Appendix 1: Analyses accounting for competing risks

Independence assumption

Standard techniques for analysing survival data include Kaplan-Meier estimation of cumulative incidence and Cox regression models. These techniques assume that the distribution of censoring times and the time-to-event distribution are independent of each other. Often, this assumption is taken to be valid without further checks. If patients are censored administratively, then this assumption may be reasonable. However, if patients are censored through becoming lost to follow up or through experiencing another event, then censoring may be related to the time to the event of interest and the independence assumption is violated. This leads to biased estimates of survival times and overestimates of percentages experiencing the event of interest.

Competing events

In our data, this assumption is violated. For example, our event of interest is starting ART. If a patient dies before starting ART, then standard techniques would result in this patient being censored. However, this patient cannot then experience the event of interest and is not representative of those remaining in follow-up. Therefore, censoring is not appropriate. In this case, the event pre-ART death precludes the event of interest from happening and is a called a competing risk.

Appendix Figure 1: Kaplan-Meier estimates of the cumulative probabilities of starting ART and death

As seen above, if competing events are present, the Kaplan-Meier curves will overestimate the percentage experiencing the event of interest within each CD4 strata. In the group with CD4 counts <25 cells/mm^3, the Kaplan-Meier estimates that 87% initiate ARVs and 91% die first. The Kaplan-Meier method also estimates that over all eligible patients and all follow up time, 92% will start ARVs but that also, that 100% will die first. This is of course impossible and methods are needed to correct for this.

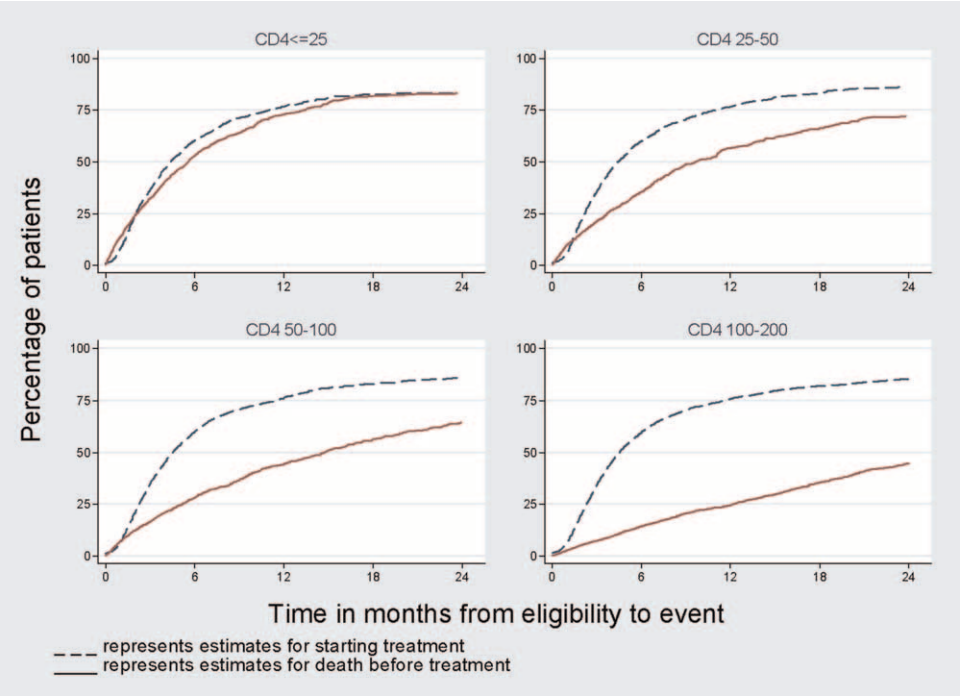
Cumulative Incidence Functions

Cumulative Incidence Functions (CIFs) provide an unbiased method of reporting the percentage of patients experiencing an event of interest in the presence of competing events. The CIF, $I_k(t)$, is defined as the probability of failing from cause k by time t. It is different from the Kaplan-Meier estimate as it does not simply treat failures from causes other than k as censored. The CIF estimates the probability of failing from cause k, in the presence of all other causes.

$$\hat{I}_k(t) = \sum_{t_i \le t} \hat{p}_k(t_i)$$

$$\hat{p}_k(t_i) = \hat{\lambda}_k(t_i)\hat{S}(t_{i-1}), \quad \hat{\lambda}_k(t_i) = \frac{d_{ki}}{n_i}$$

where $S(t_i)$ represents the estimate of the overall survivor function at time t_i



In the main text of the paper we calculate CIFs for both events within CD4 strata. We report that for patients with CD4 counts <25 cells/mm^3, 51% start ART and 48% die first. These estimates now make sense.

Risk Factor Associations

Cox Regression

When assessing associations of risk factors with event times, the standard technique is to use Cox regression modeling. Again, when using this technique, it is assumed that patients who are censored can be thought to be representative of the patients who remain in follow up. As this is not true in our data, Cox regression was not appropriate. However, we present the results from Cox regression as a comparison (Appendix Table 1).

With Cox regression we saw no effect of CD4 on the hazard of starting treatment, except in the group with CD4 counts <25 cells/mm^3. Patients with CD4 counts <25 cells/mm^3 appeared to have an increased hazard of starting treatment compared to those with CD4 counts of 100–200 cells/mm^3. However, these patients also have a much greater risk of mortality. In the Cox regression, patients who die before starting ART are censored and the patients who remain in follow up are assumed to be representative of the censored patients. This leads to the hazard of starting treatment being overestimated for this group, as patients who are not able to experience the event are treated as though they could experience the event.

To properly account for the risk of dying before treatment in our estimates of associations between

Appendix Table 1. Adjusted associations of facility and patient level characteristics with time to a) starting treatment and b) pre-ART death, using 25 imputed datasets and standard Cox regression*, not taking into account competing risks.

	N (%)	HR(95% CI) for starting ART§	HR(95% CI) for pre-ART death§
Patient characteristics			
Sex			
Female	14261 (64.6)	1 (baseline)	1 (baseline)
Male	7822 (35.4)	0.85 (0.82,0.88)	1.22 (1.15,1.30)
Age at eligibility (year)			
15–29	4928 (22.3)	0.95 (0.91,1.00)	0.89 (0.83,0.96)
30–39	9474 (42.9)	1 (baseline)	1 (baseline)
40–49	5654 (25.6)	1.01 (0.97,1.06)	1.02 (0.95,1.10)
>=50	2027 (9.2)	1.04 (0.98,1.11)	1.27 (1.15,1.40)
Weight at eligibility (kg)			
<40	1609 (7.3)	0.75 (0.68,0.83)	1.94 (1.73,2.18)
40–49	5554 (25.2)	0.92 (0.87,0.97)	1.36 (1.26,1.46)
50–59	7620 (34.5)	1 (baseline)	1 (baseline)
60–79	6400 (29.0)	1.03 (0.98,1.08)	0.74 (0.68,0.81)
>=80	900 (4.1)	1.17 (1.07,1.29)	0.48 (0.37,0.62)
CD4 value at eligibility date (cells/mm^3)			
<=25	3207 (14.5)	1.09 (1.03,1.15)	3.69 (3.42,3.98)
25–50	2698 (12.2)	1.04 (0.99,1.11)	2.39 (2.19,2.60)
50–100	5102 (23.1)	1.03 (0.99,1.08)	1.85 (1.72,2.00)
100–200	11076 (50.2)	1 (baseline)	1 (baseline)
Year of enrolment			
2004	3997 (18.1)	1 (baseline)	1 (baseline)
2005	5340 (24.2)	1.16 (1.10,1.23)	1.10 (1.01,1.19)
2006	6192 (28.0)	1.37 (1.30,1.45)	1.10 (1.02,1.20)
2007	6554 (29.7)	1.70 (1.61,1.79)	1.03 (0.94,1.13)
Facility characteristics			
Filled posts per 1000 enrolled patients per year			
<5	6805 (30.8)	0.62 (0.59,0.65)	0.98 (0.91,1.06)
5–7.5	8116 (36.8)	1 (baseline)	1 (baseline)
>7.5	7162 (32.4)	1.01 (0.96,1.07)	1.16 (1.06,1.27)
Location			
Urban/peri-urban	16970 (76.8)	1 (baseline)	1 (baseline)
Rural	5113 (23.2)	0.92 (0.86,0.99)	1.29 (1.12,1.49)
Distance to treatment site (km)			
Same site	5698 (25.8)	1 (baseline)	1 (baseline)
<8	6923 (31.3)	0.94 (0.88,0.99)	2.02 (1.80,2.26)
8–15	6563 (29.7)	0.84 (0.79,0.90)	2.11 (1.88,2.36)
>15	2899 (13.1)	0.81 (0.75,0.87)	1.52 (1.33,1.73)

*Model for starting ART: pre-ART deaths are censored. Model for pre-ART deaths: patients starting ART are censored.
§Mutually adjusted for all characteristics in the table.

Appendix 2: Adjusted sub hazard ratios (SHR) of the associations of facility and patient level characteristics with time to a) starting treatment and b) pre-ART death, using complete case data N=16,071

	N (%)	SHR(95% CI) for starting ART[§]	SHR(95% CI) for pre-ART death[§]
Patient characteristics			
Sex			
Female	10461 (65.1)	1 (baseline)	1 (baseline)
Male	5610 (34.9)	0.82 (0.78 to 0.85)	1.39 (1.30 to 1.49)
Age at eligibility (year)			
15–29	3533 (22.0)	0.99 (0.94 to 1.05)	0.92 (0.85 to 1.01)
30–39	6895 (42.9)	1 (baseline)	1 (baseline)
40–49	4188 (26.1)	0.98 (0.94 to 1.03)	0.99 (0.91 to 1.07)
>=50	1455 (9.1)	0.95 (0.89 to 1.03)	1.20 (1.07 to 1.34)
Weight at eligibility (kg)			
<40	968 (6.0)	0.58 (0.52 to 0.65)	2.41 (2.14 to 2.71)
40–49	4229 (26.3)	0.81 (0.77 to 0.86)	1.45 (1.34 to 1.57)
50–59	5757 (35.8)	1 (baseline)	1 (baseline)
60–79	4375 (27.2)	1.16 (1.10 to 1.21)	0.72 (0.65 to 0.79)
>=80	742 (4.6)	1.26 (1.15 to 1.39)	0.46 (0.35 to 0.60)
CD4 value at eligibility date (cells/mm^3)			
<=25	2263 (14.1)	0.67 (0.62 to 0.72)	2.98 (2.73 to 3.26)
25–50	1916 (11.9)	0.79 (0.73 to 0.84)	2.07 (1.87 to 2.29)
50–100	3635 (22.6)	0.86 (0.81 to 0.90)	1.75 (1.61 to 1.90)
100–200	8257 (51.4)	1 (baseline)	1 (baseline)
Year of enrolment			
2004	3687 (22.9)	1 (baseline)	1 (baseline)
2005	4574 (28.5)	1.16 (1.10 to 1.23)	0.94 (0.87 to 1.03)
2006	4506 (28.0)	1.41 (1.33 to 1.49)	0.80 (0.73 to 0.87)
2007	3304 (20.6)	2.05 (1.92 to 2.18)	0.55 (0.49 to 0.61)
Facility characteristics			
Filled posts per 1000 enrolled patients per year			
<5	4992 (31.1)	0.65 (0.62 to 0.69)	1.34 (1.23 to 1.45)
5–7.5	6233 (38.8)	1 (baseline)	1 (baseline)
>7.5	4846 (30.2)	1.12 (1.04 to 1.21)	1.02 (0.90 to 1.14)
Location			
Urban/peri-urban	12010 (74.7)	1 (baseline)	1 (baseline)
Rural	4061 (25.3)	0.74 (0.67 to 0.80)	1.50 (1.27 to 1.78)
Distance to treatment site (km)			
Same site	3703 (23.0)	1 (baseline)	1 (baseline)
<8	4911 (30.6)	0.80 (0.74 to 0.87)	1.76 (1.53 to 2.02)
8–15	5061 (31.5)	0.72 (0.66 to 0.77)	1.97 (1.71 to 2.27)
>15	2396 (14.9)	0.75 (0.68 to 0.82)	1.54 (1.32 to 1.79)

[§]Mutually adjusted for all characteristics in the table.

risk factors and hazard of starting treatment, we used competing risks regression models as defined by Fine and Gray [14] and implemented in Stata v.11 [18]. These model estimate the hazard of the subdistribution (the hazard function as would be derived from the CIF), which appropriately accounts for competing risks.

By using competing risks regression models, we see that there is an association between CD4 and the hazard of starting treatment. Patients with CD4 counts <25 cells/mm^3 have a lower hazard of starting treatment compared to those with CD4 counts of 100−200 cells/mm^3, and this is due to the fact that they are more immunosuppressed and therefore, more likely to die before initiating ART.