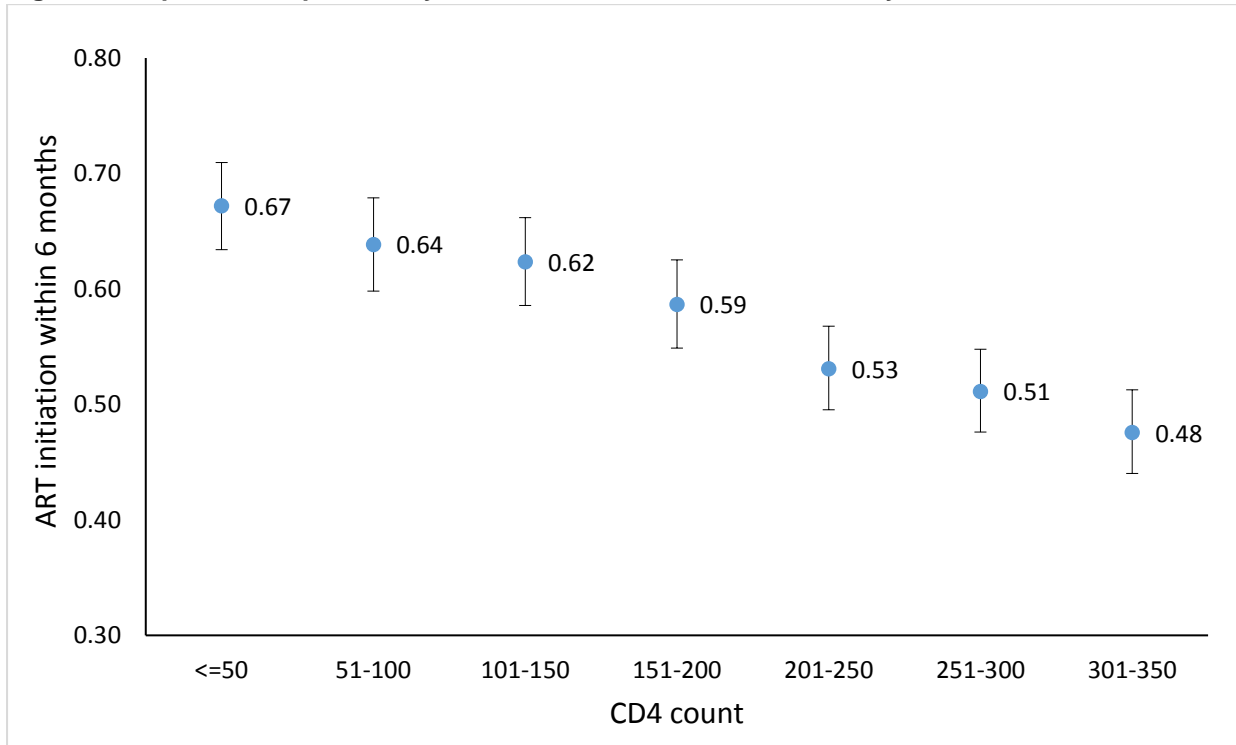


**Table 1: Descriptive statistics for all adults who received their first CD4 count between August 2011 and December 2012**

Variable	Full Sample	Population surveillance sub-sample	Population surveillance sub-sample, (200<CD4≤350)
<b>N</b>	4,630	434	193
<b>Total analysis time at risk, person years</b>	3,564	170	102
<b>N (%)</b>			
<b>Initiated ART</b>			
Within 6 months	2,626 (57%)	300 (69%)	113 (59%)
Within 12 months	2,730 (59%)	312 (72%)	121 (63%)
Within the entire study period	2,786 (60%)	320 (74%)	127 (66%)
<b>Censored</b>			
Died	47 (1%)	15 (3%)	1 (1%)
Moved out from residence at baseline	-	46 (11%)	29 (15%)
Reached end of study period	1,797 (39%)	53 (12%)	36 (19%)
<b>Earliest CD4 cell count</b>			
≤50	624 (13%)	63 (15%)	-
51-100	553 (12%)	58 (13%)	-
101-150	619 (13%)	56 (13%)	-
151-200	644 (14%)	64 (15%)	-
201-250	731 (16%)	56 (13%)	56 (29%)
251-300	741 (16%)	70 (16%)	70 (36%)
301-350	718 (16%)	67 (15%)	67 (35%)
<b>Sex is Male</b>	1,813 (39%)	157 (36%)	48 (25%)
<b>Lives more than 2 km away from the nearest clinic</b>		242 (56%)	103 (53%)
<b>Lives in a rural area</b>		206 (47%)	91 (47%)
<b>Has children under the age of 6</b>		312 (72%)	133 (69%)
<b>Has other household members linked to care</b>		226 (52%)	112 (58%)
<b>Employed or earns income</b>		163 (38%)	54 (28%)
<b>Median (IQR)</b>			
<b>Days taken to initiate ART #</b>	74 (26, .)	40 (22, .)	65 (24, .)
<b>Age at earliest CD4 count</b>	32 (27, 40)	32 (26, 41)	30 (25, 37)
<b>Years of education</b>		14 (12, 15)	14 (13, 15)
<b>Wealth index</b>		4 (1, 6)	4 (0, 7)

# Those who were censored were given an infinitely large value for days taken to initiate ART

**Figure 1: Kaplan-Meier probability of ART initiation within 6 months by earliest CD4 count**



Note: Kaplan-Meier curves were constructed for time to ART initiation, separately by 50-cell CD4 count bin. Each data point presents the probability of having initiated ART within six months from date of first CD4 count. Asymptotic 95% confidence intervals are reported. Sex-stratified estimates and numerical values are presented in Appendix Table 1 and Appendix Figure 1.

**Table 2: CD4 count and time to ART initiation in Hlabisa Cohort (n=4,630)**

Variable	Cox Proportional Hazards Models: Hazard Ratios (95% CI)			
	Crude	Adjusted for age, sex, and age*sex		Adjusted for age, sex, age*sex, year#, clinic#
<b>CD4 cell count</b>				
Per 100 cells	0.83*** (0.80, 0.86)	0.85*** (0.81, 0.88)	0.85*** (0.82, 0.88)	
51-100 v. ≤50	0.86* (0.74, 0.98)	0.87* (0.75, 0.99)	0.85* (0.73, 0.97)	
101-150 v. ≤50	0.84* (0.73, 0.96)	0.86* (0.75, 0.98)	0.84* (0.74, 0.97)	
151-200 v. ≤50	0.77*** (0.68, 0.89)	0.80** (0.69, 0.91)	0.79*** (0.69, 0.91)	
201-250 v. ≤50	0.67*** (0.59, 0.77)	0.70*** (0.61, 0.80)	0.69*** (0.61, 0.80)	
251-300 v. ≤50	0.62*** (0.54, 0.71)	0.64*** (0.56, 0.74)	0.65*** (0.57, 0.75)	
301-350 v. ≤50	0.56*** (0.49, 0.65)	0.59*** (0.52, 0.68)	0.60*** (0.52, 0.69)	
<b>Age in years</b>				
25-34 v. 18-24		1.16* (1.02, 1.32)	1.16* (1.02, 1.32)	1.22** (1.07, 1.38)
35-54 v. 18-24		1.30*** (1.13, 1.49)	1.30*** (1.13, 1.49)	1.37*** (1.19, 1.57)
55+ v. 18-24		1.51** (1.16, 1.96)	1.51** (1.16, 1.96)	1.65*** (1.27, 2.15)
<b>Male sex</b>				
		1.12 (0.88, 1.42)	1.12 (0.88, 1.42)	1.09 (0.86, 1.38)
<b>Male sex X Age in years</b>				
Male X 25-34		0.94 (0.72, 1.22)	0.90 (0.69, 1.18)	0.91 (0.70, 1.19)
Male X 35-54		0.91 (0.69, 1.19)	0.90 (0.69, 1.17)	0.90 (0.69, 1.18)
Male X 55+		0.66* (0.43, 0.99)	0.67 (0.44, 1.01)	0.67 (0.44, 1.02)

#Model includes indicator variables for year of earliest CD4 cell count and earliest registration clinic.

Coefficients are suppressed. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**Table 3: Socio-demographic predictors of ART initiation: evidence from demographic surveillance**

Variable	Cox Proportional Hazards Models: Hazard Ratios (95% CI)					
	Surveillance residents, all n=434			Surveillance residents, 200<CD4≤350 n=193		
	Crude	#Age-Sex- CD4	##All Predictors	Crude	#Age-Sex- CD4	##All Predictors
CD4 count (per 100 cells)	0.75*** (0.68, 0.84)	0.81*** (0.72, 0.91)	0.79*** (0.70, 0.89)	0.92 (0.61, 1.40)	1.09 (0.71, 1.67)	1.08 (0.69, 1.68)
>2 km from nearest clinic	0.84 (0.67, 1.05)	0.80* (0.64, 1.00)	0.91 (0.70, 1.18)	0.78 (0.55, 1.10)	0.74 (0.52, 1.06)	0.79 (0.53, 1.19)
Lives in a rural area	0.80* (0.64, 0.99)	0.77* (0.62, 0.96)	0.82 (0.63, 1.07)	0.83 (0.59, 1.18)	0.80 (0.56, 1.14)	0.83 (0.54, 1.25)
Has children under 6 years	0.90 (0.71, 1.15)	0.74* (0.56, 0.97)	0.78 (0.58, 1.04)	0.96 (0.66, 1.40)	0.71 (0.45, 1.12)	0.73 (0.45, 1.19)
Other house- hold members are in HIV care	0.85 (0.68, 1.06)	0.88 (0.70, 1.10)	0.89 (0.71, 1.12)	1.02 (0.72, 1.45)	1.08 (0.75, 1.56)	1.19 (0.81, 1.74)
Employed or earns income	1.22 (0.97, 1.54)	1.06 (0.82, 1.37)	1.02 (0.78, 1.32)	1.23 (0.84, 1.82)	0.88 (0.55, 1.40)	0.83 (0.51, 1.34)
Empl. missing	0.89 (0.59, 1.36)	1.10 (0.72, 1.69)	1.01 (0.53, 1.90)	0.87 (0.48, 1.57)	0.90 (0.50, 1.63)	0.68 (0.26, 1.76)
Education in years	0.96** (0.93, 0.99)	0.99 (0.96, 1.03)	0.99 (0.96, 1.03)	0.96 (0.92, 1.01)	1.00 (0.94, 1.06)	1.00 (0.94, 1.07)
Educ. missing	0.96 (0.54, 1.70)	1.25 (0.69, 2.25)	1.28 (0.56, 2.94)	1.28 (0.63, 2.63)	1.33 (0.64, 2.74)	1.97 (0.61, 6.41)
Wealth index	1.03 (0.99, 1.07)	1.03 (0.99, 1.07)	1.02 (0.97, 1.07)	1.02 (0.97, 1.08)	1.03 (0.97, 1.08)	1.00 (0.95, 1.07)
Wealth missing	0.90 (0.54, 1.49)	0.95 (0.57, 1.58)	0.83 (0.47, 1.48)	0.79 (0.37, 1.68)	0.75 (0.34, 1.63)	0.67 (0.29, 1.54)

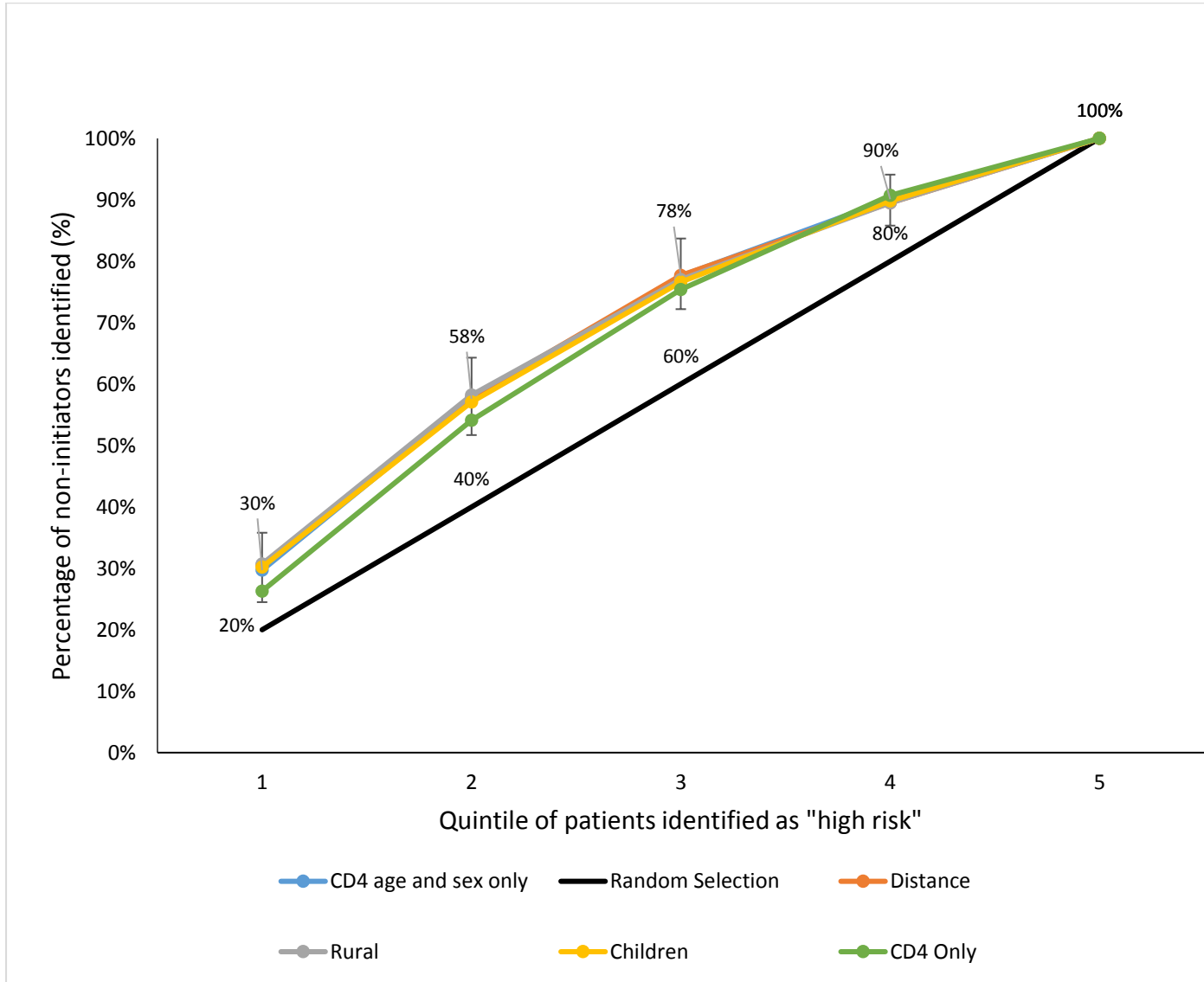
#Model adjusted for CD4 cell count, age, sex, and interaction terms between age and sex

##Coefficients are from a single model adjusted for all covariates, age, sex, and age-sex interactions

Note: Coefficients on age, sex and age-sex interaction terms are not reported. Likelihood ratio tests comparing the age/sex/CD4 model to expanded models with other predictors are shown in Appendix

Table 2. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**Figure 2: Percentage of true non-initiators identified by each predictive model at different probability thresholds**

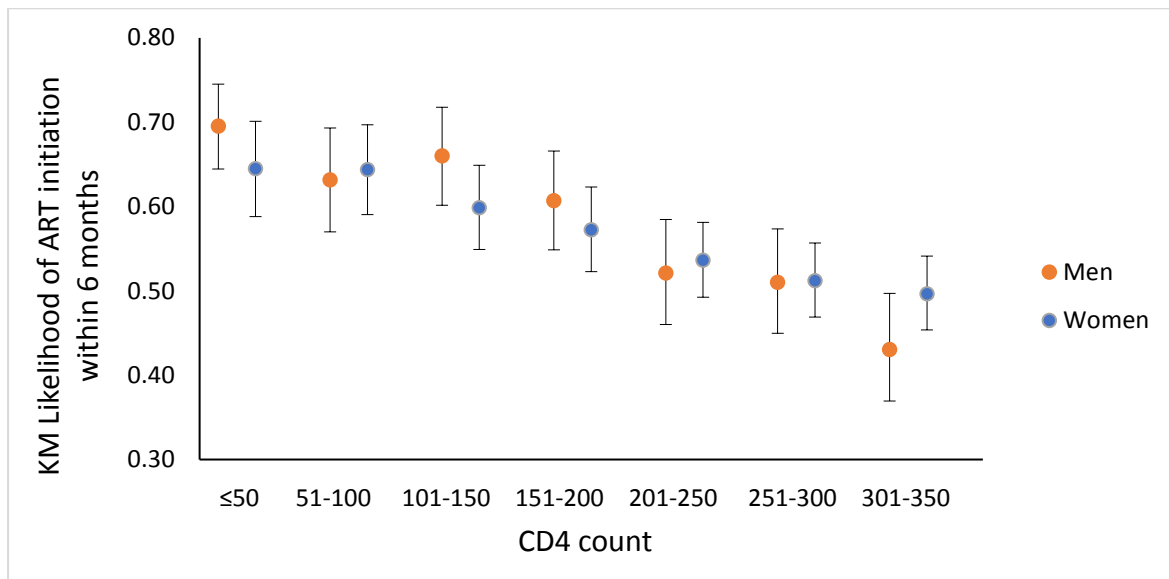


Note: Values are reported in Appendix Table 4.

**Appendix Table 1: Kaplan-Meier likelihood of ART initiation within 6 months by earliest CD4 count**

First recorded CD4 count	Kaplan-Meier likelihood of ART initiation within 6 months (95% CI)		
	Full sample	Men	Women
≤50	0.67 (0.63, 0.71)	0.70 (0.64, 0.75)	0.64 (0.59, 0.70)
51-100	0.64 (0.60, 0.68)	0.63 (0.57, 0.69)	0.64 (0.59, 0.70)
101-150	0.62 (0.59, 0.66)	0.66 (0.60, 0.72)	0.60 (0.55, 0.65)
151-200	0.59 (0.55, 0.63)	0.61 (0.55, 0.67)	0.57 (0.52, 0.62)
201-250	0.53 (0.50, 0.57)	0.52 (0.46, 0.58)	0.54 (0.49, 0.58)
251-300	0.51 (0.48, 0.55)	0.51 (0.45, 0.57)	0.51 (0.47, 0.56)
301-350	0.48 (0.44, 0.51)	0.43 (0.37, 0.50)	0.50 (0.45, 0.54)

**Appendix Figure 1: Kaplan-Meier likelihood of ART initiation within 6 months by first recorded CD4 count and sex**



**Appendix Table 2: Likelihood ratio tests for all adjusted models: incremental benefit relative to a model adjusting only for CD4, age, sex, and age-sex interactions.**

Probability that model differs from a nested model of CD4 count, age, sex, and age-by-sex (p-value from likelihood ratio test of nested models)		
Predictor(s)	All ACDIS residents n=434	Healthier sub-sample (200<CD4≤350), n=193
<i>Age, Sex, Age-Sex, CD4, plus...</i>		
Distance to nearest to nearest clinic	0.05*	0.10
Rural	0.02*	0.21
Has children under the age of 6	0.03*	0.15
Other household members linked to care	0.26	0.68
Employment	0.84	0.83
Years of education	0.71	0.77
Wealth index	0.42	0.50
All predictors <sup>#</sup>	0.25	0.57

All models adjust for age, sex, CD4 count, and age-by-sex interactions. Data are p-values from likelihood ratio tests of the hypothesis that the model differs from the nested model of just age, sex, CD4 cell count, and interaction terms between age and sex. <sup>#</sup>Model includes all predictors simultaneously. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001

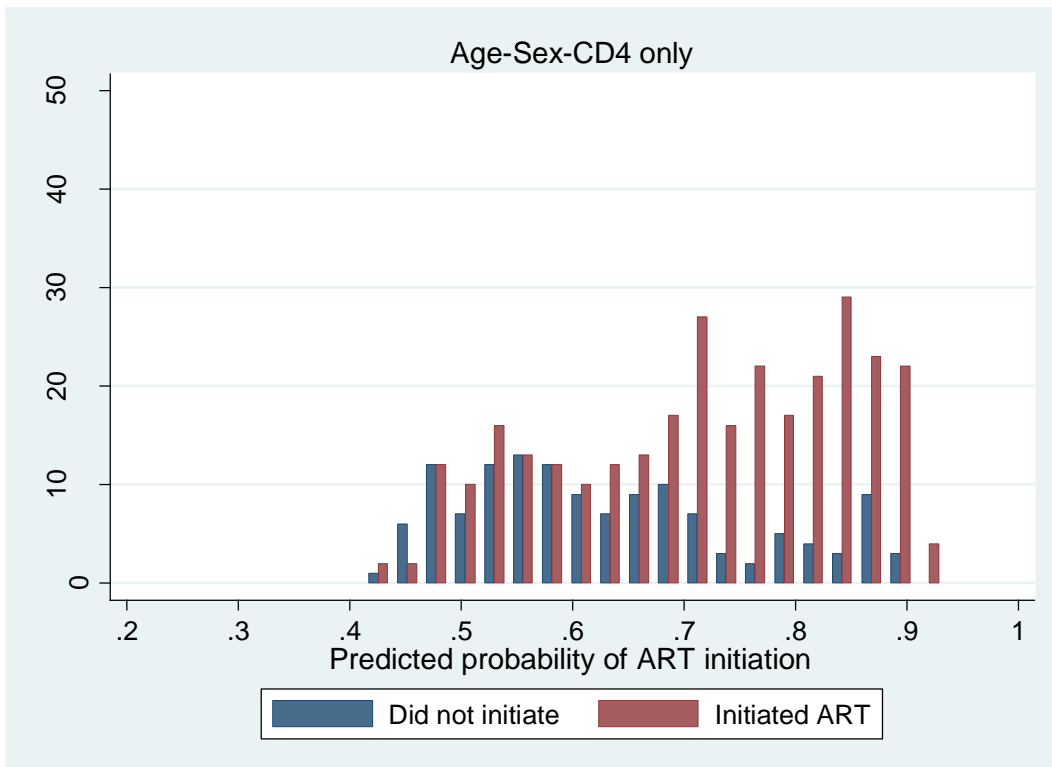
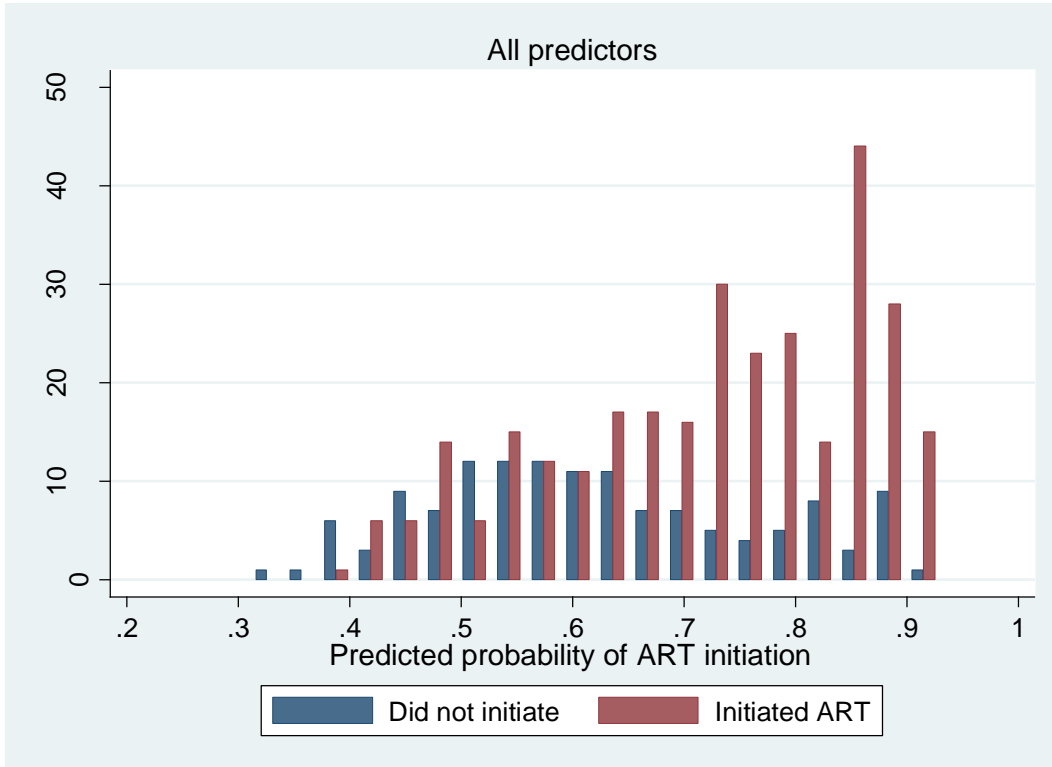


**Appendix Table 3: Logistic models for ART initiation among ACDIS residents (all CD4 counts)**

Variable	Odds ratios of ART initiation (95% CI)		
	Crude	#Age-Sex-CD4- adjusted	##All Predictors, fully adjusted
Earliest CD4 count (per 100 cells)	0.63*** (0.51, 0.78)	0.69*** (0.55, 0.86)	0.66*** (0.52, 0.84)
>2km from nearest clinic	0.91 (0.60, 1.37)	0.86 (0.56, 1.33)	1.19 (0.70, 2.01)
Rural	0.76 (0.50, 1.14)	0.72 (0.47, 1.11)	0.76 (0.44, 1.30)
Has children under 6 years	1.02 (0.65, 1.60)	0.69 (0.40, 1.18)	0.63 (0.35, 1.14)
Household members in HIV care	0.83 (0.55, 1.25)	0.93 (0.60, 1.43)	0.95 (0.60, 1.51)
Employed or earns income	1.27 (0.82, 1.97)	1.02 (0.62, 1.67)	1.02 (0.61, 1.68)
Empl. status missing	0.93 (0.45, 1.92)	1.11 (0.52, 2.37)	0.85 (0.28, 2.58)
Education in years	0.93* (0.87, 0.99)	0.98 (0.91, 1.06)	0.98 (0.91, 1.06)
Education missing	1.33 (0.42, 4.22)	1.99 (0.61, 6.48)	2.34 (0.51, 10.78)
Wealth index	1.02 (0.94, 1.10)	1.02 (0.94, 1.10)	1.01 (0.92, 1.10)
Wealth missing	0.78 (0.34, 1.82)	0.71 (0.29, 1.73)	0.64 (0.22, 1.84)

#Coefficients are from separate logistic regression models, adjusted for CD4, age, sex, and interaction terms between age and sex; age-sex coefficients are suppressed. ##Coefficients come from a single logistic regression model adjusted for all covariates and interaction terms between age and sex; age-sex coefficients are suppressed. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**Appendix Figure 2: Predicted probabilities of ART initiation by actual initiation status**



**Appendix Table 4: Percentage of true non-initiators identified by each predictive model at different probability thresholds**

For each predictive model, we used repeated sampling cross-validation techniques to estimate 95% confidence intervals for the percentage of true non-initiators identified. We randomly divided the dataset into two halves, designating one half as the training dataset and the other as the test dataset. Using the training dataset only, we estimated a logistic model for the probability of ART initiation within 6 months, the estimated regression coefficients providing a formula for a risk score. Applying these coefficients to the test dataset, we computed predicted probabilities of initiation. We then identified patients as “high risk” of non-initiation based on different risk thresholds, using quintiles of the probability distribution as thresholds. For each threshold, we then compared the group of patients identified as “high risk” with the group of patients who actually failed to initiate ART within six months. We calculated the percent of true non-initiators identified under each threshold. We repeated this process 1,000 times and present the mean and 95% confidence intervals of this performance indicator below in Table 4. The results can be interpreted as follows: with the threshold set at 40%, 40% of patients would be identified as high risk for non-initiation. If age, sex, and CD4 count were used to identify those at risk, then an intervention targeting this group would reach 58% (95% CI 52, 64) of true non-initiators.

Quintiles targeted, threshold percentile	Percentage of true non-initiators identified (95% CI)					
	CD4 Only	Age, sex, age*sex	Age, sex, age*sex and CD4	CD4, age, sex, and age*sex plus... Distance	Rural	Children
Q1, 20%	26 (21, 32)	27 (22, 33)	30 (24, 36)	31 (25, 36)	31 (25, 37)	30 (24, 36)
Q1-Q2, 40%	54 (48, 60)	53 (47, 60)	58 (52, 64)	58 (51, 64)	58 (52, 65)	57 (51, 64)
Q1-Q3, 60%	75 (70, 81)	76 (70, 82)	78 (72, 84)	78 (72, 83)	77 (72, 83)	77 (71, 82)
Q1-Q4, 80%	91 (87, 94)	91 (86, 95)	90 (86, 94)	90 (85, 94)	90 (86, 94)	90 (86, 94)

**Appendix Table 5: Proportion of patients that initiated ART at the same clinic where they received their first CD4 cell count**

<b>Variable</b>	<b>Full Sample</b>	<b>Population surveillance sub-sample</b>	<b>Population surveillance sub-sample, (200&lt;CD4≤350)</b>
<b>N</b>	4,630	434	193
<b>Initiated ART</b>			
Within the entire study period	2,786 (60%)	320 (74%)	127 (66%)
- At the same clinic where they received their first CD4 cell count	2,670 (96%)	315 (98%)	127 (100%)

**Appendix Table 6: Socio-demographic predictors of ART initiation: evidence from demographic surveillance (Competing risks model)**

Variable	Competing Risk Models: Hazard Ratios (95% CI)					
	Surveillance residents, all n=434			Surveillance residents, 200<CD4≤350 n=193		
	Crude	#Age-Sex- CD4	##All Predictors	Crude	#Age-Sex- CD4	##All Predictors
CD4 count (per 100 cells)	0.79*** (0.71, 0.88)	0.85*** (0.76, 0.95)	0.83** (0.73, 0.93)	0.87 (0.57, 1.33)	1.03 (0.66, 1.59)	1.02 (0.64, 1.62)
>2 km from nearest clinic	0.85 (0.68, 1.05)	0.82 (0.65, 1.02)	0.92 (0.71, 1.20)	0.75 (0.53, 1.06)	0.71 (0.50, 1.00)	0.76 (0.51, 1.14)
Lives in a rural area	0.80* (0.64, 0.99)	0.77* (0.62, 0.96)	0.83 (0.64, 1.07)	0.80 (0.56, 1.12)	0.76 (0.54, 1.08)	0.82 (0.56, 1.20)
Has children under 6 years	0.89 (0.69, 1.14)	0.71* (0.54, 0.94)	0.74* (0.55, 1.00)	0.94 (0.66, 1.36)	0.69 (0.45, 1.06)	0.74 (0.47, 1.16)
Other house- hold members are in HIV care	0.89 (0.72, 1.11)	0.93 (0.75, 1.17)	0.94 (0.75, 1.18)	0.97 (0.68, 1.36)	1.01 (0.70, 1.45)	1.11 (0.76, 1.62)
Employed or earns income	1.20 (0.96, 1.50)	1.08 (0.84, 1.39)	1.03 (0.79, 1.33)	1.24 (0.86, 1.80)	0.90 (0.56, 1.45)	0.85 (0.52, 1.40)
Empl. missing	0.90 (0.58, 1.39)	1.03 (0.64, 1.63)	0.96 (0.48, 1.92)	0.92 (0.52, 1.63)	0.97 (0.55, 1.70)	0.73 (0.29, 1.85)
Education in years	0.96*** (0.93, 0.98)	0.99 (0.96, 1.02)	0.99 (0.95, 1.02)	0.96 (0.91, 1.00)	1.00 (0.94, 1.05)	1.00 (0.94, 1.06)
Educ. missing	1.00 (0.58, 1.72)	1.26 (0.72, 2.21)	1.32 (0.55, 3.15)	1.37 (0.73, 2.58)	1.42 (0.76, 2.65)	2.02 (0.67, 6.04)
Wealth index	1.03 (0.99, 1.08)	1.03 (0.99, 1.08)	1.02 (0.98, 1.07)	1.03 (0.97, 1.10)	1.04 (0.97, 1.12)	1.02 (0.94, 1.10)
Wealth missing	0.87 (0.50, 1.51)	0.86 (0.49, 1.50)	0.79 (0.42, 1.46)	0.77 (0.35, 1.73)	0.73 (0.32, 1.63)	0.64 (0.26, 1.59)

#Model adjusted for CD4 cell count, age, sex, and interaction terms between age and sex

##Coefficients are from a single model adjusted for all covariates, age, sex, and age-sex interactions

Note: Coefficients on age, sex and age-sex interaction terms are not reported. Likelihood ratio tests comparing the age/sex/CD4 model to expanded models with other predictors are shown in Appendix Table 2. \* p<0.05, \*\* p<0.01, \*\*\* p<0.001