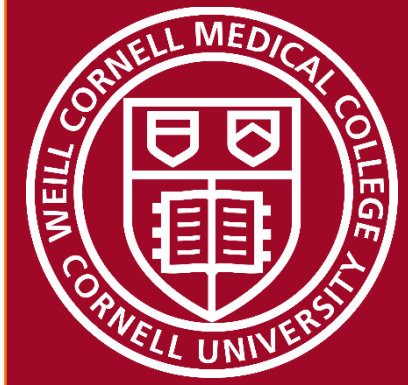


Using Multiple Imputation for Loss to Follow Up: Cohort of HIV-Positive Patients in Haiti

Deanna Jannat-Khah, DrPH, MSPH
Michelle Unterbrink

Margaret McNairy, Daniel Fitzgerald, Samuel Pierre, Jean W. Pape, Arthur Evans

2015 UK STATA Users Conference

Background

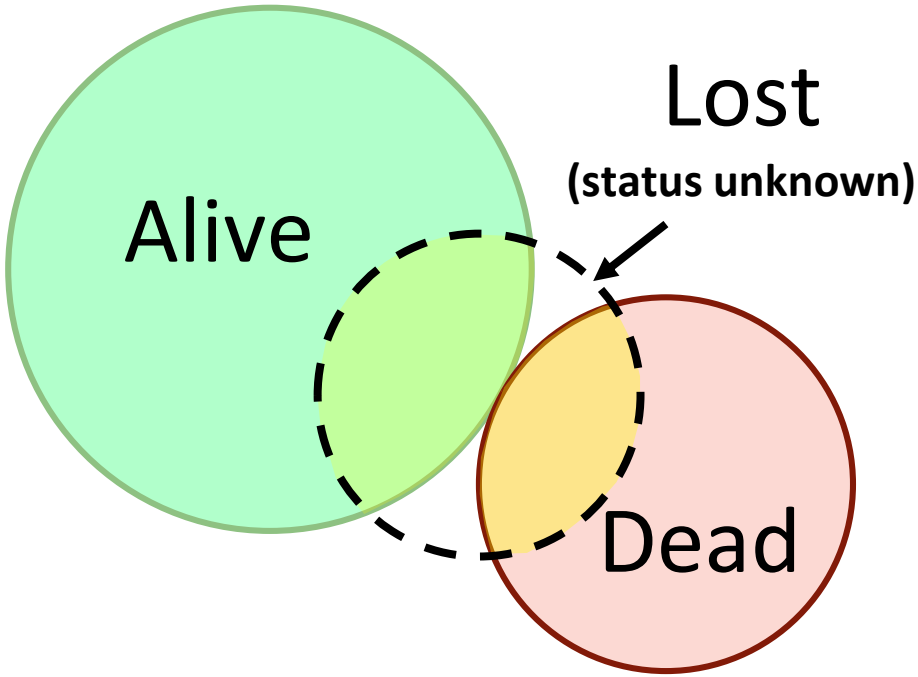
- In many public health program evaluations, cohorts of patients are followed for months-years
- A proportion of patients can not be found
- This patients are categorized as lost to follow-up (LTF)
- The challenge:
  - These patients do not have an outcome status (i.e., dead vs. alive)
  - How do programs estimate their outcomes status?

Objectives of talk

1. Presenting and comparing methods to estimate outcome status of patients who are LTF
2. Demonstrating the application of Multiple Imputation for estimating outcome status of LTF

LTF is not an outcome:
a mixture of outcome statuses

1. Undocumented death
2. Alive and in care somewhere else
3. Alive and not engaged in care



Methods used to estimate LTF outcome status

1. Survival analysis (Kaplan Meier methods)
2. Tracing with Inverse Probability Weights (IPW)
3. Multiple Imputation with Chained Equations (MICE)

Study overview

Study purpose: Estimate 10 year survival among the first cohort of HIV patients receiving treatment in Haiti

Study site: Haitian Group for the Study of Kaposi's Sarcoma and Opportunistic Infections (GHESKIO clinic)

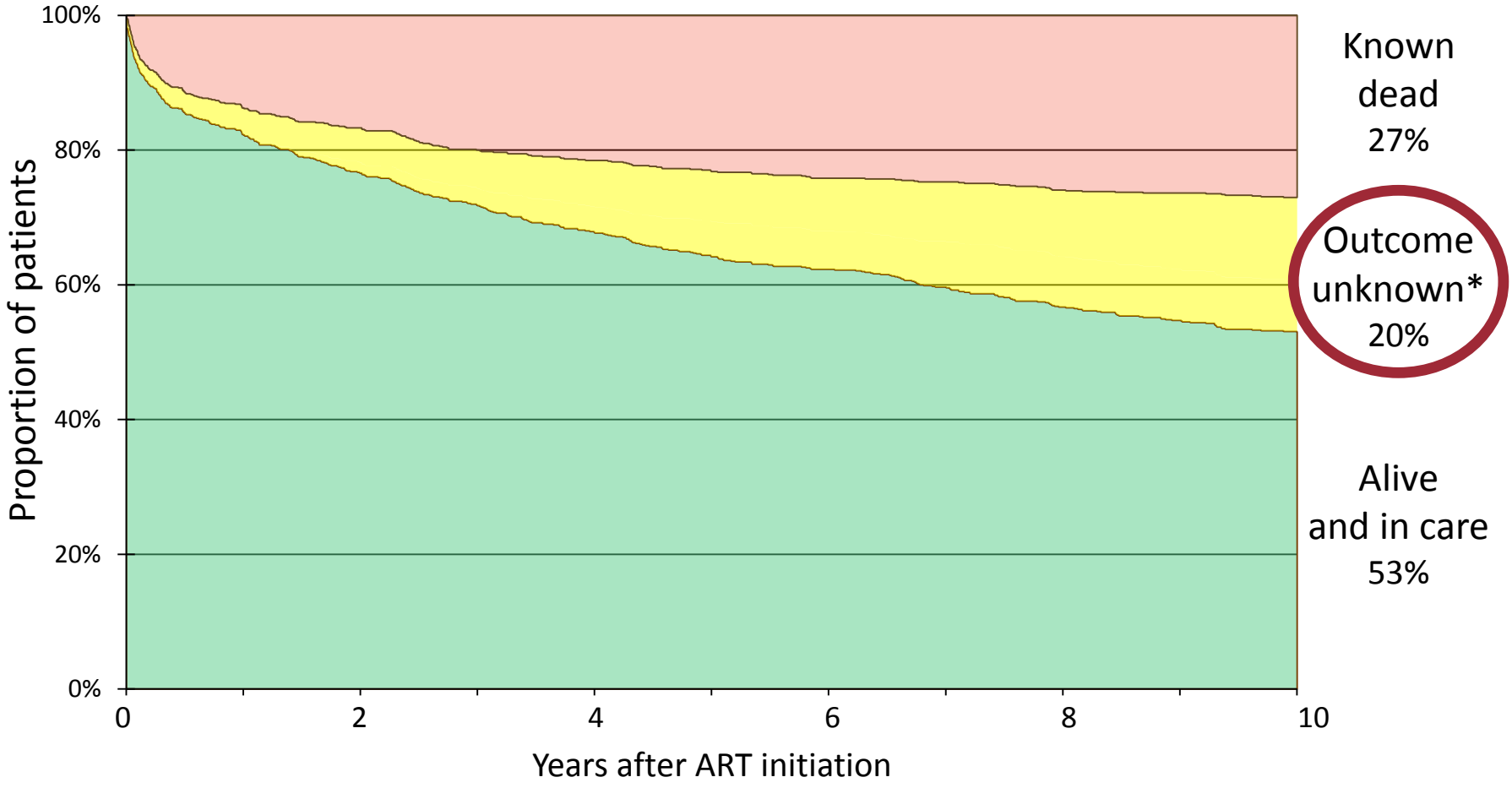
Study Population: 910 adults aged ≥ 13 years enrolled in HIV care in 2003

Study follow up period: 10 years

Primary Outcome: Survival status at 10 years

Secondary Outcome: Predictors of survival

Study outcomes at 10 years: 20% LTF



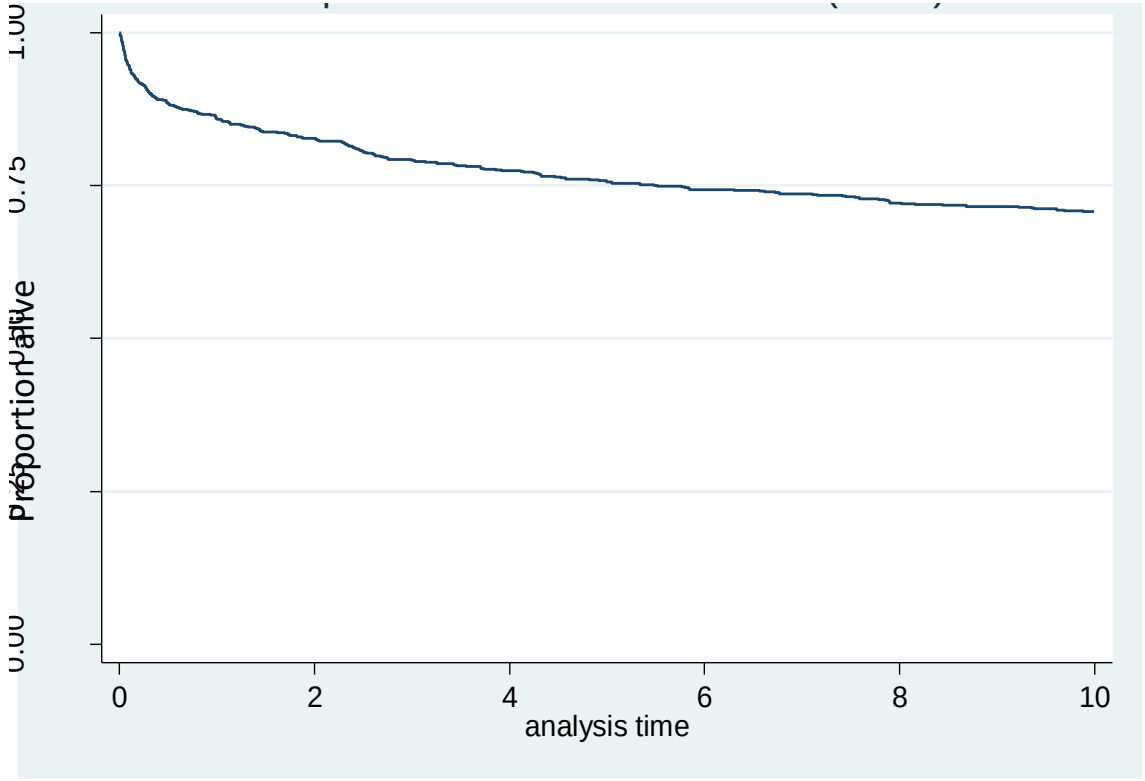
* 8% transferred, 12% lost

Applying 3 methods used to estimate survival to this cohort

1. **Survival Analysis** (Kaplan Meier methods): censor LTF[1]
2. **Tracing with Inverse Probability Weights (IPW)**: probability weights generated from tracing[2]
3. **Multiple Imputation with Chained Equations (MICE):** impute LTF and baseline characteristics that are missing[3]

1. Severe P et al. N Engl J Med. 2005.
2. Geng EH et al. J Acquir Immune Defic Syndr. 2010.
3. White IR, Royston P, Wood AM. *Stat Med*. 2011.

1. Kaplan Meier: censor LTF



Estimated alive:
71% 95%CI (68%,74%)

2. Tracing with Inverse Probability Weights

- A field worker traced patients who were LTF to determine outcome status
- Assume the ones found are a random sample of all LTF
- 156 patients categorized as LTF
- 45 were found
- Estimated alive: 71% 95%CI (68%, 74%)

45/156 of those initially LTF traced

$$\text{iweight} = \frac{1}{45/156} = 3.472$$

3. Multiple Imputation with Chained Equations

- Imputes the outcome status by using baseline covariates
- Fill in missing values present in covariates[3-5]
- Several equations are created to fill in missing values
- One must specify the number of datasets to generate, results will be averaged across datasets
- Assumptions:
  - Missing are only randomly different from patients with same set of covariates
  - LTF were assumed to have the same average survival as those not lost, conditional on covariates

3. White IR, Royston P, Wood AM. *Stat Med*. 2011.
4. White IR, Royston P. *Stat Med*. 2009.
5. Von Hippel PT. *Sociol Methods Res*. 2012.

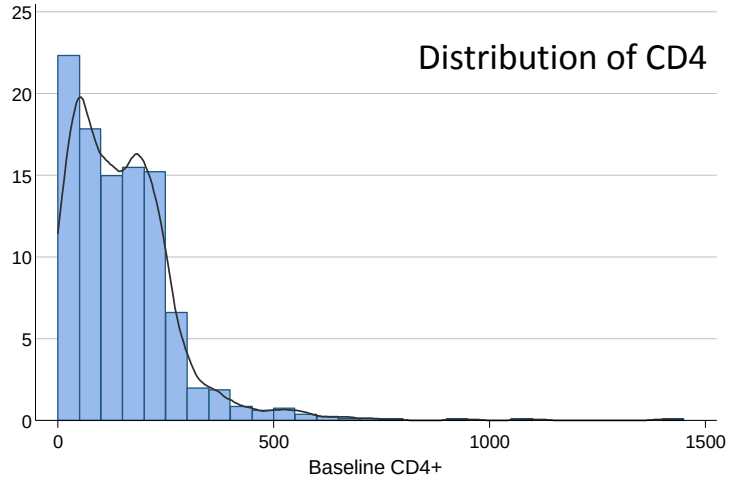
Applying this to our cohort: missing covariates

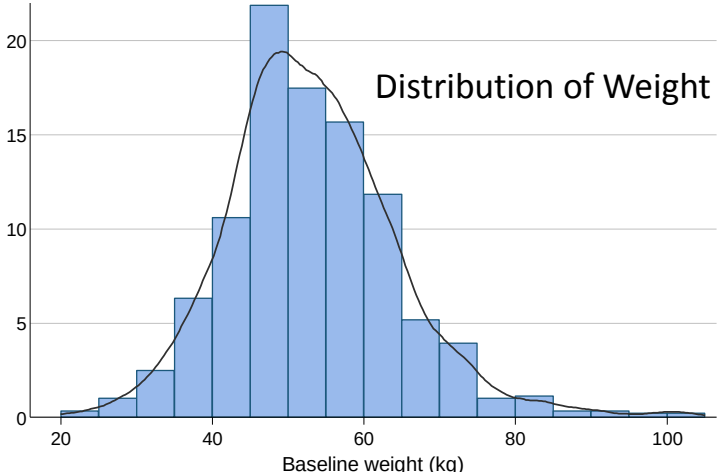
Demographic characteristics:

- Sex, age, residence, income

Clinical characteristics:

- CD4
  - Distribution 0-1400 cells/µL
  - Missing 12% of baseline CD4
  - "Missingness" associated with death: OR = 1.67 95% CI (1.09, 2.55)
- Weight
  - Distribution 20-120 kg
  - Missing 3% of baseline weight
  - "Missingness" associated with death: OR = 4.39 95% CI (1.86, 10.35)





Using Multiple Imputation with Chained Equations to impute missing covariates and outcome status

Chained Equations:

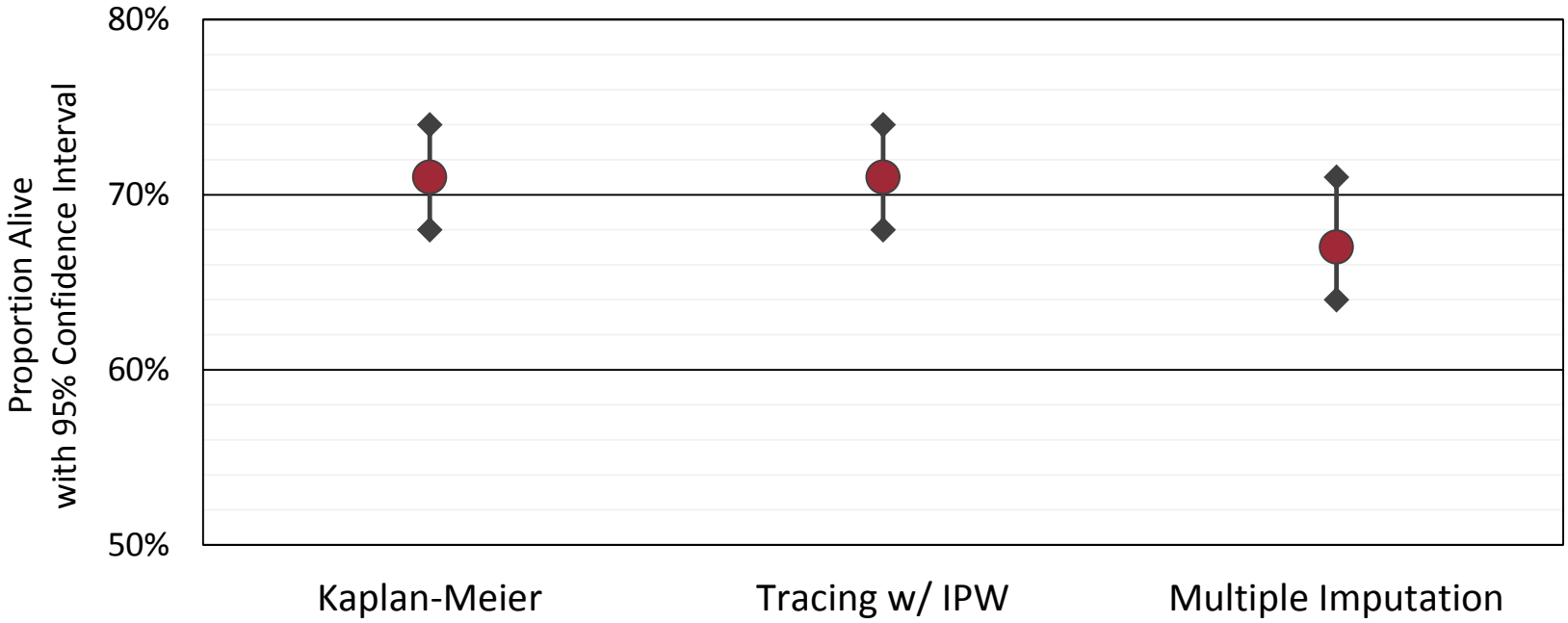
1. **Weight:** regress weight CD4 status age sex stage TB income residence
2. **CD4:** regress CD4 weight status age sex stage TB income residence
3. **10 year survival:** logit status weight CD4 age sex stage TB income residence
4. Repeated **20 times** to fill in all missing

Covariates filled in by MICE are similar to non-imputed values

Clinical Characteristic	Without Imputation	With Imputation
CD4+ count (cells/uL)		
Median (IQR) (range)	131 (51–212) [0–1400]	131(51–212) [-330–1416]
Missing	12%	N/A
Body weight (kg)		
Men median(IQR)	56 (50–63)	54 (46–61)
Women median (IQR)	49 (44–56)	47 (40–54)
Missing	3%	N/A

Estimated Survival: 67% (95% CI 64%-71%)

Primary outcome:
10 year survival estimated to be 67-71%



METHOD	ESTIMATED ALIVE % (95% CI)
Kaplan–Meier	71% (68, 74)
Tracing w/ IPW	71% (68, 74)
Multiple Imputation	67% (64, 71)

Secondary outcome: predictors of death

	Without Imputation		With Imputation	
Covariate	**Odds Ratio**	**95% CI**	**Odds Ratio**	**95% CI**
Female	**0.79**	**(0.55, 1.12)**	**0.61**	**(0.44, 0.87)**
Age	1.03	(1.01, 1.04)	1.03	(1.01, 1.05)
Residence	1.16	(0.82, 1.64)	1.14	(0.81, 1.59)
Income	**1.56**	**(1.09, 2.23)**	**1.81**	**(1.27, 2.58)**
CD4	1.00	(0.99, 1.00)	1.00	(1.00, 1.00)
Base weight	0.97	(0.95, 0.99)	0.96	(0.94, 0.98)
WHO stage	**1.51**	**(1.06, 2.14)**	**1.83**	**(1.31, 2.55)**
Baseline TB	**2.12**	**(1.24, 3.62)**	**1.59**	**(0.92, 2.73)**

Comparing methods

Method	Assumptions for LTF	How LTF is treated in the analysis	Missing Covariate Data
Survival Analysis	• LTF is unrelated to mortality • That is, they are a random sample of those who continue to be followed	• Censored	• Censored
Tracing w/ IPW	• Those unsuccessfully traced have the same mortality as those successfully traced	• Weighted	• Case-wise deletion
Multiple Imputation	• Missing are only randomly different from patients with same set of covariates	• Imputed	• Imputed • All observations used

Application of methods in our study

Method	Limitations	Strengths
Survival Analysis	• Most studies found assumption to be incorrect • Survival is usually overestimated	• Most common method • Easy to perform
Tracing w/ IPW	• Tracing was done at the end of the 10 year follow up period on everyone • Case-wise deletion if covariates are missing • Tracing can be difficult and expensive • Only as successful as your tracing success	• Common method in HIV studies • Conceptually easy to understand
Multiple Imputation	• Relies on a good prediction model • Biologically impossible values	•Use all observations •Robust standard error

Summary

1. LTF is a common category of patients in cohort studies (public health studies)
2. LTF is a mixture of patients (dead, alive)
3. Three commonly used methods estimate survival among LTF
4. Multiple Imputation with Chained Equations is a valid method that is infrequently used in public health
5. MICE estimated survival was different than the traditionally used methods
6. Potentially we could use MICE to impute survival time

Acknowledgements

GHESKIO staff

Weill Cornell Medical College
Division of Hospital Medicine

Weill Cornell Medical College Center for Global Health

Any thoughts on how to impute survival time or how to deal with violations of PH assumptions?

- Imputing survival time
- Augmenting/ limiting imputations
- Recommendations for how to deal with violations of PH assumptions: Aalen models or time varying or both

References & Resources

Multiple Imputation

White IR, Royston P. Imputing missing covariate values for the Cox model. *Stat Med*. 2009;28(15):1982-1998.

White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat Med*. 2010;29(28):2920-2931.

White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med*. 2011;30(4):377-399.

Von Hippel PT. Should a Normal Imputation Model be Modified to Impute Skewed Variables? *Sociol Methods Res*. 2012;42(1):105-138.

Rodwell L, Lee KJ, Romaniuk H, Carlin JB. Comparison of methods for imputing limited-range variables: a simulation study. *BMC Med Res Methodol*. 2014;14:57.

Marchenko, Yulia. Chained Equations and more in multiple imputation in STATA 12. 2011 UK Stata Users Group Meeting. Online at http://www.stata.com/meeting/uk11/abstracts/UK11_marchenko.pdf

GHESKIO

Severe P, Leger P, Charles M, et al. Antiretroviral therapy in a thousand patients with AIDS in Haiti. *N Engl J Med*. 2005;353(22):2325-2334.

Leger P, Charles M, Severe P, Riviere C, Pape JW, Fitzgerald DW. 5-year survival of patients with AIDS receiving antiretroviral therapy in Haiti. *N Engl J Med*. 2009;361(8):828-829

MICE code

```
mi set mlong

mi register regular age sex WHO_stage base_TB
income pap self_referred

mi register imputed base_wt CD4 10year_outcome

mi impute chained (regress) base_wt_kg (regress)
CD4 (logit) 10year_outcome= age sex WHO_stage
base_TB self_referred income pap, add(20)
rseed(1458) burnin(20) savetrace(impstats21915,
replace) dryrun
```

We chose age, sex, WHO stage, baseline and incident TB, income, residence, being self referred, weight, CD4 and outcome status at 6m and 10 years based on clinical, programmatic and research experience

MICE diagnostics

```
*check to see if the imputed values are close enough for all imputed
covariates
midiagplots base_wt, m(1/5) combine

*trace plots
use impstats21915
reshape wide *mean *sd, i(iter) j(m)
tsset iter
tsline base_wt_kg_mean*, name(graph1b) nodraw legend(off)

graph combine graph1b graph2b graph3b graph4b graph5b graph6b graph7b
graph8b graph9b graph10b, title(trace plots of summaries of imputed
values from 20 chains) rows(5)

* check for proportions and confidence intervals:
mi estimate: proportion Itdead_10
```

→ Marchenko STATA presentation great reference!