

# Supplementary Material to Survival analysis with functions of mis-measured covariate histories: the case of chronic air pollution exposure in relation to mortality in the Nurses' Health Study

Xiaomei Liao<sup>\*1,2,4</sup>, Molin Wang<sup>1,2,4</sup>, Jaime E. Hart<sup>3,4</sup>, Francine Laden<sup>2,3,4</sup>,  
and Donna Spiegelman<sup>1,2,4,5,6</sup>

<sup>1</sup>Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, 02115, USA.

<sup>2</sup>Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, 02115, USA.

<sup>3</sup>Department of Environmental Health, Harvard T. H. Chan School of Public Health, Boston, MA, 02215, USA.

<sup>4</sup>Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02115, USA.

<sup>5</sup>Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, 02115, USA.

<sup>6</sup>Department of Global Health and Population, Harvard T. H. Chan School of Public Health, Boston, MA, 02115, USA.

July 16, 2015

---

\*Email: [stxia@channing.harvard.edu](mailto:stxia@channing.harvard.edu).

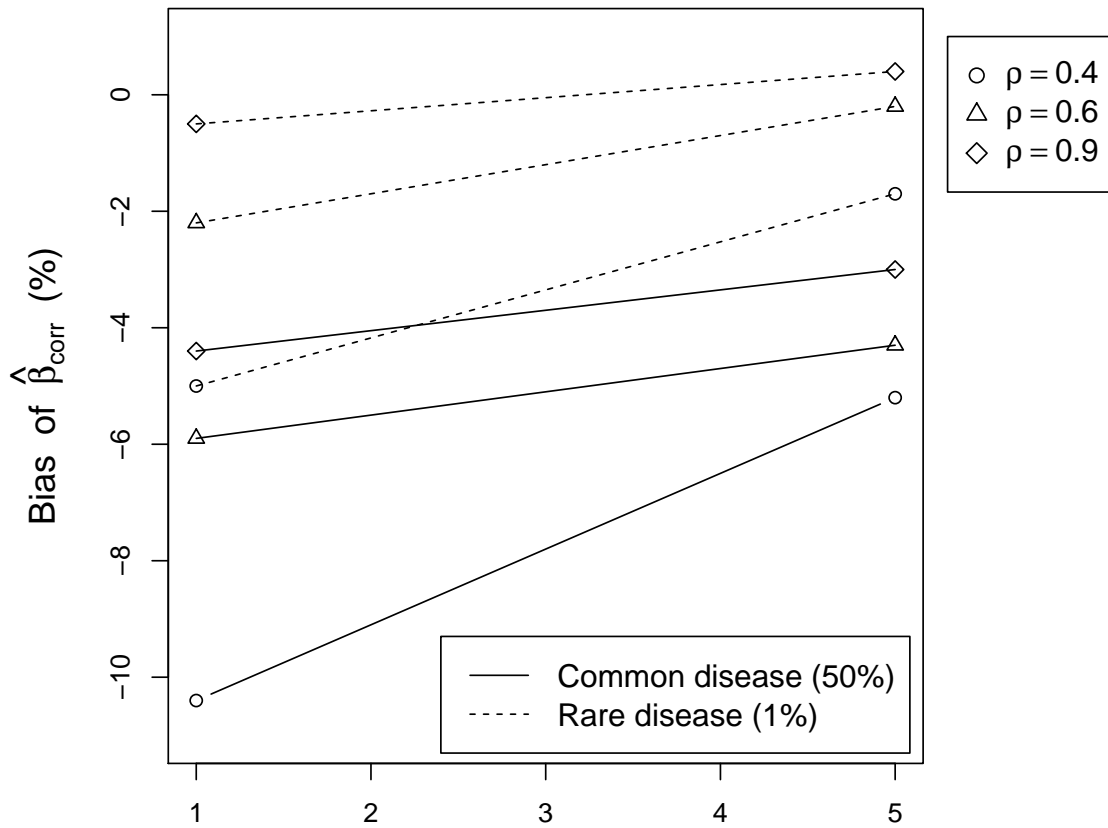
This research was supported by NIH/NIEHS grant R01ES09411.

Supplementary Table 1: Results for the simulation study of 12-month moving average exposure with a compound symmetry covariance structure, for different intra-class correlations ( $\rho_{ICS}$ ) and different amounts of measurement error ( $\rho$ ).

$\rho_{ICS}$	$\rho$	Estimated $\hat{\beta}(\hat{SE}[\hat{\beta}])$		Relative Bias(%)		95% CI Coverage(%)	
		Naive	Corrected	Naive	Corrected	Naive	Corrected
There is 1 measurement per validation study subject							
		$n_1 = 1000,$	$n_2 = 500,$	Common disease			
0.3	0.4	0.321( 0.076)	0.421( 0.143)	-35.7	-15.8	35.7	90.1
	0.6	0.328( 0.076)	0.464( 0.128)	-34.4	-7.1	37.3	95.3
	0.9	0.325( 0.076)	0.475( 0.114)	-35.0	-5.0	35.7	96.3
0.6	0.4	0.327( 0.058)	0.448( 0.126)	-34.6	-10.4	14.5	91.5
	0.6	0.326( 0.058)	0.470( 0.104)	-34.7	-5.9	14.1	93.7
	0.9	0.326( 0.058)	0.478( 0.088)	-34.9	-4.4	16.1	94.2
0.9	0.4	0.325( 0.049)	0.459( 0.121)	-34.9	-8.1	6.2	90.8
	0.6	0.328( 0.049)	0.475( 0.094)	-34.3	-5.1	5.8	95.2
	0.9	0.328( 0.049)	0.481( 0.076)	-34.5	-3.8	5.9	95.6
		$n_1 = 50000,$	$n_2 = 500,$	Rare disease			
0.3	0.4	0.344( 0.075)	0.446( 0.148)	-31.2	-10.9	44.8	91.9
	0.6	0.349( 0.075)	0.483( 0.127)	-30.3	-3.3	47.9	94.3
	0.9	0.347( 0.075)	0.498( 0.112)	-30.7	-0.4	47.6	95.6
0.6	0.4	0.349( 0.056)	0.475( 0.133)	-30.2	-5.0	21.4	94.4
	0.6	0.346( 0.056)	0.489( 0.104)	-30.8	-2.2	23.2	95.6
	0.9	0.345( 0.056)	0.497( 0.085)	-31.0	-0.5	21.8	95.1
0.9	0.4	0.348( 0.046)	0.487( 0.129)	-30.4	-2.5	9.1	93.8
	0.6	0.348( 0.047)	0.495( 0.095)	-30.5	-1.0	8.6	96.5
	0.9	0.342( 0.047)	0.494( 0.072)	-31.6	-1.1	7.2	94.3
There are 5 measurements per validation study subject							
		$n_1 = 1000,$	$n_2 = 500,$	Common disease			
0.3	0.4	0.325( 0.076)	0.457( 0.117)	-34.9	-8.6	37.6	92.8
	0.6	0.325( 0.076)	0.472( 0.114)	-35.1	-5.6	36.3	95.1
	0.9	0.332( 0.076)	0.488( 0.112)	-33.7	-2.5	39.5	95.5
0.6	0.4	0.327( 0.058)	0.474( 0.094)	-34.6	-5.2	15.6	95.0
	0.6	0.327( 0.058)	0.479( 0.089)	-34.6	-4.3	16.9	95.0
	0.9	0.329( 0.058)	0.485( 0.086)	-34.2	-3.0	16.1	95.7
0.9	0.4	0.325( 0.049)	0.477( 0.082)	-35.0	-4.5	5.9	94.0
	0.6	0.324( 0.049)	0.477( 0.076)	-35.1	-4.7	4.6	94.0
	0.9	0.329( 0.049)	0.484( 0.073)	-34.3	-3.2	5.7	94.5
		$n_1 = 50000,$	$n_2 = 500,$	Rare disease			
0.3	0.4	0.344( 0.075)	0.477( 0.115)	-31.2	-4.5	45.1	96.4
	0.6	0.347( 0.075)	0.495( 0.111)	-30.6	-1.1	46.2	95.4
	0.9	0.347( 0.075)	0.500( 0.109)	-30.6	0.1	46.7	95.6
0.6	0.4	0.346( 0.056)	0.492( 0.091)	-30.8	-1.7	20.6	94.5
	0.6	0.347( 0.056)	0.499( 0.085)	-30.7	-0.2	21.6	94.2
	0.9	0.348( 0.056)	0.502( 0.082)	-30.5	0.4	24.0	93.1
0.9	0.4	0.346( 0.046)	0.493( 0.079)	-30.9	-1.3	8.6	95.3
	0.6	0.346( 0.046)	0.499( 0.072)	-30.9	-0.3	8.5	95.2
	0.9	0.345( 0.047)	0.499( 0.068)	-31.0	-0.3	8.2	94.4

True  $\beta = 0.5$ , the study duration  $t^* = 84$ , the number of simulation replications  $B = 1000$ . In the common disease situation, the cumulative incidence was about 50% with  $n_1 = 1000$ . In the rare disease situation, the cumulative incidence was about 1% with  $n_1 = 50000$ .

CS structure (  $\rho_{l_{CS}} = 0.6$  ,  $\tau = 0$  )



The number of measurements per subject in the validation study ( $n_2 = 500$ )

Supplementary Figure 1: Plot for relative bias in relation to the validation study size for CS covariance structure,  $\beta = 0.5$ .