Application of dynamic prediction models for longitudinal biomarkers and clinical outcomes in low and middle-income settings

Frissiano Ernest Honwana

HNWFRI001

A thesis presented for the degree of

Doctor of Philosophy

in the Division of Epidemiology & Biostatistics,

School of Public Health,

Faculty of Health Sciences,

University of Cape Town

Date of submission: 3 July 2024

Revised version submitted: 16 May 2025

Supervisors: Professor Landon Myer, Professor Maia Lesosky and Associate Professor Freedom Gumedze

Abstract

Routine monitoring of individuals with chronic diseases offers valuable data for understanding disease progression and treatment effectiveness, often using biomarkers. With the modernisation of clinical care, prediction models have received greater attention in analysing such data. Prognosis prediction modelling approaches have been widely adopted, especially with digitising health records into electronic health records (EHRs). Dynamic prediction modelling has emerged as a critical approach, allowing real-time updates of prognosis predictions based on available data. However, there is a notable scarcity of dynamic prediction models applied to routine data from EHRs, particularly in contexts such as HIV and type 2 diabetes (T2DM) in resource-limited settings. Existing dynamic prediction models are typically developed and validated in data with comprehensive follow-ups and covariate collection, leading to the assumption of their universally improved predictive performance over traditional approaches such as the Cox proportional hazards-based prediction model. In addition to applying an extension of existing models to correctly model semicontinuous biomarker data (two-part joint model), this thesis challenges this assumption by applying dynamic prediction models using large routine data from EHRs generated in resource-limited settings, specifically focusing on using longitudinal biomarkers to predict probabilities of clinical outcomes in individuals with HIV or T2DM in South Africa. The predictive performance of this model is compared with that of the Cox proportional hazards-based prediction model and the two-part joint model. The prediction models had comparable predictive performances. The Cox proportional hazards-based prediction model had area under the curve (AUC) values ranging from 0.63 to 0.89 and Brier scores between 0.042 and 0.088 across routine T2DM and HIV data. The joint model had AUCs ranging between 0.66 and 0.73 and Brier scores between 0.033 and 0.089. The two-part joint model had AUCs and Brier scores closer to 0.6 and 0.1, respectively. These findings highlight the importance of adopting a conceptual approach to inform predictive performance, emphasising the need to account for context, type of disease, characteristics of a biomarker, and data characteristics. Such an approach will enhance individualised predictions using dynamic prediction models, potentially enabling recommendations for differentiated care and improving routine monitoring for individuals with chronic diseases, especially in resource-limited settings.

Declarations

This thesis is presented in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in the Division of Epidemiology & Biostatistics, School of Public Health, Faculty of Health Sciences, University of Cape Town.

The work included in this thesis is original research and has not, in whole or in part, been submitted for another degree at this or any other university. The contents of this thesis are entirely the work of the candidate, or in the case of conference outputs, constitutes work for which the candidate was the lead author.

The work has been presented at conferences and research seminars

- Honwana Frissiano, Mukonda Elton, Gumedze Freedom, Marvin Hsiao, Myer Landon, Lesosky Maia. Extended joint models for longitudinal viral load and virologic failure in HIV at Western Cape (South Africa). Poster presentation. The 44th Annual Conference of the International Society for Clinical Biostatistics, 2023, Milan, Italy.
- Honwana Frissiano, Mukonda Elton, Gumedze Freedom, Marvin Hsiao, Myer Landon, Lesosky Maia. The prospect of dynamic prediction models to inform individualized decision-making: an application to electronic health records. Oral presentation. The 10th Annual School of Public Health Research Day, 2022, Cape Town, Western Cape, South Africa.
- Honwana Frissiano, Mukonda Elton, Gumedze Freedom, Marvin Hsiao, Myer Landon, Lesosky Maia. Dynamic prediction of virologic failure in a cohort of HIV infected individuals on antiretroviral therapy in the Western Cape. Oral presentation. The 62nd annual conference of the South African Statistical Association, 2021, Stellenbosch, Western Cape, South Africa.

Date: 16 May 2025

Signed by candidate

Frissiano Ernest Honwana (HNWFRI001)
Signed by: d9214cc3-7812-416e-9e39-39a44686480f

Acknowledgements

This PhD thesis marks the culmination of a challenging and rewarding academic journey. In Siswati, there is a saying, “injobo itfungelwa ebandla”, meaning that seeking advice from others can lead to successful completion of a task and does not diminish one’s abilities. I am deeply grateful to those who provided me with support throughout this journey.

First and foremost, I would like to express my sincere gratitude to my supervisors. Maia, I am so grateful to have been your student. Your guidance has been instrumental in both my academic and personal growth, and your scientific insights and mentorship have been invaluable in shaping this work. Landon, your guidance and scientific insights have been crucial. You have consistently demanded the best from me and pushed me in the right direction while encouraging me to find my own path. Freedom, I appreciate your support and the scientific insights you have contributed to this thesis.

Thank you to all my supervisors for sharing your expertise and supporting this work.

To my colleagues, I am grateful for the hindsight and encouragement you have provided.

To my family, thank you for your unwavering love and support. My mom, Ofelia, and dad, Francisco, I know you are proud of me, and I am deeply thankful for everything you have done for me throughout my life. My little brother, Andile, you have been a source of positivity, and my older brother, Bongani, you have been my biggest cheerleader.

Auneta and Lungile: You have been a source of positivity, love, and support. Thank you for your patience and understanding throughout this academic journey.

To you all, thank you.

I would like to acknowledge Dr Jody Rusch from the South Africa’s National Health Laboratory Services in the Division of Chemical Pathology (University of Cape Town, Groote Schuur Hospital) for providing the secondary datasets used in this thesis. The content of this thesis does not represent the official views of the South Africa’s National Health Laboratory Services.

Finally, I would like to acknowledge the funding from the South African Department of Higher Education and Training (DHET) through its New Generation of Academics Programme (nGAP). The content of this thesis does not represent the official views of the DHET.

List of Abbreviations

EHRs	Electronic health records
HIV	Human immunodeficiency virus
T2DM	Type 2 diabetes
SA	South Africa
CD4	Cluster differentiation four
ART	Antiretroviral therapy
UNAIDS	United Nations Programme on HIV/AIDS
VF	Virologic failure
WHO	World Health Organisation
HbA1c	Haemoglobin A1c
LME	Linear mixed-effects
ML	Maximum likelihood
REML	Restricted maximum likelihood
GLMMs	Generalised linear mixed models
MCMC	Markov chain Monte Carlo
GLMs	Generalised linear models
KM	Kaplan-Meier
PH	Proportional hazard
SREM	Shared random effects model
CI	Credible interval or confidence interval
AUC	Area under the curve
ROC	Receiver operating curve
IPCW	Inverse probability of censoring weights

BS	Brier score
PE	Prediction error
IBS	Integrated Brier score
ROB	Risk of bias
CHARMS	Checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies
PROBAST	Prediction model risk of bias assessment tool
TRIPOD	Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis
PRISMA	Preferred reporting items for systematic reviews and meta-Analyses
USA	United States of America
JM	Joint model
JLCM	Joint latent class models
NHLS	National Health Laboratory Service
NPP	National Priority Programme
IQR	Interquartile range
HREC	Human Research Ethics Committee
HR	Hazard ratios
BIC	Bayesian information criteria
Df	Degrees of freedom
DIC	Deviance information criterion
TPM	Two-part model
MSE	Mean squared error
LKJ	Lewandowski-Kurowicka-Joe
NUTS	No-U-turn sampler

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Introduction and Background

Routine monitoring of chronic diseases using biomarkers is a core part of clinical care worldwide to optimise patient management and improve clinical outcomes. Monitoring chronic diseases using biomarkers has allowed the development of prediction models commonly used to estimate the prognosis of individuals (Steyerberg et al., 2010). Prediction models based on observed individuals' longitudinal trajectories of a biomarker are generally termed dynamic prediction models. Dynamic prediction models permit prognosis prediction that can be updated with each new data point, eventually permitting real-time prognosis prediction (Rizopoulos, 2011, 2012).

Despite the recognised advantages of dynamic prediction models over standard prediction models, there remains a lack of research on their application in settings where data quality is often difficult to assess and may be poor, such as routinely collected data from electronic health records (EHR), particularly in resource-limited settings. In such data, the improved predictive performance of dynamic prediction models over standard prediction models may not hold true. This thesis aims to address these gaps by developing and evaluating the predictive performance dynamic prediction models using routine data from EHRs from resource-limited settings, focusing on diseases with high burdens, such as human immunodeficiency virus (HIV) and type 2 diabetes (T2DM), in a resource-limited health system like South Africa (SA). The burden and context of these diseases in resource-limited settings are discussed in detail in Section 1.5.

Most prediction models are based on time-to-event analysis, exemplified by widely known risk score models like the Framingham risk score (Anderson, Odell, Wilson, & Kannel, 1991; D'Agostino Sr et al., 2008) and the QRISK (Hippisley-Cox, Coupland, Robson, & Brindle, 2010). In particular, the QRISK score is updated yearly to provide periodic updates on the risk of cardiovascular disease (Jenkins, Sperrin, Martin, & Peek, 2018). However, these prediction models are based on population or subgroup averages and cannot capture or model individual-specific dynamic changes over time. In contrast, dynamic prediction models allow for individualised predictions that can be continuously updated with new data points, facilitating real-time disease progression predictions (Rizopoulos, 2011, 2012).

Historically, dynamic prediction models have been applied to study opportunistic infections or death in HIV-infected individuals using cluster differentiation four (CD4) cell count measurements (Rizopoulos, 2011; Wang & Jeremy, 2001) and time to recurrence in prostate cancer studies using prostate-specific antigen (Proust-Lima & Taylor, 2009; Tsiatis & Davidian, 2004; Yu, Law, Taylor, & Sandler, 2004; Yu, Taylor, & Sandler, 2008). Applications were predominantly extended to conditions in resource-rich settings, for example, patients with high-grade extremity soft tissue sarcoma (Rueten-Budde, Van Praag, Van de Sande, & Fiocco, 2018); patients who received a human tissue valve in aortic position (Andrinopoulou, E. , Rizopoulos, Takkenberg, & Lesaffre, 2017) and inflammation in uveitis patients (Grand, Vermeer, Missotten, & Putter, 2019).

1.2 Landmarking and Joint Modelling for Dynamic Prediction

Joint modelling and landmarking are two popular approaches used in the literature for dynamically predicting clinical outcome probabilities based on longitudinal biomarkers. Landmarking (Van Houwelingen, 2007; Van Houwelingen & Putter, 2011) provides a simple tool for dynamic predictions that can be implemented easily. This approach is computationally efficient, as it does not require specifying the joint distribution of a longitudinal biomarker and a time-to-event (Van Houwelingen & Putter, 2011). However, for predictions to be consistent across different prediction time points, the prediction function should contain the joint distribution of the longitudinal biomarker and the time-to-event (Jewell & Nielsen, 1993; Rizopoulos, Molenberghs, & Lesaffre, 2017; Van Houwelingen & Putter, 2011). Several approaches have been proposed to incorporate a longitudinal biomarker in landmarking (Nicolaie, Van Houwelingen, de Witte, & Putter, 2013; Van Houwelingen & Putter, 2011) and the conditional expectation of a biomarker to bridge the gap between joint modelling and landmarking-based dynamic prediction models (Putter & Van Houwelingen, 2022). The findings in the study by Putter and Van Houwelingen (2022) suggested that further applications are needed to compare the extended landmarking approach with other approaches.

While landmarking provides a pragmatic approach to dynamic prediction, it avoids specifying the joint distribution of the longitudinal biomarker and the time-to-event (Rizopoulos et al., 2017). On the other hand, a joint model enables a complete specification of the joint distribution of the longitudinal biomarker and time-to-event through random effects (Henderson, Diggle, & Dobson, 2000; Rizopoulos, 2012; Tsiatis & Davidian, 2004). When correctly specified, it

provides consistent predictions (Putter & Van Houwelingen, 2022; Rizopoulos et al., 2017). While landmarking represents a valuable contribution to the literature, this thesis focuses on leveraging the strength of joint modelling to provide dynamic predictions.

1.3 Problem Statement and Rationale

Individualised prognosis prediction using dynamic prediction models is important as it presents an opportunity for personalised approaches to the care of individuals with a disease (Andrinopoulou, E., Harhay, Ratcliffe, & Rizopoulos, 2021). However, existing research predominantly focuses on developing these models using data from cohort studies conducted under controlled research conditions. This body of knowledge overlooks the challenges posed by data quality limitations in settings where data in electronic health records are routinely collected for logistical purposes only. Consequently, the assumed universal improved predictive performance of dynamic over standard prediction models may not hold true in these settings, where data quality is often challenging to assess and may be poor. This discrepancy highlights the need for research that specifically addresses the development, validation and application of dynamic prediction models using routine data, particularly in resource-limited settings where these challenges are pronounced.

Given the lack of research regarding the application of dynamic prediction for individualised prognosis in routine data from electronic health records, this thesis aimed to apply and evaluate dynamic prediction models using routine data generated in, and therefore reflective of, resource-limited settings to predict individual risk of a clinical outcome in motivating disease case studies of high burden in South Africa.

This thesis contributes to the body of knowledge on dynamic prediction models for individualised prognosis prediction by evaluating and providing insights to dynamic models for individualised prognosis prediction in large routine data from EHRs in resource-limited settings where data quality is difficult to assess and may be poor. This helps address the current shortage of research on dynamic prediction models applied in large routine data from EHRs and provide real-world individualised prognosis predictions that are reflective of resource-limited settings.

1.4 Aim and Objectives

This thesis aimed to provide insights into applying dynamic prediction models in using longitudinally measured biomarkers to predict the individual risk of a clinical outcome in selected disease cases of high burden in resource-limited health systems.

The specific objectives were:

1. To outline key reported characteristics in the conduct and quality of model evaluation of dynamic prediction models for individualised prognosis prediction.
2. To evaluate the predictive accuracy of Cox proportional hazard-based prediction models for prognosis prediction in large routine data in EHRs from resource-limited settings.
3. To compare and contrast Cox proportional hazard-based and dynamic prediction models for prognosis prediction in large routine data from EHRs from resource-limited settings regarding predictive performance.
4. To develop a novel statistical model for dynamic prediction to allow for different types of longitudinal continuous biomarker-outcome data.
5. To apply and compare the performance of dynamic prediction models with a novel dynamic prediction model in large routine data from EHRs in resource-limited settings and simulated data.

1.5 Burden of Disease in Resource-Limited Settings

Limited access to direct health care worsens health conditions in resource-limited settings. Ongoing monitoring and treatment of individuals burden the already under-resourced health systems in these settings (Beaglehole et al., 2008). With limited resources, improved clinical care and reducing unnecessary routine monitoring for individuals with well-managed chronic illnesses are important. Resource-limited countries like South Africa are facing a quadruple burden of chronic, non-chronic, perinatal, maternal and injury-related disorders (Pillay-van Wyk et al., 2016). Chronic diseases like HIV and T2DM have current and future impacts and burden the health care system in South Africa. The context for these diseases is given in sub-sections 1.5.1 and 1.5.2.

1.5.1 HIV

HIV remains a disease of significant burden in sub-Saharan Africa, although effective antiretroviral therapy (ART) is now widely available (Kharsany & Karim, 2016). Sub-Saharan Africa accounts for 71% of the global burden of HIV infection, with the highest incidence rates (23%) coming from South Africa (Kharsany & Karim, 2016). More than 20.6 million people were living with HIV in eastern and southern Africa in 2018 (United Nations Programme on HIV/AIDS [UNAIDS], 2019). In South Africa, 7.7 million people were living with HIV in 2018, of whom 62% were on ART (UNAIDS, 2019). For individuals living with HIV to remain healthy, treatment must ideally be initiated early and maintained for the remainder of life (World Health Organization [WHO], 2015). Lifelong treatment represents a significant burden to individuals as high levels of adherence to current therapeutic regimens are required to prevent developing drug resistance (WHO, 2016b). HIV RNA viral load measurement is the method of choice to monitor treatment adherence and development of resistance (WHO, 2016a), replacing clinical or CD4 cell count methods.

The 2019 South African national guidelines call for HIV viral load testing at least annually (South African National Department of Health, 2015, 2019a). HIV viral load testing tends to be suboptimal in resource-limited settings for various reasons, including complexity and associated costs (Pham, Nguyen, Anderson, Crowe, & Luchters, 2022; Roberts, Cohn, Bonner, & Hargreaves, 2016). However, South Africa's annual HIV viral load test volume increased from 1.96 million in 2013 to 6.27 million in 2022 (Hans et al., 2023). This provides an opportunity to use a longitudinal HIV viral load trajectory or multiple HIV viral load tests in detecting virologic failure, preventing drug resistance, and monitoring treatment adherence and effectiveness (Keiser et al., 2011; Mesic et al., 2021; Rowley, 2014). Figure 1.1 represents a cascade for the management of HIV viral load in individuals living with HIV on ART in South Africa. Viral non-suppressed individuals (HIV viral load ≥1000 copies/mL) can progress to adverse health outcomes such as virologic failure (VF). Virologic failure has severe implications for the quality of life of individuals living with HIV, particularly if these are not identified promptly (Saura-Lázaro et al., 2023).

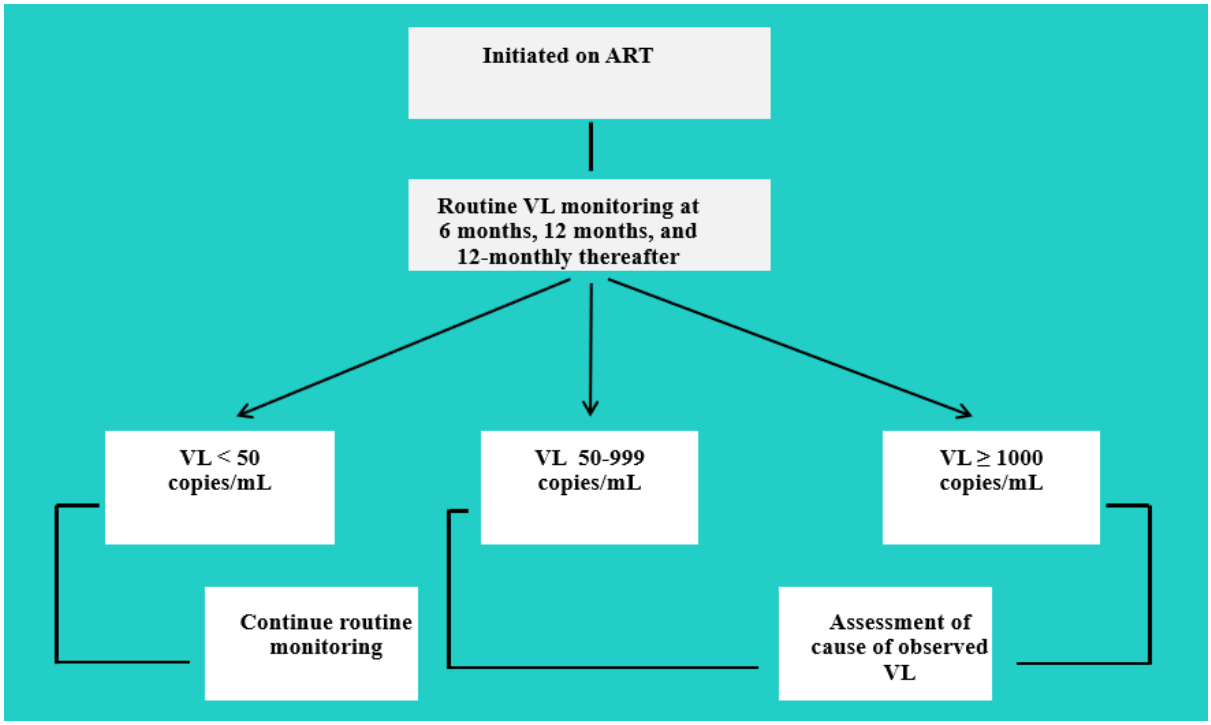

Figure 1.1: Management of HIV viral load in individuals living with HIV on ART in South Africa. Adapted from South Africa's 2019 ART clinical guidelines (South African National Department of Health, 2019a).

1.5.2 Diabetes mellitus

According to the World Health Organisation, globally, diabetes mellitus, broadly known as diabetes, is on the rise, increasing from an estimated 108 million in 1980 to 422 million adults in 2014 (WHO, 2016c). It has historically been of higher prevalence in resource-rich settings, but it is rapidly increasing in resource-limited settings (Aschner et al., 2020; Pheiffer et al., 2021). Diabetes is strongly associated with an increased risk of cardiovascular diseases, and, coupled with an ageing population, it represents a significant public health concern in many resource-limited settings (Coetzee et al., 2019; WHO, 2011, 2016c). The global prevalence of diabetes was estimated to have increased from 4.7% in 1980 to 8.5% in 2014 (WHO, 2016c). South Africa's prevalence of T2DM was estimated to be 12.8% (International Diabetes Federation, 2019). In addition to infectious diseases, the burden of non-communicable diseases adds to the strain on already resource-constrained healthcare settings such as those found in South Africa (Mayosi et al., 2009). T2DM is commonly monitored by regular measurement of haemoglobin A1c (HbA1c).

South African national guidelines recommend that HbA1c be used to diagnose T2DM and monitor glycaemic control (South African National Department of Health, 2014, 2019b, 2020). Although blood glucose level monitors are now available, they have minimal uptake in resource-limited settings, and HbA1c testing remains the common method of choice for monitoring diabetes progression (Beck, Connor, Mullen, Wesley, & Bergenstal, 2017; Chehregosha, Khamseh, Malek, Hosseinpanah, & Ismail-Beigi, 2019). Figure 1.2 shows a cascade of monitoring HbA1c in T2DM individuals and the frequency of tests in South Africa. As shown in the figure, T2DM individuals have an HbA1c test at 3-6 months until therapeutic goals have been achieved and annually if therapeutic goals are achieved (South African National Department of Health, 2014). HbA1c control, referred to hereafter as glycaemic control, is one of the therapeutic goals. Glycaemic control is associated with better long-term health outcomes among people living with diabetes and can reduce diabetes-related complications (Brennan et al., 2023; Manne-Goehler et al., 2019).

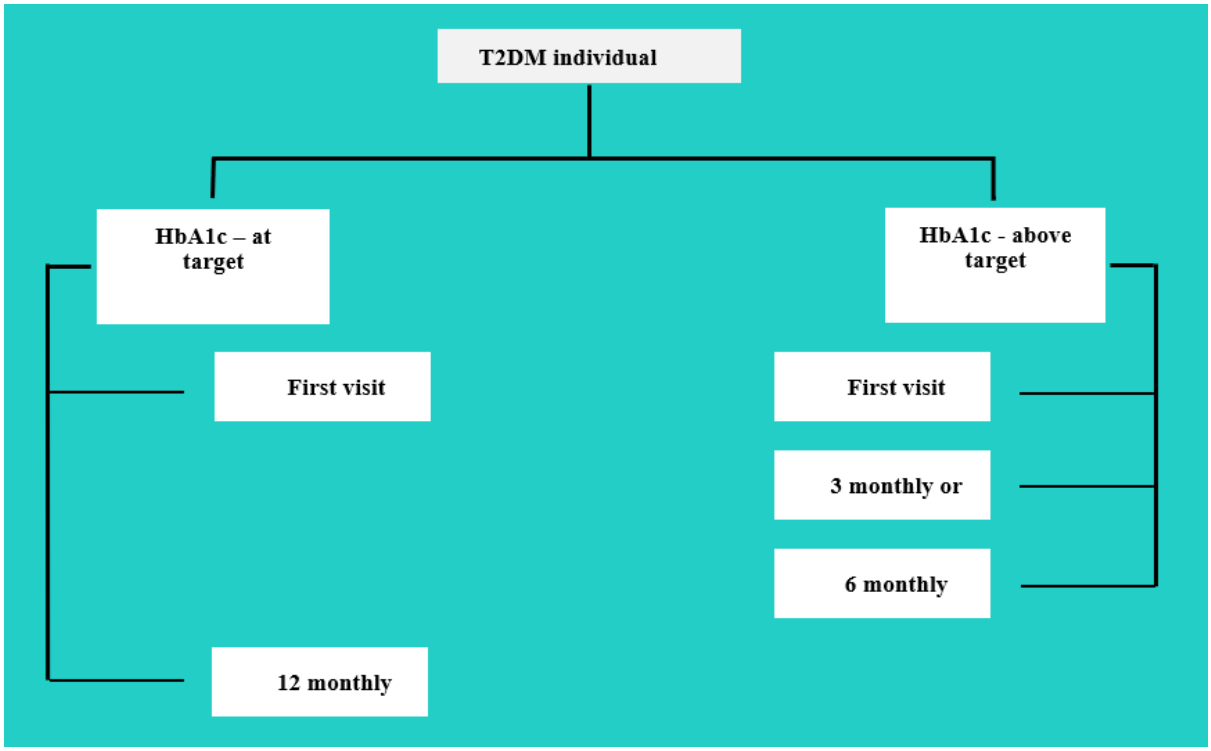

Figure 1.2: Monitoring HbA1c in T2DM individuals in South Africa. Adapted from South Africa's 2014 management of T2DM guidelines (South African National Department of Health, 2014).

1.6 Thesis Structure

This thesis comprises an introductory chapter, a theoretical background of approaches to prognosis prediction, a review of literature, three results chapters, a discussion chapter synthesising the findings and summarising the contributions of the overall thesis and supporting appendices. Table 1.1 summarises these results chapters, including objectives, clinical outcomes and corresponding biomarkers.

This introductory chapter places the application of dynamic prediction models, specifically to electronic health records data in resource-limited settings, including South Africa, into context and provides the aim and objectives.

Chapter 2 presents the theoretical background of common regression models for prognosis prediction.

In Chapter 3, the existing literature on dynamic prediction models is reviewed to identify key reported characteristics in the conduct and quality of model evaluation of dynamic prediction models for individualised prognosis prediction.

In Chapter 4, data sources for HIV and T2DM, as two high-burden chronic diseases in South Africa, are introduced to motivate the importance of prognosis prediction models introduced in Chapter 2.

Chapter 5 presents the application of Cox proportional-based prediction models for prognosis prediction of individuals with a chronic disease (from Chapter 4) in large routine data in EHRs from resource-limited settings. The predictive accuracy of these models is discussed, including the limitations.

In Chapter 6, building on the foundation of the literature review (Chapter 3), the application of dynamic prediction models for individualised prognosis prediction of individuals with a chronic disease in large routine data in EHRs from resource-limited settings is presented. Adopting dynamic prediction models is justified. These models are compared and contrasted with Cox proportional-based prediction models (Chapter 5) regarding predictive performance. The limitations and the broader implications of using dynamic prediction models for individualised prognosis prediction are discussed.

In Chapter 7, the dynamic prediction models presented in Chapter 6 are extended to allow different types of longitudinal continuous biomarker-outcome data. The adoption of the novel

dynamic prediction model is justified. This model is applied to large routine HIV and simulated data. The predictive performance of the novel and dynamic prediction models (introduced in Chapter 6) is compared using both routine and simulated data. The implications and limitations of the novel dynamic prediction model are discussed.

A discussion and summary of the thesis are presented in Chapter 8. An overview of key findings from the thesis is provided. The key findings are examined and interpreted to provide recommendations for research and practice. A concluding summary of the thesis is provided.

Table 1.1: Summary of thesis result chapters, including objectives, clinical outcomes and corresponding biomarkers.

Chapter	Objective	Disease	Biomarker
Chapter 5	Evaluate the predictive performance of Cox proportional hazard-based prediction models for prognosis prediction in routine data in EHRs from resource-limited settings.	HIV and T2DM	HIV viral load and HbA1c
Chapter 6	Compare and contrast Cox proportional hazard-based and dynamic prediction models for prognosis prediction in routine data from EHRs from resource-limited settings regarding predictive performance.	HIV and T2DM	HIV viral load and HbA1c
Chapter 7	Develop a novel statistical model for dynamic prediction to allow for different types of longitudinal continuous biomarker-outcome data.	HIV	HIV viral load
Chapter 7	Apply and compare the performance of dynamic prediction models with a novel dynamic prediction model in routine data from EHRs in resource-limited settings and simulated data.	HIV	HIV viral load

1.7 Significance for Public Health

Monitoring schedules for chronic disease conditions in resource-limited settings tend to be less frequent than in resource-rich settings, where many guidelines for chronic disease care, including monitoring, have been developed (Owolabi et al., 2018). Biomarker repeated measurements from clinical monitoring of individuals by periodic medical appointments provide an opportunity to detect progression to positive or negative health outcomes earlier than by cross-sectional biomarker evaluation. Prediction modelling approaches with good predictive performance are important when accounting for these biomarker monitoring

schedules in resource-limited settings. These approaches can potentially optimise and address individualised care. In resource-limited settings, including South Africa, where the burden of diseases such as HIV and T2DM is high, applying dynamic prediction models using routine data from electronic health records can provide quantitative evidence to identify virologic failure in individuals living with HIV on ART timeously and potentially improve management of glycaemic control in T2DM individuals.

1.8 References

Anderson, K. M., Odell, P. M., Wilson, P. W., & Kannel, W. B. (1991). Cardiovascular disease risk profiles. *American heart journal, 121*(1), 293-298.

Andrinopoulou, E., Harhay, M., Ratcliffe, S., & Rizopoulos, D. (2021). Reflection on modern methods: dynamic prediction using joint models of longitudinal and time-to-event data. *Int J Epidemiol, 50*(5), 1731-1743.

Andrinopoulou, E., Rizopoulos, D., Takkenberg, J., & Lesaffre, E. (2017). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Stat Methods Med Res, 26*(4), 1787-1801. doi:10.1177/0962280215588340

Aschner, P., Gagliardino, J. J., Ilkova, H., Lavalle, F., Ramachandran, A., Mbanya, J. C., . . . Chan, J. C. N. (2020). Persistent poor glycaemic control in individuals with type 2 diabetes in developing countries: 12 years of real-world evidence of the International Diabetes Management Practices Study (IDMPS). *Diabetologia, 63*(4), 711-721. doi:10.1007/s00125-019-05078-3

Beaglehole, R., Epping-Jordan, J., Patel, V., Chopra, M., Ebrahim, S., Kidd, M., & Haines, A. (2008). Improving the prevention and management of chronic disease in low-income and middle-income countries: a priority for primary health care. *The lancet, 372*(9642), 940-949. doi:https://doi.org/10.1016/S0140-6736(08)61404-X

Beck, R. W., Connor, C. G., Mullen, D. M., Wesley, D. M., & Bergenstal, R. M. (2017). The Fallacy of Average: How Using HbA1c Alone to Assess Glycemic Control Can Be Misleading. *Diabetes Care, 40*(8), 994-999. doi:10.2337/dc17-0636

Brennan, A. T., Lauren, E., Bor, J., George, J. A., Chetty, K., Mlisana, K., . . . Crowther, N. J. (2023). Gaps in the type 2 diabetes care cascade: a national perspective using South

Africa's National Health Laboratory Service (NHLS) database. *BMC Health Serv Res, 23*(1), 1452. doi:10.1186/s12913-023-10318-9

Chehregosha, H., Khamseh, M. E., Malek, M., Hosseinpanah, F., & Ismail-Beigi, F. (2019). A View Beyond HbA1c: Role of Continuous Glucose Monitoring. *Diabetes Therapy, 10*(3), 853-863. doi:10.1007/s13300-019-0619-1

Coetzee, A., Beukes, A., Dreyer, R., Solomon, S., van Wyk, L., Mistry, R., . . . van de Vyver, M. (2019). The prevalence and risk factors for diabetes mellitus in healthcare workers at Tygerberg hospital, Cape Town, South Africa: a retrospective study. *Journal of Endocrinology, Metabolism and Diabetes of South Africa, 24*(3), 77-82. doi:10.1080/16089677.2019.1620009

D'Agostino Sr, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation, 117*(6), 743-753.

Grand, M. K., Vermeer, K. A., Missotten, T., & Putter, H. (2019). A joint model for dynamic prediction in uveitis. *Stat Med, 38*(10), 1802-1816. doi:10.1002/sim.8069

Hans, L., Cassim, N., Sarang, S., Hardie, D., Ndlovu, S., Venter, W. D. F., . . . Stevens, W. (2023). HIV Viral Load Testing in the South African Public Health Setting in the Context of Evolving ART Guidelines and Advances in Technology, 2013-2022. *Diagnostics (Basel), 13*(17). doi:10.3390/diagnostics13172731

Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics, 1*(4), 465-480.

Hippisley-Cox, J., Coupland, C., Robson, J., & Brindle, P. (2010). Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. *bmj, 341*, c6624.

International Diabetes Federation. (2019). *IDF Diabetes Atlas* (9th ed.): International Diabetes Federation.

Jenkins, D. A., Sperrin, M., Martin, G. P., & Peek, N. (2018). Dynamic models to predict health outcomes: current status and methodological challenges. *Diagn Progn Res, 2*, 23. doi:10.1186/s41512-018-0045-2

Jewell, N. P., & Nielsen, J. P. (1993). A framework for consistent prediction rules based on markers. *Biometrika, 80*(1), 153-164. doi:10.1093/biomet/80.1.153

Keiser, O., Chi, B. H., Gsponer, T., Boulle, A., Orrell, C., Phiri, S., . . . Egger, M. (2011). Outcomes of antiretroviral treatment in programmes with and without routine viral load

monitoring in Southern Africa. *Aids, 25*(14), 1761-1769. doi:10.1097/QAD.0b013e328349822f

Kharsany, A. B., & Karim, Q. A. (2016). HIV infection and AIDS in sub-Saharan Africa: current status, challenges and opportunities. *The open AIDS journal, 10*, 34.

Manne-Goehler, J., Geldsetzer, P., Agoudavi, K., Andall-Brereton, G., Aryal, K. K., Bicaba, B. W., . . . Jaacks, L. M. (2019). Health system performance for people with diabetes in 28 low- and middle-income countries: A cross-sectional study of nationally representative surveys. *PLoS Med, 16*(3), e1002751. doi:10.1371/journal.pmed.1002751

Mayosi, B. M., Flisher, A. J., Lalloo, U. G., Sitas, F., Tollman, S. M., & Bradshaw, D. (2009). The burden of non-communicable diseases in South Africa. *The lancet, 374*(9693), 934-947.

Mesic, A., Spina, A., Mar, H. T., Thit, P., Decroo, T., Lenglet, A., . . . Oo, H. N. (2021). Predictors of virological failure among people living with HIV receiving first line antiretroviral treatment in Myanmar: retrospective cohort analysis. *AIDS Res Ther, 18*(1), 16. doi:10.1186/s12981-021-00336-0

Nicolaie, M. A., Van Houwelingen, J. C., de Witte, T. M., & Putter, H. (2013). Dynamic prediction by landmarking in competing risks. *Stat Med, 32*(12), 2031-2047. doi:10.1002/sim.5665

Owolabi, M. O., Yaria, J. O., Daivadanam, M., Makanjuola, A. I., Parker, G., Oldenburg, B., . . . Osundina, M. A. (2018). Gaps in guidelines for the management of diabetes in low- and middle-income versus high-income countries—a systematic review. *Diabetes Care, 41*(5), 1097-1105.

Pham, M. D., Nguyen, H. V., Anderson, D., Crowe, S., & Luchters, S. (2022). Viral load monitoring for people living with HIV in the era of test and treat: progress made and challenges ahead - a systematic review. *BMC Public Health, 22*(1), 1203. doi:10.1186/s12889-022-13504-2

Pheiffer, C., Pillay-van Wyk, V., Turawa, E., Levitt, N., Kengne, A. P., & Bradshaw, D. (2021). Prevalence of Type 2 Diabetes in South Africa: A Systematic Review and Meta-Analysis. *Int J Environ Res Public Health, 18*(11). doi:10.3390/ijerph18115868

Pillay-van Wyk, V., Msemburi, W., Laubscher, R., Dorrington, R. E., Groenewald, P., Glass, T., . . . Bradshaw, D. (2016). Mortality trends and differentials in South Africa from 1997 to 2012: second National Burden of Disease Study. *The Lancet Global Health, 4*(9), e642-e653. doi:https://doi.org/10.1016/S2214-109X(16)30113-9

Proust-Lima, C., & Taylor, J. M. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics, 10*(3), 535-549.

Putter, H., & Van Houwelingen, H. C. (2022). Landmarking 2.0: Bridging the gap between joint models and landmarking. *Stat Med, 41*(11), 1901-1917. doi:10.1002/sim.9336

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics, 67*(3), 819-829.

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*: Chapman and Hall/CRC.

Rizopoulos, D., Molenberghs, G., & Lesaffre, E. M. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal, 59*(6), 1261-1276.

Roberts, T., Cohn, J., Bonner, K., & Hargreaves, S. (2016). Scale-up of Routine Viral Load Testing in Resource-Poor Settings: Current and Future Implementation Challenges. *Clinical Infectious Diseases, 62*(8), 1043-1048. doi:10.1093/cid/ciw001

Rowley, C. F. (2014). Developments in CD4 and viral load monitoring in resource-limited settings. *Clin Infect Dis, 58*(3), 407-412. doi:10.1093/cid/cit733

Rueten-Budde, A. J., Van Praag, V. M., Van de Sande, M. A. J., & Fiocco, M. (2018). Dynamic prediction of overall survival for patients with high-grade extremity soft tissue sarcoma. *Surg Oncol, 27*(4), 695-701. doi:10.1016/j.suronc.2018.09.003

Saura-Lázaro, A., Bock, P., Bogaart, E. V. D., Van Vliet, J., Granés, L., Nel, K., . . . López-Varela, E. (2023). Field performance and cost-effectiveness of a point-of-care triage test for HIV virological failure in Southern Africa. *J Int AIDS Soc, 26*(10), e26176. doi:10.1002/jia2.26176

South African National Department of Health. (2014). Management of type 2 diabetes in adults at primary care level. Retrieved from https://knowledgehub.health.gov.za/elibrary/management-type-2-diabetes-adults-primary-care-level

South African National Department of Health. (2015). National consolidated guidelines for the prevention of mother-to-child transmission of HIV (PMTCT) and the management of HIV in children, adolescents and adults. In. Pretoria: South African National Department of Health.

South African National Department of Health. (2019a). 2019 ART Clinical Guidelines for the Management of HIV in Adults, Pregnancy, Adolescents, Children, Infants and Neonates. Retrieved from https://www.health.gov.za/hiv-and-aids/

South African National Department of Health. (2019b). Adult primary care, symptom-based integrated approach to the adult in primary care. Retrieved from https://knowledgehub.health.gov.za/elibrary/adult-primary-care-apc-guide-20192020-updated

South African National Department of Health. (2020). Standard treatment guidelines and essential medicines list for South Africa, primary healthcare level. Retrieved from https://knowledgehub.health.gov.za/content/standard-treatment-guidelines-and-essential-medicines-list

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. *Epidemiology, 21*(1). Retrieved from https://journals.lww.com/epidem/Fulltext/2010/01000/Assessing_the_Performance_of_Prediction_Models__A.22.aspx

Tsiatis, A. A., & Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 809-834.

UNAIDS. (2019). *South Africa's National Strategic Plan for HIV, TB and STIs 2017-2022*. Retrieved from https://www.unaids.org/en/regionscountries/countries/southafrica

United Nations Programme on HIV/AIDS [UNAIDS]. (2019). *Global AIDS Update 2019 [Fact Sheet]* Retrieved from https://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf

Van Houwelingen, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics, 34*(1), 70-85.

Van Houwelingen, H. C., & Putter, H. (2011). *Dynamic prediction in clinical survival analysis*: CRC Press.

Wang, Y., & Jeremy, M. G. T. (2001). Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome. *Journal of the American Statistical Association, 96*(455), 895-905. Retrieved from http://www.jstor.org/stable/2670229

WHO. (2011). Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus: abbreviated report of a WHO consultation. In. Geneva: World Health Organization.

WHO. (2016a). Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations–2016 update. Retrieved from https://www.who.int/publications/i/item/9789241511124

WHO. (2016b). *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach* (- 2nd ed.): World Health Organization.

WHO. (2016c). *Global report on diabetes*. Geneva: World Health Organization.

World Health Organization [WHO]. (2015). Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/186275/9789241509565_eng.pdf

Yu, M., Law, N. J., Taylor, J. M., & Sandler, H. M. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 835-862.

Yu, M., Taylor, J. M. G., & Sandler, H. M. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *Journal of the American Statistical Association, 103*(481), 178-187.

Chapter 2 Statistical Theory Review

2.1 Introduction

This chapter introduces the foundation of prediction models for clinical outcomes in cohort studies. The chapter briefly highlights some fundamental theoretical backgrounds of regression models used in this thesis to develop prognosis predictions of clinical outcomes. Since this thesis focuses on prognosis predictions of clinical outcomes using longitudinal continuous biomarkers, regression models for the analysis of longitudinal continuous biomarkers and time-to-event data are presented here. Section 2.2 introduces longitudinal biomarkers and time-to-event data, and common terminology used in the regression models to model these data. The formulation of separate modelling of longitudinal biomarker and time-to-event data is given in Sections 2.3 and 2.4, respectively. Section 2.5 introduces joint modelling of longitudinal biomarkers and time-to-event data. Section 2.6 presents prognosis prediction under regression models. The prediction performance of the prediction models is discussed in Section 2.7. Finally, a chapter conclusion is given in Section 2.8.

To outline the theoretical background of the regression models for analysing data in the motivating disease case studies (introduced in Chapter 4), throughout the thesis, n individuals with m_i $(i = 1, \cdots, n)$ repeated biomarker measures are considered. Let $N = \sum_{i=1}^{n} m_i$ denote the total number of observations, y_{ij} be an observed continuous biomarker value of individual i at observed follow-up time j $(j = 1, \cdots, m_i)$ and y_i denote a vector $(m_i \times 1)$ of a longitudinal continuous biomarker for the ith individual. Accordingly, y_{ij}^* is the true unobserved value of longitudinal continuous biomarker. For the ith individual, $T_i = \min(T_i^*, C_i)$ is defined as the observed event time, with true event time T_i^* and censoring time C_i assumed to be independent. An event indicator δ_i variable is given a value of 1 if the event is observed $(T_i^* \leq C_i)$ and 0 if censored.

2.2 Longitudinal Continuous Biomarker and Time-to-Event Data

In biomedical studies, there is often longitudinal follow-up of individuals. Depending on the study objectives and the availability of resources, these studies may take months to years to complete. During the follow-up, numerous biological and molecular composition changes can occur in individuals. These changes can be captured by biomarkers or blood values recorded

over time for an individual, for example, repeated measurements of plasma HIV viral load or HbA1c. Longitudinal biomarker outcomes arise when there are repeated biomarker measures on the same individuals over time. Repeated biomarker measurements are often collected along with a time-to-event of interest.

A time-to-event (or survival) outcome is the time until something happens to an individual, for example, the onset or relapse of a disease. Individuals are often lost to follow-up, withdraw from the study, or do not experience the event of interest during follow-up in a longitudinal study. In these cases, the event is unobserved and referred to as censored. Multiple types of censoring can occur depending on the relative timing of observation. However, in general, right censoring, where the event is presumed to happen sometime after the follow-up period stops, is of primary interest (Collett, 1994). The censoring can be classified as informative or non-informative. Informative censored results from individuals dropping out of a study are due to reasons related to the study. In contrast, non-informative censored results are due to reasons unrelated to the study (Ranganathan & Pramesh, 2012). Censored data that is not handled appropriately leads to biased estimates and can lead to misleading inferences about the distribution of the event times (Rizopoulos, Dimitris, 2012). Standard statistical tools, such as group comparisons and generalised linear models, do not account for censoring. Therefore, time-to-event analysis with non-informative right censoring was considered in this thesis. Throughout the thesis, names such as time-to-event and survival data, and event and survival times are used interchangeably.

The interest lies in using the variation in the observed longitudinal trajectories of a biomarker (HIV viral load or HbA1c) to provide improved prognosis predictions of a clinical outcome (virologic failure or glycaemic control). As mentioned in 0, one of the well-established approaches that provide improved prognosis predictions is a joint model for longitudinal and time-to-event data (Albert & Follmann, 2009; Diggle, Henderson, & Philipson, 2009; Faucett & Thomas, 1996; Henderson, Diggle, & Dobson, 2000; Tsiatis & Davidian, 2004; Wulfsohn & Tsiatis, 1997). Understanding well-established statistical regression models for separately analysing longitudinal and time-to-event data is needed to develop prognosis predictions under joint modelling. The most common approaches to analyse longitudinal data appearing in the literature to date are the mixed-effects models (Laird & Ware, 1982; Verbeke, Geert & Molenberghs, 2000) and generalised estimating equations (Zeger, Liang, & Albert, 1988). Likewise, time-to-event data can be analysed using Cox models (Collett, 1994; Cox, 1972; Cox

& Oakes, 1984). The mixed-effects and Cox models for longitudinal biomarker and time-to-event data are discussed in Sections 2.3 and 2.4, respectively.

2.3 Mixed-Effects Model for Longitudinal Continuous Biomarker Data

Mixed-effects models for longitudinal data are based on extensions of multiple regression models. Regression is a common term that includes various statistical models, including the most common models used in clinical research and prognosis prediction. Specifically, linear regression has been used in many research areas, including medicine, to describe the linear association between a continuous outcome and one or more covariates. Linear regression techniques do not simultaneously examine group and individual-specific changes over time (Verbeke, Geert 1997). Linear mixed-effects (LME) models have been developed to manage the disadvantages of linear regression models. These models are briefly discussed in this section.

2.3.1 Linear mixed-effects model specification

Repeated continuous biomarker measurements observed on the same individuals over time are generally assumed to be correlated. Therefore, the statistical model used for estimation in such a context must be able to manage non-independent observations. The mixed-effects model is one of the most common models used to analyse longitudinal continuous biomarker data. Mixed-effects models permit associations to vary by group or individual and can vary in intercept and slope. Group effects or interactions and adjusted associations can be estimated in the usual way with linear models. The critical feature of this model is that it enables prediction at group and individual levels while accounting for the non-independent nature of longitudinal continuous biomarker values taken on the same individuals.

A two-stage approach (Verbeke, Geert & Molenberghs, 2000) allows the specification of a mixed-effects model for repeated continuous biomarker measurements. The first stage of the approach uses a linear regression model to describe the individual-specific evolution of the repeated, continuous biomarker values over time. The second stage uses a multivariate regression model to describe individual differences. This allows an understanding of how variable the individual specific trajectory of repeated, continuous biomarker measurements is

and how the measurements are related. Individual-specific trajectories and variations between individuals are combined by the two-stage analysis and presented as a mixed-effects model. The change over time in the observed continuous biomarker noisy values can be modelled given covariates using a linear mixed-effects model, which is given as:

$$y_{ij} = X'_{Cij}\beta + Z'_{Cij}b_i + \epsilon_{ij} = y^*_{ij} + \epsilon_{ij}. \quad 2.1$$

where X'_{Cij} and Z'_{Cij} represents row-vectors of covariates, C represents the continuous index of a continuous biomarker with associated vectors of fixed effects coefficient β and random effects b_i, respectively.

The fixed effects coefficient describes the average association between the repeated, continuous biomarker values and the kth ($k = 1, \cdots, p$) covariate across all individuals. This coefficient can be split into a group average from Equation 2.1 as follows:

$$\beta = \beta_0 + \beta_k,$$

where the intercept β_0 represents the individual-specific average value of a continuous biomarker when all covariates are zero. Similarly, the slope β_k represents the individual-specific effect size of each covariate on the average value of a continuous biomarker.

The random effects characterise individual differences in β coefficient given as set l ($l = 1, \cdots, q$) of covariates. Informally, the random effects could be considered to capture variation due to factors not modelled directly, including genetic, environmental and biological factors. They can also be split into a deviation-based group average from Equation 2.1 as follows:

$$b_i = b_{i0} + b_{il},$$

where the random intercept b_{i0} and slope b_{il} represents individual-specific deviations from the intercept β_0 and slope β_k. The random effects are assumed to be unobserved, induce correlation

and vary randomly across individuals (Verbeke, Geert & Molenberghs, 2000). They are also assumed to be Gaussian distributed and often independent, with mean zero and variance-covariance matrix D, i.e., $b_i \sim N(0, D)$. The variance-covariance matrix D captures the correlation between observations from the individuals. The random or measurement errors ϵ_{ij} are assumed to be independent from the random effects and are also assumed to follow a Gaussian distribution, i.e., $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$.

2.3.2 Estimation of effects in a linear mixed-effects model

Given m_i repeated continuous biomarker values, y_i, and one or more covariates, a linear mixed-effects model framework (Laird & Ware, 1982) can be represented in vector form. Let X_i and Z_i be design matrices corresponding to row-vectors of covariates X'_{Cij} and Z'_{Cij}, respectively. The LME model in vector form can be specified as:

$$y_i = X_{Ci}\beta + Z_{Ci}b_i + \varepsilon_i = y_i^* + \varepsilon_i, \qquad b_i \sim N(0, D), \varepsilon_i \sim N\left(0, \sigma_\epsilon^2 I_{m_i}\right), \quad 2.2$$

where I_{m_i} represents an mth dimensional identity matrix. The $X_{Ci}\beta$ describes the average longitudinal evolution at the group level over the individuals and $Z_{Ci}b_i$ describes how specific individuals deviate from others. Conditional on the random effects b_i, all the repeated, continuous biomarker values y_i from each individual are assumed to be independent. Thus, integrating out the random effects and assuming y_i has a Gaussian distribution, the LME model reduces to a classical linear regression model on correlated data, i.e., $y_i \sim N(X_{Ci}\beta,\ V_i)$, where $V_i = Z_{Ci}DZ'_{Ci} + \sigma_\epsilon^2 I_{m_i}$ is a positive definite marginal covariance matrix. The likelihood function corresponding to the probability distribution function of the observed longitudinal continuous biomarker is given as:

$$L_i(\theta_C) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} p(y_{ij}),$$

where $\theta_C = (\beta', \sigma_\epsilon, D)$ denotes the parameter vector for the LME model for a continuous biomarker. The tractable solution of the log-likelihood of the model can be written as:

$$\log\big(L_i(\theta_C)\big) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\{n_i \log(2\pi) + \log|V_i| + (y_i - X_{Ci}\beta)' V_i^{-1}(y_i - X_{Ci}\beta)\}.$$

Maximum likelihood (ML) or restricted maximum likelihood (REML) is used to estimate the parameters in V_i (σ_ϵ, D). The main difference between the two likelihood approaches is that, in small samples, REML is unbiased compared to ML (Verbeke, Geert 1997). Given the known covariance matrix V_i the fixed effects parameter β can be estimated generalised least squares and given as:

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_{Ci}' V_i^{-1} X_{Ci}\right)^{-1} \sum_{i=1}^{n} X_{Ci}' V_i^{-1} y_i.$$

Variance components V_i enable valid inferences about the mean structure in a mixed model, $X_{Ci}\beta$ (Verbeke, Geert & Molenberghs, 2000). Commonly used classes of variance-covariance matrices include the independent structure, compound symmetry (also known as exchangeable), autoregressive, Toeplitz and unstructured variance-covariance matrix (Littell, Ramon. C. , Milliken, Stroup, Wolfinger, & Oliver, 2006; Littell, Ramon. C., Pendergast, & Natarajan, 2000; Mikkonen et al., 2008; West, Welch, & Galecki, 2014). Information criteria are often used as a basis for model selection and can be used to select the appropriate variance-covariance structure (Akaike, 1974; Schwarz, 1978). Assuming an LME model of the longitudinal continuous biomarker y_i for individuals has been fitted, estimates for the random effects b_i are based on the mean of a posterior distribution having a closed form expression:

$$b_i|y_i \sim N\{DZ_{Ci}' V_i^{-1}(y_i - X_{Ci}\beta), DZ_{Ci}' K Z_{Ci} D)\},$$

where $K = V_i^{-1} - V_i^{-1}X_{Ci}(\sum_{i=1}^{n} X'_{Ci}V_i^{-1}X_{Ci})^{-1}X'_{Ci}V_i^{-1}$. The means of these posterior distributions provide estimates of the random effects, and their tractable solution can be written as:

$$E(b_i|y_i) = DZ'_{Ci}V_i^{-1}(y_i - X_{Ci}\beta).$$

These random effects estimates are known as empirical Bayes estimates, providing individual-specific predictions (Rizopoulos, Dimitris, 2012; Verbeke, Geert & Molenberghs, 2009).

2.3.3 Extension of linear mixed-effects models

The linear mixed-effects model introduced so far describes Gaussian continuous longitudinal outcomes. In practice, often, not all outcomes are continuous and Gaussian distributed. Inferences for these different types of longitudinal outcomes are needed. Generalised linear mixed models (GLMMs) provide a generalisation of the LME to permit longitudinal outcomes other than continuous. Keeping the same notation in Equation 2.2 but omitting the continuous index, the generalised linear mixed model can be specified as:

$$g(y_{ij}) = X'_{ij}\beta + Z'_{ij}b_i + \epsilon_{ij}. \qquad 2.3$$

The corresponding likelihood functions have the form:

$$L_i = \prod_{i=1}^{n}\prod_{j=1}^{m_i}\int p(y_{ij}|b_i)p(b_i)\mathrm{d}b_i,$$

where $p(\cdot)$ denotes probability density functions. Principles used in mixed models for continuous outcomes are adopted to estimate GLMMs. Unlike in the linear mixed-effects model, the integral over the random effects b_i in the log-likelihood expression for GLMMs

does not have a closed-form solution due to the non-linear nature of the link function (Liu, 2015). Numerical integration techniques such as penalised quasi-likelihood (Breslow & Clayton, 1993), Laplace approximation (Wolfinger, 1993), Markov chain Monte Carlo (MCMC) methods (Schafer, 1997), and Gaussian quadrature rules (Pinheiro & Bates, 1995) can be used for the numerical approximation of the random effects.

2.4 Cox Models for Time-to-Event Data

In biomedical studies, clinical outcomes or events are normally pre-specified, and the time it takes to occur is often of interest. Information about the event and how long it took for it to happen is collected. The motivating disease cases in this thesis assumed that individuals were enrolled and followed forward in time (days, months, or years). During follow-up time, one of two things could happen to the individuals by the end of the study: they could have the event of interest or could not have the event (censoring) for various reasons. The resulting information on the individuals are components of time-to-event data: i) the presence or absence of the event and ii) the time it took to get there.

The presence or absence of an event is often modelled using a logistic regression model. Logistic regression models are a class of generalised linear models (GLMs). Unlike linear regression models, their interest lies in describing the association between a binary outcome (presence or absence of event) and one or more covariates (Hosmer & Lemeshow, 2000; Peng, Lee, & Ingersoll, 2002). For example, in HIV, the binary outcome is whether an individual initiated on ART treatment had virologic failure or not. Key to prognosis prediction of most clinical outcomes in chronic diseases is predicting when the event of interest might occur. Both linear and logistic regression techniques provide a foundation for prediction. However, both approaches are limited to cross-sectional data where the outcome is already known. Although other GLMs can estimate risk or predict an outcome, they do not manage missing or censored data, nor can they manage dependent data, for example, longitudinal continuous biomarker data. A set of statistical regression models to analyse these types of data broadly termed time-to-event analysis are briefly discussed in this section.

2.4.1 Descriptive tools for time-to-event data

In time-to-event analysis, the interest is on estimating the hazard, i.e., the rate or timing of an event of interest for an individual and for populations or subgroups and estimating the average survival time from a known point in disease progression, such as diagnosis. Time-to-event analysis approaches are necessary when considering censored data to ensure unbiased estimates of the hazard of the event (Collett, 1994). These statistical analysis approaches are useful in predicting when the event of interest might occur. Terminology and notations introduced earlier in this chapter (Section 2.1) are extended to introduce components of time-to-event analysis; after that, look at the idea behind this type of analysis through basic descriptive tools, building to regression models for time-to-event data.

One of the critical steps in time-to-event analysis is to display the distribution of observed event time T while accounting for censoring in the data before the modelling takes place. An estimate of the survival or hazard function is needed, and ordinally descriptive analyses or plots are used to describe the data using these estimates. The hazard function $h(t)$ estimates the sudden (instantaneous) rate of occurrence of an event at a particular time point t, given that the event has not occurred before this point (Collett, 1994). The hazard function is given as:

$$h(t) = \lim_{s \to 0} \frac{p(t \leq T < t + s | T \geq t)}{s}, s > 0, \tag{2.4}$$

where $p(t \leq T < t + s | T \geq t)$ is the probability of having an event of interest in each small interval $[t, t + s]$ with width s, given that the event has not occurred before this time t.

Following the hazard function, a cumulative risk (cumulative hazard) function, $H(t)$, which describes the expected number of events to be observed by time, t, is obtained. The cumulative hazard function can be expressed as:

$$H(t) = \int_0^t h(s)ds.$$

The hazard or cumulative hazard function can be used to estimate a survival probability characterised by the survival function $S(t)$. The survival function describes the probability of surviving (not experiencing an event) up to a time point, t. It is a non-decreasing function of time because, as time goes by, the survival probability decreases because individuals experience the event. Survival can be expressed as:

$$S(t) = \exp\{-H(t)\} = \exp\left\{\int_0^t h(s)\, ds\right\}. \qquad 2.5$$

Based on whether there was the presence or absence of the event and the corresponding time it took to get there, various quantities of the survival function, $S(t)$, can be estimated.

Life tables are the origin of time-to-event analysis and were introduced to estimate the survival function (Cutler & Ederer, 1958). Life tables involve strong assumptions, are tedious and cannot deal with covariate adjustments. The Kaplan-Meier (KM) estimator (Kaplan & Meier, 1958) extended the idea behind the life table method to estimate the survival function. Based on observed time-to-event times for individual i, T_i, the KM estimates the probabilities at each unique time point, t, while accounting for censoring by incorporating the number of individuals at risk. The KM estimator is commonly used to estimate the survival function and it is given by:

$$\hat{S}_{KM}(t) = \prod_{i:t_i \leq t} \frac{r_i - d_i}{r_i},$$

where r_i denotes the number of individuals who have not experienced the event at the time t_i, and d_i is the number of events at time t for individual i. Using a similar approach to the KM, the Nelson-Aalen estimator can be used to estimate the expected number of events by time point t, i.e., the cumulative hazard function $H(t)$ (Aalen, 1976; Altshuler, 1970; Fleming & Harrington, 1991; Nelson, 1972). The Nelson-Aalen estimator is given as:

$$\widehat{H}_{NA}(t) = \sum_{i:t_i \leq t} \frac{d_i}{r_i}.$$

Based on this Nelson-Aalen estimator, the Breslow estimator (Fleming & Harrington, 1984) can also be used to estimate the survival functions. Nelson-Aalen and Kaplan-Meier estimate the survival function without distributional assumptions on the survival times (Andersen & Gill, 1982). There is minimal difference between these two survival function estimators, and the Kaplan-Meier estimator is often used. Following this, the KM plot can be used to compare the survival between groups of individuals in conjunction with the log-rank test to assess the statistical significance between the groups (Cox, 1972).

In many biomedical studies, there might be more research questions other than comparing survival functions or curves of groups for individuals, for example, the effect of treatment on survival controlling for other covariates. Regression models for time-to-event data are needed to manage such research questions.

2.4.2 Regression models for time-to-event data specification

Different types of statistical regression models have been proposed in the literature to see if the time-to-event is related to covariates in the analysis, more specifically, regression models for modelling the hazard of an event. The most common approach to modelling the hazard of an event appearing in the literature is the Cox proportional hazards model (Cox, 1972). The Cox proportional hazard (relative risk) model is a semiparametric regression model for time-to-event data. It is considered semiparametric because it does not make assumptions about the outcome distribution (there is no assumption about shape or form of the hazard function) but does make parametric assumptions about the relationships between covariates and the hazard function. Examples of distributions often used for the event times are log-normal, log-logistic, Gamma, Weibull and exponential. The Cox proportional hazards model to model the hazard of an event (Equation 2.4) at time t, $h_i(t)$ for an ith individual can be specified as:

$$h_i(t) = h_0(t) \exp(w_k' \gamma), \qquad 2.6$$

where w_k' $(k = 1, \cdots, l)$ represent a row vector of covariates with an associated vector of regression coefficients, γ. The baseline hazard or risk function, $h_0(t)$, denotes the instantaneous risk of the event at time t when all covariates have no effect ($\gamma = 0$). This means that, if a covariate has a beneficial effect, the baseline hazard would decrease ($\gamma < 0$). Likewise, the baseline hazard would increase if a covariate had a harmful effect ($\gamma > 0$). The regression coefficients for the covariates are exponentiated to represent the change in the hazard of an event. In general, the comparison of individual d's hazard, $h_d(t)$ at time t and individual e's hazard, $h_e(t)$, at time t can be given as a ratio of hazards as

$$\frac{h_d(t|w_d')}{h_e(t|w_e')} = \exp\{\gamma(w_d' - w_e')\}.$$

where w_d' and w_e' are row vectors of covariates for individuals d and e, respectively. The multiplicative effect on the hazard for an event and the time-independent nature of covariates (baseline covariates) make up the proportional hazard (PH) assumption of the Cox proportion hazards model. The vital characteristic of the PH assumption is that the relative hazard of an event between different strata is constant over time (Cox, 1972).

2.4.3 Estimation of effects in Cox models for time-to-event data

The non-parametric part of the Cox proportional hazards model is one in which the baseline hazard function, $h_0(t)$, is not specified. The estimation is primarily on parameters γ based on maximising log-partial likelihood function and is given by:

$$p\ell(\beta) = \sum_{i=1}^{n} \delta_i \left[w_i'\gamma - \log\left\{ \sum\nolimits_{T_j \geq T_i} \exp\left(w_j'\gamma\right) \right\} \right], \qquad 2.7$$

where w_i' is a row vector of covariates associated with the individual whose observed event time is T_i. Resulting consistent estimates of γ are usually denoted as $\hat{\gamma}$ and are assumed to be asymptotically normal, expressed as

$$\hat{\gamma} \sim N(\gamma_0, [E\{\mathcal{I}(\gamma_0)\}]^{-1}),$$

where γ_0 and $E\{\mathcal{I}(\gamma_0)\}$ are the true values of and expected information corresponding to the partial log-likelihood (Equation 2.7).

2.4.4 Extended Cox model for time-to-event data

The covariates in the Cox proportional hazards model (Equation 2.6) were assumed to be fixed during follow-up time, i.e., time independent. In practice, some covariates change during follow-up (biomarkers). The time-changing covariates may be exogenous or endogenous. Exogenous covariates are variables whose future values at any time point, t, are not influenced by the occurrence of an event at a preceding time, s (Kalbfleisch & Prentice, 2002). Endogenous covariates are the opposite of exogenous time-varying covariates. Most continuous biomarkers are endogenous because the occurrence of an opportunistic infection may influence the subsequent biomarker values. Therefore, the nature of these time-changing covariates dictates the appropriate analysis. If the interest is on how strongly related the time-changing covariates are with an event of interest, the standard Cox model is no longer informative (Rizopoulos, Dimitris, 2012). A model that can connect the observed time-to-event with the time-varying covariates is postulated to remedy this. A time-varying or extended Cox model handles exogenous time-varying covariates (Andersen & Gill, 1982). Given observed longitudinal continuous values, y_i, for the individuals up to follow-up time, t, the hazard of an event can be specified as:

$$h_i(t \mid y_i(t),\, w_k) = h_0(t) R_i(t) \exp\{w_k'\gamma + \varphi y_i(t)\}, \qquad 2.8$$

where γ and φ are the regression coefficients and respectively represent the change in log hazard for the baseline covariates, w_k, and time-varying covariate, $y_i(t)$, for the i^{th} individual. Moreover, $R_i(t)$ indicates the risk process of an individual. More importantly, φ is the association parameter that quantifies the strength of association between the observed value of a continuous at time t and the hazard $h_i(t)$ of an event at that same time for an individual. Exponentiating φ, i.e., $\exp(\varphi)$, gives the relative increase in the hazard of an event given a unit increase in the continuous value.

Extended Cox model and Joint modelling for time-dependent covariates

The extended Cox model incorporates longitudinal biomarkers using a last-measurement-carried-forward approach, assuming biomarker values remain constant between follow-up times and only change when measurements are taken at follow-up times (Rizopoulos, Dimitris, 2012). This step-function assumption is inappropriate for biomarkers such as HIV viral load or HbA1c, which change continuously over time and are endogenous in nature (Papageorgiou, Mauff, Tomer, & Rizopoulos, 2019; Rizopoulos, Dimitris, 2012; Rizopoulos, D. & Takkenberg, 2014). Moreover, the model treats biomarkers as exogenous and does not account for measurement error, resulting in biased estimates (Campbell et al., 2019; Dupuy & Mesbah, 2002; Prentice, 1982; Sweeting & Thompson, 2011). Due to these limitations, survival predictions often rely on baseline biomarker values using Cox proportional hazard models (Rizopoulos, Dimitris, Molenberghs, & Lesaffre, 2017). In contrast, joint models account for the measurement error and endogenous nature of biomarkers and link their true trajectories to event hazard over time, making them more appealing for survival predictions (Rizopoulos, Dimitris, 2012). Given this, and consistent evidence that they outperform extended Cox models (Arisido, Antolini, Bernasconi, Valsecchi, & Rebora, 2019; Campbell et al., 2019; Rizopoulos, D. & Takkenberg, 2014), this thesis focused on a joint modelling approach. A brief overview of joint modelling is provided in Section 2.5.

2.5 Joint Models for Longitudinal Continuous Biomarker and Time-to-Event Data

The focus in Sections 2.3 and 2.4 has been on separate analyses of longitudinal continuous biomarkers and time-to-event data. This section focuses on simultaneous analysis of longitudinal continuous biomarkers and time-to-event data using joint models.

2.5.1 Joint model specification

The joint model for longitudinal continuous and time-to-event data constructs a joint density, $p(y_i, T_i)$, between the longitudinal continuous biomarker, y_i, and the observed time-to-event, T_i (Diggle et al., 2009; Faucett & Thomas, 1996; Henderson et al., 2000; Tsiatis & Davidian, 2004; Wulfsohn & Tsiatis, 1997). A shared parameter joint model is commonly used in the

literature for constructing the distribution of the association between the underlying trajectories of a longitudinal continuous biomarker and the hazard of an event captured by random effects (Rizopoulos, Dimitris, 2011; Tsiatis & Davidian, 2004). Accordingly, a shared random effects model (SREM) was adopted in this thesis.

If y_i and T_i are vectors of a longitudinal continuous biomarker and an observed time-to-event process for the ith individual, then the specification of a joint distribution between the two processes is given by:

$$p(T_i, y_i, \delta_i | b_i; \theta) = p(T_i, \delta_i | b_i, \theta) p(y_i | b_i; \theta), \quad 2.9$$

where $\theta = \{\theta_T, \theta_c, \theta_b\}$ represents the overall parameter vector with θ_T representing the time-to-event process, θ_C representing the longitudinal continuous biomarker process and θ_b representing parameters of the random effects variance-covariance matrix. In Equation 2.9, b_i is a vector of random effects that accounts for the correlation between the repeated continuous biomarker measurements and the dependency between the longitudinal continuous biomarker and time-to-event processes (Rizopoulos, Dimitris, 2012).

The idea of a joint model is to use a linear mixed-effects model (Equation 2.2), commonly referred to as the longitudinal submodel, to describe underlying longitudinal trajectories of continuous biomarker values for each individual and use the estimated profiles of the individuals in a proportional hazards model, i.e., event submodel. Specifically, the proportional hazards model to model the hazard of an event at time t, $h_i(t)$ for an ith individual can be specified as:

$$h_i(t) = h_0(t) \exp\{w_k' \gamma + \eta_i(\cdot)' \varphi\}, \quad 2.10$$

where terms $h_0(t)$, w_k' and γ have specifications as in Equation 2.6.

A distribution such as exponential, Gamma and Weibull is specified for the baseline hazard to avoid underestimated standard errors of regression parameters (Hsieh, Tseng, & Wang, 2006). In the formulation, φ is an association parameter that quantifies the association between the

underlying trajectories of a longitudinal continuous biomarker and the hazard of an event. Accordingly, $\eta_i(\cdot)$ is a link function or association structure of the underlying trajectories of a longitudinal continuous biomarker and the hazard of an event. A variety of association structures in joint modelling literature have been suggested.

Association structures

An essential step after describing the underlying trajectory of a longitudinal continuous biomarker is selecting an appropriate association structure, $\eta_i(\cdot)$, for the trajectory and the time-to-event outcome (Brown, 2009; Rizopoulos, Dimitris, 2012). The parameterisation of this association structure can take many forms. The simplest functional form is given as:

$$\eta_i(\cdot) = y_i^*(t) = X_{Ci}(t)\beta + Z_{Ci}(t)b_i, \qquad 2.11$$

and it is commonly referred to as the current value or level association structure. This structure assumes that the current value of the longitudinal continuous biomarker at time t, is associated with the hazard $h_i(t)$ for a time-to-event at that same time t. In the joint model with this structure, an estimate of the underlying true biomarker value, which is continuously updated over time is used instead of the observed value of the biomarker (Arisido et al., 2019; Rizopoulos, Dimitris, 2012). The exponent φ in the extended Cox model denotes the relative increase in the hazard of an event at any time t that results from one unit increase in the observed value of the biomarker at the same time point. In contrast, in the joint model the exponent φ denotes the relative increase in the hazard of an event that results from one unit increase in the underlying (function) true biomarker value. This current value association structure is considered in Chapters 6 and 7.

Another common specification of the functional form of $\eta_i(\cdot)$ that is often used is the slope association structure given as:

$$\eta_i(\cdot) = y_i^{*'}(t). \quad 2.12$$

On this association structure, it is assumed that the rate of change of a longitudinal continuous biomarker is associated with the hazard $h_i(t)$ for a time-to-event.

A rarely used association structure is a lagged effect parameterisation. This association structure will be considered in Chapter 6. Under this parameterisation of the joint model, it is assumed that the value of the longitudinal continuous biomarker at a previous time $t - c$, where c is a time lag, is associated with the hazard $h_i(t)$ for a time-to-event at time t (Rizopoulos, Dimitris, 2012). The lagged effects association structure can be specified as:

$$\eta_i(\cdot) = y_i^*\{\max(t - c, 0)\}. \quad 2.13$$

Another association structure is the cumulative effects. This association structure will also be considered in Chapter 6. Under this parameterisation of the joint model, the cumulative effect of the longitudinal continuous biomarker is associated with the hazard $h_i(t)$ for a time-to-event at time t (Sylvestre & Abrahamowicz, 2009). The cumulative effects association structure can be specified as:

$$\eta_i(\cdot) = \int_0^t y_i^*(s)ds. \quad 2.14$$

Additional association structures are available and discussed in detail in the literature (Rizopoulos, Dimitris, 2012; Ye, Lin, & Taylor, 2008).

2.5.2 Estimation of effects in joint models

Under the assumption that the longitudinal continuous biomarker and time-to-event processes are conditionally independent through the random effects b_i, the log-likelihood contribution for the ith individual is given by:

$$\log p(T_i, y_i, \delta_i; \theta) = \log \int p(T_i, y_i, \delta_i, b_i; \theta) db_i, \qquad 2.15$$
$$= \log \int p(T_i, \delta_i | b_i; \theta_T) \left[\prod_j p\{y_{ij} | b_i; \theta_Y\} \right] p(b_i; \theta_b) db_i,$$

where the conditional density for the time-to-event process is given by:

$$p(T_i, \delta_i | b_i; \theta_T) \qquad 2.16$$
$$= [h_0(T_i) \exp(w_k' \gamma + \varphi y_i^*(T_i))]^{\delta_i} \times \exp\left[- \int_0^{T_i} h_0(s) \exp(w_k' \gamma + \varphi y_i^*(s)) \, ds \right].$$

The joint density of the longitudinal continuous biomarker and the random effects is given by:

$$p(y_i | b_i; \theta) p(b_i; \theta) = \prod_j p\{y_{ij} | b_i; \theta_Y\} p(b_i; \theta_b), \qquad 2.17$$

where the longitudinal continuous biomarker conditional density is given by:

$$p(y_{ij} | b_i; \theta_Y) = (2\pi\sigma_\epsilon^2)^{-1/2} \exp\left[- \frac{(y_{ij} - y_{ij}^*)^2}{2\sigma_\epsilon^2} \right], \qquad 2.18$$

and the density of the random effects is given by:

$$p(b_i; \theta_b) = (2\pi)^{-q_b/2} |D|^{-1/2} \exp\left[- \frac{b_i' D^{-1} b_i}{2} \right], \qquad 2.19$$

where q_b represents the random effects dimension.

Maximising the log-likelihood $\log p(T_i, y_i, \delta_i; \theta)$ function concerning the random effects is computationally demanding and does not have a closed-form solution. Thus, numerical integration techniques are used to fit joint models (Rizopoulos, Dimitris, 2012). Following the estimation of model parameters of a joint model, the survival function introduced in Equation 2.5 is extended to get estimates of survival probabilities of clinical outcomes. The following Section briefly introduces and describes the rationale for using joint modelling to predict survival probabilities and clinical outcomes.

2.6 Dynamic Prediction Under Joint Models

Given observed data from individuals, the interest is on what would happen to them from the start to end of the observed follow-up time. Based on the first measurements, it is interesting to know the chances that individuals experience an event of interest by the next follow-up time. Likewise, the predictions should be updated for subsequent measurements as new information becomes available. In general, it should be possible to do the same for any duration of follow-up of these individuals. As new information from individuals is obtained during follow-up time, it is desirable to continuously update individual predictions. These individualised predictions are dynamic. The dynamic aspect of the predictions comes from the fact that they can be seen as predictions that change over time (Rizopoulos, Dimitris, 2011, 2012; Van Houwelingen & Putter, 2011).

Predictions under the Cox and extended Cox models use a fixed or the last observed biomarker value as a baseline covariate for every individual on each prediction. As a result, predictions based on standard models such as the Cox proportional hazards model are static or periodic, while predictions based on joint modelling are dynamic, i.e., progressively update over time with new available information (Rizopoulos, Dimitris, 2012; Rizopoulos, Dimitris et al., 2017).

Let $\mathcal{y}_j$ be all the available observed longitudinal continuous biomarker measurements for individual, j, up to observed follow-up time, t. The interest is on the conditional event-free probability, $\pi_j(u|t)$, where u is a future time window. Accordingly, given all the observed measurements from an individual up to time t and that an event had not occurred, the

probability that this individual will be event-free at $u > t$ can be estimated (Garre, Zwinderman, Geskus, & Sijpkens, 2008; Rizopoulos, Dimitris, 2011, 2012). The event-free probability can be specified as:

$$\pi_j(u|t) = p\big(T_j^* \geq u | T_j^* > t, \mathcal{y}_j(t), \mathcal{D}_n\big), \qquad 2.20$$

where $\mathcal{D}_n$ represents the observed data on which the joint model (Equation 2.10) was fitted. Event-free conditional probabilities are calculated by relating a new individual for whom the predictions are required with individuals with similar characteristics from the observed data used to fit the joint model. Intuitively, the model explains what led to the information observed for the individuals. Predictions for the new individuals would be obtained based on this initial information.

Developments in the literature on dynamic predictions relied on the Bayesian framework to get the estimated event-free conditional probabilities for an individual, $\hat{\pi}_j(u|t)$ (Andrinopoulou, Harhay, Ratcliffe, & Rizopoulos, 2021; Li & Luo, 2019). Moreover, the Monte Carlo simulation scheme is used (Proust-Lima & Taylor, 2009; Rizopoulos, Dimitris, 2011), and the corresponding posterior predictive distribution of $\pi_j(u|t)$ can be defined as:

$$\pi_j(u|t) = \int P\big(T_j^* \geq u | T_j^* > t, \mathcal{y}_j(t); \theta\big) p(\theta|\mathcal{D}_n) d\theta,$$

where θ is the overall parameter vector as defined in Equation 2.9. Accordingly, for $m = 1,2,\cdots,M$ MCMC samples, the event-free probability and its 95% credible interval (CI) can be estimated using a Monte Carlo simulation scheme with the following steps:

Step 1: draw $\theta^{(m)}$ from the MCMC sample of the posterior $p(\theta|\mathcal{D}_n)$,

Step 2: draw $b_j^{(m)}$ from $P\big(b_j | T_j^* > t, \mathcal{y}_j(t); \theta^{(m)}\big)$,

Step 3: compute $\pi_j\Big(t, u, b_j^{(m)}; \theta^{(m)}\Big) = \frac{S\left(u|b_j^{(m)};\theta^{(m)}\right)}{S\left(t|b_j^{(m)};\theta^{(m)}\right)}$,

Step 4: repeat steps 1-3 and derive the estimates of the event-free conditional probability as

$$\hat{\pi}_j(u|t) = \frac{1}{M}\sum_{m=1}^{M} \pi_j^m\left(t, u, b_j^{(m)}; \theta^{(m)}\right).$$

Step 5: Use the Monte Carlo sample percentiles to obtain the 95% CI.

2.7 Predictive Performance of the Prediction Models

The predictive performance of these prediction models should be evaluated for the prognosis of clinical outcomes, for example, virologic failure and glycaemic control, to be reliable. Predictive performance measures are used to assess the predictive performance of prediction models. Most popular measures are based on discrimination, calibration, and overall performance (Steyerberg et al., 2010).

2.7.1 Assessing predictive performance by discrimination

In general, common discrimination metrics such as the area under the curve (AUC) of the receiver operating curve (ROC) evaluate how well a prediction model discriminates between individuals who will have an outcome and those who will not (Steyerberg et al., 2010). Time-dependent AUCs based on inverse probability of censoring weights (IPCW) enable discrimination in time-to-event data (Heagerty & Zheng, 2005). In joint modelling, model-based weights estimate time-dependent AUCs (Rizopoulos, Dimitris, 2011). Mathematically, the extended time-dependent AUCs under joint modelling are given as:

$$\mathrm{AUC}_t^{\Delta t} = \int_0^1 \mathrm{ROC}_t^{\Delta t}(p)dp, 0 \leq p \leq 1.$$

where p is a particular threshold value $\in [0, 1]$ and $\mathrm{ROC}_t^{\Delta t}(p)$ is the corresponding ROC curve. The ROC curve over time window Δt can be plotted to indicate the overall discrimination ability of the predictive joint model. Similarly, the corresponding AUCs can be calculated, where AUC values closer to 1 suggest better discrimination ability.

2.7.2 Assessing predictive performance by calibration scores

Traditionally, calibration measures such as a plot of observed versus estimated probabilities of an event are used to assess how well a prediction model can accurately predict observed events (Schoop, Graf, & Schumacher, 2008). In time-to-event analysis, calibration refers to agreement between observed and estimated event probabilities over specific time points (Austin, Harrell Jr, & van Klaveren, 2020). This recent development in the literature for assessing the calibration of prediction models for time-to-event data used variations of the Cox model to construct smoothed calibration curves (Austin et al., 2020).

2.7.3 Assessing the overall predictive performance

The discrepancy between an observed event denoted by an event indicator, δ_i, and prediction model-based event-free probability, $\hat{\pi}_i$, for an ith individual at a given time point, t, evaluated on independent or validation data, is used to quantify overall model performance (Steyerberg et al., 2010). Traditionally, on each time point, t, the Brier score (BS) quadratic scoring rule can be calculated to quantify the discrepancy δ_i and $\hat{\pi}_i$ for individual i (Graf, Schmoor, Sauerbrei, & Schumacher, 1999). Gerds and Schumacher (2006) used inverse probability of censoring weights to derive an expected Brier score to account for censoring. The expected Brier score is given as:

$$\widehat{BS}(t, \hat{\pi}_i) = \frac{1}{M} \sum_{i \in \widetilde{D}_M} \widehat{W}_i(t)\{\delta_i(t) - \hat{\pi}_i(t)\}^2,$$

where $\widetilde{D}_M$ is a validation dataset with M individuals and $\widehat{W}_i(t)$ is the inverse probability of censoring weights at time t. As a result, time-dependent prediction error (PE) curves given by the integrated Brier score (IBS) can be used to summarise the overall measure of prediction error (Mogensen, Ishwaran, & Gerds, 2012). Mathematically, the integrated Brier score is defined as:

$$\text{IBS}(\text{predErr}, \tau) = \frac{1}{\tau} \int_0^{\tau} \text{predErr}(u, \hat{\pi}) du, \ \tau > 0.$$

where predErr is the prediction error and $\tau > 0$ is a constant that can be set to be any value between 0 and a maximum value of time t.

2.7.4 Validation of prognosis prediction models

Internal and external validation approaches are often used to validate the predictive performance of a prediction model. Internal validation often uses re-sampling techniques such as cross-validation and Bootstrap to validate the performance of a prediction model (Harrell, 2001). On the other hand, external validation uses a dataset from an independent cohort to evaluate the performance of a prediction model.

2.8 Discussion

This chapter provided an overview of the theoretical background of regression models for analysing longitudinal continuous biomarker and time-to-event data and their use as practical tools to generate prognosis predictions of clinical outcomes. As mentioned in Section 2.6, prognosis predictions of clinical outcomes are often based on the generalisation of a Cox model and joint modelling. These two modelling approaches are applied to routine HIV and diabetes data from the Western Cape in Chapter 5 and Chapter 6. A review of the literature for the application of dynamic prediction models is given in Chapter 3.

2.9 References

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian Journal of Statistics*, 15-27.

Akaike, H. (1974). A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike* (pp. 215-222): Springer.

Albert, P. S., & Follmann, D. A. (2009). Shared-parameter models. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal Data Analysis* (pp. 433-452). Boca Raton, FL: Chapman &

Hall.

Altshuler, B. (1970). Theory for the measurement of competing risks in animal experiments. *Mathematical Biosciences, 6*, 1-11.

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics*, 1100-1120.

Andrinopoulou, E., Harhay, M., Ratcliffe, S., & Rizopoulos, D. (2021). Reflection on modern methods: dynamic prediction using joint models of longitudinal and time-to-event data. *Int J Epidemiol, 50*(5), 1731-1743.

Arisido, M. W., Antolini, L., Bernasconi, D. P., Valsecchi, M. G., & Rebora, P. (2019). Joint model robustness compared with the time-varying covariate Cox model to evaluate the association between a longitudinal marker and a time-to-event endpoint. *BMC Med Res Methodol, 19*(1), 222. doi:10.1186/s12874-019-0873-y

Austin, P. C., Harrell Jr, F. E., & van Klaveren, D. (2020). Graphical calibration curves and the integrated calibration index (ICI) for survival models. *Statistics in medicine, 39*(21), 2714-2742. doi:https://doi.org/10.1002/sim.8570

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*(421), 9-25.

Brown, E. R. (2009). Assessing the association between trends in a biomarker and risk of event with an application in pediatric hiv/aids. *Ann Appl Stat, 3*(3), 1163-1182. doi:10.1214/09-aoas251

Campbell, K. R., Juarez-Colunga, E., Grunwald, G. K., Cooper, J., Davis, S., & Gralla, J. (2019). Comparison of a time-varying covariate model and a joint model of time-to-event outcomes in the presence of measurement error and interval censoring: application to kidney transplantation. *BMC Med Res Methodol, 19*(1), 130. doi:10.1186/s12874-019-0773-1

Collett, D. (1994). *Modelling Survival Data in Medical Research*: Springer US.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187-202.

Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data*: Chapman and Hall/CRC.

Cutler, S. J., & Ederer, F. (1958). Maximum utilization of the life table method in analyzing survival. *Journal of chronic diseases, 8*(6), 699-712.

Diggle, P., Henderson, R., & Philipson, P. (2009). Random-effects models for joint analysis of repeated-measurement and time-to-event outcomes. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal Data Analysis* (pp. 349-366). Boca Raton: FL: Chapman & Hall.

Dupuy, J.-f., & Mesbah, M. (2002). Joint modeling of event time and nonignorable missing longitudinal data. *Lifetime data analysis, 8*, 99-115.

Faucett, C. L., & Thomas, D. C. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in medicine, 15*(15), 1663-1685.

Fleming, T. R., & Harrington, D. P. (1984). Nonparametric estimation of the survival distribution in censored data. *Communications in Statistics-Theory and Methods, 13*(20), 2469-2486.

Fleming, T. R., & Harrington, D. P. (1991). *Counting Processes and Survival Analysis*: Wiley.

Garre, F. G., Zwinderman, A. H., Geskus, R. B., & Sijpkens, Y. W. (2008). A joint latent class changepoint model to improve the prediction of time to graft failure. *Journal of the Royal Statistical Society Series A: Statistics in Society, 171*(1), 299-308.

Gerds, T. A., & Schumacher, M. (2006). Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. *Biometrical Journal, 48*(6), 1029-1040. doi:https://doi.org/10.1002/bimj.200610301

Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine, 18*(17-18), 2529-2545. doi:https://doi.org/10.1002/(SICI)1097-0258(19990915/30)18:17/18<2529::AID-SIM274>3.0.CO;2-5

Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*: Springer.

Heagerty, P., & Zheng, Y. (2005). Survival Model Predictive Accuracy and ROC Curves. *Biometrics, 61*(1), 92-105. doi:https://doi.org/10.1111/j.0006-341X.2005.030814.x

Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics, 1*(4), 465-480.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*: Wiley.

Hsieh, F., Tseng, Y. K., & Wang, J. L. (2006). Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics, 62*(4), 1037-1043.

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*: Wiley.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457-481.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963-974.

Li, K., & Luo, S. (2019). Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimer's disease. *Statistical methods in medical research, 28*(2), 327-342.

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Oliver, S. (2006). *SAS for Mixed Models, Second Edition*: SAS Publishing.

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in medicine, 19*(13), 1793-1819.

Liu, X. (2015). *Methods and applications of longitudinal data analysis*: Elsevier.

Mikkonen, S., Rahikainen, M., Virtanen, J., Lehtonen, R., Kuikka, S., & Ahvonen, A. (2008). A linear mixed model with temporal covariance structures in modelling catch per unit effort of Baltic herring. *ICES Journal of Marine Science, 65*(9), 1645-1654.

Mogensen, U. B., Ishwaran, H., & Gerds, T. A. (2012). Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *Journal of statistical software, 50*(11), 1-23.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics, 14*(4), 945-966.

Papageorgiou, G., Mauff, K., Tomer, A., & Rizopoulos, D. (2019). An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual review of statistics and its application, 6*(1), 223-240.

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3-14.

Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics, 4*(1), 12-35.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika, 69*(2), 331-342.

Proust-Lima, C., & Taylor, J. M. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics, 10*(3), 535-549.

Ranganathan, P., & Pramesh, C. (2012). Censoring in survival analysis: potential for bias. *Perspectives in clinical research, 3*(1), 40.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics, 67*(3), 819-829.

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*: Chapman and Hall/CRC.

Rizopoulos, D., Molenberghs, G., & Lesaffre, E. M. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal, 59*(6), 1261-1276.

Rizopoulos, D., & Takkenberg, J. J. (2014). Tools & techniques--statistics: Dealing with time-varying covariates in survival analysis--joint models versus Cox models. *EuroIntervention, 10*(2), 285-288. doi:10.4244/eijv10i2a47

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*: CRC press.

Schoop, R., Graf, E., & Schumacher, M. (2008). Quantifying the Predictive Performance of Prognostic Models for Censored Survival Data with Time-Dependent Covariates. *Biometrics, 64*(2), 603-610. doi:https://doi.org/10.1111/j.1541-0420.2007.00889.x

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461-464.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology, 21*(1), 128.

Sweeting, M. J., & Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biom J, 53*(5), 750-763. doi:10.1002/bimj.201100052

Sylvestre, M. P., & Abrahamowicz, M. (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in medicine, 28*(27), 3437-3453.

Tsiatis, A. A., & Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 809-834.

Van Houwelingen, H. C., & Putter, H. (2011). *Dynamic prediction in clinical survival analysis*: CRC Press.

Verbeke, G. (1997). Linear Mixed Models for Longitudinal Data. In *Linear Mixed Models in Practice: A SAS-Oriented Approach* (pp. 63-153). New York: Springer New York.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*: Springer

Verbeke, G., & Molenberghs, G. (2009). Linear mixed models for longitudinal data. In *Linear mixed models for longitudinal data*. New York: Springer Science & Business Media.

West, B. T., Welch, K. B., & Galecki, A. T. (2014). *Linear mixed models: a practical guide using statistical software*: Chapman and Hall/CRC.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika, 80*(4), 791-795.

Wulfsohn, M. S., & Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 330-339.

Ye, W., Lin, X., & Taylor, J. M. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data–a two-stage regression calibration approach. *Biometrics, 64*(4), 1238-1246.

Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics, 44*(4), 1049-1060. doi:10.2307/2531734

Chapter 3 Literature Review of Dynamic Prediction Models

There is an abundance of dynamic prediction models for clinical outcomes in the literature. The quality of evaluation and reporting of these models varies. This chapter reviews the conduct and reporting of dynamic prediction models in the literature with a specific focus on HIV.

3.1 Background

HIV is a disease of significant burden worldwide. Almost 33 million individuals have died since the start of the epidemic (United Nations Programme on HIV/AIDS [UNAIDS], 2020). In 2019, it was estimated that 38 million people were living with HIV globally (United Nations Programme on HIV/AIDS [UNAIDS], 2020). For individuals living with HIV to remain healthy, public health responses such as antiretroviral therapy treatment should be initiated early and maintained for the remainder of life (World Health Organization [WHO], 2015). Regular monitoring of people living with HIV can provide ongoing markers of HIV progression, including HIV viral load or CD4 cell count (WHO, 2016).

Individualised predictions for clinical outcomes using biomarkers, sometimes called personalised medicine, are promoted to improve care by permitting individualised medication schedules follow-up (Hemingway et al., 2013). Dynamic prediction models could supplement informed decision-making on individualised medication schedules and follow-ups. These prediction models enable the prognosis of clinical events, where the prediction can be updated anytime during follow-up with new clinical information available at the time of prediction (Rizopoulos, 2011, 2012).

Various measures of predictive performance should be used to estimate discrimination, calibration, and overall performance to validate predictions from dynamic prediction models (Proust-Lima & Blanche, 2016). Common discrimination metrics include area under the curve of the receiver operating curve or C-statistics, sensitivity, and specificity (Steyerberg et al., 2010). Calibration measures such as plots of observed data for individuals against predicted data are used to estimate how well a model predicts the observed clinical outcome (Schoop, Graf, & Schumacher, 2008).

Statistical regression modelling for dynamic prediction models is typically based on multistate or landmark regression models (Schumacher, Hieke, Ihorst, & Engelhardt, 2020). Landmark regression models include time-to-event analysis (Van Houwelingen, 2007), direct binomial regression (Grand, de Witte, & Putter, 2018) and joint models for longitudinal data and time-to-event data (Proust-Lima & Taylor, 2009; Rizopoulos, 2011; Yu, Taylor, & Sandler, 2008). The first application of these models was in oncology, where individualised predictions of prostate cancer disease progression were made (Proust-Lima & Taylor, 2009; Yu et al., 2008). Extensions include implementing an online calculator for individualised predictions of clinical recurrence of prostate cancer for up to three years (Taylor et al., 2013). Other applications of dynamic prediction models have been in chronic diseases such as HIV (Rizopoulos, 2011) and cardiovascular disease (Paige et al., 2018). In non-chronic disease, dynamic prediction models have been applied to cardio-thoracic surgery (Andrinopoulou, Rizopoulos, Takkenberg, & Lesaffre, 2015) and intensive care unit settings (Heyard, Timsit, Essaied, Held, & consortium, 2019; Musoro, Zwinderman, Abu-Hanna, Bosman, & Geskus, 2018).

While individualised care for HIV is advocated, it is not clear if there are adequate models or algorithms in this setting (El-Sadr, Rabkin, & DeCock, 2016). A literature review was conducted to identify published dynamic prediction models applied to HIV, understand what models are available and assess the quality of model evaluation and reporting.

Section 3.2 outlines the methodology of the literature review, and Section 3.3. summaries the findings, a discussion is provided in Section 3.4, and Section 3.5 concludes the Chapter.

3.2 Methods

3.2.1 Sources

A literature search was conducted in PubMed, focusing on research publications between 1 January 1990 and 21 January 2021, restricted to publications in English. A combination of search terms was used to identify dynamic prediction models applied to HIV. The bibliographies of all included studies were manually searched to identify additional studies missed from the original search. The complete search strategy is available in Appendix 1.

3.2.2 Study inclusion and exclusion criteria

The following inclusion criteria were used:

- Statistical modelling method applied to HIV,
- Modelling approach from multistate or landmark regression models (including time-to-event, direct binomial, and joint models).

Exclusion criteria were determined in advance and included:

- Unrelated to HIV,
- Full-text not available,
- Unpublished conference proceedings,
- Cross-sectional study design,
- Static prediction models,
- Meta-analysis or systematic review,
- Simulation study without application of empirical data,
- Screening tool application data.

3.2.3 Selection of studies

Two screening stages were conducted to determine if the identified articles met the inclusion criteria. The first stage consisted of screening titles and abstracts to assess their relevance. The relevant articles were taken through a full-text review on the eligibility assessment stage. If it was unclear whether to include an article after title and abstract screening, it was automatically brought forward to a full-text review for eligibility.

3.2.4 Study quality assessment

The aspects of the checklist for critical appraisal and data extraction for systematic reviews of prediction modelling studies (CHARMS) and prediction model risk of bias assessment tools (PROBAST) were used to assess the quality of reporting in the studies and the risk of bias (ROB) of individual studies (Wolff et al., 2019). A ROB is introduced when the design and analysis of prediction model studies lead to distorted estimates of predictive performance (Moons et al., 2014). Potential sources of ROB in the studies were grouped into four domains:

participants, covariates, outcome, and analysis (Wolff et al., 2019). The ROB in each domain was assessed and assigned a rating of low, high, or unclear, and an overall rating for each prediction model study was subsequently assigned.

3.2.5 Data extraction strategy

The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) (Collins, Reitsma, Altman, & Moons, 2015) and CHARMS checklists were used to guide data extraction in identified prediction model studies. Information extracted from included studies was organised into the following: (1) study setting; (2) HIV-related outcome; (3) modelling approach and span of prediction; (4) predictive accuracy; and (5) model validation. A list of extracted information is provided in Appendix 2 and Appendix 3. The preferred reporting items for systematic reviews and meta-analyses (PRISMA) checklist was used as a basis for reporting this review (Page et al., 2021).

3.2.6 Data synthesis and presentation

Quantitative analysis of the included prediction model studies was not conducted. A narrative analysis was used to synthesise findings and summarise what the different studies found.

3.3 Results

A total of 242 studies were identified from the literature search. Of the 242 articles, 198 were excluded during the title and abstract screening stage for the following reasons: (i) not HIV-related (n=3); (ii) systematic review (n=11); (iii) simulation study (n=11); (iv) cross-sectional analysis (n=38); (v) modelling type not related to dynamic prediction (n=92); (vi) model development with no predictions (n=29); and (vii) identification of predictors (n=14). The remaining 44 studies had a full-text assessment with 38 subsequently excluded based on: (i) statistical technique to dynamic predictions with no prognosis (n=27); (ii) identification of predictors only (n=4); (iii) prognosis not HIV-related (n=1); (iv) estimation of rates (n=1); (v) parameter estimation (n=3); and (vi) optimisation with no prediction (n=2). Only six studies meeting the inclusion criteria were eligible for inclusion. The inclusion criteria are summarised

in Figure 3.1. Data extraction and critical appraisal were performed in these eligible studies (Table 3.1 and Table 3.2).

3.3.1 Study design and population

The prediction model studies properties are summarised in Table 3.1. Reviewed prediction models were based on data from cohort studies, randomised multicentre trials and retrospective surveys, primarily from the United States of America (USA) (Aldrete, 2020; Barbieri, 2020; Barrett, 2017; Rizopoulos, 2011), with one each from Iran (Khorashadizadeh, 2020) and Senegal (Tournoud, 2010). Dates of publication ranged from 2010 to 2020. Two prediction models were developed using the same cohort, a randomised multicentre trial in the USA (Barbieri, 2020; Rizopoulos, 2011). The sample size used to create the models varied from n=111 to n=2242 (median n=467). The study populations included multiple sex, apart from one study focusing on women living with HIV (Barrett, 2017).

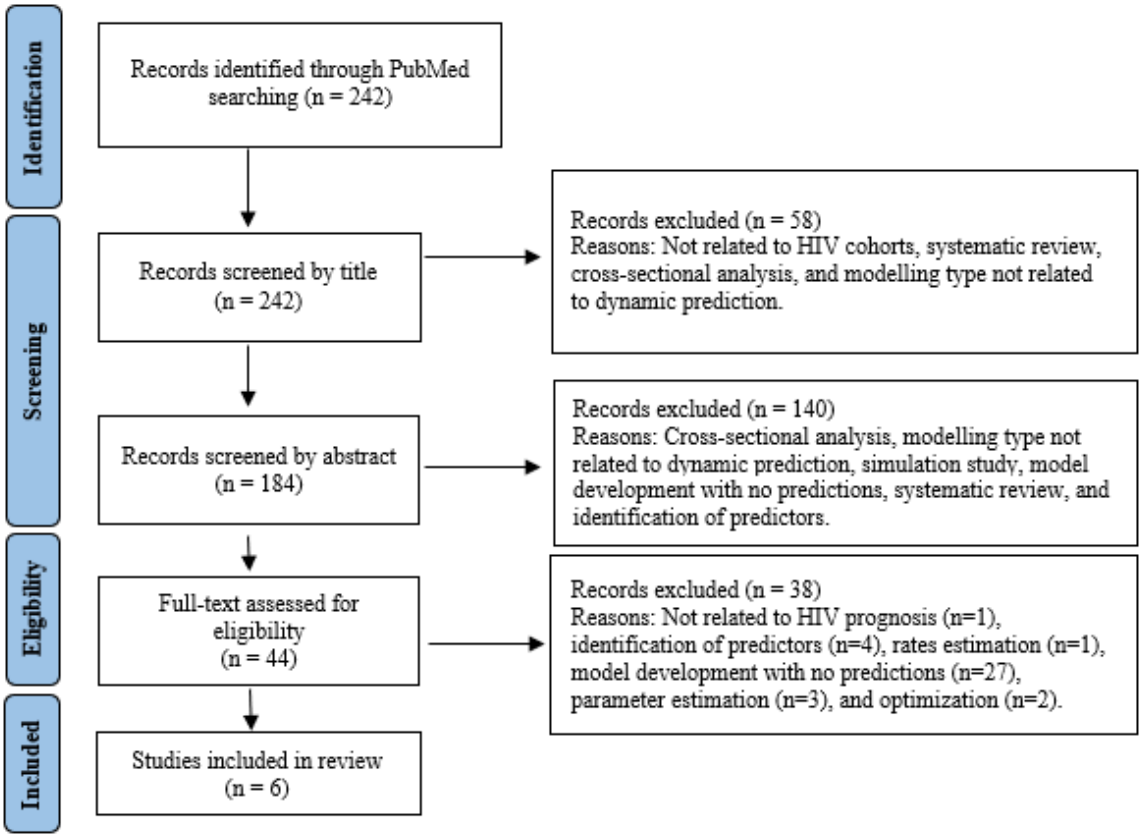

Figure 3.1: PRISMA flow diagram of study inclusions and exclusions.

3.3.2 Frequency of observations

From five of the six studies (83%) that reported study dates, individuals were followed up between 1989 and 2014. There was a considerable variation in the duration of follow-up between studies, from two (Barbieri, 2020; Rizopoulos, 2011) to 24 years (Khorashadizadeh, 2020). Four studies (67%) reported median follow-ups of 5.3 years (Aldrete, 2020), 15.9 months (Barbieri, 2020; Rizopoulos, 2011), 22.3 months (Tournoud, 2010), while one reported the frequency of measurements every six months after enrolment (Barrett, 2017) but not follow-up duration. The remaining study did not report the median follow-up times or the frequency of measures but reported the number of repeated measurements per individual (five) instead (Khorashadizadeh, 2020).

3.3.3 HIV-related outcomes and covariates

Apart from two models, the prediction outcome was mortality (Table 3.1). One of the two studies predicted a composite outcome (acquired immunodeficiency syndrome, other serious medical events, or mortality) (Aldrete, 2020), while the other predicted the timing of detection of drug resistance (Tournoud, 2010). Generally, few covariates were included in the prediction models (median n=4). Common covariates were sex and age at enrolment, CD4 cell count (Aldrete, 2020; Barbieri, 2020; Barrett, 2017; Khorashadizadeh, 2020; Rizopoulos, 2011) and HIV viral load as a time-varying covariate (Tournoud, 2010).

Table 3.1: Prediction model study characteristics.

Study; Country; Study design	Participants and setting	Recruitment dates	Follow-up	HIV related outcome	Covariates
Aldrete et al. (Aldrete, 2020); USA*; Three cohort studies	**2242 HIV-infected individuals on ART	From 1995 to 2007	12 years	Composite outcome (AIDS***, cardiovascular disease events, deaths)	Age group, sex, race, and time-varying CD4 measurements
Barbieri and Legrand (Barbieri, 2020); USA; Randomised multicentre trial	467 HIV-infected individuals on ART	1990-1991	Two years	Death	Treatment, sex, previous opportunistic infection, zidovudine treatment indicator, and CD4 cell count
Barrett and Su (Barrett, 2017); USA; Prospective cohort study	850 HIV-infected women	1993-1990	Seven years	Death	Age, time at the start of an interval, HIV viral load group, and CD4 count
Khorashadizadeh et al. (Khorashadizadeh, 2020); Iran; Retrospective survey	213 HIV-infected individuals	1989-2014	24 Years	Death	Age, sex, addiction, follow-up time, and time-varying CD4 count measurements
Rizopoulos (Rizopoulos, 2011); USA; Randomised multicentre trial	467 HIV-infected individuals on ART	1990-1991	Two years	Death	Treatment, sex, previous opportunistic infection, zidovudine treatment indicator, and CD4 cell count
Tournoud et al. (Tournoud, 2010); Senegal; Cohort study	111 HIV-1 infected individuals	1998-2002	Five years	Drug resistance	Virological and adherence response models

*USA=United States of America; ** HIV=human immunodeficiency virus, ART=antiretroviral therapy; ***AIDS=acquired immunodeficiency syndrome.

3.3.4 Modelling approach and time span of predictions

The modelling approaches in the prediction model studies are outlined in Table 3.2. Variations of the joint model (JM) approach from the frequentist and Bayesian framework used in the included studies consisted of shared random effects and joint latent class models (JLCM). In

one of the studies, a two-stage modelling approach was compared with a joint modelling approach (Aldrete, 2020). The span of the prediction time frame ranged from two months to ten years.

3.3.5 Model performance and evaluation

At least one measure of predictive performance was reported in each of the included studies (Table 3.2). Time-varying AUCs of the ROC ranging from 0.64 to 0.75 were used for discrimination (Aldrete, 2020; Khorashadizadeh, 2020; Rizopoulos, 2011), while prediction plot (Barbieri, 2020; Barrett, 2017) and posterior predictive checks (Tournoud, 2010) were carried out to assess the accuracy of the predictions. A Brier score was used to determine overall performance in one study; however, the results were not presented (Khorashadizadeh, 2020), and the rest of the studies did not report any measure of overall performance. There was no information on whether internal validation techniques were applied in all the studies. Only one study (Aldrete, 2020) used three different HIV cohorts to develop a prediction model but did not validate the predictions.

Table 3.2: Properties of prediction model studies modelling approaches.

Study	Approach	Outcome, Span of predictions	Predictive Performance validation			
			Type of Validation	*Discrimination	**Calibration	Overall Performance
Aldrete et al. (Aldrete, 2020)	Two-stage modelling and JM***	Composite outcome: 5-10 years	Not specified	AUC/Uno's C statistics (AUC=0.7; C=0.64 - 0.66)	Not reported	Not reported
Barbieri and Legrand (Barbieri, 2020)	†JLCM and SREM	Death, 20 months	Simulation	Not reported	Prediction plot	Not reported
Barrett and Su (Barrett, 2017)	JM	Death, two years	Simulation	Not reported	Prediction plot	Not reported
Khorashadizadeh et al. (Khorashadizadeh, 2020)	JLCM and SREM	Death, three years	Not reported	AUC (0.75 for JLCM; 0.64 for SREM)	Not reported	Unspecified Brier score
Rizopoulos (Rizopoulos, 2011)	JM	2-8 months	Simulation	AUC (0.67-0.72)	Not reported	Not reported
Tournoud et al. (Tournoud, 2010)	JM; Markov model and logistic mixed effects	six months	Simulation	Not reported	Posterior predicted checks	Not reported

*Time-dependent area under the curve (AUC) of the receiver operating curve (ROC) or C statistic as discrimination metrics; **calibration plot as a calibration measure; ***JM=Joint Model; †JLCM=Joint Latent Class Model, SREM=Shared Random Effect Model.

3.3.6 Risk of bias

The overall risk of bias was determined for each prediction model across the six studies using PROBAST (see Table 3.3 and Table 3.4). All studies had a high risk of bias based on the four domains (participants, covariates, outcomes, and analysis).

Table 3.3: Overview of prediction models for the prognosis of HIV-related outcomes.

Study, data and outcome	Covariates in the final model	Sample size (no of individuals with outcomes)	Predictive performance validation		Overall risk of bias**
			Type of validation	Performance*	
***Aldrete et al. (Aldrete, 2020); HIV-infected individuals on ART from three cohort studies; composite outcome.	Age group, sex, race, and time-varying CD4 measurements.	2242 (267)	Not specified	AUC; Uno's C	High
Barbieri and Legrand (Barbieri, 2020); HIV-infected individuals on ART from a randomised multicentre trial; death.	Treatment, sex, previous opportunistic infection, zidovudine treatment indicator, and time-varying CD4 cell count.	467 (188)	Simulation	Prediction plot	High
Barrett and Su (Barrett, 2017); HIV-infected women from a prospective cohort study; HIV-related death.	Age, time at the start of an interval, HIV viral load group, and CD4 count.	850 (105)	Simulation	Prediction plot	High
Khorashadizadeh et al. (Khorashadizadeh, 2020); HIV-infected individuals from a retrospective survey; death.	Age, sex, addiction, follow-up time, and time-varying CD4 count measurements.	213 (51)	Not reported	AUC	High
Rizopoulos (Rizopoulos, 2011); HIV-infected individuals on ART from a randomised multicentre trial; death.	Treatment, sex, previous opportunistic infection, zidovudine treatment indicator, and time-varying CD4 cell count.	467 (188)	Simulation	AUC	High
Tournoud et al. (Tournoud, 2010); data from HIV-1 infected individuals from a cohort study in Senegal; drug resistance.	Virological and adherence response models	111 (57)	Simulation	Posterior predicted checks	High

*Performance in the form of the area under the curve (AUC) of the receiver operating curve (ROC), calibration or overall performance; **Risk of bias using PROBAST and TRIPOD; ***HIV=human immunodeficiency virus, ART=antiretroviral therapy, CD4=cluster differentiation four.

Only one of these studies (Aldrete, 2020) had an unclear risk of bias for the participants' domain (Table 3.4). This study used combined data from a multicentre study of three cohorts for model development. The cohorts had distinct characteristics, which could result in residual confounding (Aldrete, 2020). The rest of the studies used one representative population of the target population in the cohort or a randomised multicentre trial study design for model development. Except for one study (Tournoud, 2010), the prediction model studies used common individual demographics and clinical characteristics (Table 3.3). The lack of candidate covariates in this study suggests a high risk of bias for the covariate domain induced by difficulty in summarising an individual's outcome. Using a composite outcome in one study (Aldrete, 2020) was a cause for concern about bias in the outcome domain. Composite outcomes in studies have been associated with difficulty in interpreting findings, which may reduce the credibility of the research (Ross, 2007). All studies had a high risk of bias in the analysis domain. Nonetheless, there was a reasonable number of individuals with an outcome that warranted the development of a prediction model in the studies. Four of the six studies handled continuous covariates appropriately (Barbieri, 2020; Barrett, 2017; Khorashadizadeh, 2020; Rizopoulos, 2011). They used square root CD4 count values to account for the skewed shape distribution of CD4 cell count. In one study (Tournoud, 2010), participants were missing based on the number of repeated measurements and resistance tests. The missing data was treated as missing at random and did not require special missing data techniques in a Bayesian framework for the joint model. Predictive performance was only reported in the form of discrimination (Aldrete, 2020; Khorashadizadeh, 2020; Rizopoulos, 2011) and predictive plots (Barbieri, 2020; Tournoud, 2010) in the developed models.

Table 3.4: Risk of bias assessment on four domains across six prediction model studies for HIV.

Study	Risk of bias				
	Participants	Covariates	Outcome	Analysis	*Overall
Aldrete et al. (Aldrete, 2020)	Unclear	Low	Unclear	High	†High
Barbieri and Legrand (Barbieri, 2020)	Low	Low	Low	High	†High
Barrett and Su (Barrett, 2017)	Low	Low	Low	High	†High
Khorashadizadeh et al. (Khorashadizadeh, 2020)	Low	Low	Low	High	†High
Rizopoulos (Rizopoulos, 2011)	Low	Low	Low	High	†High
Tournoud et al. (Tournoud, 2010)	Low	High	Low	High	†High

*Risk of bias using PROBAST and TRIPOD; †Risk of bias is high owing to calibration not being evaluated or not using external populations for validation

3.4 Discussion

This literature review identified only a small number of dynamic prediction models applied to HIV from the extensive HIV-associated modelling literature. Two methods have been applied: (1) variation of the joint model and (2) two-stage modelling. Most models used mortality as an HIV-related outcome of interest, reflecting the time of data and follow-up. Only one model evaluated a non-mortality-associated outcome, which was drug resistance.

The advantage of predictions that continuously update with new profile information is complemented by the time span in which these predictions are made. There was a disparity in the span of these predictions, ranging from months to ten years. Predictions made up to ten years on individuals are ideal for chronic conditions like HIV, where individuals remain on treatment for the remainder of their lives, and clinical decision-making on treatment takes place.

Predictions from dynamic models should be reliable to inform individuals and guide clinical decision-making. The included studies had only moderate discrimination ability as estimated by time-varying AUCs of the ROC curve. Given the proportion of individuals with the outcome of interest in the included studies, the discrimination ability of these models is adequate. However, it may not be good enough to consider including as part of clinical care.

Prediction plots and posterior predictive model checks did not reveal any significant problem with the fit of the modelling approaches used in the included studies. In general, at least one measure of predictive performance was reported, but there was no standardisation of measures for assessing the performance of prediction models. Most validation approaches did not use independent clinical data from different settings, potentially reflecting the lack of access to large data on independent and external populations.

3.4.1 Strengths and limitations

As more dynamic prediction models emerge over the years in HIV, the quality of reporting in published studies is an issue of concern. A significant strength of this review is the use of tools facilitating quality assessment for prediction models in these studies. A two-step quality assessment was conducted using TRIPOD and PROBAST for reporting and risk of bias. Studies were critically appraised based on the TRIPOD checklist by extracting information reflecting aspects of good practice for reporting prediction model studies. The review explicitly assessed the risk of bias in the studies using the PROBAST and CHARMS checklists. A careful review of biographies in the included studies ensured that additional articles were considered. Interpretation of the results should be done, keeping in mind the limitations of the study.

First, dynamic prediction models have become increasingly abundant in the medical literature over the years, with a lack of uniformity in language when referring to them. Some dynamic prediction model studies are at risk of not being retrieved in the literature due to the disparity in the language. Truncation and combination of all search terms, including free-text words using Boolean operators, was used in the search strategy to mitigate this risk. Second, one researcher may introduce subjectivity and bias in the screening and assessment for eligibility and inclusion stages. Carrying articles forward to the next stage if it was unclear whether they should be excluded was managed carefully to minimise the risk of subjectivity. Finally, new HIV-related publications often enter the medical literature with the increasing interest in dynamic prediction models. Therefore, this review cannot be regarded as an up-to-date list of

all currently available HIV-related dynamic prediction models. Also, studies not written in English might improve the literature when translated; these published studies will be reassessed in future iterations. Due to the slight variation in modelling approaches to dynamic prognosis prediction, the main conclusions and recommendations are implausible to change if these new or translated articles are included.

3.4.2 Implications

Prediction models are fast-growing tools to augment clinical decision-making. Specification of the clinical question and clear reporting of the model development and validation are needed to adopt these models in practice (Moons, Royston, Vergouwe, Grobbee, & Altman, 2009; Steyerberg et al., 2013). For model development and validation, representative data, ideally from nested case-control or prospective cohort studies, are recommended (Steyerberg, 2008). Most of the studies included in the review adequately described the study population. None of the predictions from the studies were validated by independent data. Nonetheless, the studies would be in little doubt if the review assessed the models' applicability based on the first three domains of the PROBAST (Wolff et al., 2019). In most cases with prognosis prediction, the interest lies in predicting the risk of an individual developing an outcome over time, and this outcome is presented as binary (Steyerberg et al., 2010). In the study by Tournoud (2010), predictions were not presented as a simple binary classification task but instead as a combination of longitudinal models. The lack of optimal ways to summarise the outcome in this model should be addressed.

It is recommended that discrimination and calibration should be included as minimum predictive performance measures for prediction model studies (Steyerberg et al., 2010). Unfortunately, prediction model studies included in the review often lacked quality in model evaluation and reporting. The lack of reporting was supported by the high risk of bias in the analysis domain of PROBAST owing to calibration not being adequately evaluated. The frequency of measurement was specified in the studies. Given the stochastic nature of time-dependent covariates, for example, biomarkers (HIV viral load or CD4) in included studies, a sensitivity analysis on varying frequencies of measurements would be of interest. Authors should adopt recommendations such as PROGRESS, TRIPOD statement and PROBAST to develop and validate prediction models to leverage the opportunities presented by prediction model studies. Given these frameworks in the literature, there is a need for more applications

of dynamic prediction models that forecast time-to-HIV-related clinical events in addition to mortality. From the large number of studies identified in our search, very few of them (only six from 242 in our studies) predict the prognosis of HIV-related outcomes in individuals with HIV. The lack of appropriate dynamic prediction models indicates a disparity in the literature regarding using the term dynamic prediction. One must exercise caution when using the limited dynamic prediction models identified in this review because they have not been validated in large independent and external populations.

3.5 Conclusion

This review demonstrated a disparity in using the term dynamic prediction with the exclusion of many identified studies. Moreover, most dynamic prediction models identified in the search were without application, while those with application did not use a consistent framework for evaluation and reporting. Despite individualised predictions being of growing interest in medical applications, there is no published empirical research on dynamic prediction models implemented as prognostic tools in HIV. Dynamic prediction models could provide an avenue for HIV-related clinical prognosis with further development. HIV-related clinical prognosis under dynamic prediction models needs large data on independent and external populations to test models for generalisable predictions to become useful and practical prognosis tools (Moons et al., 2009; Steyerberg, 2008; Tournoud, 2010). Currently, none of the dynamic prediction models identified in this review are recommended for use as practical prognosis tools. This review identified study setting, outcome type, prediction span, predictive performance, and model validation as crucial characteristics that inform the conduct and quality of model evaluation of dynamic prediction models for individualised prognosis prediction using longitudinally measured biomarkers. These characteristics are used as a basis for prognosis prediction for clinical outcomes in Chapter 5, Chapter 6 and Chapter 7.

3.6 References

Aldrete, S. J., J. H.;Easley, K. A.;Okulicz, J.;Dai, T.;Chen, Y. N.;Pino, M.;Agan, B. K.;Maves, R. C.;Paiardini, M.;Marconi, V. C. (2020). CD4 rate of increase is preferred to CD4 threshold for predicting outcomes among virologically suppressed HIV-infected adults on antiretroviral therapy. *PLoS One, 15*(1), e0227124. doi:10.1371/journal.pone.0227124

Andrinopoulou, E., Rizopoulos, D., Takkenberg, J. J. M., & Lesaffre, E. (2015). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical methods in medical research, 26*(4), 1787-1801. doi:10.1177/0962280215588340

Barbieri, A. L., C. (2020). Joint longitudinal and time-to-event cure models for the assessment of being cured. *Stat Methods Med Res, 29*(4), 1256-1270. doi:10.1177/0962280219853599

Barrett, J. S., L. (2017). Dynamic predictions using flexible joint models of longitudinal and time-to-event data. *Stat Med, 36*(9), 1447-1460. doi:10.1002/sim.7209

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC medicine, 13*(1). doi:10.1186/s12916-014-0241-z

El-Sadr, W. M., Rabkin, M., & DeCock, K. M. (2016). Population health and individualized care in the global AIDS response: synergy or conflict? *AIDS (London, England), 30*(14), 2145-2148. doi:10.1097/QAD.0000000000001192

Grand, M. K., de Witte, T. J. M., & Putter, H. (2018). Dynamic prediction of cumulative incidence functions by direct binomial regression. *Biometrical Journal, 60*(4), 734-747. doi:https://doi.org/10.1002/bimj.201700194

Hemingway, H., Croft, P., Perel, P., Hayden, J. A., Abrams, K., Timmis, A., . . . Steyerberg, E. W. (2013). Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *bmj, 346*, e5595.

Heyard, R., Timsit, J.-F., Essaied, W. I., Held, L., & consortium, o. b. o. t. C.-M. (2019). Dynamic clinical prediction models for discrete time-to-event data with competing risks—A case study on the OUTCOMEREA database. *Biometrical Journal, 61*(3), 514-534. doi:https://doi.org/10.1002/bimj.201700259

Khorashadizadeh, F. T., H.;Parsaeian, M.;Esmaily, H.;Rahimi Foroushani, A. (2020). Predicting the Survival of AIDS Patients Using Two Frameworks of Statistical Joint

Modeling and Comparing Their Predictive Accuracy. *Iran J Public Health, 49*(5), 949-958. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7475620/pdf/IJPH-49-949.pdf

Moons, K. G. M., de Groot, J. A. H., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., . . . Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med, 11*(10), e1001744-e1001744. doi:10.1371/journal.pmed.1001744

Moons, K. G. M., Royston, P., Vergouwe, Y., Grobbee, D. E., & Altman, D. G. (2009). Prognosis and prognostic research: what, why, and how? *bmj, 338*.

Musoro, J. Z., Zwinderman, A. H., Abu-Hanna, A., Bosman, R., & Geskus, R. B. (2018). Dynamic prediction of mortality among patients in intensive care using the sequential organ failure assessment (SOFA) score: a joint competing risk survival and longitudinal modeling approach. *Statistica Neerlandica, 72*(1), 34-47. doi:https://doi.org/10.1111/stan.12114

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj, 372*, n71. doi:10.1136/bmj.n71

Paige, E., Barrett, J., Stevens, D., Keogh, R. H., Sweeting, M. J., Nazareth, I., . . . Wood, A. M. (2018). Landmark Models for Optimizing the Use of Repeated Measurements of Risk Factors in Electronic Health Records to Predict Future Disease Risk. *Am J Epidemiol, 187*(7), 1530-1538. doi:10.1093/aje/kwy018

Proust-Lima, C., & Blanche, P. (2016). Dynamic Predictions. *Wiley StatsRef: Statistics Reference Online*, 1-6. doi:10.1002/9781118445112.stat07876

Proust-Lima, C., & Taylor, J. M. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics, 10*(3), 535-549. Retrieved from https://doi.org/10.1093/biostatistics/kxp009

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics, 67*(3), 819-829. doi:10.1111/j.1541-0420.2010.01546.x

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*: Chapman and Hall/CRC.

Ross, S. (2007). Composite outcomes in randomized clinical trials: arguments for and against. *American journal of obstetrics and gynecology, 196*(2), 119. e111-119. e116.

Schoop, R., Graf, E., & Schumacher, M. (2008). Quantifying the Predictive Performance of Prognostic Models for Censored Survival Data with Time-Dependent Covariates. *Biometrics, 64*(2), 603-610. Retrieved from www.jstor.org/stable/25502096

Schumacher, M., Hieke, S., Ihorst, G., & Engelhardt, M. (2020). Dynamic prediction: A challenge for biostatisticians, but greatly needed by patients, physicians and the public. *Biometrical Journal, 62*(3), 822-835. doi:https://doi.org/10.1002/bimj.201800248

Steyerberg, E. W. (2008). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*: Springer New York.

Steyerberg, E. W., Moons, K. G. M., Van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., . . . Group, P. (2013). Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med, 10*(2), e1001381-e1001381. doi:10.1371/journal.pmed.1001381

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. *Epidemiology, 21*(1). Retrieved from https://journals.lww.com/epidem/Fulltext/2010/01000/Assessing_the_Performance_of_Prediction_Models__A.22.aspx

Taylor, J. M., Park, Y., Ankerst, D. P., Proust-Lima, C., Williams, S., Kestin, L., . . . Sandler, H. (2013). Real-Time Individual Predictions of Prostate Cancer Recurrence Using Joint Models. *Biometrics, 69*(1), 206-213. doi:https://doi.org/10.1111/j.1541-0420.2012.01823.x

Tournoud, M. E., J. F.;Ecochard, R.;DeGruttola, V. (2010). Adherence to antiretroviral therapy, virological response, and time to resistance in the Dakar cohort. *Stat Med, 29*(1), 14-32. doi:10.1002/sim.3779

United Nations Programme on HIV/AIDS [UNAIDS]. (2020). Latest global and regional statistics on the status of the AIDS epidemic [Fact Sheet]. Retrieved from https://www.unaids.org/en/resources/fact-sheet

Van Houwelingen, H. C. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics, 34*(1), 70-85.

WHO. (2016). Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations–2016 update. Retrieved from https://www.who.int/publications/i/item/9789241511124

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., . . . Mallett, S. (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med, 170*(1), 51-58. doi:10.7326/m18-1376

World Health Organization [WHO]. (2015). Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/186275/9789241509565_eng.pdf

Yu, M., Taylor, J. M. G., & Sandler, H. M. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *Journal of the American Statistical Association, 103*(481), 178-187.

Chapter 4 Data Sources for Motivating Disease Case Studies

Longitudinal continuous biomarker and time-to-event data from motivating cohorts are briefly introduced in this chapter. HIV and diabetes cohorts were formed using routine test results administered by the South African National Health Laboratory Service (NHLS). As the largest public-sector diagnostic pathology service in South Africa, the NHLS offers laboratory testing for public-sector health facilities, serving 80% of the nation's population (National Health Laboratory Service, 2021a). Data from the HIV and diabetes cohorts were used for Objectives 2, 3, 4, and 5. Detailed descriptions and explorations of these cohorts are provided in this chapter.

4.1 Routine HIV Data

4.1.1 Cohort description

Study design, setting and population

The data is based on the routine HIV viral load tests administered by the NHLS in the Western Cape province. The cohort consists of individuals attending routine HIV care across public sector healthcare facilities between 1 January 2008 and 30 September 2018 in the Western Cape, South Africa. The Western Cape HIV viral load tests were in the National Priority Programme (NPP) from the NHLS. Data were available from six components of the NPP HIV section: (i) CD4 unit, (ii) HIV viral load unit, (iii) early infant diagnosis unit, (iv) HIV genotyping unit, (v) Research Diagnostic Group, (vi) Business Analysis and Monitoring, and Evaluation (National Health Laboratory Service, 2021b). The data used in this thesis were extracted from the HIV viral load unit. The data contains individual unique identifiers, clinical visit information (date of specimen test or calendar year and healthcare facility), demographic characteristics (age or date of birth, sex), and laboratory test results (HIV viral load).

Inclusion exclusion

The exclusion criteria for the data were as follows: (i) individuals < 16 years of age at their first HIV viral load test, ii) individuals from independent care, correctional and private healthcare facilities, and (iii) individuals with < 3 repeated HIV viral load tests. To ensure that

the covariance structure for the individuals' random effects in the longitudinal model is well-defined, a minimum of three repeated HIV viral load measurements were employed. The exclusion criteria are summarised in Figure 4.1.

Data quality checks and missing data

Separate unlinked individual HIV viral load test results were linked using an innovative linkage procedure as previously described elsewhere (Mukonda, Hsiao, Vojnov, Myer, & Lesosky, 2020). Briefly, a custom record linkage procedure was used to identify participants in the NHLS HIV database. It separated them into a reference set (those with matching first and last names, date of birth, and anonymized individual identifier) and comparative sets. Summary scores were evaluated in a sensitivity analysis, and the record linkage procedure attained an exceptional linkage (Mukonda et al., 2020). Reliable HIV viral load test results were extracted from the NHLS electronic database, and no missing data was found.

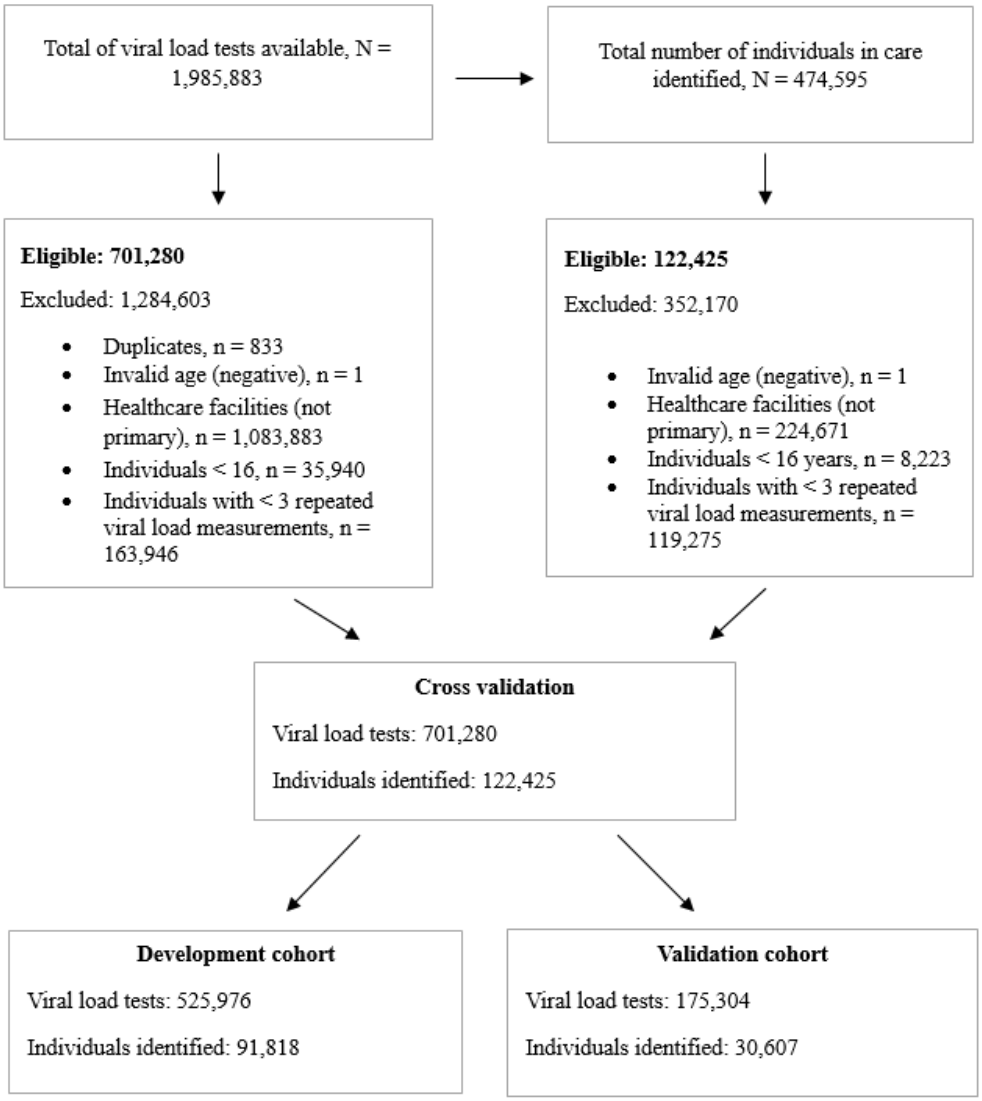

Figure 4.1: Flow diagram of individuals in the routine HIV Western Cape cohort.

4.1.2 Data exploration

Between 1 January 2008 and 30 September 2018, there were 474 595 individuals living with HIV on ART from the overall cohort. After excluding 352 170 (74%) individuals, a cohort of 122 425 (26%) individuals was included in the analysis (Figure 4.1). Of the 122 425 individuals included in the analysis, 91 818 (75%) and 30 607 (25%) were randomly allocated to the development and validation cohorts, respectively.

Longitudinal continuous HIV viral load biomarker and time-to-virologic failure

The longitudinal outcome was the repeated measurements of the continuous HIV viral load. An excess of large values often characterises these HIV viral load values. Thus, for the remainder of this thesis, log base 10-transformed HIV viral load was used. Figure 4.2 shows the distribution of the observed (log base 10) longitudinal continuous HIV viral load (copies/mL) values of individuals in the routine HIV Western Cape cohort. A notable feature of the HIV viral load data was the high proportion of values below a limit detection of 50 (1.7 on the log base 10 scale) copies/mL (referred to in this thesis as undetectable HIV viral load) and exhibiting a right-skewed distribution. This distributional characteristic of the HIV viral load data will be accommodated in a two-part joint model (to be introduced in Chapter 7).

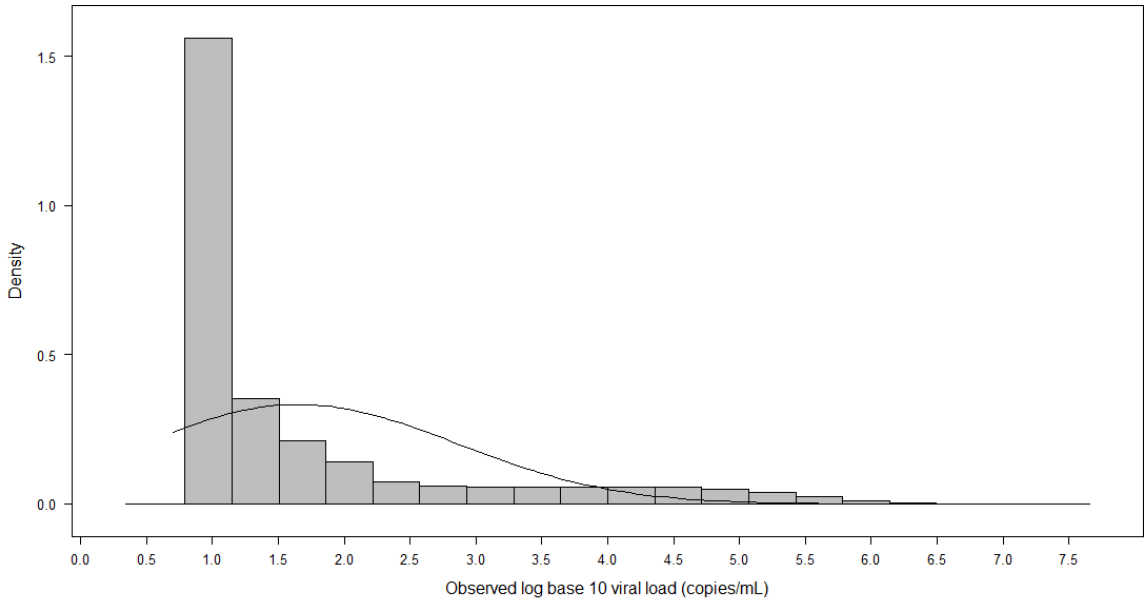

Figure 4.2: Observed (log base 10) longitudinal continuous HIV viral load (copies/mL) biomarker values of individuals in the historic routine HIV Western Cape cohort.

The time-to-event was the time to the first virologic failure. Using the 2019 South African guidelines (South African National Department of Health, 2019a, 2019b), virologic failure was defined as two consecutive HIV viral load tests ≥1000 copies/mL six months apart after ART initiation for at least six months. The duration of observed follow-up time was calculated from enrolment to the last observed follow-up time due to virologic failure or censoring. For individuals with virologic failure, time-to-first virologic failure was defined as the duration from observed enrolment time to the minimum observed follow-up time between two

sequential visits. Individuals were considered censored if they did not have virologic failure at the end of follow-up or the last observed time if they were lost to follow-up. Follow-up time in weeks was used to improve the clinical interpretation and comparison of results.

As discussed in Chapter 2, repeated biomarker measurements are often assumed to be associated with a time-to-event of interest. For example, Figure 4.3 shows observed repeated measurements (i.e., observed longitudinal trajectories) of log base 10 HIV viral load (copies/mL) biomarker for a random sample of individuals living with HIV on ART followed-up from study entry time until virologic failure or last observed follow-up. In this figure, individuals who had virologic failure during the observed follow-up time are represented by dashed lines, and solid lines represent those who did not. An important feature of the observed trajectories is considerable variability in the trajectories of the HIV viral load, with each individual appearing to have their own trajectory.

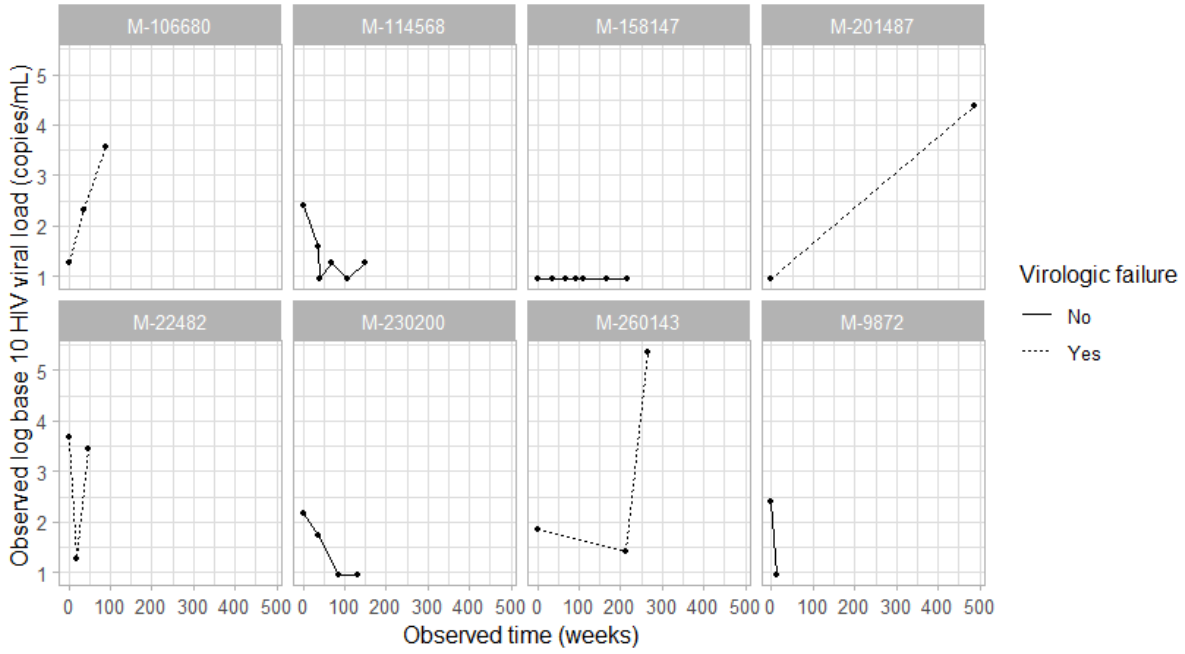

Figure 4.3: Observed (log base 10) longitudinal HIV viral load (copies/mL) measurements for eight randomly selected individuals followed-up from study entry until first virologic failure or censoring time in the Western Cape, South Africa.

Demographic characteristics

Baseline (study entry time) was considered the first HIV viral load measurement date for each individual in the cohort, between January 1, 2008, and September 30, 2018. Baseline characteristics of the Western Cape HIV cohort are shown in Table 4.1. A large proportion (70%) of the individuals in the cohort were females. The median (interquartile range; IQR) age of individuals at enrolment was 33 (28; 40). In the 11 years of follow-up time, 2018 was the year fewer (0.2%) of the individuals were followed up, and 2015 was the year where most (16%) of the individuals were followed up. As shown in Figure 4.2 and Table 4.1, a large proportion (71%) of the individuals in the cohort had HIV viral load values <50 copies/mL, while 14% had an elevated HIV viral load.

The characteristics of individuals who had virologic failure during observed follow-up time in the overall development and validation cohorts are shown in Table 4.2. Overall, individuals in the cohort had a median of three HIV viral load measurements (IQR: 1–5). By the end of the observed follow-up time, 109 878 (89.8%) were censored, and 12 547 (10.2%) had virologic failure with a median follow-up time of 41 weeks (IQR: 16 - 119) in the overall cohort. For the development cohort, 9357 (10.2%) individuals had virologic failure with a median follow-up time of 40 weeks (IQR: 16 - 118). In the validation cohort, 3190 (10.4%) individuals had virologic failure with a median follow-up time of 42 weeks (IQR: 16 - 121).

Table 4.1: Individual characteristics at baseline in the routine HIV Western Cape cohort.

Characteristic	Overall	Viral load at baseline (copies/mL)			
		< 50	≥ 50 & < 400	≥ 400 & < 1000	≥ 1000
	N = 122 425[1]	N = 87 050[1]	N = 15 596[1]	N = 2645[1]	N = 17 134[1]
Baseline age (years), Median (IQR)	33 (28, 40)	33 (28, 40)	34 (28, 41)	33 (28, 40)	33 (27, 39)
Sex, n (%)					
Male	37 185 (30.4)	25 240 (29)	5704 (36.6)	799 (30.2)	5442 (31.8)
Female	85 240 (69.6)	61 810 (71)	9892 (63.4)	1846 (69.8)	11 692 (68.2)
Cohort enrolment, n (%)					
2008	9061 (7.4)	7676 (8.8)	312 (2)	172 (6.5)	901 (5.3)
2009	7352 (6)	5971 (6.9)	457 (2.9)	147 (5.6)	777 (4.5)
2010	7079 (5.8)	5442 (6.3)	622 (4)	151 (5.7)	864 (5)
2011	11 777 (9.6)	8466 (9.7)	1379 (8.8)	206 (7.8)	1726 (10.1)
2012	14 675 (12)	10 727 (12.3)	1848 (11.8)	254 (9.6)	1846 (10.8)
2013	16 243 (13.3)	11 635 (13.4)	2037 (13.1)	314 (11.9)	2257 (13.2)
2014	15 426 (12.6)	11 045 (12.7)	1819 (11.7)	289 (10.9)	2273 (13.3)
2015	19 325 (15.8)	13 485 (15.5)	2747 (17.6)	437 (16.5)	2656 (15.5)
2016	16 352 (13.4)	10 107 (11.6)	3445 (22.1)	467 (17.7)	2333 (13.6)
2017	4913 (4)	2420 (2.8)	906 (5.8)	193 (7.3)	1394 (8.1)
2018	222 (0.2)	76 (0.1)	24 (0.2)	15 (0.6)	107 (0.6)

[1]Statistics presented: N = total number of individuals; IQR = interquartile range; n = frequency

Table 4.2: Characteristics of individuals at the observed time of virologic failure in the routine HIV Western Cape overall, development and validation cohorts.

Characteristic	Overall cohort N = 12 547[1]	Development cohort N = 9357[1]	Validation cohort N = 3190[1]
$\log_{10}$ viral load (copies/mL), Median (IQR)	4.2 (3.6, 4.8)	4.2 (3.6, 4.8)	4.2 (3.6, 4.8)
Baseline age (years), Median (IQR)	33 (28, 39)	33 (27, 39)	33 (28, 39)
Sex, n (%)			
Male	4009 (32)	2986 (31.9)	1023 (32.1)
Female	8538 (68)	6371 (68.1)	2167 (67.9)
Cohort year, n (%)			
2008	205 (1.6)	149 (1.6)	56 (1.8)
2009	478 (3.8)	351 (3.8)	127 (4.0)
2010	523 (4.2)	397 (4.2)	126 (3.9)
2011	1282 (10.2)	953 (10.2)	329 (10.3)
2012	1249 (10)	933 (10)	316 (9.9)
2013	1584 (12.6)	1182 (12.6)	402 (12.6)
2014	1448 (11.5)	1075 (11.5)	373 (11.7)
2015	1373 (10.9)	1008 (10.8)	365 (11.4)
2016	1747 (13.9)	1294 (13.8)	453 (14.2)
2017	1702 (13.6)	1276 (13.6)	426 (13.4)
2018	956 (7.6)	739 (7.9)	217 (6.8)
Follow-up time (week), Median (IQR)	40.9 (15.9, 119)	40.1 (15.9, 118.1)	42.2 (16, 121)

[1]Statistics presented: N = total number of individuals; IQR = interquartile range; n = frequency

Duration between the observed follow-up times and virologic failure patterns

The duration between the observed follow-up times for individuals with at least one consecutive elevated HIV viral load in the overall cohort is shown in Figure 4.4. These individuals had a median duration of 26.3 weeks (6 months) between observed follow-up times, and only 25% had less than 16.3 weeks of duration between observed follow-up times. Thus, a sufficient subset of individuals met the definition of time to first virologic failure.

The virologic failure-free probability (with 95% confidence interval; CI) in the individuals was 0.952 (95% CI: 0.951, 0.954) within 13 weeks (3 months) of observed follow-up time, 0.949 (95% CI: 0.948, 0.95) within 26 weeks (6 months) of observed follow-up time, 0.937 (95% CI: 0.936, 0.938) within 52 weeks (12 months) of observed follow-up time, and 0.92 (95% CI: 0.918, 0.921) within 104 weeks (24 months) of observed follow-up time from the overall routine HIV cohort. These virologic failure-free probabilities are summarised in the Kaplan-Meier plot (Figure 4.5). A large proportion of the individuals did not experience virologic failure during the observed follow-up time. The next chapter considers clinical applications of statistical regression models (introduced in Chapter 2) for prognosis prediction to individuals living with HIV to estimate the hazard of virologic failure.

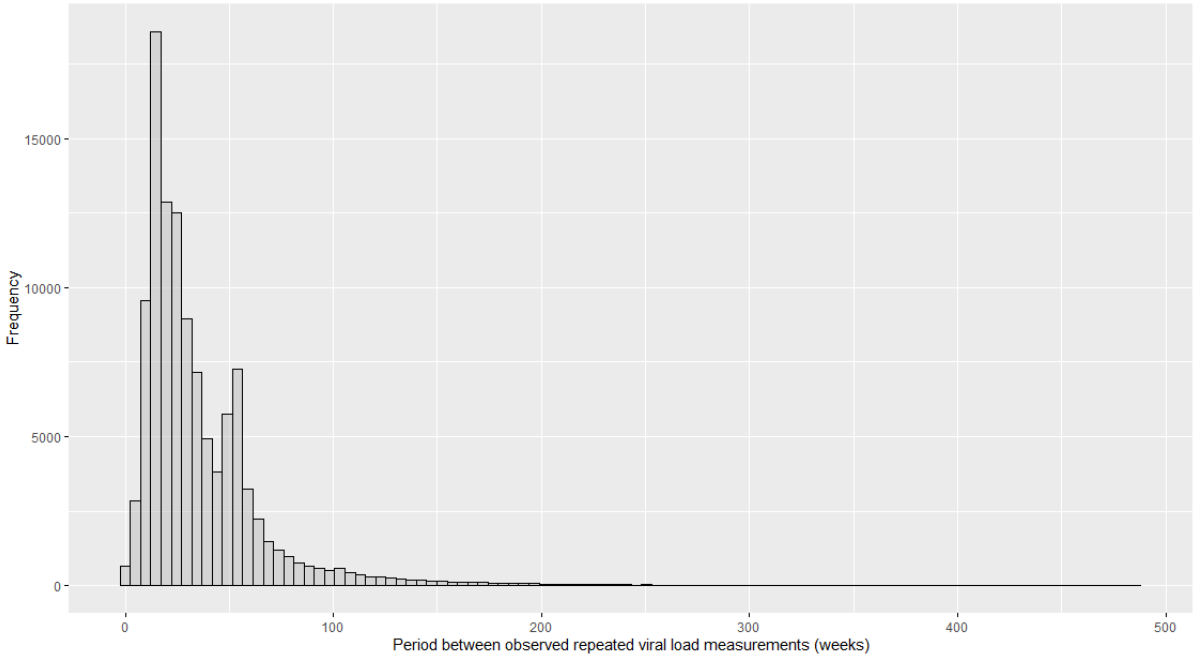

Figure 4.4: Duration (weeks) between observed follow-up time for individuals who had at least one consecutive elevated HIV viral load in the routine HIV Western Cape overall cohort.

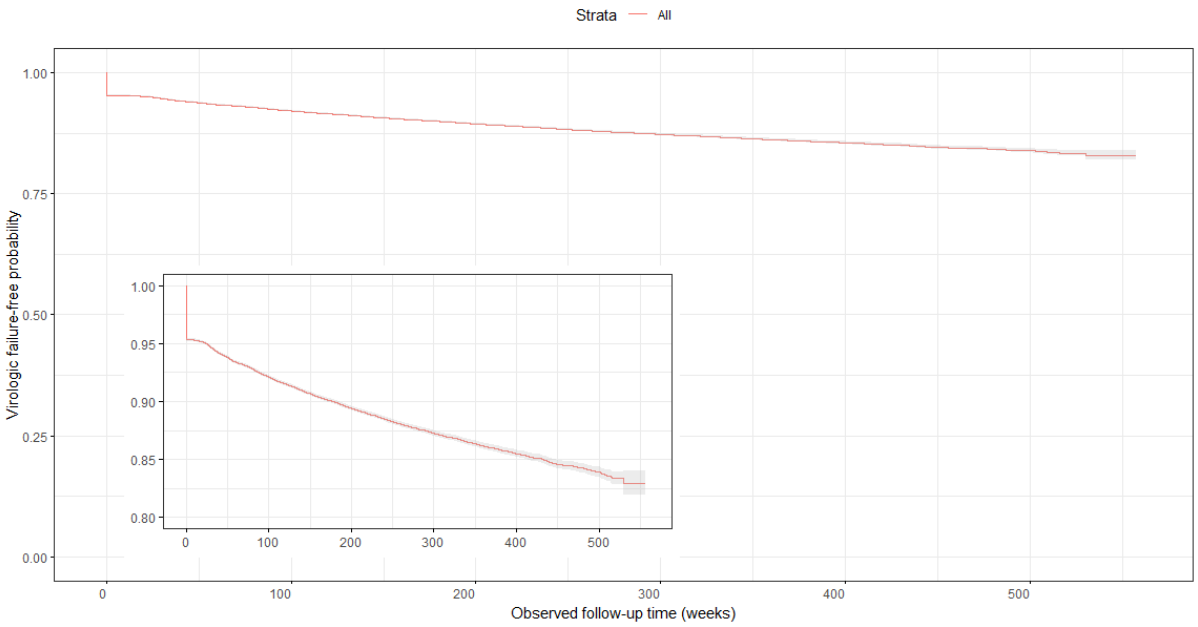

Figure 4.5: Kaplan-Meier estimate of the virologic failure-free probability for individuals living with HIV in the routine HIV Western Cape overall cohort. The inserted figure represents a zoomed-in version of the original figure.

4.2 Routine Diabetes Data

4.2.1 Cohort description

Study design, setting and population

The data is based on the Western Cape province routine HbA1c tests conducted by the NHLS. The cohort consists of individuals attending routine diabetes care across public sector healthcare facilities between 1 January 2016 and December 2021 in Western Cape, South Africa. The data contains individual unique identifiers, demographic characteristics (age or date of birth, sex), and laboratory test results (HbA1c).

Inclusion exclusion

The exclusion criteria for the data were as follows: (i) individuals < 18 years of age at their first HbA1c test, and (ii) individuals with < 3 repeated HbA1c tests. To ensure that the covariance structure for the individuals' random effects in the longitudinal model is well-defined, a minimum of three repeated HbA1c measurements were employed. The exclusion criteria are summarised in Figure 4.6.

Data quality checks and missing data handling

As previously reported, a reliable individual identifier was used to link multiple HbA1c test results. The individual identifier was used to link and track multiple HBA1c test results from the same individuals (Odayar et al., 2024). There was no need to address missing data because complete, reliable HbA1c test results were extracted from the NHLS electronic database.

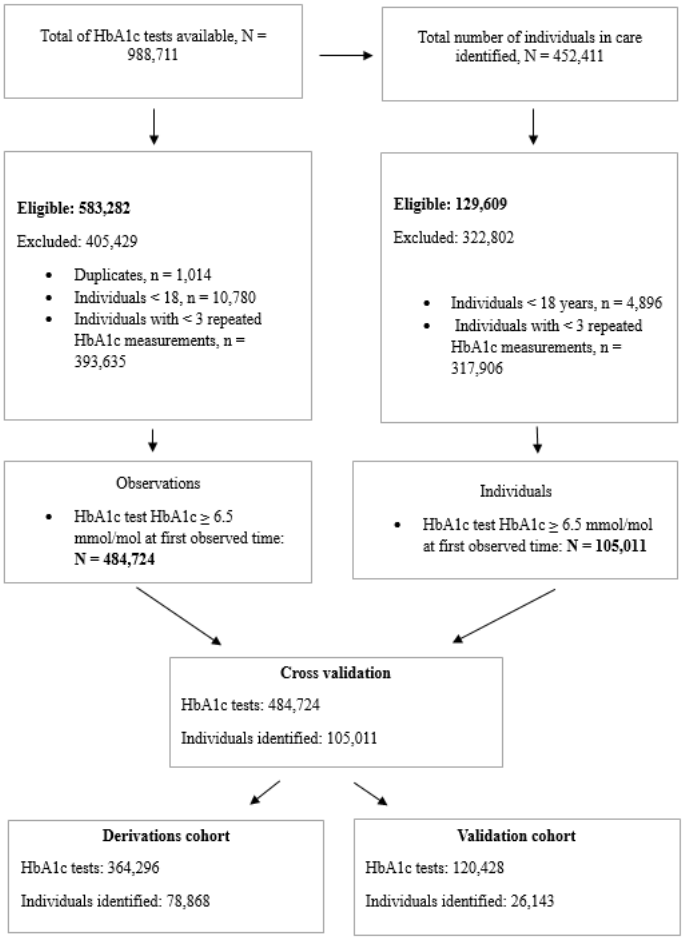

Figure 4.6: Flow diagram of individuals in the routine diabetes cohort.

4.2.2 Data exploration

There were 452 411 T2DM individuals from the Western Cape cohort between 1 January 2016 and December 2021. Of these individuals, 129 609 (28.6%) were 18 years and older and had more than two HbA1c test results. After excluding 24 598 (19%) individuals who were glycaemic controlled (index HbA1c < 6.5) at enrolment, a cohort of 105 011 (81%) individuals

was included in the analysis (Figure 4.6). Of the 105 011 individuals included in the analysis, 78 868 (75%) and 26 143 (25%) were randomly allocated to the development and validation cohorts, respectively.

Longitudinal continuous HbA1c biomarker and time to glycaemic control

The longitudinal outcome was the repeated measurements of the continuous HbA1c biomarker. Figure 4.7 shows the distribution of the observed longitudinal continuous HbA1c (mmol/mol) measures of individuals in the routine diabetes Western Cape cohort. As shown in the figure, a high proportion of values above an HbA1c glycaemic control threshold of 6.5 mmol/mol was used in this thesis.

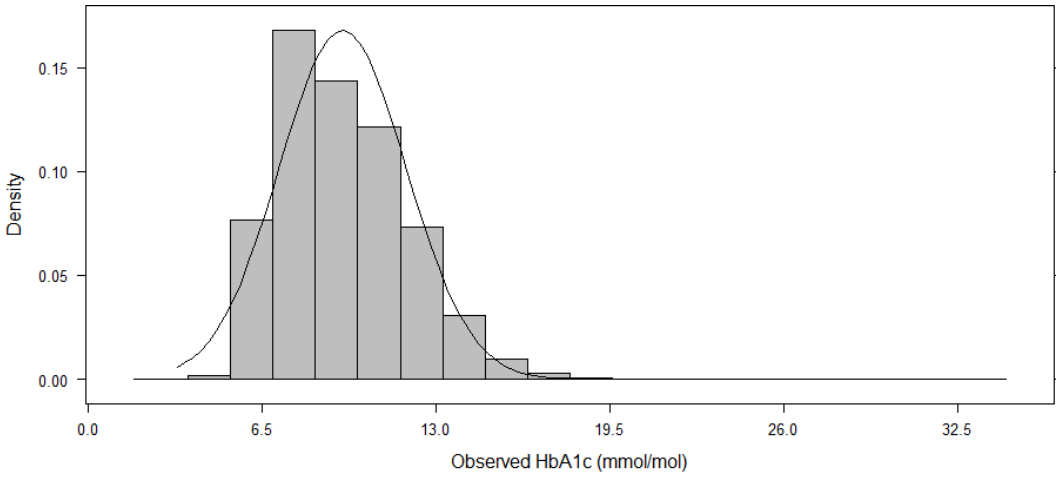

Figure 4.7: Observed longitudinal continuous HbA1c (mmol/mol) biomarker values of individuals in the routine diabetes Western Cape cohort.

Time to glycaemic controlled HbA1c

The time-to-event was the time to first glycaemic control, where glycaemic control was defined as an HbA1c < 6.5 mmol/mol based on the 2014 South African guidelines (South African National Department of Health, 2014). Duration of observed follow-up time (months) was calculated from enrolment to the last observed time due to achieving glycaemic control or censoring. Individuals were considered censored if they did not achieve glycaemic control at the end of follow-up or the last observed time if they were lost to follow-up.

Figure 4.8 shows observed longitudinal trajectories of the HbA1c (mmol/mol) biomarker for a random sample of T2DM individuals followed up from study entry until glycaemic control or last observed follow-up time. In this figure, individuals with glycaemic control during observed follow-up time are represented by dashed lines, while solid lines represent individuals who did not. The figure shows that the observed trajectories vary considerably.

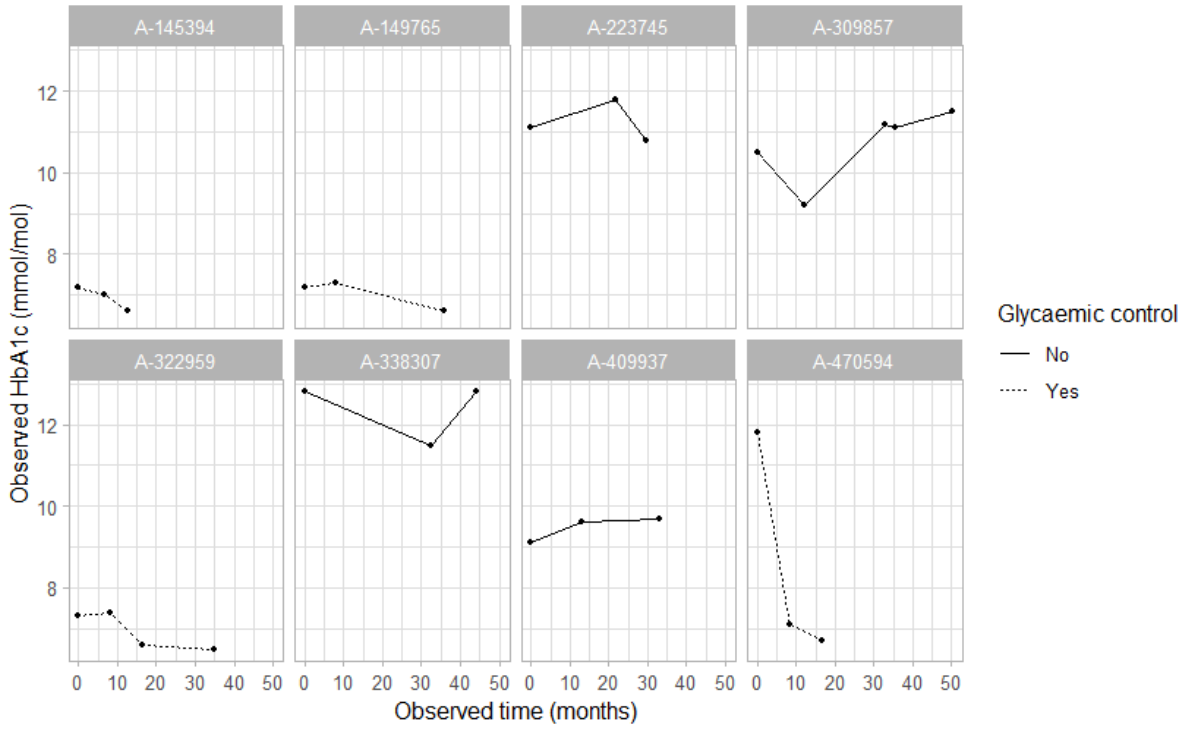

Figure 4.8: Observed longitudinal HbA1c (mmol/mol) measurements for eight randomly selected individuals followed-up from study entry time (not glycaemic controlled) until first glycaemic control or censoring time in the Western Cape, South Africa.

Demographic characteristics

Baseline was considered the first HbA1c measurement date for each individual in the cohort between 1 January 2016 and December 2021. Baseline characteristics of the Western Cape diabetes cohort are shown in Table 4.3. A large proportion (66%) of the individuals were females. Most individuals (56%) were in the 35-49 age group; 7.3% were in the 18-34 age group at enrolment. During the six years of follow-up, 2016 was when most (57%) individuals were followed up, and fewer (0.3%) individuals were followed up in 2021.

Overall, individuals in the cohort had a median of one HbA1c measurements (IQR: 1–3). The individuals' characteristics at the end of the follow-up of the overall development and

validation cohorts are shown in Table 4.4. By the end of the observed follow-up, 87 937 (83.7%) of the individuals were censored, and 17 074 (16.3%) achieved glycaemic control with a median follow-up time of 43 months (IQR: 28 - 57) in the overall cohort. For the development cohort, 12 811 (16.2%) individuals achieved glycaemic control, with a median follow-up time of 43 months (IQR: 28 - 57). In the validation cohort, 4263 (16.3%) individuals achieved glycaemic control with a median follow-up time of 43 months (IQR: 28 - 57).

Table 4.3: Individual characteristics at baseline in the historic routine diabetes Western Cape cohort.

Characteristic	**N = 129 609**[1]
Baseline HbA1c level (mmol/mol), n (%)	
< 6.5	24 598 (19)
6.5-8	31 768 (24.5)
> 8	73 243 (56.5)
Baseline Age groups, n (%)	
18-34	9512 (7.3)
35-59	72 893 (56.2)
60+	47 204 (36.4)
Sex, n (%)	
Male	44 027 (34)
Female	85 582 (66)
Cohort enrolment, n (%)	
2016	73 600 (56.8)
2017	31 502 (24.3)
2018	13 951 (10.8)
2019	7727 (6)
2020	2426 (1.9)
2021	403 (0.3)

[1]Statistics presented: N = total number of individuals; n = frequency

Table 4.4: Characteristics of individuals by the end of follow-up in the historic routine diabetes Western Cape overall, development and validation cohorts.

Characteristic	Overall cohort N = 105 011[1]	Development cohort N = 78 868[1]	Validation cohort N = 26 143[1]
HbA1c level, n (%)			
< 6.5	17 074 (16.3)	12 811 (16.2)	4263 (16.3)
6.5-8	20 473 (19.5)	15 428 (19.6)	5045 (19.3)
> 8	67 464 (64.2)	50 629 (64.2)	16 835 (64.4)
Baseline age groups, n (%)			
18-34	4989 (4.8)	3733 (4.7)	1256 (4.8)
35-59	50 817 (48.4)	38 066 (48.3)	12 751 (48.8)
60+	49 205 (46.9)	37 069 (47.0)	12 136 (46.4)
Sex, n (%)			
Male	35 117 (33.4)	26 326 (33.4)	8791 (33.6)
Female	69 894 (66.6)	52 542 (66.6)	17 352 (66.4)
Cohort year, n (%)			
2016	1175 (1.1)	896 (1.1)	279 (1.1)
2017	5149 (4.9)	3851 (4.9)	1298 (5)
2018	8466 (8.1)	6385 (8.1)	2081 (8)
2019	17 551 (16.7)	13 129 (16.6)	4422 (16.9)
2020	20 236 (19.3)	15 196 (19.3)	5040 (19.3)
2021	52 434 (49.9)	39 411 (50)	13 023 (49.8)
Follow-up time (months), Median (IQR)	43.1 (28.4, 56.8)	43.1 (28.3, 56.8)	43.2 (28.4, 56.9)

[1]Statistics presented: N = total number of individuals; IQR = interquartile range; n = frequency

The duration between the observed follow-up times and glycaemic-control patterns

The duration between the observed follow-up times for individuals who achieved glycaemic control in the overall cohort is shown in Figure 4.9. These individuals had a median duration of 11.5 months, and only 25% had less than 6.5 months between observed follow-up times.

The glycaemic control-free probability in the individuals was 0.985 (95% CI: 0.984, 0.986) within six months of observed follow-up time, 0.957 (95% CI: 0.956, 0.959) within 12 months and 0.903 (95% CI: 0.901-0.904) within 24 months from the overall routine diabetes cohort.

These glycaemic control-free probabilities are summarised in the Kaplan-Meier plot (Figure 4.10). A large proportion of the individuals did not achieve glycaemic control during the observed follow-up time. In Chapter 5, the hazard of achieving glycaemic control is estimated by applying statistical regression models for prognosis prediction to T2DM individuals.

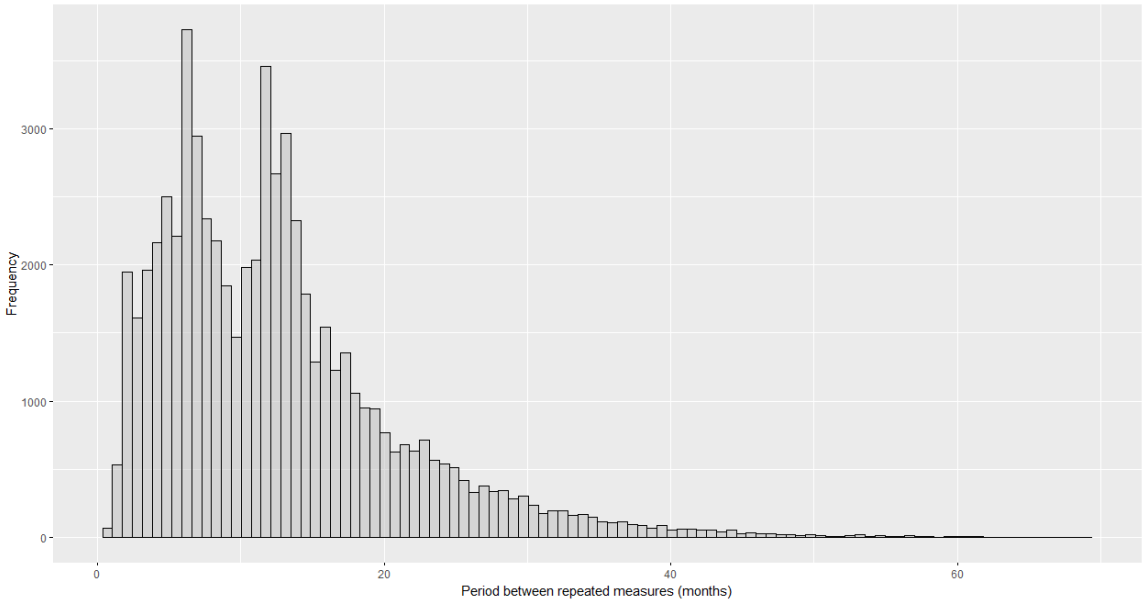

Figure 4.9: Duration (months) between observed follow-up time for individuals who achieved glycaemic control in the routine diabetes Western Cape overall cohort.

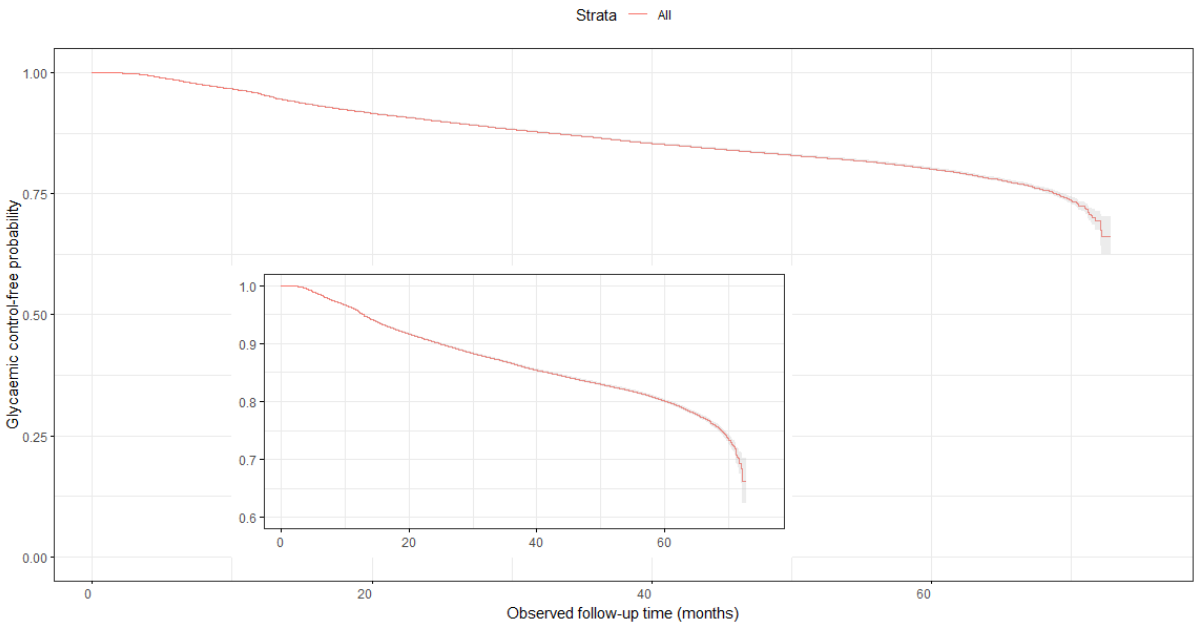

Figure 4.10: Kaplan-Meier estimate of the glycaemic control-free probability for T2DM individuals in the routine diabetes Western Cape overall cohort. The inserted figure represents a zoomed-in version of the original figure.

4.3 Ethical Approval

Existing NHLS laboratory data routinely collected during the clinical care of patients in health facilities from Western Cape, South Africa, were analysed in this thesis. Approval for analysis of anonymised NHLS HIV and diabetes data was granted by the Human Research Ethics Committee (HREC) of the University of Cape Town (Protocol No HREC 436-2020). See Appendix 12.

4.4 Discussion

The Western Cape routine HIV and diabetes datasets contained a longitudinal outcome, a time to event and demographic characteristics of patients. In the routine HIV data, the longitudinal outcome was the HIV viral load, and the time to event was the time to virologic failure. For the routine diabetes data, HbA1c and the time to glycaemic control were the longitudinal outcome and time-to-event, respectively. The longitudinal outcome and time-to-event data for each disease are modelled separately (Chapter 5) and jointly (Chapter 6 and Chapter 7) to predict probabilities of clinical outcomes (virologic failure and glycaemic control).

4.5 References

Mukonda, E., Hsiao, N.-Y., Vojnov, L., Myer, L., & Lesosky, M. (2020). Mixed-method estimation of population-level HIV viral suppression rate in the Western Cape, South Africa. *BMJ Global Health, 5*(8), e002522. doi:10.1136/bmjgh-2020-002522

National Health Laboratory Service. (2021a). Annual Performance Plan 2021–2022. Retrieved from https://www.nhls.ac.za/key-documents/strategic-documents/

National Health Laboratory Service. (2021b). NPP HIV Viral Load Unit. Retrieved from https://www.nhls.ac.za/priority-programmes/hiv-viral-load/

Odayar, J., Rusch, J., Dave, J. A., Van Der Westhuizen, D. J., Mukonda, E., Lesosky, M., & Myer, L. (2024). Transfers between health facilities of people living with diabetes attending primary health care services in the Western Cape Province of South Africa: A retrospective cohort study. *Trop Med Int Health*. doi:10.1111/tmi.13990

South African National Department of Health. (2014). Management of type 2 diabetes in adults at primary care level. Retrieved from https://knowledgehub.health.gov.za/elibrary/management-type-2-diabetes-adults-primary-care-level

South African National Department of Health. (2019a). 2019 ART Clinical Guidelines for the Management of HIV in Adults, Pregnancy, Adolescents, Children, Infants and Neonates. Retrieved from https://www.health.gov.za/hiv-and-aids/

South African National Department of Health. (2019b). Adult primary care, symptom-based integrated approach to the adult in primary care. Retrieved from https://knowledgehub.health.gov.za/elibrary/adult-primary-care-apc-guide-20192020-updated

Chapter 5 Application of Cox Models for Prediction

5.1 Introduction

In Chapter 2, regression models for analysis of longitudinal continuous biomarker and time-to-event data were introduced. Two datasets were briefly described in Chapter 4 to provide context and motivation for the regression models. These datasets are used to illustrate the application of regression models as practical tools to generate prediction of clinical outcomes. In this Chapter, a Cox proportional hazards model was used to predict probabilities of clinical outcomes in the two datasets. In Section 5.2, a Cox proportional hazards-based prediction model for individuals living with HIV in the Western Cape is developed and evaluated. Similarly, a Cox proportional hazards-based prediction model for T2DM individuals in the Western Cape is presented in Section 5.3. A summary of the findings and concluding remarks are given in Section 5.4.

5.2 Predicting Virologic Failure for Individuals Living with HIV

This section illustrates the application of a Cox proportional hazards model to the motivating routine HIV Western Cape data. The development of a cohort of the routine HIV Western Cape data was used for developing a prediction model, and the validation cohort was used for evaluating predictive performance. The development details and predictive performance of the Cox proportional hazards-based prediction model are given in sub-sections 5.2.1 and 5.2.2, respectively. A nomogram for prognosis predictions of virologic failure probabilities is provided in sub-section 5.2.3, and findings are summarised in sub-section 5.2.4.

5.2.1 Cox model for time-to-virologic failure

The Cox proportional hazards model described in Chapter 2 was used to model the hazard of virologic failure, adjusting for baseline continuous HIV viral load values, baseline age category (age category 30-39 as a reference category) and sex. The Cox proportional hazards model is given as:

$$h_i(t) = h_0(t)\exp(\gamma^*\text{age}_i + \gamma_5\text{male}_i + \varphi\log_{10}\text{vload}_i), \quad 5.1$$

$$\gamma^*\text{age}_i = \gamma_1\text{age}_{16-19} + \gamma_2\text{age}_{20-29} + \gamma_3\text{age}_{40-49} + \gamma_4\text{age}_{50+}.$$

where $h_i(t)$ represents the hazard of virologic failure for individual i at observed follow-up time t, $h_0(t)$ describes the unspecified baseline hazard, γ^* is the vector of regression coefficients for baseline age categories, γ_2 is the regression coefficient for the male dummy variable, and φ is the regression coefficient for baseline continuous HIV viral load values for the ith individual. The primary interest is in the regression coefficient φ, which quantifies the association between baseline continuous HIV viral load values and the hazard of virologic failure. The Cox proportional hazards model in Equation 5.1 relied on baseline continuous HIV viral load values.

A Cox proportional hazards model described in Chapter 2 was used to incorporate the last observed continuous HIV viral load values instead of baseline values. In this aternative Cox proportional hazards model, the interest lies in quantifying the association between the last observed continuous HIV viral load values at time t, $\log_{10}\text{vload}_i(t)$ and the hazard of virologic failure, adjusting for baseline age categories and sex. The Cox proportional hazards model is given as:

$$h_i(t) = h_0(t)\exp(\gamma_1\text{age}_{16-19} + \gamma_2\text{age}_{20-29} + \gamma_3\text{age}_{40-49} + \gamma_4\text{age}_{50+} + \gamma_5\text{male}_i + \varphi\log_{10}\text{vload}_i(t)). \quad 5.2$$

where φ quantifies the association between the last observed continuous HIV viral load value before virologic failure at time t and other model terms have the same specification as in Equation 5.2. The R code used to fit these Cox proportional hazards models is given in Appendix 4.

Overall, the Cox proportional hazards model with baseline HIV viral load values had virologic failure probabilities closer to the observed virologic failure probabilities (from a Kaplan-Meier) than the one with last observed HIV viral load values (Figure 5.1). As a result, the Cox proportional hazards model with baseline HIV viral load values was considered in this thesis for predictions.

Table 5.1 shows the hazard ratios (HR) and the 95% confidence intervals of the Cox proportional hazards model with baseline HIV viral load, adjusted for baseline age categories and sex applied to the routine HIV Western Cape development cohort. The results indicate that the baseline continuous HIV viral load values were strongly associated with the hazard of virologic failure (p-value <0.001). For an increase in baseline continuous HIV viral load value (log base 10), the hazard of virologic failure increased by 1.99-fold (95% CI: 1.96, 2.01). The predictive performance of this Cox proportional hazards model was evaluated in the routine HIV Western Cape validation cohort in subsection 5.2.2.

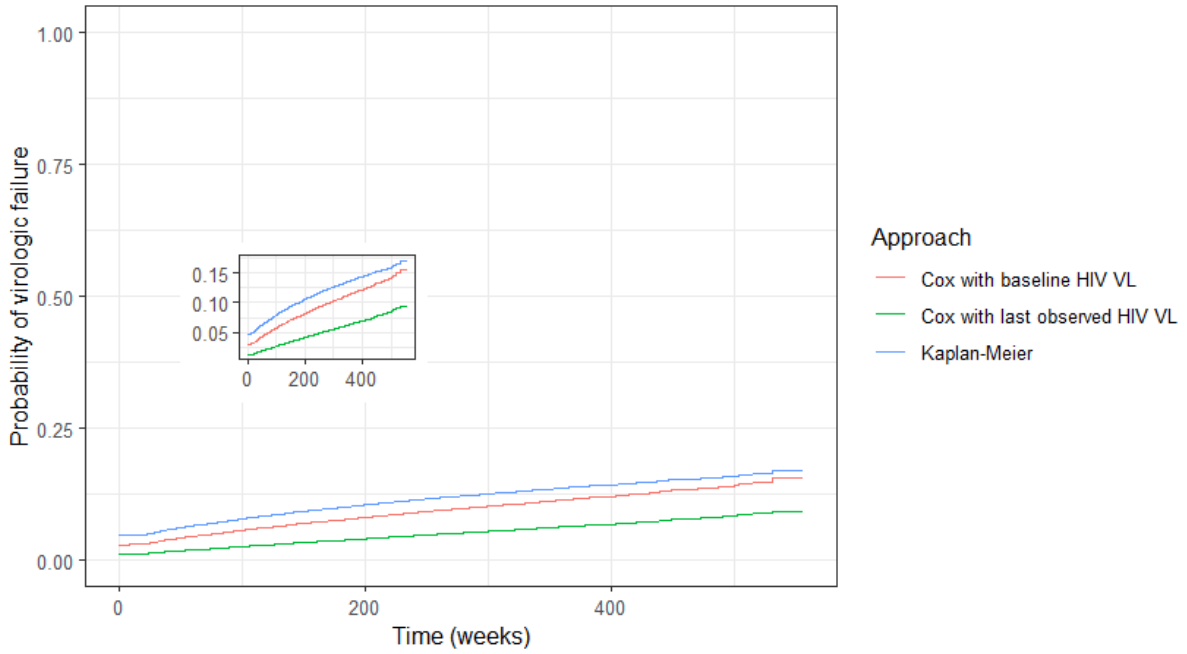

Figure 5.1: Comparison between a Cox proportional hazards model with baseline HIV viral load, Cox proportional hazards model with last observed HIV viral load and Kaplan-Meier (reference curve) on the routine HIV Western Cape development cohort. The inserted figure represents a zoomed-in version of the original figure.

Table 5.1: Parameter estimates under a Cox proportional hazards model with baseline HIV viral load applied to the routine HIV Western Cape development cohort.

	HR[1]	95% CI[1]	p-value
Baseline age group (years)			
16-19	1.22	1.05, 1.41	0.011
20-29	1.10	1.04, 1.15	<0.001
40-49	0.89	0.84, 0.95	<0.001
50+	0.80	0.72, 0.87	<0.001
Sex			
Male	0.98	0.94, 1.03	0.405
$\log_{10}$ HIV viral load (copies/mL) at baseline	1.99	1.96, 2.01	<0.001

[1]*HR = Hazard Ratio, CI = Confidence Interval*

5.2.2 Predictive performance

Time-dependent AUCs and Brier scores were used to evaluate the predictive performance of the Cox proportional hazards-based prediction model to the routine HIV Western Cape validation cohort. Figure 5.2 shows time-dependent AUCs from ROC curves among individuals living with HIV in the validation cohort at 12 and 24-month time windows. The Cox proportional hazards-based prediction model had good discrimination ability with 0.89 and 0.84 AUCs at 12 and 24-month time windows.

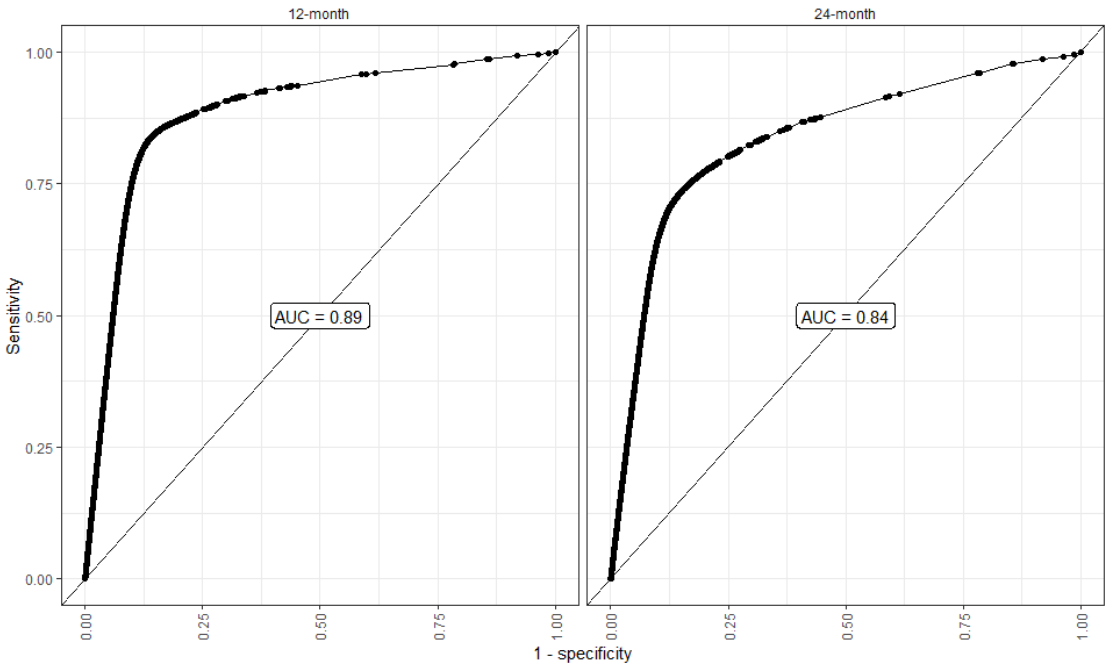

Figure 5.2: Time-dependent AUCs from ROC curves among individuals living with HIV in the routine HIV Western Cape validation cohort at 12 and 24 months.

On the other hand, time-dependent Brier scores and a plot of observed virologic failure probabilities were used to evaluate the model's prediction error and overall calibration. The model had low prediction error, evidenced by low Brier score values of 0.046 and 0.059 at 12 and 24-month time windows, respectively. The agreement between the Cox proportional hazard-based predictions and the observed virologic failure probabilities (Kaplan-Meier estimator) in the routine HIV Western Cape validation cohort was further evaluated in the calibration plot Figure 5.3. The plot suggests that the model had suboptimal calibration ability.

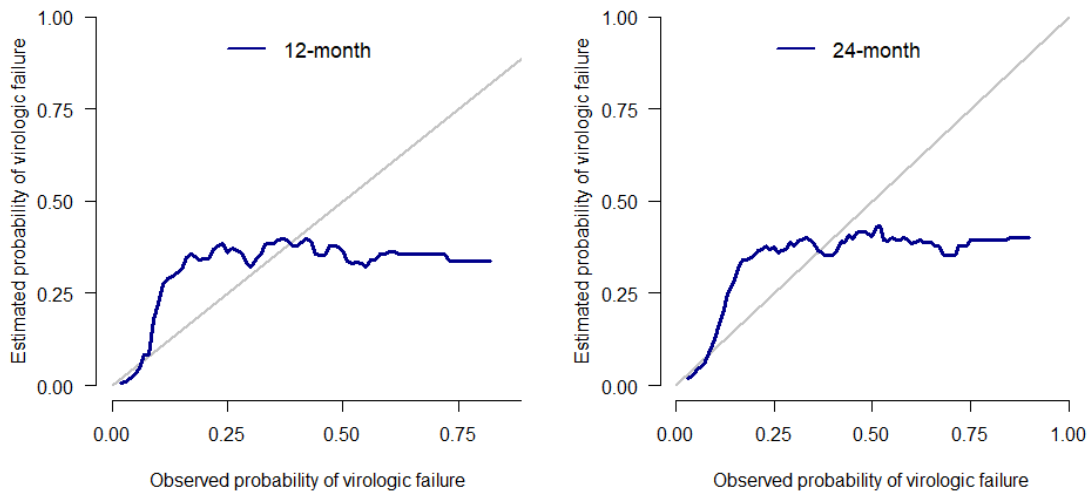

Figure 5.3: Calibration curve at 12 and 24-month prediction time windows for probabilities of virologic failure in the routine HIV Western Cape validation cohort. The diagonal line denotes the line of perfect calibration.

5.2.3 Predicting the probability of virologic failure

For illustration, the fitted Cox proportional hazard-based prediction model was used to predict the probability of virologic failure for individuals in the routine HIV Western Cape validation cohort. Based on the Cox proportional hazard model (Equation 5.1), a nomograph (Van Zee et al., 2003) from the R package RMS (Harrell, 2017) was constructed, and the routine HIV Western Cape validation cohort was used to visualise predictions of virologic failure in individuals living with HIV. Figure 5.4 shows a nomograph that visualises estimated probabilities of virologic failure at 12 and 24-month time windows based on an individual's baseline continuous log base 10 HIV viral load values, baseline age category and sex. Each covariate was assigned points between 0 and 100, and these points were added up to produce total points used to estimate the corresponding probability within 12 and 24-month time windows. For example, an elderly (50+-year old) female with an elevated baseline HIV viral load (≥1000 copies/mL), i.e., log base 10 HIV viral load value of 3 copies/mL, had a total of 35 points resulting in 12 and 24-month time windows estimated virologic failure probabilities of 0.063 (95% CI: 0.053, 0.072), and 0.08 (95% CI: 0.068, 0.092), respectively.

A web-based dynamic nomogram from the R package DynNom (Jalali, Alvarez-Iglesias, & Newell, 2017) provides a user-friendly online calculator to visualise nomograms. An example of the prediction of virologic failure probabilities of individuals in the routine HIV validation cohort is given in Figure 5.5. In Figure 5.5, predictions of virologic failure during observed follow-up for an elderly (50+-year old) female with baseline log base 10 HIV viral load value of 3 copies/mL are shown for illustration.

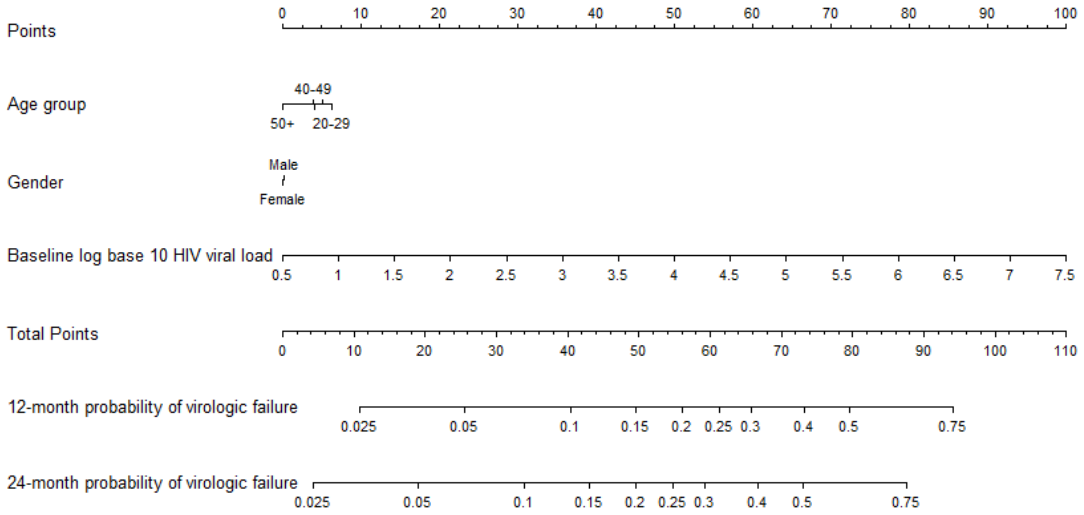

Figure 5.4: A nomogram for predicting the 12 and 24-month probabilities of virologic failure in the validation cohort. Each of the three variables (age group, baseline log base 10 HIV viral load and sex) is associated with points from 0 to 100. The total points, the sum of the points for each of all the three variables, are reported on the bottom scale of the nomogram.

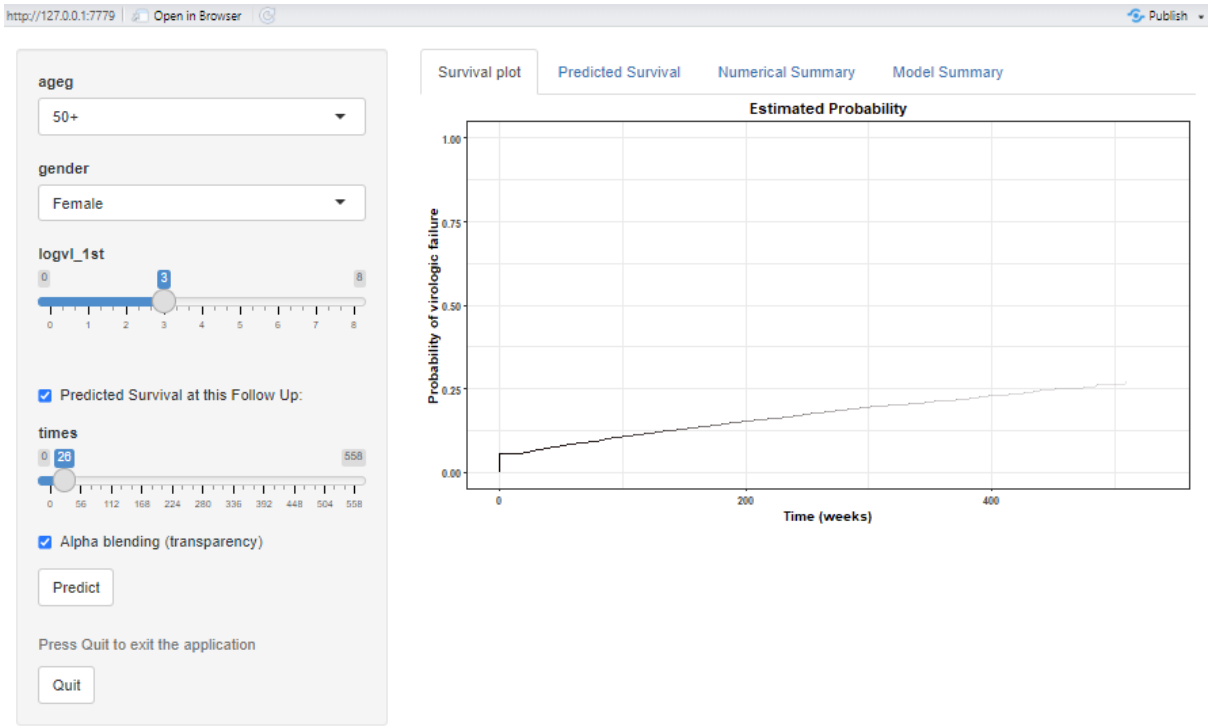

Figure 5.5: An interactive web-based calculator that calculates the probability of virologic failure for easier clinical application. It expects an input for each of the three variables (age group, baseline log base 10 HIV viral load and sex). It uses the information to produce a plot and summary for the probability of virologic failure with associated 95% confidence intervals at specified time windows. In the figure, ageg represents the baseline age group, and logvl_1st represents the baseline log base 10 HIV viral load.

5.2.4 Discussion

The findings in this study indicate that virologic failure was rare in the Western Cape (10.2% of individuals had virologic failure by the end of follow-up). This aligns with patterns observed in similar resource-limited settings (Boender et al., 2015; Mesic et al., 2021). Previous studies from the South African national treatment programme reported that 8-22% of individuals experienced virologic failure (Fox et al., 2012; Hermans et al., 2018; Rohr et al., 2016). Studies specific to the Western Cape province range from 2% to 14%, indicating potential provincial variations.

The implication of a lower proportion of virologic failure in the Western Cape highlights the effectiveness of ART in resource-limited settings. However, South Africa faces a multitude of health challenges, including communicable and non-communicable diseases, emphasising the

importance of addressing virologic failure to mitigate change of regimen associated costs and improve the quality of life for affected individuals (Agegnehu, Merid, & Yenit, 2020; SeyedAlinaghi et al., 2023). Despite abundant studies in resource-constrained settings that identified demographics and clinical variables associated with virologic failure, there remains a notable gap in developing prediction models for virologic failure, even in resource-rich settings.

Existing prediction models primarily focus on HIV-related outcomes such as immune reconstitution inflammatory syndrome (Han, X. et al., 2022), death (Hou et al., 2019; Jiang et al., 2022), CD4/CD8 ratio restoration (Li et al., 2022) and cutaneous T-Cell lymphoma (Yang, Gong, Huang, Sun, & Hu, 2022), with limited attention to virologic failure. These studies were predominantly based on retrospective cohort studies from China (Han, X. et al., 2022) and population-based cohorts across China and the United States. Only one previous study in South Africa developed a prognosis prediction model for virologic failure using routine electronic health records (Rohr et al., 2016). This thesis extends that work by predicting virologic failure, specifically in the Western Cape province, using baseline HIV viral load as a predictor instead of CD4 count and incorporating time-dependent AUCs and Brier scores for discrimination and calibration. Notably, a nomogram for the prognosis of virologic failure is introduced for practical clinical application.

In terms of study design and population, the prediction model in that study leveraged routine electronic health records data from a sizeable cohort of 72 181 individuals spanning healthcare facilities in Gauteng and Mpumalanga provinces from 2004 to 2013. The median follow-up duration in the routine HIV data in this thesis was shorter compared to the previous study (11.8 vs 21.5 months).

The previous study predicted a five-year probability of individuals on ART experiencing virologic failure in Gauteng and Mpumalanga provinces. Given the shorter median follow-up duration of individuals in the HIV routine data, the focus of this thesis was on predicting virologic failure within a 1-2 year time window for individuals on ART and living with HIV in the Western Cape province.

The prediction model in this thesis demonstrated better discrimination ability than the previous study. Time-dependent AUCs implied that individuals at higher risk of virologic failure were correctly identified with higher probabilities of 0.89 and 0.84 at 12 and 24-month time windows than 0.6 at a five-year time window. Lower time-dependent Brier scores suggested a match

between predicted and observed virologic failure patterns, consistent with the previous study. Furthermore, a nomogram in this study showed that, on average, one to two-year virologic failure probabilities were low, consistent with the five-year predictions of virologic failure probabilities in the previous study. The consistency in the findings indicates that, on average, virologic failure is low for shorter and longer time windows of prediction.

Consistent with previous research in South Africa, the findings revealed that older age at baseline was associated with decreased hazard of virologic failure (Fox et al., 2012; Hermans et al., 2018; Kehoe et al., 2020). Additionally, higher baseline HIV viral load values were associated with an increased hazard of virologic failure (Cevik, Orkin, & Sax, 2020; Fox et al., 2012; SeyedAlinaghi et al., 2023). This finding highlights the importance of routine HIV viral load monitoring to assess progression in individuals living with HIV. Chapter 6 extends the Cox proportional hazards-based prediction model for virologic failure to incorporate repeated HIV viral load measurements, recognising their dynamic nature in predicting virologic failure.

There are several limitations of the current prediction model applied to HIV routine data. First, the routine electronic health HIV data from the Western Cape lacked crucial demographics and clinical variables, such as adherence behaviour (Agegnehu, Techane, Mersha, & Atalell, 2022; Cevik et al., 2020; SeyedAlinaghi et al., 2023), ART regimen (Rohr et al., 2016), and co-infection (Agegnehu et al., 2022), known to influence virologic failure. The lack of these variables may have contributed to suboptimal agreement between observed and predicted virologic failure probabilities. Secondly, an independent HIV cohort was not available to externally validate the prediction model, restricting the generalisability of the prediction model to other resource-limited settings. The findings should be interpreted cautiously when extended to individuals in other provinces in South Africa because the prediction model was specific to populations in the Western Cape. Finally, due to the scarcity of prediction models for virologic failure in the literature, particularly in resource-limited settings, only one study was available for direct predictive performance comparison, limiting comprehensive assessment.

The developed prediction model has several strengths. A major strength was the large sample size covering most of the Western Cape population and a substantial follow-up period. Another strength is that demographics and biomarkers used in the prediction model are readily accessible from routine electronic health records. This enhances the model's potential applicability to other resource-limited healthcare settings. The development of a user-friendly nomograph was another strength as it offers clinicians a practical tool for predicting

individualised probabilities of virologic failure among individuals living with HIV in the Western Cape, potentially enabling personalised therapeutic interventions.

5.3 Predicting Glycaemic Control for T2DM Patients

In this Section, a Cox proportional hazards model was applied to the motivating routine diabetes Western Cape data. The model was developed and evaluated in the development and validation cohorts of the routine diabetes Western Cape data. Sub-sections 5.3.1 and 5.3.2 provide details on the development and evaluation of the predictive performance of the Cox proportional hazards-based prediction model. A nomogram for prognosis prediction of glycaemic control probabilities is provided in sub-section 5.3.3. The findings are summarised in sub-section 5.3.4.

5.3.1 Cox model for time to glycaemic control

In this sub-section, a Cox proportional hazards model was applied to the routine diabetes Western Cape development cohort, adjusted for baseline continuous HbA1c values, age categories (age category 35-59 as reference category) and sex. The model is given as:

$$h_i(t) = h_0(t)\exp(\gamma^*\text{age}_i + \gamma_3\text{male}_i + \varphi A1c_i), \qquad 5.3$$
$$\gamma^*\text{age}_i = \gamma_1\text{age}_{35-59} + \gamma_2\text{age}_{60+}.$$

where $h_i(t)$ denotes the hazard of glycaemic control for individual i at observed follow-up time t, $h_0(t)$ represents the unspecified baseline hazard, γ^* is the vector of regression coefficients for baseline age categories, γ_2 is the regression coefficient for the male dummy variable, and φ is the regression coefficient for baseline continuous HbA1c values for the ith individual. This regression coefficient (φ) quantifies the association between the hazard of glycaemic control and baseline continuous HbA1c values.

A Cox proportional hazards model was alternatively used to quantify the association between the last observed continuous HbA1c values at time t, $\varphi A1c_i(t)$ and the hazard of glycaemic control, adjusting for baseline age and sex. The Cox proportional hazards model is given as:

$$h_i(t) = h_0(t)\exp(\gamma_1 \text{age}_{35-59} + \gamma_2 \text{age}_{60+} + \gamma_3 \text{male}_i + \varphi A1c_i(t)). \quad 5.4$$

where φ quantifies the association between the last observed continuous HbA1c value before glycaemic control HbA1c at time t and other model terms are like the specification in Equation 5.3. The R code for fitting the Cox proportional hazards models is provided in Appendix 4.

Figure 5.6 compares the glycaemic control probabilities between the Kaplan-Meier estimator and Cox models. In this figure, predictions for both Cox models were poor. However, the Cox proportional hazards model with baseline HbA1c had glycaemic control probabilities that were more aligned with the observed glycaemic control probabilities (Kaplan-Meier estimator) compared to the Cox proportional hazards model with last observed HbA1c values. As a result, the Cox proportional hazards model with baseline HbA1c values was used for predictions in this thesis.

Parameter estimates under a Cox proportional hazards model with baseline HbA1c values applied to the routine diabetes Western Cape development cohort are shown in Table 5.2. The findings show that, for a unit increase in baseline continuous HbA1c value, the hazard of glycaemic control decreased by 0.777-fold (95% CI: 0.770, 0.785). In sub-section 5.3.2, the Cox proportional hazard-based prediction model was evaluated regarding predictive performance in the routine diabetes Western Cape validation cohort.

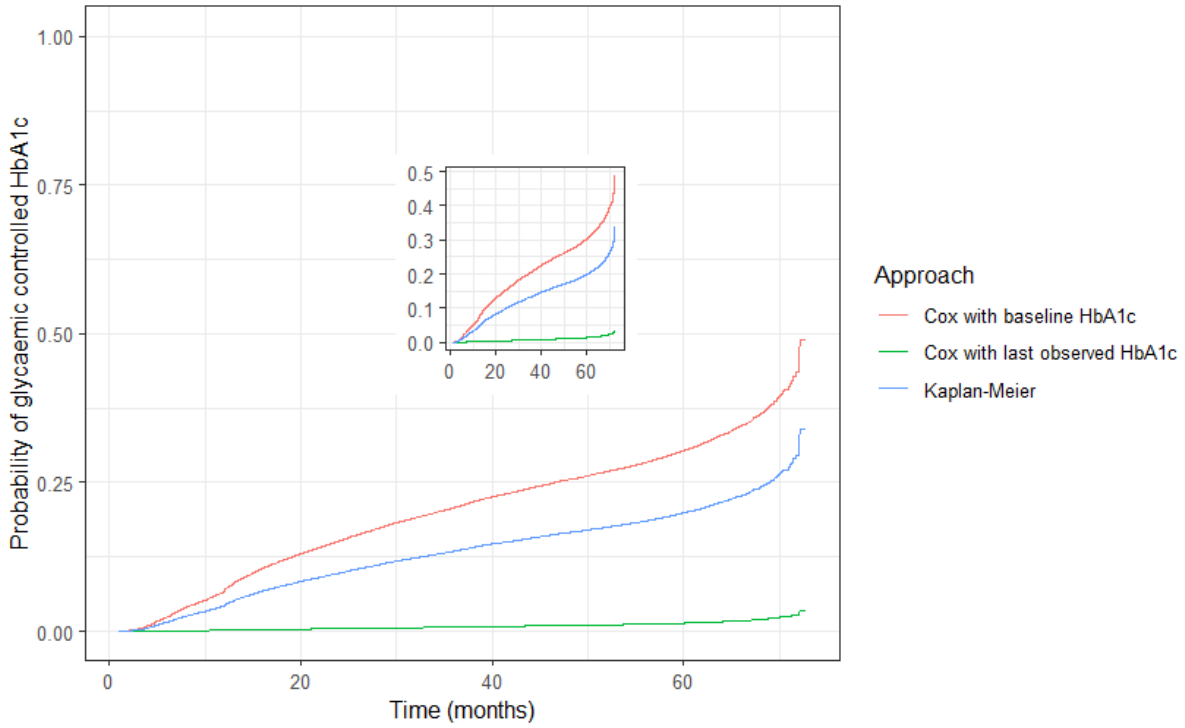

Figure 5.6: Comparison between a Cox proportional hazards model with baseline HbA1c, Cox proportional hazards model with last observed HbA1c and Kaplan-Meier (reference curve) on the routine diabetes Western Cape development cohort. The inserted figure represents a zoomed-in version of the original figure.

Table 5.2: Parameter estimates under a Cox proportional hazards model with baseline HbA1c applied to the routine diabetes Western Cape development cohort.

	HR[1]	**95% CI[1]**	**p-value**
Baseline age group (years)			
18-34			
35-59	0.432	0.407, 0.459	<0.001
60+	0.546	0.513, 0.581	<0.001
Sex			
Male	1.202	1.160, 1.246	<0.001
HbA1c (mmol/mol) at baseline	0.777	0.770, 0.785	<0.001

[1]HR = Hazard Ratio, CI = Confidence Interval

5.3.2 Predictive performance

The predictive performance of the Cox proportional hazards-based prediction model was evaluated in terms of time-dependent AUCs and Brier scores in the routine diabetes Western Cape validation cohort. The time-dependent AUCs from ROC curves at 12 and 24 months are shown in Figure 5.7. The Cox proportional hazards-based prediction model had suboptimal discrimination (0.63 at 12 months and 0.67 at 24 months) at these times. The prediction error and overall calibration were evaluated using Brier scores and a plot of observed and predicted glycaemic control probabilities. The Cox proportional hazards-based prediction model had low prediction error with Brier scores of 0.042 and 0.088 at 12 and 24 months, respectively. Figure 5.8 suggests the model had moderate calibration, evidenced by low agreement between the Cox proportional hazards-based predictions and the observed glycaemic control probabilities.

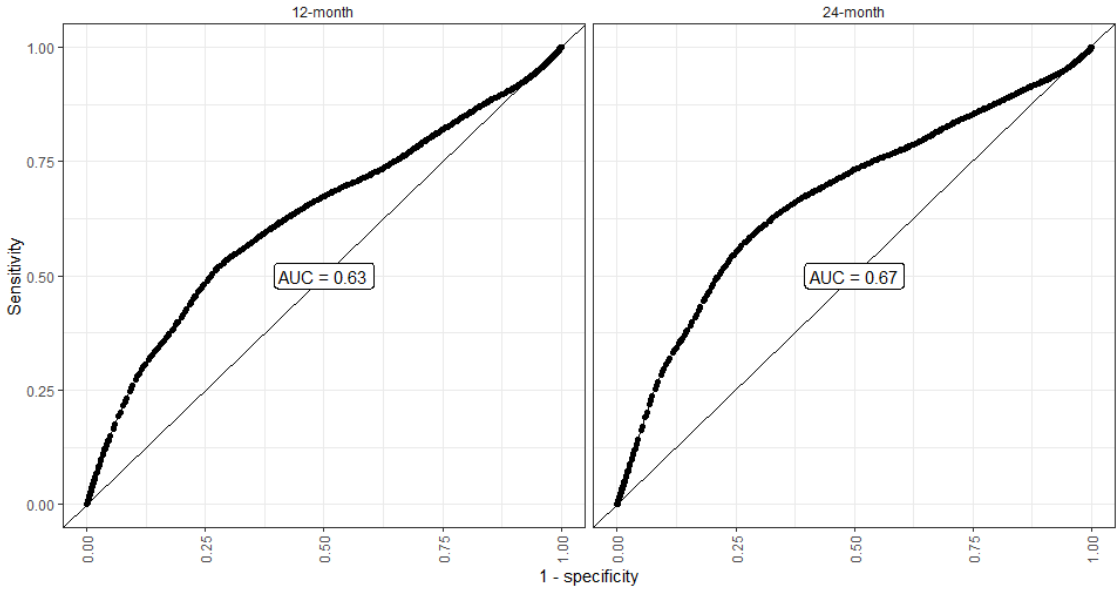

Figure 5.7: Time-dependent AUCs from ROC curves among T2DM individuals in the routine Western Cape validation cohort at 12 and 24 months.

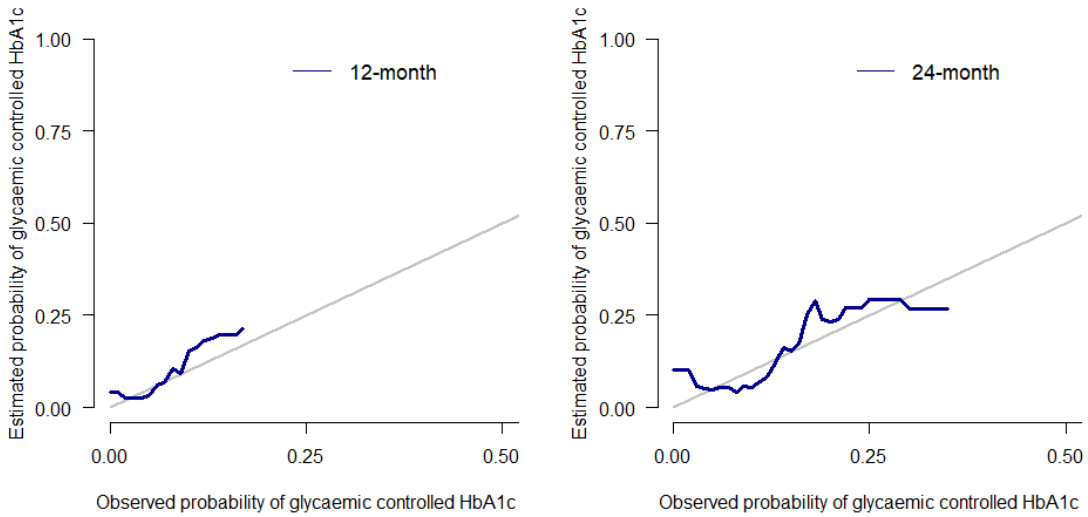

Figure 5.8: Calibration curve at 12 and 24-month prediction windows for probabilities of glycaemic control in the routine Western Cape validation cohort. The diagonal line denotes the line of perfect calibration.

5.3.3 Prediction of the probability of glycaemic control

This Cox proportional hazards-based model from Equation 5.3 predicted glycaemic control for individuals in the routine diabetes Western Cape validation cohort. Specifically, a nomogram was used to visualise predictions of glycaemic control in T2DM individuals in the Western Cape. The monogram is given in Figure 5.9. It visualises estimated probabilities of glycaemic control at 12 and 24 months, based on an individual's baseline continuous HbA1c values, age group and sex. For example, an elderly (60+) female with a baseline continuous HbA1c value of 7 mmol/mol had a total of 102 points resulting in 12 and 24-month estimated glycaemic control probabilities of 0.073 (95% CI: 0.068, 0.079) and 0.162 (95% CI: 0.153, 0.171), respectively.

A web-based dynamic nomogram providing a user-friendly online calculator was used to visualise probabilities of glycaemic control. For example, Figure 5.10 shows a dynamic nomogram for predicting glycaemic control probabilities for an elderly (60+) female with a baseline HbA1c value of 7 mmol/mol in the routine diabetes Western Cape validation cohort.

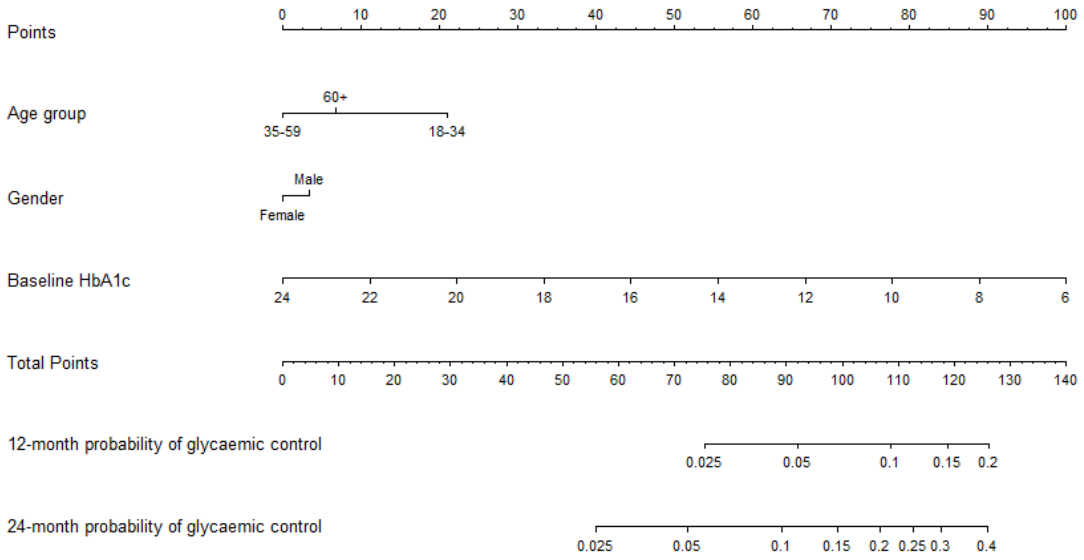

Figure 5.9: A nomogram for predicting the 12 and 24-month probabilities of glycaemic control using the Western Cape routine T2DM validation cohort. Each of the three variables (age group, baseline HbA1c and sex) is associated with points from 0 to 100. The total points (the sum of the points for each of the three variables) are reported on the bottom scale of the nomogram.

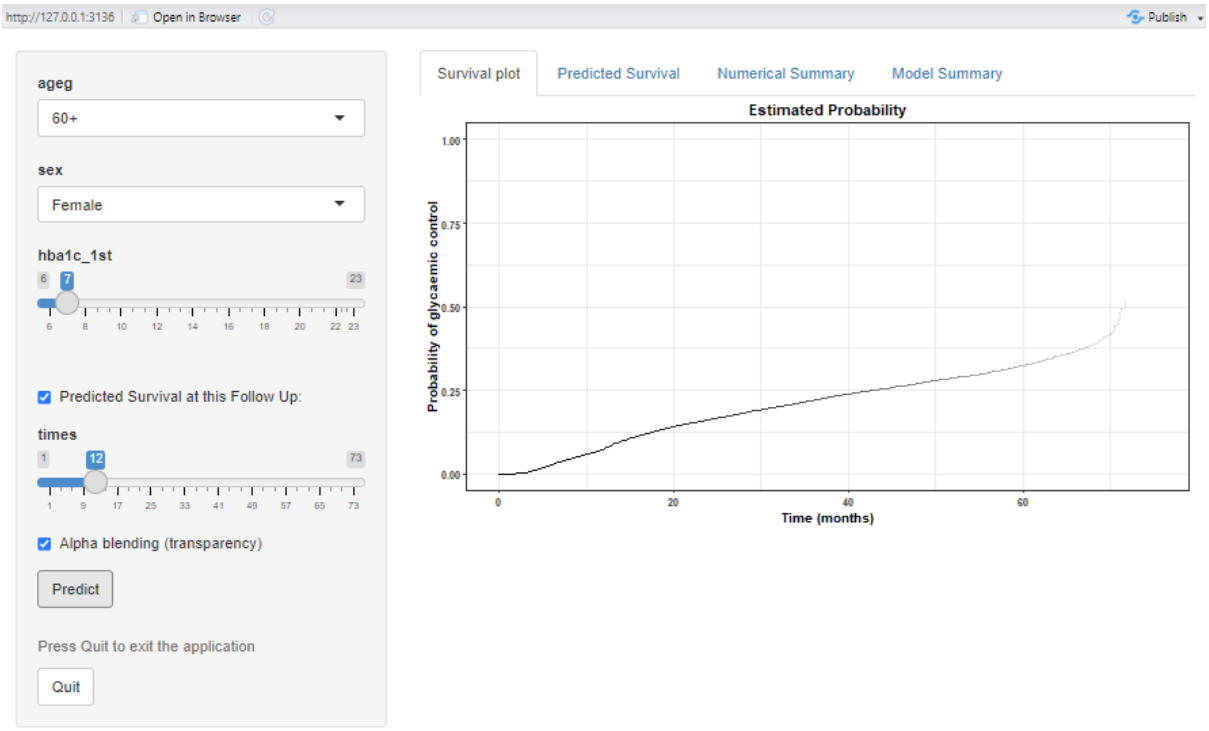

Figure 5.10: An interactive web-based calculator that calculates the probability of glycaemic control for easier clinical application. It expects an input for each of the three variables (age group, baseline HbA1c and sex). It uses the information to produce a plot and summary for the probability of glycaemic control with associated 95% confidence intervals at specified times. In the figure, ageg represents the baseline age group, and hba1c_1st represents the baseline HbA1c.

5.3.4 Discussion

Glycaemic control is essential for better health outcomes. The findings in the present study showed that 16.2% of the individuals in the Western Cape province achieved glycaemic control. This magnitude of glycaemic control is lower than previously reported (23%) in a systematic review of cross-sectional studies that analysed the performance of health systems in managing diabetes across resource-limited settings (Manne-Goehler et al., 2019). The findings of this systematic review are similar (23-29%) to findings in other cross-sectional studies in South African populations (Folb et al., 2015; Mhlaba, Mpanya, & Tsabedze, 2023; Ngassa Piotie, Webb, & Rheeder, 2021). The lower proportion of glycaemic control in this current study than in other studies in resource-limited settings could be that the cross-sectional nature of the studies in the systematic review tended to overestimate the proportion of

glycaemic control (Brennan et al., 2023). In contrast, the magnitude of glycaemic control in this current study is higher (8.7%) than in a study that used cohort routine data from EHR in South Africa (Brennan et al., 2023). The higher proportion is most likely due to the more extended follow-up period of individuals in this current study (2015-2021) compared to the other study (2012-2015).

The study highlights the significant challenge of achieving glycaemic control among individuals with T2DM, particularly in resource-limited settings. Detecting T2DM individuals less likely to achieve glycaemic control early on is crucial to reducing adverse health outcomes. While previous research has identified demographics and clinical variables associated with glycaemic control (Bahizi et al., 2022; Brennan et al., 2023; Demoz et al., 2019; Leulseged & Ayele, 2019; Manne-Goehler et al., 2019; Mhlaba et al., 2023), the development of prediction models for glycaemic control specifically tailored to T2DM individuals in resource-limited settings remain scarce, even in other income settings.

Existing studies from other income settings predominantly focused on developing prediction models for various T2DM-related complications such as stroke (Kim et al., 2020), severe hypoglycaemia (Han, K. et al., 2018), and end-stage renal disease (Dong et al., 2021).

These prediction models were developed using population-based cohorts from Chinese and Korean administrative databases. The sample size in these studies varied, ranging from 141 516 to 1 676 885 individuals, with follow-up conducted between 2009 and 2017. Moreover, there was a large variability in the follow-up period of individuals in the studies. The duration of follow-up varied notably, from one year (Han, K. et al., 2018) to a median follow-up period of 9.8 years (Dong et al., 2021).

The prediction time windows varied across the studies. Han et al. (2018) focused on a short-term prediction window of one year, while Kim et al. (2020) presented predictions for a mid-term window of five years. In contrast, Dong et al. (2021) offered a long-term perspective with a prediction window of ten years. The prediction model in this study focused on a one to two-year probability of glycaemic control in T2DM individuals in the Western Cape.

These studies exhibit varying predictive performances, as evidenced by reported C-statistics. For instance, Kim et al. (2020) used a nationwide Korean database to develop and internally validate a Cox-based prediction model that yielded a C-statistic of 0.703 for predicting stroke in middle-aged individuals with T2DM. Similarly, a population-based cohort from a Hong Kong hospital electronic health records database was used to develop and internally validate a

Cox-based prediction model with a C-statistic of 0.889 for predicting a ten-year risk of end-stage renal disease (Dong et al., 2021).

However, these prediction models do not specifically address the prediction of glycaemic control and may not be directly applicable to resource-limited settings. The present study addresses this gap by constructing a tailored prognosis prediction model for glycaemic control among South African individuals with T2DM. Leveraging provincial-level routine diabetes data from the Western Cape, the prediction model was internally validated and demonstrated moderate discrimination (time-dependent AUCs of 0.63 and 0.67 at 12 and 24 months) and good calibration. The prediction model was used to construct a nomogram to predict the one and two-year probability of glycaemic control for T2DM individuals in South Africa, especially those in Western Cape province.

To the best of the available literature, the present study represents the first prediction model for glycaemic control in a South African T2DM population. Baseline age groups, sex, and continuous HbA1c biomarker values were included in the prediction model.

Consistent with previous research, the findings revealed that older individuals (>34 years) were associated with lower glycaemic control. An Ethiopian study also found that achieving glycaemic control decreases with age (Leulseged & Ayele, 2019). In contrast, a previous study in South Africa reported that older people (≥60 years) were associated with higher glycaemic control (Brennan et al., 2023). The age group differences in achieving glycaemic control underscore the role of lifestyle factors such as medical adherence and diet (Chetty & Pillay, 2022). Males were associated with higher glycaemic control. Studies in resource-limited settings suggested that poorer glycaemic control can be attributed to body mass index, glucose metabolism and hormonal differences between males and females (Demoz et al., 2019; Kautzky-Willer, Leutner, & Harreiter, 2023). In line with a study reported in Rwanda, the current study found higher levels of baseline continuous HbA1c biomarkers were associated with lower glycaemic control (Bahizi et al., 2022). This is likely because repeated HbA1c biomarker measurements and lifestyle changes over time are required to achieve glycaemic control (Ceriello, 2010). As a result, Chapter 6 extends the prediction model for glycaemic control to include repeated HbA1c measurements.

The current prediction model applied to routine diabetes data has several limitations. First, the score of the prediction model developed in the current study was limited by the availability of demographics and clinical variables in the routine electronic diabetes data from the Western

Cape NHLS database. This resulted in the exclusion of essential factors such as socioeconomic status, lifestyle changes, blood pressure, body mass index, diabetes neuropathy, hypertension, and dyslipidaemia, known to influence glycaemic control (Ceriello, 2010; Demoz et al., 2019; Leulseged & Ayele, 2019; Manne-Goehler et al., 2019; Wei, 2019). Additionally, this may have compromised the discrimination ability of the prediction model. Secondly, while the study used routine diabetes data from South Africa's national database, the generalisability of findings to other healthcare systems, particularly those in resource-limited settings, remains uncertain and necessitates further validation using external cohorts from populations in these settings. Third, the study relied solely on routine data from the Western Cape, and, as a result, the findings should be interpreted with caution when extrapolated to individuals from other provinces in South Africa. Finally, due to the scarcity of comparable literature, direct predictive performance comparison with existing prediction models focusing on glycaemic control was not feasible. Nonetheless, internal validation demonstrated good calibration of the developed prediction model in the present study.

Despite the limitations, routine data covering most of South Africa's Western Cape population enhanced the robustness and representativeness of the findings from the developed prediction model. Including demographics and clinical variables readily available in routine data from healthcare systems enhances the applicability of the prediction model to other resource-limited settings. Furthermore, developing a user-friendly nomogram for individualised prognosis of achieving glycaemic control enables easy accessibility and implementation of the prediction model.

In conclusion, this section underlines the need for tailored prediction models for individualised glycaemic control prediction to ensure efficient use of resources in resource-limited healthcare settings. While existing research predominantly focuses on prediction models for T2DM-related complications, the present study fills a significant gap by developing a prediction model for glycaemic control that can be applied in South African populations. The prediction model can be extended to other resource-limited healthcare settings because the demographics and clinical variables used in the model are routinely collected in clinical care. The prediction model in the present study demonstrated moderate discrimination and good calibration, offering valuable insights for guiding clinical decision-making in South Africa's clinical care and similar resource-limited settings. The predictive performance of this model will be compared with that of an extended prediction model for glycaemic control in Chapter 6.

5.4 Conclusion

The development and application of prediction models for virologic failure in HIV and glycaemic control in T2DM is a significant advance in the literature, particularly in resource-limited settings like South Africa. Despite the pressing need, such models are scarce, even in resource-rich settings. In South Africa, the prediction model for virologic failure in this thesis is only the second of its kind, using routine data from electronic health records. Similarly, the prediction model for glycaemic control is the first explicitly tailored to South African populations using routine data from electronic health records.

Virologic failure and glycaemic control are distinct clinical outcomes. Virologic failure signifies a health concern, while glycaemic control represents a favourable clinical outcome. This contrast underscores the versatility and importance of prediction models in this thesis for predicting a range of clinical outcomes. The limitations of these models lie in assuming baseline values for a continuous biomarker (HIV viral load and HbA1c) for a given individual. In practice, the repeated biomarker values over time provide more insights into the disease progression for a given individual. Dynamic prediction models are a direct solution to enable the inclusion of repeated biomarker values over time, allowing for predictions that use all available biomarker measurements instead of baseline measurements for a given individual. These dynamic prediction models are applied in Chapter 6 for the individuals living with HIV and T2DM individuals. The strengths of these models lie in using readily available demographic and clinical variables from routine data in electronic health records, enhancing their potential applicability in similar resource-limited settings.

In conclusion, developing and applying prediction models for virologic failure and glycaemic control in South Africa presents an opportunity to improve healthcare outcomes in resource-limited settings. These prediction models offer valuable insights into routine monitoring of chronic diseases. However, further validation of these prediction models is needed to improve their predictive performance and generalisability across limited-resource settings.

5.5 References

Agegnehu, C. D., Merid, M. W., & Yenit, M. K. (2020). Incidence and predictors of virological failure among adult HIV patients on first-line antiretroviral therapy in Amhara regional referral hospitals; Ethiopia: a retrospective follow-up study. *BMC Infect Dis, 20*(1), 460. doi:10.1186/s12879-020-05177-2

Agegnehu, C. D., Techane, M. A., Mersha, A. T., & Atalell, K. A. (2022). Burden and Associated Factors of Virological Failure Among People Living with HIV in Sub-Saharan Africa: A Systematic Review and Meta-Analysis. *AIDS Behav, 26*(10), 3327-3336. doi:10.1007/s10461-022-03610-y

Bahizi, S., Mugeni, R., Banhart, D., Mukankuranga, C., Makiriro, G., Kirk, C., . . . Cubaka, V. K. (2022). Glycemic control among patients with type 2 diabetes in a low resource setting in Rwanda: a prospective cohort study. *Pan Afr Med J, 43*, 74. doi:10.11604/pamj.2022.43.74.35639

Boender, T. S., Sigaloff, K. C., McMahon, J. H., Kiertiburanakul, S., Jordan, M. R., Barcarolo, J., . . . Bertagnolio, S. (2015). Long-term Virological Outcomes of First-Line Antiretroviral Therapy for HIV-1 in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. *Clin Infect Dis, 61*(9), 1453-1461. doi:10.1093/cid/civ556

Brennan, A. T., Lauren, E., Bor, J., George, J. A., Chetty, K., Mlisana, K., . . . Crowther, N. J. (2023). Gaps in the type 2 diabetes care cascade: a national perspective using South Africa's National Health Laboratory Service (NHLS) database. *BMC Health Serv Res, 23*(1), 1452. doi:10.1186/s12913-023-10318-9

Ceriello, A. (2010). The glucose triad and its role in comprehensive glycaemic control: current status, future management. *Int J Clin Pract, 64*(12), 1705-1711. doi:10.1111/j.1742-1241.2010.02517.x

Cevik, M., Orkin, C., & Sax, P. E. (2020). Emergent Resistance to Dolutegravir Among INSTI-Naïve Patients on First-line or Second-line Antiretroviral Therapy: A Review of Published Cases. *Open Forum Infectious Diseases, 7*(6). doi:10.1093/ofid/ofaa202

Chetty, R., & Pillay, S. (2022). The relationship between age and glycaemic control in patients living with diabetes mellitus in the context of HIV infection: a scoping review. *Journal of Endocrinology, Metabolism and Diabetes of South Africa, 27*(1), 1-7.

Demoz, G. T., Gebremariam, A., Yifter, H., Alebachew, M., Niriayo, Y. L., Gebreslassie, G., . . . Shibeshi, W. (2019). Predictors of poor glycemic control among patients with type 2

diabetes on follow-up care at a tertiary healthcare setting in Ethiopia. *BMC Res Notes, 12*(1), 207. doi:10.1186/s13104-019-4248-6

Dong, W., Wan, E. Y. F., Fong, D. Y. T., Kwok, R. L. P., Chao, D. V. K., Tan, K. C. B., . . . Lam, C. L. K. (2021). Prediction models and nomograms for 10-year risk of end-stage renal disease in Chinese type 2 diabetes mellitus patients in primary care. *Diabetes Obes Metab, 23*(4), 897-909. doi:10.1111/dom.14292

Folb, N., Timmerman, V., Levitt, N. S., Steyn, K., Bachmann, M. O., Lund, C., . . . Fairall, L. R. (2015). Multimorbidity, control and treatment of noncommunicable diseases among primary healthcare attenders in the Western Cape, South Africa. *S Afr Med J, 105*(8), 642-647. doi:10.7196/samjnew.8794

Fox, M. P., Cutsem, G. V., Giddy, J., Maskew, M., Keiser, O., Prozesky, H., . . . Boulle, A. (2012). Rates and predictors of failure of first-line antiretroviral therapy and switch to second-line ART in South Africa. *J Acquir Immune Defic Syndr, 60*(4), 428-437. doi:10.1097/QAI.0b013e3182557785

Han, K., Yun, J. S., Park, Y. M., Ahn, Y. B., Cho, J. H., Cha, S. A., & Ko, S. H. (2018). Development and validation of a risk prediction model for severe hypoglycemia in adult patients with type 2 diabetes: a nationwide population-based cohort study. *Clin Epidemiol, 10*, 1545-1559. doi:10.2147/clep.S169835

Han, X., Liu, H., Wang, Y., Wang, P., Wang, X., Yi, Y., & Li, X. (2022). A nomogram for predicting paradoxical immune reconstitution inflammatory syndrome associated with cryptococcal meningitis among HIV-infected individuals in China. *AIDS Res Ther, 19*(1), 20. doi:10.1186/s12981-022-00444-5

Harrell, F. E. (2017). rms: regression modeling strategies. R package version 5.1-2. *Dept. Biostatist., Vanderbilt Univ., Nashville, TN, USA*.

Hermans, L. E., Moorhouse, M., Carmona, S., Grobbee, D. E., Hofstra, L. M., Richman, D. D., . . . Wensing, A. M. J. (2018). Effect of HIV-1 low-level viraemia during antiretroviral therapy on treatment outcomes in WHO-guided South African treatment programmes: a multicentre cohort study. *Lancet Infect Dis, 18*(2), 188-197. doi:10.1016/s1473-3099(17)30681-3

Hou, X., Wang, D., Zuo, J., Li, J., Wang, T., Guo, C., . . . Mao, G. (2019). Development and validation of a prognostic nomogram for HIV/AIDS patients who underwent antiretroviral therapy: Data from a China population-based cohort. *EBioMedicine, 48*, 414-424. doi:10.1016/j.ebiom.2019.09.031

Jalali, D. A., Alvarez-Iglesias, A., & Newell, J. (2017). DynNom: dynamic nomograms for linear, generalized linear and proportional hazard models. R package version 4.1. 1. In.

Jiang, F., Xu, Y., Liu, L., Wang, K., Wang, L., Fu, G., . . . Peng, Z. (2022). Construction and validation of a prognostic nomogram for predicting the survival of HIV/AIDS adults who received antiretroviral therapy: a cohort between 2003 and 2019 in Nanjing. *BMC Public Health, 22*(1), 30. doi:10.1186/s12889-021-12249-8

Kautzky-Willer, A., Leutner, M., & Harreiter, J. (2023). Sex differences in type 2 diabetes. *Diabetologia, 66*(6), 986-1002. doi:10.1007/s00125-023-05891-x

Kehoe, K., Boulle, A., Tsondai, P. R., Euvrard, J., Davies, M. A., & Cornell, M. (2020). Long-term virologic responses to antiretroviral therapy among HIV-positive patients entering adherence clubs in Khayelitsha, Cape Town, South Africa: a longitudinal analysis. *J Int AIDS Soc, 23*(5), e25476. doi:https://doi.org/10.1002/jia2.25476

Kim, M. K., Han, K., Cho, J. H., Kwon, H. S., Yoon, K. H., & Lee, S. H. (2020). A model to predict risk of stroke in middle-aged adults with type 2 diabetes generated from a nationwide population-based cohort study in Korea. *Diabetes Res Clin Pract, 163*, 108157. doi:10.1016/j.diabres.2020.108157

Leulseged, T. W., & Ayele, B. T. (2019). Time to optimal glycaemic control and prognostic factors among type 2 diabetes mellitus patients in public teaching hospitals in Addis Ababa, Ethiopia. *PLoS One, 14*(7), e0220309. doi:10.1371/journal.pone.0220309

Li, B., Zhang, L., Liu, Y., Xiao, J., Li, C., Fan, L., . . . Zhao, H. (2022). A novel prediction model to evaluate the probability of CD4+/CD8+ cell ratio restoration in HIV-infected individuals. *Aids, 36*(6), 795-804. doi:10.1097/qad.0000000000003167

Manne-Goehler, J., Geldsetzer, P., Agoudavi, K., Andall-Brereton, G., Aryal, K. K., Bicaba, B. W., . . . Jaacks, L. M. (2019). Health system performance for people with diabetes in 28 low- and middle-income countries: A cross-sectional study of nationally representative surveys. *PLoS Med, 16*(3), e1002751. doi:10.1371/journal.pmed.1002751

Mesic, A., Spina, A., Mar, H. T., Thit, P., Decroo, T., Lenglet, A., . . . Oo, H. N. (2021). Predictors of virological failure among people living with HIV receiving first line antiretroviral treatment in Myanmar: retrospective cohort analysis. *AIDS Res Ther, 18*(1), 16. doi:10.1186/s12981-021-00336-0

Mhlaba, L., Mpanya, D., & Tsabedze, N. (2023). HbA1c control in type 2 diabetes mellitus patients with coronary artery disease: a retrospective study in a tertiary hospital in South Africa. *Front Clin Diabetes Healthc, 4*, 1258792. doi:10.3389/fcdhc.2023.1258792

Ngassa Piotie, P., Webb, E. M., & Rheeder, P. (2021). Suboptimal control for patients with type 2 diabetes in the Central Chronic Medicine Dispensing programme in South Africa. *Afr J Prim Health Care Fam Med, 13*(1), e1-e7. doi:10.4102/phcfm.v13i1.2648

Rohr, J. K., Ive, P., Horsburgh, C. R., Berhanu, R., Shearer, K., Maskew, M., . . . Fox, M. P. (2016). Developing a predictive risk model for first-line antiretroviral therapy failure in South Africa. *J Int AIDS Soc, 19*(1), 20987. doi:10.7448/ias.19.1.20987

20987

SeyedAlinaghi, S., Afsahi, A. M., Moradi, A., Parmoon, Z., Habibi, P., Mirzapour, P., . . . Dadras, O. (2023). Current ART, determinants for virologic failure and implications for HIV drug resistance: an umbrella review. *AIDS Res Ther, 20*(1), 74. doi:10.1186/s12981-023-00572-6

Van Zee, K. J., Manasseh, D.-M. E., Bevilacqua, J. L. B., Boolbol, S. K., Fey, J. V., Tan, L. K., . . . Kattan, M. W. (2003). A Nomogram for Predicting the Likelihood of Additional Nodal Metastases in Breast Cancer Patients With a Positive Sentinel Node Biopsy. *Annals of Surgical Oncology, 10*(10), 1140-1151. doi:10.1245/ASO.2003.03.015

Wei, F. (2019). Correlation between glycosylated hemoglobin level of patients with diabetes and cardiovascular disease. *Pak J Med Sci, 35*(2), 454-458. doi:10.12669/pjms.35.2.589

Yang, Z., Gong, D., Huang, F., Sun, Y., & Hu, Q. (2022). Epidemiological Characteristics and the Development of Prognostic Nomograms of Patients With HIV-Associated Cutaneous T-Cell Lymphoma. *Front Oncol, 12*, 847710. doi:10.3389/fonc.2022.847710

Chapter 6 Application of Joint Models for Dynamic Prediction

6.1 Introduction

In Chapter 5, a Cox proportional hazards-based prediction model was used to develop a prognosis prediction tool for estimating probabilities of clinical outcomes in HIV and T2DM individuals for a setting where the continuous biomarker was assumed to be measured only at baseline for a given individual. However, repeated, continuous biomarker measurements are more useful in practice to understand disease progression. In this chapter, an extension of the Cox proportional hazards-based prediction model is considered to enable the inclusion of repeated, continuous biomarker measurements. Specifically, a joint model for longitudinal continuous biomarker and time-to-event data was used to generate dynamic predictions of probabilities of clinical outcomes in two disease case studies described in Chapter 4. The context and theoretical background of dynamic prediction under a joint modelling framework were provided in 0 and Chapter 2. Moreover, progress made in the conduct and quality of model evaluation and reporting of published dynamic prediction models was highlighted in Chapter 3. This chapter describes a joint model for dynamic virologic failure and glycaemic control prediction. In Section 6.2, a dynamic prediction model for virologic failure probabilities for individuals living with HIV on ART in the Western Cape is developed and internally validated in the routine HIV development and validation cohorts, respectively. Similarly, in Section 6.3, the routine diabetes development and validation cohorts were used to develop and internally validate a dynamic prediction model for glycaemic control probabilities for T2DM individuals in the Western Cape, South Africa. Concluding remarks are given in Section 6.4.

6.2 Dynamic Prediction of Virologic Failure for Individuals Living with HIV

This section illustrates the application of joint models to the routine HIV development cohort. The model was used for dynamic prediction of virologic failure, and its predictive performance was evaluated in the routine HIV validation cohort. The joint model development details and predictive performance are given in sub-sections 6.2.1 and 6.2.2. An illustration of dynamic

predictions of virologic failure probabilities is described in sub-section 6.2.3. Findings are summarised in sub-section 6.2.4.

6.2.1 Joint model for longitudinal continuous HIV viral load and time to virologic failure

The joint model specified a linear mixed-effects model to describe underlying longitudinal trajectories of continuous HIV viral load values for each individual and used these estimated trajectories in a Cox proportional hazards model for time to virologic failure. Longitudinal trajectories for randomly selected individuals in the motivating routine HIV Western Cape development cohort are presented in Figure 6.1. The figure shows variation and nonlinear trends in the longitudinal trajectories of the randomly selected individuals. The characteristics of the longitudinal trajectories warrant a mixed-effects model (Chapter 2) with the nonlinear effect of follow-up time in the fixed effects and both random intercepts and slopes in the random effects.

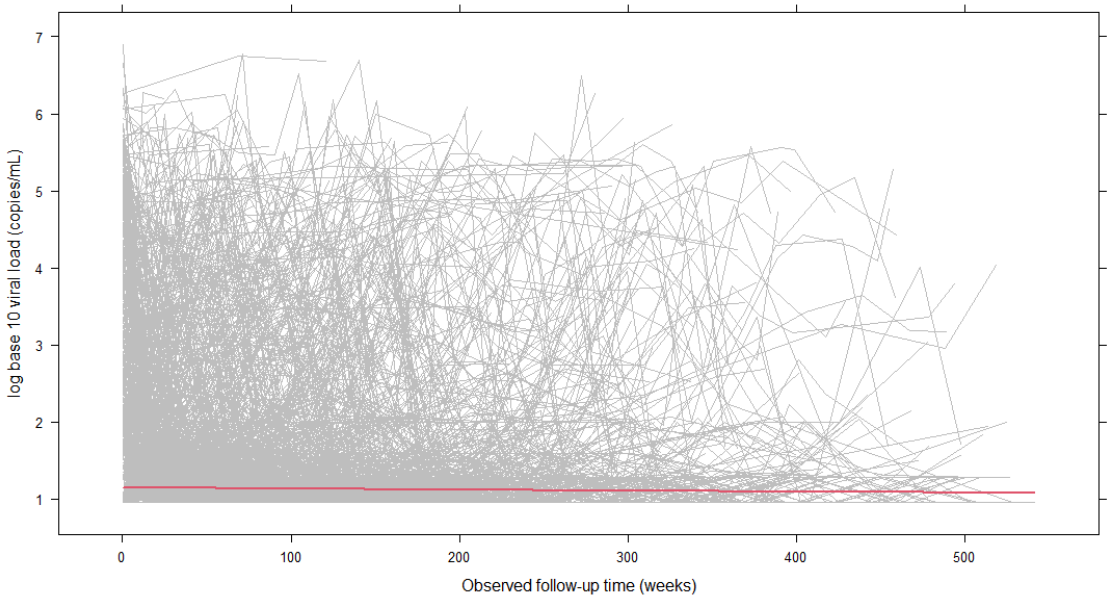

Figure 6.1: Observed longitudinal continuous HIV viral load trajectories for randomly selected individuals in the routine HIV Western Cape overall cohort.

Mixed-effects model for longitudinal continuous HIV viral load data

Two mixed-effects models, i) a mixed-effect model with random intercepts only and (ii) a mixed-effect model with linear random intercepts and slopes, were compared to determine which structure is a better fit for the routine HIV Western Cape development cohort.

The mixed-effect model with random intercepts only was given as:

$$\begin{aligned}\log_{10}\text{vload}_{ij} &= \log_{10}\text{vload}^{*}_{ij} + \epsilon_{ij} \\ &= (\beta_0 + b_{i0}) + \beta_1 N(\text{t}_{ij})_1 + \beta_2 N(\text{t}_{ij})_2 + \beta_3 N(\text{t}_{ij})_3 \\ &+ \beta_4 \text{age}_{16-19} + \beta_5 \text{age}_{20-29} + \beta_6 \text{age}_{40-49} + \beta_7 \text{age}_{50+} \\ &+ \beta_8 \text{male}_{\text{i}} + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2) \text{ and } b_i \sim N(0, D),\end{aligned} \tag{6.1}$$

The mixed-effect model with linear random intercepts and slopes was given as:

$$\begin{aligned}\log_{10}\text{vload}_{ij} &= \log_{10}\text{vload}^{*}_{ij} + \epsilon_{ij} \\ &= (\beta_0 + b_{i0}) + b_{i1}\text{t}_{ij} + \beta_1 N(\text{t}_{ij})_1 + \beta_2 N(\text{t}_{ij})_2 + \beta_3 N(\text{t}_{ij})_3 \\ &+ \beta_4 \text{age}_{16-19} + \beta_5 \text{age}_{20-29} + \beta_6 \text{age}_{40-49} + \beta_7 \text{age}_{50+} \\ &+ \beta_8 \text{male}_{\text{i}} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2) \text{ and } b_i \sim N(0, D).\end{aligned} \tag{6.2}$$

where $\log_{10}\text{vload}_{ij}$ is continuous HIV viral load value for an ith individual at the jth observed follow-up time (in weeks), $\log_{10}\text{vload}^{*}_{ij}$ represents underlying longitudinal continuous HIV viral load trajectory for the individuals, β_0 is the intercept, β_k, $k = 1,2,3$ are regression parameters for the B-splines, β_k, $k = 4,5,\cdots,8$ are the regression coefficients, b_{ik}, $k = 0,1$ are the random effects assumed to have a normal distribution with a mean of zero and variance-covariance matrix D, and ϵ_{ij} are error terms assumed to be normally distributed with mean zero and variance σ_ϵ^2. In both equations (Equations 6.1 and 6.2), $N(\cdot)$ represents the nonlinear effect of follow-up time through B-splines having two internal knots placed at the 5% (week 0.14) and 95% (week 317) percentiles of observed follow-up time. The R code used to fit the mixed-effects models is given in Appendix 5.

Table 6.1 compares the two mixed-effects models given in Equations 6.1 and 6.2. The corresponding information criterion (Chapter 2), i.e., Bayesian information criteria (BIC) low value, suggests that a mixed-effect model with linear random intercepts and linear random slopes provides a better fit to the routine HIV Western Cape development cohort.

Figure 6.2 presents an effects plot showing the mixed-effects model-predicted HIV viral load trajectories from baseline to 572 weeks across males and females of all age groups in the routine HIV Western Cape development cohort. The plot provides an interpretable visualization of the non-linear time trends captured by the spline terms in the model. Unlike a plot of mean HIV viral load over time, it reflects model-based predictions, allowing for interpretation of the average HIV viral load trajectory across subgroups. Specifically, the fitted values show a consistent decline in HIV viral load over time across all age groups and both sexes.

Table 6.1: Comparison between a random intercepts LME model (model 1) and an LME model (model 2) with random intercepts and slopes.

Model	df[1]	BIC[1]
Model 1	11	986 529.6
Model 2	13	982 410.7

[1]*df = degrees of freedom, BIC = Bayesian information criterion*

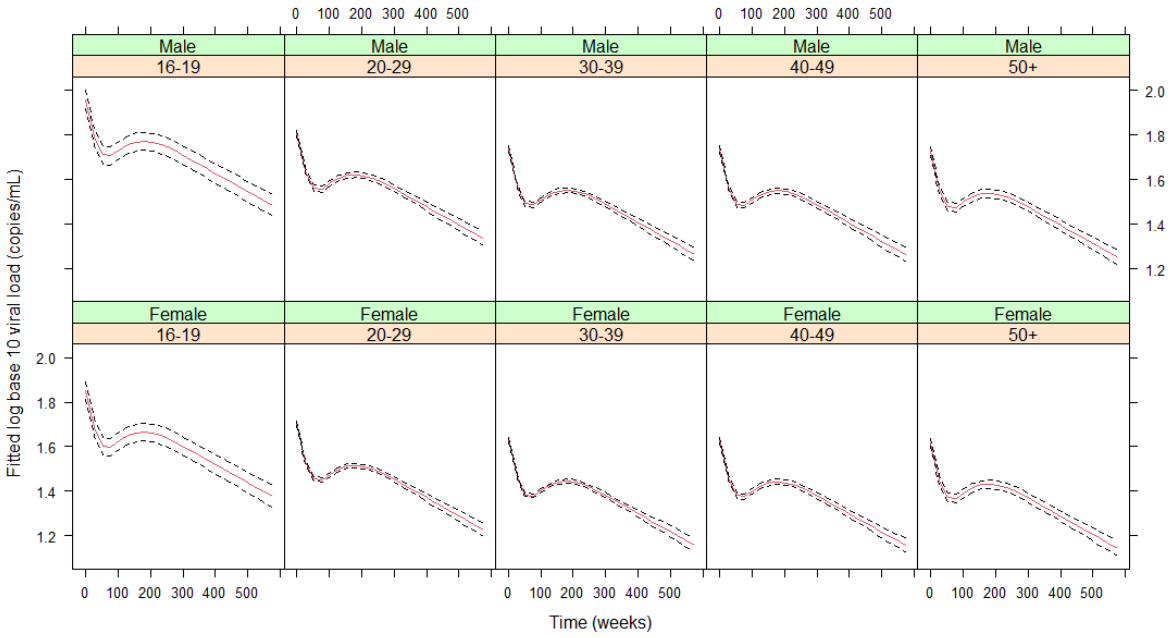

Figure 6.2: Fitted average longitudinal continuous HIV viral load profiles from an LME model with random intercepts and linear slopes in the random effects part, adjusted for baseline differences in age group and sex, and non-linearity through splines in the fixed effects part. The broken lines denote the corresponding 95% confidence intervals for each fitted continuous HIV viral load value in red.

A joint model for longitudinal continuous HIV viral load and time to virologic failure data

The estimated longitudinal continuous HIV viral load trajectory for the individuals from the mixed-effects model with random intercepts and linear slopes was used in a joint model (Chapter 2) for dynamic prediction of virologic failure probabilities for individuals living with HIV in the Western Cape. Two joint models, i) one with a current value association structure and ii) one with a lagged effects association structure, were compared to determine which provides a better fit to the routine HIV Western Cape development cohort.

The joint model with a current value association structure was given as:

$$h_i(t) = h_0(t)\exp(\gamma_1 \text{age}_{16-19} + \gamma_2 \text{age}_{20-29} + \gamma_3 \text{age}_{40-49} + \gamma_4 \text{age}_{50+} + \gamma_5 \text{male}_i + \varphi_1 \log_{10} \text{vload}_i^*(t)), \quad 6.3$$

The current value association structure used the association parameter φ_1 to quantify the association between the current level of continuous HIV viral load at follow-up time t, i.e., $\log_{10}\text{vload}_i^*(t)$ and the hazard $h_i(t)$ of virologic failure at this follow-up time for individual i.

The joint model with a lagged effects association structure was given as:

$$h_i(t) = h_0(t)\exp(\gamma_1\text{age}_{16-19} + \gamma_2\text{age}_{20-29} + \gamma_3\text{age}_{40-49} + \gamma_4\text{age}_{50+} + \gamma_5\text{male}_i + \varphi_2\log_{10}\text{vload}_i^*\{\max(t-c,0)\}), \quad 6.4$$

The lagged effects association structure used the association parameter, φ_2, to quantify the association between the underlying trajectory of the continuous HIV viral load at a previously observed follow-up time $t-c$, where c is a six-month time lag, i.e., $\log_{10}\text{vload}_i^*\{\max(t-c,0)\}$ and the hazard $h_i(t)$ of virologic failure at this follow-up time for individual i.

In the two joint models (Equations 6.3 and 6.4), $h_0(t)$ denotes a B-spline baseline hazard function, $\gamma_1,\cdots,\gamma_5$ are regression coefficients for baseline age groups and male dummy variable.

Bayesian inference

A Bayesian inference through the Monte Carlo simulation scheme was used to infer unknown parameters of interest for the two joint models in Equations 6.3 and 6.4. The R package JMbayes2 (Rizopoulos, Papageorgiou, & Miranda Afonso, 2022) was used to fit the two joint models. Default JMbayes2 non-informative priors for model parameters were used. Three chains with 2000 after burn-in MCMC posterior samples were generated for each model parameter. The MCMC estimation of the current value parameterised joint model (Equation 6.3) took 7.3 hours on a machine with a 3.6GHz (Intel Core i7) processor and 16 GB RAM running 64-bit Windows 10 Enterprise (version 21H2). On the same machine, the MCMC estimation of the lagged effects parameterised joint model (Equation 6.4) took 3.4 hours. The R code used to fit the current value and lagged effects parameterised joint models is given in Appendix 5.

Table 6.2 compares the two joint models. The current value parameterisation had a low deviance information criterion (DIC) value. This suggested that the current value-parameterised joint model provided an improved fit to the routine HIV Western Cape development cohort. In addition, diagnostics plots (trace and density plot) for the association parameter in the current value-parameterised joint model applied to the routine HIV development cohort are shown in Figure 6.3. A bell-shaped density plot indicates unsatisfactory convergence to a target posterior distribution of the association parameter. However, the random scatter around the mean value 6.2 suggests satisfactory convergence to the association parameter.

Table 6.2: Comparison between current value and lagged effects parameterised joint models.

Model	DIC[1]	WAIC[1]	LPML[1]
Current value	1 075 953	1 075 957	-537 978.6
Lagged effects	1 255 066	1 085 774.97	-542 893.5

[1]*DIC = deviance information criterion, WAIC = Watanabe-Akaike information criterion, LPML = log-pseudo-marginal-likelihood value*

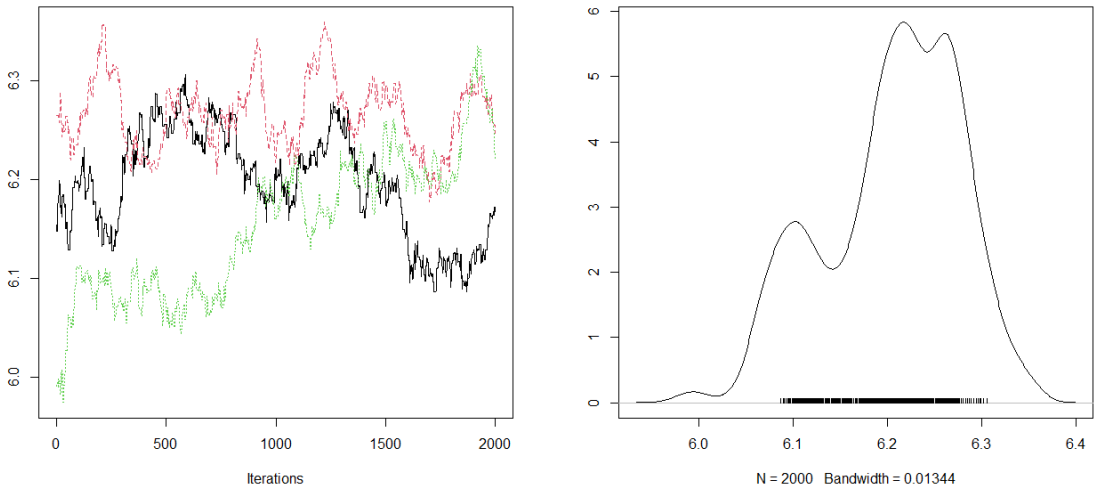

Figure 6.3: MCMC diagnostics plots for the association parameter of the current value-parameterised joint model applied to the routine HIV Western Cape development cohort.

Table 6.3 shows the parameter estimates and the 95% credible intervals of the current value-parameterized joint model applied to the routine HIV Western Cape development cohort.

Males in the cohort were statistically associated with higher viral loads compared to females. On average, after adjusting for baseline age group and follow-up time, males had HIV viral loads that were 0.11 copies/mL higher than females. Individuals younger than 30 years had, on average, higher viral loads than those aged 30 and above. There was a strong association between continuous HIV viral load and the hazard of virologic failure. Adjusting for baseline age group and sex and an increase in the current level of HIV viral load value at time t, the hazard of virologic failure at that time increased by a factor of 6.2 (95% credible interval: 6.1, 6.3). The predictive performance of the current value-parameterised joint model was evaluated in the routine HIV Western Cape validation cohort in sub-section 6.2.2.

Table 6.3: Parameter estimates, and the 95% credible intervals of the current value parameterised joint model applied to the routine HIV Western Cape cohort.

	Mean	2.5%[1]	97.5%[1]
Event process			
Baseline age group (years)			
30-39	-	-	-
16-19	-0.4177	-0.9121	0.0380
20-29	-0.1401	-0.2352	-0.0475
40-49	-0.2359	-0.3676	-0.1172
50+	-0.2751	-0.4804	-0.0565
Male	-0.6140	-0.7160	-0.5165
$\log_{10}$ HIV viral load (copies/mL)	6.2051	6.0637	6.3237
Longitudinal process			
Baseline group (years)			
30-39	-	-	-
16-19	0.2293	0.1858	0.271
20-29	0.0744	0.0626	0.0861
40-49	-0.0048	-0.0176	0.0086
50+	-0.0177	-0.0378	0.1236
Male	0.1128	0.1018	0.145
Spline terms			
B-spline 1 (β_1)	-0.0068	-0.0193	0.0046
B-spline 2 (β_2)	-0.3799	-0.3949	-0.3647
B-spline 3 (β_3)	-0.0409	-0.0567	-0.0253
Variance components			
σ_ϵ	0.7222	0.7201	0.7243
D_{11}	0.6180	0.607	0.629
D_{21}	-0.00166	-0.00172	-0.00162
D_{22}	0.00000964	0.0000093	0.00000993

[1] *$D[r, c]$ is the rc-element of the D covariance matrix of the random effects, 2.5% = lower limit of credible interval, 97.5% = upper limit of credible interval.*

6.2.2 Predictive performance

The predictive performance of the current value-parameterised joint model was evaluated using discrimination and calibration. Time-dependent AUCs and Brier scores were used to assess discrimination and calibration of the model in the routine HIV Western Cape validation cohort. During the first six months of observed follow-up, a subset of virologic-failure-free individuals was used to assess predictive performance at two selected clinically relevant future times (12 and 24 months). The resulting longitudinal data (up to 6 months) of the individuals was used to calculate time-dependent AUCs and Brier scores at the selected times. A summary of the time-dependent AUCs and Brier scores for the current value-parameterised joint model is provided in Table 6.4. Figure 6.4 and Figure 6.5**Table *6.4*** show the time-dependent AUCs and Brier scores at 12 and 24 months, respectively.

Time-dependent AUCs at 12 and 24 months were 0.69 and 0.73, respectively. The value of the AUCs suggests that the current value-parameterised joint model moderately distinguishes individuals of low and high risk of virologic failure.

The Brier scores were 0.033 and 0.051 at 12 and 24 months, suggesting that the current value-parameterised joint model had a very low prediction error and thus provided adequate dynamic predictions of virologic failure. However, as the calibration plot (Figure 6.5) shows, most of the estimated virologic failure probabilities are generally very low at the selected times.

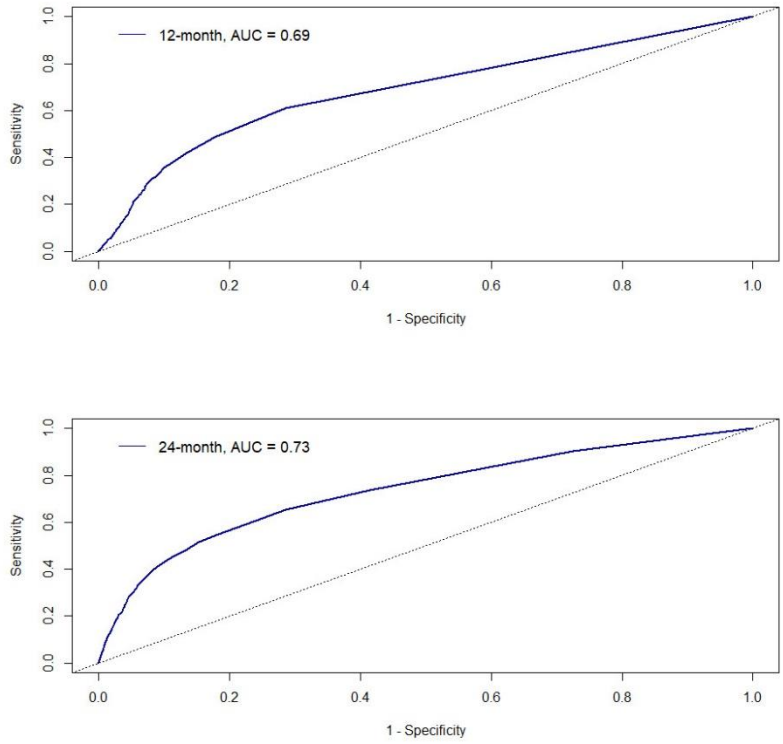

Figure 6.4: Time-dependent AUCs from ROC curves among individuals living with HIV in the routine HIV Western Cape validation cohort at 12 and 24 months.

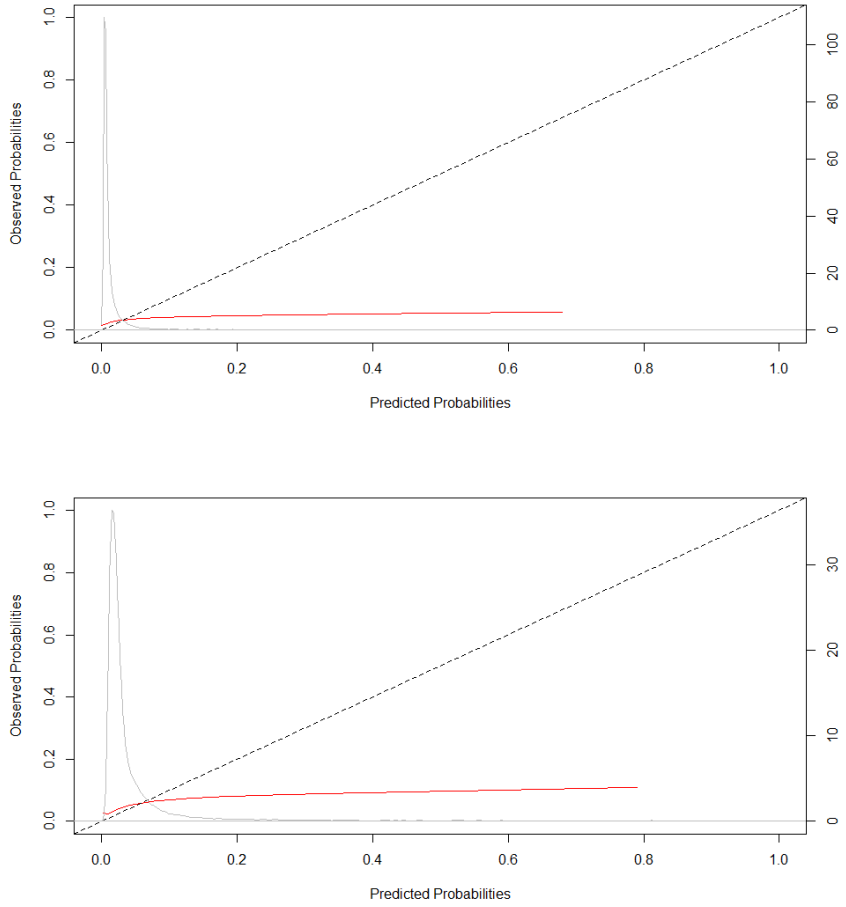

Figure 6.5: Calibration curves at 12 (top panel) and 24 months (bottom panel) prediction time windows for probabilities of virologic failure in the routine HIV Western Cape validation cohort. The diagonal line denotes the line of perfect calibration. The density function (grey curve) denotes a non-parametric estimate of the distribution of predicted virologic failure across the validation cohort.

Table 6.4: Time-dependent AUCs and Brier scores for the current value-parameterised joint model at selected time windows in the routine HIV Western Cape validation cohort.

Time window	AUC[1]	Brier score
12-month	0.69	0.033
24-month	0.73	0.051

[1]*AUC = Area under the curve*

6.2.3 Dynamic prediction of the probability of virologic failure

The fitted current value-parameterised joint model was used to generate dynamic predictions of virologic failure probabilities for individuals living with HIV in the Western Cape. Two individuals (M-9872 and M-100240) were randomly selected from the cohort for illustrative purposes. These individuals were virologic failure-free in the first six months of follow-up. The first individual (M-9872) was a 38-year-old male with a continuous HIV viral load value (log base 10) of 2.40 copies/mL at baseline and did not have virologic failure by the end of observed follow-up time (0.1, 13.6 and 116.1 weeks). The second individual (M-100240) was a 37-year-old female with a continuous HIV viral load value of 2.02 copies/mL at baseline. She did not have virologic failure by the end of the observed follow-up time (0.1, 36.1 and 60.1 weeks). Figure 6.6 shows the HIV viral load values during the first six months of observed follow-up for two randomly selected individuals. M-100240 had increasing HIV viral load values in the first six months of follow-up, while M-9872 had decreasing HIV viral load values at the same follow-up time. The fitted longitudinal trajectory for the two individuals did not align well with the observed HIV viral load values (asterisks) possibly because the sparse data is difficult to adequately capture in the joint model's mixed effects model.

Dynamic predictions of virologic failure probabilities for the two individuals were calculated and presented in Figure 6.7 and Figure 6.8. The plots show the estimated conditional probabilities of virologic failure for the two individuals with data up to time point s prior to their last observed follow-up time t. Given that individual M-9872 had no virologic failure by week 13.6 prior to the previous observed follow-up (116.1 weeks), Figure 6.7 presents the estimated probabilities of the first virologic failure for the individual. Similarly, given that individual M-100240 had no virologic failure by week 36.1 before the last observed follow-up time (60.1 weeks), Figure 6.8 presents the estimated probabilities of virologic failure for the individual.

The conditional probabilities of virologic failure for M-9872 were much smaller when more longitudinal continuous HIV viral load values accumulated over follow-up time. M-100240 had a steady progression to virologic failure as more longitudinal continuous HIV viral load values accumulated over follow-up time. These conditional probabilities of virologic failure were much smaller on average. In addition, the 95% Monte Carlo prediction intervals for M-9872 were narrower than M-100240, suggesting inconsistent precision in the estimated

probabilities, potentially due to suboptimal estimates of the individual-specific longitudinal trajectories leading to virologic failure from the joint model's mixed effects model.

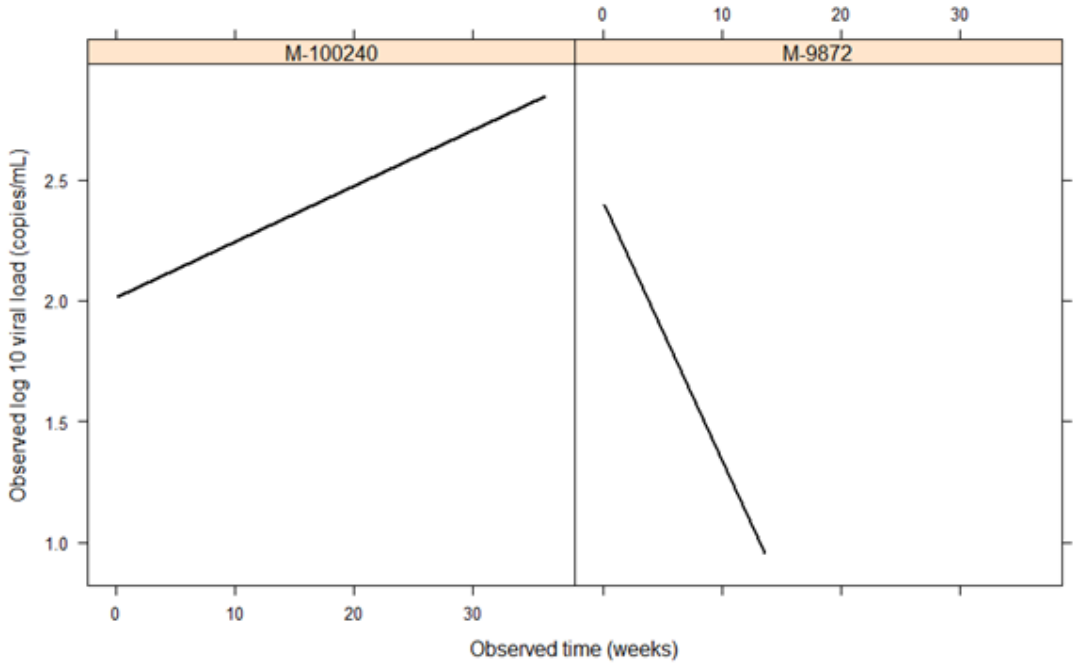

Figure 6.6: Observed HIV viral load values during the first six months of observed follow-up for two randomly selected individuals from the routine HIV Western Cape validation cohort.

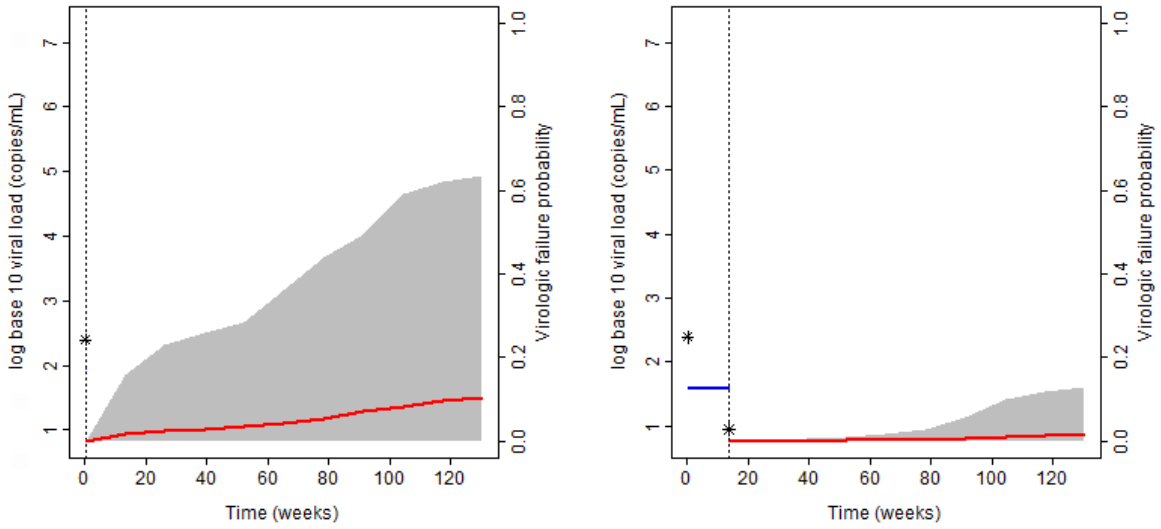

Figure 6.7: Estimated probabilities of virologic failure, conditional on being failure-free during the first six months of follow-up for M-9872 in the validation cohort. The black asterisks

represent observed log base 10 HIV viral load values during the first six months of observed follow-up time. The vertical dotted lines represent the observed follow-up time. A solid blue line depicts the fitted longitudinal trajectory from a mixed-effects model. The solid red line represents the estimated probability of virologic failure, and the shaded grey area is the corresponding 95% confidence interval of the probabilities. The panels represent predictions made over time, with the left panel showing predictions at the baseline and the right panel showing predictions at the second follow-up time.

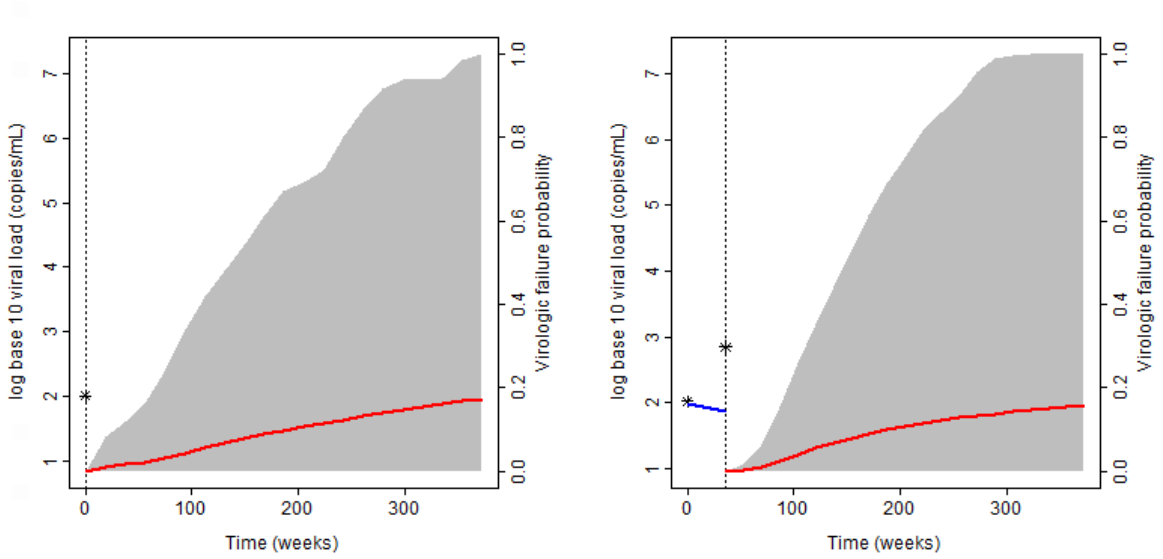

Figure 6.8: Estimated probabilities of virologic failure, conditional on being failure-free during the first six months of follow-up for M-100240 in the validation cohort. The black asterisks represent observed log base 10 HIV viral load values during the first six months of observed follow-up time. The vertical dotted lines represent the observed follow-up time. A solid blue line depicts the fitted longitudinal trajectory from a mixed-effects model. The solid red line represents the estimated probability of virologic failure, and the shaded grey area is the corresponding 95% confidence interval of the probabilities. The panels represent predictions made over time, with the left panel showing predictions at the baseline and the right panel showing predictions at the second follow-up time.

6.2.4 Discussion

A joint model was used to model longitudinal HIV viral load and time to virologic failure data in this section. This presents a novel approach to individualised prognosis of virologic failure probabilities. Compared with the Cox proportional hazard-based prediction approach in Chapter 5, this approach integrates repeated HIV viral load measurements over time through a mixed-effects model, providing a more comprehensive understanding of HIV progression dynamics. The model assumed that the hazard of virologic failure depends on the current viral load. This parameterisation is known as the current value-parameterisation of the joint model (Rizopoulos, 2012). The current value-parameterised joint model aligns with the pathophysiological process of HIV viral load, characterised by abrupt variations due to non-adherence to treatment (WHO, 2016).

The joint model was applied to the routine data from electronic HIV health records from the Western Cape introduced in Chapter 4. In particular, the development and validation cohorts were used to develop and validate model development. A mixed-effects model was used to model individual-specific average HIV viral load change over time and subsequently include an estimated longitudinal profile of individuals as a covariate in a Cox model, adjusting for baseline age and sex.

The findings showed that the increase in the current value of HIV viral load (log base 10 copies/mL) was associated with an increased hazard of virologic failure after adjusting for baseline age and sex. This finding is consistent with an African study which reported that HIV viral load changes over time were associated with increased hazard of an unfavourable outcome (death or treatment failure) (Zakaria, Ayele, Kebede, Jaldo, & Lajore, 2022). However, this study used a joint model with a lagged effects parameterisation, which implied viral load in the previous six months was associated with an unfavourable outcome. Nonetheless, the findings of this thesis demonstrate the importance of incorporating longitudinal HIV viral load measurements in estimating the hazard of virologic failure and subsequently predicting virologic failure probabilities, as opposed to relying solely on baseline values (Chapter 5).

The joint modelling approach considers variability individually to get prognosis prediction of virologic failure probabilities based on an individual's characteristics. Individual-specific average HIV viral load changes over time were used for individualised prognosis prediction of virologic failure that was updated dynamically with data at hand. The findings indicate that the Cox proportional hazard-based prediction model (Chapter 5) had superior discrimination ability compared to the joint modelling-prediction model (0.89 vs 0.69 AUCs at 12 months and 0.84 vs 0.73 AUCs at 24 months). In contrast, the joint modelling-based prediction model had better calibration than the Cox proportional hazard-based prediction model (0.046 vs 0.033 Brier scores at 12 months and 0.059 vs 0.051 Brier scores at 24 months).

These surprising findings underscore the sparseness and timing mismatch between HIV viral load measurements and virologic failure in the routine electronic HIV data. The individual-specific trajectories leading to virologic failure from the joint model's mixed effects model may be poorly estimated, resulting in reduced the predictive performance of the model. Nonetheless, joint modelling-based predictions provide a differentiated approach that can be useful in viral load monitoring to adequately and timely identify a higher probability of virologic failure in resource-limited settings (Mesic et al., 2021).

The joint modeling-based predictions for individuals living with HIV had limitations. During HIV viral load testing, a biomarker can be below or above a detection limit. One limitation of joint modeling-based predictions is the inability to appropriately handle biomarker values below the detection limit of an essay. In Chapter 7, a two-part joint model is applied to enable predictions that consider biomarker values below the lower limit of detection.

Limitations of applying prediction models to the routine electronic HIV health data were discussed in Chapter 5. These limitations included i) unavailability of essential variables, ii) generalisability of the findings to other resource-limited settings due to the lack of an external validation cohort, and iii) findings were specific to populations living with HIV in the Western Cape province only in South Africa.

Another limitation is that resampling approaches were not used to verify the generalisability of the findings further due to computational constraints associated with using such approaches in the joint modeling framework. This is an area for future research.

The strengths of the findings lie in the use of routine electronic HIV data with a large sample size, which allowed for the internal validation of the dynamic prediction model. The dynamic prediction for virologic failure had several advantages over the Cox proportional hazards-based prediction model. It managed the longitudinal nature of HIV viral load values and explored their association with the hazard of virologic failure. Another strength is that it enables individual-specific prognosis prediction of virologic failure that dynamically updates with the available data over time. Finally, these predictions can be used to identify individuals likely to experience virologic failure timeously, enabling personalised therapeutic interventions.

In conclusion, following the current literature, this thesis presents the first attempt to compare the predictive performance of Cox proportional hazard and joint modelling-based prediction models using routine HIV data from electronic health records in a resource-limited setting. The findings indicate that, while dynamic prediction models offer advantages in handling longitudinal HIV viral load data and providing individual-specific prognosis predictions of virologic failure, they do not consistently outperform Cox proportional hazard-based models. Factors such as the availability of variables and data characteristics are central to predictive performance. These findings contribute to a clearer understanding of the predictive performance of dynamic prediction models applied to routine HIV data from electronic health records in resource-limited settings and highlight the need for a balanced approach in

evaluating predictive performance between these and Cox proportional hazard-based prediction models.

6.3 Dynamic Prediction of Glycaemic Control for T2DM Individuals

A joint model is applied to the routine diabetes development cohort in this section. The predictive performance and dynamic prediction of glycaemic control were calculated from the routine diabetes validation cohort. The specification of the joint model is given in sub-section 6.3.1. The predictive performance of the joint model is given in sub-section 6.3.2, and dynamic predictions of glycaemic control probabilities are shown in sub-section 6.3.3. The findings are summarised in sub-section 6.3.4.

6.3.1 Joint model for longitudinal continuous HbA1c and time to glycaemic control

A mixed-effects model was used to describe underlying longitudinal trajectories of continuous HbA1c values for each individual. Longitudinal trajectories for randomly selected individuals in the routine diabetes Western Cape development cohort are presented in Figure 6.9. The longitudinal trajectories varied and had nonlinear trends. The characteristics of the longitudinal trajectories suggest that a mixed-effects model with a nonlinear effect of follow-up time in the fixed effects and random intercepts and slopes in the random effects is needed.

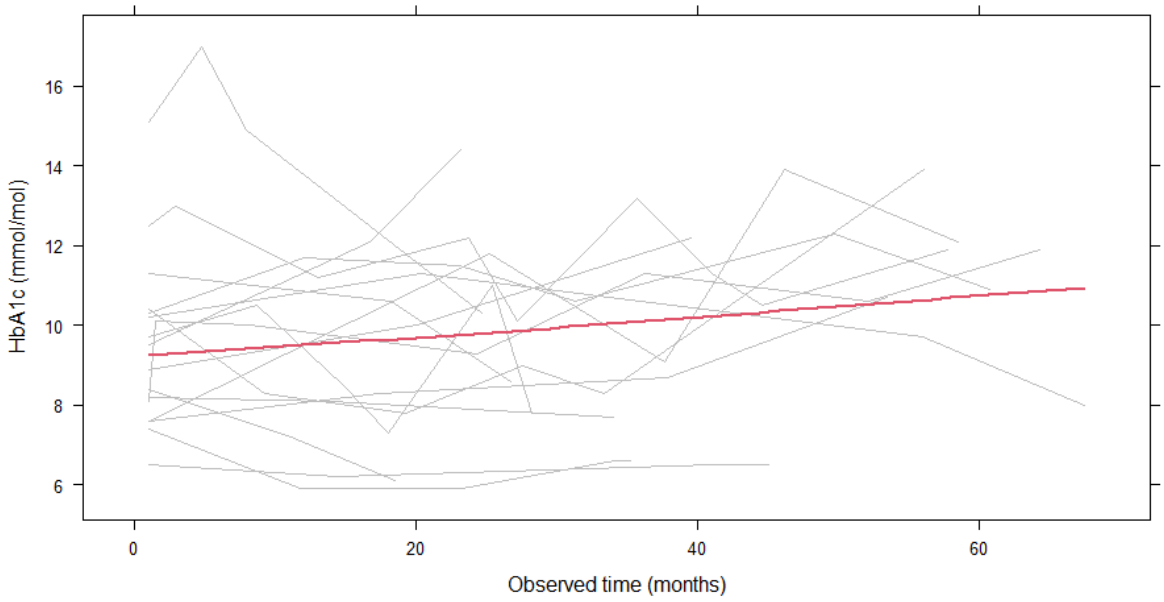

Figure 6.9: Observed longitudinal continuous HbA1c trajectories for randomly selected individuals in the routine diabetes Western Cape overall cohort.

Mixed-effects model for longitudinal continuous HbA1c data

B-splines with two internal knots placed at the 5% (month 1) and 95% (month 71) percentiles of observed follow-up time were used to represent the nonlinear effects of follow-up time on continuous HbA1c. Two mixed effects models, i) one with random intercepts only and ii) one with linear random intercepts and slopes, were considered to model the changes over time of continuous HbA1c biomarker values in the routine diabetes development cohort.

The mixed-effect model with random intercepts only was given as:

$$A1c_{ij} = A1c_{ij}^{*} + \epsilon_{ij} \qquad 6.5$$
$$= (\beta_0 + b_{i0}) + \beta_1 N(\mathrm{t}_{ij})_1 + \beta_2 N(\mathrm{t}_{ij})_2 + \beta_3 N(\mathrm{t}_{ij})_3 + \beta_4 \mathrm{age}_{35-59}$$
$$+ \beta_5 \mathrm{age}_{60+} + \beta_6 \mathrm{male}_{\mathrm{i}} + \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2) \text{ and } b_i \sim N(0, D).$$

The mixed-effect model with linear random intercepts and slopes was given as:

$$
\begin{aligned}
A1c_{ij} &= A1c_{ij}^{*} + \epsilon_{ij} \\
&= (\beta_0 + b_{i0}) + b_{i1}\mathrm{t}_{ij} + \beta_1 N\left(\mathrm{t}_{ij}\right)_1 + \beta_2 N\left(\mathrm{t}_{ij}\right)_2 + \beta_3 N\left(\mathrm{t}_{ij}\right)_3 \\
&+ \beta_4 \mathrm{age}_{35-59} + \beta_5 \mathrm{age}_{60+} + \beta_6 \mathrm{male}_{\mathrm{i}} \\
&+ \epsilon_{ij}, \qquad \epsilon_{ij} \sim N(0, \sigma_\epsilon^2) \text{ and } b_i \sim N(0, D),
\end{aligned}
\tag{6.6}
$$

where $A1c_{ij}$ is the continuous HbA1c value for an ith individual at the jth observed follow-up time (in months), $A1c_{ij}^{*}$ represent the underlying longitudinal continuous HbA1c trajectory for each individual, β_0 is the intercept, β_k, $k = 1,2,3$ are regression parameters for the B-splines, β_k, $k = 4,5,6$ are the regression coefficients, b_{ik}, $k = 0,1$ are the random effects assumed to be normally distributed with a mean of zero and variance-covariance matrix D, ϵ_{ij} are error terms assumed to have a normal distribution with mean zero and variance σ_ϵ^2, and $N(\cdot)$ represent the B-splines. The appendix (Appendix 5) contains the R code used to fit the two mixed-effects models.

The two mixed-effects models are compared in Table 6.5. The mixed-effects model with random intercepts and linear random slope provides an improved fit to the routine diabetes Western Cape development cohort.

Figure 6.10 presents an effects plot showing the mixed-effects model–predicted HbA1c trajectories from baseline to the end of follow-up for individuals in the routine diabetes development cohort. The plot illustrates trajectories across males and females of all age groups. Notably, for individuals aged 60 years and older, regardless of sex, the fitted HbA1c values increased steadily over time but remained, on average, lower than those observed in younger age groups.

Table 6.5: Comparison between a random intercept LME (model 1) and an LME model with random intercepts and slopes (model 2).

Model	df[1]	BIC[1]
Model 1	9	989 446.47
Model 2	11	985 038.34

[1]*df = degrees of freedom, BIC = Bayesian information criterion*

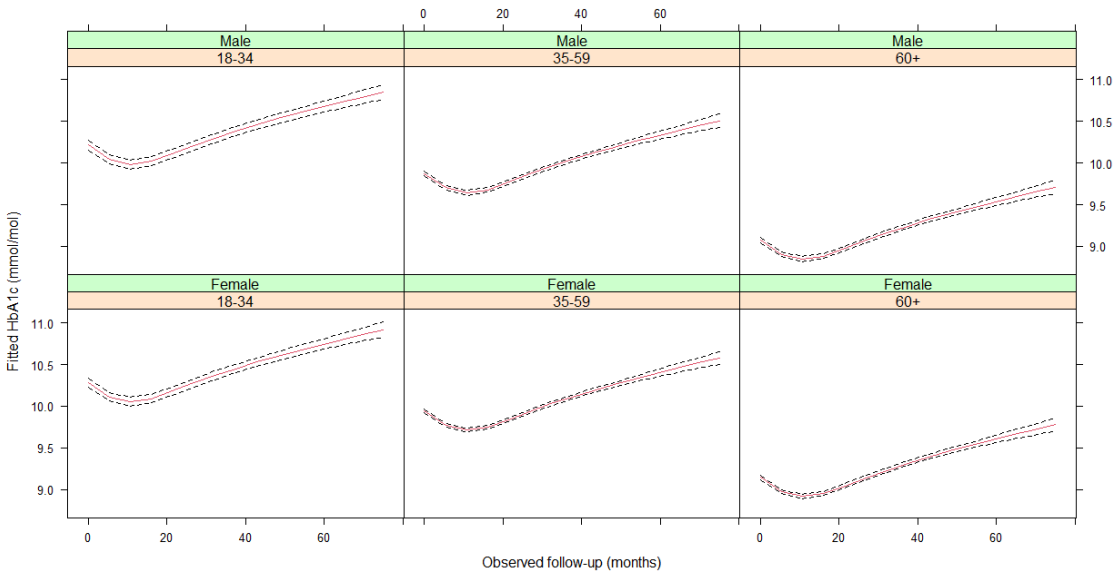

Figure 6.10: Fitted average longitudinal continuous HbA1c profiles from an LME model with random intercepts and linear slopes in the random effects part, adjusted for baseline differences in age group and sex, and non-linearity through splines in the fixed effects part. The broken lines denote the corresponding 95% confidence intervals for each fitted continuous HbA1c value in red.

A joint model for longitudinal continuous HbA1c and time-to-glycaemic control data

A joint model incorporated the estimated longitudinal continuous HbA1c trajectory for the individuals from the mixed-effects model with random intercepts and linear slopes. This joint model was used to calculate dynamic predictions of glycaemic control probabilities for T2DM individuals in the Western Cape. Two variations of the joint model parameterisation were

compared. First, a joint model with a current value association structure parameterisation was considered. The second joint model parameterisation was a cumulative effects association structure.

The joint model with a current value association structure was given as:

$$h_i(t) = h_0(t)\exp\left(\gamma_1 \text{age}_{i_{35-59}} + \gamma_2 \text{age}_{i_{60+}} + \gamma_3 \text{male}_i + \varphi_1 A1c_i^*(t)\right), \quad 6.7$$

The current value association structure used the association parameter, φ_1, to quantify the association between the current level of continuous HbA1c at follow-up time t, i.e., $A1c_i^*(t)$ and the hazard $h_i(t)$ of glycaemic control at this follow-up time for individual i.

The joint model with a lagged effects association structure was given as:

$$h_i(t) = h_0(t)\exp\left(\gamma_1 \text{age}_{i_{35-59}} + \gamma_2 \text{age}_{i_{60+}} + \gamma_3 \text{male}_i + \varphi_2 \int_0^s A1c_i^{**}(s)ds\right) \quad 6.8$$

The cumulative effects association structure used the association parameter φ_2 to quantify the association between a summary of all previous continuous HbA1c up to observed follow-up time t and the hazard $h_i(t)$ of glycaemic control at this follow-up time for individual i.

In the two joint models (Equations 6.7 and 6.8), γ_k, $k = 1,2,\cdots,3$ regression coefficients for baseline age groups and male dummy variable, and $h_0(t)$ is B-spline baseline hazard functions.

Bayesian inference

The R package JMbayes2 was used to estimate the parameters in the current value and cumulative effects parameterised joint models. Default JMbayes2 non-informative priors for model parameters were used to run three chains with 2000 after burn-in MCMC posterior samples for each model parameter. The MCMC estimation of the current value parameterised joint model and the cumulative effects parameterised joint model took the same time (2.5 hours). The R code used to fit the models is given in Appendix 5.

Table 6.6 compares the parameterisations of the two joint models. The current value joint model provides an improved fit to the routine diabetes Western Cape development cohort. However, continuous HbA1c values summarise average blood glucose in the last months (South African National Department of Health, 2014). For this reason, the cumulative effects parameterisation provides more insights about the hazard of achieving glycaemic control than the current value parameterisation. Specifically, a joint model assuming a cumulative effect association structure was used to infer and predict glycaemic control. Figure 6.11 shows the trace and density plots for the association parameter in the cumulative effects parameterised joint model applied to the routine diabetes cohort. There was convergence to a target posterior distribution of the association parameter evidenced by a caterpillar pattern and unimodal bell shape for the trace and density plot, respectively.

Table 6.6: Comparison between current value and cumulative effects parameterised joint models.

Model	DIC[1]	WAIC[1]	LPML[1]
Current value	1 243 706	1 243 698	-621 849.2
Cumulative effects	1 244 045	1 244 041	-622 020.5

[1]DIC = deviance information criterion, WAIC = Watanabe-Akaike information criterion, LPML = log-pseudo-marginal-likelihood value

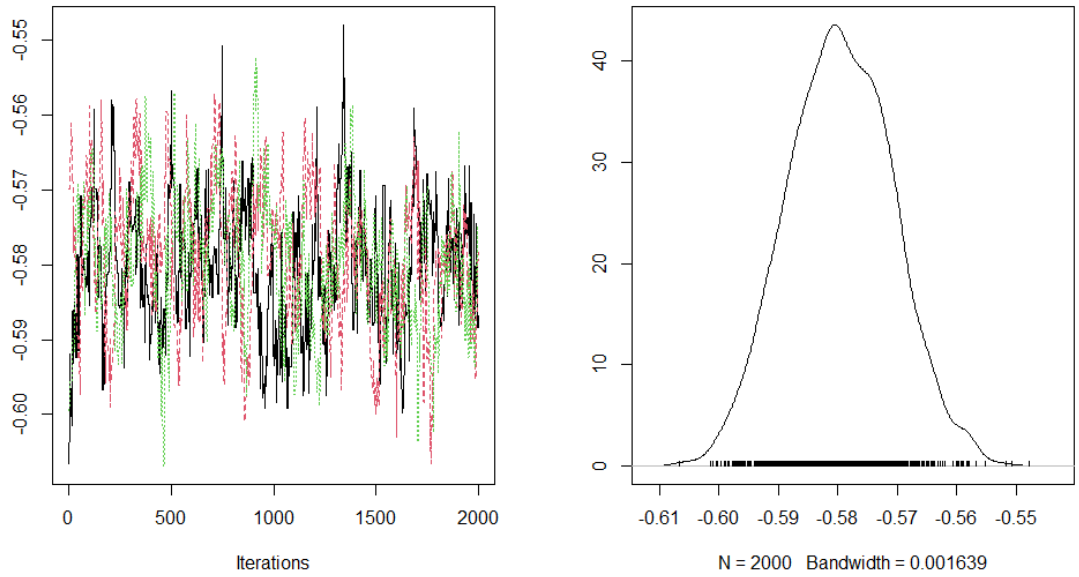

Figure 6.11: MCMC diagnostic plots for the association parameter of the cumulative effects parameterised joint model applied to the routine diabetes development cohort.

Parameter estimates and the 95% credible intervals of the cumulative effects parameterised joint model are given in Table 6.7. Males were statistically found to have lower HbA1c levels compared to females. After adjusting for baseline age group and follow-up duration, males had HbA1c values that were 0.08 mmol/mol lower than females. Additionally, individuals aged 35 and above had lower HbA1c levels compared to those aged 34 and below.. There was a strong association between continuous HbA1c and the hazard of glycaemic control. Adjusted for baseline age groups and sex, a one-unit increase in the summary of all past continuous HbA1c values, the hazard of achieving glycaemic control decreases by a factor of 0.58 (95% credible interval: -0.597, -0.563). In sub-section 6.3.2, the predictive performance of the fitted cumulative effects parameterised joint model is evaluated in the validation cohort.

Table 6.7: Parameter estimates, and the 95% credible intervals of the cumulative effects parameterised joint model applied to the routine diabetes development cohort.

	Mean	2.5%[1]	97.5%[1]
Event process			
Baseline age group (years)			
18-34	-	-	-
35-59	-0.956	-1.051	-0.863
60+	-0.955	-1.041	-0.871
Male	0.170	0.131	0.206
HbA1c (mmol/mol)	-0.580	- 0.597	- 0.563
Longitudinal process			
Basline age group (years)			
18-34	-	-	-
35-59	-0.319	-0.372	-0.265
60+	-1.125	-1.179	-1.068
Male	-0.078	-0.105	-0.049
Spline terms			
B-spline 1 (β_1)	0.236	0.197	0.278
B-spline 2 (β_2)	0.042	-0.0019	0.087
B-spline 3 (β_3)	0.6696	0.6056	0.730
Variance components			
σ_ϵ	1.403	1.398	1.408
D_{11}	3.40	3.334	3.455
D_{21}	-0.0268	-0.0283	-0.0256
D_{22}	0.00122	0.00118	0.00127

[1] *$D[r, c]$ is the rc-element of the D covariance matrix of the random effects, 2.5% = lower limit of credible interval, 97.5% = upper limit of credible interval.*

6.3.2 Predictive performance

Time-dependent AUCs and Brier scores were used to evaluate the predictive performance of the cumulative effects parameterised joint model to the validation cohort. Data from individuals who did not achieve glycaemic control during the first six months of observed follow-up were

used to evaluate predictive performance at two-time windows, 12 and 24 months. The predictive performance by time-dependent AUCs and Brier scores is given in Figure 6.12 and Figure 6.13. The time-dependent AUCs and Brier scores are summarised in Table 6.8.

As shown in Figure 6.12, the cumulative effects parameterized joint model had moderate discrimination ability given by time-dependent AUCs of 0.66 and 0.68 at 12 and 24 months.

Low Brier scores of 0.058 and 0.089 at 12 and 24 months suggest that the cumulative effects parameterised joint model had a low prediction error and provided adequate predictions of glycaemic control. These low Brier scores were demonstrated by the agreement between the observed and estimated probabilities of glycaemic control from the cumulative effects parameterised joint model at the 12 and 24 months (Figure 6.13).

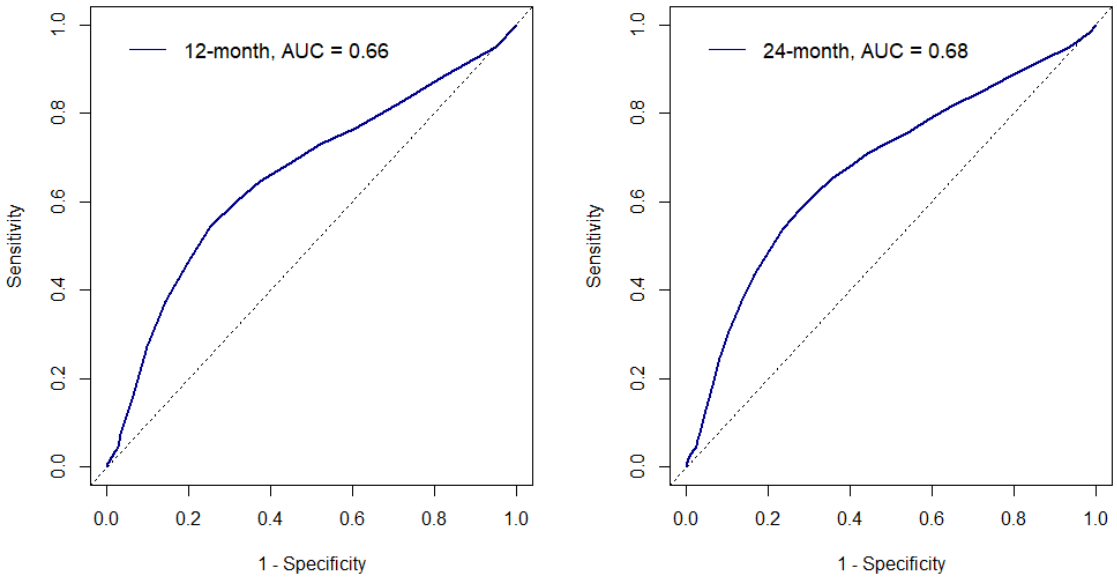

Figure 6.12: Time-dependent AUCs from ROC curves among T2DM individuals in the routine diabetes Western Cape validation cohort at 12 and 24 months.

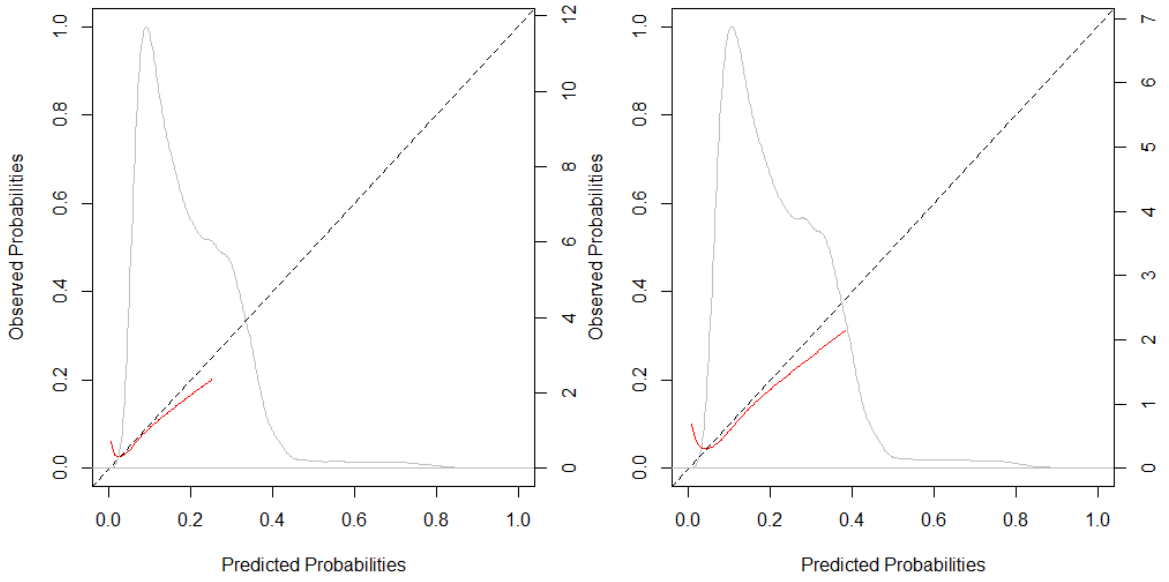

Figure 6.13: Calibration curves at 12 (left panel) and 24 months (right panel) prediction time windows for probabilities of glycaemic control in the routine diabetes validation cohort. The diagonal line denotes the line of perfect calibration. The density function (grey curve) denotes a non-parametric estimate of the distribution of predicted glycaemic control across the validation cohort.

Table 6.8: Time-dependent AUCs and Brier scores for the cumulative effect parameterised joint model at selected time windows in the routine validation cohort.

Time window	AUC[1]	Brier score
12-month	0.66	0.058
24-month	0.68	0.089

[1]*AUC = Area under the curve*

6.3.3 Dynamic prediction of glycaemic control probabilities

In this sub-section, the fitted cumulative effects parameterised joint model is used for dynamic predictions of glycaemic control for two randomly selected individuals (A-144452 and A-215470) from the routine diabetes Western Cape validation cohort. These individuals did not achieve glycaemic control in the first six months of follow-up. One of these individuals (A-

215470) achieved glycaemic control, while the other did not. The first individual (A-144452) was a 56-year-old female with an HbA1c biomarker value of 13.1 mmol/mol at baseline and did not have glycaemic control by the end of follow-up (observed follow-ups, 1, 5.57, and 25.62 months). The second individual (A-215470) was a 58-year-old female with an HbA1c of 11.1 mmol/mol at baseline and achieved glycaemic control by the end of follow-up (observed follow-ups, 1, 6.16, 9.65, 16.39, 19.51, and 23.06 months). The HbA1c values observed during the first six months of follow-up for the individuals in the cohort are shown in Figure 6.14. Both individuals had decreasing HbA1c values in the first six months of follow-up, with A-215470's HbA1c values decreasing at a faster and steeper rate than A-144452.

Figure 6.15 shows the dynamic predictions of glycaemic control probabilities for the two individuals. The estimated probabilities of glycaemic control for A-144452, conditional on not achieving glycaemic control by month 5.57 before the last observed follow-up (25.62 months), are given in this figure. The conditional estimated probabilities of glycaemic control for A-215470, who had a last follow-up time of 23.06 months, are shown in the same figure. The estimated conditional probabilities of achieving glycaemic control were lower for both individuals, on average. A-215470 had a steady progression to glycaemic control compared to A-144452 as more longitudinal continuous HbA1c values accumulated over follow-up time. On average, the 95% Monte Carlo prediction intervals for both individuals were narrower.

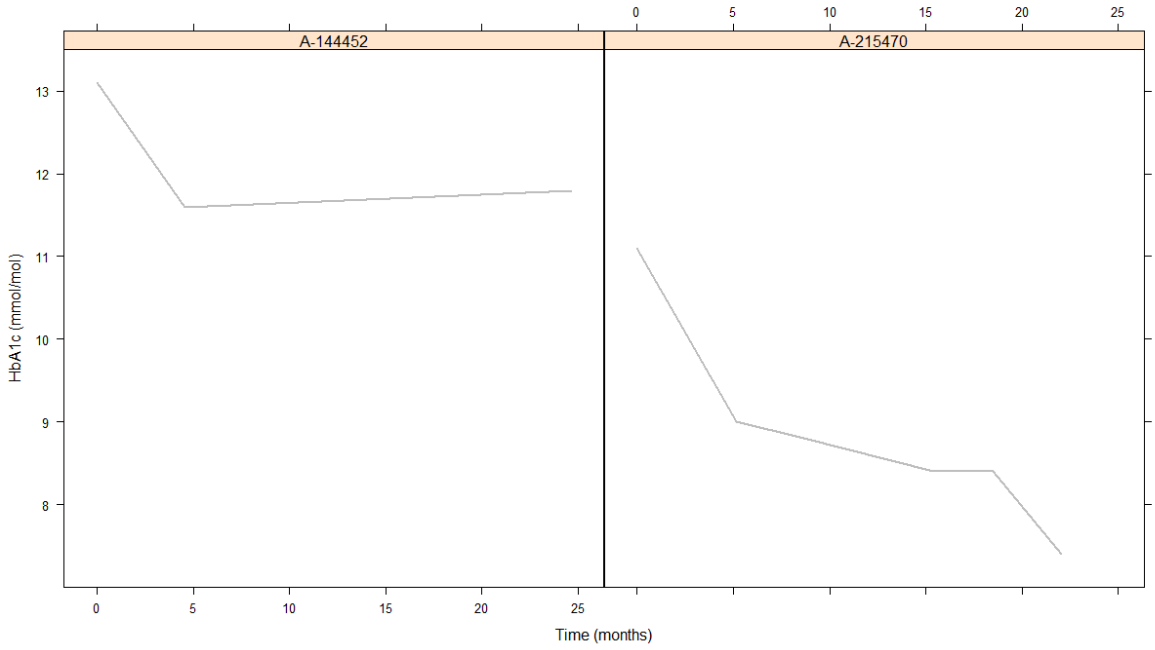

Figure 6.14: Observed HbA1c values during the first six months of observed follow-up for two randomly selected individuals from the routine diabetes Western Cape validation cohort.

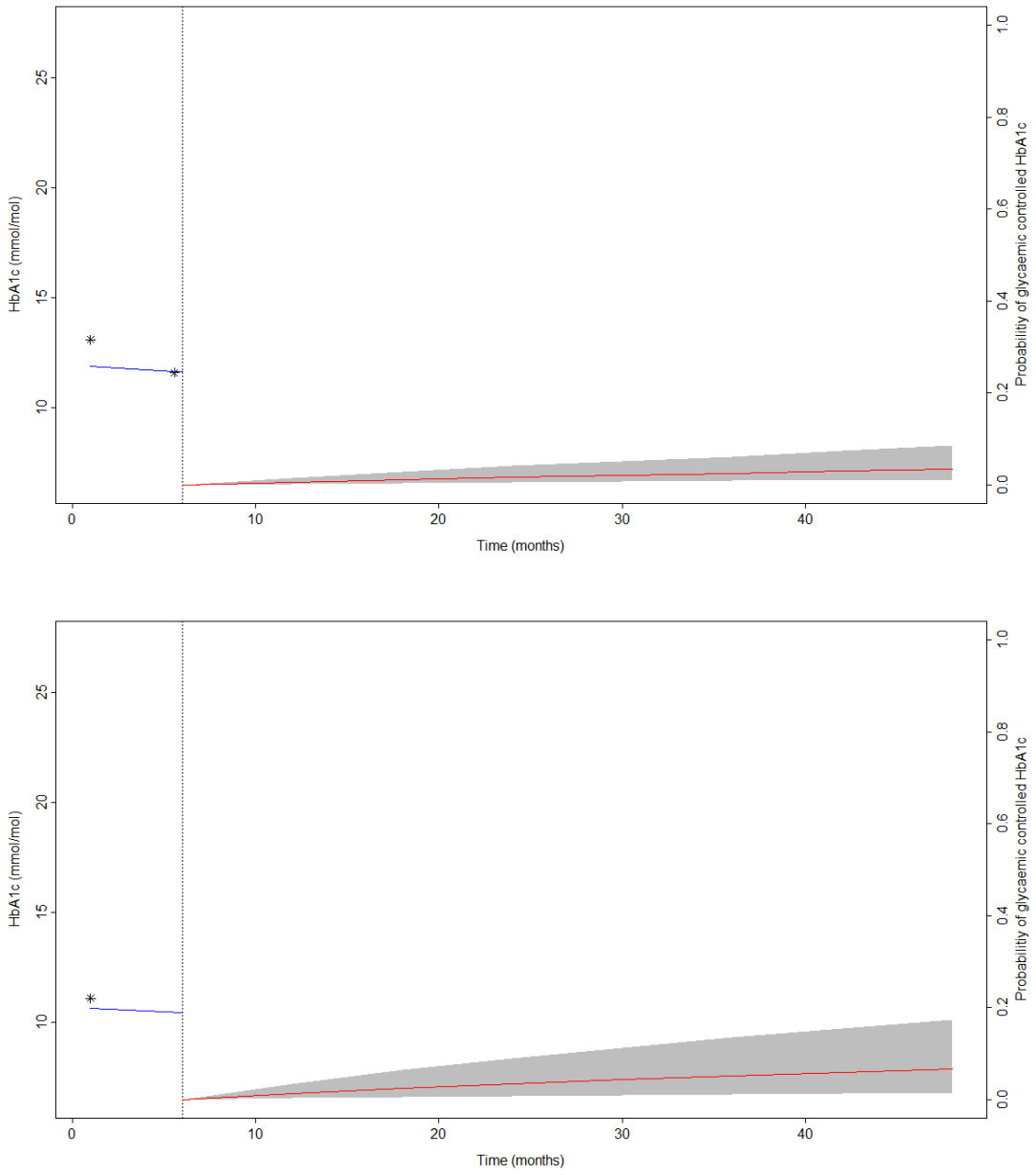

Figure 6.15: Estimated probabilities of glycaemic control, conditional on not having glycaemic control during the first six months of follow-up for A-144452 (top panel) and A-215470 (bottom panel) in the validation cohort. The black asterisks represent HbAlc values during the first six months of observed follow-up. The vertical dotted lines represent observed follow-up time. The solid blue line shows the fitted longitudinal trajectory from a mixed-effects model, and the solid red line represents the estimated probability of glycaemic control, with the shaded grey area corresponding to 95% confidence intervals of the probabilities.

6.3.4 Discussion

A joint model was used for longitudinal HbA1c and time to glycaemic control data. The joint modelling approach offers a significant advantage over the Cox proportional hazards-based prediction model in Chapter 5 as it incorporates repeated measurements of HbA1c over time, providing a comprehensive understanding of T2DM progression dynamics compared to using baseline values. The joint model assumed that the hazard of glycaemic control depends on all past HbA1c values. This parameterisation is the cumulative effects parameterised joint model (Rizopoulos, 2012). This parameterisation was practical because HbA1c reflects a summary of the previous 3-6 months of blood glucose values, as highlighted in 0.

This cumulative effects' model was applied to the routine electronic health records diabetes data from the Western Cape introduced in Chapter 4, with separate cohorts for model development and validation. Individual-specific average HbA1c changes over time were estimated using a mixed-effects model, and these estimated longitudinal profiles of individuals were subsequently included as a covariate in a Cox model, adjusting for baseline age and sex.

The findings demonstrated that an increase in HbA1c reduces the hazard of glycaemic control after adjustment for the covariates, including baseline age and sex. The hazard of glycaemic control changed from 0.777 to 0.56 when longitudinal HbA1c was compared to using baseline HbA1c values under the Cox proportional hazard model in Chapter 5. The predictive performance of these approaches was compared to highlight the differences in the joint and Cox proportional hazard-based prediction modelling approaches.

A previous study compared the predictive accuracy of the joint modelling approach with the Cox proportional hazard modelling approach in individuals with T2DM (Oulhaj et al., 2023). The study found that a joint model improved predictive performance for prognosis prediction of major adverse cardiovascular events compared to a Cox proportional hazard-based prediction model. However, the study was applied to a multinational controlled trial. Electronic health records data in South Africa's NLHS database was used in this study. Following current literature, this thesis is the first to compare the predictive performance of joint modelling and Cox proportional hazard-based prediction models using electronic health diabetes data from resource-limited settings. The findings indicate comparable discrimination between a dynamic prediction and the Cox proportional hazard-based prediction model (0.63 vs 0.66 AUCs at 12 months; 0.67 vs 0.68 AUCs at 24 months). Similarly, the dynamic prediction and Cox proportional hazard-based prediction model had comparable calibration (0.042 vs 0.058 Brier

scores at 12 months and 0.088 vs 0.089 Brier scores at 24 months). These findings highlight the notable measurement noise in the routine electronic diabetes data, characterized by inconsistent measurements due to variations in timing. The noise may not be adequately captured by estimated individual-specific trajectories leading to glycaemic control from the joint model's longitudinal model, resulting in comparable predictive performance of the model with to that of a Cox proportional hazards-based prediction model.

The joint modeling-based predictions for individuals with T2DM had limitations. The limitations of the routine electronic health records diabetes data were discussed in Chapter 5. However, an overarching limitation included the inability to generalise the findings to other T2DM populations in South Africa because the data only contained the population in the Western Cape.

The strengths regarding the routine diabetes data discussed in Chapter 5 were as follows: i) the large data with substantial follow-up of individuals to enable comprehensive validation, and ii) the use of readily available routinely collected variables in healthcare settings. An additional strength is that the dynamic prediction model enabled individual-specific prognosis prediction of glycaemic control that is updated in real-time when data accumulates. Another major strength is that the individual-specific predictions of glycaemic control can potentially inform personalised approaches to care for T2DM individuals.

In conclusion, this thesis provides the first attempt to compare the predictive performance of dynamic and Cox proportional hazard-based prediction models using routine diabetes data from resource-limited settings. The dynamic prediction model, leveraging longitudinal HbA1c data, offers a promising approach for individualised prognosis prediction of glycaemic control probabilities.

However, the findings indicate that the dynamic prediction model did not necessarily outperform the Cox proportional hazard-based prediction model in terms of predictive performance. These findings are unexpected and highlight the notable challenges in using routine electronic health diabetes data from resource-limited settings where there is a need for personalised strategies to treat diabetes better (Brennan et al., 2023). The findings highlight the need for further research to understand the factors central to predictive performance across different contexts.

6.4 Conclusion

In conclusion, the application of dynamic prediction models for virologic failure in individuals living with HIV and glycaemic control in T2DM individuals represents a novel and significant advancement in the literature of these models, particularly in resource-limited settings. The applications of these models to individualised prognosis prediction of virologic failure or glycaemic control are scarce even in resource-rich settings. In South Africa, these dynamic prediction models for virologic failure and glycaemic control are the first of their kind, particularly using routine data from electronic health records.

The strengths of these dynamic prediction models include their ability to capture the nature of HIV viral load and HbA1c over time and use them to predict individual-specific probabilities of virologic failure and glycaemic control that update in real-time with accumulating data, which is crucial for informing timely therapeutic interventions to improve the quality of life of individuals. These strengths are discussed further in Chapter 8.

Importantly, the findings represent a significant step in utilising routine data from electronic health records to develop dynamic prediction models for virologic failure and glycaemic control. Additionally, they contribute to the literature on dynamic prediction models by providing an understanding of predictive performance between dynamic prediction and Cox proportional-based prediction models in resource-limited settings, where such prediction models are needed to individualise care.

In summary, comparing predictive performance using routine HIV and diabetes data from resource-limited settings offers valuable insights into the universal assumption that dynamic prediction models have improved predictive performance than Cox proportional hazard-based prediction models. The novelty of these models in resource-limited settings warrants ongoing investigation. The findings in this thesis lay a foundation for future advancements in dynamic predictive modelling applied to chronic diseases in resource-limited settings by addressing the limitations of joint modelling-based prediction approaches.

6.5 References

Brennan, A. T., Lauren, E., Bor, J., George, J. A., Chetty, K., Mlisana, K., . . . Crowther, N. J. (2023). Gaps in the type 2 diabetes care cascade: a national perspective using South Africa's National Health Laboratory Service (NHLS) database. *BMC Health Serv Res, 23*(1), 1452. doi:10.1186/s12913-023-10318-9

Mesic, A., Spina, A., Mar, H. T., Thit, P., Decroo, T., Lenglet, A., . . . Oo, H. N. (2021). Predictors of virological failure among people living with HIV receiving first line antiretroviral treatment in Myanmar: retrospective cohort analysis. *AIDS Res Ther, 18*(1), 16. doi:10.1186/s12981-021-00336-0

Oulhaj, A., Aziz, F., Suliman, A., Iqbal, N., Coleman, R. L., Holman, R. R., & Sourij, H. (2023). Joint longitudinal and time-to-event modelling compared with standard Cox modelling in patients with type 2 diabetes with and without established cardiovascular disease: An analysis of the EXSCEL trial. *Diabetes Obes Metab, 25*(5), 1261-1270. doi:10.1111/dom.14975

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*: Chapman and Hall/CRC.

Rizopoulos, D., Papageorgiou, G., & Miranda Afonso, P. (2022). JMbayes2: extended joint models for longitudinal and time-to-event data. *R package version 0.3-0, ed*.

South African National Department of Health. (2014). Management of type 2 diabetes in adults at primary care level. Retrieved from https://knowledgehub.health.gov.za/elibrary/management-type-2-diabetes-adults-primary-care-level

WHO. (2016). Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations–2016 update. Retrieved from https://www.who.int/publications/i/item/9789241511124

Zakaria, H. F., Ayele, T. A., Kebede, S. A., Jaldo, M. M., & Lajore, B. A. (2022). Joint Modeling of Incidence of Unfavorable Outcomes and Change in Viral Load Over Time Among Adult HIV/AIDS Patients on Second-Line Anti-Retroviral Therapy, in Selected Public Hospitals of Addis Ababa, Ethiopia. *HIV AIDS (Auckl), 14*, 341-354. doi:10.2147/hiv.S368373

Chapter 7 Extending a Joint Model for Longitudinal Continuous HIV Viral Load and Time to Virologic Failure Data with Application

7.1 Introduction

In Chapter 5 and Chapter 6, two approaches for prognosis prediction were proposed. Unlike the Cox proportional hazards-based prediction model used in Chapter 5, a joint model applied in Chapter 6 could use continuous biomarkers as time-dependent covariates. However, it was assumed that the continuous biomarker values were reliably measured to a particular detection limit. In practice, not all biomarker values can be reliably measured. For example, in HIV studies, HIV viral load data are often left truncated below the lower limit of quantification and right-skewed. These characteristics may lead to methodological issues when describing HIV viral load data (Brilleman, 2016; Dagne, 2017). In Chapter 6, a Gaussian mixed effects model was used in a joint model to describe longitudinal continuous biomarker data. An extension of the Gaussian mixed effects model is considered in this chapter. Specifically, a two-part joint for longitudinal semicontinuous biomarker and time-to-event data was used to generate predictions of probabilities of virologic failure.

Section 7.2 provides context, motivation, formulation and estimation of a two-part joint model. The performance of the two-part joint model is evaluated in simulated and routine HIV Western Cape data, with illustrative prediction of virologic failure probabilities in Sections 7.3 and 7.4. A summary of the chapter is provided in Section 7.5.

7.2 Two-part Joint Model for Longitudinal Semicontinuous HIV Viral Load and Time to Virologic Failure Data

7.2.1 Context

In biomedical studies, some biomarkers are often characterised as semicontinuous, i.e., they are made up of two components: i) excess values below a limit of detection (commonly known

as zero) and right-skewed positive (non-zero) continuous values (Olsen & Schafer, 2001). These semicontinuous biomarkers are often associated with a time-to-event in biomedical studies. The dependency between longitudinal continuous and time-to-event data was modelled by a shared parameter joint models approach (Henderson, Diggle, & Dobson, 2000; Tsiatis & Davidian, 2004) in Chapter 6. A Gaussian mixed effects model was used in a shared parameter joint model to describe longitudinal continuous biomarker data. However, a Gaussian mixed effects model is not flexible in dealing with longitudinal semicontinuous biomarker data, which are often truncated or right skewed (Brilleman, 2016). Two approaches are often used to model semicontinuous data: (i) the Tobit model (Tobin, 1958) and (ii) the two-part model (TPM) (Manning et al., 1981). The main difference between the two approaches is that, in the Tobit model, zero values are treated as censored observations, while in the TPM, zeros are treated as true observations (Liu et al., 2019). In this thesis, a TPM for the longitudinal semicontinuous biomarker data was considered.

A TPM model was initially proposed by Manning et al. (1981) to account for the distributional characteristics of semicontinuous responses from cross-sectional and later extended to longitudinal data (Olsen & Schafer, 2001). The first part of this model estimates the probability of observing a positive value (i.e., nonzero), while the second part estimates the mean of these positive values (Liu, 2009). Logistic and linear models are commonly used for the first and second parts of the two-part model. Some studies have included two-part longitudinal models for longitudinal semicontinuous biomarkers and time-to-event data in the joint modelling framework (Brilleman, 2016; Dagne, 2017; Liu, 2009).

Studies on a two-part joint modelling framework of longitudinal semicontinuous biomarker and time-to-event data were motivated by medical costs and time-to-event data (Liu, 2009). This study used a logistic mixed effects model for the probability of attaining non-zero costs as the first part of the two-part mixed effects model. A linear mixed effects model was used for non-zero log costs in the second part of the two-part mixed effects model. The two-part mixed effects model and time-to-death components of the joint model were linked by shared random intercepts from the two-part mixed effects model. Similarly, in oncology, random intercepts from each part of the two-part mixed effects component of a two-part joint model of longitudinal semicontinuous patient-reported outcomes and disease progression or death were used from a Bayesian approach (Hatfield, Boye, Hackshaw, & Carlin, 2012). For the joint analysis of a longitudinal semicontinuous HIV viral load and time-to-event data, Brilleman (2016) used a two-part joint model and showed that the longitudinal semicontinuous HIV viral

load and time to stopping or modifying treatment components can be linked by an extension of the current value association structure outlined in Chapter 2. This chapter proposes a two-part joint model to jointly model longitudinal semicontinuous HIV viral load and time to virologic failure data. This two-part joint model is briefly introduced in Section 7.22.

7.2.2 Two-part joint model specification

The notation introduced in Chapter 2 is extended in this chapter. Let w_{ij}, i.e., $I(y_{ij} > 0)$ represent a binary indicator of whether y_{ij} is positive and y_{ij}^{+}, i.e., $y_{ij}|y_{ij} > 0$ represent positive values of y_{ij} for the longitudinal semicontinuous biomarker data. For individual i at the jth observed follow-up time, $\boldsymbol{x}'_{Bij}$ and $\boldsymbol{x}'_{Cij}$ represent vectors of covariates. In the time-to-event data, T_i represents the observed time-to-event such that an event indicator δ_i takes a value of 1 if the event is observed and 0 if censored. For individual i at time t, $\boldsymbol{x}'_{T_i}$ represents the vector of baseline covariates, and $h_i(t)$ is the hazard of an event. To quantify the association between a longitudinal semicontinuous biomarker y_{ij} and the hazard of an event $h_i(t)$, a two-part joint model is introduced.

A two-part mixed-effects model characterised by i) a logistic mixed-effects model for modelling the probability of observing a nonzero value and ii) a Gaussian mixed-effects model for modelling the expected value of positive values were included in the two-part joint model. The two-part mixed-effects model was given as:

$$\begin{cases} \text{logit}\left(p(y_{ij} > 0)\right) = x'_{Bij}\alpha + z'_{Bij}a_i, & \text{(Binary part)} \\ g(y_{ij}^{+}) = E\left[g(y_{ij}^{+})\right] + \epsilon_{ij} = x'_{Cij}\beta + z'_{Cij}b_i + \epsilon_{ij}. & \text{(Continuous part)} \end{cases} \quad 7.1$$

where $g(\cdot)$ is a non-linear transformation that corrects for Gaussian departures (Su, Tom, & Farewell, 2009). In the formulation α and β represent fixed effects parameters for the vector of covariates in the binary (x'_{Bij}) and continuous (x'_{Cij}) parts, a_i and b_i represent random effects for the vector of covariates in the binary (z'_{Bij}) and continuous (z'_{Cij}) parts, and ϵ_{ij} represent the error term in the continuous part and assumed to be normally and independently distributed with mean zero and variance σ_ϵ^2, i.e., $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$. The correlation between the binary and continuous parts and between repeated biomarker measurements within an individual is

captured by the random effects a_i and b_i. These random effects have a multivariate normal with a mean vector of zero and variance-covariance matrix D, i.e.,

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b_i} & \rho\sigma_{b_i}\sigma_{a_i} \\ \rho\sigma_{b_i}\sigma_{a_i} & \sigma_{a_i} \end{bmatrix}\right). \quad 7.2$$

Conditional on the random effects, the binary and continuous parts are assumed to be independent (Brilleman, 2016). Thus, the expected value of a biomarker is given as:

$$p(y_{ij} > 0) \times E\left[g(y_{ij}^{+})\right] = \left(\frac{\exp\left(x'_{Bij}\alpha + z'_{Bij}a_i\right)}{1 + \exp\left(\alpha x'_{Bij} + a_i z'_{Bij}\right)}\right) \times \left(x'_{Cij}\beta + z'_{Cij}b_i\right). \quad 7.3$$

The formulation in Equation 7.3 is a current value association structure outlined in Chapter 2. Other association structures, for example, shared random effects and the sum of the current probability of nonzero value and expected positive value, have been proposed for two-part joint models (Rustand, Briollais, Tournigand, & Rondeau, 2022). In this thesis, the focus was on the current value association structure. Thus, using the expected value of a biomarker given in Equation 7.3, a two-part joint model assuming a current value association structure (product of the current probability of observing a nonzero value and expected value of positive values) is given by:

$$h_i(t) = h_0(t)\exp(\gamma x'_{Ti} + \varphi p(y_i(t) > 0) \times E[y_i^{+}(t)]). \quad 7.4$$

was used to quantity the association φ between the expected value of a biomarker, i.e., $p(y_i(t) > 0) \times E[y_i^{+}(t)]$ and hazard of an event $h_i(t)$ at time t. This two-part joint model includes a baseline hazard function $h_0(t)$, a vector of baseline covariates x'_{Ti} with corresponding regression parameter vector γ.

7.2.3 Estimation of effects in two-part joint models

Let ϕ represent all parameters included in the binary and continuous parts of the longitudinal semicontinuous biomarker process, variance-covariance components, and time-to-event process. The joint distribution full likelihood of the observed longitudinal semicontinuous biomarker measurements y_{ij}, random effects a_i and b_i, time-to-event T_i, and event indicator δ_i for an ith individual is given by:

$$L_i(\phi) = \int_{a_i}\int_{b_i}\left(\prod_{j=1}^{n_i}\left(\exp\left(x'_{Bij}\alpha + z'_{Bij}a_i\right)\right)^{w_{ij}}\left(1 - \frac{\exp\left(x'_{Bij}\alpha + z'_{Bij}a_i\right)}{1+\exp\left(\alpha x'_{Bij} + a_i z'_{Bij}\right)}\right)\right. \tag{7.5}$$
$$\times\left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}}\exp\left(-\frac{\left(y_{ij}^{+} - \left(x'_{Cij}\beta + z'_{Cij}b_i\right)\right)^2}{2\sigma_\epsilon^2}\right)\right)^{w_{ij}}$$
$$\left.\times\, h_i(T_i|\Phi_i)^{\delta_i}\exp\left(-\int_0^{T_i} h_i(t|\phi_i)\mathrm{d}t\right)p(a_i,b_i)\right)db_i da_i.$$

The joint distribution full likelihood is an intractable and requires numerical approximations outlined in Chapter 2. Specifically, a Gauss-Legendre quadrature (Abbott, 2005) is commonly used to estimate parameters in the joint distribution full likelihood of the two-part joint model. Following the estimation of model parameters in the two-part joint model, predictions under joint models are extended to two-part joint modelling in Section 7.2.4.

7.2.4 Prediction under two-part joint models

Predictions under joint models relying on Bayesian inference were introduced in Chapter 2 and illustrated in Chapter 6. Using MCMC sampling, the posterior distribution of event-free probability at a future time window $u > t$ was given as:

$$\pi_j(u|t) = \int P\left(T_j^* \geq u | T_j^* > t, \mathcal{y}_j(t); \theta\right)p(\theta|\mathcal{D}_n)d\theta.$$

Under two-part joint models, $\mathcal{y}_j(t)$ represents observed longitudinal semicontinuous biomarker values for individual j up to observed follow-up time t and θ represents the overall

parameter vector containing parameters from the binary and continuous parts of the longitudinal semicontinuous process and parameters from the time-to-event process. Accordingly, based on the Monte Carlo simulation scheme, the event-free conditional probabilities for an individual $\hat{\pi}_j(u|t)$ can be estimated using the same procedure given in Chapter 2.

7.3 Simulation and Application of Two-part Joint Models to Simulated Longitudinal Semicontinuous and Time-to-Event Data

This section presents a simulation study conducted to evaluate the performance of the proposed two-part joint model in the context of longitudinal semicontinuous biomarker and time-to-event data. In particular, the focus was to assess whether a two-part joint model flexibility provides better inference and predictive performance than a joint model in data subject to limit detection. Three simulation scenarios with different proportions (low, moderate, and high) of values below the limit of detection were used to generate data motivated by the routine HIV Western Cape data described in Chapter 4. For each scenario, 100 datasets were generated for $n = 400$ individuals in R version 4.0.5. These data were randomly divided into development and validation cohorts for model development and validation. Specifically, the inference was based on the development cohort of each simulation scenario, and predictive performance was based on the validation cohort of each simulation scenario.

7.3.1 Data generation

In each simulation scenario, individuals were assumed to be followed-up for ten years or until the time-to-event with semicontinuous biomarker values measured every 12 months. A binary covariate (sex) was used in the longitudinal semicontinuous and time-to-event processes. This binary covariate was generated from the Bernoulli distribution with a 0.4 probability of success. The two-part joint model described in Section 7.2, which considers random intercepts in the binary part and random intercepts and slopes in the continuous part, was used for data generation. This model assumed a current value association structure to link the longitudinal semicontinuous and time-to-event processes, and it is given by:

$$\begin{cases} \text{logit}\left(p(y_{ij}>0)\right) = (\alpha_0 + a_{0i}) + \alpha_1 \text{t}_{ij} + \alpha_2 \text{sex}, & \text{(Binary part)} \\ y_{ij}^{+} = E\left[y_{ij}^{+}\right] + \epsilon_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})\text{t}_{ij} + \beta_2 \text{sex} + \epsilon_{ij}, & \text{(Continuous part)} \\ h_i(t) = h_0(t)\exp(\gamma \text{sex} + \varphi p(y_i(t)>0) \times E[y_i^{+}(t)]). & \text{(Time to event process)} \end{cases} \quad 7.6$$

where random effects for each individual were described as follows:

$$\begin{pmatrix} b_{0i} \\ b_{1i} \\ a_{0i} \end{pmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, D\right) \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b_{0i}} & \rho\sigma_{b_{0i}}\sigma_{b_{1i}} & \rho\sigma_{b_{0i}}\sigma_{a_{0i}} \\ \rho\sigma_{b_{0i}}\sigma_{b_{1i}} & \sigma_{b_{1i}} & \rho\sigma_{b_{1i}}\sigma_{a_{0i}} \\ \rho\sigma_{b_{0i}}\sigma_{a_{0i}} & \rho\sigma_{b_{1i}}\sigma_{a_{0i}} & \sigma_{a_{0i}} \end{bmatrix}\right), \text{ with}$$

$$D = \begin{bmatrix} 0.45 & 0.2 & 0.2 \\ 0.2 & 0.25 & 0.7 \\ 0.2 & 0.7 & 0.5 \end{bmatrix}.$$

The R package mvtnorm (Genz et al., 2021) was used to generate random effects from a Gaussian distribution, a zero-mean vector and the D variance covariance matrix. Given the values of the random effects, continuous longitudinal data y_{ij}^{+} were generated from a Gaussian distribution with mean $E\left[y_{ij}^{+}\right]$ and variance $\sigma_{\epsilon_{ij}}^2 = 0.6$. In the mean $E\left[y_{ij}^{+}\right]$, the regression coefficients were given as:

$$\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 3 \\ -0.3 \\ -1 \end{pmatrix}.$$

For the binary longitudinal data, i.e., an indicator for whether a continuous value was positive or not (below or above the limit of detection) was generated from a Bernoulli distribution with probability of success $p(y_{ij}>0)$ to mimic varying proportions of values below the limit of detection. Thus, three simulation scenarios, namely low, moderate, and high proportions of values below the limit of detection, were considered to generate longitudinal semicontinuous biomarker and time-to-event data.

Scenario 1: Data for a high proportion of values below the limit of detection, which was regarded as Scenario 1, were generated by assuming the intercept $\alpha_0 = 1$ (representing the log odds of observing a continuous value above zero). Thus, the regression coefficients for the binary part's probability of success $p(y_{ij}>0)$ are given as:

$$\begin{pmatrix}\alpha_0\\ \alpha_1\\ \alpha_2\end{pmatrix} = \begin{pmatrix}1\\ -1\\ -1\end{pmatrix}.$$

The event times T_i were generated from the R package PermAlgo (Sylvestre, Edens, MacKenzie, Abrahamowicz, & Sylvestre, 2010) conditional on the binary covariate and time-dependent semicontinuous values. The regression coefficients for the binary covariate and time-dependent semicontinuous values were given as follows:

$$\begin{pmatrix}\gamma\\ \varphi\end{pmatrix} = \begin{pmatrix}-1\\ 0.1\end{pmatrix}.$$

Non-informative censoring times C_i were generated from a uniform distribution ranging between 1 and 10. Under this setting, the proportion of individuals with an event during follow-up time was 27%. Any longitudinal semicontinuous measurements observed after the observed event time T_i were excluded.

Scenario 2: Keeping the same settings in simulation Scenario 1, data for a moderate proportion of values below the limit of detection, which was regarded as Scenario 2, were generated by assuming the intercept $\alpha_0 = 2.5$. Under this scenario, 27% of the individuals had the event during follow-up time.

Scenario 3: With the same settings in simulation Scenario 1, data for the final simulation (Scenario 3) with a low proportion of values below the limit of detection were generated by assuming the intercept $\alpha_0 = 5$. Under this scenario, 27% of the individuals had the event during follow-up time.

7.3.2 Methods considered, estimation and predictive performance

For each generated dataset under each simulation scenario, the interest was on the parameters and measures of the predictive performance of joint and two-part joint models. These parameter estimates and predictive performance measures were averaged across the 100 datasets generated in each simulation scenario. Specifically, the standard error (SE), mean squared error (MSE), and coverage probability were calculated to evaluate model performance in inference. The true values of the parameters used for data generation are given in Table 7.1. Average time-

dependent AUCs and Brier scores were used to summarise the predictive performance of the models in each simulation scenario.

Table 7.1: True values of parameters for data generation in simulation studies.

	Scenario 1	**Scenario 2**	**Scenario 3**
Parameter	**True[1]**	**True[1]**	**True[1]**
α_0	1	2.5	5
α_1	-1	-1	-1
α_2	-1	-1	-1
β_0	3	3	3
β_1	-0.3	-0.3	-0.3
β_2	-1	-1	-1
γ_1	-1	-1	-1
φ	0.1	0.1	0.1
σ_ϵ	0.6	0.6	0.6
$\sigma_{a_{0i}}$	0.5	0.5	0.5
$\sigma_{b_{0i}}$	0.45	0.45	0.45
$\sigma_{b_{1i}}$	0.25	0.25	0.25

True[1]: True parameter

Bayesian inference

MCMC posterior sampling based on the no-U-turn sampler (NUTS) (Hoffman & Gelman, 2014) on R package Rstan (Stan Developent Team, 2016) was used to estimate unknown parameters of the two-part joint model. R was used for data manipulation, and Stan statistical software through Rstan was used for MCMC posterior sampling. The Stan and R code used to fit the two-part joint model are given in Appendix 8 and Appendix 9. Non-informative priors (Gelman, Simpson, & Betancourt, 2017) were adopted for regression coefficient vectors (α, β, γ) and association parameter φ. Specifically, a normal distribution with mean 0 and scale 1 was used for the regression parameters and association parameter. A gamma prior with shape 2 and scale 1 was used for the shape parameter of the Weibull distribution baseline hazard in the two-part joint model. To enable efficient exploration of the parameter space for the variance

of the error term (σ_ϵ), non-centred parametrisation (Betancourt & Girolami, 2015) was adopted. Lewandowski-Kurowicka-Joe (LKJ) correlation prior (Lewandowski, Kurowicka & Joe, 2009) with shape 1 was adopted for the Cholesky decomposition of the combined random effects. r_i variance-covariance matrix D parameterisation. The convergence of posterior samples of the all-unknown parameters of the two-part joint model was assessed using trace plots and potential scale reduction factor $\hat{R}$ (Gelman & Rubin, 1992).

Assessing predictive performance

As in Chapter 5 and Chapter 6, the predictive performance of the two-part joint model was assessed using discrimination (discriminate between individuals who will experience virologic failure and those who will not) and calibration (how will the two-part joint model predict the observed virologic failures). Specifically, time-dependent AUCs were used for discrimination and expected Brier scores were used for calibration. The R packages survival ROC (Heagerty, Saha-Chaudhuri, & Saha-Chaudhuri, 2013) and pec (Mogensen, Ishwaran, & Gerds, 2012) were used to estimate these time-dependent AUCs and Brier scores, respectively.

7.3.3 Simulation results

The results of the simulation scenarios with 100 replications in inference and predictive ability are summarised in Table 7.2, Table 7.3 and Table 7.4 and Figure 7.1 and Figure 7.2. The R package JMBayes2 and Stan were used to fit the joint and two-part joint models, respectively. In both JMBayes2 and Stan, three chains, each with 1000 after warm-up iterations (1000 warm-up iterations), were used. The joint and two-part joint models were run on a machine with a 3.6 GHz (Intel Core i7) processor and 16 GB RAM running 64-bit Windows 10 Enterprise (version 21H2). The computing time to fit and evaluate the predictive performance of the joint introduced in Chapter 2 and the two-part joint (Equation 7.4) models under simulation Scenario 1 was 3.5 hours and 25.6 hours, respectively. Under Scenario 2, the computing time for the joint model was 4.6 hours and 26.4 hours for the two-part joint model. In Scenario 3, the computing time for the joint and two-part joint models were 2.8 hours and 28.8 hours, respectively. In all simulation scenarios, a joint model was made up of i) a Gaussian mixed effects model with semicontinuous values as the outcome observed follow-up time and binary covariate in the fixed effects and random intercepts and slopes in the random effects, and ii) a

proportional hazards model with a binary covariate and assuming a B-spline baseline hazard and a current value association structure compare to the two-part joint model given in Equation 7.4.

Table 7.2, Table 7.3 and Table 7.4 present the inference results for each simulation scenario. In each scenario, the distribution of continuous values below the limit of detection was not included in the joint model. Thus, the comparison focus for parameter estimates between the joint and two-part joint models should be on the time-to-event process (Rustand et al., 2022). As a result, the effect of the binary covariate γ_1 and association parameter φ between the longitudinal and time-to-event processes were compared. The inference results under simulation Scenario 1 with a high proportion of values below the limit of detection are given in Table 7.2. Under this scenario, the two-part joint model outperformed the joint model concerning SEs and MSEs for the time-to-event process. It can be observed that the two-part joint model tends to have smaller SEs and MSEs.

Table 7.3 and Table 7.4 summarise the simulation results under scenarios 2 and 3, respectively. These simulation scenarios were used to check if the proportion of values below the limit of detection poorly affects the performance of the joint and two-part joint models in inference. Under all simulation scenarios (different proportions of values below the limit of detection), the two-part joint model outperformed the joint model in SEs and MSEs.

Table 7.2: Simulation results for Scenario 1 with a high proportion (73% zeros) of values below the limit of detection.

		Standard Joint model			Two-part Joint model		
Parameter	**True[1]**	**Est.[2]**	**SE[3]**	**MSE[4]**	**Est.[2]**	**SE[3]**	**MSE[4]**
Longitudinal process							
Binary part							
α_0	1				0.963	0.0026	0.0173
α_1	-1				-1.014	0.0013	0.0027
α_2	-1				-0.972	0.0029	0.0261
Continuous part							
β_0	3	1.488	0.0243	2.290	2.971	0.0009	0.0041
β_1	-0.3	-0.249	0.0061	0.003	-0.210	0.0006	0.0111
β_2	-1	-0.388	0.0102	0.378	-0.930	0.0013	0.0242
Time-to-event process							
γ_1	-1	-1.089	0.0272	0.097	-0.992	0.0991	0.0020
φ	0.1	-0.090	0.0496	0.157	0.036	0.0005	0.0042
Standard deviations							
σ_{ϵ}	0.6	0.924	0.0036	0.106	0.602	0.00004	0.0042
$\sigma_{a_{0i}}$	0.5				0.502	0.0004	0.0002
$\sigma_{b_{0i}}$	0.45	0.532	0.532	0.117	0.456	0.0003	0.0004
$\sigma_{b_{1i}}$	0.25	0.024	0.024	0.001	0.249	0.0002	0.0001

True[1]: True parameter; Est.[2]: Average of the parameter estimates; MSE[3]: Standard error; MSE[4]: Mean squared error

Table 7.3: Simulation results for Scenario 2 with a moderate proportion (52% zeros) of values below the limit of detection.

		Standard Joint model			Two-part Joint model		
Parameter	**True[1]**	**Est.[2]**	**SE[3]**	**MSE[4]**	**Est.[2]**	**SE[3]**	**MSE[4]**
Longitudinal process							
Binary part							
α_0	2.5				2.488	0.0029	0.0231
α_1	-1				-1.014	0.0010	0.0024
α_2	-1				-0.995	0.0026	0.0334
Continuous part							
β_0	3	2.340	0.008	0.440	2.950	0.0007	0.0042
β_1	-0.3	-0.361	0.003	0.004	-0.203	0.0004	0.0110
β_2	-1	-0.695	0.007	0.100	-0.957	0.0009	0.0112
Time-to-event process							
γ_1	-1	-1.294	0.033	0.178	-0.998	0.099	0.0027
φ	0.1	-0.426	0.069	0.412	0.014	0.0002	0.0075
Standard deviations							
σ_{ϵ}	0.6	0.887	0.001	0.083	0.607	0.00004	0.0075
$\sigma_{a_{0i}}$	0.5				0.496	0.0004	0.0003
$\sigma_{b_{0i}}$	0.45	0.321	0.321	0.019	0.451	0.0003	0.0005
$\sigma_{b_{1i}}$	0.25	0.014	0.014	0.002	0.250	0.0002	0.00005

True[1]: True parameter; Est.[2]: Average of the parameter estimates; MSE[3]: Standard error; MSE[4]: Mean squared error

Table 7.4: Simulation results for Scenario 3 with a low proportion (24% zeros) of values below the limit of detection.

		Standard Joint model			Two-part Joint model		
Parameter	**True[1]**	**Est.[2]**	**SE[3]**	**MSE[4]**	**Est.[2]**	**SE[3]**	**MSE[4]**
Longitudinal process							
Binary part							
α_0	5				4.791	0.0048	0.0676
α_1	-1				-1.009	0.0010	0.0025
α_2	-1				-1.046	0.0029	0.0207
Continuous part							
β_0	3	2.966	0.003	0.003	2.972	0.0006	0.0023
β_1	-0.3	-0.344	0.001	0.002	-0.274	0.0002	0.0018
β_2	-1	-1.027	0.004	0.007	-0.979	0.0006	0.0037
Time-to-event process							
γ_1	-1	-1.024	0.029	0.077	-0.993	0.1004	0.0016
φ	0.1	0.016	0.031	0.049	0.006	0.0001	0.0089
Standard deviations							
σ_{ϵ}	0.6	0.614	0.001	0.002	0.596	0.00003	0.0089
$\sigma_{a_{0i}}$	0.5				0.507	0.0004	0.0003
$\sigma_{b_{0i}}$	0.45	0.213	0.213	0.001	0.451	0.0003	0.0002
$\sigma_{b_{1i}}$	0.25	0.013	0.013	0.002	0.253	0.0002	0.0001

True[1]: True parameter; Est.[2]: Average of the parameter estimates; MSE[3]: Standard error; MSE[4]: Mean squared error

Predictive performance of a two-part joint model to simulate longitudinal semicontinuous and time-to-event data

Time-dependent AUCs and Brier scores were used to summarise the predictive performance of the models under each simulation scenario. Given that individuals in the development cohort were event-free during the first 12 months of follow-up, two future time windows (12 and 24 months) were chosen to assess the predictive performance. Specifically, time-dependent AUCs and Brier scores were used to evaluate the predictive performance of the proposed two-part joint model and compare its predictive performance with a joint model. These AUCs and Brier scores within given future time windows were averaged across 100 datasets in each scenario. Figure 7.1 and Figure 7.2 provide the AUCs and Brier scores under all simulation scenarios.

These scenarios were used to check if the proportion of values below the limit of detection poorly affects the predictive ability of the joint and two-part joint models. Under all simulation scenarios, the two-part joint model performed well and outperformed the joint model concerning discrimination within 12 and 24-month future time windows (Figure 7.1). It can be observed that Brier scores were smaller for the joint model, suggesting good calibration for this model (Figure 7.2). Nevertheless, the Brier scores for the two-part joint model were still acceptable (less than 0.25, which represents a dummy prediction). Following the comparison in both inference and predictive performance of the joint and two-part joint models under all scenarios, the following section (Section 7.4) illustrates the application of the two-part joint model to the routine HIV Western Cape data.

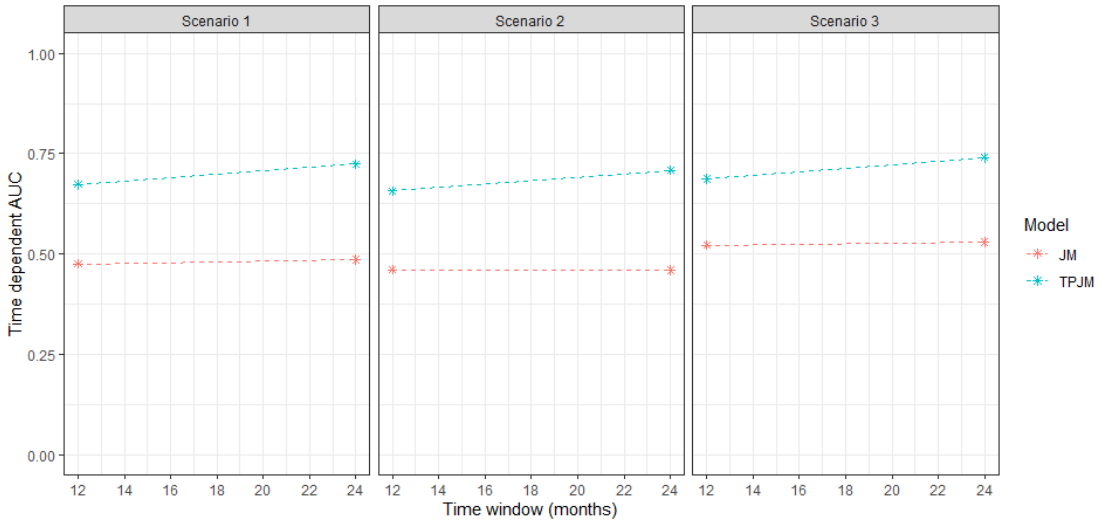

Figure 7.1: Comparison of time-dependent AUCs between joint and two-part joint models under three simulation scenarios: i) low, ii) moderate, and iii) high proportion of values below the limit of detection.

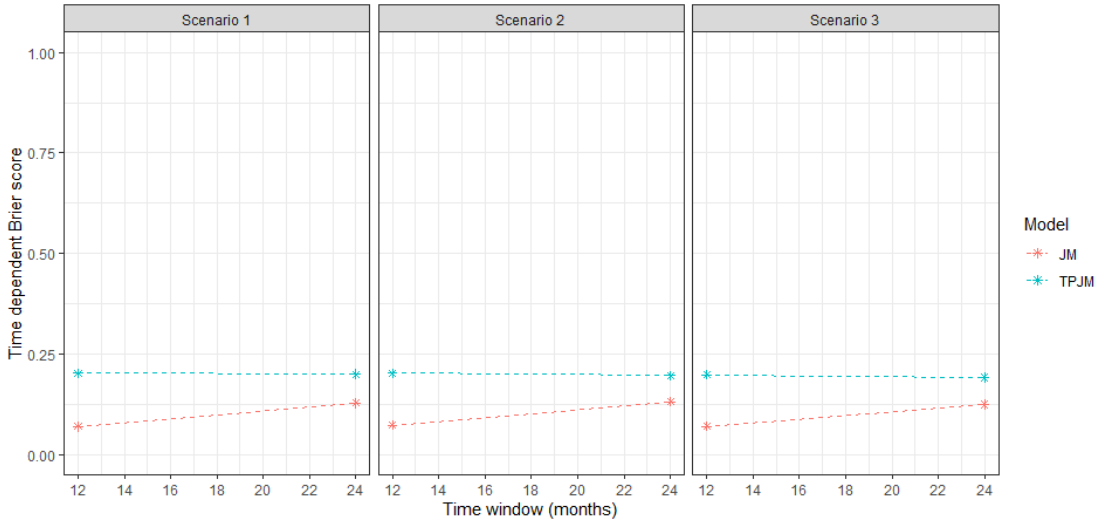

Figure 7.2: Comparison of time-dependent Brier scores between standard and two-part joint models under three simulation scenarios: i) low, ii) moderate, and iii) high proportion of values below the limit of detection.

7.4 Application of Two-part Joint Models to Longitudinal Semicontinuous HIV Viral Load and Time to Virologic Failure Data

This Section illustrates the application of the proposed two-part joint model to the motivating routine HIV Western Cape data. Only 1% (4858) of the individuals in the motivating data were used due to the computation costs of fitting the two-part joint model in Stan. The resulting data were randomly split into development (3613) and validation (1245) cohorts for model development and validation, respectively. The development cohort was used for inference, and the validation cohort was used for predictive performance. The two-part joint model was evaluated in inference and predictive performance in sub-sections 7.4.1 and 7.4.2, respectively. Prediction under this model is presented in sub-section 7.4.3.

7.4.1 Two-part joint model for longitudinal semicontinuous HIV viral load and time to virologic failure data

The motivating routine HIV viral load data from the Western Cape were considered as semicontinuous, i.e., containing excess (81%) HIV viral load values below the limit of detection of 1.7 log base 10 copies/mL, as mentioned in Chapter 4. Figure 7.3 presents log base 10 HIV viral load values in which the top panel represents all log base 10 HIV viral load values, and the bottom panel represents those above the limit of detection. It can be seen in the plot (top panel) that the HIV viral load values are skewed to the right and a large portion of them are below the limit of detection. When HIV viral load values below the limit of detection were excluded, the histogram (bottom panel) appeared to be slightly normal. The interest lies in examining the association between HIV viral load and the hazard of virologic failure. In Figure 7.4, longitudinal trajectories for randomly selected individuals in the motivating routine HIV Western Cape development cohort are presented. It can be seen that there is variation in the trajectories of the individuals over time. This suggests that a two-part mixed-effects model needs random intercepts and slopes to account for between-individual variation.

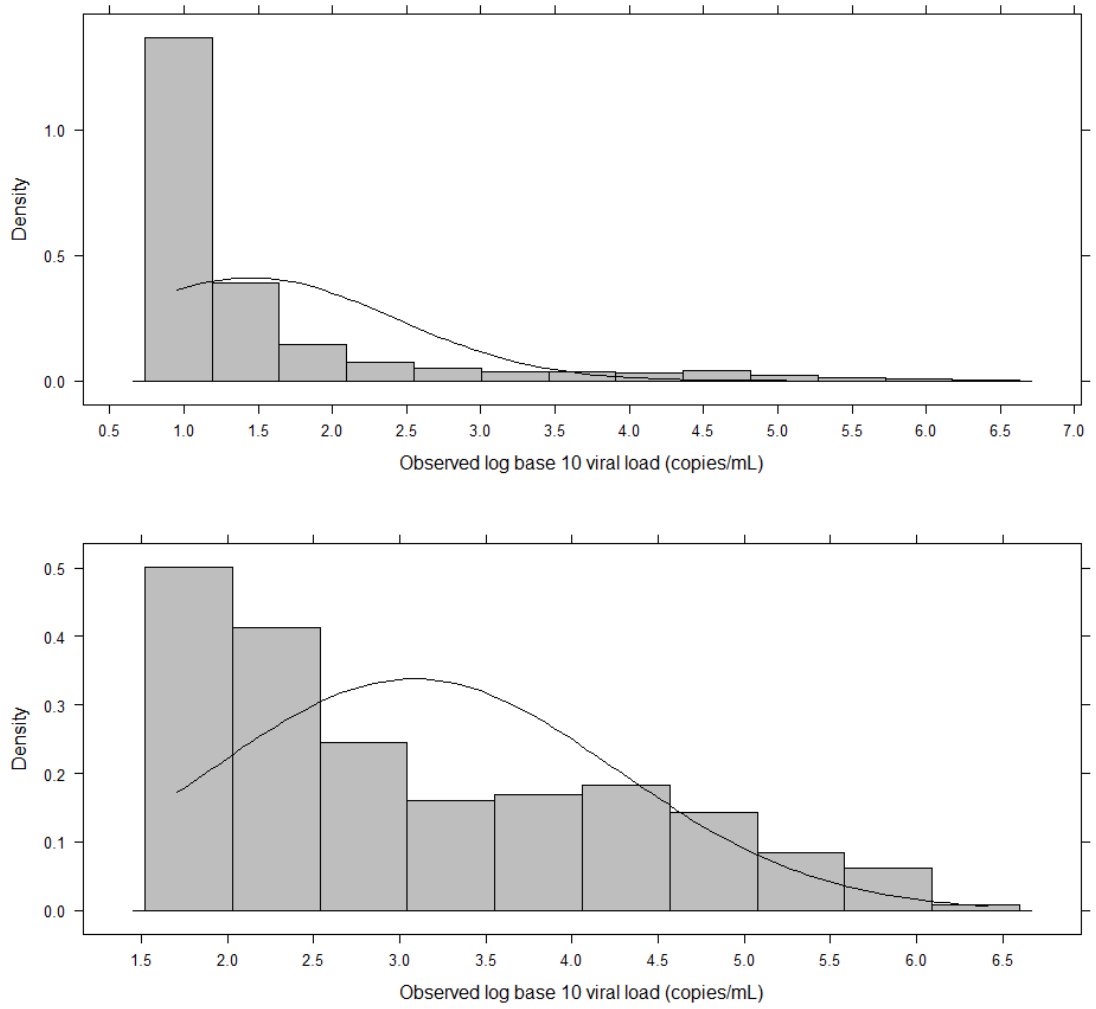

Figure 7.3: Observed log base 10 HIV viral load (copies/mL) values in the routine HIV Western Cape development cohort.

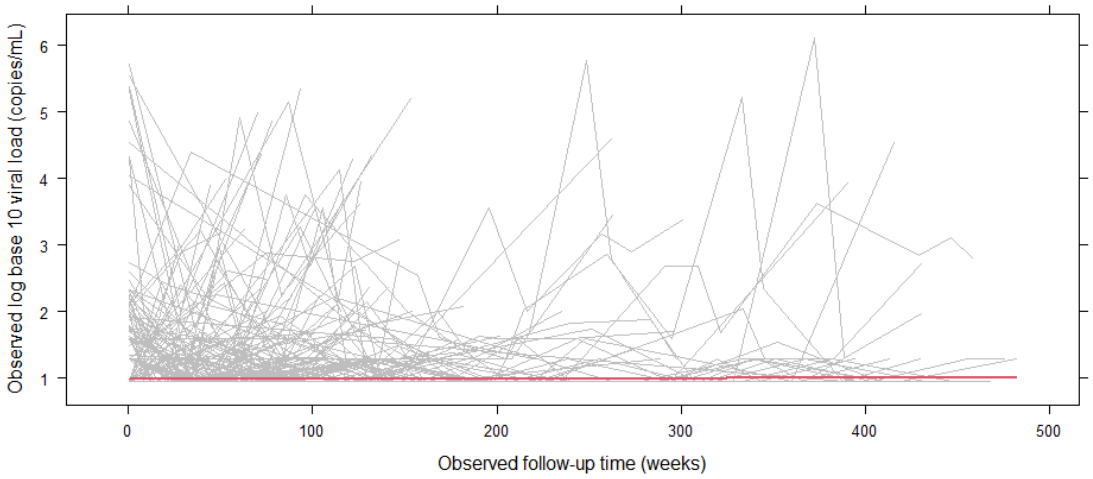

Figure 7.4: Observed longitudinal HIV viral load trajectories for randomly selected individuals in the routine HIV Western Cape subset development cohort.

Two-part mixed-effects for longitudinal semicontinuous HIV viral load data

A two-part mixed-effects model characterised by i) a logistic mixed-effects model and ii) a Gaussian mixed-effects model for modelling the expected value of HIV viral load values above the limit of detection was used. Specifically, a two-part mixed-effects model containing i) random intercepts in the logistic mixed-effects model for the binary part and ii) random intercepts and slopes in the Gaussian mixed-effects model had the following specification:

$$\begin{cases} \text{logit}\left(p\left(\log_{10}\text{vload}_{ij} > 1.7\right)\right) = (\alpha_0 + a_{0i}) + \alpha_1 \text{t}_{ij} + \alpha_2 \text{sex}, \text{(Binary part)} \\ \log_{10}\text{vload}_{ij}^{+} = E\left[\log_{10}\text{vload}_{ij}^{+}\right] + \epsilon_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})\text{t}_{ij} + \beta_2 \text{sex} + \epsilon_{ij}. \text{(Continuous part)} \end{cases} \quad 7.7$$

where α_0 and β_0 are intercepts, α_1 and β_1 are the effects of observed follow-up time, α_2 and β_2 are regression coefficients for a binary covariate contained in both binary and continuous parts, and ϵ_{ij} is the error term assumed to be normally distributed with mean 0 and variance σ_ϵ^2. On the other hand, a_{0i} are random intercepts in the binary part and b_{0i} and b_{1i} are random intercepts and slopes in the continuous part. These random effects have a multivariate normal with a mean vector of 0 and variance-covariance matrix D, i.e.,

$$\begin{pmatrix} b_{0i} \\ b_{1i} \\ a_{0i} \end{pmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, D\right) \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b_{0i}} & \rho\sigma_{b_{0i}}\sigma_{b_{1i}} & \rho\sigma_{b_{0i}}\sigma_{a_{0i}} \\ \rho\sigma_{b_{0i}}\sigma_{b_{1i}} & \sigma_{b_{1i}} & \rho\sigma_{b_{1i}}\sigma_{a_{0i}} \\ \rho\sigma_{b_{0i}}\sigma_{a_{0i}} & \rho\sigma_{b_{1i}}\sigma_{a_{0i}} & \sigma_{a_{0i}} \end{bmatrix}\right).$$

A two-part joint model for longitudinal semicontinuous HIV viral load and time to virologic failure data

The estimated longitudinal semicontinuous HIV viral load trajectory for individuals from the two-part mixed-effects model containing random intercepts in the logistic mixed-effects model binary part and random intercepts and slopes in the Gaussian mixed-effects model continuous part was used in a two-part joint model. The two-part joint model with a current value association structure was given by:

$$h_i(t) = h_0(t)\exp\big(\gamma \text{sex} + \varphi p(\log_{10} \text{vload}_i(t) > 1.7) \times E\big[\log_{10} \text{vload}_{ij}^{+}(t)\big]\big), \qquad 7.8$$

where used φ quantity is the association between the expected value of HIV viral load, i.e., $p(\log_{10} \text{vload}_i(t) > 1.7) \times E\big[\log_{10} \text{vload}_{ij}^{+}(t)\big]$ and hazard of virologic failure $h_i(t)$. This model included a binary covariate and a Weibull baseline hazard $h_0(t)$ given by:

$$h_0(t) = \lambda t^{\lambda-1},$$

where λ is the shape parameter of the Weibull distribution. This Weibull baseline hazard allowed for virologic failure times to change over time.

Bayesian inference

As described in Section 7.3, MCMC posterior sampling in Stan was used to estimate each parameter in the two-part joint model. The Stan code used to fit the model is given in Appendix 8. The two-part joint model was run with three chains, each with 1000 after warm-up iterations (1000 warm-up iterations) on a machine with a 3.6 GHz (Intel Core i7) processor and 16 GB RAM running 64-bit Windows 10 Enterprise (version 21H2). The computation time to fit and evaluate the predictive performance of the model was 1.7 hours. Rhat values (between 1.00 – 1.01) for each parameter and trace plots in Figure 7.5, Figure 7.6, Figure 7.7, Figure 7.8, Figure 7.9, Figure 7.10 and Figure 7.11 indicated convergence of the posterior samples. The parameter

estimates of the two-part joint model applied to the routine HIV Western Cape subset development cohort are presented in Table 7.5. Unlike in Section 7.3, HIV viral load values had a weak association with the hazard of virologic failure. Specifically, an expected HIV viral load above the limit of detection tends to increase the risk of virologic failure by a very small and negligible fold (HR = 1.0011; 95% CI, 1.000, 1.0037). In sub-section 7.4.2, the predictive performance of the two-part joint model was evaluated in the subset validation cohort.

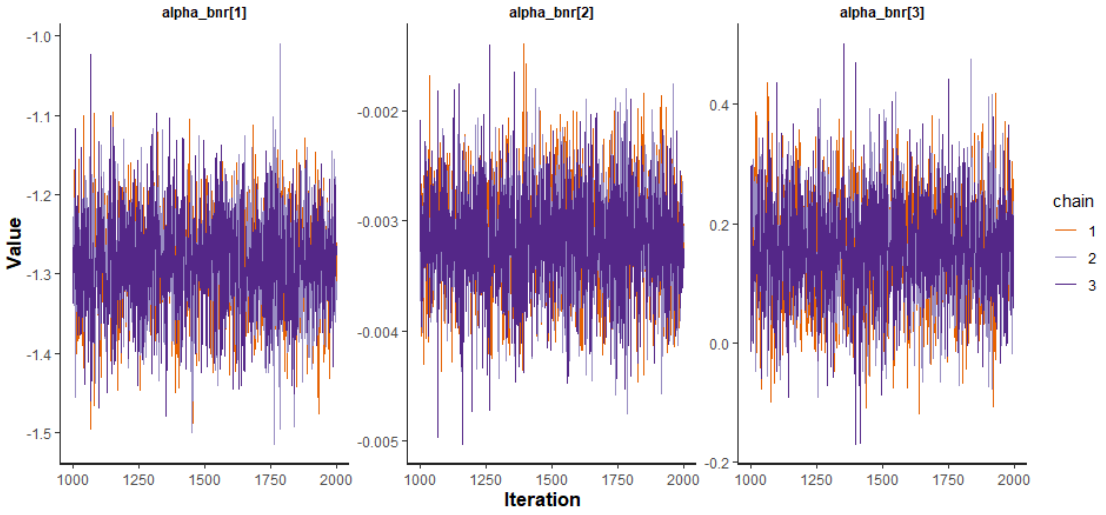

Figure 7.5: Trace plots for alpha parameters of the two-part joint model applied to the routine HIV Western Cape subset development cohort.

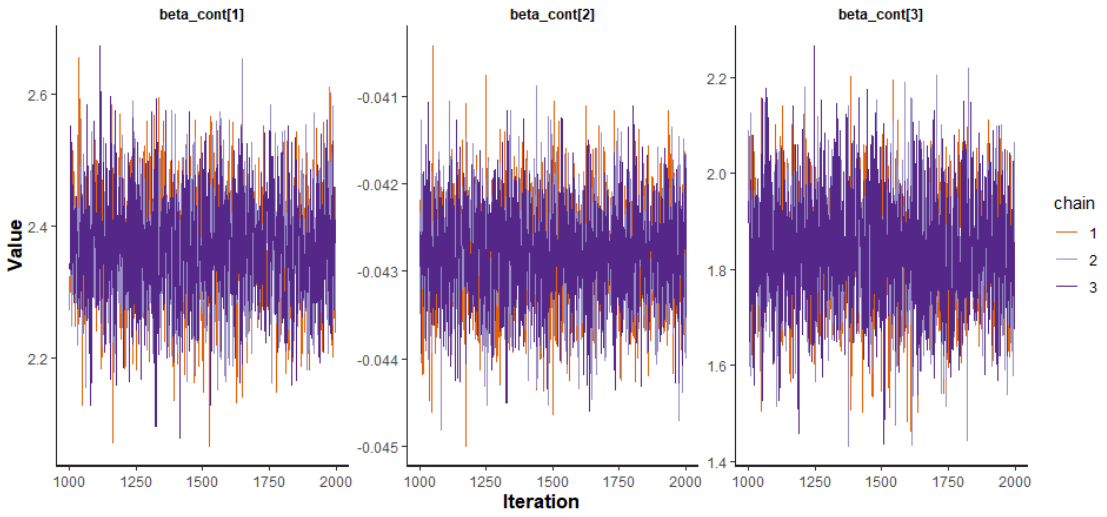

Figure 7.6: Trace plots for beta parameters of the two-part joint model applied to the routine HIV Western Cape subset development cohort.

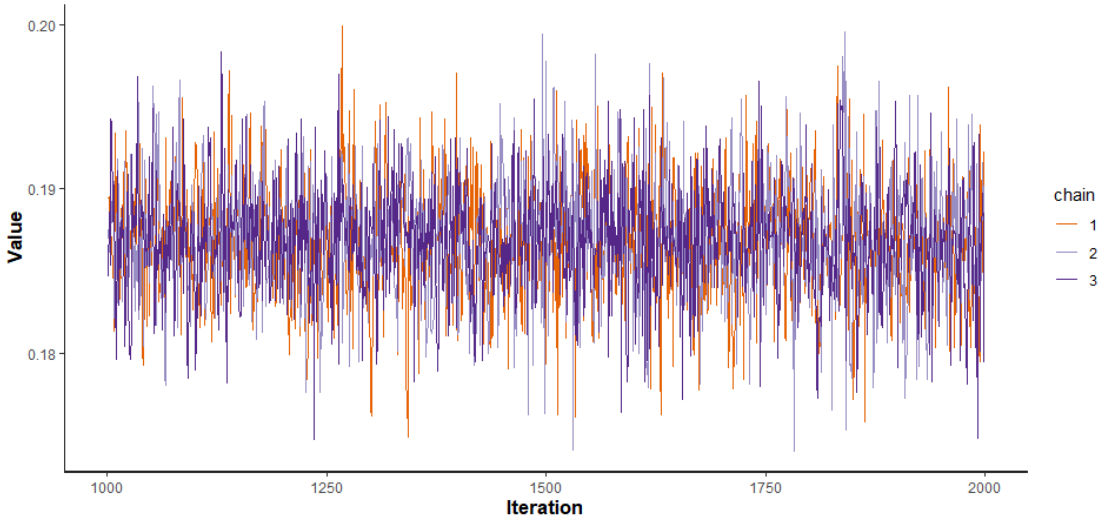

Figure 7.7: Trace plot of the Weibull shape parameter of the two-part joint model applied to the routine HIV Western Cape subset development cohort.

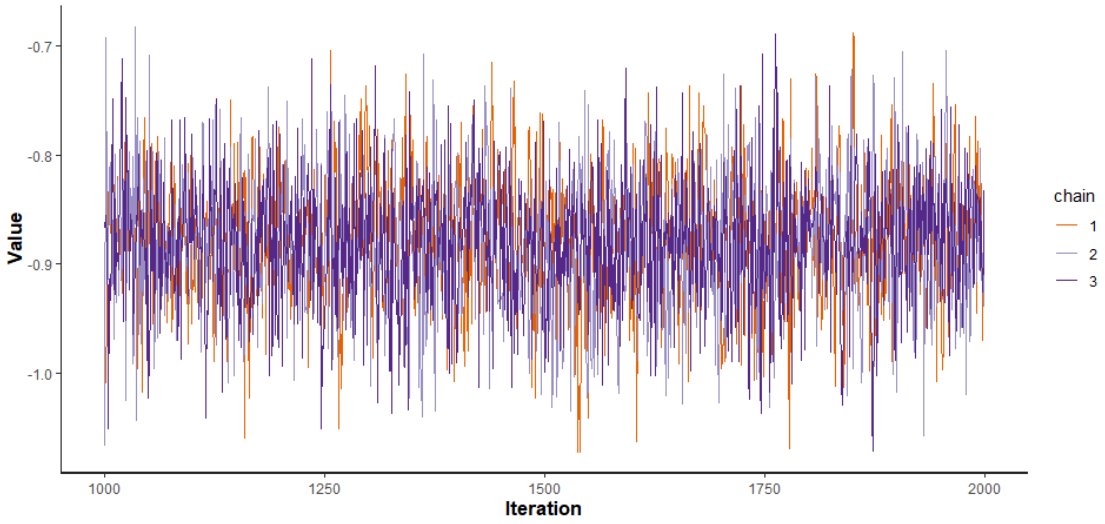

Figure 7.8: Trace plot of the gamma parameter of the two-part joint model applied to the routine HIV Western Cape subset development cohort.

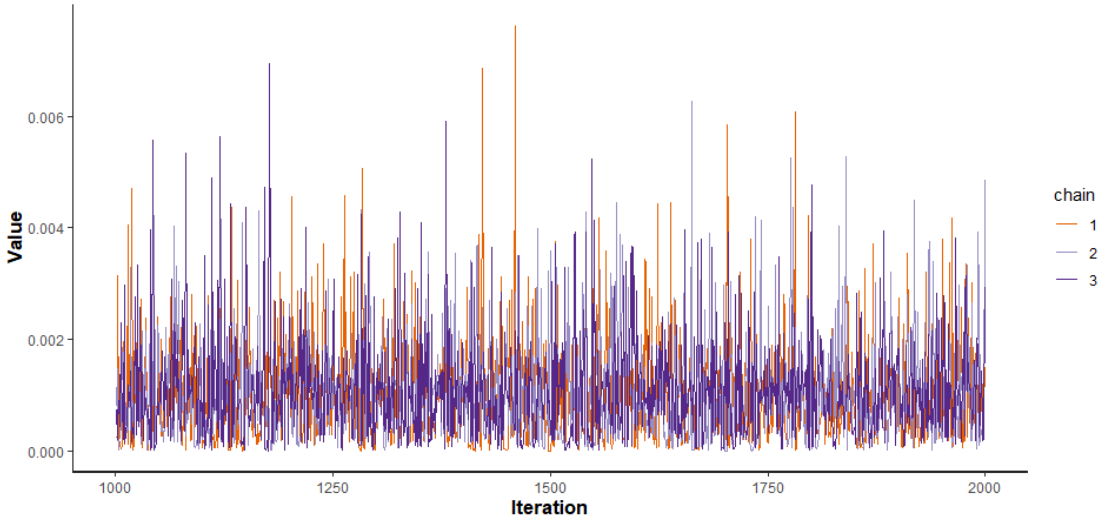

Figure 7.9: Trace plot of the association parameter of the two-part joint model applied to the routine HIV Western Cape subset development cohort.

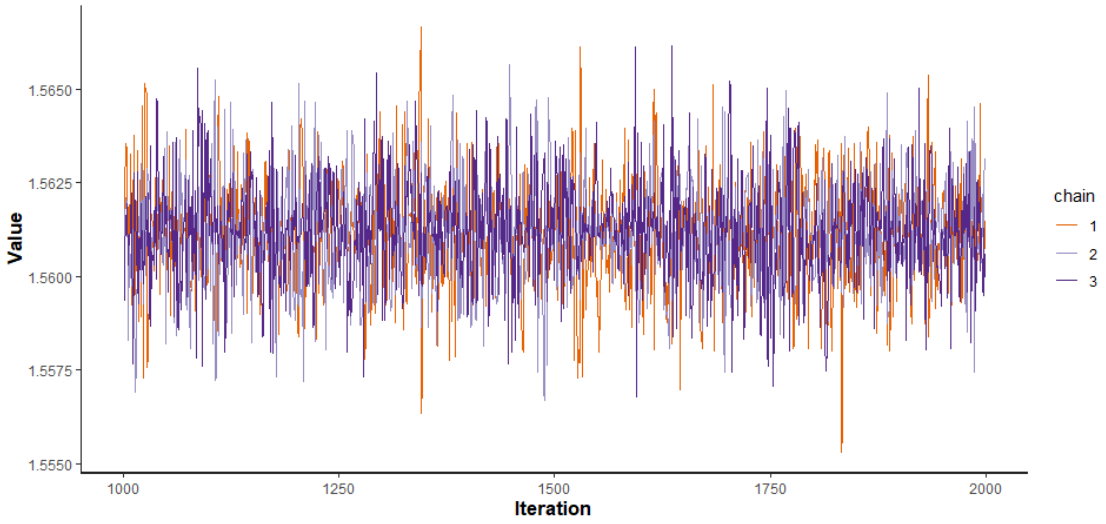

Figure 7.10: Trace plot of the standard deviation parameter of the error term of the two-part joint model applied to the routine HIV Western Cape subset development cohort.

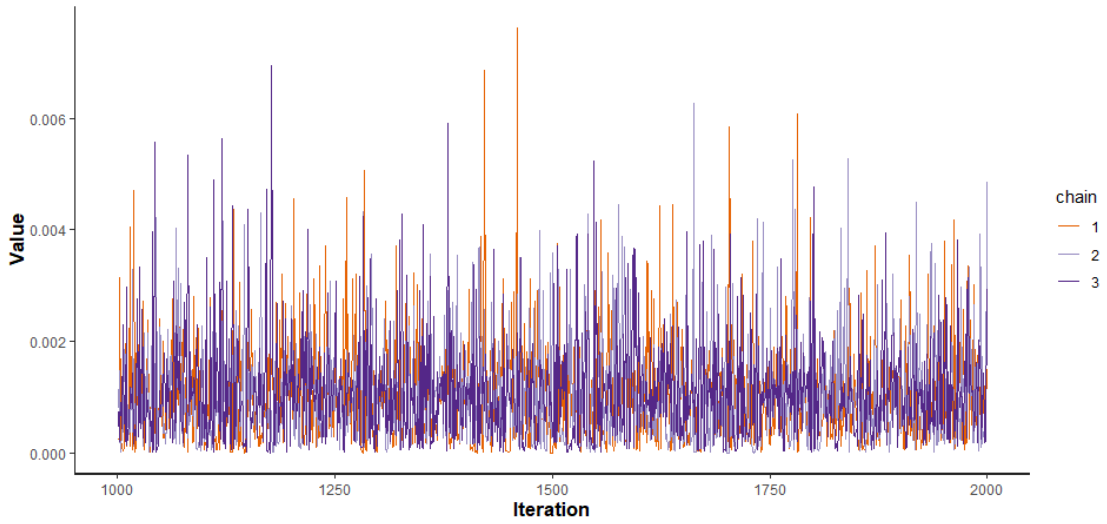

Figure 7.11: Trace plots of the standard deviation parameters of the random effects of the two-part joint model applied to the routine HIV Western Cape subset development cohort.

Table 7.5: Parameter estimates for the two-part joint model applied to the routine HIV Western Cape subset development cohort.

Parameter	Mean	SD[1]	2.5.%[2]	97.5%[3]
Longitudinal process				
Binary part				
α_0	-1.2878	0.0641	-1.4129	-1.1601
α_1	-0.0032	0.0005	-0.0041	-0.0022
α_2	0.1579	0.0898	-0.0164	0.336
Continuous part				
β_0	2.3663	0.0848	2.2026	2.5361
β_1	-0.0428	0.0006	-0.044	-0.0416
β_2	1.8401	0.1226	1.6089	2.0782
Time-to-event process				
λ	0.1868	0.0038	0.1795	0.1944
γ	-0.8789	0.0631	-1.0045	-0.7585
φ	0.0011	0.001	0.0000	0.0037
Standard deviation				
σ_{ϵ}	1.5612	0.0015	1.5582	1.5641
$\sigma_{a_{0i}}$	0.4453	0.0104	0.4252	0.467
$\sigma_{b_{0i}}$	0.2453	0.0058	0.2343	0.2569
$\sigma_{b_{1i}}$	0.4917	0.0118	0.469	0.5158

SD[1]: Standard deviation; 2.5%[2]: lower limit of the 95% credible interval; 97.5%[3]: upper limit of the 95% credible interval

7.4.2 Predictive performance

The predictive performance of the proposed two-part joint model was evaluated and compared with a joint model (defined in Chapter 6) in the motivating routine HIV Western Cape subset validation cohort. The joint model adapted from Chapter 6 had i) a Gaussian mixed effects model with semicontinuous HIV viral load values as the outcome, observed follow-up time and binary covariate in the fixed effects and random intercepts and slopes in the random effects, and ii) a proportional hazards model with a binary covariate and assuming a B-spline baseline hazard and a current value association structure. On the other hand, the two-part joint had i) a two-part mixed-effects model containing random intercepts in the logistic mixed-effects model for the binary part and random intercepts and slopes in the Gaussian mixed-effects model, and ii) a proportional hazards model with a binary covariate and assuming a Weibull baseline hazard and a current value association structure.

Given that individuals in the routine HIV Western Cape subset validation cohort were virologic failure-free during the first 12 months of follow-up, two clinically relevant future times (12 and 24 months) were chosen to assess the predictive performance. Specifically, time-dependent AUCs and Brier scores were used to evaluate the predictive performance of the proposed two-part joint model and compare its predictive performance with the joint model. The AUCs and Brier scores at 12 and 24 months are given in Figure 7.12 and Figure 7.13, respectively. Unlike the simulation study, the two-part joint model had lower AUCs than the joint mode for all given time windows (Figure 7.12). This suggests that the joint model distinguishes individuals with low and high risk of virologic failure better than the two-part joint model. Figure 7.13 displays Brier scores for the given time windows. Like the simulation study, the proposed two-part joint model has higher Brier scores than the joint model at all time windows, suggesting that the joint model had better calibration than the two-part joint model. Nevertheless, the Brier scores for the two-part joint model remain acceptable (less than 0.25); thus, this model is adequate for calculating dynamic predictions of virologic failure.

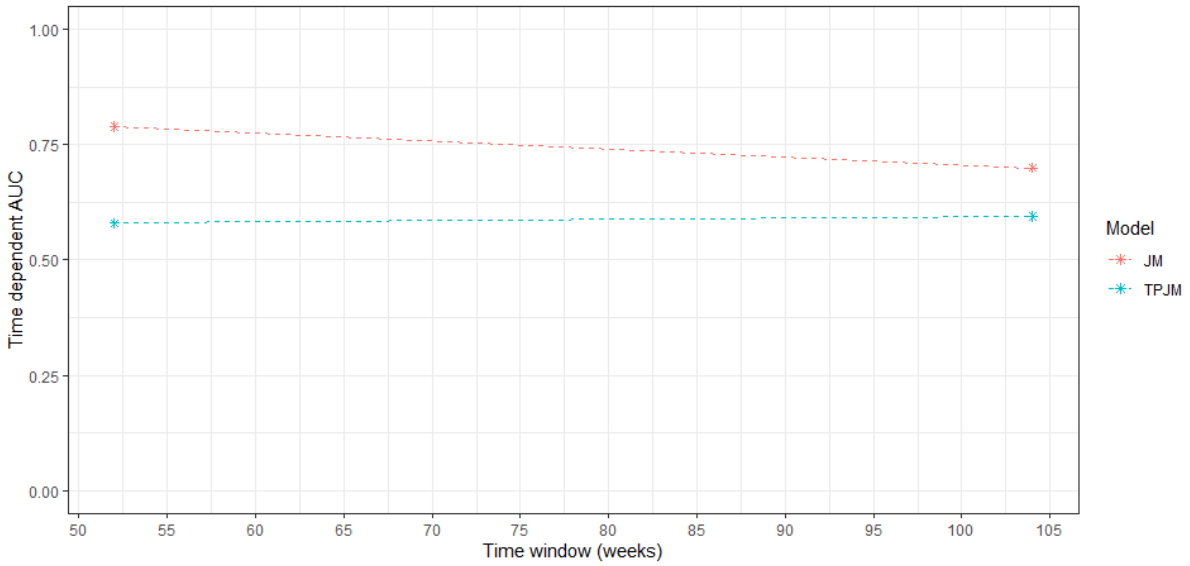

Figure 7.12: Time-dependent AUCs for a joint and two-part joint model applied to the routine HIV Western Cape subset validation cohort.

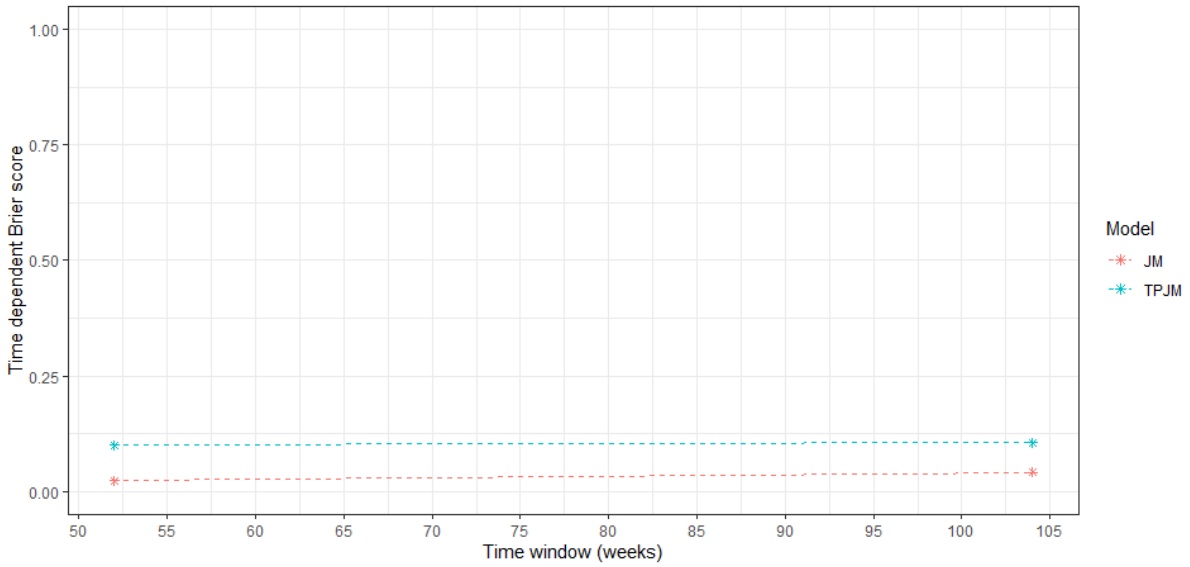

Figure 7.13: Time-dependent Brier scores for a joint and two-part joint model applied to the routine HIV Western Cape subset validation cohort.

7.4.3 Prediction of the probability of virologic failure

The two-part joint model was used to calculate dynamic predictions for virologic failure for three exemplary individuals. These individuals (32, 149, and 95) were randomly selected from

the routine HIV Western Cape subset validation cohort and their HIV viral load values during observed follow-up time are given in Figure 7.14. In the first 12 months of follow-up, the individuals were virologic failure-free. Individual 32 was a 49-year-old female with a low proportion (25%) of HIV viral load values below the 1.7 copies/mL detection limit. She experienced virologic failure at a later observed follow-up time. Individual 149 was a 30-year-old female with a moderate proportion (50%) of HIV viral load values below the limit of detection. She did not have virologic failure at a later observed follow-up time. Individual 95 was a 32-year-old male with a high proportion (75%) of HIV viral load values below the limit of detection and did not have virologic failure at a later observed follow-up time. Dynamic predictions of the three individuals were calculated from 52 weeks (12 months) to 104 weeks (24 months) by an increment of 52 weeks. These dynamic predictions are given in Figure 7.15, Figure 7.16 and Figure 7.17. Individuals 32 and 149 had higher probabilities of virologic failure than individual 95. The 95% Monte Carlo prediction intervals for individuals 32 and 149 are narrower than that of individual 95.

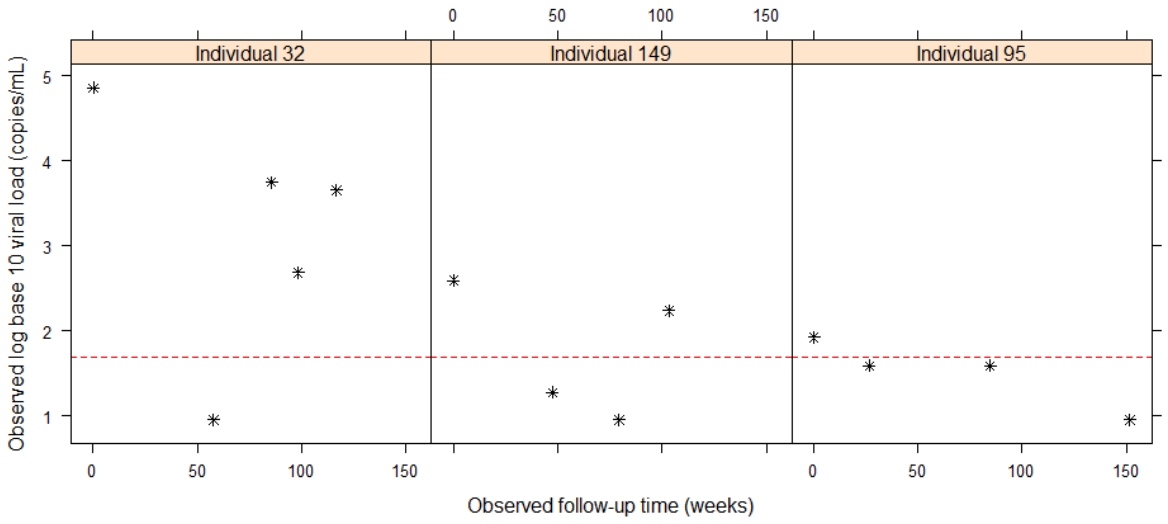

Figure 7.14: Observed HIV viral load biomarker values during observed follow-up time for three randomly selected people from the historic routine HIV validation cohort.

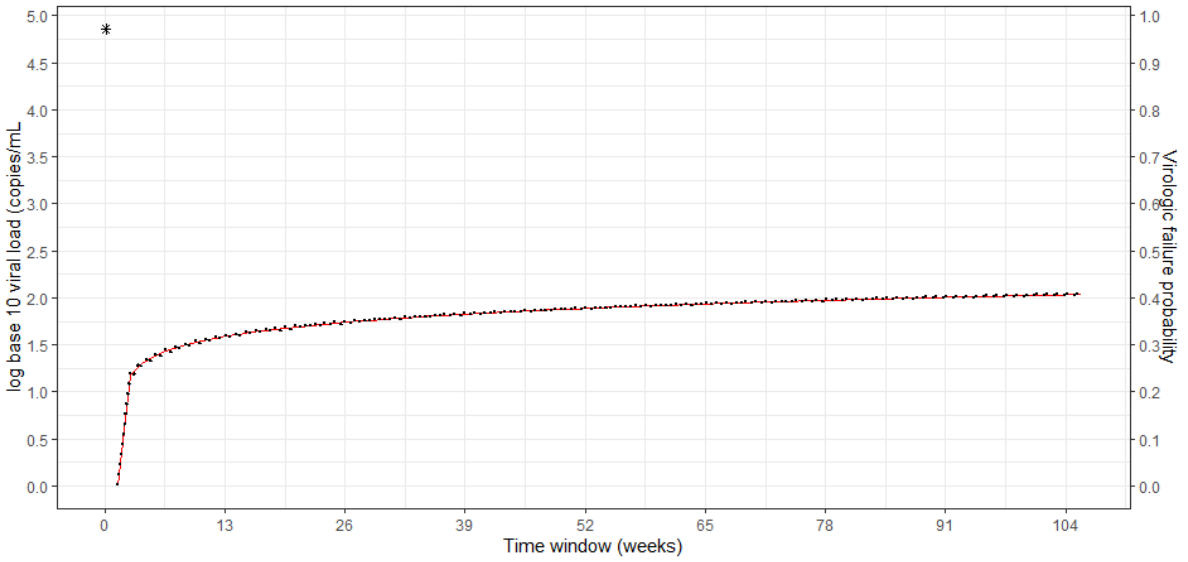

Figure 7.15: Estimated probability of virologic failure, conditional on being virologic failure-free during the first 12 months of follow-up for individual 32 having a low proportion of HIV viral load values below the limit of detection. The black asterisks represent observed log base 10 HIV viral load biomarker values during the first 12 months of observed follow-up. The solid red line is the estimated probability of virologic failure up to 104 weeks (24 months). The broken black lines are the 95% credible intervals for the estimated probabilities.

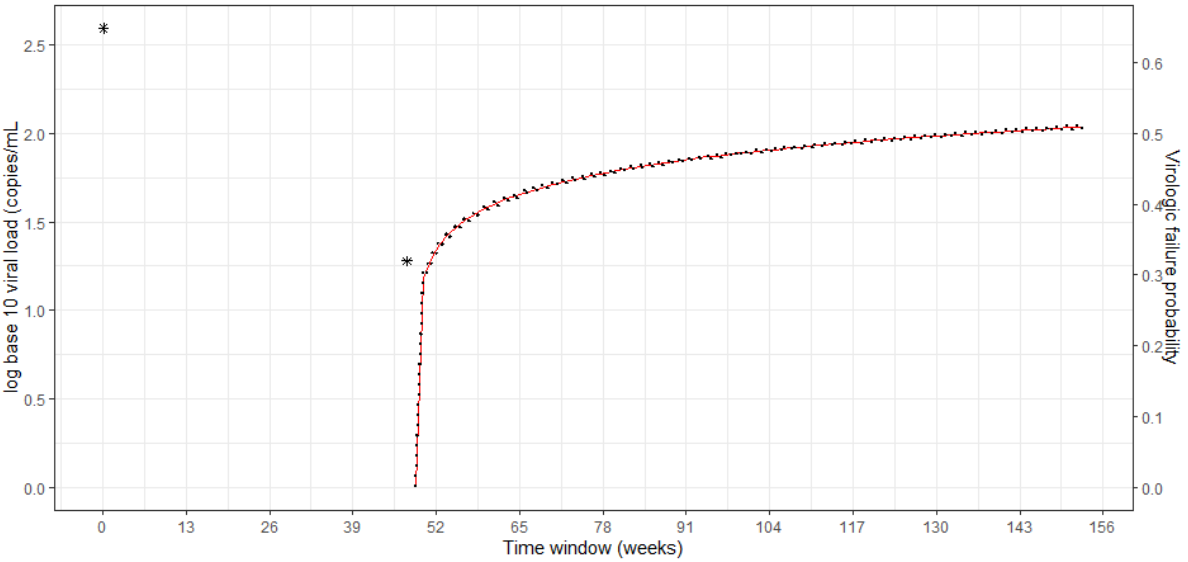

Figure 7.16: Estimated probability of virologic failure, conditional on being virologic failure-free during the first 12 months of follow-up for individual 149 having a moderate proportion

of HIV viral load values below the limit of detection. The black asterisks represent observed log base 10 HIV viral load biomarker values during the first 12 months of observed follow-up. The solid red line is the estimated probability of virologic failure up to 104 weeks (24 months). The broken black lines are the 95% credible intervals for the estimated probabilities.

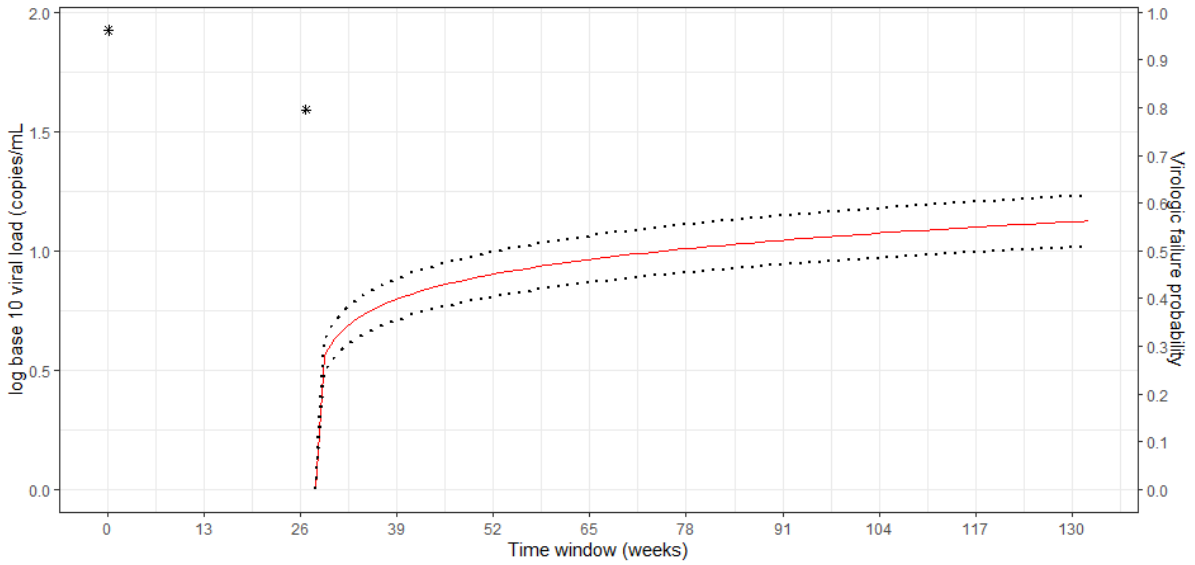

Figure 7.17: Estimated probability of virologic failure, conditional on being virologic failure-free during the first 12 months of follow-up for individual 95 having a high proportion of HIV viral load values below the limit of detection. The black asterisks represent observed log base 10 HIV viral load biomarker values during the first 12 months of observed follow-up. The solid red line is the estimated probability of virologic failure up to 104 weeks (24 months). The broken black lines are the 95% credible intervals for the estimated probabilities.

7.5 Discussion and Conclusions

In this chapter, a two-part joint model was used to analyse longitudinal HIV viral load and time to virologic failure data. Compared with the joint model in Chapter 6, the advantage of this modelling approach is that using a two-part mixed effects model provides robust modelling of HIV viral load values below the detection limit. HIV viral load values below a detection limit of 50 copies/mL were presented as zero to allow for comparable results between the simulation study and application to routine HIV Western Cape data. Due to the difference in the

parameterisation of the joint and two-part joint models, only the association parameter was compared between the simulation study and application to routine HIV Western Cape data.

The simulation study showed that the two-part joint model had a lower mean squared error than a joint model regardless of the proportion of zeros. The two-part joint model generally had better discrimination ability than a joint model, although the Brier scores were suboptimal. These suboptimal Brier scores may potentially be due to the sample size used in the simulation study.

The two-part joint modelling approach could contribute to research in HIV, where an assay used in HIV viral load testing may not detect HIV viral load values below a limit of detection. The contribution is not limited to HIV research because it can be extended to other diseases where a biomarker may not be detectable below a certain limit of detection. In this thesis, the model provides a novel approach to modelling longitudinal semicontinuous HIV viral load and time to virologic failure as it accounts for excess HIV viral load values below a limit of detection while estimating the mean of the HIV viral load values above the limit of detection (Brilleman, 2016). The findings showed that expected HIV viral load values above a detection limit were associated with an increased hazard of virologic failure. This finding is consistent with a study by Brilleman (2016), which used a two-part joint model for longitudinal semicontinuous HIV viral load and time-to-stopping or modifying treatment in individuals living with HIV on ART. This study used a two-part model parameterisation different to the two-part joint model presented in this thesis. The study reported that HIV viral load changes over time were associated with increased hazard of stopping or modifying treatment after adjusting for age, sex, and year of ART initiation.

Finally, the two-part joint model was used to calculate prognosis predictions of the probability of virologic failure, using the history of an individual's baseline demographics and HIV viral load changes over time. The two-part joint modelling-based prediction model had suboptimal discrimination ability and better calibration than a joint model in the routine HIV data. This surprising finding warrants further investigation. A potential reason for this finding might be that the data generated in the simulation used equally spaced follow-up times, while in the routine HIV data, there were intermittent follow-up times. This resulted in a complexity for the average changes in the longitudinal outcome. This notable feature of intermittent follow-up times is flexibly captured in a joint model (Rizopoulos, 2012) and may be captured similarly in a two-part joint model.

The two-part joint modelling approach used in this thesis had limitations. Firstly, the model assumed a linear trajectory for the longitudinal HIV viral load biomarker. Although the HIV viral load trajectories in the routine HIV data had evidence of nonlinearity, the two-part joint model applied in this thesis assumed linearity in both the fixed and random effects. This has likely contributed to the suboptimal predictive performance of the model. Misspecification of the longitudinal trajectory can lead to biased parameter estimates and reduced predictive performance, as shown in previous studies. Arisido et al. (2019) demonstrated, through a comprehensive simulation study, that joint models assuming linear trajectories in the presence of true nonlinear patterns produced biased estimates. Similarly, Davies et al. (2024) showed that joint models assuming linearity in both the fixed and random effects had higher prediction error compared to models that incorporated nonlinearity, particularly when the biomarker trajectory have evidence of nonlinearity across individuals. This thesis did not account for such nonlinearities, which may have limited the two-part joint model's ability to capture the true underlying HIV viral load and its association with virologic failure. While the two-part joint model provided a useful foundation, its adequacy could be improved by incorporating alternative specifications of the trajectory for the longitudinal HIV viral load biomarker. Future work would explore alternative specifications of the longitudinal trajectory, such as the use of splines or polynomial terms in both fixed and random effects, to better accommodate the nonlinear patterns often observed in biomarkers. Additionally, comparative analysis of predictive performance under different trajectory specifications would help identify the most appropriate longitudinal biomarker trajectory in this context.

The second key limitation of this thesis was the specified of the association structure between the longitudinal HIV viral load and the hazard of virologic failure. In the joint modelling framework for longitudinal and time-to-event data, carefully specifying the association structure or functional form is crucial to avoid biased results (Rizopoulos, 2012). Studies have primarily focused on the benefits of using alternative association structures to characterize this association. A comprehensive overview of such association structures is provided by (Rizopoulos, 2012). Moreover, Mchunu et al. (2022), in a study of individuals co-infected with HIV and tuberculosis, compared four commonly used association structures describing the association between longitudinal CD4 count and the hazard of death. Their findings demonstrated that the current value association structure is not always the most suitable choice.

This thesis fitted a two-part joint model with a current value association structure. The model's predictive performance was suboptimal, and this may have been partly attributable to the use

of the current value association structure. Restricting the association to the current value alone may lead to misleading conclusions about the underlying disease process (Rizopoulos, 2012; Vacek, 1997). Accordingly, the results from the two-part joint model should be interpreted with caution, particularly given the potential for misspecification of the association structure between the longitudinal HIV viral load and virologic failure. A potential improvement for future work would be to investigate alternative functional forms such as the time-dependent slope (Brown, 2009; Rizopoulos, 2012) or cumulative effects (Brown, 2009) to better characterize the association between the biomarker trajectory and the event hazard. Further research could also evaluate the predictive performance of two-part joint models incorporating different association structures to identify the most appropriate structure.

Thirdly, due to the computation and memory challenges for Bayesian inference in Stan software, a smaller subset of the routine HIV data was used for the development and validation of the two-part joint model. A natural extension would be to use a large subset of the routine HIV data and apply techniques that allow for better computational time. Techniques such as the embarrassingly parallel MCMC will be employed in future research (Ren, Wang, & Luo, 2021).

Finally, similar to the joint model in Chapter 6, external validation was not feasible due to the lack of an external cohort to validate the two-part joint model. External validation needs to be explored in future research.

In conclusion, the two-part joint modelling approach used in this thesis generally had good or similar predictive performance compared to a joint model discussed in Chapter 6. The utility of this model was illustrated by prognosis prediction of virologic failure probabilities using individual-specific HIV viral load values and baseline characteristics. Following the current literature, this study is the first to compare predictive performance between a two-part joint model and a joint model applied to routine data from electronic health records in a resource-limited setting and provide prognosis predictions for virologic failure. Therefore, two-part joint modelling-based predictions could be useful for prognosis, even when there is an excess of HIV viral load values below detection.

7.6 References

Abbott, P. (2005). Tricks of the trade: Legendre-gauss quadrature. *Mathematica Journal, 9*(4), 689-691.

Arisido, M. W., Antolini, L., Bernasconi, D. P., Valsecchi, M. G., & Rebora, P. (2019). Joint model robustness compared with the time-varying covariate Cox model to evaluate the association between a longitudinal marker and a time-to-event endpoint. *BMC Med Res Methodol, 19*(1), 222. doi:10.1186/s12874-019-0873-y

Betancourt, M., & Girolami, M. (2015). Hamiltonian Monte Carlo for hierarchical models. *Current trends in Bayesian methodology with applications, 79*(30), 2-4.

Brilleman, S. L. C., M. J.;May, M. T.;Gompels, M.;Abrams, K. R. (2016). Joint longitudinal hurdle and time-to-event models: an application related to viral load and duration of the first treatment regimen in patients with HIV initiating therapy. *Stat Med, 35*(20), 3583-3594. doi:10.1002/sim.6948

Brown, E. R. (2009). Assessing the association between trends in a biomarker and risk of event with an application in pediatric hiv/aids. *Ann Appl Stat, 3*(3), 1163-1182. doi:10.1214/09-aoas251

Dagne, G. A. (2017). Joint two-part Tobit models for longitudinal and time-to-event data. *Stat Med, 36*(26), 4214-4229. doi:10.1002/sim.7429

Davies, A. L., Coolen, A. C., & Galla, T. (2024). Delayed kernels for longitudinal survival analysis and dynamic prediction. *Stat Methods Med Res, 33*(10), 1836-1858. doi:10.1177/09622802241275382

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 457-472.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy, 19*(10), 555.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., . . . Hothorn, M. T. (2021). Package 'mvtnorm'. *Journal of computational and Graphical Statistics, 11*(950-971), 155.

Hatfield, L. A., Boye, M. E., Hackshaw, M. D., & Carlin, B. P. (2012). Multilevel Bayesian models for survival times and longitudinal patient-reported outcomes with many zeros. *Journal of the American Statistical Association, 107*(499), 875-885.

Heagerty, P., Saha-Chaudhuri, P., & Saha-Chaudhuri, M. P. (2013). Package 'survivalROC'. *San Francisco: GitHub*.

Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics, 1*(4), 465-480.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res., 15*(1), 1593-1623.

Liu, L. (2009). Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in medicine, 28*(6), 972-986.

Liu, L., Shih, Y.-C. T., Strawderman, R. L., Zhang, D., Johnson, B. A., & Chai, H. (2019). Statistical analysis of zero-inflated nonnegative continuous data: a review. *Statistical Science, 34*(2), 253-279.

Manning, W. G., Morris, C. N., Newhouse, J. P., Orr, L. L., Duan, N., Keeler, E. B., . . . Phelps, C. E. (1981). A two-part model of the demand for medical care: preliminary results from the health insurance study. *Health, economics, and health economics, 137*, 103-123.

McHunu, N. N., Mwambi, H. G., Rizopoulos, D., Reddy, T., & Yende-Zuma, N. (2022). Using joint models to study the association between CD4 count and the risk of death in TB/HIV data. *BMC Med Res Methodol, 22*(1), 295. doi:10.1186/s12874-022-01775-7

Mogensen, U. B., Ishwaran, H., & Gerds, T. A. (2012). Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *Journal of statistical software, 50*(11), 1-23.

Olsen, M. K., & Schafer, J. L. (2001). A Two-Part Random-Effects Model for Semicontinuous Longitudinal Data. *Journal of the American Statistical Association, 96*(454), 730-745. Retrieved from http://www.jstor.org/stable/2670310

Ren, X., Wang, J., & Luo, S. (2021). Dynamic prediction using joint models of longitudinal and recurrent event data: a Bayesian perspective. *Biostatistics & epidemiology, 5*(2), 250-266.

Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data: With applications in R*: Chapman and Hall/CRC.

Rustand, D., Briollais, L., Tournigand, C., & Rondeau, V. (2022). Two-part joint model for a longitudinal semicontinuous marker and a terminal event with application to metastatic colorectal cancer data. *Biostatistics, 23*(1), 50-68.

Stan Developent Team. (2016). RStan: the R interface to Stan. R package version. 2016; 2 (1). In.

Su, L., Tom, B. D., & Farewell, V. T. (2009). Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics, 10*(2), 374-389.

Sylvestre, M.-P., Edens, T., MacKenzie, T., Abrahamowicz, M., & Sylvestre, M. M.-P. (2010). Package 'PermAlgo'. In: Last accesses: Sep-10-2014.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: journal of the Econometric Society*, 24-36.

Tsiatis, A. A., & Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 809-834.

Vacek, P. M. (1997). Assessing the effect of intensity when exposure varies over time. *Statistics in medicine, 16*(5), 505-513. doi:10.1002/(sici)1097-0258(19970315)16:5<505::aid-sim424>3.0.co;2-z

Chapter 8 Discussion and Recommendations

8.1 Introduction

This thesis challenges the prevailing understanding that dynamic prediction models applied to data in all settings have improved predictive performance over Cox proportional hazards-based prediction models. This thesis contributes to the body of dynamic prediction models for individualised prognosis prediction by providing insights into the conduct and quality of model evaluation for dynamic prediction models. According to the current literature, this thesis is the first combined body of work in the growing area of dynamic prediction models to i) apply dynamic prediction models using the joint modelling approach to routine chronic disease data from electronic health records in resource-limited settings where data quality is often difficult to assess and may be poor, and ii) compare predictive performance between a two-part joint and a joint model and provide prognosis predictions. Dynamic prediction models were applied to routine data from electronic health records in South Africa with particular emphasis on using longitudinal biomarkers for two chronic diseases of high burden in this resource-limited setting for individualised prognosis prediction of clinical outcomes.

Individualised prognosis prediction can help inform healthcare providers in therapeutic decision-making (Collins, Reitsma, Altman, & Moons, 2015). Many modelling approaches can be used to obtain predictions, including a Cox proportional hazard and a joint model. Cox proportional hazards-based predictions are popular in biomedical research, mainly because they are easy to implement and adopt as simple tools for calculating probabilities of clinical outcomes. Joint modelling-based predictions (dynamic predictions) enable prognosis predictions that can be updated in real-time as more information on individuals living with a disease accumulates over time (Rizopoulos, 2011b). This thesis sought to provide insights into dynamic prediction models in the context of their development and reporting for individualised prognosis prediction using routine data from electronic health records in the Western Cape, South Africa. This final chapter concludes the thesis. The main results of each chapter, strengths and limitations, and recommendations are provided. First, a synopsis of the key findings in relation to the objectives of the thesis is provided in Section 8.2. Secondly, strengths and limitations of the thesis are discussed in Section 8.3. Third, a discussion of the key findings of this thesis is presented in Section 8.4. Recommendations from this thesis as a combined body of work are provided in Section 8.5, and Section 8.6 concludes the thesis.

8.2 Synopsis

Different modelling approaches were applied to routine data throughout the thesis to get prognostic prediction probabilities for clinical outcomes. The theoretical background of these models is provided in Chapter 2. Consolidated guidelines such as PROGRESS (Steyerberg et al., 2013) and TRIPOD (Collins et al., 2015) are recommended for developing and validating prediction models. These consolidated guidelines were used in a literature review described in Chapter 3 to outline the characteristics that inform the development and reporting of published dynamic prediction models. The factors included attributes of data, type of disease and availability of covariates. The literature review (Chapter 3) addressed the first objective of this thesis by providing characteristics that inform the development and reporting of dynamic prediction models.

Motivated by two chronic diseases (HIV and T2DM) of high burden in South Africa described in Chapter 4, TRIPOD and PROGRESS guidelines were used as a basis for developing and validating prediction models for prognosis prediction in Chapter 5, Chapter 6 and Chapter 7. In Chapter 5, Cox proportional hazard models were applied to routine HIV and diabetes data from electronic health records in the Western Cape, South Africa. The second objective of the thesis was addressed by highlighting key demographics and clinical variables, such as repeated biomarker measurements central to understanding predictive performance in Cox proportional hazard models.

In Chapter 6, the Cox proportional hazards model was extended to include repeated biomarker measurements using joint modelling. The chapter addressed the third objective of this thesis by exploring the predictive performance of joint modelling and comparing it with the Cox proportional hazards-based prediction model. The prevailing assumption that dynamic prediction models have better predictive performance than Cox proportional hazards-based prediction models was challenged and discussed. Quantitative and novel evidence against this assumption was provided. It is suggested that improved predictive performance is not always guaranteed for dynamic prediction models applied to routine data in electronic health records from resource-limited settings where data quality is often difficult to assess and may be poor. In addition, the findings presented suggest that characteristics such as type of disease, biomarker type and availability of covariates inform predictive performance. Finally, in Chapter 7, the joint model introduced in Chapter 6 was extended by developing a two-part joint model for longitudinal semicontinuous biomarker and time-to-event data. The extension was

motivated by routine HIV viral load data, regarded as semicontinuous. Specifically, a two-part joint model for longitudinal semicontinuous biomarker and time-to-event data was developed and validated in simulated and routine HIV data. This model was used for the prognosis of virologic failure probabilities in individuals living with HIV in the Western Cape. Chapter 7 addressed the fourth and fifth objectives by i) extending joint models to allow for different types of longitudinal continuous biomarker-outcome data and ii) applying and comparing the predictive performance of joint and two-part joint models in simulated and routine HIV data from electronic health records in resource-limited settings, respectively. The findings in this chapter suggested that, although the two-part joint model provides flexible modelling of longitudinal semicontinuous and time-to-event data, a joint model offers comparable predictive performance for such data. Taken together, the findings in the thesis have limitations. The strengths and limitations of the thesis as a body of work are presented in Section 8.3 below.

8.3 Strengths and Limitations

Strengths and limitations were presented in the discussion section of each chapter. However, additional strengths and limitations were considered before discussing the key findings of the thesis and providing recommendations.

8.3.1 Limitations

First and foremost, the prediction models presented in this thesis were developed and validated from two provincial-level (Western Cape) populations: people with HIV on ART and T2DM individuals. This is a limitation of the prediction models presented in this thesis because, in South Africa, the proportion of individuals living with HIV varies geographically, with the Western Cape province having the lowest prevalence (8.2%) (Human Sciences Research Council, 2023). Similarly, this province has the lowest prevalence of T2DM individuals compared to other provinces in South Africa, with a prevalence of 14% (Sahadew, Pillay, & Singaram, 2022). Coupled with this, South Africa has an under-resourced healthcare system (Brennan et al., 2023; Maphumulo & Bhengu, 2019) with inequalities in healthcare expenditure. However, the Western Cape is one of the South African provinces with the most significant healthcare expenditure (Statistics South Africa, 2021). Therefore, the findings in this thesis are not generalisable to all provinces in South Africa. Despite this limitation, these

findings can be used preliminarily at a provincial level to point to the need for early identification of individuals at a higher probability of having virologic failure or achieving glycaemic control to enhance chronic disease management in the current South African under-resourced healthcare system.

Secondly, the efficacy of the prediction models was not further verified by external validation and resampling techniques such as bootstrapping and repeated cross-validation. However, recommended guidelines such as TRIPOD and PROGRESS (Collins et al., 2015; Steyerberg et al., 2013) were used as a basic means to develop and internally validate the prediction models on individuals living with HIV and T2DM individuals in the Western Cape province, South Africa. Given that the South African NHLS database provides nationwide coverage of the country's diverse population, provincial-level data from other provinces can be used in future research to facilitate the external validation and adoption of other resampling techniques, further verifying the prediction models.

Thirdly, the definition of virologic failure was restricted to South Africa's 2019 national guidelines for managing HIV in adults (South African National Department of Health, 2019). The findings should be interpreted with caution for updated national guidelines. A sensitivity analysis using updated national guidelines is an area for future research.

Another significant concern was the absence of essential demographic and clinical variables for individuals living with HIV and those with T2DM in the routine electronic health datasets. This deficiency significantly compromised the predictive performance of the prediction models presented in the thesis.

Finally, random effects were used to formulate the joint model for longitudinal biomarker and time-to-event data in Chapter 6 and Chapter 7. The dimension of random effects structures poses a computational challenge because numerical integration over the random effects is often required to estimate them (Rizopoulos, 2011a). Alternative numerical integration approaches such as pseudo-adaptive Gaussian quadrature (Rizopoulos, 2012) and parallelising MCMC (Wang, Guo, Heller, & Dunson, 2015) are proposed to improve computation efficiency. Given the large routine datasets used in this thesis, it was not feasible to further validate the dynamic prediction models using bootstrapping and parallelising MCMC without using high-performance resources. Moreover, as seen in Chapter 7, a smaller subset of data was used to apply a two-part joint model because it was computationally demanding to develop and validate the model on the full dataset. The dynamic prediction models in this thesis and other dynamic

prediction models in the literature can benefit from adopting the embarrassingly parallel MCMC approach and using high-performance resources to enable improved computational time in a sensitivity analysis with a larger data subset (Ren, Wang, & Luo, 2021).

8.3.2 Strengths

Despite these limitations, this thesis contributes to the body of knowledge in research in several ways. Dynamic prediction models are well-established and are mostly based on data from research cohorts in resource-rich healthcare settings. In this thesis, routine data from electronic health records in resource-limited healthcare settings were used to develop and validate dynamic prediction models. The application of these models to routine HIV and diabetes data enabled prognosis predictions that are reflective of resource-limited healthcare systems.

Another strength of the findings in this thesis is that consolidated guidelines (TRIPOD and PROGRESS) for developing and validating prediction models were used to identify ranging characteristics, including the type of disease and the amount of readily available information in the data that inform the development and reporting of prediction models (Collins et al., 2015; Moons et al., 2014). These guidelines often exist in isolation, with developed individualised prediction models in the literature. In this thesis, the guidelines were adopted for prediction models based on both Cox proportional hazards (Chapter 5) and joint modelling-based (Chapter 6). These two models were compared to challenge and provide insights to the forebelief that the joint modelling-based prediction models (dynamic prediction models) universally have improved predictive performance compared to Cox proportional-hazards-based models.

The joint modelling approach was extended by specifying a two-part mixed-effects model to flexibly account for viral load data below a detection limit of 50 copies/mL, as shown in Chapter 7. The two-part joint model in this thesis extends findings reported by Brilleman (2016), who implemented a two-part joint model in Stan to evaluate an association between longitudinal semicontinuous viral load and time to stopping or modifying treatment. The findings in their study are extended in this thesis in several ways: i) linear changes over time in viral load biomarker measurements were assumed, ii) a normal distribution was used to model longitudinal viral load biomarker data above a detection limit of 50 copies/mL, iii) the covariates were specified in both binary and continuous parts of the two-part mixed effects model, iii) a comprehensive simulation study was used to compare the two-part joint model

with a joint model in both inference and predictive performance, and iv) the two-part joint model was used for individualised prognosis predictions on routine HIV data.

Finally, taken together as a combined body of work, predictive performance was evaluated and compared across Cox proportional hazards-based and dynamic prediction models (joint and two-part joint models). This allows for evaluating whether increasing complexity in the modelling for prognosis prediction improves or limits predictive performance. The contributions of the strengths to the key findings of the entire thesis are further discussed in Section 8.4 below.

8.4 Discussion of Key Findings

Considering the limitations noted, this section explores the key findings of the entire thesis, placing these findings in the context of developing and reporting prediction models before providing recommendations. The findings from the literature review on the application of dynamic prediction models are revisited to provide a conceptual basis for developing and reporting dynamic prediction models. Figure 8.1 provides a useful conceptual understanding of how consolidated guidelines (TRIPOD and PROGRESS) for developing and validating prediction models can inform the development and reporting of prediction models in research. This figure was used to frame the discussion on the key findings. The key findings are presented in five broad underlying themes that span across the results in the chapters of this thesis: i) the type of disease characterises length of follow-up and prediction time, ii) characteristics of longitudinal biomarkers are central to dynamic prediction models, iii) consistent follow-ups and sufficient covariate collection is critical to prediction models, iv) dynamic prediction models do not always have better prediction performance, and v) model complexity is not an answer to lack of information in data.

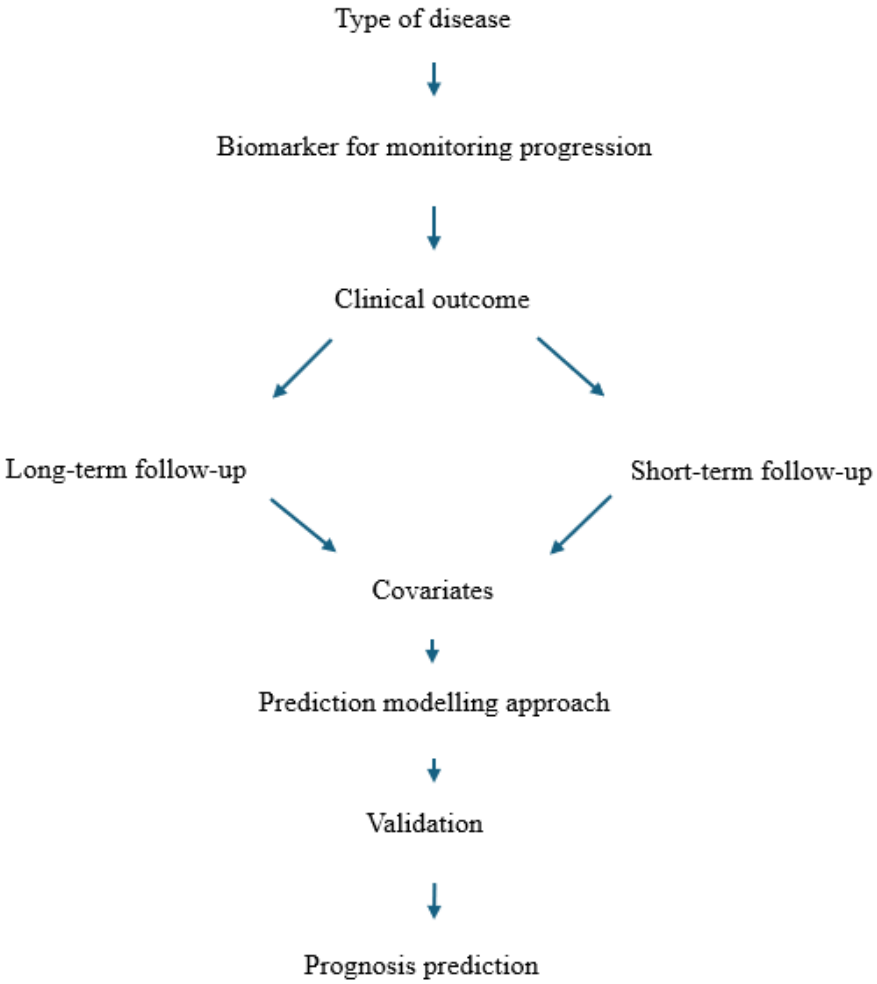

Figure 8.1: A summary of a conceptual framework for developing and reporting dynamic prediction models adapted from the literature review in Chapter 3.

8.4.1 Type of disease characterises the length of follow-up and prediction time window

Dynamic prediction models perform poorly over long-term prediction time windows. The interval between consecutive follow-ups in the data is so large and inconsistent that the change in the biomarker cannot be adequately observed within those gaps, leading to poor information for prognosis predictions. Long-term predictions were used in this thesis for individuals living with HIV and T2DM. These chronic diseases often require longer follow-ups, as seen in the routine datasets described in Chapter 4. For individuals living with HIV on ART, South Africa's 2019 national guidelines recommend routine follow-up at 6 and 12 months and annually

thereafter if they are virally suppressed (South African National Department of Health, 2019). For T2DM individuals, South Africa's 2014 guidelines recommended routine follow-up at 3-6 months until therapeutic goals have been achieved and annually after that (South African National Department of Health, 2014). The gaps between consecutive follow-ups and the inconsistent nature of the follow-ups in the resulting routinely collected data make it difficult to monitor changes in the disease biomarkers. As highlighted in the literature review in Chapter 3, the type of disease informs the length of follow-up and prediction time window, ranging from months to years. Although the chronic nature of HIV and T2DM necessitates long-term prediction time windows compared to short-term diseases, inconsistent follow-ups in the routine datasets have led to inadequate observation of changes in the biomarkers. This discussion section suggests that dynamic prediction models applied to data with inconsistent follow-ups result in poor predictions made in long-term prediction time windows.

8.4.2 Longitudinal biomarkers characteristics are central to prediction models

Biomarkers with high variance (does not show interim changes in individuals), especially high within-individual variance compared to between-individual variance, perform poorly in dynamic prediction models. This is because reliable average differences between subgroups and narrow prognostic prediction probabilities for clinical outcomes are hard to achieve. Biomarkers are widely used to evaluate disease progression, monitor therapeutic interventions and identify individuals with a high risk of adverse outcomes (Brody, 2016). These biomarkers vary from nearly direct and intermediate (for example, HIV viral load) to indirect, medium-term (for example, HbA1c) measures of changes in individuals over follow-ups. In this thesis, changes within six months of follow-up for individuals living with HIV on ART were monitored using HIV viral load (WHO, 2016), while HbA1c was used to monitor changes within 3-6 months of follow-up for T2DM individuals (South African National Department of Health, 2014). These biomarkers defined adverse outcomes (virologic failure and glycaemic control). Using biomarkers as intermediate clinical outcomes is more appropriate when they sufficiently represent a biological process of disease (Burke, 1998, 2016). Given the lack of consistent follow-ups in the routine data, the biomarkers provided limited information about the interim changes in the biology process of HIV and T2DM to the dynamic prediction models. This section of the discussion suggests that biomarkers which do not adequately reflect interim

changes in the biological processes of disease in individuals provide limited value in dynamic prediction models.

8.4.3 Availability of information in data from studies is crucial to prediction models

Large but low information (insufficient covariate collection) data are not a good substitute for smaller data with better information quality. The digitisation of routine healthcare data has led to the availability of data on a large scale and an increase in the application of prediction models to these large routine data. An attractive feature of routine data is that it is routinely collected in healthcare and provides more practical documentation of therapeutic interventions for the general population in healthcare systems. Routine data are mainly used for clinical documentation (Andrew et al., 2023; Fox et al., 2017). As highlighted in 0, Chapter 5, Chapter 6 and Chapter 7, routine data from electronic health records come with several challenges, including inconsistent follow-ups (Fox et al., 2017; Goldstein, Navar, Pencina, & Ioannidis, 2017) and limited information quality (Savitz, Savitz, Fleming, Shah, & Go, 2020; Wright, Maloney, & Feblowitz, 2011). These challenges make routine data a poor substitute for smaller data obtained from time-consuming research studies, where follow-ups are consistent and additional covariates that could enhance the prediction of clinical outcomes are conventionally collected. This section of the discussion provides compelling evidence that large data do not compensate for the lack of consistent follow-ups and covariate collection in prediction models.

8.4.4 Dynamic prediction models do not always have better prediction performance

Multiple metrics for evaluating the performance of prediction models do not consistently agree on improving predictive performance. Generalisations about the predictive performance of dynamic prediction models over other approaches, even across these metrics, should not be made. This is because the context and characteristics of the data seem to drive much of the observed predictive performance. In the literature, several examples have demonstrated that dynamic prediction models have improved predictive performance compared to Cox proportional hazards-based prediction models (Basol, Goksuluk, Sipahioglu, & Karaagaoglu, 2021; Oulhaj et al., 2023). This thesis challenged these findings by comparing predictive

performance across the Cox proportional hazards-based and joint modelling approaches in routine chronic disease data. As seen in Chapter 5 and Chapter 6, the Cox proportional hazard-based prediction modelling approach had higher discrimination ability. In comparison, the joint modelling approach had lower calibration ability in the routine HIV data. In the routine diabetes data, the Cox proportional hazards-based prediction modelling approach had higher discrimination and comparable calibration ability to the joint modelling approach.

The inconsistency in predictive performance metrics provides compelling evidence that model performance is highly context-dependent and influenced by the specific characteristics of the data. For instance, routine electronic health records for conditions such as HIV or diabetes are often characterized by measurement noise and inconsistencies due to variations in the timing of assessments or data entry errors. Consequently, prediction models based on joint modelling may fail to accurately estimate individual-specific trajectories, thereby reducing the predictive performance of these dynamic models. This finding provides persuasive evidence that improved predictive performance is not always guaranteed for dynamic predictions.

8.4.5 Model complexity is not an answer to a lack of information in the data

Additional parameterisation of dynamic prediction models does not improve predictive performance if the data do not have consistent follow-ups and collection of sufficient covariates. In Chapter 5, the Cox proportional hazards-based prediction modelling approach was used to develop nomograms for prognosis prediction of virologic failure and glycaemic control in individuals living with HIV and T2DM in the Western Cape, South Africa. The chapter demonstrated that baseline biomarker values do not offer prognosis predictions that capture individuals' changes over time. As alluded to in Section 8.4.2 above, biomarkers such as HIV viral load and HbA1c measure disease progression over time, not just at baseline. This necessitated using a joint modelling-based prediction approach for prognosis prediction. Chapter 6 demonstrated that the joint modelling approach is a novel approach for prognosis prediction of virologic failure and glycaemic control, as it can model individuals' changes over time captured by repeated HIV viral load and HbA1c measurements, respectively. A more joint model complex (joint model with increased parameters) resulting in added computation costs was developed in Chapter 7 to improve predictions of virologic failure. The model had similar predictive performance compared to less parameterized joint model. Taken together as a body of work (Chapter 5, Chapter 6 and Chapter 7), the findings in this thesis indicate that a lack of

sufficient covariates in the data cannot be compensated for, even with model complexity and more parameterisation.

8.5 Recommendations

The discussion of critical findings above and recommendations alluded to across discussions in this thesis are summarised to provide recommendations regarding adopting dynamic prediction models and their predictive performance in resource-limited settings.

8.5.1 Type of disease characterises the length of follow-up and prediction time window

Dynamic prediction over long-term prediction time requires more consistent interim data where progression between disease status (healthy to unhealthy or unhealthy to healthy) can be adequately observed for individuals for whom predictions are to be made. These models are more appropriate for short-term or acute-type than chronic diseases. Chronic diseases need data with consistent follow-up and a high-quality biomarker that sufficiently measures disease progression in individuals.

8.5.2 Characteristics of longitudinal biomarkers are central to prediction models

Biomarkers with low variance are more suitable for dynamic prediction models. Dynamic prediction models need data that sufficiently capture the changes in the biomarker over consistent follow-ups for biomarkers with high variance. Researchers should have a good understanding of whether biomarkers show interim changes and that these changes are sufficiently observed to inform the performance of these biomarkers in dynamic prediction models. When using routine data, researchers should be aware of whether national guidelines for routine data collection were aligned with biomarker testing intervals to ensure that the information about the biological process of disease in the data is reasonably consistent in the observed follow-ups.

8.5.3 Availability of information in data from studies is critical to prediction models

Even large, routine data needs consistent follow-ups and sufficient covariates to be more useful than smaller data with better information quality on follow-ups and covariate collection. Smaller data with sufficient covariates and consistent follow-ups is more appropriate for dynamic prediction over large, low-quality information data. There should be consistent follow-ups and adequate covariates for dynamic prediction models to be more useful in predicting clinical outcomes for routine data.

8.5.4 Dynamic prediction models do not always have better prediction performance

The PROGRESS and TRIPOD guidelines should be used to identify data characteristics that inform a more balanced and contextualised development and validation of dynamic prediction models with improved predictive performance. Although dynamic prediction models using joint modelling is a novel approach for individualised prognosis prediction, these models universally do not have improved predictive performance in all contexts (Rizopoulos, Molenberghs, & Lesaffre, 2017). Researchers should not always assume that a dynamic prediction model always outperforms other models. More focus should be on when and how predictive performance can be limited or improved in dynamic prediction models. Alongside this recommendation, a conceptual understanding of the type of disease, biomarker type and variability, and sufficient covariates in data is needed to inform predictive performance.

8.5.5 Model complexity is not an answer to a lack of information in the data

Increasing the complexity and adding more parameters in prediction models do not make up for the lack of covariates and consistent follow-ups in data. Techniques and high-performance computing resources to overcome the computational burden of applying more complex and parameterised dynamic prediction models are unnecessary in settings where data quality is difficult to assess and may be poor.

8.6 Conclusion

Dynamic prediction models can potentially supplement decision-making in healthcare, but it is essential to consider their needs in resource-limited settings. There is a future for routine data in these resource-limited settings. Prospective thinking about information or covariates is needed to ensure routine data with better information quality.

This thesis addressed the gap in the literature concerning the scarcity of dynamic prediction models applied to routine data, particularly in the contexts of HIV and T2DM. Existing dynamic prediction models are often developed and validated in well-designed studies with consistent follow-ups and sufficient covariate collection. This leads to the assumption of universally improved predictive performance for dynamic prediction models over traditional methods like the Cox proportional hazards-based prediction model. This thesis challenged this assumption by comparing the predictive performance of the Cox and joint model on routine HIV and diabetes data. The findings in this thesis underscored the need for a balanced, conceptual rather than a universal approach to inform predictive performance, highlighting the central role that data characteristics, disease type and characteristics that biomarkers play in the development of prediction models.

Critical discussions emerged across five overarching themes: the type of disease characterises the length of follow-up and prediction time window, characteristics of longitudinal biomarkers are central to prediction models, consistent follow-ups and sufficient covariate collection are crucial to prediction models, dynamic prediction models do not always have better prediction performance, and model complexity is not an answer to lack of information in data.

Based on the thesis findings, the following recommendations are made. Firstly, a balanced and contextualised approach, such as the TRIPOD statement, should be used to identify data characteristics to inform the development and validation of dynamic prediction models. Secondly, high-quality biomarkers with low variability are more suitable for dynamic prediction. Third, large routine data is not a substitute for small data with consistent follow-ups and sufficient covariate collection. Fourth, if dynamic predictions are used for chronic diseases, there should be a high-quality biomarker for the disease with consistent follow-ups. Finally, ways to improve computation costs are not more meaningful if there is a lack of information in data.

In conclusion, this thesis contributes valuable insights into the application of dynamic prediction models in resource-limited settings, emphasising the need for a conceptual and balanced understanding of disease dynamics, biomarkers and data characteristics to predictive performance. Applying dynamic prediction models that incorporate these elements can potentially inform decision-making and subsequently improve the quality of life for individuals in care. Furthermore, despite the limited value of the prediction models due to inherent uncertainties, the findings reveal significant gaps in chronic disease monitoring, with many individuals having fewer than three documented visits. This underscores the need for health departments to strengthen monitoring efforts, ensure better access to care, and effectively communicate the importance of regular follow-up for individuals with chronic conditions.

8.7 References

Andrew, N. E., Beare, R., Ravipati, T., Parker, E., Snowdon, D., Naude, K., & Srikanth, V. (2023). Developing a linked electronic health record derived data platform to support research into healthy ageing. *Int J Popul Data Sci, 8*(1), 2129. doi:10.23889/ijpds.v8i1.2129

Basol, M., Goksuluk, D., Sipahioglu, M. H., & Karaagaoglu, E. (2021). Effect of Serum Albumin Changes on Mortality in Patients with Peritoneal Dialysis: A Joint Modeling Approach and Personalized Dynamic Risk Predictions. *Biomed Res Int, 2021*, 6612464. doi:10.1155/2021/6612464

Brennan, A. T., Lauren, E., Bor, J., George, J. A., Chetty, K., Mlisana, K., . . . Crowther, N. J. (2023). Gaps in the type 2 diabetes care cascade: a national perspective using South Africa's National Health Laboratory Service (NHLS) database. *BMC Health Serv Res, 23*(1), 1452. doi:10.1186/s12913-023-10318-9

Brilleman, S. L. C., M. J.;May, M. T.;Gompels, M.;Abrams, K. R. (2016). Joint longitudinal hurdle and time-to-event models: an application related to viral load and duration of the first treatment regimen in patients with HIV initiating therapy. *Stat Med, 35*(20), 3583-3594. doi:10.1002/sim.6948

Brody, T. (2016). *Clinical trials: study design, endpoints and biomarkers, drug safety, and FDA and ICH guidelines*: Academic press.

Burke, H. B. (1998). Integrating Multiple Clinical Tests to Increase Predictive Power. *Tumor Marker Protocols*, 3-10.

Burke, H. B. (2016). Predicting clinical outcomes using molecular biomarkers. *Biomarkers in cancer, 8*, BIC. S33380.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC medicine, 13*(1). doi:10.1186/s12916-014-0241-z

Fox, M. P., Maskew, M., Brennan, A. T., Evans, D., Onoya, D., Malete, G., . . . Sanne, I. (2017). Cohort profile: the Right to Care Clinical HIV Cohort, South Africa. *BMJ Open, 7*(6), e015620. doi:10.1136/bmjopen-2016-015620

Goldstein, B. A., Navar, A. M., Pencina, M. J., & Ioannidis, J. P. (2017). Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc, 24*(1), 198-208. doi:10.1093/jamia/ocw042

Human Sciences Research Council. (2023). The Sixth South African National Hiv Prevalence, Incidence, Behaviour And Communication Survey, 2022. Retrieved from https://hsrc.ac.za/special-projects/sabssm-survey-series/sabssmvi-media-pack-november-2023/

Maphumulo, W. T., & Bhengu, B. R. (2019). Challenges of quality improvement in the healthcare of South Africa post-apartheid: A critical review. *Curationis, 42*(1), e1-e9. doi:10.4102/curationis.v42i1.1901

Moons, K. G. M., de Groot, J. A. H., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D. G., . . . Collins, G. S. (2014). Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med, 11*(10), e1001744-e1001744. doi:10.1371/journal.pmed.1001744

Oulhaj, A., Aziz, F., Suliman, A., Iqbal, N., Coleman, R. L., Holman, R. R., & Sourij, H. (2023). Joint longitudinal and time-to-event modelling compared with standard Cox modelling in patients with type 2 diabetes with and without established cardiovascular disease: An analysis of the EXSCEL trial. *Diabetes Obes Metab, 25*(5), 1261-1270. doi:10.1111/dom.14975

Ren, X., Wang, J., & Luo, S. (2021). Dynamic prediction using joint models of longitudinal and recurrent event data: a Bayesian perspective. *Biostatistics & epidemiology, 5*(2), 250-266.

Rizopoulos, D. (2011a). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics, 67*(3), 819-829. doi:10.1111/j.1541-0420.2010.01546.x

Rizopoulos, D. (2011b). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics, 67*(3), 819-829.

Rizopoulos, D. (2012). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis, 56*(3), 491-501. doi:https://doi.org/10.1016/j.csda.2011.09.007

Rizopoulos, D., Molenberghs, G., & Lesaffre, E. M. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal, 59*(6), 1261-1276.

Sahadew, N., Pillay, S., & Singaram, V. S. (2022). Diabetes in the public healthcare sector of four South African provinces: A comparative analysis. *S Afr Med J, 112*(11), 855-859. doi:10.7196/SAMJ.2022.v112i11.16546

Savitz, S. T., Savitz, L. A., Fleming, N. S., Shah, N. D., & Go, A. S. (2020). How much can we trust electronic health record data? *Healthcare, 8*(3), 100444. doi:https://doi.org/10.1016/j.hjdsi.2020.100444

South African National Department of Health. (2014). Management of type 2 diabetes in adults at primary care level. Retrieved from https://knowledgehub.health.gov.za/elibrary/management-type-2-diabetes-adults-primary-care-level

South African National Department of Health. (2019). 2019 ART Clinical Guidelines for the Management of HIV in Adults, Pregnancy, Adolescents, Children, Infants and Neonates. Retrieved from https://www.health.gov.za/hiv-and-aids/

Statistics South Africa. (2021). Financial statistics of provincial government 2019/2020. Retrieved from https://www.statssa.gov.za/?p=14755

Steyerberg, E. W., Moons, K. G. M., Van der Windt, D. A., Hayden, J. A., Perel, P., Schroter, S., . . . Group, P. (2013). Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med, 10*(2), e1001381-e1001381. doi:10.1371/journal.pmed.1001381

Wang, X., Guo, F., Heller, K. A., & Dunson, D. B. (2015). *Parallelizing MCMC with Random partition trees.* Paper presented at the 28th International Conference on Neural Information Processing Systems, Montreal, Canada.

WHO. (2016). Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations–2016 update. Retrieved from https://www.who.int/publications/i/item/9789241511124

Wright, A., Maloney, F. L., & Feblowitz, J. C. (2011). Clinician attitudes toward and use of electronic problem lists: a thematic analysis. *BMC Med Inform Decis Mak, 11*, 36. doi:10.1186/1472-6947-11-36

Appendices

Appendix 1: Search Strategy used in PubMed Database.

1	"dynamic model*" [tw]
2	"dynamic prediction*" [tw]
3	"landmark model*" [tw]
4	"landmark prediction*" [tw]
5	HIV [tw]
6	spatial [tw]
7	1 OR 2 OR 3 OR 4
8	7 AND 5
9	8 NOT 6

Appendix 2: Data Extraction Guided by TRIPOD and CHARMS (Form 1).

Study; Country; Study design	Participants and setting	Recruitment dates	Follow-up	HIV related outcome	Covariates

Appendix 3: Data Extraction Guided by TRIPOD and CHARMS (Form 2).

Prediction model study	Modelling Approach	Outcome; Span of predictions	Predictive Performance Validation			
			Type of Validation	Discrimination	Calibration	Overall Performance

Appendix 4: R Code to Fit Cox PH Models

```
# Packages

library(survival)


##############################################

# Cox PH model - HIV routine data

cox_HIV_base <- coxph(Surv(times, vfailure == 1) ~ ageg + sex + logvl_1st, data = tdf.surv)

# Alternative Cox model - HIV routine data

cox_HIV_last <- coxph(Surv(times, vfailure == 1) ~ ageg + sex + logvl, data = tdf.surv)

##############################################


##############################################

# Cox PH model - T2DM routine data

cox_T2DM_base <- coxph(Surv(times, hba1c_contr == 1) ~ ageg + sex + hba1c_1st, data =
tdf.surv)

# Alternative Cox model - T2DM routine data

cox_T2DM_last <- coxph(Surv(times, hba1c_contr == 1) ~ ageg + sex + hba1c, data = tdf.surv)

##############################################


# Nomograms produced using the rms and DynNom packages

# AUCs and Brier scores produced using "pec", and "survivalROC" packages
```

Appendix 5: R Code to Fit JM Models to Routine HIV and T2DM Data

```
# Packages

library(nlme)

library(spline)

library(survival)

library(JMbayes2)


##############################################

# JM model - HIV routine data

##############################################


# *** Selecting an LME model

## ** nonlinear effect of follow-up time through B-splines having two internal knots

## ** random intercepts

tlme.int.JM <- lme(

  logvl ~ ageg + ns(tstart, 3, B = c(0.1, 317.1)) + sex,

  data = tdf.long,

  random = ~ 1 | pid,

  control = lmeControl(opt = "optim"))

## ** random intercepts & random slopes

tlme.slo.JM <- update(tlme.int.JM, random = ~ tstart | pid)

## * compare the LME models

anova(tlme.int.JM, tlme.slo.JM)
```

```
# ** Cox PH model
tcox.JM <- coxph(Surv(times, vfailure == 1) ~ ageg + sex, data = tdf.surv, model = TRUE)

# *** Selecting a JM model
## ** current value association structure
tjm.current <- jm(
  tcox.JM, tlme.slo.JM, time_var = "tstart",
  control = list(n_chains = 3L, n_iter = 4000L, n_burnin = 2000L, cores = 3L))
## ** Lagged effects association structure
fForms <- list("logvl" = ~ slope(logvl, direction = "back", eps = 26))
tjm.6mon <- jm(
  tcox.JM,
  tlme.slo.JM,
  time_var = "tstart",
  functional_forms = fForms,
  n_chains = 3L, n_iter = 4000L, n_burnin = 2000L, cores = 3L
)
## * compare the JM models
anova(tjm.current, tjm.6mon)

# *** Diagnostics
## ** Trace plot for the association parameter
traceplot(tjm.current, parm = "alphas", main = "")
```

```
## ** density plot for the association parameter
densplot(tjm.current, parm = "alphas", main = "")

# *** AUC and Brier scores and predictions calculated from tjm.current using "JMbayes2" R package

##############################################
# JM model - T2DM routine data
##############################################

# *** Selecting an LME model
## ** nonlinear effect of follow-up time through B-splines having two internal knots
## ** random intercepts
tlme.int.JM <- lme(
  A1c ~ ageg + ns(tstart, 3, B = c(1, 71.4)) + sex,
  data = tdf.long,
  random = ~ 1 | pid,
  control = lmeControl(opt = "optim"))
## ** random intercepts & random slopes
tlme.slo.JM <- update(tlme.int.JM, random = ~ tstart | pid)
## * compare the LME models
anova(tlme.int.JM, tlme.slo.JM)

# ** Cox PH model
tcox.JM <- coxph(Surv(times, hba1c_contr == 1) ~ ageg + sex, data = tdf.surv, model = TRUE)
```

```
# *** Selecting a JM model

## ** current value association structure

tjm.current <- jm(
  tcox.JM, tlme.slo.JM, time_var = "tstart",
  control = list(n_chains = 3L, n_iter = 4000L, n_burnin = 2000L, cores = 3L))

## ** cumulative effects association structure

fForms <- list("A1c" = ~ area(A1c))

tjm.cumulative <- jm(
  tcox.JM,
  tlme.slo.JM,
  time_var = "tstart",
  functional_forms = fForms,
  n_chains = 3L, n_iter = 4000L, n_burnin = 2000L, cores = 3L)

## * compare the JM models

anova(tjm.current, tjm.cumulative)


# *** Diagnostics

## ** Trace plot for the association parameter

traceplot(tjm.cumulative, parm = "alphas", main = "")

## ** density plot for the association parameter

densplot(tjm.cumulative, parm = "alphas", main = "")


# *** AUC and Brier scores and predictions calculated from tjm.cumulative using "JMbayes2"
R package
```

Appendix 6: R Code to Simulate Longitudinal Semi Continuous and Time-to-Event Data

```
# Packages

library(dplyr)

library(mvtnorm)

library(PermAlgo)


rm(list = ".Random.seed", envir = globalenv())

##############################################

# Simulation - From a Two-part Joint Model in Chapter 7


# Contain 3 Parts:

# Part 1 - Dimensions

# Part 2 - Semi continuous longitudinal data simulation

# Part 3 - Time to event simulation

##############################################


set.seed(101)

# 101:200 for replications


# *** Part 1 - Dimensions

n_ind <- 400

follow_dur <- 10

follow_gap <- 1
```

```
repeated_num <- (follow_dur/follow_gap) + 1 # no. repeated measurements
N_measures <- repeated_num * n_ind # total measurements
follow_times_temp <- seq(0, follow_dur, follow_gap) # follow-up times
follow_times <- rep(follow_times_temp, n_ind)

pid <- rep(seq_len(n_ind), each = repeated_num) # unique identifier
sex <- rep(rep(rbinom(n_ind, 1, 0.4)), each = repeated_num) # sex covariate

df <- data.frame(
  pid,
  follow_times,
  sex = sex
)

# *** Part 2 - Semi continuous longitudinal data simulation
# ** From a Two-part mixed-effects model

# * Design matrices for the binary and continuous parts
X_cont <- unname(model.matrix(~ follow_times + sex, data = df))
attr(X_cont, "assign") <- NULL
Z_cont <- unname(model.matrix(~ follow_times, data = df))
attr(Z_cont, "assign") <- NULL
X_bnr <- unname(model.matrix(~ follow_times + sex, data = df))
attr(X_bnr, "assign") <- NULL
```

```
Z_bnr <- unname(model.matrix(~ 1, data = df))
attr(Z_bnr, "assign") <- NULL

# * Parameters
beta_cont <- c(3, -0.3, -1)

alpha_bnr <- c(1, -1, -1) # scenario 1 - High
# alpha_bnr <- c(2.5, -1, -1) # scenario 2 - Moderate
# alpha_bnr <- c(5, -1, -1) # scenario 3 - Low

# * Random effects *
sigma_betaInt <- 0.45
sigma_betaSlope <- 0.25
sigma_alphaInt <- 0.5

corr_betaIntSlope <- 0.2
corr_betaInt_alphaInt <- 0.2
corr_betaSlope_alphaInt <- 0.7

cov_betaIntSlope <- sigma_betaInt * sigma_betaSlope * corr_betaIntSlope
cov_betaInt_alphaInt <- sigma_betaInt * sigma_alphaInt * corr_betaInt_alphaInt
cov_betaSlope_alphaInt <- sigma_betaSlope * sigma_alphaInt * corr_betaSlope_alphaInt

sigma_reff <- matrix(
```

```
  c(
    sigma_betaInt^2, cov_betaIntSlope, cov_betaInt_alphaInt,
    cov_betaIntSlope, sigma_betaSlope^2, cov_betaSlope_alphaInt,
    cov_betaInt_alphaInt, cov_betaSlope_alphaInt, sigma_alphaInt^2
  ),
  ncol = 3,
  nrow = 3)

mean_reff <- rep(0, 3)
reff <- mvtnorm::rmvnorm(
  n_ind,
  mean = mean_reff,
  sigma = sigma_reff)

# * error terms
sigma_e <- 0.6

# * Binary and continuous outcomes
lp_bnr <- as.vector(X_bnr %*% alpha_bnr + rowSums(Z_bnr * reff[df$pid, 3]))
df$Y_bnr <- rbinom(N_measures, size = 1, prob = plogis(lp_bnr))
lp_cont <- as.vector(X_cont %*% beta_cont + rowSums(Z_cont * reff[df$pid, 1:2]))
Y <- rnorm(nrow(df), lp_cont, sigma_e)

df$Y <- ifelse(lp_cont < 0, 0, lp_cont)
```

```
df <- df %>%
  dplyr::mutate(
    Y = ifelse(Y_bnr == 1, Y, 0)
  )

# * Simulated semi continuous longitudinal data
longdf_temp <- df

# *** Part 2 - Time to event simulation

# * Matrix of covariates
X_eventmat <- matrix(ncol = 2, nrow = N_measures)
X_eventmat[, 1] <- longdf_temp[, "sex"]
X_eventmat[, 2] <- longdf_temp[, "Y"]

# * Parameters
gamma_event1 <- -1 # regression parameters
assoc_par <- 0.1 # association parameter

# ** Event and censoring times **
eventRandom <- round(rexp(n_ind, 0.08) + 1, 0)
censorRandom <- runif(n_ind, 1, repeated_num)

# ** Time to event data conditional on estimated longitudinal outcome
```

```
survdf_temp <- permalgorithm(
  numSubjects = n_ind,
  maxTime = repeated_num,
  Xmat = X_eventmat,  XmatNames = c("sex", "Y"),
  eventRandom = eventRandom, censorRandom = censorRandom,
  betas = c(gamma_event1, assoc_par)
)

# ** Event and censoring indicator and times
indic <- c(which(diff(survdf_temp[, "Id"]) == 1), dim(survdf_temp)[1])
survdf_temp2 <- survdf_temp[indic, c("Id", "Event", "Stop", "sex", "Y")]
survdf_temp2$times <- follow_times_temp[survdf_temp2$Stop + 1]

# ** Simulated time to event data
survdf <- survdf_temp2[, c("Id", "Event", "times", "sex", "Y")]
names(survdf) <- c("pid", "event", "times", "sex", "Y")
longdf_temp2 <- survdf_temp[, c("Id", "Event", "Start", "sex", "Y")]
longdf_temp2$tstart <- follow_times_temp[longdf_temp2$Start+1]
# * Simulated time to event data - tstart-stop format
longdf_temp3 <- longdf_temp2[, c("Id", "Event", "tstart", "sex", "Y")]
names(longdf_temp3) <- c("pid", "event", "tstart", "sex", "Y_hurdle")

# ** Simulated semi continuous longitudinal and time to event data
visits <- follow_times_temp[which(follow_times_temp -
round(follow_times_temp/follow_gap, 0) * follow_gap == 0)]
```

```
longdf <- longdf_temp3[longdf_temp3$tstart %in% visits, ]
longdf$Y_bnr <- ifelse(longdf$Y_hurdle > 0, 1, 0)
names(survdf) <- c("pid", "event", "times", "sex", "Y_hurdle")

# * Extract binary part longitudinal data
longdf %>% dplyr::mutate(
  pid_bnr = pid
) %>% dplyr::select(
  pid, pid_bnr,
  tstart, Y_hurdle, Y_bnr, sex) -> dataB

# * Extract continuous part longitudinal data
dataB %>% filter(
  Y_bnr == 1 # above detection or positive
) %>% dplyr::mutate(
  pid_cont = pid_bnr
) %>% dplyr::select(
  pid, pid_bnr, pid_cont,
  tstart, Y_hurdle, Y_bnr, sex) -> dataC
names(dataC) <- c("pid", "pid_bnr", "pid_cont", "tstart", "Y_cont", "Y_bnr", "sex")

# ** Data preparation for modelling
# * update PIDs
# binary part
```

```
dataB %>%
  dplyr::group_by(pid_bnr) %>%
  dplyr::slice(n()) %>%
  dplyr::select(pid_bnr) %>%
  dplyr::ungroup() -> dataB_1st
dataB_1st <- dataB_1st %>% dplyr::mutate(ids = 1:nrow(dataB_1st))
dataB_1st %>%
  dplyr::left_join(dataB, by = "pid_bnr") %>%
  dplyr::select(-pid_bnr) -> dataB
dataB <- dataB %>%
  dplyr::group_by(pid) %>%
  dplyr::mutate(pid_bnr = ids) %>%
  dplyr::ungroup() %>%
  dplyr::select(pid_bnr, tstart, Y_hurdle, Y_bnr, sex)

# continuous part
dataC %>%
  dplyr::group_by(pid_cont) %>%
  dplyr::slice(n()) %>%
  dplyr::select(pid_cont) %>%
  dplyr::ungroup() -> dataC_1st
dataC_1st <- dataC_1st %>% dplyr::mutate(ids = 1:nrow(dataC_1st))
dataC_1st %>% dplyr::left_join(dataC, by = "pid_cont") -> dataC
dataC <- dataC %>%
```

```
  dplyr::group_by(pid_cont) %>%
  dplyr::mutate(pid_contt = ids) %>%
  dplyr::ungroup() %>%
  dplyr::select(-pid_cont)
dataC <- dataC %>%
  dplyr::group_by(pid_contt) %>%
  dplyr::mutate(pid_cont = pid_contt) %>%
  dplyr::ungroup() %>%
  dplyr::select(pid_cont, tstart, Y_bnr, Y_cont, sex)

# ** Development (75%) and validation sets (25%)
set <- sample(unique(survdf$pid), 100)

tlongdf <- longdf[!longdf$pid %in% set, ]
tdataB <- dataB[!dataB$pid_bnr %in% set, ]
tdataC <- dataC[!dataC$pid_cont %in% set, ]
vlongdf <- longdf[longdf$pid %in% set, ]
vdataB <- dataB[dataB$pid_bnr %in% set, ]
vdataC <- dataC[dataC$pid_cont %in% set, ]

tsurvdf <- survdf[!survdf$pid %in% set, ]
vsurvdf <- survdf[survdf$pid %in% set, ]

# * update PIDs
```

```
# Development set - binary part
tdataB %>%
  dplyr::group_by(pid_bnr) %>%
  dplyr::slice(n()) %>%
  dplyr::select(pid_bnr) %>%
  dplyr::ungroup() -> tdataB_1st
tdataB_1st <- tdataB_1st %>% dplyr::mutate(ids = 1:nrow(tdataB_1st))
tdataB_1st %>%
  dplyr::left_join(tdataB, by = "pid_bnr") %>%
  dplyr::select(-pid_bnr) -> tdataB
tdataB <- tdataB %>%
  dplyr::mutate(pid_bnr = ids) %>%
  dplyr::select(pid_bnr, tstart, Y_hurdle, Y_bnr, sex)

# * update PIDs
# Development set - continuous part
tdataC %>%
  dplyr::group_by(pid_cont) %>%
  dplyr::slice(n()) %>%
  dplyr::select(pid_cont) %>%
  dplyr::ungroup() -> tdataC_1st
tdataC_1st <- tdataC_1st %>% dplyr::mutate(ids = 1:nrow(tdataC_1st))
tdataC_1st %>% dplyr::left_join(tdataC, by = "pid_cont") -> tdataC
tdataC <- tdataC %>%
```

```
  dplyr::group_by(pid_cont) %>%
  dplyr::mutate(pid_contt = ids) %>%
  dplyr::ungroup() %>%
  dplyr::select(-pid_cont)
tdataC <- tdataC %>%
  dplyr::group_by(pid_contt) %>%
  dplyr::mutate(pid_cont = pid_contt) %>%
  dplyr::ungroup() %>%
  dplyr::select(pid_cont, tstart, Y_bnr, Y_cont, sex)

# Development set - Time to event submodel
tsurvdf %>% dplyr::mutate(ids = 1:nrow(tsurvdf)) %>%
  dplyr::select(-pid) -> tsurvdf
tsurvdf <- tsurvdf %>% dplyr::mutate(pid = ids) %>%
  dplyr::select(pid, event, times, Y_hurdle, sex)

# * update PIDs
# Validation set - binary part
vdataB %>%
  dplyr::group_by(pid_bnr) %>%
  dplyr::slice(n()) %>%
  dplyr::select(pid_bnr) %>%
  dplyr::ungroup() -> vdataB_1st
vdataB_1st <- vdataB_1st %>% dplyr::mutate(ids = 1:nrow(vdataB_1st))
```

```
vdataB_1st %>%
  dplyr::left_join(vdataB, by = "pid_bnr") %>%
  dplyr::select(-pid_bnr) -> vdataB
vdataB <- vdataB %>%
  dplyr::mutate(pid_bnr = ids) %>%
  dplyr::select(pid_bnr, tstart, Y_hurdle, Y_bnr, sex)

# Validation set - continuous part
vdataC %>%
  dplyr::group_by(pid_cont) %>%
  dplyr::slice(n()) %>%
  dplyr::select(pid_cont) %>%
  dplyr::ungroup() -> vdataC_1st
vdataC_1st <- vdataC_1st %>% dplyr::mutate(ids = 1:nrow(vdataC_1st))
vdataC_1st %>% dplyr::left_join(vdataC, by = "pid_cont") -> vdataC
vdataC <- vdataC %>%
  dplyr::group_by(pid_cont) %>%
  dplyr::mutate(pid_contt = ids) %>%
  dplyr::ungroup() %>%
  dplyr::select(-pid_cont)
vdataC <- vdataC %>%
  dplyr::group_by(pid_contt) %>%
  dplyr::mutate(pid_cont = pid_contt) %>%
  dplyr::ungroup() %>%
```

```
  dplyr::select(pid_cont, tstart, Y_bnr, Y_cont, sex)


# Validation set - Time to event submodel
vsurvdf %>%
  dplyr::mutate(ids = 1:nrow(vsurvdf)) %>%
  dplyr::select(-pid) -> vsurvdf
vsurvdf <- vsurvdf %>% dplyr::mutate(pid = ids) %>%
  dplyr::select(pid, event, times, Y_hurdle, sex)


# random effects
cbind(reff, as.numeric(1:nrow(survdf))) -> u_hurdle
t_uhurdle <- u_hurdle[!u_hurdle[, 4] %in% set, 1:3]
v_uhurdle <- u_hurdle[u_hurdle[, 4] %in% set, 1:3]


rm(list = c(
  "alpha_bnr", "assoc_par", "beta_cont", "censorRandom",
  "corr_betaInt_alphaInt", "corr_betaIntSlope", "corr_betaSlope_alphaInt",
  "cov_betaInt_alphaInt", "cov_betaIntSlope", "cov_betaSlope_alphaInt",
  "dataB_1st", "dataC_1st", "df",
  "eventRandom", "follow_dur", "follow_gap", "follow_times", "follow_times_temp",
  "gamma_event1", "indic",
  "longdf_temp", "longdf_temp2", "longdf_temp3",
  "lp_bnr", "lp_cont",
  "mean_reff",
```

"n_ind", "N_measures",

"pid",

"repeated_num",

"sex",

"sigma_alphaInt", "sigma_betaInt", "sigma_betaSlope", "sigma_e",

"survdf_temp", "survdf_temp2",

"visits",

"X_bnr", "X_cont", "X_eventmat", "Y", "Z_bnr", "Z_cont",

"reff", "set",

"tdataB_1st", "tdataC_1st", "u_hurdle", "vdataB_1st", "vdataC_1st"))

Appendix 7: R Code to Fit the JM Model to Simulated Data

```
load("C:/Users/Documents/PhD/Appendix/Chap7_Simulation.RData")

# Packages

library(dplyr)

library(nlme)

library(survival)

library(JMbayes2)


################################################

# *** Data preparation

################################################

set.seed(101)

# 101:200 for replications


tdataB$pid <- tdataB$pid_bnr

tsurvdf_updated <- tsurvdf %>% dplyr::select(pid, event, times)

tsurvdf_updated %>%

  dplyr::left_join(tdataB, by = "pid") %>%

  dplyr::select(pid, event, times, tstart, Y_hurdle, Y_bnr, sex) -> tdataBB


vdataB$pid <- vdataB$pid_bnr

vsurvdf_updated <- vsurvdf %>% dplyr::select(pid, event, times)

vsurvdf_updated %>%

  dplyr::left_join(vdataB, by = "pid") %>%
```

```
  dplyr::select(pid, event, times, tstart, Y_hurdle, Y_bnr, sex) -> vdataBB


##############################################
# *** Fitting JM
##############################################


# *** Random Intercepts and Slopes LME
model.lme <- lme(
  Y_hurdle ~ tstart + sex,
  random = ~ tstart | pid,
  data = tdataBB,
  control=lmeControl(maxIter=1000, msMaxIter=1000, niterEM=1000, opt='optim'))
# *** Cox model
model.Cox <- coxph(Surv(times, event == 1) ~ sex, data = tsurvdf, model = TRUE)
# *** JM model with current value association structure
model.JM <- jm(
  model.Cox, model.lme, time_var = "tstart",
  control = list(n_chains = 3L, n_iter = 2000L, n_burnin = 1000L, n_thin = 1L, cores = 3L))


# Mean Bias, MSE, and CP for parameters extracted from 100 repetitions of the simulation and
model.JM fit
# Mean AUC and Brier scores extracted from 100 repetitions of the simulation and model.JM
fit using "JMbayes2" R package
```

Appendix 8: Stan Code to Fit Two-Part JM Model to Simulated and Routine HIV Data

```
data {
  // Binary part dimensions
  int<lower=0> N_obs;                          // Number of observations
  int<lower=0> n_ind;                          // Number of individuals
  int<lower=1, upper=n_ind> pid_bnr[N_obs];        // Individual PID
  int<lower=0, upper=1> Y_bnr[N_obs];              // Binary outcome (0 or 1)

  // Continuous part dimensions
  int<lower=0, upper=N_obs> N_cont;                // Number of observations
  int<lower=0, upper=n_ind> n_ind_cont;            // Number of individuals with Y_cont > 0
  int<lower=1, upper=n_ind_cont> pid_cont[N_cont];  // Individual PID for continuous part
  vector<lower=0>[N_cont] Y_cont;                  // Positive continuous outcomes

  // Binary part - fixed and random effects
  int<lower=1> p_bnr;                          // Number of fixed effects (including intercept)
  int<lower=1> q_bnr;                          // Number of random effects (only random   intercept)
  row_vector[p_bnr] X_bnr[N_obs];                  // Fixed effects design matrix
  row_vector[q_bnr] Z_bnr[N_obs];                  // Individual random effect design matrix

  // Continuous part - fixed and random effects
  int<lower=1> p_cont;                         // Number of fixed effects (including intercept)
  int<lower=1> q_cont;                  // Number of random effects (including random intercept)
```

```
  row_vector[p_cont] X_cont[N_cont];              // Fixed effects design matrix
  row_vector[q_cont] Z_cont[N_cont];              // Individual random effect design matrix

  // Hurdle submodel - linking the binary and continuous parts
  int<lower=3> q_hurdle;       // Number of random effects across binary and continuous parts
  vector[q_hurdle] zeros_uhurdle; // Zero mean vector across binary and continuous parts random effects
  vector[q_hurdle] u_hurdle[n_ind];    // Combined random effects

  // Time to event submodel – dimensions
  vector<lower=0> [n_ind] time;                 // event or censoring time
  real<lower=0, upper=1> event[n_ind];          // event indicator (1 if event and 0 otherwise)

  // Time to event submodel – dimensions (Integrating random effects)
  vector[15] quadnodes;                      // Gaussian quadrature K nodes
  real quadweight[15];                       // Gaussian quadrature weights

  // Time to event submodel – covariates
  int<lower=1> p_event;                      // Number of covariates
  row_vector[p_event] X_event[n_ind];           // Covariates design matrix

  // Time to event submodel – covariates at event time
  row_vector[p_bnr] X_bnr_eventtime[n_ind];      // Binary part- Fixed effects matrix
  row_vector[q_bnr] Z_bnr_eventtime[n_ind];   // Binary part- Indiv random effects matrix
  row_vector[p_cont] X_cont_eventtime[n_ind];     // Continuous part- Fixed effects matrix
```

```
row_vector[q_cont] Z_cont_eventtime[n_ind]; // Continuous part- Indiv random eff matrix

// Time to event submodel – covariates at quadrature K nodes
row_vector[p_bnr] X_bnr_quadnode[n_ind, 15];   // Binary part- Fixed effects design matrix
row_vector[q_bnr] Z_bnr_quadnode[n_ind, 15];  // Binary part- Indiv random eff matrix
row_vector[p_cont] X_cont_quadnode[n_ind, 15];  // Continuous part- Fixed effects matrix
row_vector[q_cont] Z_cont_quadnode[n_ind, 15];  // Continuous part- Indiv rand eff matrix

/*
--------------------------------------------------------------------------
**********************************
***** validation set ****
**********************************
--------------------------------------------------------------------------
*/
int<lower=0> n_ind_pred;                    // num. individuals validation set
vector<lower=0> [n_ind_pred] time_pred;        // event or censoring time
real<lower=0, upper=1> event_pred[n_ind_pred]; // event indicator

// Time to event submodel - covariates
row_vector[p_event] X_event_pred[n_ind_pred];

// Time to event submodel – covariate at event time
row_vector[p_bnr] X_bnr_eventtime_pred[n_ind_pred];   // Binary part- Fixed eff matrix
```

```
  row_vector[q_bnr] Z_bnr_eventtime_pred[n_ind_pred];    // Binary part- Individual rand eff matrix
  row_vector[p_cont] X_cont_eventtime_pred[n_ind_pred]; // Continuous part- Fixed effects design matrix
  row_vector[q_cont] Z_cont_eventtime_pred[n_ind_pred]; // Continuous part- Indiv random effect design matrix
}

transformed data {
  // priors
  real<lower=0> prior_mean_sigma_e;
  real<lower=0> prior_sd_sigma_e;

  // standardized quadrature nodes to an unstandardized value
  vector<lower=0>[n_ind] time_midpt;                              // event time midpoint
  vector[15] quadnode[n_ind];            // unstandardized Gaussian quadrature K nodes
  time_midpt = time/2;                                            // calculate event time midpoint

  // unstandardized Gaussian quadrature K nodes for each individual
  for(i in 1:n_ind){
    quadnode[i] = (quadnodes*time_midpt[i]) + time_midpt[i];
  }

  prior_mean_sigma_e = 0.6;
  prior_sd_sigma_e = 0.0025;
```

```
}

parameters {
  // Hurdle submodel
  vector[p_bnr] alpha_bnr;                    // Fixed effects intercept and slopes - binary part
  vector[p_cont] beta_cont;          // Fixed effects intercept and slopes - continuous part
  real<lower=0> sigma_e;                    // error term SD - continuous part

  cholesky_factor_corr[q_hurdle] Lcorr_uhurdle; // Cholesky factor
  vector<lower=0>[q_hurdle] sd_uhurdle;        // Combined random effects SD

  // Time to event submodel
  vector[p_event] gamma_event;                          // coefficient for covariates
  real<lower=0> assoc_par;                  // current value association parameter
  real<lower=0> weibull_shape;               // Weibull shape parameter
}

transformed parameters {
  // linear predictor for binary part
  real mu_bnr[N_obs];
  // linear predictor for binary part - event time
  real mu_bnr_eventtime[n_ind];
  // linear predictor for binary part - unstandardized Gaussian quadrature K nodes
  vector[15] mu_bnr_quadnode[n_ind];
```

```
// linear predictor for continuous part
vector[N_cont] mu_cont;
// linear predictor for continuous part - event time
real mu_cont_eventtime[n_ind];
// linear predictor for continuous part - unstandardized Gaussian quadrature K nodes
vector[15] mu_cont_quadnode[n_ind];

// combined linear predictor - event time
real cur_marker_eventtime[n_ind];
// combined linear predictor - quadrature K nodes
vector[15] cur_marker_quadnode[n_ind];

// log hazard likelihood - event time
real log_haz_eventtime[n_ind];
// log hazard likelihood - unstandardized Gaussian quadrature K nodes
vector[15] log_haz_quadnode[n_ind];
// log hazard likelihood - weighted unstandardized Gaussian quadrature K nodes
vector[15] log_haz_quadnode_weight[n_ind];
// log survival likelihood
real log_surv_eventtime[n_ind];
// event log likelihood
real log_event[n_ind];
```

```
  // Weibull scale parameter
  vector[n_ind] weibull_scale;

  // extract individual random effects
  vector[q_bnr] reff_bnr_inter[n_ind];
  vector[q_cont] reff_cont[n_ind];

  for (i in 1:n_ind) {
    reff_cont[i] = head(u_hurdle[i], q_cont);
    reff_bnr_inter[i] = tail(u_hurdle[i], q_bnr);
  }

  // linear predictors for the binary and continuous parts
  for (j in 1:N_obs) {
    mu_bnr[j] = dot_product(X_bnr[j], alpha_bnr) + dot_product(Z_bnr[j],
reff_bnr_inter[pid_bnr[j]]);
  }

  for (j in 1:N_cont) {
    mu_cont[j] = dot_product(X_cont[j], beta_cont) + dot_product(Z_cont[j],
reff_cont[pid_cont[j]]);
  }

  // event log likelihood based on the log survival and log hazard likelihoods
  for(i in 1:n_ind) {
```

```
  mu_bnr_eventtime[i] = inv_logit(dot_product(X_bnr_eventtime[i], alpha_bnr) +
dot_product(Z_bnr_eventtime[i], reff_bnr_inter[i]));

  mu_cont_eventtime[i] = dot_product(X_cont_eventtime[i], beta_cont) +
dot_product(Z_cont_eventtime[i], reff_cont[i]);

  cur_marker_eventtime[i] = mu_bnr_eventtime[i] * mu_cont_eventtime[i];

  // Weibull scale parameter

  weibull_scale[i] = exp(dot_product(X_event[i], gamma_event) +
cur_marker_eventtime[i]*assoc_par);

  // log hazard likelihood

  if(event[i] == 1) {

   log_haz_eventtime[i] = log(weibull_shape) + ((weibull_shape - 1) * log(time[i])) +

    dot_product(X_event[i], gamma_event) + (cur_marker_eventtime[i]*assoc_par);

  } else{

   log_haz_eventtime[i] = 0;

  }


  for(k in 1:15){

   mu_bnr_quadnode[i, k] = inv_logit(dot_product(X_bnr_quadnode[i, k], alpha_bnr) +
dot_product(Z_bnr_quadnode[i, k], reff_bnr_inter[i]));

   mu_cont_quadnode[i, k] = dot_product(X_cont_quadnode[i, k], beta_cont) +
dot_product(Z_cont_quadnode[i, k], reff_cont[i]);

   cur_marker_quadnode[i, k] = mu_bnr_quadnode[i, k] * mu_cont_quadnode[i, k];

   log_haz_quadnode[i, k] = log(weibull_shape) + ((weibull_shape - 1) * log(quadnode[i, k]))
+

    dot_product(X_event[i], gamma_event) + (cur_marker_quadnode[i, k]*assoc_par);
```

```
      // log hazard likelihood at weighted unstandardized Gaussian quadrature K nodes
      log_haz_quadnode_weight[i, k] = quadweight[k] * exp(log_haz_quadnode[i, k]);
    }
    // log survival likelihood
    log_surv_eventtime[i] = -(time_midpt[i] * sum(log_haz_quadnode_weight[i]));
    // event log likelihood
    log_event[i] = log_haz_eventtime[i] + log_surv_eventtime[i];
  }
}



model {
  // priors
  alpha_bnr ~ normal(0, 1);
  beta_cont ~ normal(0, 1);
  sigma_e ~ normal(prior_mean_sigma_e, prior_sd_sigma_e);
  sd_uhurdle ~ cauchy(0, 5);
  Lcorr_uhurdle ~ lkj_corr_cholesky(1);

  // combined random effects covariance matrix
  u_hurdle ~ multi_normal_cholesky(zeros_uhurdle, diag_pre_multiply(sd_uhurdle,
Lcorr_uhurdle));

  gamma_event ~ normal(0, 1);
  assoc_par ~ normal(0, 1);
```

```
  weibull_shape ~ gamma(2, 1);

  // Likelihoods
  Y_bnr ~ bernoulli_logit(mu_bnr);
  Y_cont ~ normal(mu_cont, sigma_e);

  target += log_event;

  for (i in 1:n_ind) {
    time[i] ~ weibull(weibull_shape, (weibull_scale[i]^(-1/ weibull_shape)));
  }

}

generated quantities {
  real mu_bnr_eventtime_pred[n_ind_pred];          // logit prob.
  real mu_cont_eventtime_pred[n_ind_pred];         // mean
  real cur_marker_eventtime_pred[n_ind_pred];      // current value

  vector[n_ind_pred] weibull_scale_pred;           // weigbull scale
  vector[n_ind] hat_time;                          // posterior
  vector[n_ind_pred] pred_time_pred;               // predicted

  // random effects variance-covariance matrix
```

```
  matrix[q_hurdle, q_hurdle] omega_uhurdle;
  matrix[q_hurdle, q_hurdle] sigma_uhurdle;
  omega_uhurdle = multiply_lower_tri_self_transpose(Lcorr_uhurdle);
  sigma_uhurdle = quad_form_diag(omega_uhurdle, sd_uhurdle);

  // for posterior predictive checks
  for(i in 1:n_ind){
    if (event[i] == 1) {
      hat_time[i] = weibull_rng(weibull_shape, weibull_scale[i]);
      hat_time[i] = fmin(hat_time[i], time[i]);
    } else {
      hat_time[i] = time[i];
    }
  }

  // for posterior predictive checks - validation
  for(i in 1:n_ind_pred){
    mu_bnr_eventtime_pred[i] = inv_logit(dot_product(X_bnr_eventtime_pred[i], alpha_bnr) +
dot_product(Z_bnr_eventtime_pred[i], reff_bnr_inter[i]));
    mu_cont_eventtime_pred[i] = dot_product(X_cont_eventtime_pred[i], beta_cont) +
dot_product(Z_cont_eventtime_pred[i], reff_cont[i]);
    cur_marker_eventtime_pred[i] = mu_bnr_eventtime_pred[i] * mu_cont_eventtime_pred[i];

    // Weibull scale parameter
```

```
    weibull_scale_pred[i] = exp(dot_product(X_event_pred[i], gamma_event) +
cur_marker_eventtime_pred[i]*assoc_par);


    if (event_pred[i] == 1) {
      pred_time_pred[i] = weibull_rng(weibull_shape, weibull_scale_pred[i]);
      pred_time_pred[i] = fmin(pred_time_pred[i], time_pred[i]);
    } else {
      pred_time_pred[i] = time_pred[i];
    }
  }
}
```

Appendix 9: R Code to Fit Two-Part JM Model to Simulated Data

```
load("C:/Users/Documents/PhD/Appendix/Chap7_Simulation.RData")

setwd("C:/Users/Documents/PhD/Appendix")

# Packages

library(statmod)

library(rstan)


options(mc.cores = parallel::detectCores())

rstan_options(auto_write = TRUE)


################################################

# *** Two-part joint model - Data preparation

################################################

set.seed(151)

# 151:250 for replications


# *** Development set

# ** Semi continuous longitudinal data

X_bnr <- model.matrix(~ tstart + sex, data = tdataB)

Z_bnr <- model.matrix(~ 1, data = tdataB)

p_bnr <- ncol(X_bnr)

q_bnr <- ncol(Z_bnr)

X_cont <- model.matrix(~ tstart + sex, data = tdataC)

Z_cont <- model.matrix(~ tstart, data = tdataC)
```

```
p_cont <- ncol(X_cont)

q_cont <- ncol(Z_cont)


# * Dimensions

N_obs <- nrow(X_bnr)

n_ind <-  nlevels(as.factor(tdataB$pid_bnr))

pid_bnr <- as.integer(tdataB$pid_bnr)

Y_bnr <- tdataB$Y_bnr

N_cont <- nrow(X_cont)

n_ind_cont <- nlevels(as.factor(tdataC$pid_cont))

pid_cont <- as.integer(tdataC$pid_cont)

Y_cont <- tdataC$Y_cont


# * Random effects

zeros_uhurdle <- rep(0, 3)

q_hurdle <- 3


# ** Time to event data

time <- tsurvdf$times

event <- tsurvdf$event

X_event <- model.matrix(~ -1 + sex, data = tsurvdf)

p_event <- ncol(X_event)

X_bnr_eventtime <- model.matrix(~ times + sex, data = tsurvdf)

Z_bnr_eventtime <- model.matrix(~ 1, data = tsurvdf)
```

```
X_cont_eventtime <- model.matrix(~ times + sex, data = tsurvdf)

Z_cont_eventtime <- model.matrix(~ times, data = tsurvdf)


# * Gauss-Legendre quadrature

gau_quad <- gauss.quad(15, kind = "legendre")

quadnodes <- gau_quad$nodes

quadweight <- gau_quad$weights

quads <- table(rep(1:nrow(tsurvdf), each = 15))

X_bnr_quadnode <- array(1, dim = c(nrow(tsurvdf), max(quads), 3))

X_bnr_quadnode[, , 2] <- tsurvdf$times

X_bnr_quadnode[, , 3] <- tsurvdf$sex

Z_bnr_quadnode <- array(1, dim = c(nrow(tsurvdf), max(quads), 1))

X_cont_quadnode <- array(1, dim = c(nrow(tsurvdf), max(quads), 3))

X_cont_quadnode[, , 2] <- tsurvdf$times

X_cont_quadnode[, , 3] <- tsurvdf$sex

Z_cont_quadnode <- array(1, dim = c(nrow(tsurvdf), max(quads), 2))

Z_cont_quadnode[, , 2] <- tsurvdf$times


# *** Validation set

# * Dimensions

n_ind_pred <-  nlevels(as.factor(vdataB$pid_bnr))

time_pred <- vsurvdf$times

event_pred <- vsurvdf$event

# * Covariates
```

```
X_bnr_eventtime_pred <- model.matrix(~ times + sex, data = vsurvdf)

Z_bnr_eventtime_pred <- model.matrix(~ 1, data = vsurvdf)

X_cont_eventtime_pred <- model.matrix(~ times + sex, data = vsurvdf)

Z_cont_eventtime_pred <- model.matrix(~ times, data = vsurvdf)

X_event_pred <- model.matrix(~ -1 + sex, data = vsurvdf)


stanData_hurdle <- list(

  X_bnr = X_bnr,

  Z_bnr = Z_bnr,

  p_cont = p_cont,

  q_cont = q_cont,

  X_cont = X_cont,

  Z_cont = Z_cont,

  p_bnr = p_bnr,

  q_bnr = q_bnr,

  N_obs = N_obs,

  n_ind = n_ind,

  pid_bnr = pid_bnr,

  Y_bnr = Y_bnr,

  N_cont = N_cont,

  n_ind_cont = n_ind_cont,

  pid_cont = pid_cont,

  Y_cont = Y_cont,

  zeros_uhurdle = zeros_uhurdle,
```

```
q_hurdle = q_hurdle,

u_hurdle = t_uhurdle,

time = time,

event = event,

X_event = X_event,

p_event = p_event,

X_bnr_eventtime = X_bnr_eventtime,

Z_bnr_eventtime =  Z_bnr_eventtime,

X_cont_eventtime = X_cont_eventtime,

Z_cont_eventtime = Z_cont_eventtime,

quadnodes =  quadnodes,

quadweight = quadweight,

X_bnr_quadnode = X_bnr_quadnode,

Z_bnr_quadnode = Z_bnr_quadnode,

X_cont_quadnode =  X_cont_quadnode,

Z_cont_quadnode = Z_cont_quadnode,

X_bnr_eventtime_pred = X_bnr_eventtime_pred,

Z_bnr_eventtime_pred = Z_bnr_eventtime_pred,

X_cont_eventtime_pred = X_cont_eventtime_pred,

Z_cont_eventtime_pred = Z_cont_eventtime_pred,

n_ind_pred = n_ind_pred,

time_pred = vsurvdf$times,

event_pred  = vsurvdf$event,

X_event_pred = X_event_pred)
```

```
rm(list = c(
  "event", "event_pred", "gau_quad", "N_cont", "n_ind", "n_ind_cont", "n_ind_pred",
  "N_obs", "p_bnr", "p_cont", "p_event",
  "pid_bnr", "pid_cont", "q_bnr", "q_cont", "q_hurdle", "quadnodes", "quads", "quadweight",
  "time", "time_pred",
  "X_bnr", "X_bnr_eventtime", "X_bnr_eventtime_pred", "X_bnr_quadnode",
  "X_cont", "X_cont_eventtime", "X_cont_eventtime_pred", "X_cont_quadnode",
  "X_event", "X_event_pred",
  "Y_bnr", "Y_cont",
  "Z_bnr", "Z_bnr_eventtime", "Z_bnr_eventtime_pred", "Z_bnr_quadnode",
  "Z_cont", "Z_cont_eventtime", "Z_cont_eventtime_pred", "Z_cont_quadnode",
  "zeros_uhurdle"))

Stanmod <- stan_model("TPJM_Stancode.stan")

TPJM_sim <- sampling(
  Stanmod,
  data = stanData_hurdle,
  chains = 3,
  iter = 2000,
  verbose = FALSE,
  cores = getOption("mc.cores", 3L))
```

Mean Bias, MSE, and CP for parameters extracted from 100 repetitions of the simulation and TPJM_sim fit

Mean AUC and Brier scores extracted from 100 repetitions of the simulation and model TPJM_sim fit

through riskRegression, "pec", and "survivalROC" R packages

Appendix 10: R Code to Fit the JM Model to a Subset Routine HIV Data

```
load("C:/Users/Documents/PhD/Appendix/Chap7_HIVRoutine.RData")

# Packages

library(dplyr)

library(nlme)

library(survival)

library(JMbayes2)


##############################################

# *** Data preparation

##############################################


tdataBnrr$pid <- tdataBnrr$pid_bnr

tdataBnrr_temp <- tdataBnrr %>% dplyr::select(pid, tstart, logvl, vldl, sex)

tdataSurvv_temp <- tdataSurvv %>% dplyr::select(pid, vfailure, times)

tdataSurvv_temp %>% dplyr::left_join(tdataBnrr_temp, by = "pid") -> tdataBB

tdataSSurv <- tdataSurvv %>%

  dplyr::select(pid, vfailure, times, tstart, logvl, vldl, sex)


vdataBnrr$pid <- vdataBnrr$pid_bnr

vdataBnrr_temp <- vdataBnrr %>% dplyr::select(pid, tstart, logvl, vldl, sex)

vdataSurvv_temp <- vdataSurvv %>% dplyr::select(pid, vfailure, times)

vdataSurvv_temp %>% dplyr::left_join(vdataBnrr_temp, by = "pid") -> vdataBB
```

```
vdataSSurv <- vdataSurvv %>%
  dplyr::select(pid, vfailure, times, tstart, logvl, vldl, sex)

##############################################
# *** Fitting JM
##############################################

# *** Random Intercepts and Slopes LME
model.lme <- lme(
  logvl ~ tstart + sex,
  random = ~ tstart | pid,
  data = tdataBB,
  control=lmeControl(maxIter=1000, msMaxIter=1000, niterEM=1000, opt='optim'))
# *** Cox model
model.Cox <- coxph(Surv(times, vfailure == 1) ~ sex, data = tdataSSurv, model = TRUE)
# *** JM model with current value association structure
model.JM <- jm(
  model.Cox, model.lme, time_var = "tstart",
  control = list(n_chains = 3L, n_iter = 4000L, n_burnin = 2000L, n_thin = 1L, cores = 3L))

# AUC and Brier scores extracted from model.JM fit using "JMbayes2" R package
```

Appendix 11: R Code to Fit the Two-Part JM Model to a Subset Routine HIV Data

```
load("C:/Users/Documents/PhD/Appendix/Chap7_HIVRoutine.RData")

setwd("C:/Users/Documents/PhD/Appendix")

# Packages

library(statmod)

library(rstan)


options(mc.cores = parallel::detectCores())

rstan_options(auto_write = TRUE)


################################################

# *** Two-part joint model - Data preparation

################################################

set.seed(7341)


# *** Development set

# ** Semi continuous longitudinal data

X_bnr <- model.matrix(~ tstart + sex, data = tdataBnrr)

Z_bnr <- model.matrix(~ 1, data = tdataBnrr)

p_bnr <- ncol(X_bnr)

q_bnr <- ncol(Z_bnr)

X_cont <- model.matrix(~ tstart + sex, data = tdataContt)

Z_cont <- model.matrix(~ tstart, data = tdataContt)
```

```
p_cont <- ncol(X_cont)

q_cont <- ncol(Z_cont)


# * Dimensions

N_obs <- nrow(X_bnr)

n_ind <-  nlevels(as.factor(tdataBnrr$pid_bnr))

pid_bnr <- as.integer(tdataBnrr$pid_bnr)

Y_bnr <- tdataBnrr$vldl

N_cont <- nrow(X_cont)

n_ind_cont <- nlevels(as.factor(tdataContt$pid_cont))

pid_cont <- as.integer(tdataContt$pid_cont)

Y_cont <- tdataContt$ldl_vload


# * Random effects

zeros_uhurdle <- rep(0, 3)

q_hurdle <- 3


# ** Time to event data

time <- tdataSurvv$times

event <- tdataSurvv$vfailure

X_event <- model.matrix(~ -1 + sex, data = tdataSurvv)

p_event <- ncol(X_event)

X_bnr_eventtime <- model.matrix(~ times + sex, data = tdataSurvv)

Z_bnr_eventtime <- model.matrix(~ 1, data = tdataSurvv)
```

```
X_cont_eventtime <- model.matrix(~ times + sex, data = tdataSurvv)

Z_cont_eventtime <- model.matrix(~ times, data = tdataSurvv)


# * Gauss-Legendre quadrature

gau_quad <- gauss.quad(15, kind = "legendre")

quadnodes <- gau_quad$nodes # quadrature nodes

quadweight <- gau_quad$weights # quadrature weights

quads <- table(rep(1:nrow(tdataSurvv), each = 15))

X_bnr_quadnode <- array(1, dim = c(nrow(tdataSurvv), max(quads), 3))

X_bnr_quadnode[, , 2] <- tdataSurvv$times

X_bnr_quadnode[, , 3] <- tdataSurvv$sex

Z_bnr_quadnode <- array(1, dim = c(nrow(tdataSurvv), max(quads), 1))

X_cont_quadnode <- array(1, dim = c(nrow(tdataSurvv), max(quads), 3))

X_cont_quadnode[, , 2] <- tdataSurvv$times

X_cont_quadnode[, , 3] <- tdataSurvv$sex

Z_cont_quadnode <- array(1, dim = c(nrow(tdataSurvv), max(quads), 2))

Z_cont_quadnode[, , 2] <- tdataSurvv$times


# *** Validation set

n_ind_pred <-  nlevels(as.factor(vdataBnrr$pid_bnr))

time_pred <- vdataSurvv$times

event_pred <- vdataSurvv$vfailure

X_bnr_eventtime_pred <- model.matrix(~ times + sex, data = vdataSurvv)

Z_bnr_eventtime_pred <- model.matrix(~ 1, data = vdataSurvv)
```

```
X_cont_eventtime_pred <- model.matrix(~ times + sex, data = vdataSurvv)

Z_cont_eventtime_pred <- model.matrix(~ times, data = vdataSurvv)

X_event_pred <- model.matrix(~ -1 + sex, data = vdataSurvv)



stanData_hurdle <- list(

  X_bnr = X_bnr,

  Z_bnr = Z_bnr,

  p_cont = p_cont,

  q_cont = q_cont,

  X_cont = X_cont,

  Z_cont = Z_cont,

  p_bnr = p_bnr,

  q_bnr = q_bnr,

  N_obs = N_obs,

  n_ind = n_ind,

  pid_bnr = pid_bnr,

  Y_bnr = Y_bnr,

  N_cont = N_cont,

  n_ind_cont = n_ind_cont,

  pid_cont = pid_cont,

  Y_cont = Y_cont,

  zeros_uhurdle = zeros_uhurdle,

  q_hurdle = q_hurdle,
```

u_hurdle = t_hurdle,

time = time,

event = event,

X_event = X_event,

p_event = p_event,

X_bnr_eventtime = X_bnr_eventtime,

Z_bnr_eventtime = Z_bnr_eventtime,

X_cont_eventtime = X_cont_eventtime,

Z_cont_eventtime = Z_cont_eventtime,

quadnodes = quadnodes,

quadweight = quadweight,

X_bnr_quadnode = X_bnr_quadnode,

Z_bnr_quadnode = Z_bnr_quadnode,

X_cont_quadnode = X_cont_quadnode,

Z_cont_quadnode = Z_cont_quadnode,

X_bnr_eventtime_pred = X_bnr_eventtime_pred,

Z_bnr_eventtime_pred = Z_bnr_eventtime_pred,

X_cont_eventtime_pred = X_cont_eventtime_pred,

Z_cont_eventtime_pred = Z_cont_eventtime_pred,

n_ind_pred = n_ind_pred,

time_pred = time_pred,

event_pred = event_pred,

X_event_pred = X_event_pred)

```
TPJM_Routine <- stan(
  file = "TPJM_Stancode.stan",
  data = stanData_hurdle,
  chains = 3,
  iter = 4000,
  seed = 7341,
  verbose = FALSE,
  cores = getOption("mc.cores", 3L))

# Trace plots and other diagnostics plot extracted from TPJM_Routine fit
# AUC and Brier scores extracted from TPJM_Routine fit
# through riskRegression, "pec", and "survivalROC" R packages
# predictions from TPJM_Routine fit
```

Appendix 12: University of Cape Town Human Research Ethics Committee approval for this thesis

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

FACULTY OF HEALTH SCIENCES
Human Research Ethics Committee

FHS016: Annual Progress Report / Renewal

HREC office use only (FWA00001637; IRB00001938)			
This serves as notification of annual approval, including any documentation described below.			
☑ Approved	Annual progress report	Approved until/next renewal date	30.8.2024
☐ Not approved	See attached comments		
Signature Chairperson of the HREC/ Designee		Date Signed	19/8/23

Note: Please email this form and supporting documents (if applicable) in a combined pdf-file to hrec-enquiries@uct.ac.za.
Please clarify your plan for research-related activities during COVID-19 lockdown.
Please use the latest form found on our website:
http://www.health.uct.ac.za/fhs/research/humanethics/forms

HUMAN RESEARCH ETHICS COMMITTEE
17 AUG 2023
HEALTH SCIENCES FACULTY
UNIVERSITY OF CAPE TOWN

Comments to PI from the HREC

Principal Investigator to complete the following:

1. Protocol information

Date (when submitting this form)	August 17 2023		
HREC REF Number	436/2020	Current Ethics Approval was granted until	30/08/2023
Protocol title	Application of dynamic prediction models for longitudinal biomarkers and clinical outcomes in low-and middle-income settings		
Protocol number (if applicable)	Version 1.0		
Are there any sub-studies linked to this study?		☐ Yes	☑ No
If yes, could you please provide the HREC Reference number for all sub-studies? **Note:** A separate FHS016 must be submitted for each sub-study.			

(Note: Please complete the Closure form (FHS010) if the study is completed within the approval period)