Survival Methods in Electronic Health Records Data in the Presence of System Migration

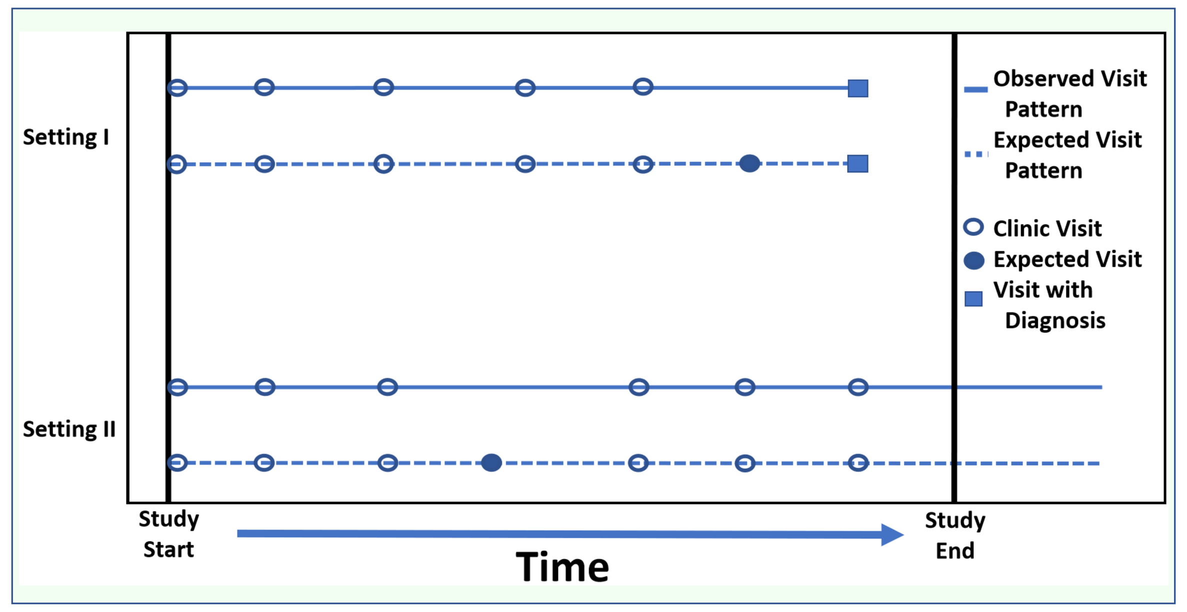
Website & Contact

Kyle Conniff[1], Luohua Jiang, PhD[2], Joan O'Connell, PhD[3], Daniel L. Gillen, PhD[1]

1 University of California, Irvine Department of Statistics
2 University of California, Irvine Department of Epidemiology and Biostatistics
3 Centers for American Indian and Alaska Native Health, University of Colorado Anschutz Medical Campus

Introduction

- Electronic Health Record data are subject to several biases (confounding bias, selection bias, informed-presence bias, and misclassification bias to name a few).
- In time-to-diagnosis studies (e.g. time-to-dementia diagnosis), patients may have multiple insurance and healthcare provider options.
- When patients switch between multiple providers, valuable information may be missing within a single EHR database.
- We define System Migration (SM): patients using multiple healthcare systems.



Motivation

This methodology is motivated by research into Alzheimer's disease and related dementias in Indian Health Services (IHS) data. The IHS is a federally-funded health system for qualifying American Indians and Alaskan Natives. In fiscal year 2010, 93.5% of IHS-enrolled participants aged 65+ also had Medicare coverage[1]. This led to concern that patients may be using multiple healthcare systems for treatment. The application of this method to IHS data is undergoing IHS and tribal review as a key component data sovereignty and decolonization of Indigenous health data.

Methodology

To account for SM in EHR data, we propose a two-step solution:

1. Build predictive return-to-system model to identify patients who are potentially migrating between systems.
2. For the patients identified in (1), use multiple imputation to impute return-to-system times that are more "normal" for the given patient.

References

[1] O'Connell et al. ARRA ACTION: Comparative Effectiveness of Health Care Delivery Systems for American Indians and Alaska Natives Using Enhanced Data Infrastructure. 2014.

[2] Law and Brookmeyer. Effects of Mid-point Imputation on the Analysis of Doubly Censored Data. 1992.

Acknowledgements

This research was supported by the National institutes of Health's National Institute of Aging Diversity Supplement PA-21-071 under the R01 parent grant (5R01AG061189) awarded to Drs. Joan O'Connell and Luohua Jiang.

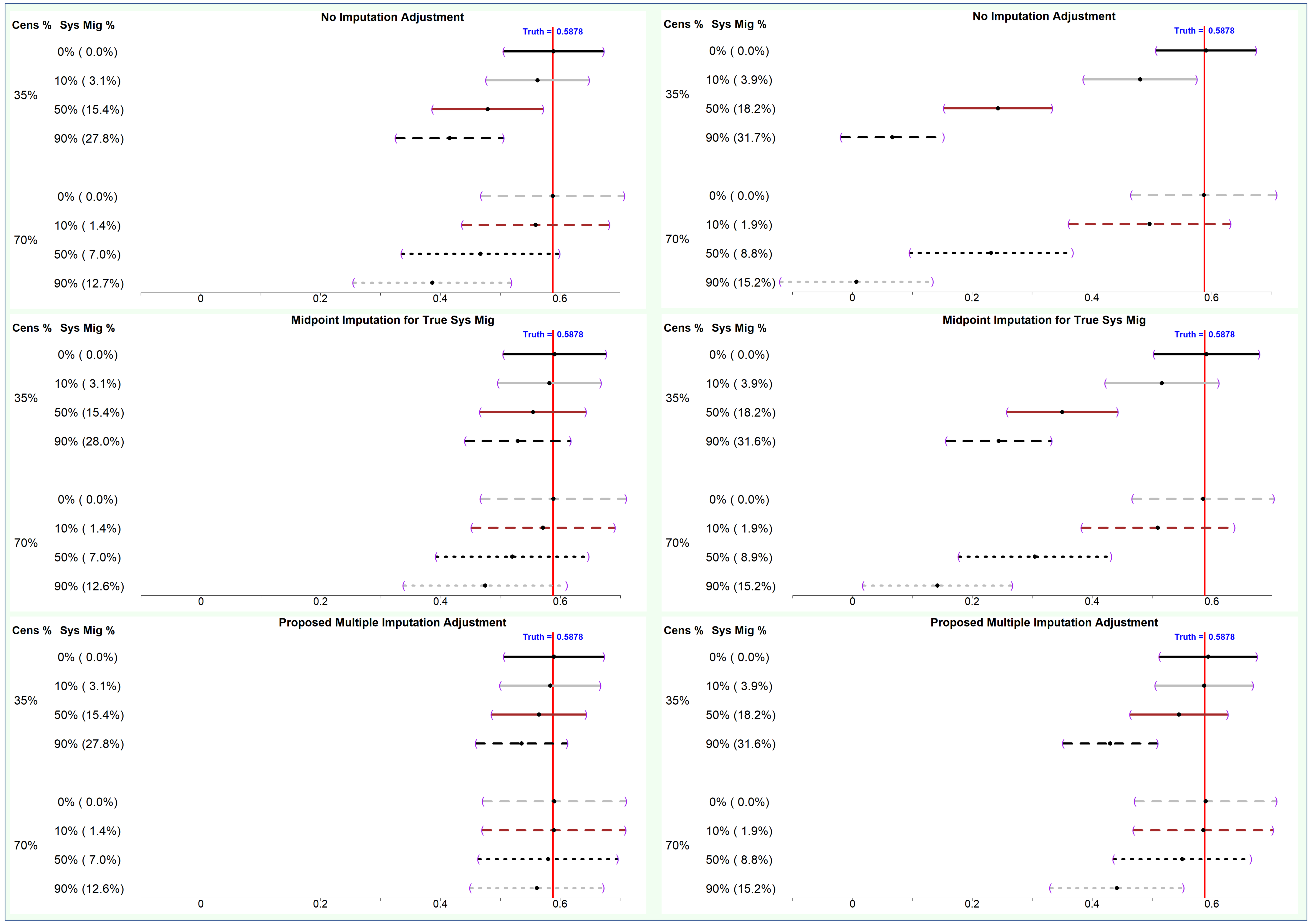
Results

MAR1

MAR1: SM does not depend on any covariates.

MAR2

MAR2: SM depends only on collected covariates.



Cens Rate	SM% (SMD%)	No Imputation: Coef Est (% Bias)	No Imputation: Mod Var	No Imputation: Emp Var	No Imputation: Cov %	Midpoint Imputation: Coef Est (% Bias)	Midpoint Imputation: Mod Var	Midpoint Imputation: Emp Var	Midpoint Imputation: Cov %	Multiple Imputation: Coef Est (% Bias)	Multiple Imputation: Mod Var	Multiple Imputation: Emp Var	Multiple Imputation: Cov %
0.35	0 (0.0)	0.59 (0.2)	1.8	1.8	96	0.59 (0.5)	1.8	1.9	94	0.59 (0.3)	1.8	1.8	95
0.35	10 (3.1)	0.56 (-4.3)	1.8	1.9	90	0.58 (-1.1)	1.8	1.9	94	0.58 (-0.8)	1.8	1.8	95
0.35	50 (15.4)	0.48 (-18.5)	1.7	2.2	29	0.55 (-5.7)	1.8	2.0	85	0.56 (-4.0)	1.8	1.6	93
0.35	90 (27.8)	0.42 (-29.3)	1.7	2.1	3	0.53 (-10.0)	1.7	2.0	70	0.54 (-8.9)	1.9	1.5	80
0.70	0 (0.0)	0.59 (0.0)	3.7	3.7	95	0.59 (0.1)	3.7	3.8	94	0.59 (0.4)	3.7	3.7	96
0.70	10 (1.4)	0.56 (-4.9)	3.7	3.9	91	0.57 (-2.9)	3.7	3.7	94	0.59 (0.3)	3.7	3.7	95
0.70	50 (7.0)	0.47 (-20.5)	3.6	4.5	46	0.52 (-11.7)	3.6	4.2	77	0.58 (-1.4)	3.8	3.5	95
0.70	90 (12.7)	0.39 (-34.2)	3.6	4.5	11	0.47 (-19.3)	3.6	4.8	51	0.56 (-4.6)	4.0	3.2	95

Cens Rate	SM% (SMD%)	No Imputation: Coef Est (% Bias)	No Imputation: Mod Var	No Imputation: Emp Var	No Imputation: Cov %	Midpoint Imputation: Coef Est (% Bias)	Midpoint Imputation: Mod Var	Midpoint Imputation: Emp Var	Midpoint Imputation: Cov %	Multiple Imputation: Coef Est (% Bias)	Multiple Imputation: Mod Var	Multiple Imputation: Emp Var	Multiple Imputation: Cov %
0.35	0 (0.0)	0.59 (0.5)	1.8	1.8	95	0.59 (0.6)	1.8	2.0	94	0.59 (1.0)	1.8	1.7	96
0.35	10 (3.9)	0.48 (-18.3)	1.7	2.3	29	0.52 (-12.1)	1.7	2.3	57	0.59 (-0.2)	1.8	1.7	97
0.35	50 (18.2)	0.24 (-58.6)	1.6	2.1	0	0.35 (-40.4)	1.5	2.2	0	0.54 (-7.3)	2.1	1.7	86
0.35	90 (31.7)	0.07 (-88.7)	1.6	1.9	0	0.24 (-58.5)	1.5	2.0	0	0.43 (-26.8)	2.2	1.6	4
0.70	0 (0.0)	0.59 (-0.2)	3.7	3.8	94	0.59 (-0.4)	3.7	3.6	96	0.59 (0.3)	3.7	3.6	95
0.70	10 (1.9)	0.50 (-15.6)	3.6	4.7	64	0.51 (-13.3)	3.6	4.2	72	0.59 (-0.4)	3.7	3.5	96
0.70	50 (8.8)	0.23 (-60.6)	3.6	4.8	0	0.30 (-48.2)	3.5	4.2	1	0.55 (-6.4)	4.2	3.4	94
0.70	90 (15.2)	0.01 (-98.9)	3.7	4.2	0	0.14 (-75.9)	3.5	4.0	0	0.44 (-24.9)	4.7	3.2	40

Tables 1 & 2: Coefficient estimates of a continuous covariate (n = 1000) from Cox PH model under MAR1 (left) and MAR2 (right) SM (1000 simulations). Variances are all x10^{-3}. SMD% represents the percent of patients whose SM resulted in displaced diagnosis time.

System migration multiple imputation method

Step 0:	Given longitudinal survival data, a number of imputations, *n.imps*, and a cutoff for system migration, *p.cut*
Step 1:	Estimate $S_i(t)$ with $exp\{\hat{\Lambda}_{0i} e^{x_i \hat{\gamma}}\}$
Step 2:	Predict $p_{m_i} = \hat{S}_i(T_{K_i} - T_{(K-1)_i})$
Step 3:	If $(p.cut > p_{m_i})$ then set $p_{m_i} = 0$. While j in *n.imps*
Step 4:	$MI_i \sim Bern(p_{m_i})$
Step 5:	If$(MI_i == 1)$, then draw $U_i \sim \text{Unif}(0,1)$
Step 6:	Set $T_{K_i} = \min(T_{(K-1)_i} + \hat{S}_i^{-1}(U_i), T_{K_i})$
Step 7:	set $\beta_j =$ Estimate β
Step 8:	set $\hat{\beta} = \sum \frac{\beta_j}{n.imps}$

where $\hat{\Lambda}_{0i} = \sum_{j:t_j \le t} \frac{d_j}{\sum_{i \in R_j} exp\{x_i \hat{\gamma}\}}$ is the Nelson-Aalen estimator

$\hat{\gamma}$ are Cox PH coefficients

p_{m_i} is individual i's probability of system migration,

$\hat{S}_i(t)$ is subject i's probability of returning after time t,

T_{K_i} is the last observation (time of event) for individual i,

MI_i is an indicator of individual i's system migration status.

Results

- SM can induce bias with as little as 4% of patients being diagnosed in another system.
- Knowing who is migrating across systems is not enough for effective imputation.
- The proposed model can reduce bias resulting from SM.
- Variances are not impacted by SM.

Conclusions

- SM can induce a missing at random mechanism even when an individual's SM does not depend directly on covariates.
- Hazard ratio estimates are biased toward the null
  - Patients with earlier event times look similar to patients with later event times.
- Midpoint imputation[2] on SM patients is not enough to completely remove bias.
  - It is important to account for the time at which patients are predicted to leave the system.
- The proposed multiple imputation method:
  - removes bias with low-to-moderate amounts of SM.
  - reduces bias in more extreme SM settings.

Next steps

- Extend this methodology for robustness to model misspecification.
- Extend time-varying area under the receiver operating characteristic curve to the recurrent event setting to improve the prediction modeling of this method.