Many features were recorded at baseline in the ADD-ART sample, as detailed in Section 5.1.2. These include factors relating to socio-demographic, psychological and self-reported medical history characteristics. In addition, some clinical data was recorded including baseline tenofovir diphosphate (TFV-DP in DBS), emtricitabine (FTC) and hematocrit levels (HCT), and viral load.

Table 1 summarises baseline demographic information for participants in the ADD-ART sample. The average age was around 34-35 years, with a minimum of 18 and a maximum of 72 years. 22% of the participants were men. The sample almost exclusively consisted of South African born (97%) people who self-identified as African (98%). On average, the individuals reported four people living in their household but with a fair amount of variation between subjects. Around 78% of the sample reported having at least one child, with an average of two and a maximum of eight. Less than a third of the sample reported being employed and 26% reported often experiencing at least some food insecurity in their household. The mean maximum educational attainment was reflects a grade 4-5 level, though with much between subject variation. Overall, the information is nearly complete - though one individual had data for neither race nor education.

Table 1: ADD-ART Descriptive Statistics for Demographic Variables

Variable	**n**	**Min**	**Max**	**Mean**	**Median**	**SD**
Age	250	18	72	35.52	34.00	10.42
Sex: Male	250	0	1	0.22	-	-
Birth country: SA	250	0	1	0.97	1	-
Race: African	249	0	1	0.98	-	-
Household size	250	1	13	4.2	4	2.46
Have children	250	0	1	0.78	1	-
Number of children	194	1	8	2.1	2	1.2
Employed	250	0	1	0.31	0	-
Maximum Education	249	1	13	4.93	4	3.44
Any food insecurity	250	0	1	0.26	0	-

Table 2 presents some baseline clinical data for the ADD-ART sample. 236 individuals (94%) were virally suppressed at baseline ($VL < 50$ *copies/ml*). Of the 14 virally non-suppressed at baseline, the viral loads ranged between 55 and 1000, with a mean of 230 *copies/ml*. Baseline TFV levels ranged between 12.5 and 3504 *fmol/punch*, with a mean of 1160 *fmol/punch*. The median FTC level was around 0.32. The upper quartile for FTC was around 0.42, but the maximum was 12.5. The upper limit here was likely a flag value for the maximum quantifiable limit, since it was the exact same number found at the lower limit of TFV and these observations were extremely outlying - though this could not be confirmed. All of this information as complete at baseline, but body mass index (BMI) was only recorded for 88 participants. Throughout the study, attempts were made to collect further BMI observations which enabled the variable to be recorded for a total of 198 subjects (80%) in the sample. This first recorded BMI ranged between around 17 and 76 kg/m^2, with a mean of around 29 kg/m^2.

Table 2: ADD-ART Descriptive Statistics for Clinical Measures

Variable	**n**	**Min**	**Max**	**Mean**	**Median**	**SD**
Virally suppressed	250	0	1	0.94	1	-
HCT	250	21.00	49.00	36.84	37.00	4.33
TFV	250	12.50	3504.21	1160.49	1094.96	528.37
FTC	250	0.12	12.50	1.35	0.32	3.39
BMI baseline	88	18.72	61.88	29.16	27.52	8.41
BMI first recorded	198	16.98	75.75	28.82	27.07	8.69

Table 3 summarises some self-reported baseline variables that relate to medical history. 249 participants reported the year in which they first tested positive for HIV. This ranged between 1980 and 2018, though 75% of the individuals reported a year greater than 2014 and 50% reported a year greater than 2015. Participants were asked to rate their overall health on a scale between 1 and 5, with 5 representing very good health. The average overall health rating was 3.77 and the median was 4. Only 7% out of 248 participants reported every experiencing an HIV related illness, though 18% of 249 and 34% of 250 reported every having had tuberculosis and anemia, respectively. Only 107 participants reported whether they had ever had a CD4 count, and of these 60% said they had. Only 58 reported their last CD4 count, which ranged between 6 and 8000 $cells/m^3$. The mean was 952 and the median 518 $cells/m^3$. 109 out of 154 individuals reported having ever had a viral load test, while the other 96 did not answer the question. Two measures of self-reported adherence were recorded: the number of missed doses and a subjective rating of 'overall adherence' in the past 30 days. The former ranged between 0 and 14 doses, with a mean of around 1 doses and a median of 0 missed doses. The subjective rating was on a scale of 1 (very poor) to 6 (excellent). The responses ranged from 2 to 6, with the average response of around 4-5 representing 'good' to 'very good' overall adherence.

Table 3: ADD-ART Descriptive Statistics for Self-reported Medical History Variables

Variable	**n**	**Min**	**Max**	**Mean**	**Median**	**SD**
Year first tested HIV positive	249	1980	2018	2015	2016	-
Overall health	250	1.00	5.00	3.77	4.00	-
Ever had HIV illness	248	0	1	0.07	-	-
Ever had TB	249	0	1	0.18	-	-
Ever had anemia	250	0	1	0.34	-	-
Ever had CD4 counted	107	0	1	0.60	-	-
Latest CD4 count	58	6	8000	952	518	1353
Ever had VL test	154	0	1	0.71	-	-
Missed doses in past 30-days	248	0	14	1.10	0	1.89
Rate overall adherence in past 30-days	248	2	6	4.41	4	-

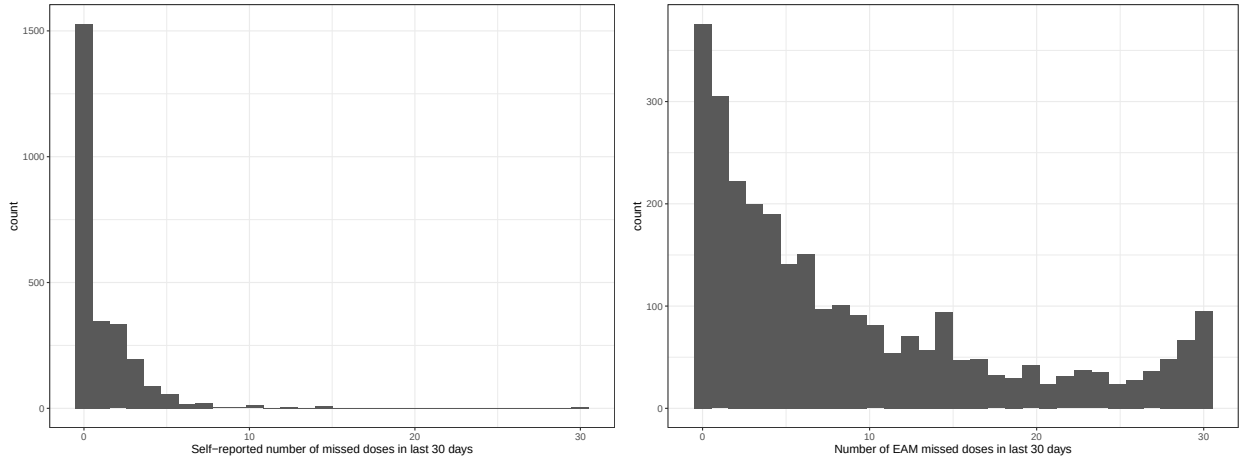
6.3 Outcomes

6.3.1 Adherence measures

In this section, the distributional properties and longitudinal profiles of the adherence outcomes as manifested in the ADD-ART sample are explored. These include: (a) the number of missed doses in the past 30 days according to self-report; (b) the number of missed doses in the past 30 days according to the EAM device; and (c) the concentration of tenofovir diphosphate as measured in dry blood spots. This exploratory analysis serves a few major purposes. The empirical distributions of the longitudinal outcomes help to guide the choice of model with respect to distributional assumptions and aid in identifying outlying observations. The longitudinal profiles help determine an appropriate functional form in terms of the progression of the mean adherence outcomes over time. The variation in outcomes within and between individuals, and over time also are important factors to explore when building regression models.

The histograms in Figures 3a and 3b shows the distribution of the number of missed doses in the previous 30 days across all visits, according to (a) self-report and (b) the EAM device. Considering the former, there are a few outlying responses recording around the maximum of 30 missed doses. This distribution, however, is clearly right-skewed with the vast majority of participants reporting fewer than five missed doses. Furthermore, the mass of the distribution is clearly concentrated around zero missed doses. Over 1500 of the 2631 ($\sim 57\%$) of the responses recorded across all visits were zero. The picture is very different when considering the latter figure. According to the EAM device, there was much more variation in the number of missed doses. Although the mode is still situated at zero, this distribution is much more evenly spread between that minimum and the maximum of thirty missed doses. While the median and upper quartile according to self-report are zero and two missed doses, the corresponding summary statistics are five and thirteen according to EAM. The empirical distributions illustrated by these figures suggest that

one may encounter some problems if attempting to frame the number of missed doses as being generated by a Poisson process. The self-reported distribution is clearly underdispersed due to zero-inflation, while the EAM distribution appears to be overdispersed at the upper tail.

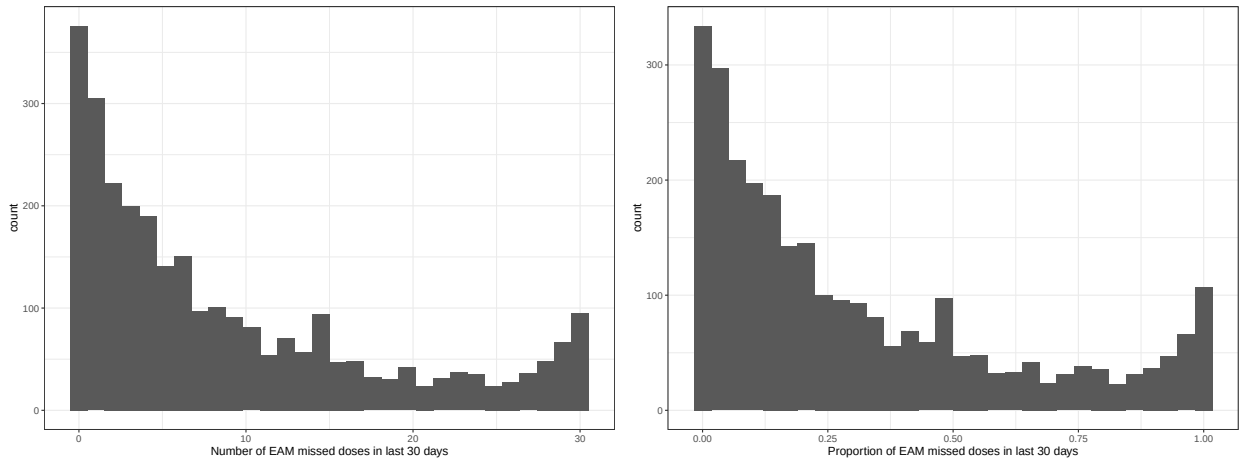


(a) Distribution of self-reported missed doses

(b) Distribution of EAM missed doses

Figure 3: Histogram showing the number of missed doses across all visits

As outlined in 6.1.2, the number of signals received from the EAM device varies between individuals and between visits. It is thus important to consider how impactful this potential source of bias is in the sample. One way of exploring this is to normalize the counts by the number of signals received in the last 30 days, to arrive at a non-adherence proportion, and examine how the distribution changes. The figures in 4 illustrate this, suggesting that there is not a dramatic influence of this source of bias. There are certainly subtle differences though. The number of zero counts is reduced, suggesting that some of the zeroes result from months where no signals were recorded from the device at all. Another difference is that the number of maximum missed doses increases slightly, suggesting that there were cases where, for example 27 missed doses were recorded but out of 27 total signals in the previous 30 days. These subtle sources of bias are thus indeed present in the sample and thus need to be accounted for in the modelling process.

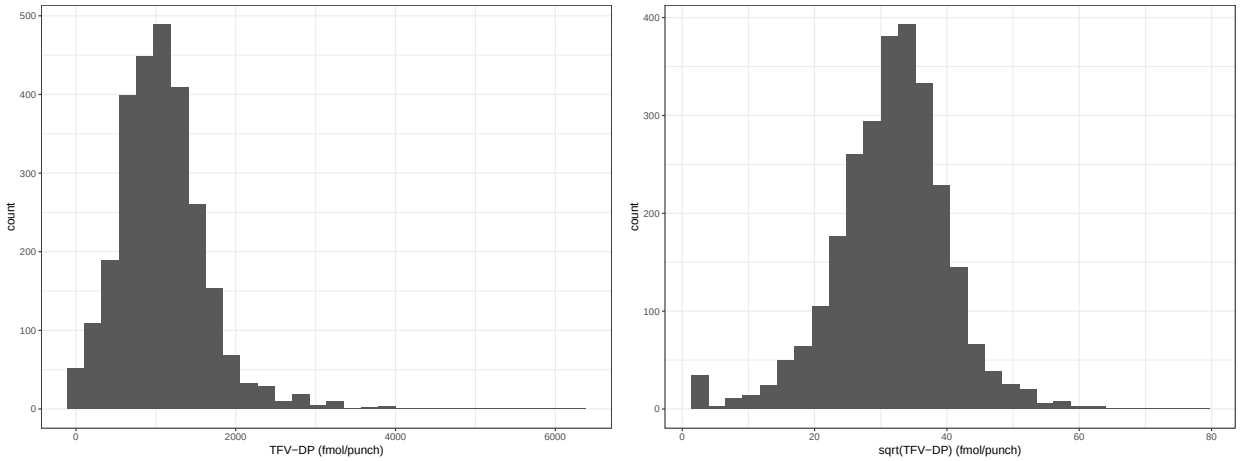


(a) EAM number of missed doses

(b) EAM missed doses proportion

Figure 4: Histograms comparing the distribution of EAM non-adherence before and after normalization

The histogram in Figure 5a presents the distribution of TFV-DP concentrations as measured in DBS over all visits. The minimum quantifiable limit of this assay results in a sharp lower tail cutoff at 12.5 *fmol/punch*, while there is a noticeable outlying observation extending the range to 6285 *fmol/punch*. The drug concentrations are right-skewed, with a mean and median of 1074 and 1036 *fmol/punch*, respectively. Figure 5b shows the distribution after a square-root transformation. This brings more symmetry to the distribution, and further highlights the censoring induced by the lower limit of quantification.

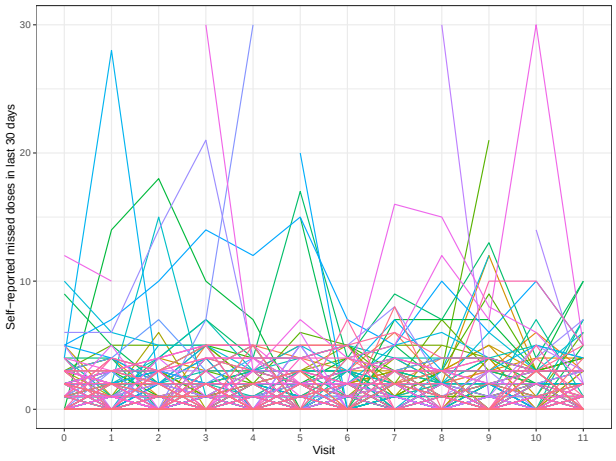


(a) Distribution of TFV-DP

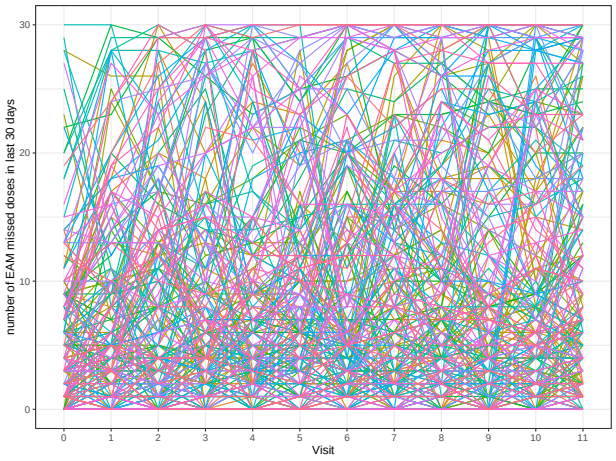
(b) Distribution of square root of TFV-DP

Figure 5: Histogram showing the distribution of TFV-DP as measured in DBS across all visits

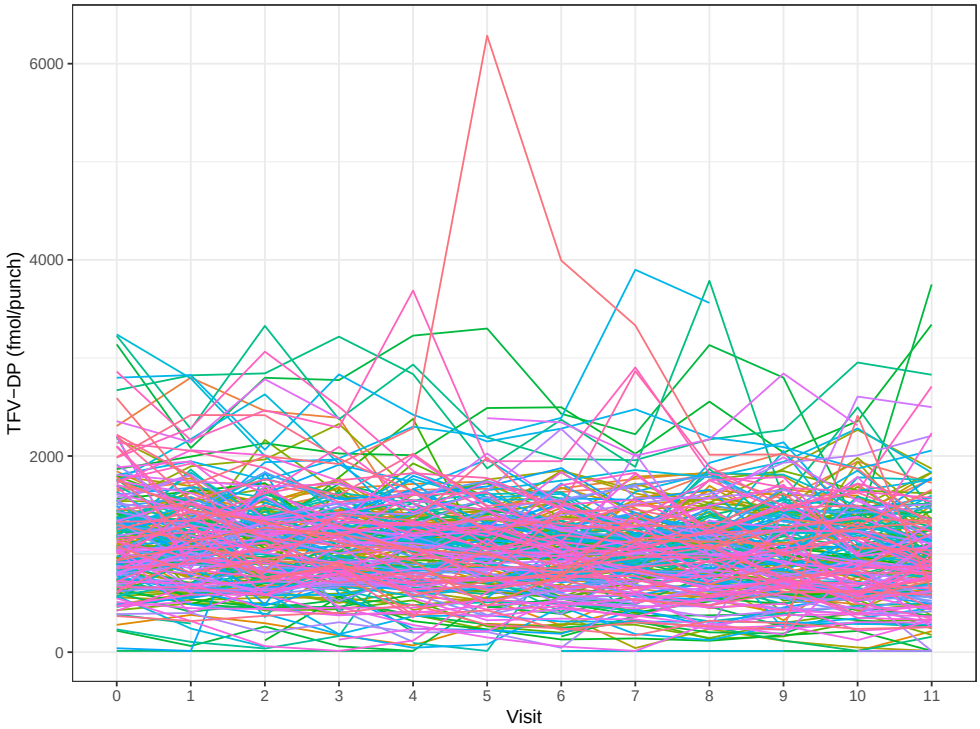
Next, attention is turned to the longitudinal nature of the data. For each participant, the adherence measures were recorded at each of the twelve visits which were attended or until loss-to-follow-up. For the EAM data, data was still recorded in many cases where visits where missed (see Section 6.1.2). Figure 6 presents the individual trajectories for all 238 participants included in analysis. Figure 6a presents the trajectories for self-reported missed doses. This shows that the majority of individuals consistently reported missing fewer than 5 doses at each visit, although the number of missed doses does seem to fluctuate between 0 and 5 missed doses over time. No participants reported consistently missing many doses - for the few observations above 15 missed doses, the tendency appears to be that this is not sustained. Considering the number of missed doses according to EAM data (Figure 6b), it seems that there is much more variability between individuals' trajectories and over time when compared to self-report. Finally, Figure 6c presents the individual longitudinal TFV-DP levels in DBS. As seen through the histograms, the distribution of drug concentrations is much more symmetrical at each visit. While difficult so see how the center of the distribution behaves over time here, it is interesting that the TFV-DP observations at the tails of the distribution appear to persist over time. In other words, an individual who has relatively high drug levels at one visit, tends to have relatively high concentrations at the next. This being said, there is still significant variability within individuals, as well as between individuals and between visits. Regarding the obvious outlying drug concentration seen at visit 5 - this was checked with subject matter experts and confirmed to be within a possible range of values, and thus is likely not a result of measurement error.



(a) Self-reported missed doses over time



(b) EAM missed doses over time

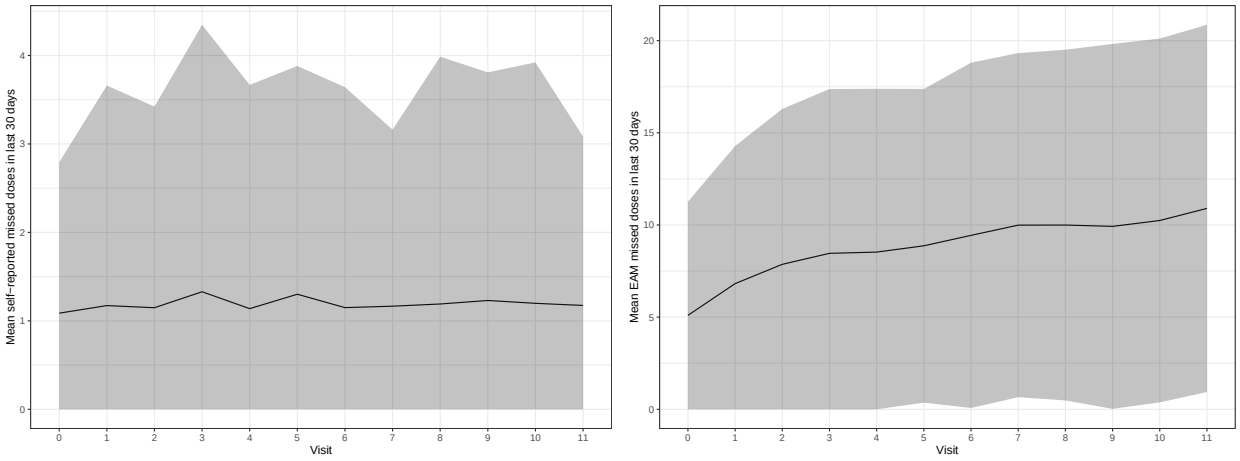


(c) TFV-DBS over time

Figure 6: Individual adherence trajectories

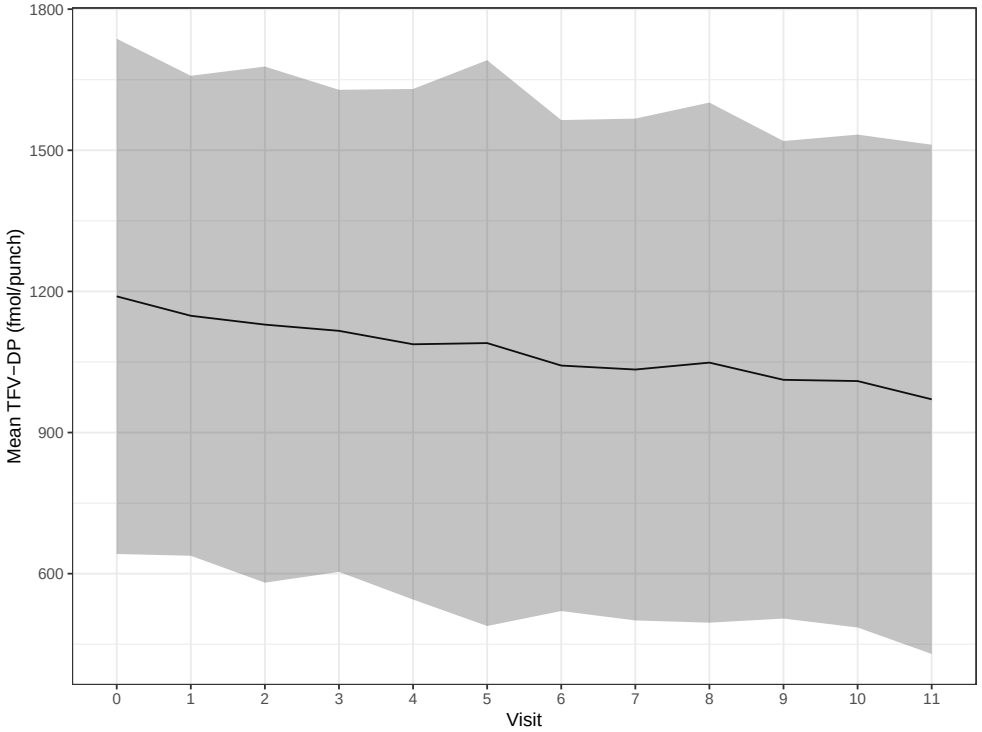
While the individual trajectories are useful, there are simply too many participants to explore aggregate behaviour in this sample with any degree of confidence. Figure 7 thus presents line plots of the per-visit sample means for each adherence outcome. Additionally, the variability is represented by the shaded area - which represent one sample standard deviation from the mean (with a lower bound of zero). Figure 7a reveals that the mean self-reported missed doses were consistently slightly above 1 at every visit. It also shows that there was very little between-individual variability, with the shaded area being bound between zero and 4 missed doses for all but the third visit. By contrast, the mean number of missed doses according to the EAM data (Figure 7b) seems to trend upwards over time. Initially, the mean number of missed doses is around 5 and climbs towards 11 doses at visit 11. This upwards trend is not strictly linear, and appears

to level off between visits 3 and 4, and visits 7 and 9. This mean trajectory could likely be approximated by a low order polynomial of time. The shaded area shows that the sample standard deviation is much higher when compared to the self-reported data and looks to be between five and ten units depending on the visit. Lastly, the mean profile of the TFV-DP in DBS levels is downward trending over time. The drug concentrations measure adherence, while the number of missed doses measure non-adherence. These two mean profiles thus seem to agree with each other, both indicating an average decrease in adherence over the course of the study but with much between-individual variability. One important note here is that the use of the sample standard deviation to represent variability could potentially be misleading if not interpreted with caution. As seen previously, all of the above distributions are right-skewed and thus the variability is not evenly distributed on either side of the mean but dependant on the degree of skewness.



(a) Self-reported missed doses

(b) EAM missed doses

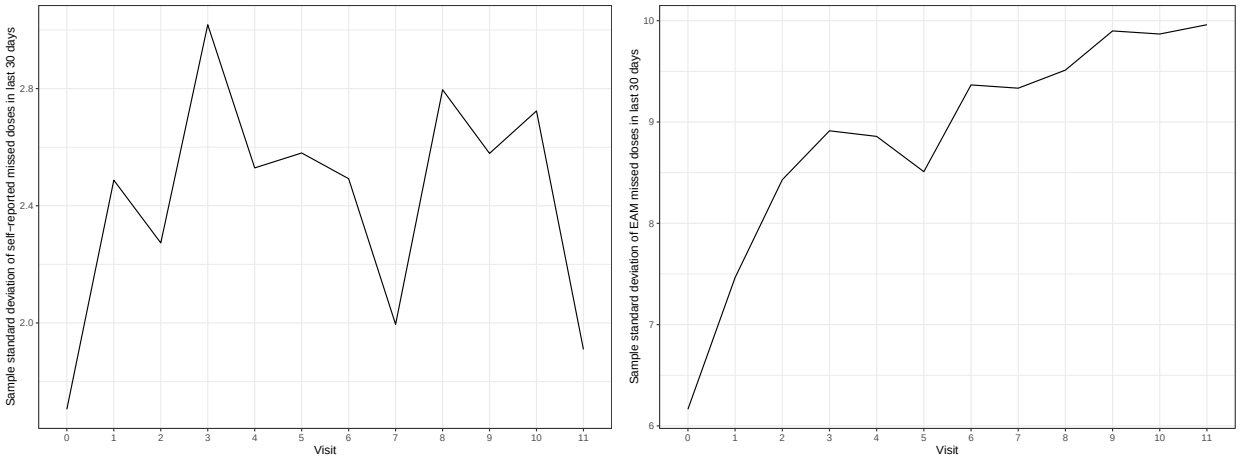


(c) TFV-DP in DBS

Figure 7: Mean adherence trajectories with shaded area showing 1 standard deviation from mean

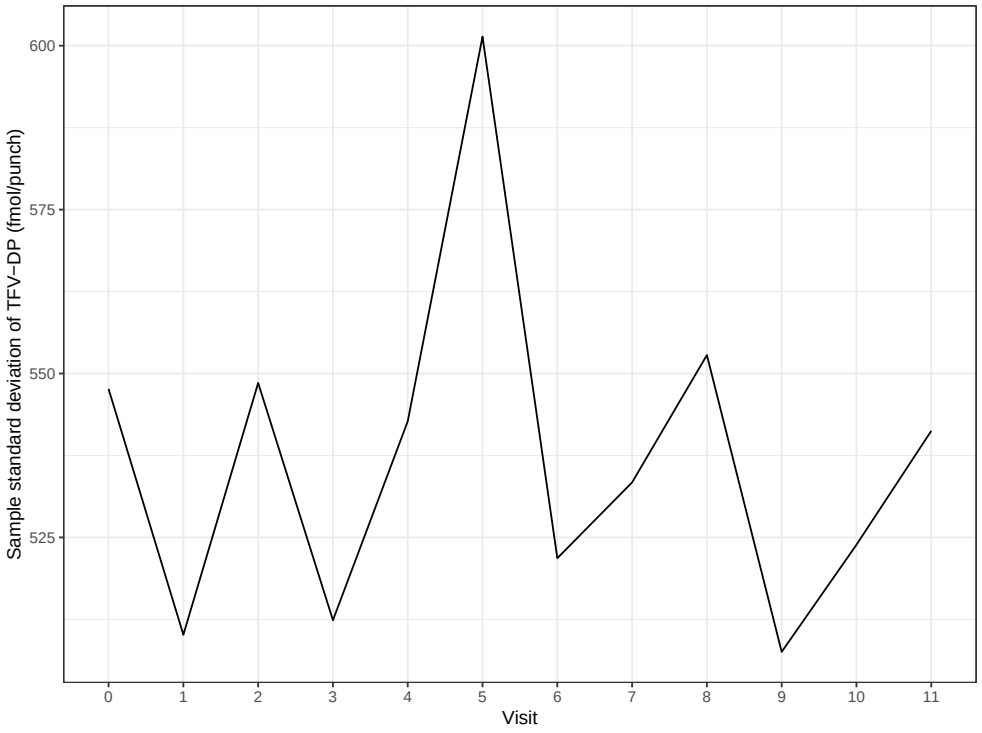
Another important consideration is the longitudinal nature of variability in this sample. This is primarily because many regression models assume homoskedasticity and non-constant variance over time can violate this assumption if the additional variability is not accounted for in the model specification. Figure 8 thus illustrates the sample standard deviation at each visit in each of the adherence outcomes. Considering Figure 8a, the standard deviation of self-reported missed doses appears to fluctuate over time without any apparent pattern - although not by a large magnitude. It is in fact clear that the 'spikes' in variability at visits 1, 3, 6 and 10 correspond to visits with large outlying values (see Figure 6a). The sample standard deviation of EAM missed doses, Figure 8b, clearly trends upwards - though not monotonically. This suggests that, as the study progressed, the number of missed doses varied more and more between-participants. While

important to note and consider, this does not necessarily present a problem. As touched on previously, a fundamental characteristic of Poisson processes is the assumption that the mean equals the variance. As seen previously, the sample mean for these counts do trend upwards so one would expect to see the variance trend upwards if the missed doses are generated by a Poisson process (although the sample standard deviation summarises many sources of variability). Lastly, Figure 8c shows the sample standard deviation of the DBS TFV-DP levels at each visit. This clearly fluctuates between visits with no consistent pattern. There is a large deviation from the range at visit 5, which is clearly a result of the large single outlying observation previously noted (see Figure 6c).



(a) Self-reported missed doses

(b) EAM missed doses



(c) TFV-DBS

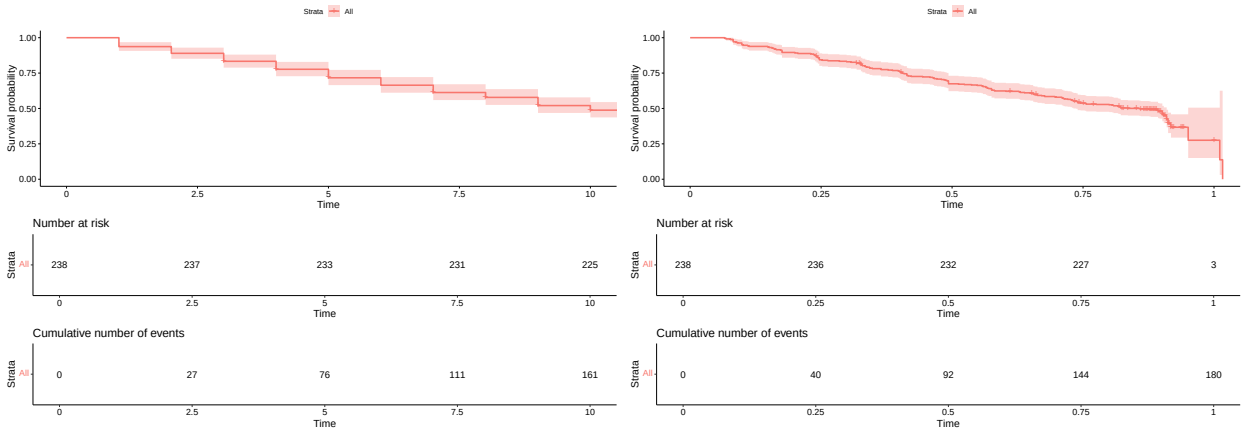
Figure 8: Sample standard deviation profiles of adherence measures at each visit

6.3.2 Viral non-suppression

In this section, the time to the repeated viral non-suppression events among the ADD-ART sample is explored. Firstly, Kaplan-Meier curves are presented to illustrate the survival experiences across the whole sample. Importantly, the data was structured in counting process format and the Kaplan-Meier curves constructed under the recurrent events framework. The survival probabilities are assessed with respect to the discrete visit-time as well as continuous time-in-years from Visit 1, since this information is available and it provides more granularity. The continuous time-in-years differs because the discrete visits occur at different intervals for each individual, due to the nature of the study. Figure 9 presents these Kaplan Meier curves (with complementary log-log 95% confidence intervals), as well as tables that indicate the number at risk and cumulative number of events at a selection of time points.

Figure 9a presents the former, where one can see that the events are considered to occur at the 11 discrete visit times. At the first visit, all the participants are at risk and there had been no events. At each subsequent visit, a number of events are recorded which causes the survival probability to decrease in a step-wise fashion. As indicated by the plot, around half of all participants had experienced at least one non-suppression event by visit 10. This suggests a median survival time of around 10 visits. Since the events are recurrent, individuals may remain in the risk set until drop out or the end of the study, and thus the number at risk does not decrease linearly with the number of events. By Visit 10, 161 events had occurred and 225 individuals remained at risk - clearly indicating that (a) 225 individuals made it to visit 10 without dropping out, and (b) that many individuals experienced multiple events. In fact, of those who experienced a non-suppression event, the mean number of events was 2.4. Across all participants, the mean number of events was around 0.75.

Figure 9b, on the other-hand, illustrates estimates of the survival probabilities given continuous time since Visit 1. This provides more granularity, since the time between visits varied between individuals. Despite this, events are still clustered in pockets that correspond to the discrete visits, approximately 1 month apart, resulting in steps in the survival probability. The median survival time, according to this curve, is around 0.8-0.9 years from Visit 1 ($\sim$ 10 months). There is thus agreement here to the discrete visit view - a median survival time of around 10 months from the first visit. Besides the granularity, the main difference between these approaches to generating the Kaplan-Meier estimates can be seen towards the ends of the curves. Unlike with the discrete visit-time, the survival probability diminishes to zero with the continuous time. This is because only a handful of participants had their last visit more that 1 year after their first, and thus most of the individuals fall out of the risk set. This should thus not be interpreted as a literal survival probability.



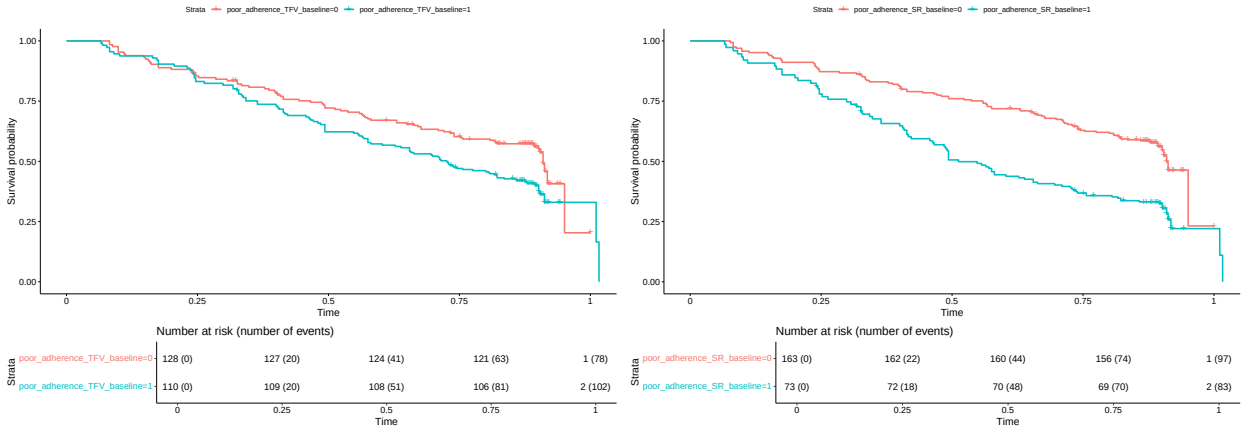
(a) Discrete visit time

(b) Continuous time since visit 1

Figure 9: Kaplan-Meier curves for repeated time-to-viral non-suppression

It is important to investigate potential associations between survival times for repeated viral non-suppression and baseline covariates in the data. As presented in the introductory chapter (Section 2), viral non-suppression is broadly thought to be influenced by pre-treatment resistance, acquired resistance, and pharmacokinetic/dynamic related factors. No genetic information (viral nor host) is available for this sample. There is also no data on either previous exposure to ARVs or concurrent use of potentially conflicting medications. Pre-treatment resistance and pharmacokinetic/dynamic factors can thus not be included in the analysis, and it is possible that variation in the sample along these missing variables may confound the results.

It is, however, possible to investigate the association between the survival probabilities and adherence - as measured through the different monitoring mechanisms. Firstly, the sample was stratified by baseline adherence according to: (a) the median of TFV-DP DBS concentrations at baseline; and (b) self reported missed doses at baseline (in the past 30 days). Respectively, individuals were classed as poorly adherent if: (a) their baseline TFV-DP DBS concentration was less than the median (1060.4 *fmol/punch*); their baseline self-reported missed doses was more than the upper quartile of missed doses at baseline (2 missed doses). There was no EAM data at baseline, so that factor could not be assessed. Kaplan-Meier curves were constructed under these strata, presented in Figures 10a and 10b. Log-rank tests could not be used to formally assess differences in survival distributions given the recurrent nature of the events. Despite this, even these crude measures provide an indication that poor adherence at baseline, represented by the blue curves, was associated with more rapidly deteriorating survival probabilities. Considering the strata according to baseline TFV-DP concentrations, the survival probabilities appear to start diverging after around 0.25 years from the first visit. Those classed as poorly adherent at baseline, according to TFV-DP levels, had a median survival time of around 0.75 years compared to around 0.85 years for those classed as adherent. The strata according to baseline self-reported missed doses suggest a similar pattern, although the survival probabilities diverge almost immediately and more dramatically. Under these strata, the poorly adherent had a median survival time of around 0.5 years compared to around 0.85 for the adherent group. It is worth noting that the strata according to each measure are not directly comparable, since one may be a more lenient definition of poor adherence, and both are pretty crude. It should also be noted that the strata according to self-report were not balanced, since the upper quartile was used as a cut-off. This does mean the probabilities could be impacted by uneven risk sets, so the Kaplan-Meier curves should be interpreted with this in mind.



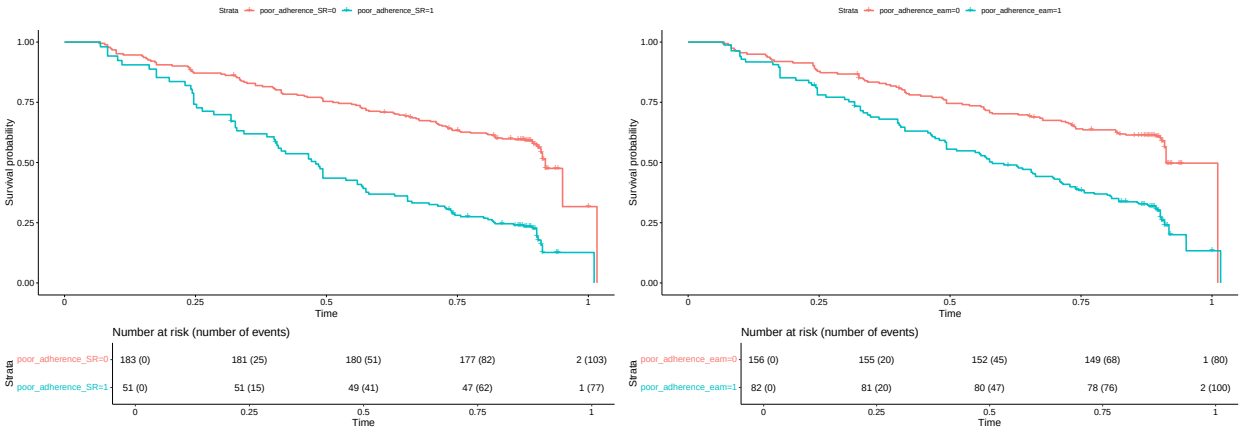
(a) Stratified by baseline TFV-DP DBS measure

(b) Stratified by baseline self-reported missed doses

Figure 10: Kaplan-Meier curves for recurring time-to-viral non-suppression for sample strata

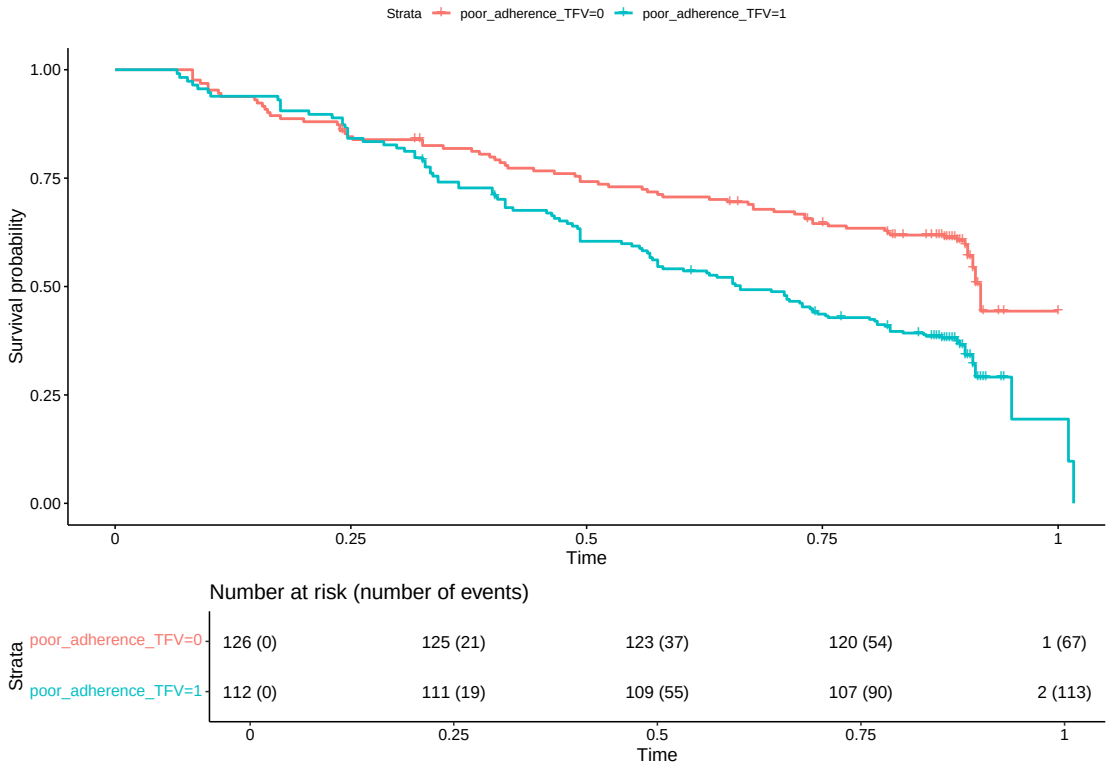
One can also stratify the sample by their aggregate adherence over the course of the study. By summarising each individual's longitudinal adherence data, and stratifying the sample according to the distribution of this, one can compare those who had poor adherence longitudinally with those who had better adherence. To enable this, each individual's mean number of missed doses according to self-report and the EAM device,

as well as mean TFV-DP concentrations were computed. Individual's longitudinal adherence was then categorized, according to each measure, depending on their mean adherence in relation to the global mean (across all individuals and all time points). For TFV-DP, poor adherence was defined for those below the global mean, and for the number of missed doses, this was defined as above the global mean. The specific values used to categorize are indicated in Figure 11, which shows the survival curves according to these strata for each measure of adherence. The non-adherent strata are again represented by the blue curves, suggesting that longitudinal non-adherence was indeed associated with more rapidly deteriorating survival probabilities. The median time to viral non-suppression for the non-adherent groups were around 0.45, 0.55, and 0.6 years according to self-reported missed doses, EAM missed doses and TFV-DP concentrations, respectively. The corresponding median survival times for the adherent groups were much larger, according to all of these measures, and in fact not reliably estimable because far fewer events had occurred by the end of the observation period.



(a) Mean self-reported missed doses > 2.2

(b) Mean EAM missed doses > 12.2

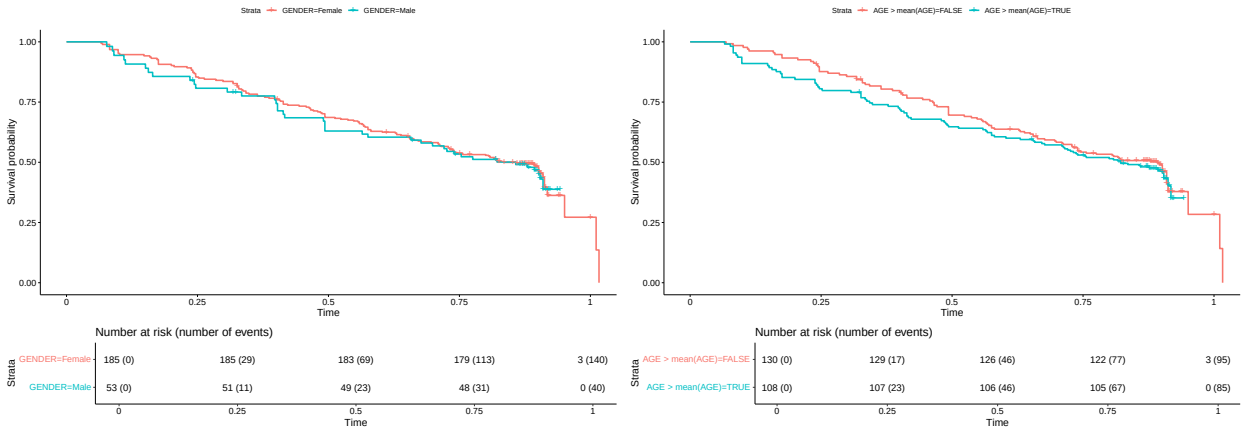


(c) Mean TFV-DP DBS $< 900.2 fmol/punch$

Figure 11: Kaplan-Meier curves for the sample stratified by average longitudinal adherence

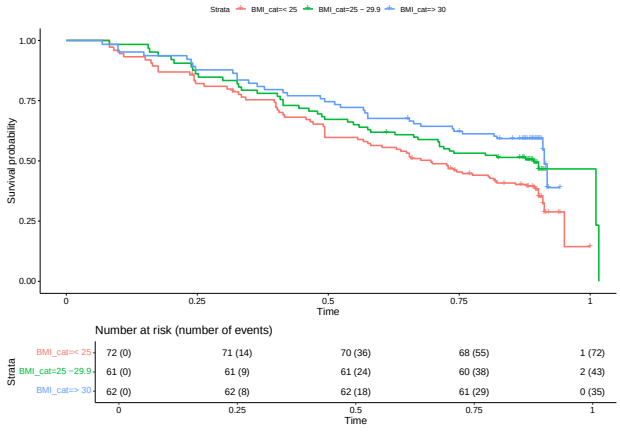
Similar exploratory analyses as those implemented above for adherence metrics, were also implemented for the sample stratified by a selection of baseline demographic and medical history information. Figure 12a shows the curves for males and females. The curves do not diverge in any significant manner, providing no indication that the sexes had differences in their survival distributions. Figure 12b shows the sample stratified by the mean age of 35.7 years. The group comprising of those older than the mean do experience more events initially. They thus have lower survival probabilities over much of the observation period, although the difference is not dramatic. Lastly, Figure 12c shows the sample stratified by three category levels of baseline Body Mass Index (BMI). As previously noted, BMI was collected through an amendment

to the study protocol - and thus as inconsistent times and with incompleteness. The 195 individuals who did have BMI recorded, were categorized into three levels. The sample was stratified by these categories, and Kaplan-Meier estimates generated - as visualized in Figure 12c. The curves suggest that there may have been a difference in the survival times according to BMI, with a lower BMI being associated with lower probabilities of viral-suppression over time.



(a) Stratified by sex

(b) Stratified by mean age (35.7 years)



(c) Stratified by BMI

Figure 12: Kaplan-Meier curves for the sample stratified by selected demographic factors

Another consideration is that the time-to-viral non-suppression may be affected by differences in baseline viral loads. Figure 13 provides an illustration of the survival probabilities for the sample stratified by baseline viral suppression status. For those with unsuppressed viral loads (n = 13), plotted in blue, there appear to be proportionally more events early on in the observation window compared to the rest of the sample. This results in more rapidly deteriorating survival probabilities for the unsuppressed sub-sample. These strata are extremely unbalanced though, so this should be interpreted with caution.

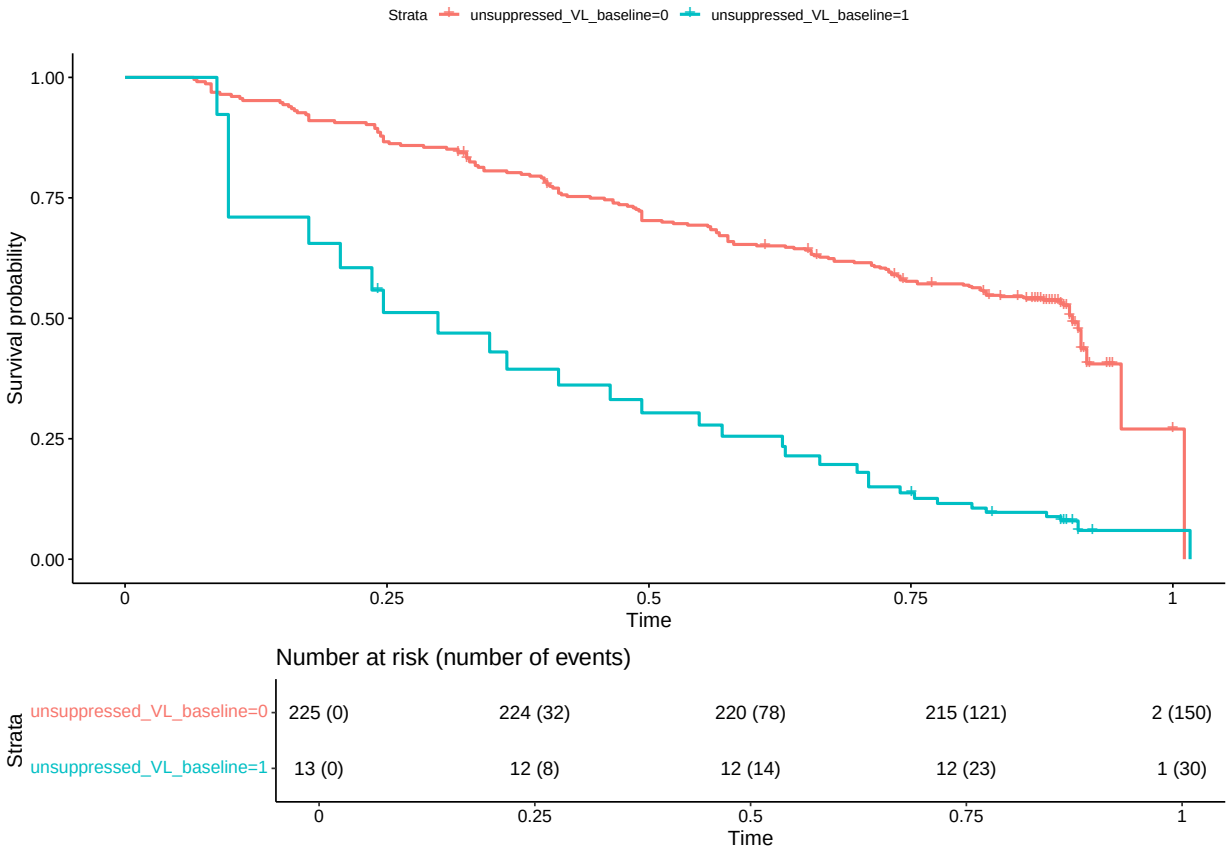
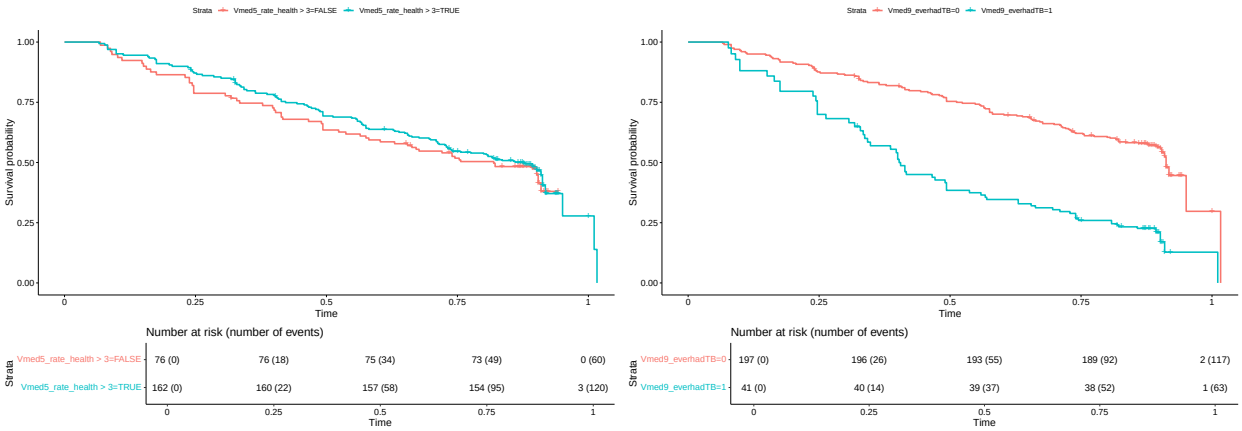


Figure 13: Kaplan-Meier curves for the sample stratified by baseline viral suppression status

As a final exploration into stratified survival times, Figures 14a and 14b present Kaplan-Meier estimates for the sample according to two self-reported factors that relate to medical history. The former asked the participants to rate their overall health at baseline on a discrete scale from 1 to 5, with higher numbers indicating better health. The blue curve represents those who rated their health highly (n = 76), while the orange curve represents all else (n = 162). The survival probabilities were estimated to be greater for the 'healthier' participants throughout the observation window, although the differences are relatively small in magnitude. Figure 14b asked participants to disclose whether they had ever been diagnosed with Tuberculosis (TB). Only 41 of the 238 responses indicated that they had previous TB exposure, shown by the blue curve. Although the risk sets were imbalanced, the Kaplan-Meier estimates indicate that these individuals experienced proportionately more events - resulting in more rapidly deteriorating survival probabilities. The median survival time for this group was around 0.4 years compared to around 0.85 years for the other participants.



(a) Overall health rating (1-5)

(b) Tuberculosis exposure

Figure 14: Kaplan-Meier curves for the samples stratified by selected self-reported medical history factors

7 Main results

7.1 Marginal survival models

7.1.1 Model building

In this section, applications of the marginal Cox model for recurrent events to the ADD-ART sample are presented. The recurring survival times are primarily conditioned against each measure of adherence. Given the available data, and the insights gained from the exploratory analysis, a selection of baseline covariates are chosen to include in the model building procedure. A forward selection procedure was implemented, whereby additional fixed effects were retained if they were strongly related to the outcome. This informal strategy was adopted instead of likelihood ratio tests since the marginal models are computed with robust variances, negating the possibility of using these formal hypothesis tests. The final chosen adjusted models are presented in Tables 4 - 6.

Firstly, Cox-PH models which condition against mean observed adherence over time are considered. This model follows the structure presented in 4.2.1, with the model specified as:

$$h_i(t|\bar{y}_i, \mathbf{w}_i) = h_0(t) \exp^{\alpha \bar{y}_i + \mathbf{w}^T \boldsymbol{\gamma}}, \tag{31}$$

where the hazard function for the i-th individual, h_i, is conditioned on their observed mean adherence over time, $\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$, as measured by method y; and a vector of other covariates $\mathbf{w}_i$.

This method is then extended to follow the approach presented in 4.2.2, where the hazard function is conditioned on the time-varying observed adherence. Under this method, the model is specified as:

$$h_i(t|y_i(t), \mathbf{w}_i) = h_0(t) \exp^{\alpha y_i(t) + \mathbf{w}^T \boldsymbol{\gamma}}, \tag{32}$$

where $y_i(t)$ denotes the observed adherence according to monitoring method y at t.

The population-level mean factor change in the instantaneous risk of viral non-suppression associated with a one unit increase in the observed mean adherence measure y over time is thus represented by $\exp^{\alpha}$ in equation 31; while the corresponding association with the observed value of y at t is represented by $\exp^{\alpha}$ in equation 32.

Estimates of the exponentiated coefficients, along with 95% confidence intervals, are presented in Tables 4, 5 and 6. These tables present the model results for the hazard as a function of: (a) mean adherence; (b) time-varying adherence; and (c) time-varying adherence, adjusted for the square-root of baseline viral load and the binary self-reported factor indicating TB exposure. Consider the results associated with the number of missed doses in the last 30 days according to self-report, as presented in Table 4. The results from model *(a)* indicate that there is evidence to support an association between the mean number of self-reported missed doses over time and the hazard of viral non-suppression. In particular, there is an estimated 15.3% increase in the average hazard of non-suppression for each unit increase in mean missed doses per 30 days (HR = 1.153, 95% CI: [1.093; 1.216]). Considering the estimates associated with the time-varying SR missed doses in model *(b)*, there is an average increase in the hazard of non-suppression of around 9.5% for each unit increase in self-reported missed doses at the time of the event (HR = 1.095, 95% CI = [1.057; 1.133]). The corresponding estimate, after adjusting for differences in the baseline viral load and exposure to TB, is around 8.7% - as evidenced by model *(c)* (HR = 1.087, 95% CI = [1.047; 1.129]).

Table 4: Cox model results: Self-reported missed doses

Model name	Variable(s)	Coef	e^{coef}	Robust SE	p-value	e^{coef} 95% CI
(a) SR mean	$\bar{y}$	0.142	1.153	0.027	< 0.001	[1.093; 1.216]
(b) SR time-varying	$y_i(t)$	0.09	1.095	0.018	< 0.001	[1.057; 1.133]
(c) SR time-varying adjusted	$y_i(t)$	0.084	1.087	0.019	< 0.001	[1.047; 1.129]
	sqrt(Baseline VL)	0.094	1.098	0.012	< 0.001	[1.072; 1.125]
	everhadTB	0.683	1.979	0.270	0.011	[1.166; 3.359]

Table 5 presents the Cox model results associated with the number of missed doses in the last 30 days according to EAM. Model *(a)* suggests that there is an average increase in the hazard of non-suppression of around 5% for each unit increase in the mean number of missed doses over time (HR = 1.05, 95% CI = [1.027; 1.073]). According to model *(b)*, the hazard increases by an average factor of 1.043 per unit increase in 30 day missed doses (95% CI = [1.023; 1.064]). Model *(c)* estimates this factor change to be around 1.035 after adjusting to the effects of baseline viral load and TB exposure (95% CI = [1.014; 1.057]). Considering these three models, there is significant evidence to suggest that on average the instantaneous risk of viral non-suppression was positively associated with an increasing number of EAM missed doses in the past 30 days.

Table 5: Cox model results: EAM missed doses

Model name	Variable(s)	Coef	e^{coef}	Robust SE	p-value	e^{coef} 95% CI
(a) EAM mean	$\bar{y}$	0.048	1.05	0.011	< 0.001	[1.027; 1.073]
(b) EAM time-varying	$y_i(t)$	0.042	1.043	0.01	< 0.001	[1.023; 1.064]
(c) EAM time-varying adjusted	$y_i(t)$	0.034	1.035	0.011	0.001	[1.014; 1.057]
	sqrt(Baseline VL)	0.064	1.066	0.010	< 0.001	[1.046; 1.086]
	everhadTB	0.922	2.514	0.256	< 0.001	[1.523; 4.150]

The results associated with TFV-DP concentrations as a measure of adherence are presented in Table 6. Given the large range of TFV values (see Figure 5a), it is crucial to note that a unit change in TFV-DP levels is not very meaningful. Therefore, in this case, the relative hazard is regressed against a transformation of the drug levels. Specifically, adherence (y) in equations 31 and 32 is represented by the values of $log_{10}(TFV)$. The coefficients thus represent associations with a ten-fold change in TFV-DP levels and not unit changes.

Another important distinction from the missed dose models above: increases in TFV-DP concentrations measure adherence and not non-adherence. If one exponentiated the negation of the coefficient, this would correspond to the factor change in the hazard of non-suppression associated with a ten-fold decrease in TFV-DP concentrations - a measure of non-adherence. Following this logic, model *(a)* suggests that the hazard of non-suppression is increased by an average factor of around 2.375 for each ten-fold decrease in mean TFV-DP concentration over time (HR 95% CI = [1.525; 3.699]). A ten-fold decrease in TFV-DP at the time of event is estimated to increase the hazard of non-suppression by around 2.379, according to model *(b)* (HR 95% CI = [1.740; 3.254]). This estimate is very close to the former, but is estimated with more precision. After adjusting for baseline viral load and TB exposure, model *(c)* estimates this factor change in the hazard at around 2.088 (HR 95% CI = [1.465; 2.977]).

Table 6: Cox model results: TFV-DP in DBS concentration

Model name	Variable	Coef	e^{-Coef}	Robust SE	p-value	e^{-Coef} 95% CI
(a) $log_{10}(TFV)$ mean	$\bar{y}$	-0.865	2.375	0.226	< 0.001	[1.525; 3.699]
(b) $log_{10}(TFV)$ time-varying	$y_i(t)$	-0.867	2.379	0.160	< 0.0001	[1.740; 3.254]
(c) $log_{10}(TFV)$ time-varying adjusted	$y_i(t)$	-0.736	2.088	0.181	< 0.0001	[1.465; 2.977]
	sqrt(Baseline VL)	0.060	0.942	0.010	< 0.0001	[0.923; 0.960]
	everhadTB	0.992	0.371	0.252	< 0.0001	[0.226; 0.607]

7.1.2 Model diagnostics

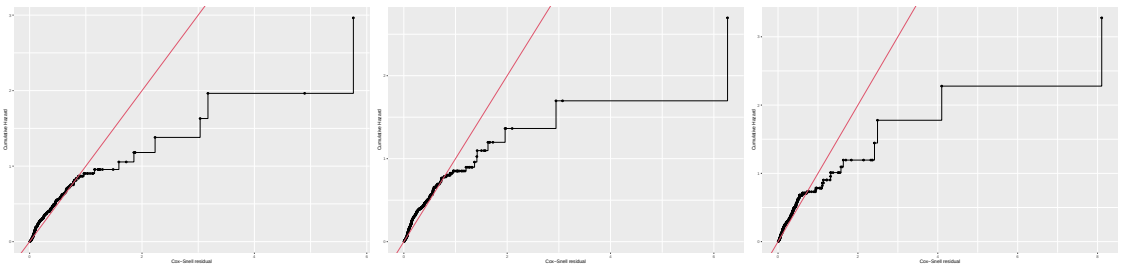
In this section, model diagnostics relating to model *(c)* associated with each of the adherence measures are presented. The PH assumption is assessed graphically via a plot of the scaled Schoenfeld residuals against follow-up time along with a formal test for a non-zero slope. Cox-Snell residual plots are used as a diagnostic tool to evaluate overall model fit. In a standard Cox-PH model one would investigate the functional form of continuous covariates using diagnostic plots, and this can be assessed for the baseline viral load and the adherence covariates. Lastly, diagnostic plots using deviance residuals are used to identify outlying observations and score residual plots are used to identify influential outlying observations.

Firstly, the PH assumption is assessed for each of the three covariates in each of the three models. The graphical assessment is provided via Figures A.1 - A.3 in Appendix A.2.1, while Table 7 provides results associated with formal tests for a non-zero slope. A small p-value would suggest rejecting the null hypothesis of a zero-slope and imply that there is evidence to reject the PH assumption. Firstly, for the SR model, there is no clear trend over time for any of the three variables which is supported by non-significant test statistics for all variables and for the global test. This suggests that the PH assumption is not violated in this model. For the EAM model, the test statistic associated with the baseline viral load suggests that the marginal association is not constant over time which implies a violation of the PH assumption. This may suggest that the marginal impact of baseline viral load on the time to viral non-suppression is not constant over time, which may be an intuitive result since one would expect it to have a greater impact at times closer to baseline. Looking at the graphical representation, however, it seems that this result may be due to a few outlying observations towards the end of the observation period which 'pull' the trend line downwards. Furthermore, the residuals associated with the EAM counts, though not resulting in a significant p-value, do appear to slightly trend upwards over time. This may suggest that the effect of EAM adherence on time to viral non-suppression is not constant over time. Overall, the global test statistic indicates the PH assumption is not satisfied for this model. Lastly, for the TFV model, the PH assumption is also only violated for the baseline viral load covariate which results in the global test also being rejected.

Table 7: Formal test for assessing PH assumption in marginal models

	SR model (c)		EAM model (c)		TFV model (c)	
Variable	chi-square test statistic	p-value	chi-square test statistic	p-value	chi-square test statistic	p-value
y(t)	1.56	0.21	2.08	0.15	0.30	0.58
EverhadTB	0.08	0.78	0.04	0.85	0.05	0.83
sqrt(Baseline VL)	0.06	0.80	6.14	0.01	11.80	0.001
GLOBAL	1.99	0.58	11.04	0.01	14.00	0.003

The Cox-Snell residuals are plotted against the Nelson-Aalen estimate of the cumulative hazard of these residuals, for each of the three models, in Figure 15. Similar patterns are seen for all three models: most of the points appear to follow the 45-degree line reasonably closely, indicating a generally good overall fit of the Cox model to the data. There are clearly some deviations from this expected distribution at larger values of the residuals. Larger Cox-Snell residuals are associated with observed survival times which are longer than what the fitted model would suggest, and the model is thus underestimating observed survival probabilities for these observations. This may arise if there are influential observations distorting the estimated marginal effect of one or more covariates, or if there are omitted variables which would otherwise account for this unexplained heterogeneity. Another possible explanation may be due to inherent limitations of the Cox-Snell residuals. As shown by Klein and Moeschberger [48], deviations from the expected distribution can arise from uncertainty in the estimation of the coefficients and cumulative hazard. These uncertainties are largest in the upper tail of the distribution, a pattern that is consistent with these diagnostic plots.



(a) SR model c (b) EAM model c (c) TFV-DP model c

Figure 15: Cox-Snell residuals used to assess overall model fit

The functional form of the numerical covariates in each model is assessed through the deviance residual plots

provided in Figure A.4 in Appendix A.2.1. With regards to all three models, the residuals do not exhibit any clear trend when plotted against the adherence measure - indicating that the functional form provides an appropriately linear fit. With respect to the square root of baseline viral load, there is no alarming vertical trend in the residuals associated with this covariate across all three models and thus it is likely appropriately specified too.

Outliers can be investigated by assessing the deviance residuals plotted against the linear predictor, as shown in Figure A.5 for all three models. The plots do not suggest that there were any obviously outlying observations in the data. There is however a clear strip of negative deviance residuals which indicate a systematic overestimation of hazards. These observations were further investigated by colouring observations by censoring status, which revealed that that the majority of these points were associated with censored observations. This means that the model predicted a higher hazard, or a shorter time to event for the censored observations, which is to be expected since the event times were never observed in these cases. This is a common artifact of censoring in survival analysis diagnostics.

Observations can be influential without being outlying. The score residuals presented in Figure A.6 in Appendix A.2.1 are used to investigate influential observations. In the self-report model, two observations seem to have outlying influence according to the score residuals across all three plots. These observations were associated with two subjects with unsuppressed viral loads at 0.34 and 0.86 years, and their data was confirmed to be valid and reasonable.

In the EAM model, the score residuals associated with two observations deviate significantly from the distributions of the rest in each plot. These are associated with two subjects, one of which experienced an event. The observations were confirmed to contain reasonable values for all variables.

In the TFV model, there are perhaps three disproportionately influential observations associated with the baseline viral load variable, and one clear observation stands out with respect to the TFV variable. One of these observations pops up in both cases, and is associated with a subject who only had one event over a year from the first visit, despite recording unsuppressed viral load at baseline. Never-the-less, all the data points were confirmed to be reasonable and not obviously outlying.

Overall, the diagnostics suggest that all of these marginal models suffer from some issues which may introduce bias into the parameter estimates. The overall fits, as illustrated by the Cox-Snell residuals suggest that the models do not fit the data well across the entire expected distribution. Furthermore, the PH assumptions were violated for the baseline viral load covariate and globally so in the EAM and TFV models. These models' fits could possibly be improved by introducing an interaction term between the baseline TFV variable and observation time, or by excluding some outlying and influential observations. In the next section, the focus is directed at the EAM adherence measure. The analysis is moved towards conditional models in the hope that this will provide a model which fits the data better and can be used as a valid submodel in the subsequent joint modelling section.

7.2 Conditional survival models and further extensions

The above models were all marginal models, solving the problem of recurrent events through robust estimation of the standard errors. The marginal models provide estimates which are interpretable as population-averaged marginal effects, useful when the objective is to assess the average association between adherence and time-to-viral non-suppression events.

In Section 7.4, attention is turned to a joint model of this survival outcome and longitudinal adherence. This requires a conditional survival submodel. What follows below is thus the presentation of a recurrent events frailty survival model as applied to the time to viral non-suppression conditioned on the observed EAM adherence counts. This model is specified as:

$$h_i(t|y_i(t), \mathbf{w}_i) = h_0(t)\omega_i \exp^{\alpha y_i(t) + \mathbf{w}^T \boldsymbol{\gamma}}, \tag{33}$$

where the variables are defined as in equation 32 with the additional inclusion of the unobserved, subject-specific frailty term ω_i.

This model is fit to the data with the same fixed effects as presented in model (c) in Table 5, with the results presented in Table 8. Since this is a conditional model, this is interpreted as the average effect for the same individual - not an average effect across the population. According to this model, for any individual an additional missed dose in the past 30 days was associated with an average 1.014 factor change in the hazard of viral non-suppression, ceteris paribus (HR 95% CI = [0.993; 1.035]). In this case, this estimated individual-level association is not statistically significant with a p-value of around 0.2. An individual with prior TB exposure was estimated to have a hazard of non-suppression around 2.843 times higher than the same individual without TB exposure, holding all else constant (HR 95 % CI = [1.547; 5.224]). Similarly, a unit increase in the square root of baseline viral load was associated with an approximate 1.105 times increase in the hazard of non-suppression, holding all else constant (HR 95% CI = [1.036; 1.178]).

The point estimate relating to the longitudinal missed doses is smaller in magnitude at the individual level (conditional model) than at the population level (marginal model). The 95% confidence intervals do overlap by some margin though, with the lower bound of the conditional model estimate corresponding to the point estimate of the marginal model - at a hazard ratio of around 1.035. For the TB exposure term and the square-root of baseline VL, the estimated coefficients are very similar between those obtained from the conditional model and the marginal model. This may suggest that these baseline covariates have a similar impact on time to viral-non-suppression for the individual as they do across the population. Lastly, the standard deviation of the frailty term was estimated at 0.769 and was significantly different from zero at the 0.1% significance level.

Table 8: Cox Model Results: Frailty Model with EAM

Model Name	Variable(s)	Coef	e^{coef}	SE	p-value	e^{coef} 95% CI
(a) Frailty Model	$y(t)$	0.013	1.014	0.011	0.200	[0.993; 1.035]
	everhadTB	1.045	2.843	0.310	0.0008	[1.547; 5.224]
	sqrt(baseline VL)	0.100	1.105	0.033	0.0025	[1.036; 1.178]
	frailty(PID)	-	-	-	< 0.001	-

The Schoenfeld residuals plot for the coefficients in the above model are presented in Figure 16. The top panel, which corresponds to the adherence measure, exhibits the slight upward trend over time that was seen in the marginal model. This seems to suggest that the conditional effect of missed doses increases slightly over time, which would be a violation of the PH assumption. The score test for a zero slope yielded a p-value of 0.034 which suggests rejection of the PH assumption at 5% significance. The bottom two panels are associated with the other two covariates. The TB exposure effect seems to be relatively constant over time (p-value = 0.423), though the last few observations do deviate upwards. The residuals associated with the square-root of baseline VL seem to trend downward, which suggests that the impact of this covariate diminishes over the observation period and thus violates the PH assumption (p-value $\leq$ 0.001). Overall, these diagnostics suggest that the point estimates for the number of missed doses and the square root of baseline VL cannot be interpreted as conditionally proportional hazard ratios over the observation period.

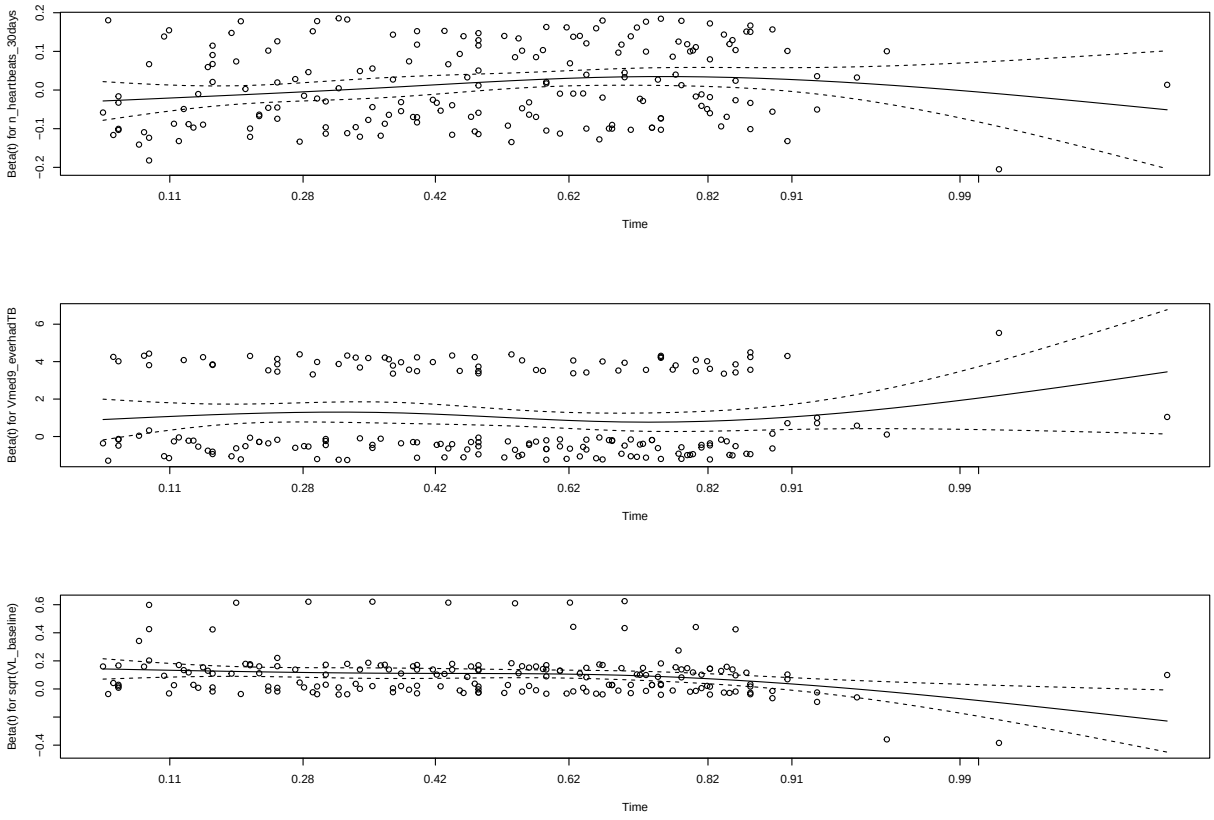


Figure 16: Scaled Schoenfeld residuals to assess PH assumption

The Cox-Snell residuals for this model are presented in Figure 17. These deviate from the expected distribution ($Exp(1)$) rather significantly. There is far more mass between the values of 0 and 1 in the observed distribution and fewer residuals in the upper tail than expected, resulting in the S-shaped curve seen in the plot. This suggests that the model underestimates the hazard of non-suppression for earlier survival times and overestimates the hazards for relatively longer survival times. Though not ideal for the overall fit of the model, compared to the marginal model these deviations may be more tolerable. In marginal model, the deviation seemed to exponentially grow with the magnitude of the residuals, indicating increasingly larger overestimation of hazards for increasingly longer survival times - while in this conditional model the deviations are consistently offset at the upper tail.

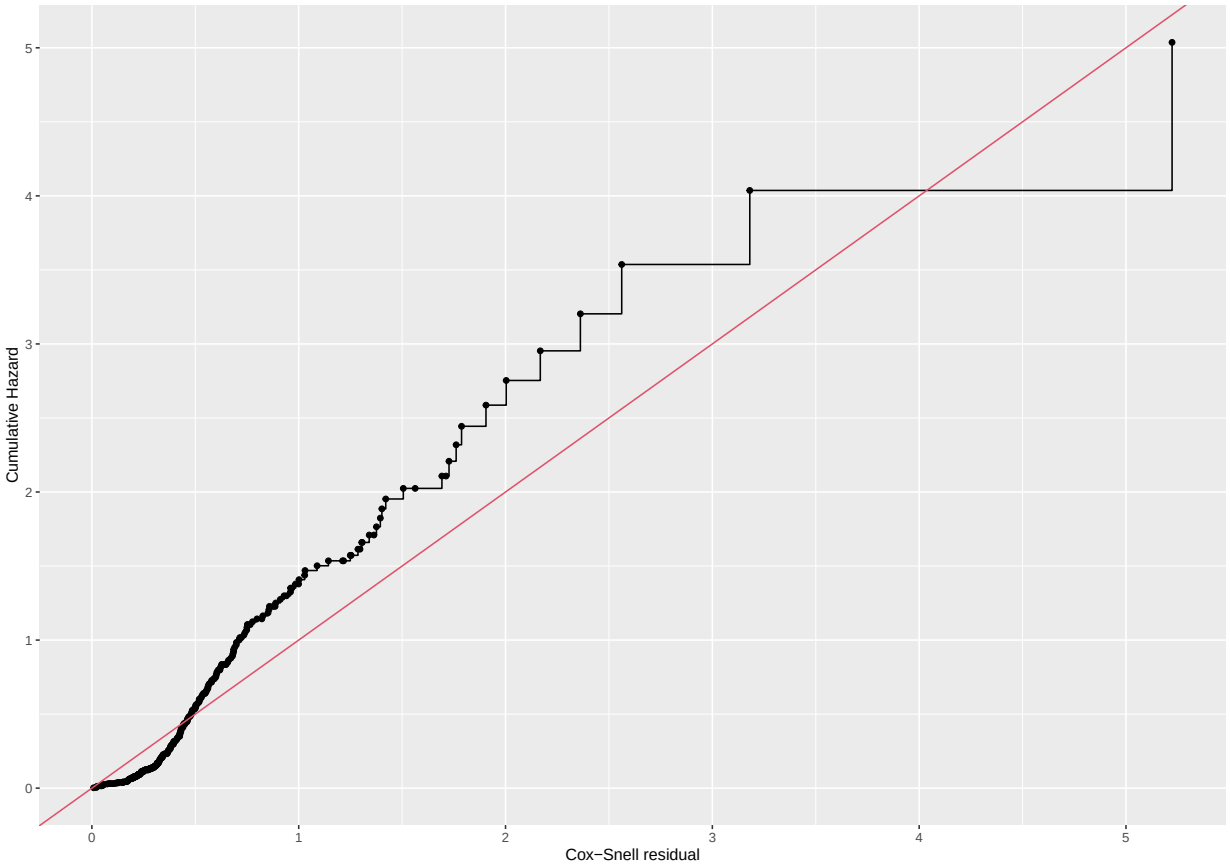


Figure 17: Cox-Snell residuals for frailty model

Figure 18 plots the deviance residuals against the linear predictions for the frailty model. In this case, the deviance residuals are pretty evenly distributed around zero for the positive linear predictions but exclusively negative for the negative linear predictions. Linear predictions of greater magnitude represent observations with a large hazard of non-suppression, and thus the even dispersion of deviance residuals at positive linear predictions suggests that the model seems to estimate individuals with shorter survival times pretty well. The negative linear predictions, however, are all associated with negative deviance residuals. This suggests that the model systematically overestimates the hazards, or equivalently underestimates the survival probabilities, for observations with relatively longer survival times.

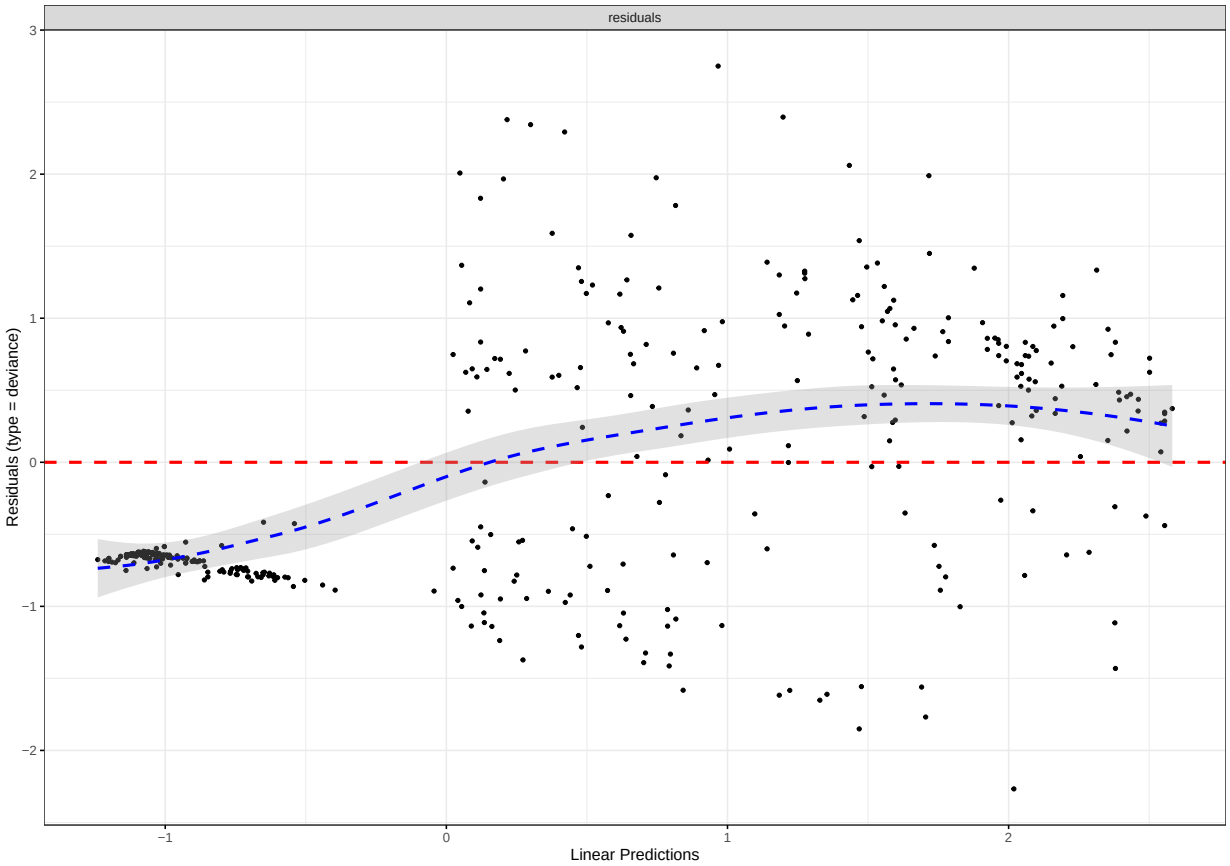


Figure 18: Deviance residuals vs linear predictor for frailty model

Overall, the diagnostics relating to the frailty model provided by equation 33 suggest that the specification of the square root of baseline VL and the longitudinal 30-day missed doses result in a violation of the PH assumption. This means that the estimates cannot be interpreted as conditionally proportional hazard ratios. This likely introduces a source of bias into the model which results in a systematic overestimation of the hazards for longer survival times.

Possible extensions to the model were applied in an attempt to improve the model fit. Firstly, discrete time-interaction terms were introduced to allow the conditional effect of baseline VL and missed doses to be estimated separately for observations before and after t years from the first visit. Secondly, a continuous time-interaction was introduced to allow these terms to vary linearly with time. However, when considered in the context of this conditional model with both fixed and random effects, both of these approaches provided over-parameterised models. Given the available sample size, the parameters in these models could thus not be estimated.

Another possibility would be to move from a Cox-PH model to a fully parametric Accelerated Failure Time (AFT) model which does not assume proportional hazards. While this could theoretically have solved some of the problems seen in the frailty model here, software constraints were encountered. To elaborate, the only parametric AFT models supported by the JMBayes software which is needed to fit the joint model in Section 7.4 are those fit via the *survreg* function from the R package *survival* [83]. This function (*survreg*) does not support data in the counting process format [94] and thus cannot be used in the context of time-varying covariates - as is necessary here since the primary goal is to assess the association between longitudinal adherence and time to viral-non-suppression. For these reasons, the analysis which follows in Section 7.4 uses the conditional model specified by equation 17 while acknowledging the possible bias introduced by the

violations of the PH assumption.

7.3 Longitudinal submodel

7.3.1 Model building

In this section, the procedure for building the generalized linear mixed effects model used to capture the longitudinal adherence profiles of the sample is presented. This procedure follows an approach that is informed by: (a) background knowledge; (b) insights gained from the exploratory data analysis; and (c) model fit statistics and likelihood ratio tests enabling formal model comparisons.

The number of missed doses in a period, as measured by the EAM data, is a finite count variable. The data generating process at any point in time could thus be reasonably assumed to follow a Poisson process, or some more flexible distribution such as the negative binomial, with some offset. This offset reflects the number of possible missed doses in a time period - which accounts for varying exposure times between individuals and observation times - as measured by the number of signals recorded on an EAM device. The most general formulation for this mixed effects model could thus be represented as in equation 7:

$$log(\boldsymbol{\lambda}_i) = \boldsymbol{X_i\beta} + \boldsymbol{Z_i b_i} + log(\boldsymbol{o}_i).$$

where $\boldsymbol{X}_i$ includes terms of the intercept and possible polynomial terms for time in years from baseline.

The number of missed doses are initially assumed to be Poisson-distributed and modeled with a log-link function. The focus is first placed on finding an appropriate functional form for time, which is specified as years from the first visit. A null model is fit with only the offset term, an intercept as the fixed effect ($\boldsymbol{X}_i = \mathbf{1}$), and a subject-specific random effect on the intercept ($\boldsymbol{Z}_i = \mathbf{1}$). The mean adherence is assumed to follow a polynomial of time. The structure of $\boldsymbol{X}_i$ is thus amended to reflect: (a) linear time; (b) quadratic time; (c) cubic time.

These models were all fit to the data. Model fit statistics used to determine the best parameterisation of time are thus presented in Table 9. Considering firstly the information criteria statistics, these decrease with each subsequent model and are both minimized by the cubic polynomial ($AIC \approx 17900$; $BIC \approx 17918$. Consider the likelihood ratio test statistic (LRT), which formally compared the reduction in the log-likelihood (logLik) between subsequently fitted models. While each increase in the degree of the polynomial provides a statistically significant increase in the logLik, the magnitude of this change diminishes as the model complexity increases. According to these model fit statistics, the model fit improves with each increase in complexity. It is, however, important to consider the inherent trade-off between complexity and parsimony when building statistical models. Higher order polynomials may thus provide a better model fit, at the cost of reduced interpretability. Furthermore, the purpose of this model is to form a sub-model for the joint model - where overparameterisation can be a concern, in which case simpler models are preferred. This is noted, and both the quadratic and cubic models are taken forward in the analysis.

Table 9: Model fit statistics: EAM fixed effects

Fixed effects	AIC	BIC	logLik	Comparison	LRT	df	p-value
1	18807.48	18814.42	-9401.738	-	-	-	-
$1 + time$	17993.04	18003.46	-8993.520	a vs. b	816.44	1	< 0.0001
$\sum_{k=0}^{2} time^k$	17937.46	17951.34	-8964.728	b vs. c	57.58	1	< 0.0001
$\sum_{k=0}^{3} time^k$	17900.16	17917.52	-8945.081	c vs. d	39.29	1	< 0.0001

Next, attention was turned to the random effects parameterisation. Table 10 provides information criteria and formal comparisons between successive models with different random effects. The quadratic and cubic models were thus fit with an additional subject-specific random effect on the linear term in the polynomial (the slope). The random effects were initially assumed to be independent of each other (indicated by the $\perp\!\!\!\perp$ symbol in the table). This assumption was then relaxed and a null hypothesis of independent random effects was tested using likelihood ratio tests to compare the models. For both the quadratic model and the cubic model, the inclusion of a random effect on the linear component significantly improves the model

fit. This is evidenced by the large reductions in magnitude of the information criteria, and significantly LRT statistics, when comparing models a and b in Table 10. Similarly, when comparing the models with independent random effects and those where this assumption is relaxed (b vs. c), there is a reduction in information criteria and a significant LRT statistic regardless of the fixed effects parameterisation. This suggests that there is sufficient evidence to reject the null hypothesis of independent random effects.

Table 10: Model fit statistics: EAM random effects

Fixed effects	Random effects	AIC	BIC	logLik	Comparison	LRT	df	p-value
$\sum_{k=0}^{2} time^k$	1	17937.46	17951.34	-8964.73	-	-	-	-
	$1 + time(\perp\!\!\!\perp)$	16579.50	16596.87	-8284.75	a vs. b	1359.95	1	< 0.0001
	$1 + time$	16522.81	16543.64	-8255.40	b vs. c	58.70	1	< 0.0001
$\sum_{k=0}^{3} time^k$	1	17900.16	17917.52	-8945.08	-	-	-	-
	$1 + time$ $(\perp\!\!\!\perp)$	16550.98	16571.82	-8269.49	a vs. b	1351.18	1	< 0.0001
	$1 + time$	16495.50	16519.81	-8240.75	b vs. c	57.49	1	< 0.0001

The final model, in scalar form, thus expressed the expected number of missed doses after accounting for the offset for subject i at time t as:

$$log(m_i|t, \boldsymbol{b}) = \beta_0 + b_{0,i} + (\beta_1 + b_{1,i})t_i + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3 + log(o_{ij}), \tag{34}$$

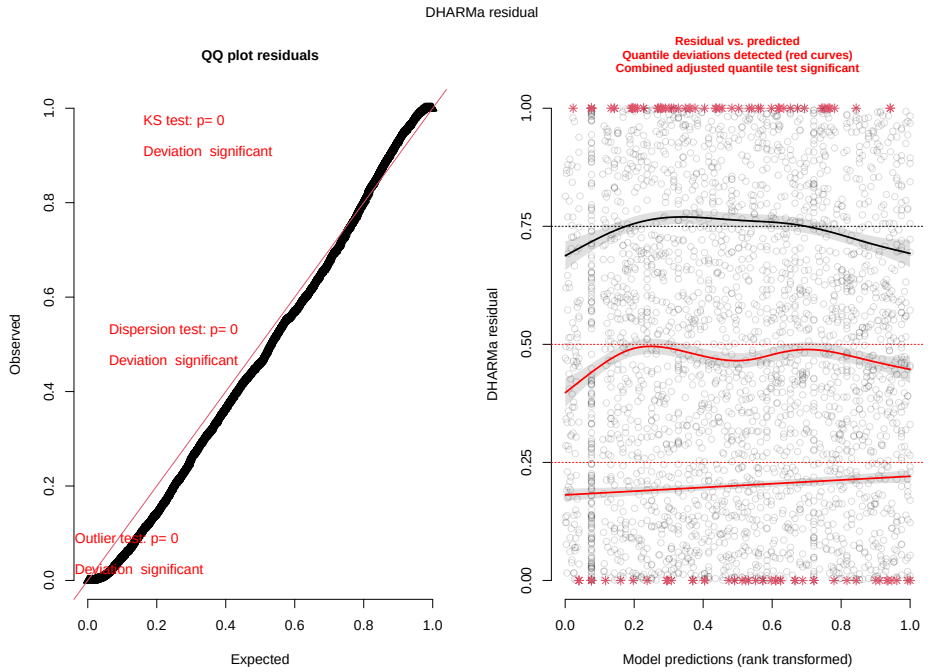
where $b_{0,i} \not\perp\!\!\!\perp b_{1,i}$ and $\beta_3 = 0$ when the quadratic time is used.

7.3.2 Model diagnostics

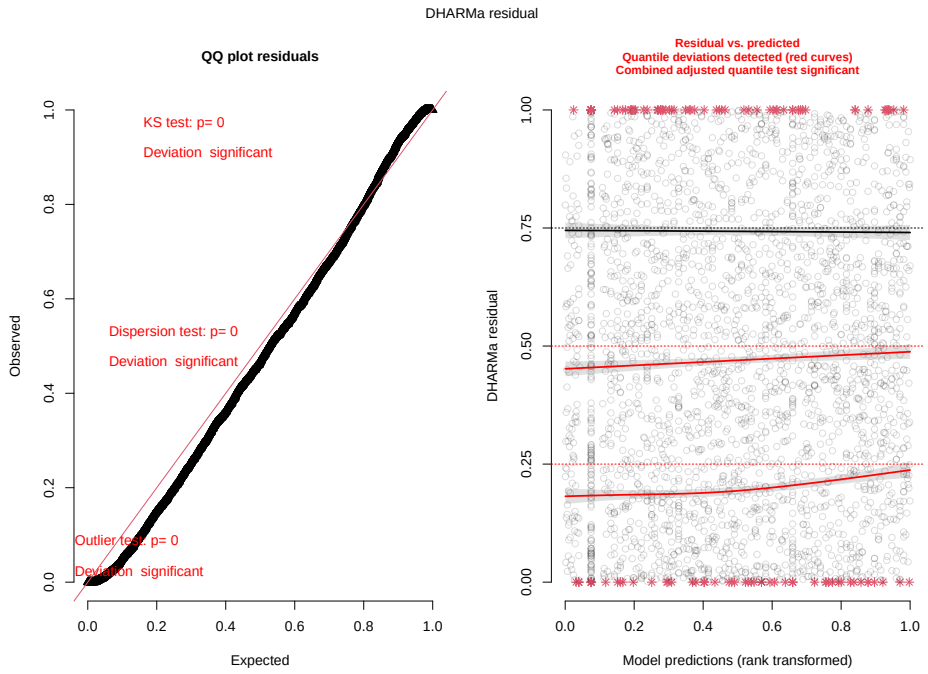
In this section, diagnostic plots based on DHARMa residuals are presented for the quadratic and cubic models as defined by equation 34. In Figure 19, the left panel provides a QQ plot to assess the overall distribution of the residuals, while the right panel plots the residuals against the rank-transformed predicted values with quantile regression lines overlayed.

Considering the former, the observed residuals exhibit some evidence of overdispersion in both models - with a larger proportion of relatively small residuals than expected. There seem to also be more residuals in the upper tail than expected. Three formal tests are provided in the plots which test for: (a) overall deviation from the expected distribution (KS test); deviation from the expected dispersion; deviation from the number of expected outliers. These tests all reject their respective null hypotheses in this case, however it must be noted that these formal tests can often be very significant to small deviations. In this analysis, more emphasis is placed on the subjective interpretation of the visual patterns in the plots.

Considering the right-hand panel in Figure 19a, there are no obvious trends across either axis. The quantile regression lines do deviate slightly at the tails of the predicted distribution, perhaps indicating a systematic under prediction of the observed extreme values. When considering the corresponding plot for the cubic model (19b), the quantile lines align more to the expectation, although they also indicate a slight linear trend among the lower residual values. There is also clear strip of residuals around a rank-transformed prediction value of around 0.08-0.1 under both models. The residuals associated with this are pretty evenly distributed however, and thus this is likely generated from some inflated observed integer value. Lastly, in both models there are many observations which result in outlying residuals (indicated by the red stars).



(a) Quadratic Poisson model



(b) Cubic Poisson model

Figure 19: Dharma residual plots for the Poisson EAM models

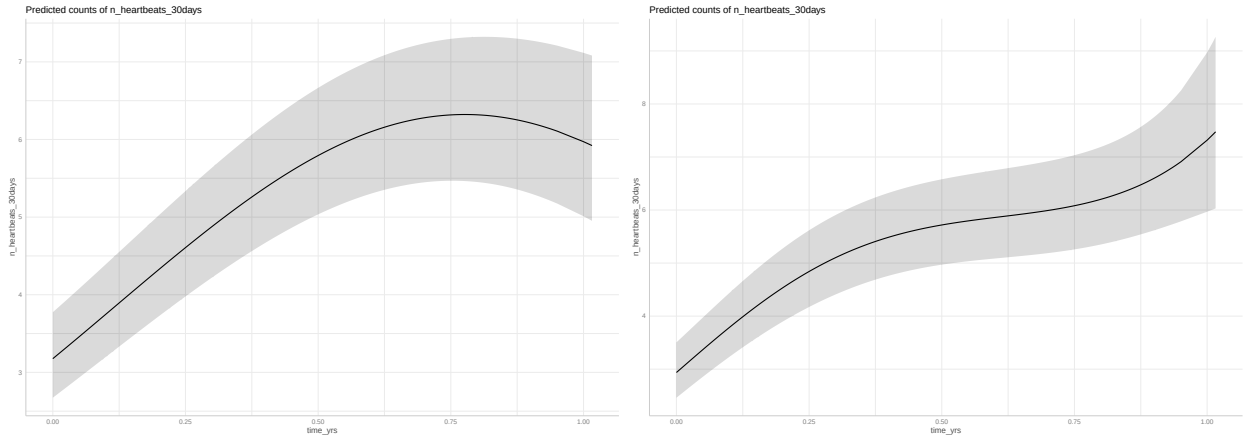
Overall, the left-hand panels suggest that there may be some overdispersion in the data, and the right-hand panels indicate some heteroskedasticity. These observations suggest that perhaps a model which can more flexibly model the dispersion may fit the data better. A negative binomial model is thus fit to the data,

under both the quadratic and cubic parameterisations of time. DHARMa residuals associated with these two models are presented in Appendix A.2.2 (Figure A.7). For both negative binomial models, the observed residuals deviate from the expected distribution in a manner that is indicative of underdispersion. This is shown by a smaller than expected proportion of residuals at the lower tail, and the converse at the upper tail. This underdispersion may be more acceptable than overdispersion, since it would tend to bias the estimated standard errors to be larger rather than smaller [31]. Considering the right-panel in the Figures, these are similarly distributed as with the Poisson model, with no major visual trends along either axis. Noticeably, there are many fewer outlying observations than observed among the Poisson model residuals.

7.3.3 Key Results

A few key points should be taken from the model building procedure and model diagnostics. Firstly, the model as specified by equation 34 likely provides a relatively good fit to the data when compared to other specifications tried in the section. Secondly, the assumption of an underlying negative binomial data generating process is likely more suitable than a simple Poisson process. The model fit is best with a cubic parameterisation of time and a negative binomial assumption, although this increased complexity comes at the cost of reduced parsimony - an important note in the context of joint modelling. This section will thus present the estimated model parameters associated with the negative binomial models.

The estimated fixed effects are provided numerically in Table A.1, however these are not easily interpretable and are thus relegated to Appendix A.2.3. Instead, the predicted marginal mean number of missed doses over time and 95% confidence intervals are illustrated in Figure 20. The quadratic model estimates an average of around 3.2 missed doses at time zero. This increases rapidly to a maximum of around 6.3 after around 0.75 years from the first visit. The mean missed doses then decreases rapidly to around 5.9 at the end of the observation period. In contrast, the cubic model is monotonic and estimates increasing mean missed doses over the entire observation window. Initially it estimates a mean of around 3 units, increasing rapidly to around 5.5 over the first third of a year. From this time until around 0.75 years after the first visit, the model estimates the average number of missed doses to increase at a slow rate, before increasing rapidly over the final few months to a maximum of around 7.5 missed doses. The mean pattern captured by the cubic seems to align more with the empirical trend seen in the exploratory analysis.



(a) Quadratic model

(b) Cubic model

Figure 20: Predicted marginal means and associated 95% CIs for negative binomial count models

Table 11 presents the estimated standard deviation of the random effects associated with these models: $\hat{\sigma}_{b_0} \approx 1.24$ and $\hat{\sigma}_{b_1} \approx 1.17$. These magnitude of these estimates clearly do not differ much between the two models. The negative correlation between the random effects, $Corr(\hat{b}_0, \hat{b}_1) \approx -0.51$, suggests that subjects with higher initial missed doses experience slower rates of increase over time, and vice-versa for

lower initial missed doses. The magnitude of these random effects, when considered in the context of the dispersion parameter estimates presented in Table 12 (which is supposed to capture any additional unexplained dispersion), are certainly not insubstantial. This indicates that there is considerable between-subject variability in the number of missed doses at time 0 and in the rate of change of this over time.

Table 11: Random Effects Covariance Matrix for Quadratic and Cubic Negative Binomial Models

Model	Parameter	StdDev	Corr
Quadratic	Intercept	1.2383	
	time_yrs	1.1736	-0.5077
Cubic	Intercept	1.2357	
	time_yrs	1.1709	-0.5049

Lastly, Table 12 presents the point estimates and standard errors of the dispersion parameter, δ from the negative binomial distribution, on the log-scale, which should capture all dispersion not accounted for by the random effects. These estimates are relatively large for both models, which suggest that there was indeed sufficient overdispersion in the data to warrant the use of the negative binomial model over the Poisson model. The standard errors are also much smaller than the point estimates, providing precision for the point estimates.

Table 12: log(dispersion) Parameter for Quadratic and Cubic Negative Binomial Models

Model	Estimate	Std.Err
Quadratic	1.9052	0.0645
Cubic	1.9184	0.0648

7.4 Joint model

7.4.1 Model building

The generalized linear mixed effects model built and presented in Section 7.3 provides the basis to extend the conditional survival model presented in Section 7.2 by harmonizing the two models within the joint modelling framework. Specifically, the longitudinal submodels provide a continuous estimate for the true and unobserved number of missed doses in the past 30 days on which the recurrent time to viral non-suppression is conditioned.

These joint models can be specified by:

$$\begin{aligned} \log(m_i|t, \boldsymbol{b}) &= \beta_0 + b_{0,i} + (\beta_1 + b_{1,i})t_i + \beta_2 t_i^2 + \beta_3 t_i^3 + \log(o_i), \\ m_i(t) &= \exp(\beta_0 + b_{0,i} + (\beta_1 + b_{1,i})t_i + \beta_2 t_i^2 + \beta_3 t_i^3 + \log(o_i)), \\ h_i(t) &= h_0(t)\exp(\mathbf{w}_i^T\boldsymbol{\gamma} + \alpha m_i(t) + v_i), \end{aligned} \tag{35}$$

where the frailty term is represented by $v_i \sim N(0, \sigma_v^2)$; the random effects, by $\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{D})$; and the true, unobserved number of missed doses at t is represented by $m_i(t)$. For the quadratic models, $\beta_3 = 0$; while $\boldsymbol{Y_i}|\boldsymbol{b_i} \sim negbin(\boldsymbol{\lambda}_i, \theta\boldsymbol{I})$ for the negative binomial models; and $\boldsymbol{Y_i}|\boldsymbol{b_i} \sim Poisson(\boldsymbol{\lambda}_i)$ for the Poisson models. The current value of m_i was used in the model specification, which represents cumulative adherence in the 30 days preceding t.

Model fit statistics for the joint models, fit with these different specifications for the longitudinal sub-model, are shown in Table 13. The statistics presented are the marginal and conditional DIC, WAIC and logarithm of the pseudo maximum likelihood (LPML) as provided by the *JMBayes2* software. These joint models were fit to the data using the Bayesian estimation approach outlined in Section 4.3. 6 parallel chains were used for the MCMC sampling with 10000 iterations per chain and a burn-in of 500 samples. There is a large

difference in magnitude across all statistics between the Poisson and Negative Binomial models, suggesting that the latter fits the data much better. The cubic models also seem to describe the data better than the quadratic models, although the difference in magnitude among the test statistics is relatively small. Overall, the negative binomial model with a cubic parameterisation of time minimises the information criteria and has the largest LPML, suggesting that this provides the best overall fit.

Table 13: Model Fit Statistics for Four Joint Models

Model	**DIC**		**WAIC**		**LPML**	
	Marginal	**Conditional**	**Marginal**	**Conditional**	**Marginal**	**Conditional**
Poisson Quadratic	18135.09	18347.53	18028.32	18157.20	-9253.25	-9364.43
Poisson Cubic	18095.54	18306.51	17991.99	18117.54	-9209.88	-9358.17
Negative Binomial Quadratic	16854.00	17572.87	16761.00	17378.67	-8537.40	-8986.08
Negative Binomial Cubic	16838.14	17554.62	16745.35	17360.52	-8509.80	-8982.69

Given the model fit statistics, as well as the insights gained from the residual diagnostics in Section 4.1.3, the negative binomial model with a cubic polynomial is chosen as the longitudinal sub-model. The joint model can thus be expressed through the following specification:

$$\begin{aligned} \boldsymbol{Y}_i|\boldsymbol{b_i} &\sim negbin(\boldsymbol{\lambda}_i, \theta\boldsymbol{I}), \\ \log(m_i|t, \boldsymbol{b}) &= \beta_0 + b_{0,i} + (\beta_1 + b_{1,i})t_i + \beta_2 t_i^2 + \beta_3 t_i^3 + \log(o_i), \\ m_i(t) &= \exp(\beta_0 + b_{0,i} + (\beta_1 + b_{1,i})t_i + \beta_2 t_i^2 + \beta_3 t_i^3 + \log(o_i)), \\ h_i(t) &= h_0(t)\exp(\mathbf{w}_i^T\boldsymbol{\gamma} + \alpha m_i(t) + v_i), \end{aligned} \tag{36}$$

where the frailty term is represented by $v_i \sim N(0, \sigma_v^2)$; the random effects, by $\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{D})$; and the true, unobserved number of missed doses at t is represented by $m_i(t)$.

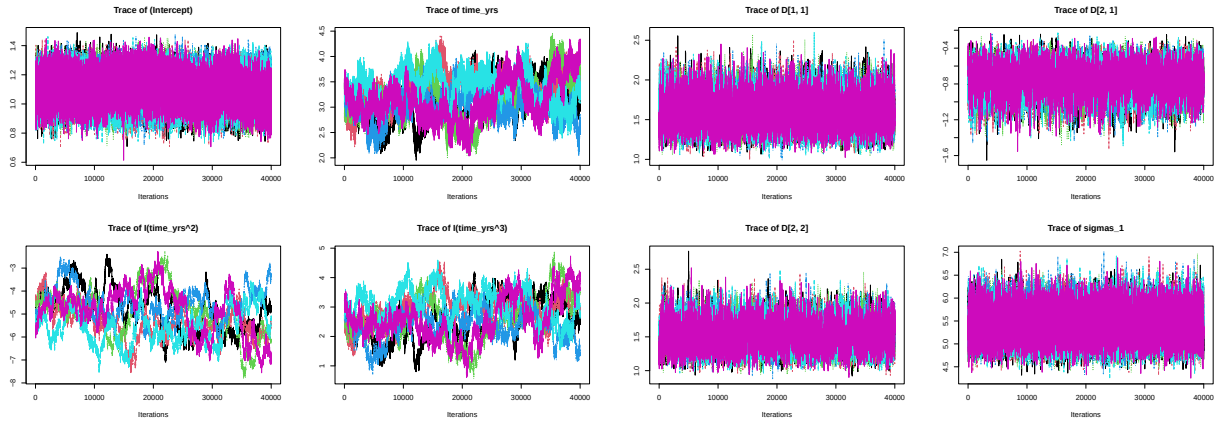
The joint model specified by equation 36 was refit to the data with a larger number of iterations to further support the convergence of the algorithm. This was not done for all candidate models since it would have been too computationally demanding. 6 parallel chains were again used for the MCMC sampling but this time with 50000 iterations per chain and a burn-in of 10000 samples.

7.4.2 Model diagnostics

The main diagnostic task with joint models is assessing convergence of the estimation of the model parameters. Prior to this, however, the individual submodels should be evaluated against their underlying assumptions. This was done for the longitudinal submodel in Section 7.3.2 (see Figure A.7b), which suggested that the model fit was acceptable. The conditional survival submodel fit was assessed in Section 7.2 (see Figures 16 - 18), which revealed a violation of the PH assumption and a resultant systematic bias in overestimating the hazards for relatively longer observed survival times and censored observations. This section proceeds with these limitations in mind.

Next, attention is turned to assessing convergence in the parameter estimates - firstly with respect to the longitudinal sub-model. Figure 21a presents traceplots for the estimation of each $\boldsymbol{\beta}$ in equation 36 over the 40000 iterations post burn-in, with each chain represented by a different colour in the plot. The point estimate and standard error for the parameters can be somewhat seen here by the mean across all chains at the last iteration and the dispersion around this mean. The intercept term clearly converges early somewhere between 0.8 and 1.3 (on the log scale). The estimate for the linear term also seems to converge although not as quickly and with some more variation between the chains. The quadratic and cubic term estimates display even more between-chain variation. However, this variation is constrained within a reasonable range and there is no global trend in the estimation over the iterations. This illustrates the necessity of estimating across more than one chain if adequate convergence is to be achieved. Figure 21b presents the traceplots for the negative binomial dispersion term (sigmas_1 = θ in equation 36), and the elements of the random-effects

covariance matrix $\boldsymbol{D}$. In this case, $D[1,1]$, $D[2,2]$ and $D[2,1]$ correspond to $\sigma_{b1,i}$, $\sigma_{b2,i}$ and $Corr(\sigma_{b1,i}, \sigma_{b2,i})$, respectively. Clearly, the estimation of all four parameters display suitable convergence.

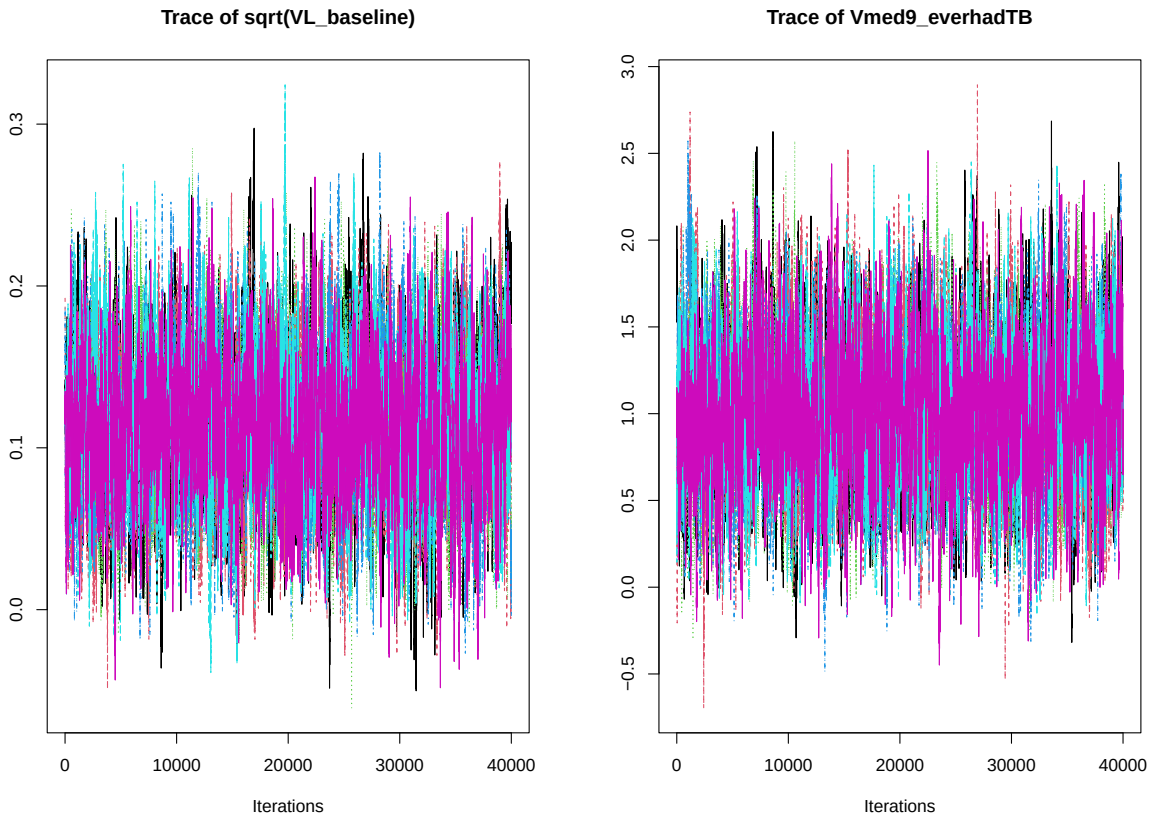


(a) Fixed effect parameter estimates

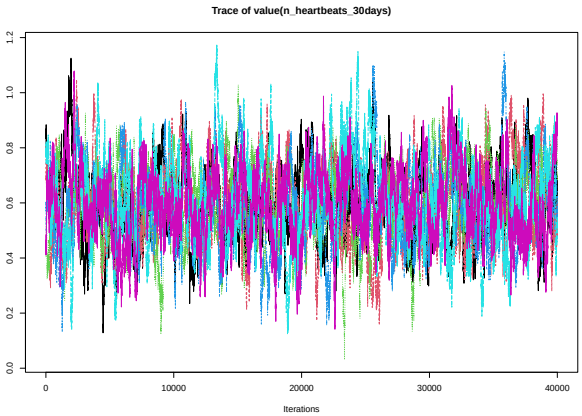
(b) Heterogeneity parameter estimates

Figure 21: Traceplots to assess convergence for longitudinal sub-model

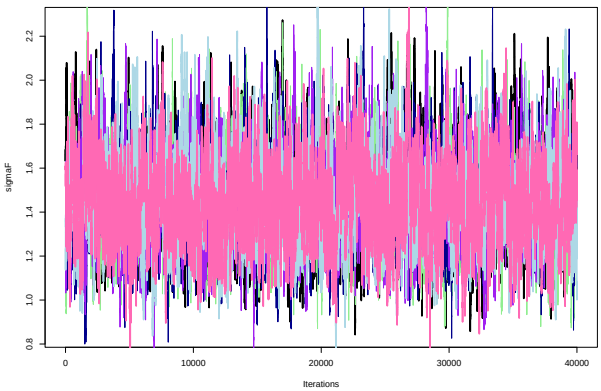
Figure 22 illustrates the MCMC estimation for the main parameters in the survival submodel. The first of these, Figure 22a, suggests that those estimates relating to the two covariates in the model - the logarithm of baseline viral load, and historical TB exposure - likely converged sufficiently. Figure 22b provides the traceplot for the association parameter, α in equation 36, the main effect of interest in this model. There is little evidence of a global trend, concerning pattern, or outlying chain/iteration(s). The estimation of this parameter thus seemed to converge. Figure 22c illustrates the estimation of σ_v, the standard deviation of the frailty term. This plot also suggests that the estimation converged. Lastly, the figures in A.8 in Appendix A.2.4 are traceplots of the parameters used to estimate the baseline hazard. These display no patterns to suggest that convergence was an issue for these parameters



(a) Covariate parameter estimates



(b) Association parameter estimate



(c) Standard deviation of frailty estimate

Figure 22: Traceplots to assess convergence for the survival sub-model

7.4.3 Results

Overall, the diagnostic plots in the previous section provide much confidence in the estimation of the parameters in this joint model. The point estimates of these parameters, as given by the mean of the posteriors, as well as the standard deviation and resulting significance values and 95% credibility intervals are displayed in Table 15.

Consider first the estimates of the random effects covariance matrix in the longitudinal sub-model - presented in Table 14. These estimates are very similar to those from the model when fit independently of the survival outcome (see Table 11), although slightly lower in magnitude. The same interpretation thus applies in terms of the random intercept and slope terms being negatively correlated.

Table 14: Random Effects Covariance Matrix for Longitudinal sub-model

Parameter	**StdDev**	**Corr**
$b_{0,i}$	1.261	
$b_{1,i}$	1.227	-0.480

Furthermore, with respect to heterogeneity in the longitudinal component, consider the terms that capture dispersion in the data generating process (Table 15). The estimate for the negative binomial dispersion term is presented on the natural scale, and is smaller in magnitude to that obtained from the independent model fit ($log(5.42) = 1.69$ vs. 1.92). This may be because some of the additional variation in the longitudinal outcome is now captured by it's association with the survival outcome.

Table 15: Parameter Estimates for Joint Model of EAM Adherence and Time-to-Viral Non-Suppression

Parameter	**Mean**	**Standard deviation**	95% CI	**p-value**
Longitudinal outcome				
$\hat{\beta_0}$ (intercept)	1.100	0.093	[0.916; 1.279]	< 0.0001
$\hat{\beta_1}$ (t)	3.198	0.375	[2.445; 3.927]	< 0.0001
$\hat{\beta_2}$ (t^2)	-4.987	0.938	[-6.808; -3.098]	< 0.0001
$\hat{\beta_3}$ (t^3)	2.676	0.672	[1.339; 3.977]	< 0.0001
$\hat{\theta}$ (dispersion)	5.421	0.320	[4.828; 6.081]	< 0.0001
Survival outcome				
$\hat{\gamma_1}$ (sqrt(Baseline VL))	0.115	0.044	[0.031; 0.202]	0.0073
$\hat{\gamma_2}$ (TB exposure = 1)	1.029	0.388	[0.285; 1.819]	0.0069
$\hat{\alpha}$	0.596	0.140	[0.331; 0.872]	< 0.0001

The estimates of the longitudinal fixed effect parameters are presented in Table 15. The intercept term in this case includes adjustment for the offset term and thus corresponds to a mean number of missed doses at the first visit of around $exp(1.10) \approx 3.00$ units ($p < 0.001$). Although the estimate obtained from the same model when estimated independently did not include adjustment for the offset term (see Table A.1), if one adjusts the estimate by the mean exposure ($\hat{o} = 29.1$ days) it is clear that the estimates are very similar: i.e. from the original longitudinal model, $exp(-2.281) \times \hat{o} = 0.109 \times 29.1 \approx 2.96$.

The linear term (slope) is estimated as around 3.2 units/day with a 95% credibility interval of [2.45; 3.93]; the quadratic term is estimated at -4.99 (95% CI = [-6.81; -3.10]); and the cubic term is approximated as 2.68 (95% CI = [1.34; 3.98]). Overall, the magnitude of all of these estimates are larger than those obtained from the model when estimated out of the joint modelling framework.

The differences in the fixed effect estimates are illustrated via Figure A.9 in Appendix A.2. This shows the fitted curves for the longitudinal outcome according to the fixed effect components of the negative binomial cubic model fitted in Section 7.3 and the corresponding sub-model in this joint model, with twelve randomly chosen observed trajectories overlayed. Overall, the two fitted curves are very similar. The noticeable difference is that the fitted curve from the joint model exhibits a more rapid deceleration and thus has a slightly lower level over the observation window.

One can also generate subject-specific predictions from these models by making use of the random effects. This is illustrated for the twelve randomly sampled participants in Figure 23, with the observations shown by dashed black lines. The predictions for the joint sub-model and original longitudinal model shown in red and blue, respectively. Since the fixed effect and random effect estimates are both very similar between

the two models, the predictions are very similar between the two. The shaded areas represent some level of uncertainty for each model, a 95% interval according to Poisson assumptions. The uncertainty for the joint model predictions appear to be much larger and perhaps more accurately reflects true uncertainty since the red shaded area tends to more often over the observed missed doses. According to the model specification (equation 36), these individual fitted curves are considered the true and unobserved adherence trajectories upon which to condition the survival submodel.

Figure 23: Subject-level Longitudinal Predictions from Joint Sub-model and Longitudinal Model

Next, attention is turned to the results associated with the survival sub-model. The main effect of interest in this model is α, which estimates the association between the hazard of viral non-suppression and the true, unobserved number of missed doses in the preceding 30 days according to EAM. According to the joint model, this is estimated at around 0.60 (95% CI = [0.33; 0.87]). For any individual, the hazard of viral non-suppression is thus increased by a factor of approximately 1.81 for each additional missed dose in the preceding 30 days (HR 95% CI = [1.40; 2.39]), holding baseline viral load and TB exposure constant. This estimate is far higher in magnitude than the estimates obtained from both the marginal survival models (Table 5) and the conditional survival model which underlies this joint model (Table 33).

According to the (exponentiated) 95% credibility intervals, unit increases in the square root of baseline VL were associated with a conditional factor increase in the hazard of non-suppression of between 1.03 and 1.22; while prior TB exposure was associated with a factor increase in the hazard of between 1.33 and 6.17. These estimates are similar to those obtained in both the marginal and conditional survival models. Lastly, the standard deviation of the frailty term was estimated as $\sigma_v \approx 1.464$ (95% CI = [1.096; 1.905]) which is higher than the estimate obtained from the conditional model in Section 7.2.

7.5 Longitudinal latent class analysis

7.5.1 ADD-ART

In this section, the results relating to the Group-Based Trajectory Model applied to the ADD-ART sample are presented. This details the model specification, the model selection, results and some diagnostics.

Whereas the negative binomial distribution was used to account for additional variability in the GLMM presented in Section 7.3, it is assumed that additional heterogeneity here is captured by conditioning on the latent classes and underdispersion is captured through a zero-component in the model. Furthermore, in this analysis time is parameterised as discrete visits from 1 to 11.

The number of EAM missed doses for individuals i in the 30 days preceding time t, conditional on being in group g, is thus estimated using the zero-inflated Poisson GBTM as described in Section 4.4.4:

$$y_{it}^g \sim (1 - q_t^g)Poisson(\lambda_t^g \kappa_{it}) + q_t^g I(y_{it}^g = 0),$$

where λ_t^g is the expected number of missed doses at t for individuals in group g, q_t is the zero-component mixing proportion, and the offset term κ_{it} is the proportion of the days preceding t for which signals from the EAM device were recorded. q_t is modelled proportionately to λ_t^g, which is modelled as a polynomial of time, as was defined in Section 4.4.4.

This model was fit to the data with both a quadratic and a cubic polynomial, specifying 2 to 5 groups. The model fit statistics are presented in Table 16, with the best fit indicated via the green shaded cells. According to all four test statistics, the models with 5 groups always provide the best fit to the data. The log-likelihood, AIC and BIC all suggest that the quadratic model is preferred, however the CVE is minimised with the cubic polynomial. An attempt to fit a 6-group model to the data was made, but the model would not converge. The cubic polynomial model with five latent groups is thus chosen. Since the degree of the polynomial is forced upon all latent classes, this should allow both the simpler and more complex subgroup trajectories to be captured (the coefficient associated with the cubic term in the log-linear polynomial could trend towards zero for simpler trends).

Table 16: Model fit statistics: ADD-ART missed doses GBTM

Groups	logLik		AIC		BIC		CVE	
	Quadratic	Cubic	Quadratic	Cubic	Quadratic	Cubic	Quadratic	Cubic
2	-10204	-10189	20426	20399	20480	20465	4.702	4.698
3	-9232	-9217	18493	18467	18576	18569	3.993	3.982
4	-8890	-8872	17819	17791	17932	17928	3.722	3.831
5	-8674	-8695	17397	17448	17540	17620	3.529	3.522

Figure 24 plots the individual trajectories of EAM missed doses for each of the five latent classes. While there is much noise in these figures, the overall heterogeneity does seem to have been well deconstructed into distinct classes. The number of participants in each group is also indicated in the plot. Group sample sizes are relatively well balanced between groups 1 to 4, but group 5 does have 15 more members than the next largest group. Figure A.10 in Appendix A.2.6 illustrates the distribution of posterior probabilities associated with each of the five latent classes. Most of the mass of these distributions is concentrated at zero and one, which indicates that the latent groups are well separated according to the model. For some individuals, there are clearly some probabilities which are not at these bounds, which does imply that for a few individuals the assignment of group membership is more ambiguous. These are clearly exceptions to the rule, however, and thus for the most part group membership does not seem to be arbitrary.

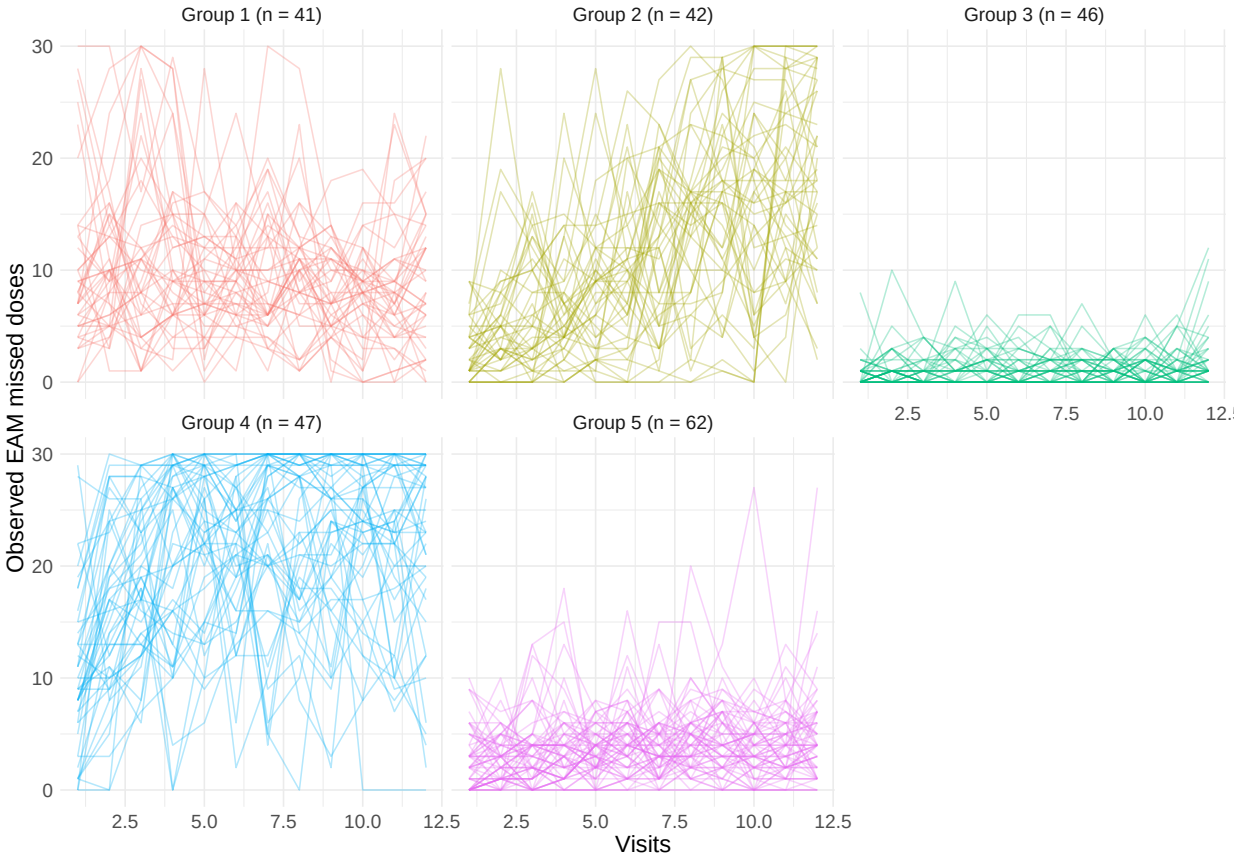


Figure 24: Individual EAM trajectories for individuals in each latent class

Figure 25 plots with solid lines, the fitted mean trajectories for each of the five latent classes according to the model as well as 95% confidence intervals estimated via a Gaussian approximation. Group 3 seems to be comprised of the most adherence individuals, who missed a stable average of around one missed dose throughout the study. Group 5 also seems to represent individuals with stable, good adherence albeit at a slightly higher level than Group 3. Group 2 appears to represent individuals who entered the study with relatively good adherence, but steadily missed more doses as the study continued - going from a mean of around four to around nineteen missed doses. Group 4 appears to represent individuals who's adherence worsened over time. This group on average missed an estimated twelve doses at the first time point - which increased rapidly to around twenty-three units after 6 visits and remained high. Lastly, Group 1 seems to represent individuals who's adherence was stable and relatively poor throughout the study. This group missed an estimated eleven units at the first observation and this did not change much throughout the rest of the study.

For each group, Figure 25 also plots the observed mean missed doses, weighted by the posterior probability of group membership, as points connected by dotted lines. This provides another view of how well the fitted model captures the observed responses - in terms of means per latent class. The weighted means seem to follow the predicted means for all groups, indicating that the model fits the data rather well.

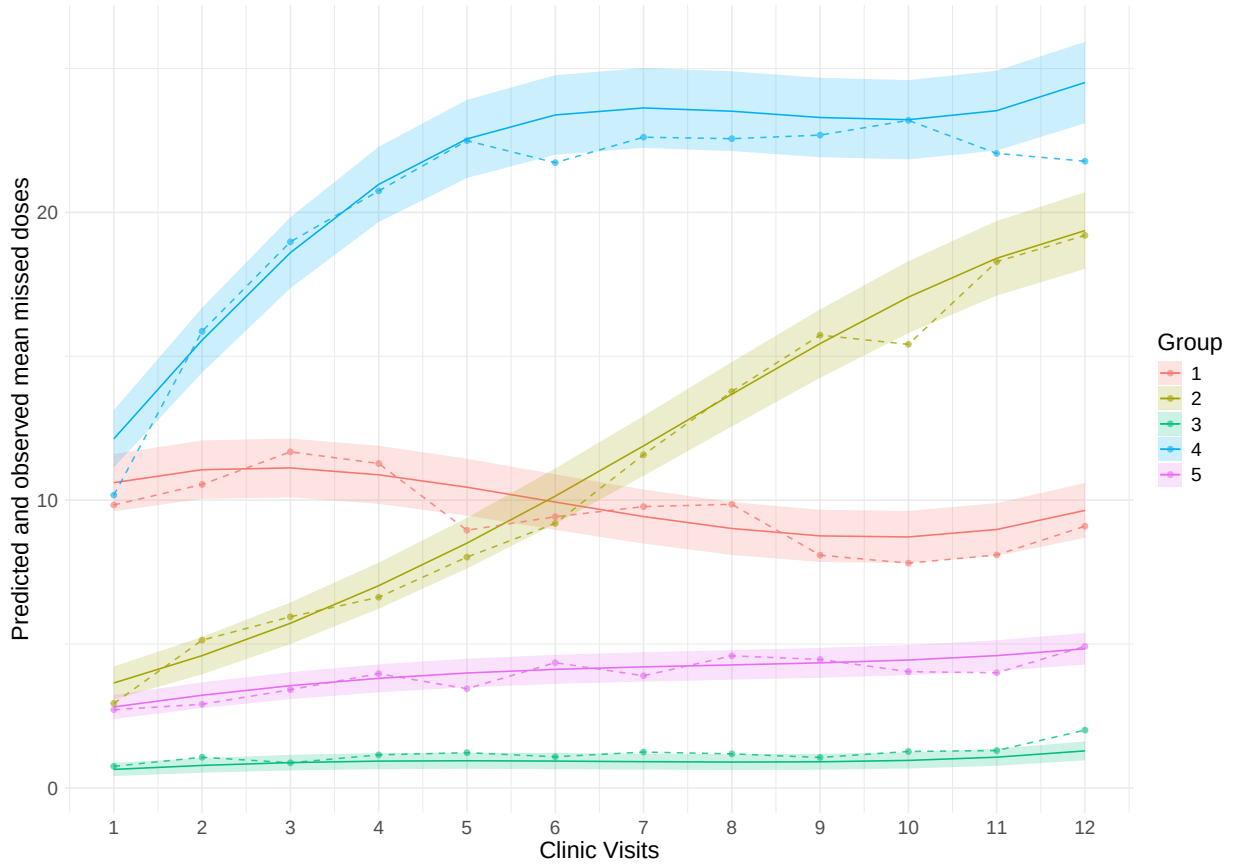


Figure 25: Predicted and weighted mean trajectories for each latent class

A Kaplan-Meier analysis was applied to the recurrent viral non-suppression events after stratifying the sample by the five latent adherence groups. This is visualised in Figure 26, which plots the estimated probability of viral suppression over time in years from Visit 1. It is clear that Groups 1 and 4, which correspond to groups with initial poor adherence, experienced many more non-suppression events than the other groups over the first 0.5 years - reflected by a sharp drop off in the estimated survival probability. The median survival time for these groups is thus both around 0.5-0.6 years. Group 1's survival probability then plateaus over the rest of the period, accumulating far fewer events over the second half of the time frame than the first. Group 4 however continued to experience many survival events over the rest of the observation period, as their average adherence worsened, contributing an eventual 85 of the total 180 events recorded in the sample. Group 2, who started with reasonable average adherence which deteriorated relatively slowly, recorded comparatively fewer events (11) over the first 0.5 years resulting in a much slower decline in the survival probability. As their average adherence worsened though, the group accumulated more non-suppression events and their survival probability correspondingly deteriorated faster. The two groups which exhibited stable and good adherence, Groups 3 and 5, accumulated the fewest non-suppression events - only 16 and 18, respectively. These groups thus had the slowest decline in survival probability over the observation window. Interestingly, although Group 3 had the best adherence according to the model, they accumulated slightly more events initially - the same as Group 2 over the first 0.5 years. The magnitude of these differences is not very large though, so this could very well be a result observed by pure chance.

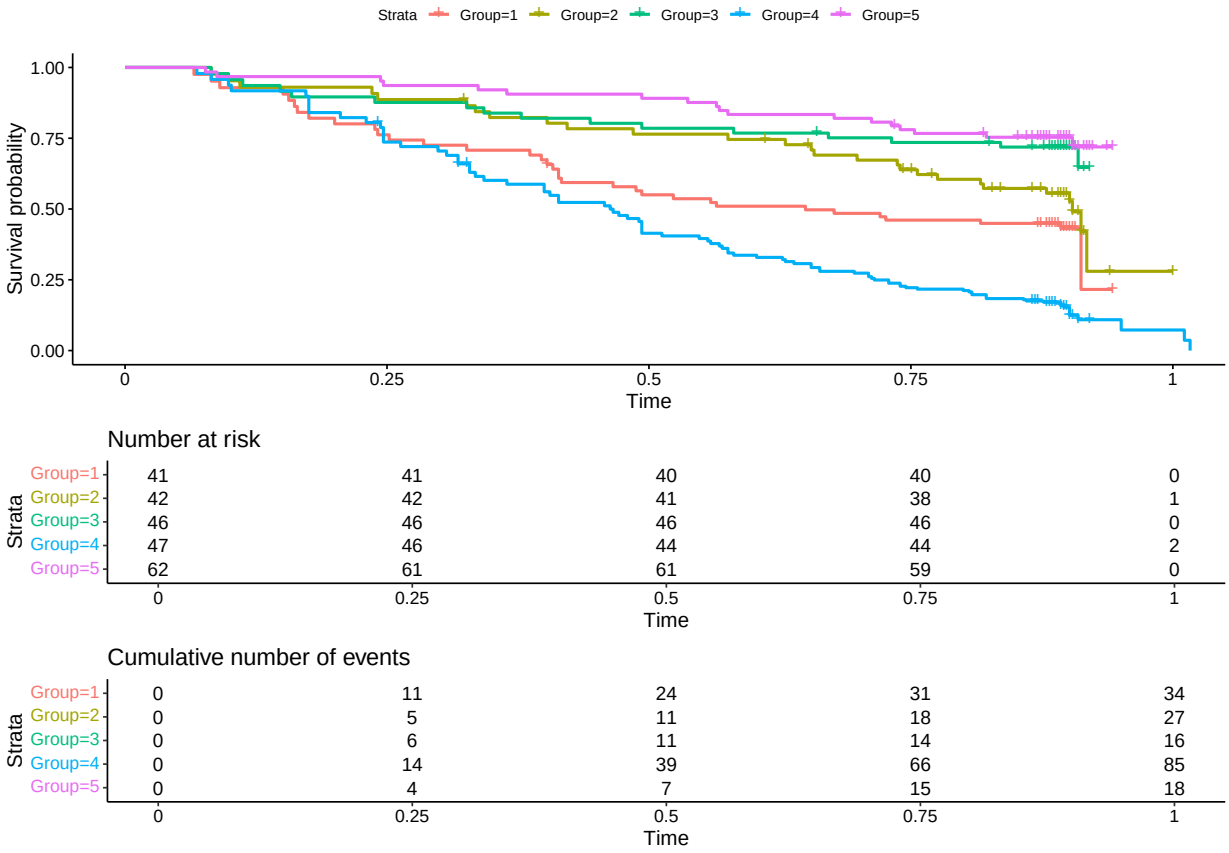


Figure 26: Time to recurrent viral non-suppression for the sample stratified by latent adherence group

Differences between the adherence groups along baseline cultural, social, and psychological characteristics were also investigated with the aim of identifying whether any of these traits are associated with the different adherence trajectories. The traits investigated include each of the scales which were described in Section 5.1.2 - after summing across the items that comprise these scales. Box plots are provided in Figure A.11 in Appendix A.2.6 for this purpose. No clear differences between the groups along any of the scales were observed - as can be seen visually and is evidenced by the p-value which corresponds to a Kruskal-Wallis test for a global difference in location.

8 Discussion

This dissertation aimed to explore the complex relationship between antiretroviral adherence and viral outcomes in people living with HIV in South Africa using advanced statistical modeling techniques. Specifically, the focus was on comparing different adherence monitoring tools and examining heterogeneity in longitudinal adherence behaviors. The analyses utilized data from the ADD-ART study, a prospective cohort study of 250 virally-suppressed adults on ART in Cape Town. Several key insights into adherence patterns and their associations with viral non-suppression were extracted through the application of survival analysis, joint modelling, and longitudinal latent class modeling to the data.

8.1 Key Results

8.1.1 Comparison of Adherence Measures

One of the primary objectives was to compare different adherence monitoring tools in terms of their association with viral outcomes. Significant associations between adherence measures and the hazard of viral non-suppression were found across all marginal survival models. This relationship persisted after adjusting for baseline viral load and TB exposure, two seemingly important clinical factors.

While both in theory measure the number of missed doses in the preceding 30 days, SR adherence consistently underestimated non-adherence compared to EAM data. This was evident in the descriptive statistics, exploratory analysis, and the marginal survival model results. The hazard ratio associated with unit increases in SR missed doses was larger than that for EAM missed doses. This could be explained by a single SR missed dose corresponding to multiple actual missed doses as more accurately measured by EAM. In this case, the association for a one unit increase in SR missed doses would actually represent more than one missed dose and thus correspond to a relatively higher hazard ratio. This finding aligns with previous research indicating that self-reported adherence often overestimates true adherence (or underestimates true non-adherence, in this case) due to recall and social desirability biases [90].

The TFV-DP concentration measure provided valuable insights as a biological marker of adherence. Unlike missed dose counts, higher TFV-DP levels indicate better adherence. The hazard ratios for TFV-DP were thus in the opposite direction compared to missed dose measures, as expected. The magnitude of association between TFV-DP levels and viral non-suppression was substantial, with a 10-fold decrease in TFV-DP concentration associated with approximately double the hazard of viral non-suppression. Since the units of measurement were very different between SR and EAM (missed doses), and TFV-DP (*fmol/punch*), the magnitude of the hazard ratios were not directly comparable.

Interestingly, the conditional survival model did not find a statistically significant association between EAM missed doses and the hazard of non-suppression. This suggests that the association may be different after conditioning on unobserved subject-specific random effects, than it is at the population level. However, the joint modeling approach, which accounted for the longitudinal nature of adherence data, revealed a stronger association between EAM adherence and viral outcomes compared to the marginal and conditional survival models. The joint model estimated that each additional missed dose in the preceding 30 days was associated with an approximate 81% increase in the hazard of viral non-suppression (HR = 1.81, 95% CI: [1.40; 2.39]). This estimate was substantially larger than those obtained from marginal or conditional survival models, highlighting the importance of appropriately modeling the time-varying nature of adherence and uncertainty inherent in the observations.

8.1.2 Impact of Baseline Medical Factors

While adherence was the primary focus, the analyses also revealed the association between certain baseline factors and viral outcomes. Baseline viral load consistently emerged as a significant predictor of subsequent viral non-suppression, even after accounting for adherence. This suggests that patients with higher baseline viral loads may require more intensive monitoring or support, regardless of their initial adherence levels.

Prior TB exposure was also associated with an increased hazard of viral non-suppression across multiple

models, with a clinically and statistically significant magnitude. This finding suggests that patients with a history of TB may face additional challenges in maintaining viral suppression, which may be linked to the complexity of HIV-TB co-infection or perhaps serve as an indicator of HIV disease progression.

8.1.3 Longitudinal Adherence Patterns

The exploratory and statistical analysis of longitudinal adherence measures revealed concerning trends over the study period. Both EAM data and TFV-DP concentrations indicated a general decline in adherence over the observation window. This trend was not captured by SR adherence, which remained relatively stable throughout the study. The discrepancy highlights the limitations of relying solely on SR adherence in clinical practice and research.

The longitudinal latent class modeling identified five distinct adherence trajectory groups based on EAM data, showing that the mean trend of deteriorating adherence did not apply to the whole sample - in fact about half of the participants exhibited consistently good to excellent adherence over the study. Importantly, the groups with poorer or declining adherence experienced significantly higher rates of viral non-suppression events. This underscores the statistical relevance of latent class analysis of longitudinal data. When there are no known or measured predictors of possibly different adherence trajectories, unobserved heterogeneity can still be identified with these methods.

In an attempt towards understanding the drivers of this heterogeneity, an analysis attempted to classify the identified latent classes based on some measured baseline characteristics. Unfortunately, no strong associations between baseline socio-demographic or psychological factors and membership to specific adherence trajectory groups were found. This suggests that predicting long-term adherence patterns based on initial patient characteristics may be challenging, emphasizing the need for ongoing adherence monitoring and support throughout treatment. This finding aligns with previous research, such as that of Mody et. al. [59] (see section 3). This may indicate that the drivers of non-adherence in this context are complicated and multi-faceted, and cannot be attributed to any single measured socio-demographic or psychological outcome. Though more thoroughly developed quantitative models designed to assess the drivers of adherence may reveal intricacies that the methods implemented in this research are unable to.

8.2 Limitations

This dissertation contains several important limitations which should be considered when interpreting the results. Notably, some model fit issues were seen due to violations of statistical assumptions, particularly in the marginal and conditional survival models. Violations of the proportional hazards assumption for baseline viral load and EAM adherence measures were observed and noted. The violation for baseline viral load is intuitive - one would expect the baseline viral load to have a larger effect on the probability of viral non-suppression at points in time closer to baseline. The reasons for the PH violation for EAM measures is less apparent. The Schoenfeld residuals implied an increasing effect over time which may suggest some sort of cumulative effect of EAM adherence on the risk of viral non-suppression. Overall, these violations resulted in survival models which were biased towards overestimation of the hazard of non-suppression for longer observed survival times and right-censored survival times. While various approaches to address this significant issue were explored, including time-dependent coefficients, these were ultimately limited by computational constraints and the requirements of the joint modeling software.

These limitations likely also apply to the joint model results, which use the conditional survival sub-model which suffers from these problems, though it is very difficult to assess this since the use of residual diagnostics directly from the joint model is still an active area of research.

Additionally, many statistical analyses were conducted in this dissertation. Some of these essentially tested for the same effects over the same sample, such as the many survival models. Implementing many statistical tests results in an increased chance of Type I errors - where the null hypothesis is incorrectly rejected in favour of the alternative due to pure chance. No adjustments for multiple comparisons were considered in this research and thus results which are only marginally statistically significant should be considered with caution.

The study sample of 238 participants may be considered relatively small for some of the more complex statistical models employed, such as the joint model and the GBTM. In complex models with many parameters, a larger sample size is generally needed to achieve adequate statistical power and precise parameter estimates. With a smaller sample size, the uncertainty of parameter estimates is inflated resulting in wider confidence intervals and inhibiting the power to detect true effects of small magnitude, potentially leading to Type II errors. This limitation is especially relevant for the longitudinal latent class analysis, where the 238 participants were divided among five groups of sizes 41 to 62. Though the groups were relatively well balanced, a larger sample size may have increased the potential to detect more subtle associations between baseline characteristics and adherence trajectory group membership.

Furthermore, the complexity of the models paired with the relatively small sample size also imposed some limitations on the exploration of more elaborate model specifications. This is a general statistical and machine learning constraint where more highly parameterised models require larger sample sizes to estimate the parameters. A natural extension would be to consider more complex latent class models, such as a Growth Mixture Model, since the GBTM does make some potentially naive simplifying assumptions. A further extension would be to fit a latent class joint model, which would identify latent subgroups by making use of both the longitudinal and survival data.

Another statistical limitation involves the analysis of psychological and social factors following the longitudinal latent class analysis. These factors relied on simple sum scores for various scales, which likely oversimplified complex latent constructs. A more nuanced approach, such as factor analysis or item response theory, could potentially reveal more subtle associations between these factors and adherence patterns. This was however beyond the scope of this research.

A further set of limitations has to do with bias introduced in the sampling procedure. Firstly, the exclusion of 12 participants due to insufficient data or missing covariates may have introduced some selection bias, potentially impacting the ability to generalize the key findings to the full population from which the sample was drawn. It was assumed that any missingness was at random, and not related to the outcomes, and this assumption is key to the validity of the statistical methods used. Secondly, the lack of sampling for genetic and pharmacokinetically relevant data limited capacity for adjusting the analyses for potential confounders which may impact drug metabolism and efficacy. These omitted variables may contribute to variation in the TFV-DP measurements and/or the resistance of the virus to the medication, and thus to variation in viral outcomes not attributable to adherence. Therefore, missing variable bias is likely introduced into these analyses by not accounting for differences in BMI, creatinine levels, as well as the phenotypes of the virus and the host. This missing data represents an important gap in the capacity of this research to fully characterize the factors influencing adherence and viral suppression.

Furthermore, some limitations on the longitudinal nature of the data were identified. Firstly, only twelve months worth of adherence and viral outcomes data were used in the analyses. Consequently, longer term adherence behaviours could not be explored. Secondly, the EAM data was only available from visit 1 and not the baseline visit - which means that some baseline data was not used and essentially the observation period was shifted by one visit with visit 1 corresponding to time 0. This necessary step had the potential to introduce some bias into the analyses since it discards some information on viral and adherence outcomes at true baseline, though efforts were taken to make this clear in the interpretation of results throughout the dissertation. Lastly, the daily EAM data was summarised into monthly counts. This does result in a smoothing over of information. More granular data, such as weekly counts or perhaps event the daily indicators, could be used to identify more subtle trends

Although several strategies to mitigate known sources of bias in EAM data were implemented, such as accounting for curiosity openings and device malfunctions, some residual bias may have remained. A notable example is the potential of the EAM device to act as an adherence intervention. Individuals using an EAM device may change their adherence behaviour and thus the adherence measured by the device may not accurately measure adherence under no monitoring. The impact of this potential bias on the findings is difficult to quantify.

The research also has some important limitations around generalizability which relate to the sampled popula-

tion. Firstly, the ADD-ART study was completed in 2019 - before the switch from efavirenz to dolutegravir-based first-line ART in South Africa. This may limit generalisation to populations in further points in time due to a number of mostly unknown temporal effects - the change in ART regimen being just one of these. Furthermore, the study sample was drawn from relatively impoverished, urban areas in the Western Cape of South Africa, likely limiting generalizability to this very specific population. These sampling constraints in time and space are always crucial to consider when attempting to generalize statistical results from sampled data.

8.3 Notes and Future Work

This research has several important implications for clinical practice and future research, despite the limitations discussed earlier. The key results underscore the significance of employing multiple methods to assess adherence in both clinical and research settings. The observed discrepancies between self-reported adherence and more objective measures (EAM and drug concentrations) highlight the potential limitations of relying solely on self-report, which may lead to overestimation of adherence.

The findings emphasize the value of joint modeling approaches in HIV research. The stronger associations found in the joint modeling approach suggest that future studies should consider employing these methods to more accurately capture the dynamic relationship between time-varying adherence and clinical outcomes. Additionally, the increased risk of viral non-suppression among patients with higher baseline viral loads and prior TB exposure indicates that these groups may benefit from more intensive monitoring and support. Clinical guidelines should consider incorporating these factors into risk assessment.

Future research directions could explore more nuanced aspects of adherence patterns, such as timing irregularities or treatment interruptions, and their impact on viral outcomes. Integration of genetic and pharmacologically relevant data in future studies could provide valuable insights into variations in drug metabolism and susceptibility to side effects, which may influence adherence patterns and treatment outcomes. Extended follow-up periods in future studies could offer insights into the long-term consequences of different adherence trajectories and the sustainability of viral suppression over time.

Implementation science research is needed to determine how best to integrate advanced adherence monitoring and modeling techniques into routine clinical care in resource-limited settings. The latent class analysis revealed that some patients experience rapid declines in adherence, emphasizing the need for regular adherence monitoring using objective measures. Implementing such monitoring could help identify these patients early, allowing for timely interventions. The feasibility of different monitoring mechanisms at scale is thus also a practical topic to be addressed.

The heterogeneity in adherence trajectories suggests that personalized approaches to adherence support may be more effective than uniform strategies. This aligns with the field of precision medicine, focusing on identifying individuals at higher risk of non-adherence and tailoring interventions to these individuals. However, the results from this research and that of Mody et al. [59] suggest that identifying drivers of adherence behavior and classifying individuals at entry to care based on their profile along these drivers may not be effective. Future research could focus on thoroughly developing quantitative models which can capture more nuanced dynamics specific to adherence. Since the quantitative methods used in this research failed to identify factors predictive of different longitudinal adherence behaviors, a psychologically-driven qualitative approach could help reveal more subtle differences between these latent groups and be used to develop more suitable quantitative models.

Another possibility for identifying higher-risk individuals is to use adherence data for dynamic predictions within the joint modeling framework. This would entail estimating individuals relative hazard of viral non-suppression based off of their adherence data and the joint model fit. A joint model that does not suffer from a biased survival submodel could reliably generate dynamic predictions of individuals' hazards of viral non-suppression using their baseline characteristics and longitudinal adherence monitoring data. This extension could have significant implications for clinical practice in South Africa, though it would require accurate measurement of adherence.

An alternative perspective is to target interventions to subpopulations rather than individuals, aligning with the developing field of precision public health. As described by Khoury et al. [46], precision public health aims to acknowledge sub-populations with different needs within broader populations and accurately tailor interventions towards sub-groups based on their needs. The latent class methods applied in this research could be a promising tool in this area. If applied to different geographic regions, such as routine adherence monitoring data from each clinic in a province, it may be possible to characterize the predominant adherence patterns in each region and tailor public health resources and strategies to suit the profiles associated with different healthcare facilities.

An extension of the latent class component of this research is in progress, applying similar techniques to a more recent adherence study called RETAIN. This research will evaluate latent adherence trajectories using more granular EAM data among a similar cohort of PLHW on the dolutegravir-based ART regimen. RETAIN includes a single-arm intervention aimed at improving adherence among individuals already identified as non-adherent. Furthermore, TFV-DP and viral outcomes at study entry and follow-up will be analyzed to evaluate the impact of the intervention on adherence and viral outcomes.

9 Conclusion

In summary, this research provides valuable insights into the complex relationship between antiretroviral adherence and viral outcomes among people living with HIV in South Africa. Through advanced statistical modeling techniques, including joint models and longitudinal latent class analysis, the research revealed important associations between adherence patterns and viral non-suppression. Key findings highlight the superiority of objective adherence measures over self-report, the impact of baseline factors like viral load and TB history on viral outcomes, and the existence of distinct adherence trajectory groups. Despite some limitations in the application, the analytical methods used in this research display clear potential for making material contributions to precision public health and precision medicine strategies in antiretroviral adherence interventions. An emphasis is placed on the need for ongoing, multifaceted adherence monitoring in clinical practice and directions for future research are suggested, including further qualitative research and more comprehensive collection of variables related to viral non-suppression. Ultimately, this dissertation underscores the critical importance of understanding and supporting antiretroviral adherence to optimize HIV treatment outcomes in resource-limited settings.

References

[1] Andersen, P. K. and Gill, R. D. "Cox's regression model for counting processes: a large sample study". In: *The annals of statistics* (1982), pp. 1100–1120.

[2] Arminger, G. and Sobel, M. E. "Pseudo-maximum likelihood estimation of mean and covariance structures with missing data". In: *Journal of the American Statistical Association* 85.409 (1990), pp. 195–203.

[3] Auguie, B. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3. 2017. URL: `https://CRAN.R-project.org/package=gridExtra`.

[4] Bangsberg, D. R. "Non-adherence to highly active antiretroviral therapy predicts progression to AIDS". In: *Aids* 15.9 (2001), pp. 1181–1183.

[5] Bellera, C. A. "Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer". In: *BMC medical research methodology* 10 (2010), pp. 1–12.

[6] Bender, A. *ldatools: Utility functions for life-time data analysis*. R package version 0.0.5, commit a0eeff1dff2635b586d4670183426c47619b200b. 2024. URL: `https://github.com/adibender/ldatools`.

[7] Bessong, P. O., Matume, N. D., and Tebit, D. M. "Potential challenges to sustained viral load suppression in the HIV treatment programme in South Africa: a narrative overview". In: *AIDS Res Ther* 18.1 (2021), p. 1.

[8] Breuer, E. "Reliability of the lay adherence counsellor administered substance abuse and mental illness symptoms screener (SAMISS) and the International HIV Dementia Scale (IHDS) in a primary care HIV clinic in Cape Town, South Africa". In: *AIDS and Behavior* 16 (2012), pp. 1464–1471.

[9] Brizzi, M. "Long-acting injectable antiretroviral therapy: will it change the future of HIV treatment?" In: *Therapeutic Advances in Infectious Disease* 10 (2023), p. 20499361221149773.

[10] Castillo-Mancilla, J. R. "Tenofovir diphosphate levels in dried blood spots are associated with virologic failure and resistance to first-line therapy in South Africa: a case–control cohort study". In: *Journal of the International AIDS Society* 24.12 (2021), e25849.

[11] Chesney, M. A. "Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: the AACTG adherence instruments". In: *AIDS care* 12.3 (2000), pp. 255–266.

[12] Chiou, S. H. "Regression Modeling for Recurrent Events Possibly with an Informative Terminal Event Using R Package reReg". In: *J. Stat. Softw.* (2023). DOI: `10.18637/jss.v105.i05`.

[13] *Clinical Guidelines for the Management of HIV AIDS in Adults and Adolescents*. Tech. rep. National Department of Health, South Africa, 2010.

[14] Collett, D. *Modelling survival data in medical research*. Chapman and Hall/CRC, 2023.

[15] Cox, D. R. "Regression models and life-tables". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2 (1972), pp. 187–202.

[16] Crowell, T. A. "Pretreatment and acquired antiretroviral drug resistance among persons living with HIV in four African countries". In: *Clinical Infectious Diseases* 73.7 (2021), e2311–e2322.

[17] Davies, C. E., Glonek, G. F., and Giles, L. C. "The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure". In: *Statistical Methods in Medical Research* 26.4 (2017), pp. 1982–1991.

[18] Dunn, P. K. and Smyth, G. K. "Randomized quantile residuals". In: *Journal of Computational and graphical statistics* 5.3 (1996), pp. 236–244.

[19] Evans, D. "Can short-term use of electronic patient adherence monitoring devices improve adherence in patients failing second-line antiretroviral therapy? Evidence from a pilot study in Johannesburg, South Africa". In: *AIDS and Behavior* 20 (2016), pp. 2717–2728.

[20] Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. *Applied Longitudinal Analysis*. Wiley, 2004.

[21] Fleming, T. R. and Harrington, D. P. *Counting Processes and Survival Analysis*. Hoboken, NJ: Wiley, 2005. ISBN: 9780471722235.

[22] Furtado dos Santos, S. "Does switching from multiple to single-tablet regimen containing the same antiretroviral drugs improve adherence? A group-based trajectory modeling analysis". In: *Aids Care* 32.10 (2020), pp. 1268–1276.

[23] Gelman, A. and Hill, J. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.
[24] Gelman, A., Hwang, J., and Vehtari, A. "Understanding predictive information criteria for Bayesian models". In: *Statistics and computing* 24 (2014), pp. 997–1016.
[25] Gill, C. J. "Importance of dose timing to achieving undetectable viral loads". In: *AIDS and Behavior* 14 (2010), pp. 785–793.
[26] *Global HIV AIDS Statistics — 2019 Fact Sheet*. Accessed May 17, 2024. 2019. URL: https://www.unaids.org/en/resources/fact-sheet.
[27] Grambsch, P. M. and Therneau, T. M. "Proportional hazards tests and diagnostics based on weighted residuals". In: *Biometrika* 81.3 (1994), pp. 515–526.
[28] Grolemund, G. and Wickham, H. "Dates and Times Made Easy with lubridate". In: *Journal of Statistical Software* 40.3 (2011), pp. 1–25. URL: https://www.jstatsoft.org/v40/i03/.
[29] Haberer, J. E. "Real-time adherence monitoring for HIV antiretroviral therapy". In: *AIDS and Behavior* 14 (2010), pp. 1340–1346.
[30] Hartig, F. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. R package version 0.4.6. 2022. URL: https://CRAN.R-project.org/package=DHARMa.
[31] Hartig, F. *DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models*. R package vignette. 2023. URL: https://cran.r-project.org/web/packages/DHARMa/vignettes/DHARMa.html (visited on 06/17/2024).
[32] Havenga, K. "The in vitro and in vivo effects of Hypoxis hemerocallidea on indinavir pharmacokinetics: modulation of efflux". In: *Planta Medica* 84.12/13 (2018), pp. 895–901.
[33] Hodes, R. "Pesky metrics: the challenges of measuring ART adherence among HIV-positive adolescents in South Africa". In: *Critical Public Health* 30.2 (2020), pp. 179–190.
[34] Horne, R., Weinman, J., and Hankins, M. "The beliefs about medicines questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication". In: *Psychology and health* 14.1 (1999), pp. 1–24.
[35] Ibrahim, J. G. and Molenberghs, G. "Missing data methods in longitudinal studies: a review". In: *Test* 18.1 (2009), pp. 1–43.
[36] Jackson, A. "Tenofovir, emtricitabine intracellular and plasma, and efavirenz plasma concentration decay following drug intake cessation: implications for HIV treatment and prevention". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 62.3 (2013), pp. 275–281.
[37] Jennings, L. "Tenofovir diphosphate in dried blood spots predicts future viremia in persons with HIV taking antiretroviral therapy in South Africa". In: *AIDS* 36.7 (2022), pp. 933–940.
[38] Jennings, L. "Comparing Predictive Ability of Two Objective Adherence Measures in a Community-Based Cohort on Antiretroviral Therapy in South Africa: Tenofovir Diphosphate Concentrations and Electronic Adherence Monitors". In: *Journal of Acquired Immune Deficiency Syndromes (1999)* 93.4 (2023), p. 327.
[39] Kaplan, E. L. and Meier, P. "Nonparametric estimation from incomplete observations". In: *Journal of the American statistical association* 53.282 (1958), pp. 457–481.
[40] Kassambara, A. *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.6.0. 2023. URL: https://CRAN.R-project.org/package=ggpubr.
[41] Kassambara, A., Kosinski, M., and Biecek, P. *survminer: Drawing Survival Curves using 'ggplot2'*. R package version 0.4.9. 2021. URL: https://CRAN.R-project.org/package=survminer.
[42] Keiser, O. "Outcomes of antiretroviral therapy in the Swiss HIV Cohort Study: latent class analysis". In: *AIDS and Behavior* 16 (2012), pp. 245–255.
[43] Kessler, R. C. "Short screening scales to monitor population prevalences and trends in non-specific psychological distress". In: *Psychological medicine* 32.6 (2002), pp. 959–976.
[44] Kessler, R. C. "Screening for serious mental illness in the general population". In: *Archives of general psychiatry* 60.2 (2003), pp. 184–189.
[45] Khorashadizadeh, F. "Predicting the survival of AIDS patients using two frameworks of statistical joint modeling and comparing their predictive accuracy". In: *Iranian Journal of Public Health* 49.5 (2020), p. 949.

[46] Khoury, M. J., Iademarco, M. F., and Riley, W. T. "Precision public health for the era of precision medicine". In: *American journal of preventive medicine* 50.3 (2016), p. 398.
[47] Kiwuwa-Muyingo, S. "Dynamic logistic regression model and population attributable fraction to investigate the association between adherence, missed visits and mortality: a study of HIV-infected adults surviving the first year of ART". In: *BMC infectious diseases* 13 (2013), pp. 1–14.
[48] Klein, J. P. and Moeschberger, M. L. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd. New York: Springer, 2005.
[49] Lambert, D. "Zero-inflated Poisson regression, with an application to defects in manufacturing". In: *Technometrics* 34.1 (1992), pp. 1–14.
[50] Little, R. J. and Rubin, D. B. *Statistical analysis with missing data*. Vol. 793. John Wiley & Sons, 2019.
[51] Lüdecke, D. "ggeffects: Tidy Data Frames of Marginal Effects from Regression Models." In: *Journal of Open Source Software* 3.26 (2018), p. 772. DOI: 10.21105/joss.00772.
[52] Luvanda, H. B., Mukyanuzi, E. N., and Akarro, R. R. "A joint survival model for estimating the association between viral load outcome and survival time to death among HIV/AIDS patients attending health care and treatment centers in Tanzania". In: *BMC Public Health* 23.1 (2023), p. 2091.
[53] Mantel, N. "Evaluation of survival data and two new rank order statistics arising in its consideration". In: *Cancer Chemother Rep* 50.3 (1966), pp. 163–170.
[54] Marcellin, F. "Assessing adherence to antiretroviral therapy in randomized HIV clinical trials: a review of currently used methods". In: *Expert Review of Anti-infective Therapy* 11.3 (2013), pp. 239–250.
[55] Mason, M. "Technologies for medication adherence monitoring and technology assessment criteria: narrative review". In: *JMIR mHealth and uHealth* 10.3 (2022), e35157.
[56] Mchunu, N. N. "Using joint models to study the association between CD4 count and the risk of death in TB/HIV data". In: *BMC Medical Research Methodology* 22.1 (2022), pp. 1–9.
[57] McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. "Finite mixture models". In: *Annual review of statistics and its application* 6 (2019), pp. 355–378.
[58] "Measuring stigma in people with HIV: Psychometric assessment of the HIV stigma scale¶". In: ().
[59] Mody, A. "Longitudinal engagement trajectories and risk of death among new ART starters in Zambia: A group-based multi-trajectory analysis". In: *PLoS medicine* 16.10 (2019), e1002959.
[60] Müller, A. C., Skinner, M. F., Kanfer, I. "Effect of the African traditional medicine, Sutherlandia frutescens, on the bioavailability of the antiretroviral protease inhibitor, atazanavir". In: *Evidence-based Complementary and Alternative Medicine* 2013 (2013).
[61] Naar-King, S. "A multisite randomized trial of a motivational intervention targeting multiple risks in youth living with HIV: Initial effects on motivation, self-efficacy, and depression". In: *Journal of Adolescent Health* 46.5 (2010), pp. 422–428.
[62] Nachega, J. B. "Adherence to highly active antiretroviral therapy assessed by pharmacy claims predicts survival in HIV-infected South African adults". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 43.1 (2006), pp. 78–84.
[63] Nagin, D. *Group-based modeling of development*. Harvard University Press, 2005.
[64] Nagin, D. S. "Group-based trajectory modeling: an overview". In: *Handbook of quantitative criminology* (2010), pp. 53–67.
[65] Nest, G. van der. "An overview of mixture modelling for latent evolutions in longitudinal data: Modelling approaches, fit statistics and software". In: *Current Perspectives on Aging and the Life Cycle* (2020). [Online] 43100323–.
[66] Nielsen, J. D. "Group-based criminal trajectory analysis using cross-validation criteria". In: *Communications in Statistics-Theory and Methods* 43.20 (2014), pp. 4337–4356.
[67] Nielsen, J. D. *crimCV: Group-Based Modelling of Longitudinal Data*. R package version 1.0.0. 2023. URL: https://CRAN.R-project.org/package=crimCV.
[68] Orrell, C. "A randomized controlled trial of real-time electronic adherence monitoring with text message dosing reminders in people starting first-line antiretroviral therapy". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 70.5 (2015), pp. 495–502.

[69] Orrell, C. "Comparison of six methods to estimate adherence in an ART-naıve cohort in a resource-poor setting: which best predicts virological and resistance outcomes?" In: *AIDS Research and Therapy* 14 (2017), pp. 1–11.

[70] Paterson, D. L. "Adherence to protease inhibitor therapy and outcomes in patients with HIV infection". In: *Annals of internal medicine* 133.1 (2000), pp. 21–30.

[71] Petersen, M. L. "Super learner analysis of electronic adherence data improves viral prediction and may provide strategies for selective HIV RNA monitoring". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 69.1 (2015), pp. 109–118.

[72] Pinheiro, J. and Bates, D. *Mixed-effects models in S and S-PLUS*. Springer science & business media, 2006.

[73] Pinheiro, J. C. and Chao, E. C. "Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models". In: *Journal of Computational and Graphical Statistics* 15.1 (2006), pp. 58–81.

[74] Plummer, M. "CODA: convergence diagnosis and output analysis for MCMC". In: *R news* 6.1 (2006), pp. 7–11.

[75] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2019. URL: `https://www.R-project.org/`.

[76] R Core Team. *foreign: Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'Systat', 'Weka', 'dBase', ...* R package version 0.8-86. 2023. URL: `https://CRAN.R-project.org/package=foreign`.

[77] Reay, R. "CYP2B6 haplotype predicts efavirenz plasma concentration in black South African HIV-1-infected children: a longitudinal pediatric pharmacogenomic study". In: *OMICS: A Journal of Integrative Biology* 21.8 (2017), pp. 465–473.

[78] Remien, R. H. "Adherence to medication treatment: A qualitative study of facilitators and barriers among a diverse sample of HIV+ men and women in four US cities". In: *AIDS and Behavior* 7 (2003), pp. 61–72.

[79] Remien, R. H. *Use of ARV Drug Levels in DBS to Assess and Manage ART Adherence in South Africa*. Study Protocol. Version 1.1. Aug. 2016.

[80] Rizopoulos, D. *Joint models for longitudinal and time-to-event data: With applications in R*. CRC press, 2012.

[81] Rizopoulos, D. "The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC". In: *arXiv preprint arXiv:1404.7625* (2014).

[82] Rizopoulos, D. *GLMMadaptive: Generalized Linear Mixed Models using Adaptive Gaussian Quadrature*. R package version 0.9-1. 2023. URL: `https://CRAN.R-project.org/package=GLMMadaptive`.

[83] Rizopoulos, D., Papageorgiou, G., and Miranda Afonso, P. *JMbayes2: Extended Joint Models for Longitudinal and Time-to-Event Data*. R package version 0.4-5. 2023. URL: `https://CRAN.R-project.org/package=JMbayes2`.

[84] Robbins, R. N. "Enhancing lay counselor capacity to improve patient outcomes with multimedia technology". In: *AIDS and Behavior* 19 (2015), pp. 163–176.

[85] Schloerke, B. *GGally: Extension to 'ggplot2'*. R package version 2.2.1. 2024. URL: `https://CRAN.R-project.org/package=GGally`.

[86] Schneider, J. "Better physician-patient relationships are associated with higher reported adherence to antiretroviral therapy in patients with HIV infection". In: *Journal of general internal medicine* 19 (2004), pp. 1096–1103.

[87] Seyoum, A., Ndlovu, P., and Temesgen, Z. "Joint longitudinal data analysis in detecting determinants of CD4 cell count change and adherence to highly active antiretroviral therapy at Felege Hiwot Teaching and Specialized Hospital, North-west Ethiopia (Amhara Region)". In: *AIDS research and therapy* 14 (2017), pp. 1–13.

[88] Sinxadi, P. Z. "Pharmacogenetics of plasma efavirenz exposure in HIV-infected adults and children in South Africa". In: *British journal of clinical pharmacology* 80.1 (2015), pp. 146–156.

[89] Smith, R. "Accuracy of measures for antiretroviral adherence in people living with HIV". In: *Cochrane Database of Systematic Reviews* 7 (2022).

[90] Spinelli, M. A. "Approaches to objectively measure antiretroviral medication adherence and drive adherence interventions". In: *Current HIV/AIDS Reports* 17 (2020), pp. 301–314.

[91] Spreafico, M. and Ieva, F. "Dynamic monitoring of the effects of adherence to medication on survival in heart failure patients: a joint modeling approach exploiting time-varying covariates". In: *Biometrical Journal* 63.2 (2021), pp. 305–322.

[92] Swart, M. "High predictive value of CYP2B6 SNPs for steady-state plasma efavirenz levels in South African HIV/AIDS patients". In: *Pharmacogenetics and Genomics* 23.8 (2013), pp. 415–427.

[93] Temesgen, A., Gurmesa, A., and Getchew, Y. "Joint modeling of longitudinal CD4 count and time-to-death of HIV/TB co-infected patients: A case of Jimma University Specialized Hospital". In: *Annals of Data Science* 5 (2018), pp. 659–678.

[94] Therneau, T. M. *A Package for Survival Analysis in R*. R package version 3.7-0. 2024. URL: `https://CRAN.R-project.org/package=survival`.

[95] *Updated recommendations on first-line and second-line antiretroviral regimens and post-exposure prophylaxis and recommendations on early infant diagnosis of HIV*. Tech. rep. World Health Organization, 2018.

[96] Viljoen, M. "Influence of CYP2B6 516G¿ T polymorphism and interoccasion variability (IOV) on the population pharmacokinetics of efavirenz in HIV-infected South African children". In: *European journal of clinical pharmacology* 68 (2012), pp. 339–347.

[97] Voisin, D. R. "A longitudinal analysis of antiretroviral adherence among young black men who have sex with men". In: *Journal of Adolescent Health* 60.4 (2017), pp. 411–416.

[98] Vrijens, B. "Modelling the association between adherence and viral load in HIV-infected patients". In: *Statistics in medicine* 24.17 (2005), pp. 2719–2731.

[99] Wagner, G. "Correlates of HIV antiretroviral adherence in persons with serious mental illness". In: *AIDS care* 16.4 (2004), pp. 501–506.

[100] Webel, A. R. "Measuring HIV self-management in women living with HIV/AIDS: a psychometric evaluation study of the HIV Self-management Scale". In: *JAIDS Journal of Acquired Immune Deficiency Syndromes* 60.3 (2012), e72–e81.

[101] Wei, L. J., Lin, D. Y., and Weissfeld, L. "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions". In: *Journal of the American Statistical Association* 84.408 (1989), pp. 1065–1073.

[102] Wheeler, A. P., Worden, R. E., and McLean, S. J. "Replicating group-based trajectory models of crime at micro-places in Albany, NY". In: *Journal of quantitative criminology* 32 (2016), pp. 589–612.

[103] Whetten, K. "A brief mental health and substance abuse screener for persons with HIV". In: *AIDS Patient Care & STDs* 19.2 (2005), pp. 89–99.

[104] Wickham, H. and Bryan, J. *readxl: Read Excel Files*. R package version 1.4.3. 2023. URL: `https://CRAN.R-project.org/package=readxl`.

[105] Wickham, H. "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43 (2019), p. 1686. DOI: `10.21105/joss.01686`.

[106] William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. R package version 2.4.3. Northwestern University. Evanston, Illinois, 2024. URL: `https://CRAN.R-project.org/package=psych`.

[107] Williams, A. B. "A proposal for quality standards for measuring medication adherence in research". In: *AIDS and Behavior* 17 (2013), pp. 284–297.

[108] Zhang, H. "Identification of self-management behavior clusters among people living with HIV in China: A latent class profile analysis". In: *Patient preference and adherence* (2021), pp. 1427–1437.

[109] Zhang, Z. "Time-varying covariates and coefficients in Cox regression models". In: *Annals of translational medicine* 6.7 (2018).

A Appendix

A.1 Additional Statistical Methods

A.1.1 Survival Model Diagnostics

To assess the overall fit of a Cox model, one can investigate the Cox-Snell residuals. For the i-th individual, this is defined as:

$$r_{Ci} = \exp(\hat{\boldsymbol{\gamma}}^T \mathbf{w}_i)\hat{H}_0(t_i)$$

where $\hat{\boldsymbol{\gamma}}$ is the vector of estimated regression coefficients, $\mathbf{w}_i$ represents the covariate values for the i-th individual, and $\hat{H}_0(t_i)$ estimates the baseline cumulative hazard function at t_i [14]. The Cox-Snell residuals should follow a censored Exp(1) distribution, as should the cumulative hazard function. To evaluate this, one can plot thus plot an estimate of the cumulative hazard function of these residuals against the residuals themselves, expecting the majority of points to fall along the 45 degree line.

Another diagnostic tool uses Martingale residuals to investigate whether the functional form of continuous covariates is appropriate in the model. These residuals are defined as:

$$r_{Mi} = \delta_i - \exp(\hat{\boldsymbol{\gamma}}^T \mathbf{w}_i)\hat{H}_0(t_i),$$

where δ_i is the censoring indicator.

Martingale residuals are useful for examining the functional form of a continuous covariate. After fitting a Cox model that excludes the covariate of interest, one can plot these residuals against the omitted covariate. A plot that is relatively flat, without a discernible trend, suggests that the functional form of the covariate is appropriately specified in the model [14].

One issue with Martingale residuals is that they are distributed asymmetrically between $-\infty$ and 1. Deviance residuals, which are transformations of Martingale residuals, are symmetrically distributed around zero under a correctly specified model [14]. These can thus be more easily used to investigate the functional forms, following the same procedure as outlined above for Martingale residuals.

Outlying observations can be identified by plotting deviance residuals against their corresponding linear predictors. Deviance residuals, which are transformations of Martingale residuals, are symmetrically distributed around zero under a correctly specified model [14]. Observations that appear as outliers in this plot may indicate a poor model fit for those data points and warrant further investigation.

Lastly, influential observations can be identified by generating score residuals, which are transformations of the Schoenfeld residuals used to assess the proportional hazards (PH) assumption. When plotted against the scaled values of the covariates, score residuals can help identify observations that disproportionately influence the parameter estimates. These influential data points should be examined further to ensure their validity and to understand their impact on the model [14].

A.2 Additional Results

A.2.1 Survival Diagnostics

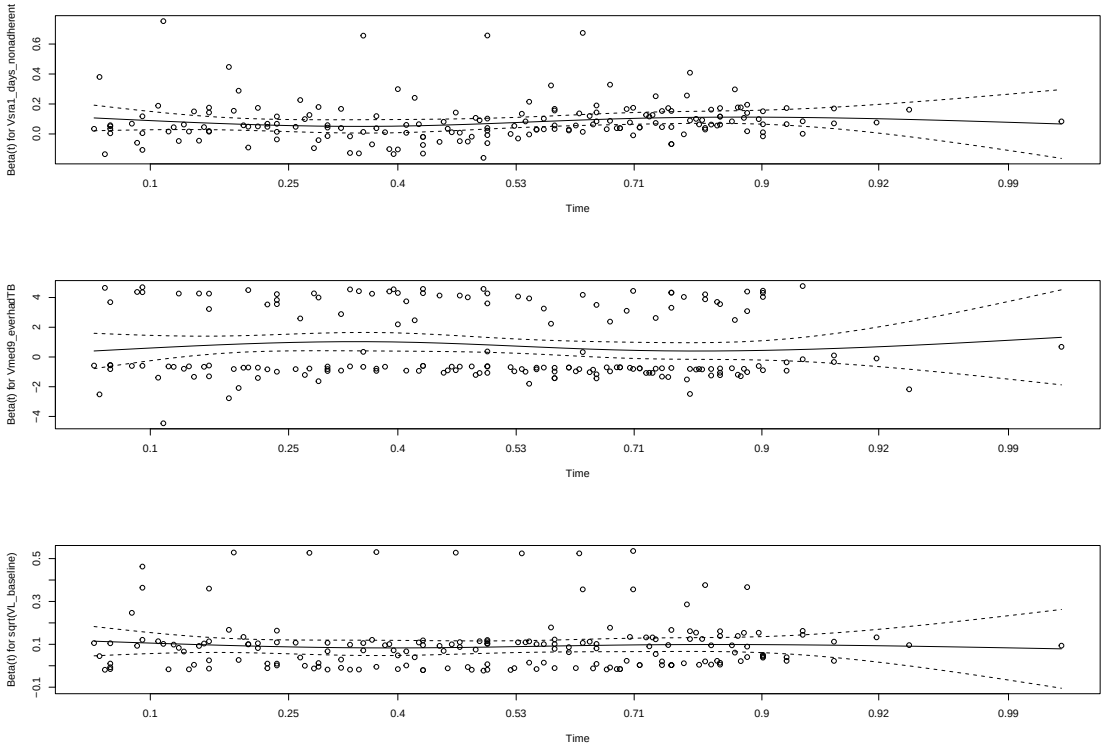


Figure A.1: SR marginal model *(c)*: Scaled Schoenfeld residuals against time

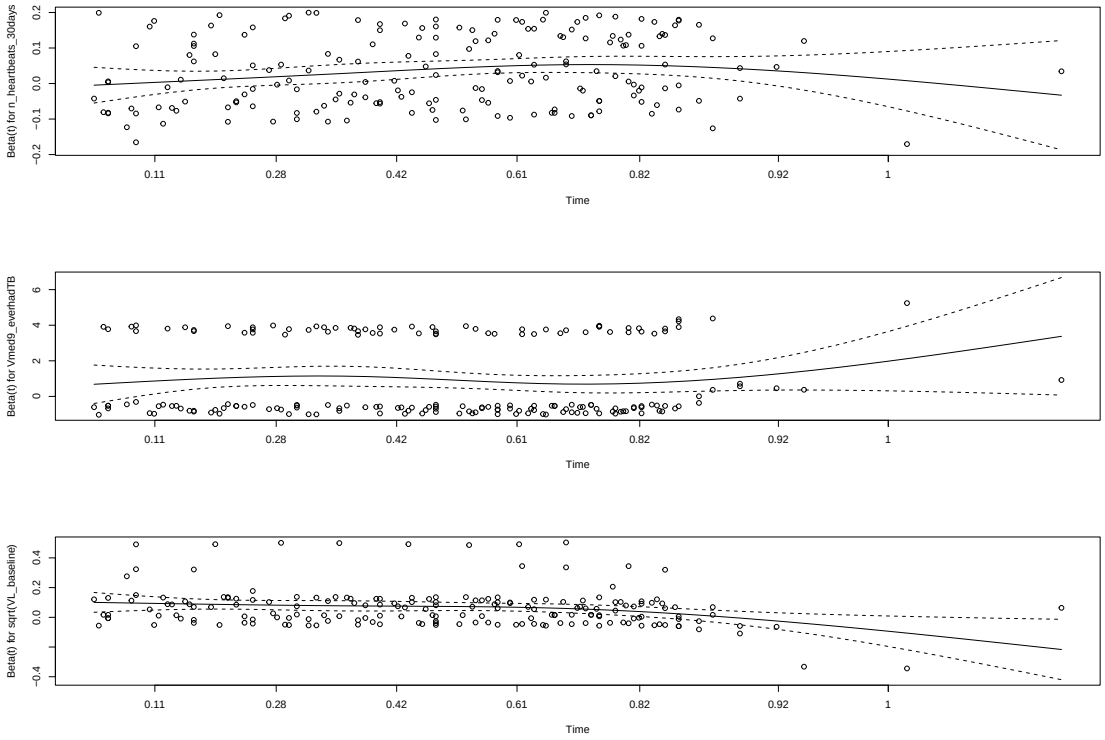


Figure A.2: EAM marginal model *(c)*: Scaled Schoenfeld residuals against time

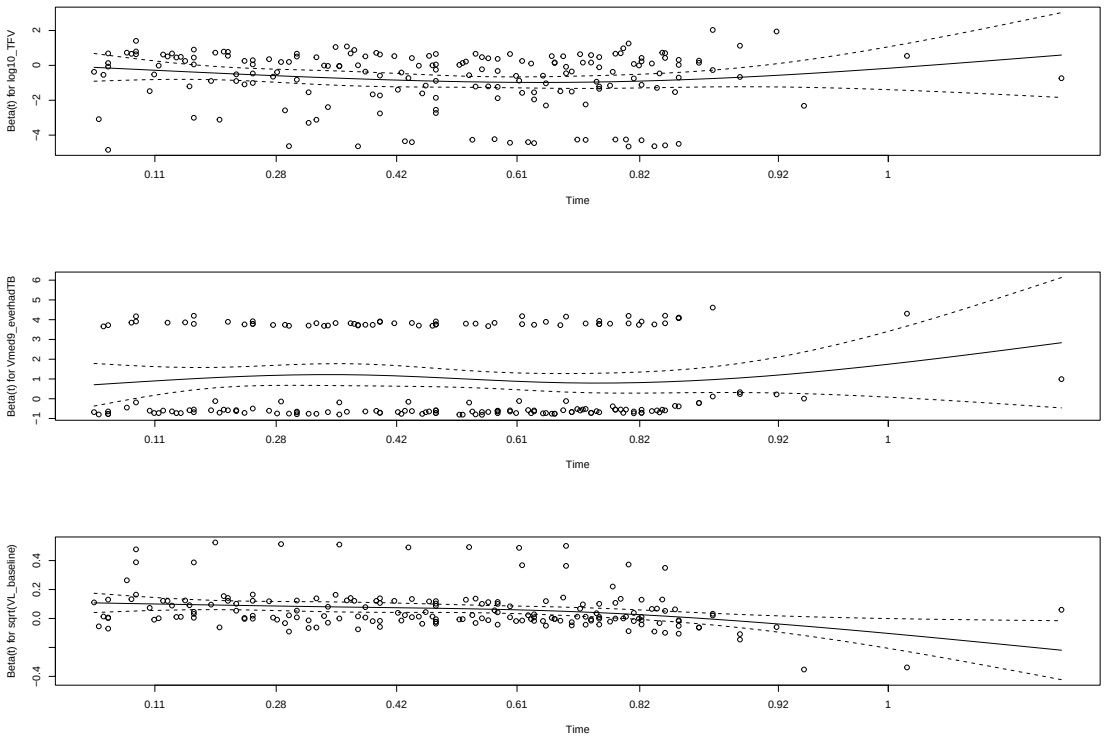
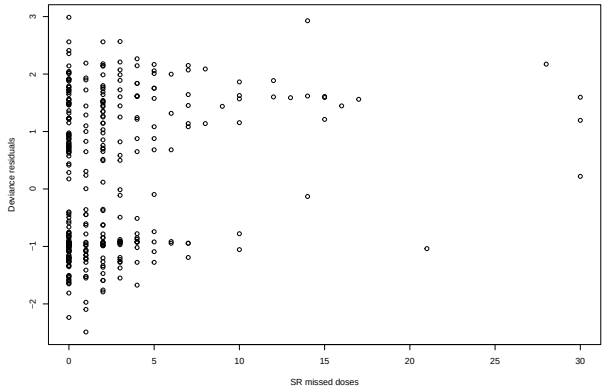
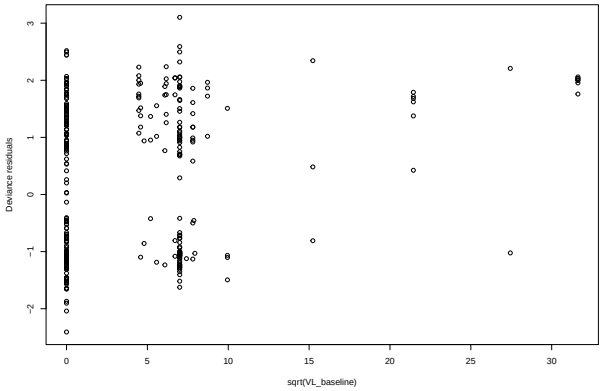


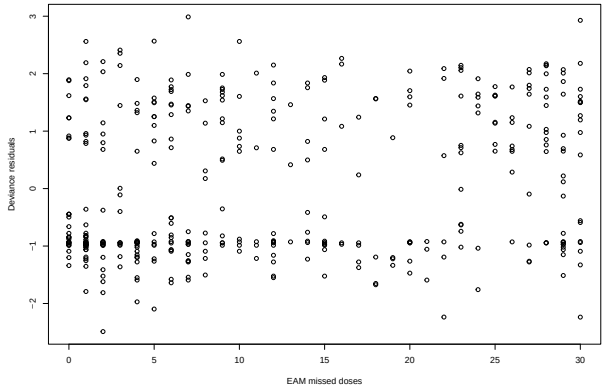
Figure A.3: TFV marginal model *(c)*: Scaled Schoenfeld residuals against time



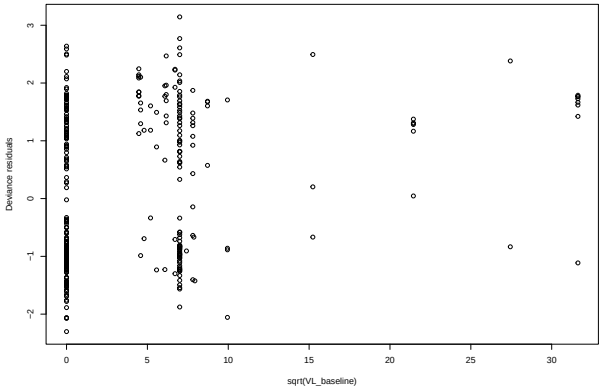
(a) SR model c: missed doses



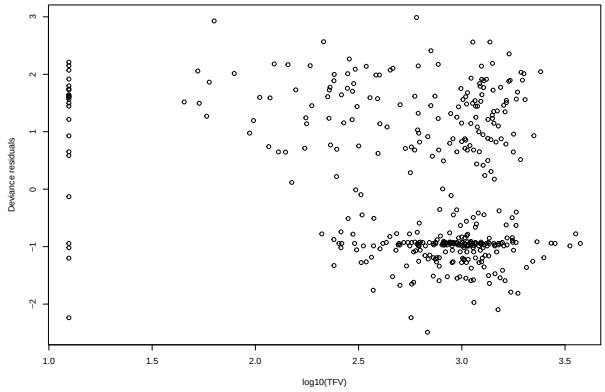
(b) SR model c: sqrt(baseline viral load)



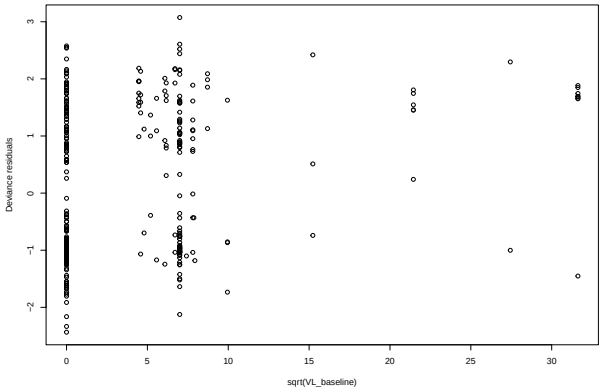
(c) EAM model c: missed doses



(d) EAM model c: sqrt(baseline viral load)

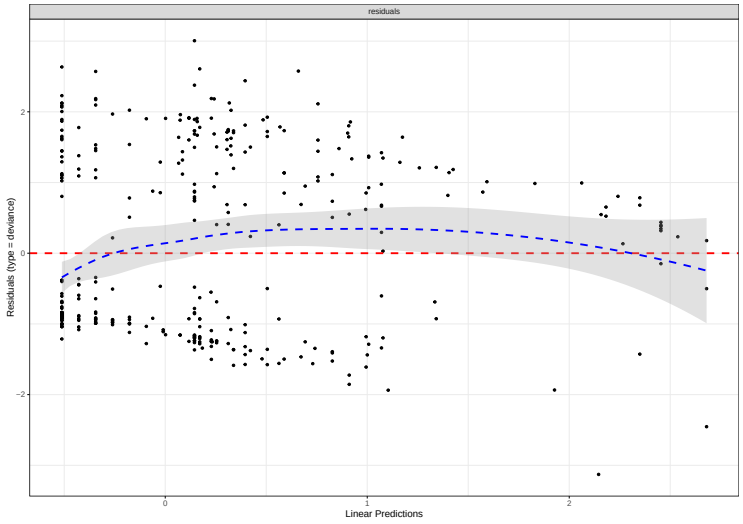


(e) TFV model c: TFV-DP

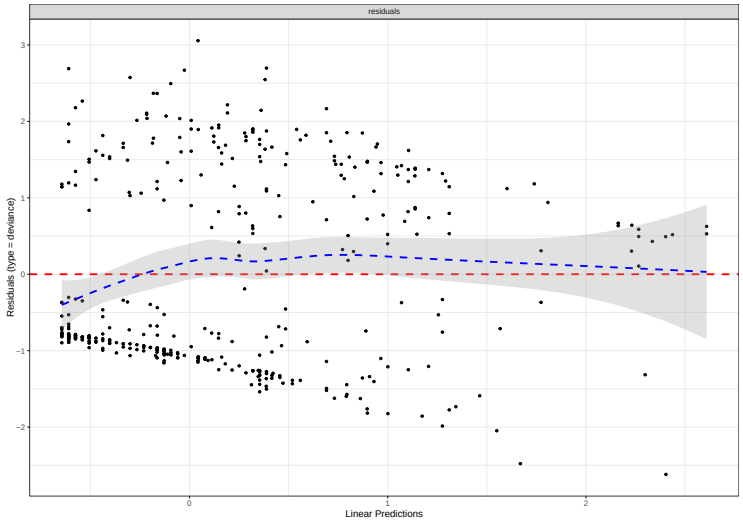


(f) TFV model c: sqrt(baseline viral load)

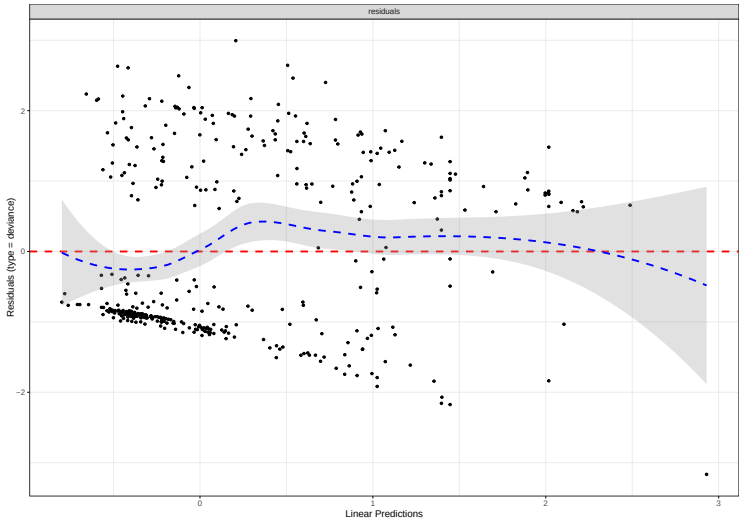
Figure A.4: Deviance residuals assessing functional form of numerical covariates



(a) SR model c

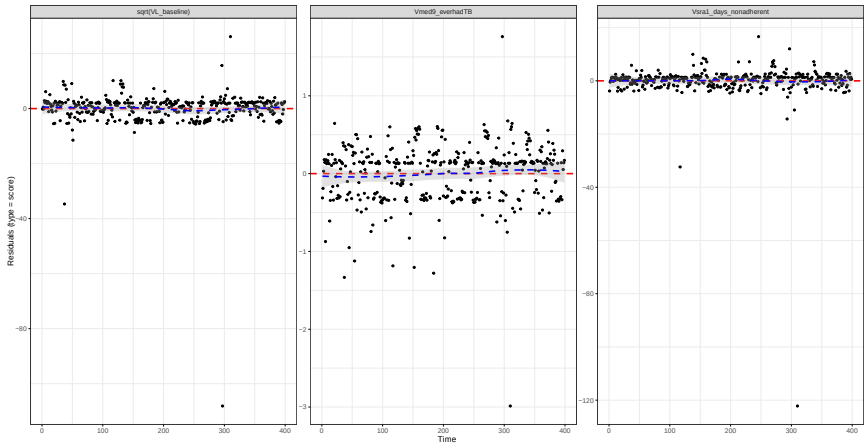


(b) EAM model c

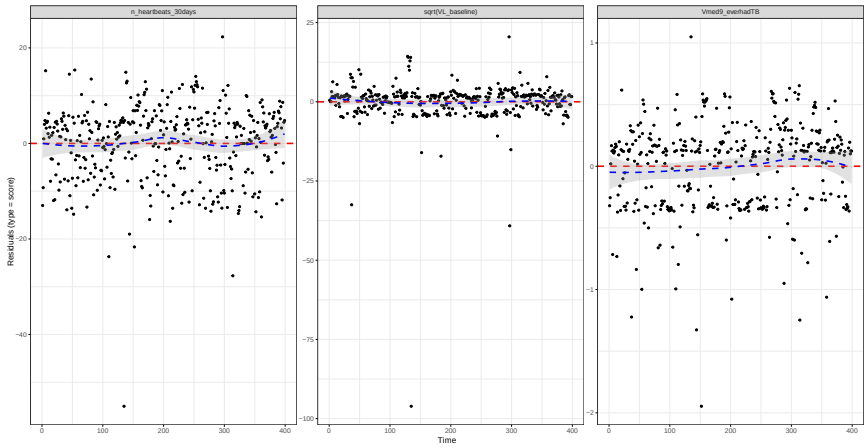


(c) TFV model c

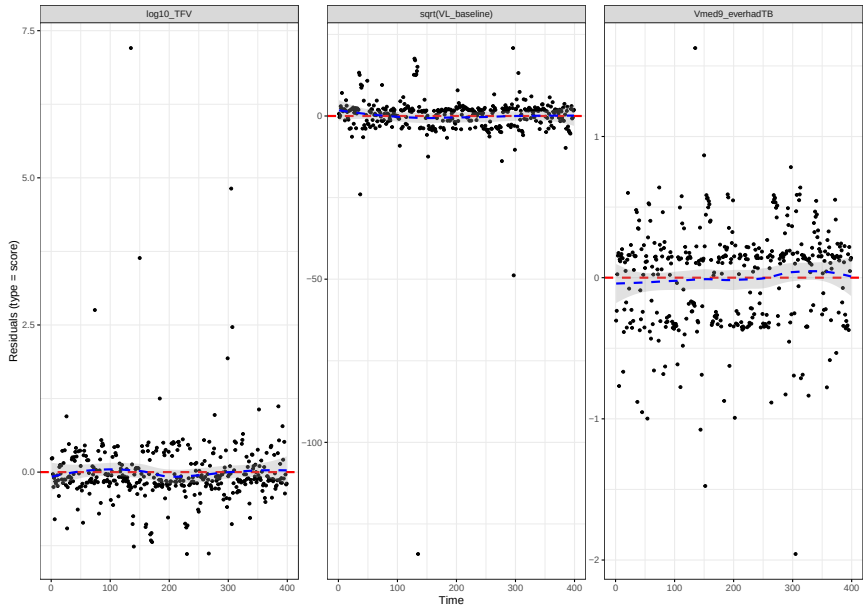
Figure A.5: Deviance residuals against linear predictors



(a) SR model c



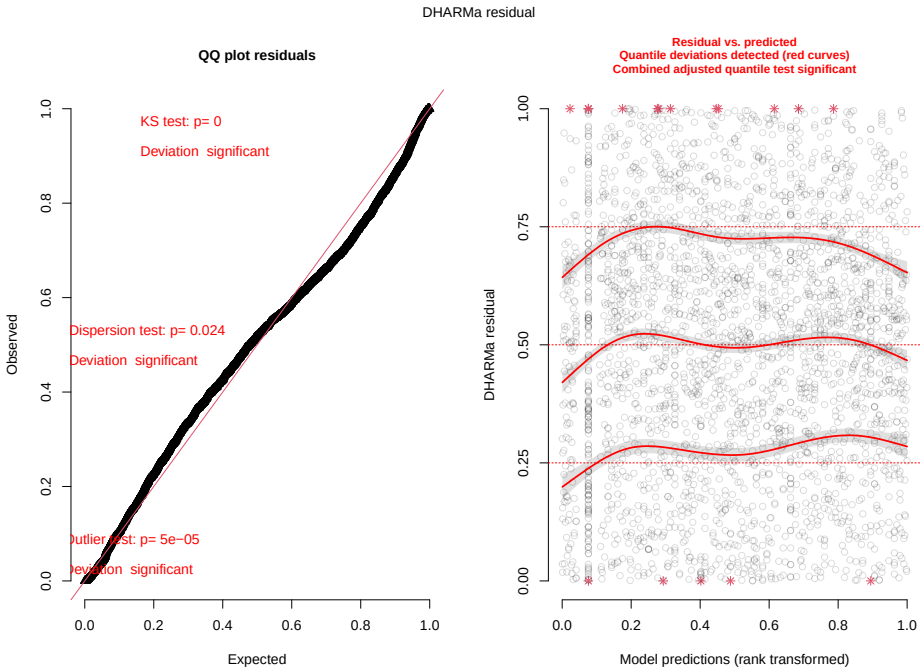
(b) EAM model c



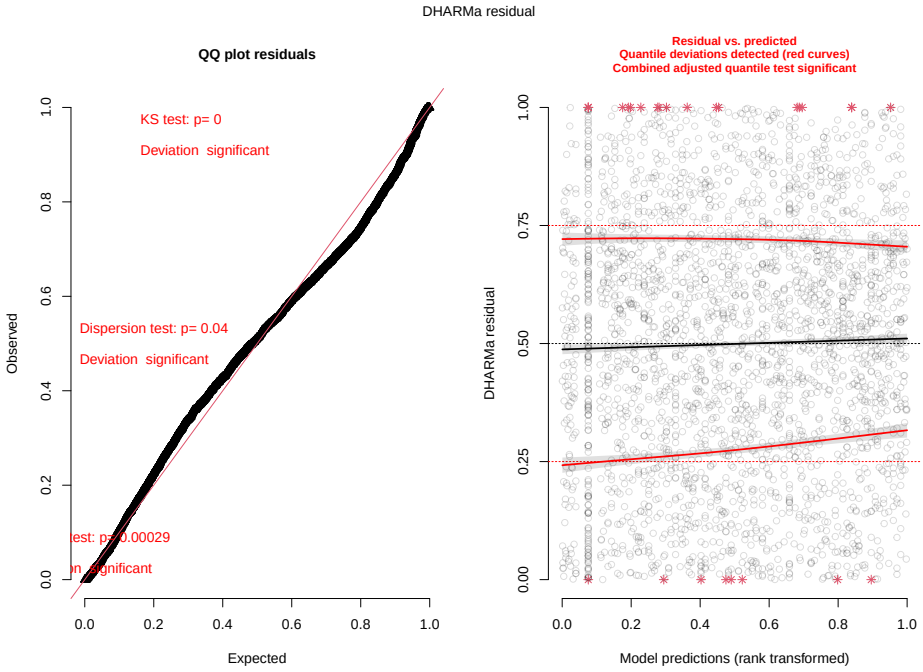
(c) TFV model c

Figure A.6: Score residuals to identify influential observations

A.2.2 Longitudinal Model Diagnostics



(a) Quadratic model



(b) Cubic model

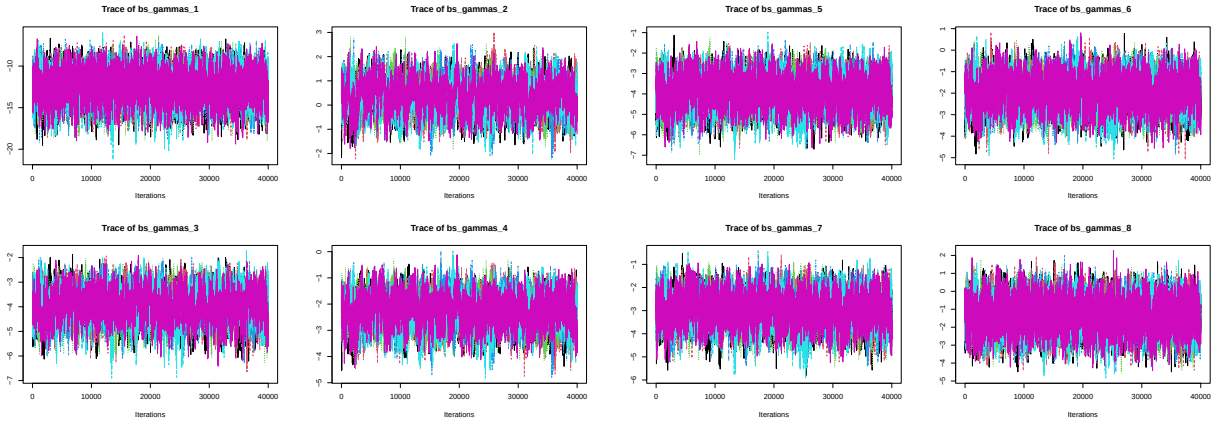
Figure A.7: Dharma residual plots for the Negative Binomial EAM models

Table A.1: Fixed Effects Estimates for Quadratic and Cubic Negative Binomial Models

Model	**Parameter**	**Estimate**	**Std.Err**	**z-value**	**p-value**
Quadratic	(Intercept)	-2.2025	0.0881	-25.0120	< 0.0001
	time_yrs	1.7725	0.1734	10.2207	< 0.0001
	I(time_yrs^2)	-1.1412	0.1588	-7.1853	< 0.0001
Cubic	(Intercept)	-2.2808	0.0902	-25.2828	< 0.0001
	time_yrs	2.9700	0.3552	8.3608	< 0.0001
	I(time_yrs^2)	-4.4927	0.8835	-5.0849	< 0.0001
	I(time_yrs^3)	2.4354	0.6323	3.8520	< 0.0002

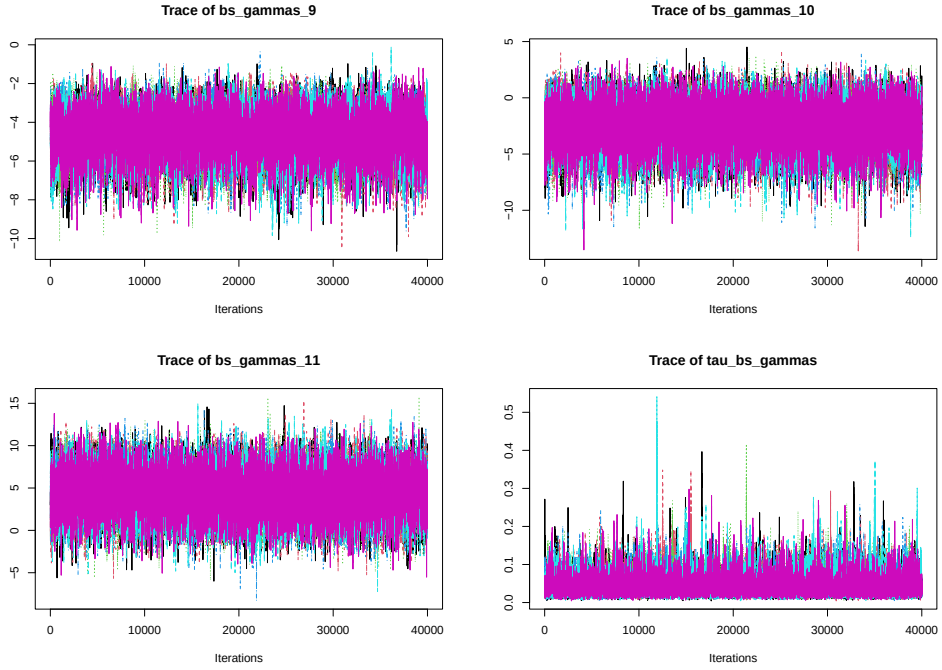
A.2.3 Longitudinal Model Results

A.2.4 Joint Model Diagnostics



(a) First set of parameters

(b) Second set of parameters



(c) Third set of parameters

Figure A.8: Traceplots to assess convergence in parameters used to estimate baseline hazard

A.2.5 Joint Model Results

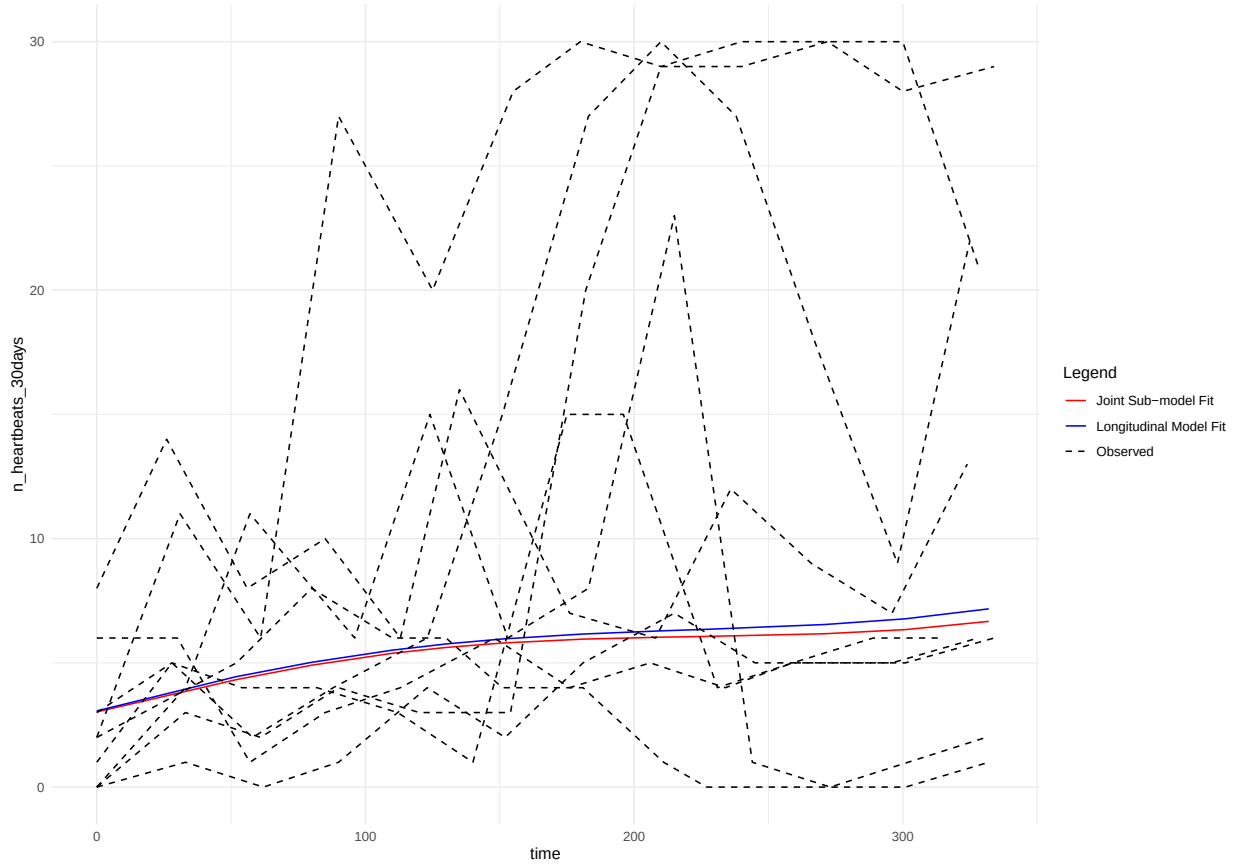


Figure A.9: Fitted curves for longitudinal sub-model and longitudinal model when fitted outside of joint model (with twelve random observed trajectories)

A.2.6 Latent Class Model Results

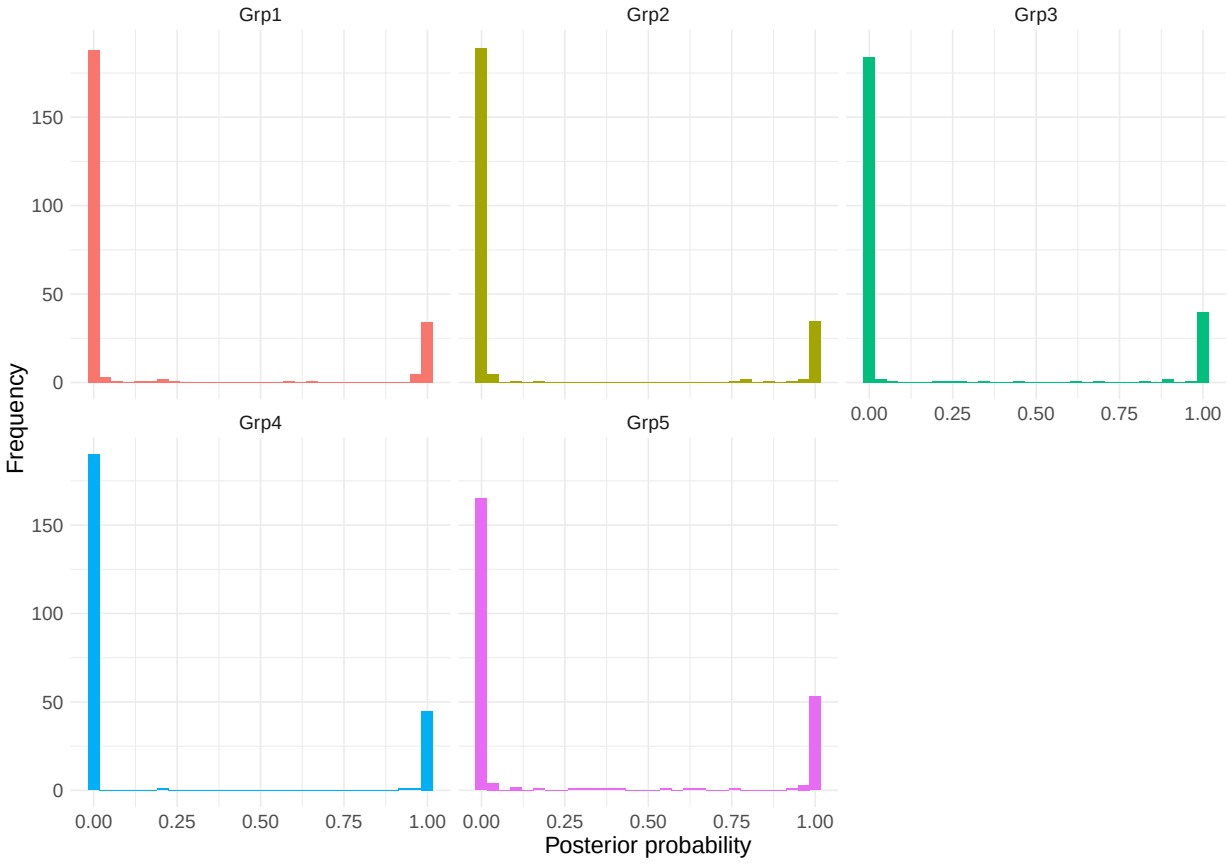
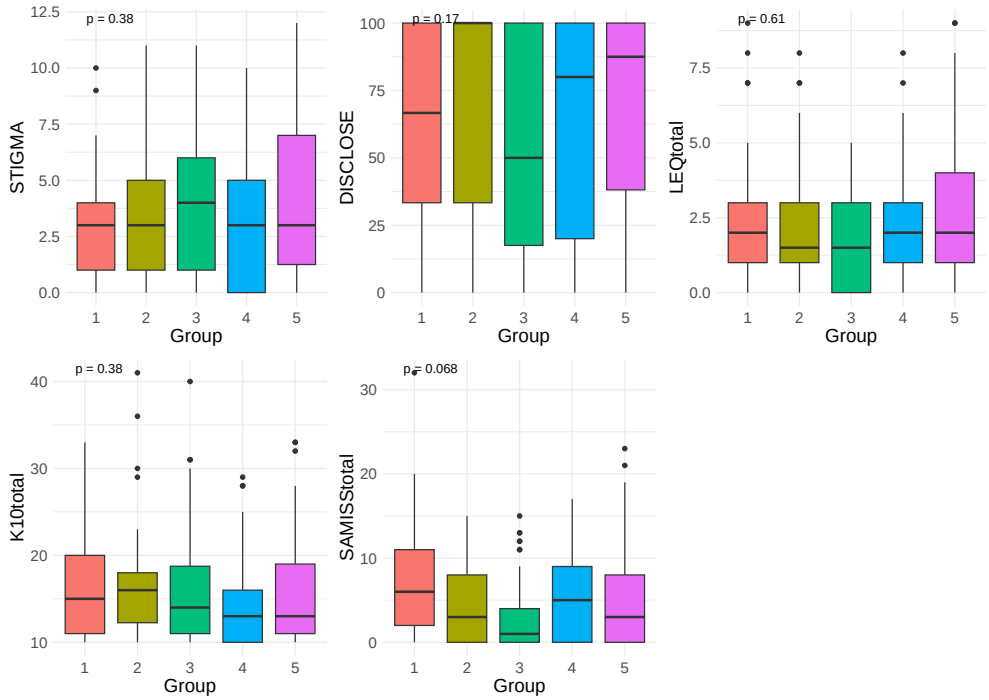
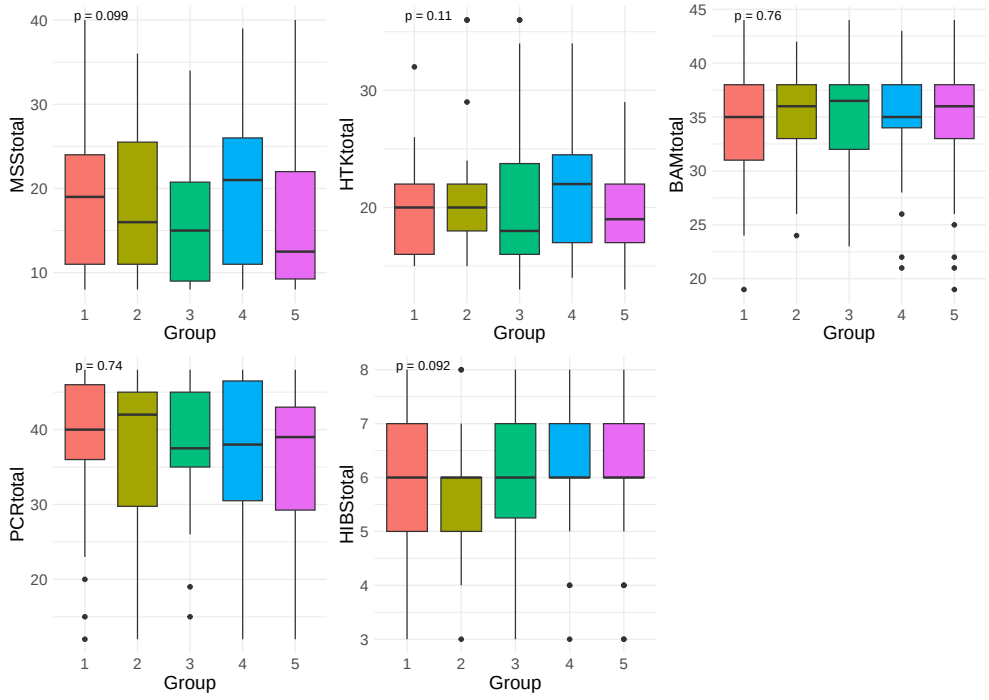


Figure A.10: ADD-ART GBTM: Distribution of posterior probabilities associated with each latent class



(a) Stigma, Disclosure, Life Events, Kessler Psychological Distress, Substance Abuse and Mental Illness



(b) Medication Social Support, HIV Treatment Knowledge, Beliefs About Medication, Patient-Clinic Relationship

Figure A.11: ADD-ART: Distribution of psychological, cultural and social scale scores for each GBTM group

A.3 Code Listing

All code used for data wrangling, data exploration and statistical modelling can be found at the following GitHub repository: `https://github.com/CampbellMcduling/MSc`