

Joint modelling of longitudinal and time-to-event data: An illustration using CD4 count and mortality in a cohort of patients initiated on antiretroviral therapy

Nobuhle Nokubonga Mchunu (✉ nobuhlemchunu21@gmail.com)

University of KwaZulu-Natal <https://orcid.org/0000-0002-1832-2863>

Henry Mwambi

University of KwaZulu-Natal - Pietermaritzburg Campus

Tarylee Reddy

South African Medical Research Council

Nonhlanhla Yende-Zuma

Centre for the Aids Programme of Research in South Africa

Kogieleum Naidoo

Centre for the Aids Programme of Research in South Africa

Research article

Keywords: Time-to-event data, longitudinal data, joint models, CD4 count, mortality, bias

Posted Date: January 6th, 2020

DOI: <https://doi.org/10.21203/rs.2.19995/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Infectious Diseases on March 30th, 2020. See the published version at <https://doi.org/10.1186/s12879-020-04962-3>.

RESEARCH

Joint modelling of longitudinal and time-to-event data: An illustration using CD4 count and mortality in a cohort of patients initiated on antiretroviral therapy

Nobuhle N Mchunu^{1,2,3*}, Henry G Mwambi², Tarylee Reddy¹, Nonhlanhla Yende-Zuma^{3,4} and Kogieleum Naidoo^{3,4}

Abstract

Background: Modelling of longitudinal biomarkers and time-to-event data are important to monitor disease progression. However, these two variables are traditionally analyzed separately or time-varying Cox models are used. The former strategy fails to recognize the shared random-effects from the two processes while the latter assumes that longitudinal biomarkers are exogenous covariates, resulting in inefficient or biased estimates for the time-to-event model. Therefore, we used joint modelling for longitudinal and time-to-event data to assess the effect of longitudinal CD4 count on mortality.

Methods: We studied 4014 patients from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) who initiated ART between June 2004 and August 2013. We used proportional hazards regression model to assess the effect of baseline characteristics (excluding CD4 count) on mortality, and linear mixed effect models to evaluate the effect of baseline characteristics on the CD4 count evolution over time. Thereafter, the two analytical approaches were amalgamated to form an advanced joint model for studying the effect of longitudinal CD4 count on mortality. To illustrate the virtues of the joint model, the results from the joint model were compared to those from the time-varying Cox model.

Results: Using joint modelling, we found that lower CD4 count over time was associated with a 1.3-fold increase in the risk of death, (HR: 1.34, 95% CI: 1.27-1.42). Whereas, results from the time-varying Cox model showed lower CD4 count over time was associated with a 1.2-fold increase in the risk of death, (HR: 1.17, 95% CI: 1.12-1.23).

Conclusions: Joint modelling enabled the assessment of the effect of longitudinal CD4 count on mortality while correcting for shared random effects between longitudinal and time-to-event models. In the era of universal test and treat, the evaluation of CD4 count is still crucial for guiding the initiation and discontinuation of opportunistic infections prophylaxis and assessment of late presenting patients. CD4 count can also be used when immunological failure is suspected as we have shown that it is associated with mortality.

Keywords: Time-to-event data; longitudinal data; joint models; CD4 count; mortality; bias

Background

Until the era of test-and-treat, CD4 count was by far the most widely used biological marker for antiretroviral therapy (ART) eligibility and HIV (human immunodeficiency virus) progression [1]. However, the introduction of universal test and treat has put less

emphasis on the importance of CD4 count and only viral load is now used to monitor HIV disease progression and virologic failure. Arguably, the role of CD4 count in the current era of HIV monitoring is still crucial, particularly for patients presenting late to care as they are at high risk of presenting with opportunistic infections and also in areas where viral load testing is not affordable [2, 3].

*Correspondence: Nobuhle.mchunu@mrc.ac.za

¹South African Medical Research Council (SAMRC), SAMRC Building, 491 Peter Mokaba Ridge Road, Overport, Durban, 4041, South Africa
Full list of author information is available at the end of the article

Although CD4 count testing is no longer recommended for stable virologically suppressed patients in South Africa, however, the CD4 count still plays an important role in stratifying the risk of death among patients with low CD4 count who are failing first line ART regimen [4]. In South Africa, it was estimated that 7.9 million people of all ages were living with HIV in 2017 [5]. The country has the largest ART programme in the world with 4.7 million on lifelong ART treatment and these efforts have been largely financed from its own domestic resources [6]. Considering that South Africa is not a rich resourced country, managing such high volume of ART patients requires resources and different biomedical strategies using different longitudinal biomarkers at different stages of the disease.

In our setting, it has been shown that patients who started ART with low CD4 count have an increased risk of death [7]. Even though longitudinal CD4 count was not used in these findings, one can deduce that these patients' CD4 count recovery over time would have been slow and hence they died. Statistical models of the association between longitudinal CD4 count and mortality have been carried out using time-varying Cox models [8]. The counting process formulation of the time-varying Cox model has a distinct flexibility that not only allows for time-dependent covariates but also for left truncation, multiple time scales, multiple events per subject and various forms of case-cohort models, among others [9]. However, this model assumes that time-dependent covariates are (i) measured without error and (ii) are external or exogenous (that is, the value of the covariate at a future time point is not affected by the occurrence of the event). The second assumption is not valid for endogenous covariates (such as clinical biomarkers) [10], since the repeatedly measured marker like CD4 count is directly related to the mortality mechanism. Longitudinal and time-to-event outcomes are traditionally analyzed separately, however this approach leads to inefficient or biased results for the time-to-event model [11], due the fact that this type of modelling fails to recognize the shared random-effects from the two processes. Thus joint models are preferred over separate analyses and time-dependent models because they account for the special features of endogenous covariates and non-random dropout in a longitudinal data analysis context [10, 12].

Joint models of longitudinal and time-to-event data have received much attention in the literature dating back to the past two decades [13–15], however their application in the HIV context is limited, particularly when data from the sub-Saharan African region is used for modelling.

In view of the shortcomings of time-varying Cox models and separate analyses of longitudinal and time-to-event processes, our objective is to use a joint modelling strategy to assess the effect of longitudinal CD4 count on mortality among patients initiated on ART.

Methods

Source of data and description

In this analyses we use data from the Centre for the AIDS Programme of Research in South Africa (CAPRISA). The CAPRISA AIDS Treatment (CAT) programme enrolled HIV positive patients and initiated them on ART between June 2004 and August 2013. Eligibility criteria was in accordance with the Department of Health guidelines throughout. Males and females at least 14 years of age from urban (eThekweni) and rural (Vulindlela) sites were enrolled. Routine demographic and clinical data were recorded at baseline and at follow-up visits. Laboratory safety assessments and CD4 counts and viral loads were conducted at baseline and every 6 months or as clinically indicated. Patients were regarded as lost to follow-up if they missed 3 consecutive scheduled visits and if all attempts to track them telephonically and physically had failed. Information on the deaths was based on hospital chart notes, death certificates or oral reports from patient's relatives.

The eThekweni and Vulindlela sites initiated the first patient on ART in October 2004 and June 2004 respectively. Patients at the eThekweni site were recruited from the Prince Cyril Zulu Clinic of Communicable Disease which is the chest clinic adjacent to the CAPRISA clinic and sometimes patients presented themselves for HIV testing. Patients at the Vulindlela site were recruited from the Mafakatini clinic which is situated near that site or similarly presented themselves seeking health care.

Statistical analysis

In our setting it has been shown that gender is associated with both CD4 count and mortality [7]. Therefore, descriptive data, which was stratified by gender were presented as medians with interquartile range, percentages and graphical exploration was used where applicable. Unpaired t-test or the Wilcoxon rank sum test was used to compare continuous demographics and clinical data for men and women. Fisher's exact test was used for the comparison of categorical data. Poisson regression was used to calculate 95% confidence intervals (CI) for mortality rates and F-test was used for their comparisons.

Baseline predictors of mortality were assessed through both univariable and multivariable proportional hazards regression. The linear mixed effects models were used to assess the effect of baseline characteristics on the CD4 count evolution post ART initiation, where the individual patient and time post ART initiation were used as random effects. All multivariable models were adjusted for the following baseline covariates: gender (male or female), age (in years), clinic site (urban and rural), \log_{10} viral load and tuberculosis (TB) status.

Proportionality was assessed by the Schoenfeld proportional hazards test which provides proportional hazards test for individual covariates and a global test for the model with all variables combined. Variables that violated the proportional hazards assumption were not included in the proportional hazards model.

Thereafter the longitudinal and time-to-event models were amalgamated to form a joint model. To illustrate the virtues of the joint model we compared it to the time-varying Cox model. Analyses were conducted using SAS, version 9.4 (SAS Institute INC., Cary) and R version 3.5.1.

Results

Exploratory data analysis at ART initiation

There were 4014 patients enrolled whose ages range from 14-76 (with an overall mean age of 34.6 years). Out of the 4014, 2557 (63.7%) were females. TB prevalence was higher in men compared to women (32.1% vs. 19.7%). Moreover, women initiated ART with slightly higher CD4 count than men (132.0 vs. 113.0 cells/mm³, $p < 0.001$) (Table 1), and this pattern persisted over time (Figure 1) and matches the probability of death (Figure 2).

Results from the longitudinal sub-model: random effects multivariable model

Men started ART with low CD4 count ($\beta = -1.90$, S.E = 0.22, $p < 0.001$) when compared to women. However, their CD4 count evolution was not significantly different and thus the interaction term between gender and time was excluded in the model (Table S1). Patients presenting without TB at ART initiation started ART with higher mean CD4 count compared to those with prevalent TB but their rate of increase in CD4 count was slower when compared to those with prevalent TB ($\beta = -1.27$, S.E = 0.17, $p < 0.001$).

Mortality rates

There were a total of 414 deaths observed over 8195.87 person-years of follow-up. Mortality rates for men and women were 6.7 (95% CI: 5.8- 7.8) and 4.3 (95% CI: 3.8- 4.9) per 100 person-years (p-y), respectively; mortality rate ratio (MRR): 1.54 (1.27-1.88) $p < 0.001$. In addition, patients from the rural site had higher mortality rates compared to those from the urban site, 4.8 per 100 p-y, (95% CI: 4.3-5.4) vs. 5.6 per 100 p-y, (95% CI: 4.7-6.6); MRR: 1.16, (95% CI: 0.95-1.42), $p = 0.156$.

Results from the survival sub-model: modelling mortality

The multivariable proportional hazards regression analysis showed that men had a 62% significantly elevated risk of death when compared to women, HR: 1.62, (95% CI: 1.16-2.26), $p = 0.005$. In addition, patients who started ART with a higher baseline viral load had a significantly higher risk of death HR: 1.57, (95% CI: 1.26-1.96), $p = 0.004$ (Table S2).

Results from the joint model of longitudinal and time-to-event data

The joint model finds a significantly strong association between the CD4 count and the risk of death, with a unit decrease in the marker corresponding to a 1.3-fold increase in the risk of death (HR: 1.34, 95% CI: 1.27-1.42). These results are statistically significant indicating that indeed CD4 count is a good predictor of mortality and in fact confirms that an increase in CD4 counts is associated with better survival (Table 3). These results were compared with those from the time-varying Cox model and we also observe a strong association between the longitudinal CD4 count and the risk of death. In particular a unit decrease in the CD4 count corresponds to a 1.2-fold increase in the risk of death (HR: 1.17, 95% CI: 1.12-1.23) (Table 3). Previous research and simulation studies have shown that the time-varying Cox model underestimates the true association size of markers [10].

Discussion

Results from the longitudinal sub-model (random effects multivariable model), showed no statistical difference between the urban and rural sites in terms of the CD4 count improvement over time, with patients from the urban site having a higher rate of change. This finding reaffirms the results obtained by [7,16,17]. Men and older people on average initiated ART with significantly lower CD4 counts. These results support those obtained by [18,19]. Patients presenting without TB at ART initiation started ART with high mean CD4 count compared to those with prevalent TB but

Table 1: Baseline characteristics of patients initiated on ART

Characteristic	Women (N= 2557)	Men (N= 1457)	p-value
Age (years), median (IQR) ^a	32.0 (28.0-39.0)	35.0 (30.0-41.0)	< 0.001
Site, n (%):			< 0.001
Rural	1507 (58.9)	692 (47.5)	
Urban	1050 (41.1)	765 (52.5)	
Prevalent TB , n (%):			< 0.001
No	2052 (80.3)	990 (67.9)	
Yes	505 (19.7)	467 (32.1)	
Body mass index (kg/m ²), median (IQR) ^b	24.2 (21.0-28.1)	21.0 (19.0-23.2)	0.415
CD4 count (cells/ μ L), median (IQR) ^c	132.0 (69.0-202.0)	113.0 (47.0-177.0)	< 0.001
CD8 count (cells/ μ L), median (IQR) ^d	818.5 (533.5-1197.5)	736.0 (462.0-1123.0)	< 0.001
Viral load (log ₁₀ copies/ml), mean (SD) ^e :	4.9(0.9)	5.0 (0.9)	< 0.001
CD4:CD8 ratio, median(IQR) ^f	0.2 (0.1-0.2)	0.1 (0.1-0.2)	< 0.001

^a4 patients had missing age, ^b237 patients had missing BMI, ^c382 patients had missing CD4 count, ^d1936 patients had missing baseline CD8 count, ^e488 patients had missing baseline viral load, ^f1929 patients had missing CD4:CD8 ratio

Table 2: Longitudinal process estimates from the joint model

	β estimate ^a	S.E.	p-value
Intercept	18.26	0.73	< 0.001
Age (years)	-0.02	0.01	0.052
Men (ref: women)	-1.86	0.22	< 0.001
Urban site (ref: rural)	-0.27	0.25	0.2735
Prevalent TB (ref: No prevalent TB)	-0.57	0.29	0.047
Log ₁₀ viral load (copies/ml)	-0.61	0.11	< 0.001
Time on ART (years)	1.89	0.37	< 0.001
Time \times Prevalent TB (ref: no prevalent TB)	1.26	0.16	< 0.001
Time \times log ₁₀ viral load	0.19	0.07	0.007

^aadjusted estimates; S.E.: standard error

Table 3: Event process estimates from the joint model and time-varying Cox model estimates

	β estimate	aHR (95% CI)	S.E.	p-value
Time-varying proportional hazards model				
Longitudinal CD4 count (per unit decrease)	0.16	1.17 (1.12-1.23)	0.03	< 0.001
Event Process				
Age (years)	0.01	1.01 (0.99-1.03)	0.01	0.148
Log ₁₀ viral load (copies/ml)	0.37	1.45 (1.21-1.69)	0.12	0.002
Men (ref: women)	0.05	1.05 (0.70-1.40)	0.18	0.775
Urban site (ref: rural)	-0.08	0.92 (0.49-1.35)	0.22	0.758
Prevalent TB (ref: no prevalent TB)	-0.21	0.81 (0.34-1.28)	0.24	0.377
Association (per unit decrease)	0.30	1.34 (1.27-1.42)	0.03	< 0.001

aHR: adjusted hazard ratios; S.E.: standard error

their rate of change in CD4 count was significantly less compared to those with prevalent TB. These results are similar to those found by [7] in the same study.

Results from the survival sub-model (Cox proportional hazards regression analysis) showed that pa-

tients from the urban site had a higher survival prognosis compared to those from the rural site, however this was not significant at 5% level of significance. Patients presenting without TB at ART initiation had an elevated risk of dying compared to those with prevalent TB. These results reaffirms the results obtained by

[16]. Prevalent TB was also previously shown to be associated with low mortality, maybe related to TB care being an access point to earlier ART initiation [7, 16]. Published literature has cited that undiagnosed TB is higher among patients accessing ART than in the general population; with the majority of incident TB diagnosed in the early weeks of ART initiation being TB prevalent but missed at baseline screening [20]. In addition male patients, older people or those with a higher mean baseline viral load had a significantly elevated risk of death (refer to Table S2). This finding is in consonance with previous research which showed that men and older patients were at an increased risk of mortality due to HIV/AIDS [16, 18, 21].

The joint model was advantageous for answering multivariate questions at the same time (in our case CD4 count and mortality). The most appealing feature of joint models is its ability to capture or take into consideration the association between the survival time and repeated measurement of a risk factor variable [10]. The joint model showed a significantly strong association between CD4 count and the risk for death, implying that CD4 count is a good predictor of mortality. The joint model also helped assess the correlation between the two response variables and gave ample opportunity to see predictors of the two response variables simultaneously. The results in this study indicated that CD4 count change due to ART and mortality had been influenced jointly by some of the covariates like gender, age, baseline viral load, time (in years) and by the interaction effects of time (in years) with TB status and baseline viral load (refer to Table 2 and Table 3). Research findings from a longitudinal study by [12] also showed that CD4 count change was affected by these covariates.

The joint model for longitudinal and time-to-event data has several advantages especially in clinical trials. In a survival analysis setting, where the covariate of interest is time-dependent, either the entire history of the covariate for every subject, or, minimally, measurements of the covariate at each time of disease occurrence for all subjects in the corresponding risk set, are necessary. This extensive measurement of covariate is rarely, if ever, executed and the values obtained are typically subject to measurement error. Thus by modelling the covariates over time, we can enhance the survival analysis since we can interpolate covariate values between the observed measurements to the specific times of disease occurrence, with the use of the entire covariate history of the subjects. Furthermore, according to [22], after accounting for measurement error, the standard error of the relative risk estimate will reflect

correctly the uncertainty in the measurements of the covariate. Conversely, utilizing the survival data in the longitudinal model will yield improved longitudinal parameter estimates by allowing adjustment for informative right censoring of the repeated measurements by the disease process. Furthermore, the joint model allows for unequally spaced measurements, or missing covariate data and censoring of survival times. The fact that the joint model has the distinct advantage of simultaneously modelling two response variables (for example in this study, CD4 count and time-to-death) allows the researcher some degree of flexibility.

We found that after ART initiation the CD4 count increases and is influenced by measured covariates such as age, gender TB status and baseline viral load. Furthermore, gender and baseline viral load were found to be significant predictors of all-cause mortality. The joint model found a strong association between CD4 count measurement process and mortality which means that the full CD4 count history is a predictor of mortality. These results are in consonant with previous research [10, 23, 24].

Conclusion

Joint modelling enabled the assessment of the effect of longitudinal CD4 count on mortality while correcting for shared random effects between longitudinal and time-to-event models. In biomedical research where measurements of various outcomes are taken over a time period in an attempt to understand patients' health or the risk of an event occurring, the joint modelling approach will be the most useful tool to consider in an effort to link the longitudinal measurement process and time-to-event outcomes. In the era of universal test and treat, the evaluation of CD4 count is still crucial for guiding the initiation and discontinuation of opportunistic infections prophylaxis and assessment of late presenting patients. The CD4 count can also be used when immunological failure is suspected as we have shown that it is associated with mortality.

Abbreviations

aHR: adjusted hazard ratio; ART: antiretroviral therapy; CAPRISA: Centre for the AIDS programme of research in South Africa; CAT: CAPRISA AIDS treatment; CI: confidence intervals; HIV: human immunodeficiency virus; HR: hazard ratio; MRR: mortality rate ratio; S.E.: standard error; TB: Tuberculosis.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

All the authors made substantial intellectual contributions to the study.

Availability of data and materials

Data will be made available upon request but access will be controlled and each request will be considered on a case by case basis.

Acknowledgements

The work from which this study emanated was funded by the South African Medical Research Council under the Biostatistics Capacity Development funded in partnership with Building Academic Partnerships For Economics Development (BAPED): The content of any Publications from any studies during this work are solely the responsibility of the authors and do not necessarily represent the official views of the South African Medical Research Council.

This work was also supported by the national research fund (NRF) and by the DELTAS Africa Initiative Grant No. 107754/Z/15/Z-DELTAS Africa SSACAB. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust (Grant No. 107754/Z/15/Z) and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Funding

This study was supported by the US President's Emergency Plan for AIDS Relief (PEPFAR), the Global Fund to fight AIDS, Tuberculosis and Malaria and the National Institutes of Health Comprehensive International Program of Research on AIDS.

Ethics approval and consent to participate

Ethics approval for use of the treatment program data was obtained from the Biomedical Research Ethics Committee of the University of KwaZulu-Natal. (Ref-E248/05). The biomedical research ethics committee reviewed the application and deemed the need for consent unnecessary in terms of national regulations: "the REC may approve a waiver of consent for secondary use of material or data where no more than minimal risk of harm is likely; and donor rights and welfare interests are unlikely to adversely affected; and research cannot be conducted if waiver is not approved". The national health Act permitted children 12 years and older to access HIV care without parental or guardian consent. Hence no parental or guardian consent was required for this age group [25]. (National Health Act 61 of 2003). Informed consent was obtained for HIV testing and treatment (including those less than 16 years) from all individuals included in the study in keeping with standard of care.

Consent for publication

Not applicable.

Author details

¹South African Medical Research Council (SAMRC), SAMRC Building, 491 Peter Mokaba Ridge Road, Overport, Durban, 4041, South Africa. ² University of KwaZulu-Natal, School of Mathematics, Statistics and Computer Science, King Edward Avenue, Pietermaritzburg, 3209, South Africa. ³Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, 719 Umbilo Road, Durban, 4041, South Africa. ⁴MRC-CAPRISA HIV-TB Pathogenesis and Treatment Research Unit, Doris Duke Medical Research Institute, University of KwaZulu-Natal, Durban, South Africa.

References

- Karim, S.S.A., Karim, Q.A.: HIV/AIDS in South Africa, Cambridge University Press (2010)
- Ford, N., Meintjes, G., Pozniak, A., Bygrave, H., Hill, A., Peter, T., Davies, M.-A., Grinsztajn, B., Calmy, A., Kumarasamy, N., *et al.*: The future role of CD4 cell count for monitoring antiretroviral therapy. *The Lancet Infectious Diseases* **15**(2), 241–247 (2015)
- Rice, B., Boulle, A., Schwarcz, S., Shroufi, A., Rutherford, G., Hargreaves, J.: The Continuing Value of CD4 Cell Count Monitoring for Differential HIV Care and Surveillance. *JMIR public health and surveillance* **5**(1), 11136 (2019)
- Bell-Gorrod, H., Fox, M.P., Boulle, A., Prozesky, H., Wood, R., Tanser, F., Davies, M.-A., Schomaker, M.: The impact of delayed switch to second-line antiretroviral therapy on mortality, depending on failure time definition and CD4 count at failure. *bioRxiv*, 661629 (2019)
- UNAIDS: Regional Maps, Treatment Cascade 90-90-90: People living with HIV who have suppressed viral loads (2017)
- on HIV/AIDS (UNAIDS), J.U.N.P., *et al.*: Ending AIDS: Progress towards the 90-90-90 targets. *Global AIDS update* (2017)
- Naidoo, K., Hassan-Moosa, R., Yende-Zuma, N., Govender, D., Padayatchi, N., Dawood, H., Adams, R.N., Govender, A., Chinappa, T., Abdool-Karim, S., *et al.*: High mortality rates in men initiated on anti-retroviral treatment in KwaZulu-Natal, South Africa. *PLoS one* **12**(9), 0184124 (2017)
- Ren, L., Li, J., Zhou, S., Xia, X., Xie, Z., Liu, P., Xu, Y., Qian, Y., Zhang, H., Ma, L., *et al.*: Prognosis of HIV patients receiving antiretroviral therapy according to CD4 counts: a long-term follow-up study in Yunnan, China. *Scientific reports* **7**(1), 9595 (2017)
- Therneau, T.M., Grambsch, P.M.: *Modeling Survival Data Extending the Cox Model*. Springer (2000)
- Rizopoulos, D.: *Joint Models for Longitudinal and Time-to-event Data: With Applications in R*, Chapman and Hall/CRC (2012)
- Asar, Ö., Ritchie, J., Kalra, P.A., Diggle, P.J.: Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *International journal of epidemiology* **44**(1), 334–344 (2015)
- Guo, X., Carlin, B.P.: Separate and joint modelling of longitudinal and event time data using standard computer packages. *The American Statistician* **58**(1), 1 (2004)
- De Gruttola, V., Tu, X.M.: Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, 1003–1014 (1994)
- Wulfsohn, M.S., Tsiatis, A.A.: A joint model for survival and longitudinal data measured with error. *Biometrics*, 330–339 (1997)
- Henderson, R., Diggle, P., Dobson, A.: Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**(4), 465–480 (2000)
- Dawood, H., Hassan-Moosa, R., Zuma, N.-Y., Naidoo, K.: Mortality and treatment response amongst HIV-infected patients 50 years and older accessing antiretroviral services in South Africa. *BMC infectious diseases* **18**(1), 168 (2018)
- Yende, N.: *Modelling CD4+ count over time in HIV positive patients initiated on HAART in South Africa using linear mixed models*. PhD thesis (2010)
- Maskew, M., Brennan, A.T., Westreich, D., McNamara, L., MacPhail, A.P., Fox, M.P.: Gender differences in mortality and CD4 count response among virally suppressed HIV-positive patients. *Journal of women's health* **22**(2), 113–120 (2013)
- Cornell, M., Schomaker, M., Garone, D.B., Giddy, J., Hoffmann, C.J., Lessells, R., Maskew, M., Prozesky, H., Wood, R., Johnson, L.F., *et al.*: Gender differences in survival among adult patients starting antiretroviral therapy in South Africa: a multicentre cohort study. *PLoS medicine* **9**(9), 1001304 (2012)
- Etard, J.-F., Ndiaye, I., Thierry-Mieg, M., Guèye, N.F.N., Gueye, P.M., Laniece, I., Dieng, A.B., Diouf, A., Laurent, C., Mboup, S., *et al.*: Mortality and causes of death in adults receiving highly active antiretroviral therapy in Senegal: a 7-year cohort study. *Aids* **20**(8), 1181–1189 (2006)
- Prins, M., Robertson, J.R., Brettle, R.P., Aguado, I.H., Broers, B., Boufassa, F., Goldberg, D.J., Zangerle, R., Coutinho, R.A., van den Hoek, A.: Do gender differences in CD4 cell counts matter? *Aids* **13**(17), 2361–2364 (1999)
- Faucett, C.L., Thomas, D.C.: Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in medicine* **15**(15), 1663–1685 (1996)
- Wang, Y., Taylor, J.M.G.: Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association* **96**(455), 895–905 (2001)
- Lim, H.J., Mondal, P., Skinner, S.: Joint modeling of longitudinal and event time data: application to hiv study. *Journal of Medical Statistics and Informatics* **1**(1), 1 (2013)
- Parliament, S.A.: National health act no. 61 of 2003. Pretoria: Government Printers (2003)

Additional Files

Table S1: Baseline predictors of CD4 count evolution using linear mixed effects multivariable model

Effect	β estimate ^a	S.E.	p-value
Intercept	17.82	0.82	< 0.001
Age (years)	-0.02	0.01	0.047
Men (ref: women)	-1.90	0.22	< 0.001
Urban site (ref: rural)	-0.33	0.25	0.189
No prevalent TB (ref: prevalent TB)	0.71	0.31	0.022
Log ₁₀ viral load	-0.65	0.12	< 0.001
Time on ART (years)	2.67	0.39	< 0.001
Time × no prevalent TB (ref: prevalent TB)	-1.27	0.17	< 0.001
Time × log ₁₀ viral load	0.24	0.07	0.005

^aadjusted estimates; S.E.: standard error

Table S2: Baseline predictors of mortality using Cox proportional hazards model

Effect	HR (95% CI)	p-value	aHR (95% CI)	p-value
Age (years)	1.01 (0.99-1.03)	0.191	1.01 (0.99-1.03)	0.237
Rural site	1.0		1.0	
Urban site	0.97 (0.66-1.43)	0.882	1.02 (0.68-1.55)	0.919
Women	1.0		1.0	
Men	1.69 (1.21-2.35)	0.002	1.62 (1.16-2.26)	0.005
No prevalent TB	1.0		1.0	
Prevalent TB	1.13 (0.75-1.71)	0.573	1.31 (0.84-2.05)	0.238
Log ₁₀ viral load (copies/ml)	1.57 (1.26-1.97)	< 0.001	1.57 (1.26-1.96)	0.004

aHR: adjusted hazard ratios

Figures

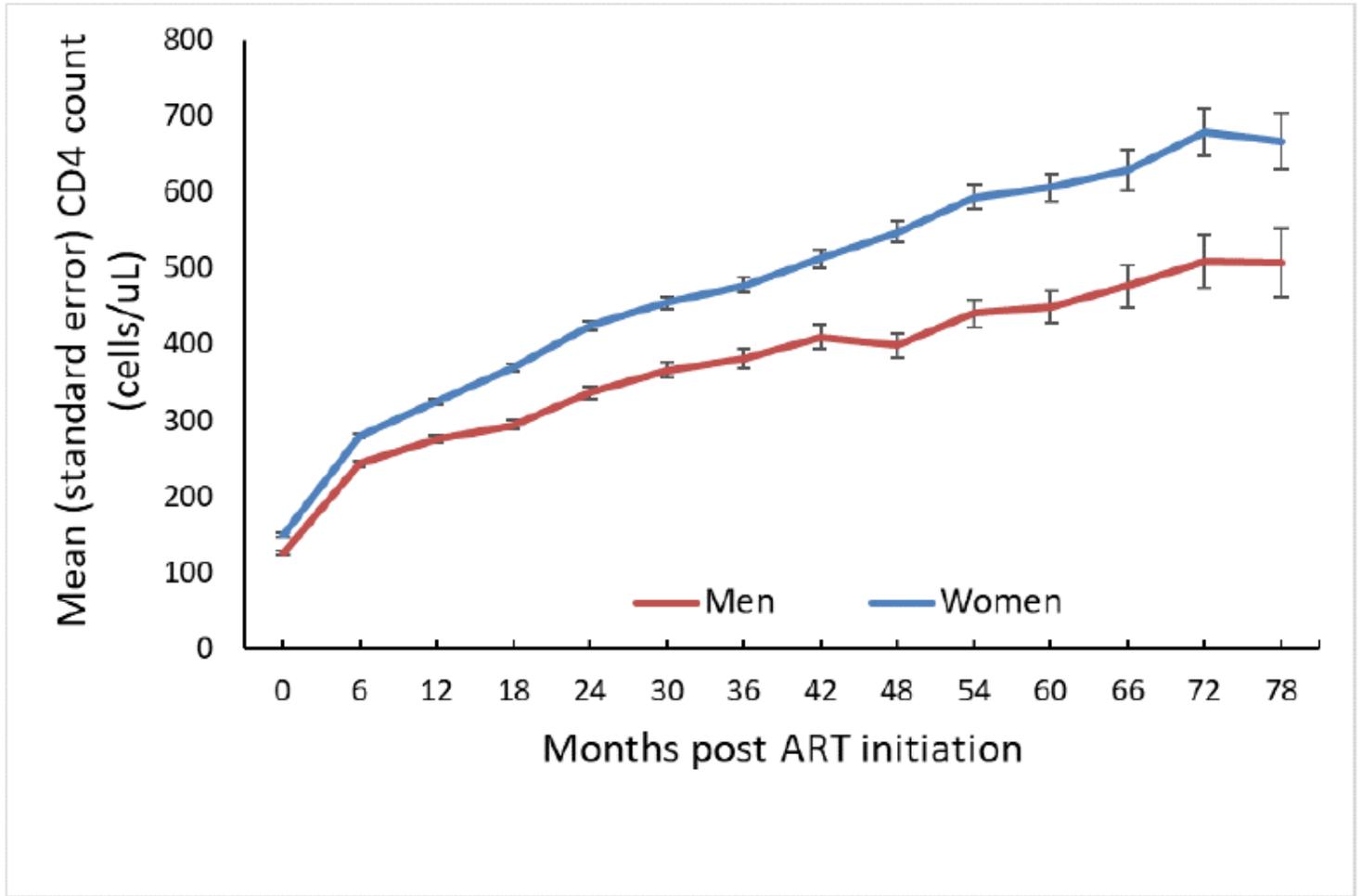
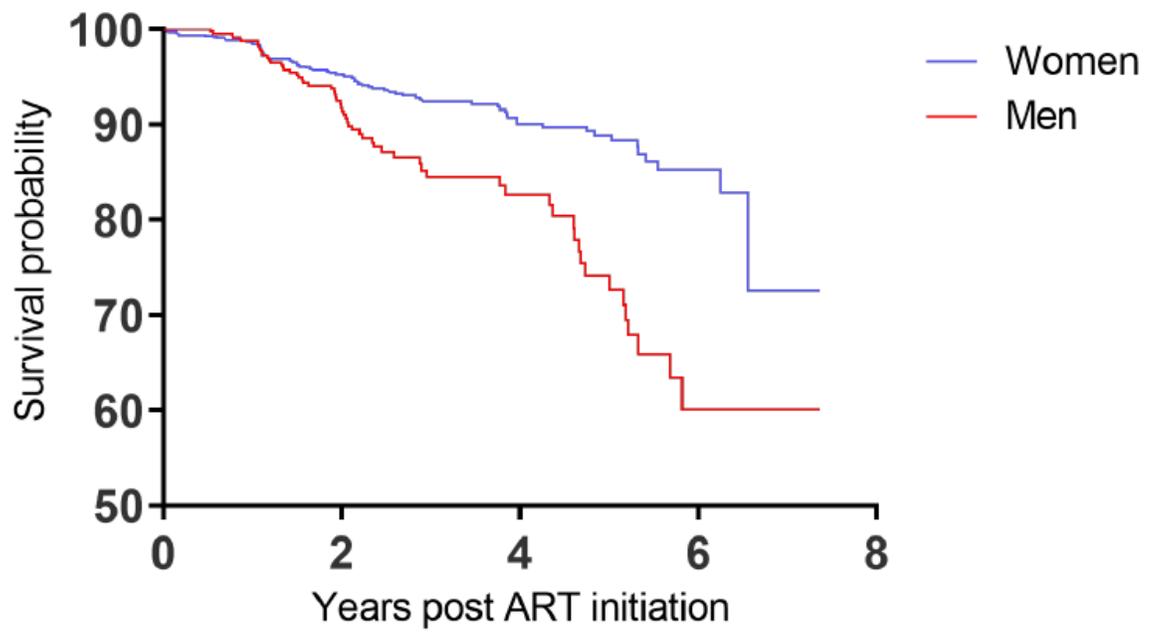


Figure 1

Mean CD4 count (cells/ μ L) over time by gender



Number at risk					
Men	555	283	77	13	0
Women	1091	625	211	40	0

Figure 2

Kaplan-Meier curve for survival