CROSS-SECTIONAL LENGTH-BIASED SEMI-COMPETING RISKS DATA

Siyuan Guo

Department of Epidemiology and Biostatistics

Arnold School of Public Health

University of South Carolina

UNIVERSITY OF
South Carolina

OUTLINE

- Background
- Data Structure
- Motivation
- Method
- Numerical Study and Data Analysis
- Discussion and future direction

U of SC South Carolina

BACKGROUND

- Infertility
  - Infertility is defined as couple taking longer than 12 months to get pregnant (some define as longer than 24 months).
  - More than 12 months was required to conceive the prevalence of infertility and results were vary from 3% to 26%(Gurunath and others,2011).

U of SC South Carolina

MOTIVATION

- An estimated 7.3 million women (12%) aged 15–44 in 2006–2010 had ever used fertility treatment.
- In current method, these women are excluded from the analysis or the fertility treatment is assumed no effect on time to pregnancy. (Keiding et al biostats 2002)
- In the 2002 National Survey of Family Growth (NSFG) data, about 17% couples reported that they had sought fertility treatment during their current attempt using the most restrictive definition.
- Aim: Find method that can estimate the population time to pregnancy with the fertility treatment may occur.

NATIONAL SURVEY OF FAMILY GROWTH (NSFG)

- Gathers information on family life, marriage and divorce, pregnancy, infertility, use of contraception, and men's and women's health.
- The survey results are used by the U.S. Department of Health and Human Services and others to plan health services and health education programs, and to do statistical studies of families, fertility, and health.
- https://www.cdc.gov/nchs/nsfg/index.htm

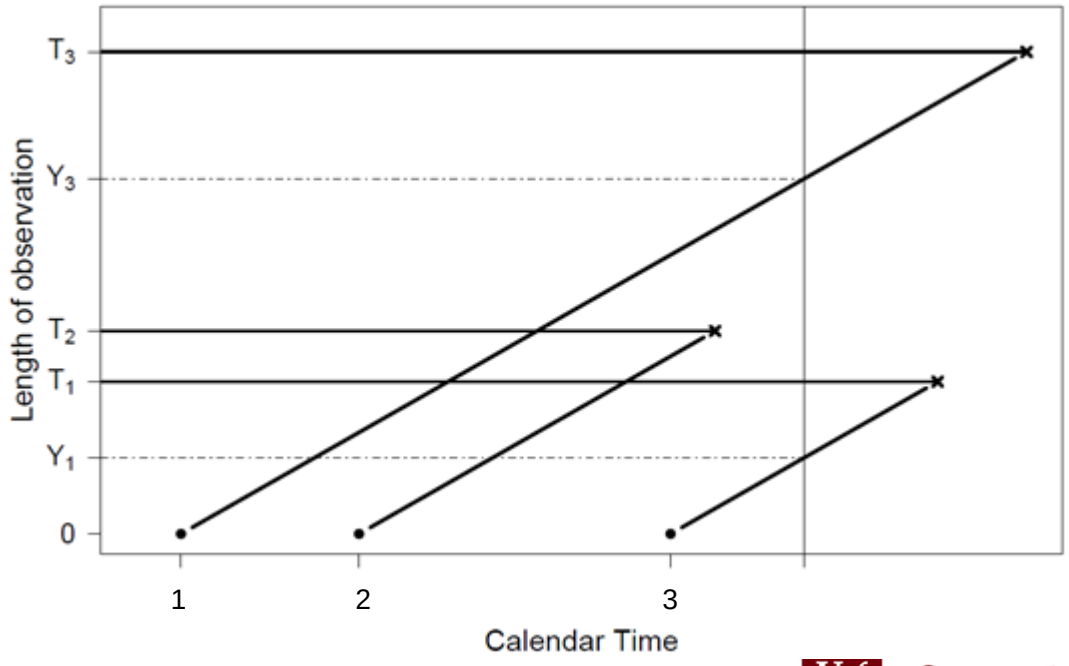
DATA SOURCE: NSGF

In order to estimate the time to pregnancy, we use the 2002, 2006-2010, 2011-2015 cycles of the NSFG data

- It contains 1047 women reported currently attempting pregnancy and gave their current duration of pregnancy attempt.
- 372 (36%) reported receiving medical treatment to help to get pregnant during current duration of pregnancy attempt.

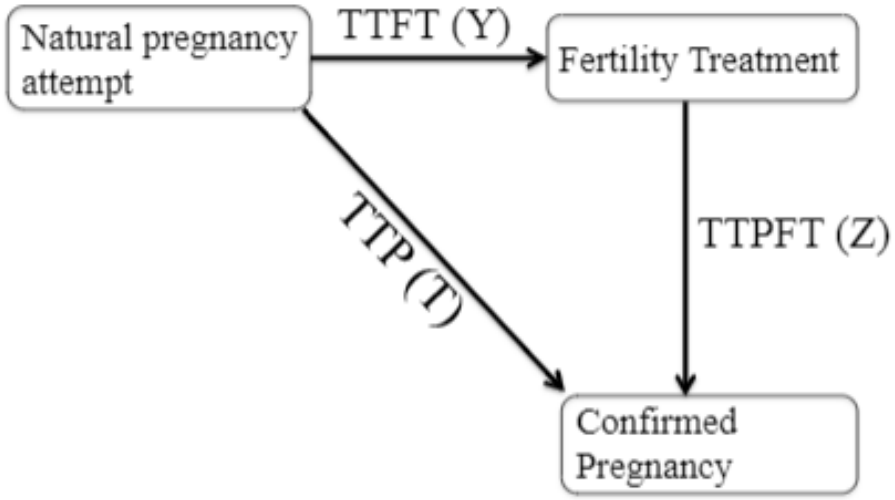
CROSS-SECTIONAL LENGTH-BIASED DATA

- Y= The current duration of pregnancy attempt (observed)
- T= The total duration of pregnancy attempt (unobserved)
- Subject 1 and 3 were observed
- Length-Biased: Subject 2



U of SC South Carolina

CROSS-SECTIONAL LENGTH-BIASED SEMI-COMPETING DATA



- T= The time to nature pregnancy (Main event)
- Y= The time to fertility treatment (intermittent event)
- Z= The time to pregnancy on fertility treatment

USC South Carolina

STANDARD APPROACH

- Fertility treatment was not included
- Current duration approach (van Es et al., 2000 JSPI)

$$g(y) = \frac{d}{dy} G(x) = \frac{d}{dy} \int_0^\infty \left(\frac{y}{x} \wedge 1 \right) f_x(x) dx = \frac{\bar{F}_T(y)}{\mu_T}$$

- With g(y) estimated, the survival function of T can be estimated as

$$\hat{\bar{F}}_T(t) = \hat{g}(t)\hat{g}^{-1}(0)$$

U of SC South Carolina

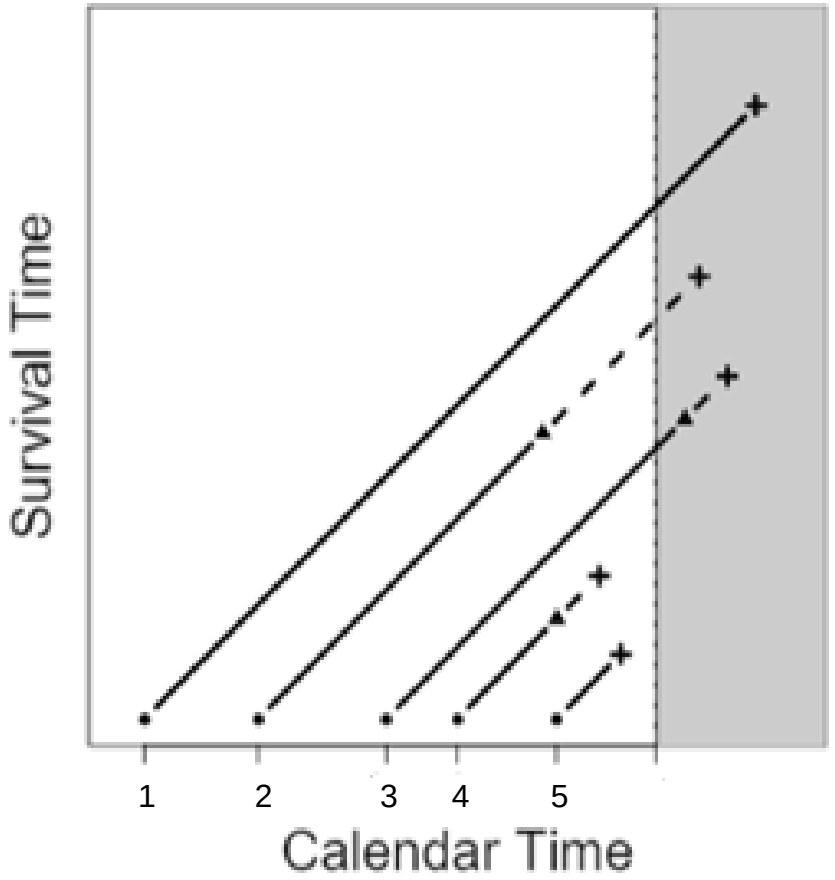
PROPOSED METHOD: OBSERVED DATA

- Cross-sectional length-biased Semi-competing Data
- $D_i = \{X_i, Z_i^*, \delta_i\}$ for i in 1,2,…,n where
  - X_i is the observed time at risk, i.e. min(T,Y).
  - Z_i^* is the observed time since the fertility treatment if the fertility treatment occurred.
  - $\delta_i = I(Z_i^* > 0)$ for subject i, indicates the occurrence of the fertility treatment.

U of SC South Carolina

PROPOSED METHOD: OBSERVED DATA

- Subject 1,2,3 were observed
- Subject 2 had fertility treatment occurred



U of SC South Carolina

PROPOSED METHOD

- Extend the standard approach to new model with likelihood function:

$$L(\mathcal{D}_i|\theta) \propto \frac{\bar{F}_{T\wedge Y}(X_i)^{(1-\delta_i)}\left\{\bar{F}_{T|Y}(X_i - Z_i|X_i - Z_i)f_Y(X_i - Z_i)\bar{F}_Z(Z_i)\right\}^{\delta_i}}{E(Z)\Pr(T > Y) + E(T \wedge Y)}$$

- The parameters are estimated using maximum likelihood analysis by maximizing $l(\theta|D) = \sum_{i=1}^{n} \log\{L(D_i|\theta)$

SIMULATION

- Assumptions:
- Hazard functions of T, Y, and Z follow piecewise constant hazard model(McLain et al. 2014)
- The dependence between variables were modeled using Gaussian Copula.
- Y|T and Z|T are independent and $\rho_{TY} < 0, \rho_{TZ} > 0$
- Parameters set: $\theta := \{\alpha_T, \alpha_Y, \alpha_Z, \rho_{TY}, \rho_{TZ}\}$
- Comparison between Naïve method results and proposed method results was done.

U of SC South Carolina

SIMULATION

- Standard method: fertility treatment not included.

		$n = 1{,}000$			$n = 2{,}000$		
		$\rho_{xy} = -0.2$ $\rho_{xz} = 0.4$	$\rho_{xy} = -0.2$ $\rho_{xz} = 0.6$	$\rho_{xy} = -0.6$ $\rho_{xz} = 0.6$	$\rho_{xy} = -0.2$ $\rho_{xz} = 0.4$	$\rho_{xy} = -0.2$ $\rho_{xz} = 0.6$	$\rho_{xy} = -0.6$ $\rho_{xz} = 0.6$
t	TRUE	BIAS(SE)	BIAS(SE)	BIAS(SE)	BIAS(SE)	BIAS(SE)	BIAS(SE)
		Standard Method					
		$\bar{F}_T(t)$					
3	0.773	-0.039(0.079)	-0.025(0.082)	-0.028(0.084)	-0.036(0.059)	-0.026(0.062)	-0.023(0.063)
6	0.578	-0.054(0.058)	-0.042(0.061)	-0.054(0.065)	-0.051(0.042)	-0.042(0.043)	-0.053(0.045)
12	0.294	-0.083(0.022)	-0.059(0.027)	-0.076(0.025)	-0.082(0.016)	-0.059(0.020)	-0.076(0.018)
24	0.060	0.028(0.009)	0.045(0.012)	0.037(0.011)	0.028(0.007)	0.045(0.009)	0.037(0.008)
36	0.012	0.036(0.005)	0.046(0.007)	0.042(0.006)	0.037(0.004)	0.046(0.005)	0.042(0.004)

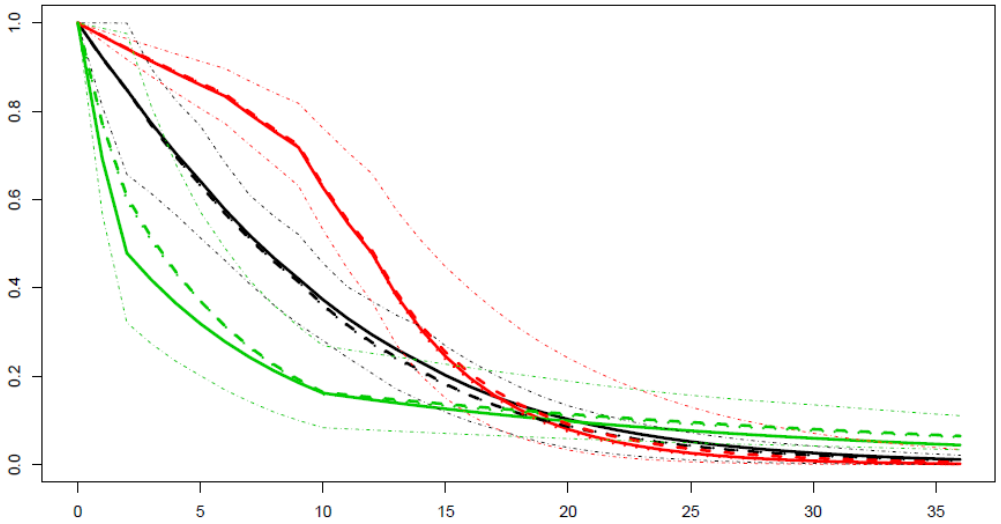
SIMULATION

- Proposed method

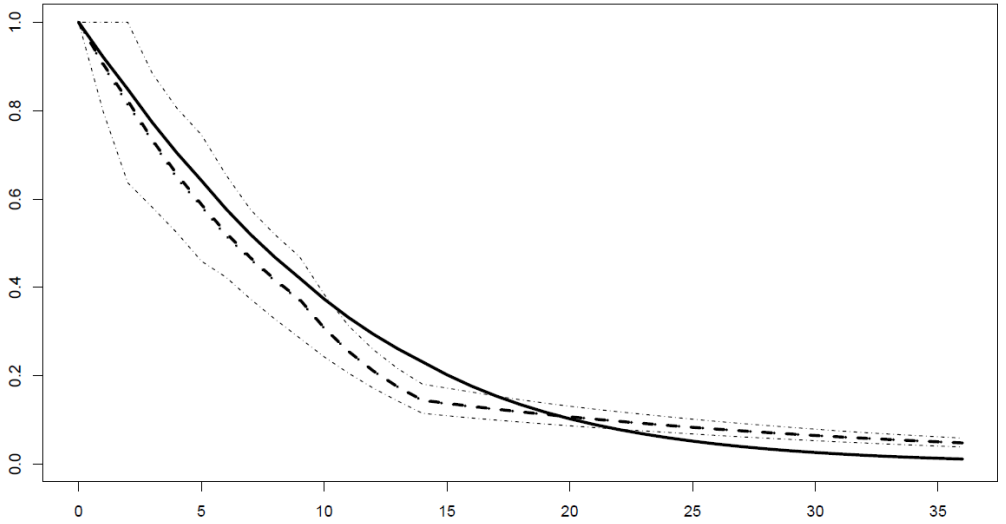
		$n = 1{,}000$			$n = 2{,}000$		
		$\rho_{xy} = -0.2$ $\rho_{xz} = 0.4$	$\rho_{xy} = -0.2$ $\rho_{xz} = 0.6$	$\rho_{xy} = -0.6$ $\rho_{xz} = 0.6$	$\rho_{xy} = -0.2$ $\rho_{xz} = 0.4$	$\rho_{xy} = -0.2$ $\rho_{xz} = 0.6$	$\rho_{xy} = -0.6$ $\rho_{xz} = 0.6$
t	TRUE	BIAS(SE)	BIAS(SE)	BIAS(SE)	BIAS(SE)	BIAS(SE)	BIAS(SE)
		Proposed Method					
		$\bar{F}_T(t)$					
3	0.773	-0.003(0.100)	0.010(0.104)	0.002(0.098)	-0.005(0.074)	0.008(0.082)	0.011(0.075)
6	0.578	-0.006(0.079)	0.003(0.084)	0.001(0.076)	-0.009(0.054)	0.003(0.062)	0.008(0.058)
12	0.294	-0.014(0.058)	-0.002(0.061)	-0.013(0.068)	-0.017(0.041)	-0.005(0.043)	-0.009(0.048)
24	0.060	-0.013(0.024)	-0.012(0.022)	-0.023(0.021)	-0.011(0.018)	-0.011(0.016)	-0.020(0.016)
36	0.012	-0.003(0.007)	-0.003(0.007)	-0.006(0.004)	-0.003(0.006)	-0.003(0.005)	-0.006(0.003)

SIMULATION

Proposed Method



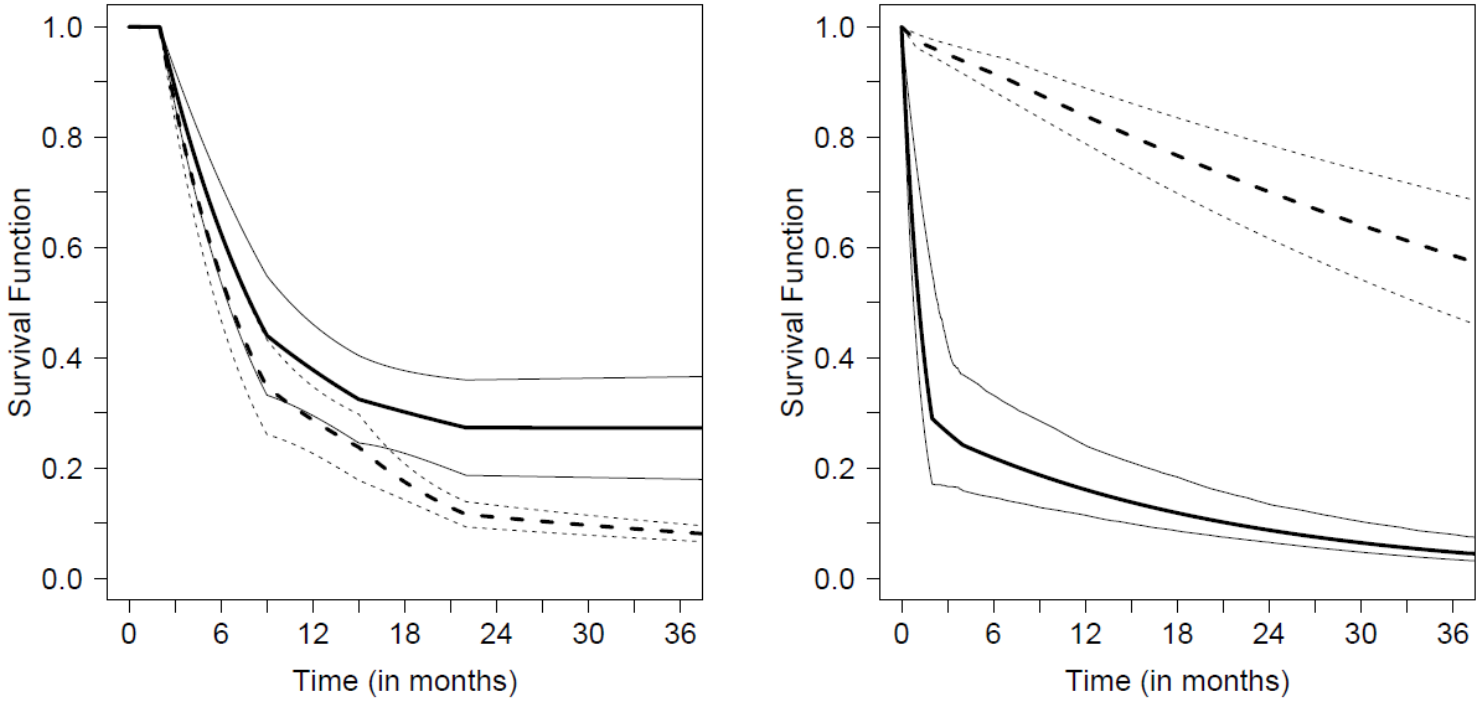
Standard Method



- Left figure: True survival function (valid) and estimated survival function (dashed) for all three events with $\rho_{XY=-0.2}, \rho_{YZ=0.4}$ and 2000 repetition using proposed method.
- Right figure: True survival function (valid) and estimated survival function (dashed) for time to get pregnant using standard method.

Univ of SC South Carolina

DATA ANALYSIS



- Left figure: Survival function estimation for time to pregnancy from proposed method (valid) and standard method (dashed).
- Right figure: Survival function estimation for time to fertility treatment (dashed) and attempt time to pregnancy from the fertility treatment (valid).

U of SC South Carolina

DISCUSSION

- The fertility treatment has help on the probability to get pregnant.
- In general, the proposed method estimate the survival function well when the intermittent event exists.
- This method can be extended to other area, such as unemployment.

U of SC South Carolina

FUTURE DIRECTION

- Our proposed method is a parametric approach, more work on nonparametric approach is needed.
- Developing methods to incorporate some covariates for prediction issues.
- There are more possible reasons that a couple stop attempting (e.g. giving up), more researches on this is needed.

REFERENCE

- van Es B., Klaassen C. A. J., and Oudshoorn K. (2000), “Survival analysis under cross-sectional sampling: length bias and multiplicative censoring.” J. Statist. Plann. Inference, 91, 295–312.
- Keiding, N., Højbjerg Hansen, O. K., Sørensen, D. N., and Slama, R. (2012), “The current duration approach to estimating time to pregnancy,” Scandinavian Journal of Statistics, 39, 185–204.
- Dauxois, J. Y., Guilloux, A., and Kirmani, S. (2014), “Estimation in a competing risks proportional hazards model under length-biased sampling with censoring,” Lifetime Data Analysis, 20, 276–302.
- McLain A. C., Sundaram R., Thoma M., and Buck Louis G. M. (2014), “Semiparametric modeling of grouped current duration data with preferential reporting.” Statistics in Medicine, 33, 3961–3972.

U of SC South Carolina

THANKS!

Collaborator:

Alexander C. McLain, University of South Carolina

Marie Thoma, University of Maryland

Jiajia Zhang, University of South Carolina