Supplementary Information of "Evaluation of the natural history of disease by combining incident and prevalent cohorts: application to the Nun Study"

In this Supplementary Information, the estimation results of the baseline intensities and the parameters for A_0 for both the simulation studies and the data analysis and the sensitivity analysis when the density of $\widetilde{W}$ are provided.

Table S1: Simulation results for estimating the baseline intensities with incident cohort data only ($m_1 = 300$) or with the combined cohort ($m_1 = 300$ and $m_2 = 300$) under different censoring rates for the prevalent cohort. For the results of the combined cohort, the results for the parameters in $g(w)$ are also provided, where $g(w) = \eta\kappa t^{\kappa-1}e^{-\eta t^{\kappa}}$ with $\eta = 0.5$ and $\kappa = 1.2$.

	$\log\lambda_{01}$	$\log\lambda_{03}$	$\log\lambda_{10}$	$\log\lambda_{12}$	$\log\lambda_{13}$	$\log\lambda_{23}$	$\log\gamma$	$\log\eta$	$\log\kappa$
Truth	-0.693	-1.609	-0.916	-1.204	-0.916	-1.204	0.405	-0.693	0.182
Incident cohort only									
Bias	0.015	0.022	0.021	0.022	0.040	0.028	-0.005		
ESE	0.166	0.260	0.334	0.303	0.300	0.408	0.042		
SE	0.166	0.259	0.326	0.297	0.290	0.401	0.043		
CP	0.949	0.956	0.944	0.949	0.950	0.948	0.951		
Combined cohorts with no censoring									
Bias	0.014	0.021	0.023	0.021	0.034	0.016	-0.005	0.011	-0.044
ESE	0.160	0.259	0.329	0.295	0.284	0.161	0.037	0.506	0.252
SE	0.159	0.255	0.321	0.285	0.273	0.157	0.037	0.618	0.260
CP	0.950	0.956	0.944	0.950	0.950	0.944	0.950	0.979	0.965
RE	1.100	1.023	1.068	1.052	1.078	6.267	1.283		
Combined cohorts with 15% censoring									
Bias	0.000	0.020	-0.008	0.027	0.025	0.008	-0.003	0.006	-0.060
ESE	0.153	0.257	0.328	0.301	0.283	0.161	0.037	0.496	0.253
SE	0.160	0.256	0.321	0.285	0.274	0.162	0.038	0.616	0.261
CP	0.954	0.946	0.937	0.937	0.948	0.943	0.949	0.982	0.958
RE	1.196	1.041	1.041	1.023	1.094	6.399	1.298		
Combined cohorts with 30% censoring									
Bias	0.000	0.017	0.004	0.012	0.027	0.013	-0.005	0.031	-0.056
ESE	0.160	0.245	0.316	0.289	0.271	0.165	0.036	0.556	0.246
SE	0.159	0.255	0.320	0.283	0.274	0.169	0.038	0.676	0.261
CP	0.945	0.954	0.947	0.945	0.949	0.955	0.966	0.976	0.970
RE	1.108	1.118	1.113	1.066	1.156	5.767	1.273		

Table S2: Simulation results for estimating the baseline intensities with incident cohort data only ($m_1 = 600$) or with the combined cohort ($m_1 = 600$ and $m_2 = 600$) under different censoring rates for the prevalent cohort. For the results of the combined cohort, the results for the parameters in $g(w)$ are also provided, where $g(w) = \eta\kappa t^{\kappa-1}e^{-\eta t^{\kappa}}$ with $\eta = 0.5$ and $\kappa = 1.2$.

	$\log\lambda_{01}$	$\log\lambda_{03}$	$\log\lambda_{10}$	$\log\lambda_{12}$	$\log\lambda_{13}$	$\log\lambda_{23}$	$\log\gamma$	$\log\eta$	$\log\kappa$
Truth	-0.693	-1.609	-0.916	-1.204	-0.916	-1.204	0.405	-0.693	0.182
Incident cohort only									
Bias	0.009	0.014	0.016	0.011	0.019	0.025	-0.004		
ESE	0.117	0.179	0.226	0.206	0.198	0.285	0.029		
SE	0.117	0.182	0.228	0.208	0.202	0.275	0.031		
CP	0.954	0.953	0.949	0.957	0.949	0.957	0.955		
Combined cohorts with no censoring									
Bias	0.009	0.014	0.016	0.010	0.016	0.010	-0.004	0.043	-0.011
ESE	0.111	0.176	0.225	0.197	0.191	0.110	0.026	0.371	0.171
SE	0.112	0.178	0.225	0.199	0.191	0.111	0.027	0.418	0.182
CP	0.954	0.957	0.950	0.957	0.946	0.952	0.953	0.975	0.974
RE	1.091	1.034	1.009	1.094	1.089	6.682	1.238		
Combined cohorts with 15% censoring									
Bias	0.007	0.010	0.018	0.011	0.019	0.005	-0.004	0.060	-0.011
ESE	0.113	0.186	0.221	0.203	0.188	0.113	0.027	0.373	0.177
SE	0.112	0.178	0.225	0.199	0.190	0.115	0.027	0.423	0.182
CP	0.952	0.940	0.948	0.938	0.943	0.946	0.945	0.974	0.961
RE	1.073	0.970	1.036	1.009	1.090	6.247	1.150		
Combined cohorts with 30% censoring									
Bias	0.003	0.021	0.015	0.010	0.005	0.004	-0.003	0.045	-0.014
ESE	0.116	0.182	0.233	0.195	0.190	0.125	0.028	0.395	0.181
SE	0.113	0.180	0.225	0.199	0.191	0.120	0.027	0.438	0.184
CP	0.945	0.945	0.940	0.959	0.950	0.926	0.930	0.973	0.959
RE	1.039	0.980	0.933	1.104	1.073	5.197	1.081		

Table S3: Simulation results with the combined cohort ($m_1 = 300$ and $m_2 = 300$) when the density of $\tilde{W}$ is incorrectly specified. Two misspecified cases are considered: (1) $g(w) = \eta e^{-\eta t}$ (*Exponential distribution*) and (2) $g(w) = 1/\tau$ with $\tau = 10$ (*Length-biased sampling*). The prevalent cohort data are generated with a Weibull density of $\tilde{W}$ with 30% censoring rates, where $g(w) = \eta\kappa t^{\kappa-1} e^{-\eta t^{\kappa}}$ with $\eta = 0.5$ and $\kappa = 1.2$.

	$\log\alpha_{01}$	$\log\alpha_{03}$	$\log\alpha_{10}$	$\log\alpha_{12}$	$\log\alpha_{13}$	$\log\alpha_{23}$	$\log\gamma$	$\beta_{01}^{(1)}$	$\beta_{01}^{(2)}$	$\beta_{03}^{(1)}$	$\beta_{03}^{(2)}$	$\beta_{10}^{(1)}$	$\beta_{10}^{(2)}$	$\beta_{12}^{(1)}$	$\beta_{12}^{(2)}$	$\beta_{13}^{(1)}$	$\beta_{13}^{(2)}$	$\beta_{23}^{(1)}$	$\beta_{23}^{(2)}$
TRUE	-0.693	-1.609	-0.916	-1.204	-0.916	-1.204	0.405	0.200	-0.200	0.100	-0.200	-0.200	0.200	0.200	-0.300	0.200	0.200	0.300	-0.200
Case I: the density of $\tilde{W}$ is misspecified (Exponential distribution)																			
Bias	-0.018	-0.050	-0.066	-0.063	-0.061	-0.074	0.027	-0.001	-0.012	0.001	0.046	-0.006	-0.015	-0.014	-0.002	-0.017	0.011	0.007	0.007
ESE	0.184	0.289	0.401	0.342	0.326	0.166	0.035	0.171	0.687	0.285	1.182	0.401	1.588	0.311	1.331	0.299	1.27	0.127	0.526
SE	0.177	0.284	0.383	0.334	0.315	0.167	0.035	0.164	0.698	0.277	1.139	0.380	1.558	0.308	1.319	0.282	1.205	0.126	0.536
CP	0.939	0.943	0.948	0.944	0.942	0.930	0.851	0.939	0.951	0.954	0.944	0.949	0.960	0.959	0.949	0.943	0.943	0.950	0.962
Case II: the density of $\tilde{W}$ is misspecified (Length-biased sampling)																			
Bias	0.025	-0.006	-0.074	0.025	0.033	-0.007	0.033	-0.005	-0.016	-0.013	0.041	-0.005	-0.012	-0.034	-0.011	-0.040	0.000	-0.015	0.034
ESE	0.180	0.291	0.406	0.337	0.321	0.152	0.035	0.171	0.684	0.290	1.199	0.404	1.606	0.310	1.332	0.295	1.264	0.110	0.451
SE	0.175	0.288	0.387	0.330	0.308	0.154	0.035	0.163	0.695	0.282	1.160	0.384	1.574	0.307	1.319	0.280	1.198	0.110	0.465
CP	0.941	0.947	0.950	0.942	0.935	0.954	0.822	0.933	0.949	0.960	0.944	0.951	0.959	0.959	0.948	0.943	0.941	0.950	0.959

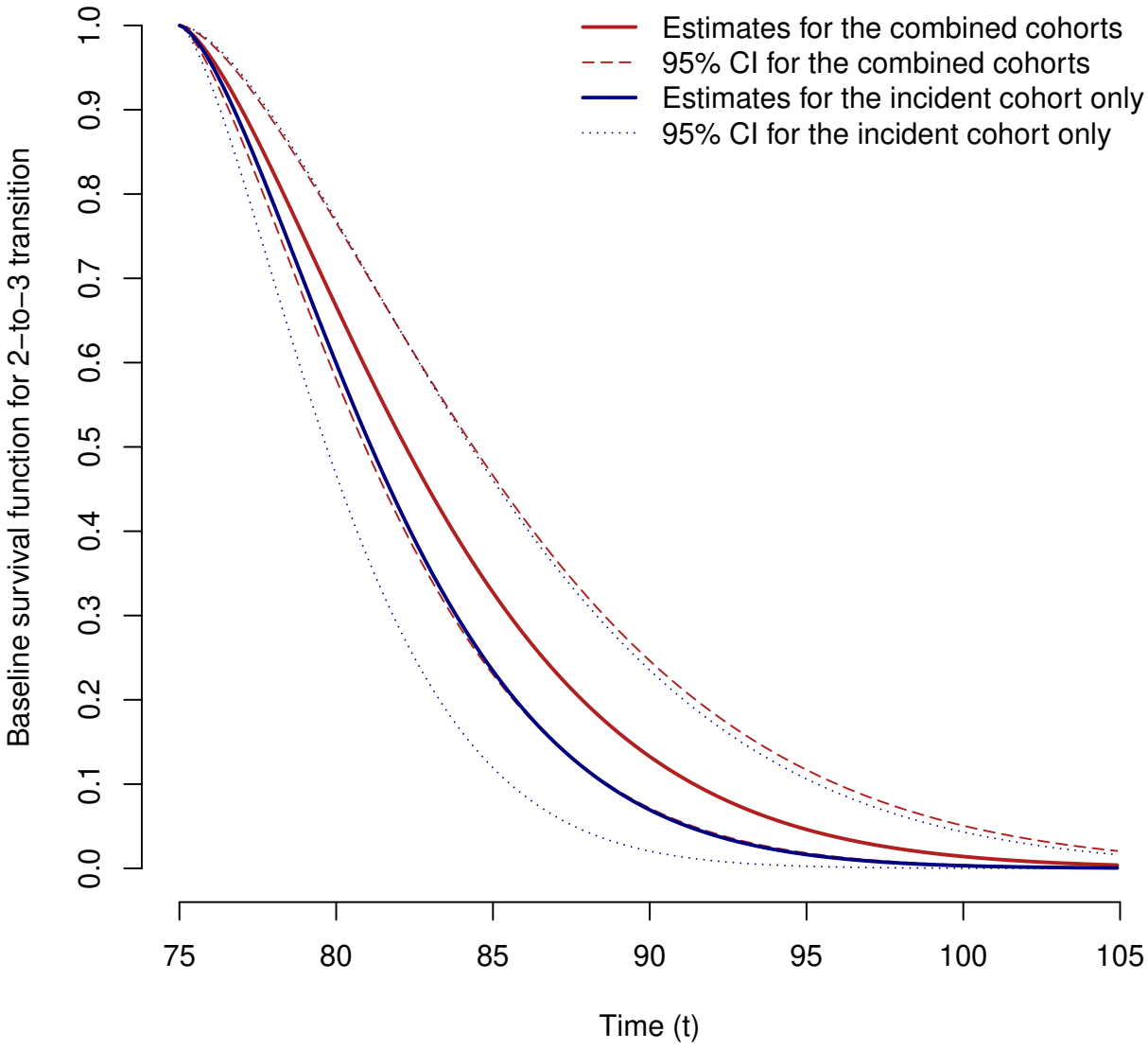

Figure S1: Flow diagram of the Nun Study that consists of both incident cohort data ($m_1 = 424$) and prevalent cohort data ($m_2 = 77$).

Table S4: Estimates of the baseline intensities and the parameters for A_0 in the Nun Study by using the combined cohorts or the incident cohort only.

	$\log \lambda_{01}$	$\log \lambda_{03}$	$\log \lambda_{10}$	$\log \lambda_{12}$	$\log \lambda_{13}$	$\log \lambda_{23}$	$\log \gamma$	$\log \eta$
Combined cohorts data								
Estimate	-1.571	-6.509	-3.907	-4.505	-5.054	-3.256	0.380	0.284
SE	0.311	0.538	0.434	0.263	0.354	0.222	0.033	0.334
Incident cohort data only								
Estimate	-1.672	-6.603	-4.010	-4.574	-5.265	-3.087	0.407	
SE	0.324	0.530	0.448	0.282	0.404	0.292	0.036	

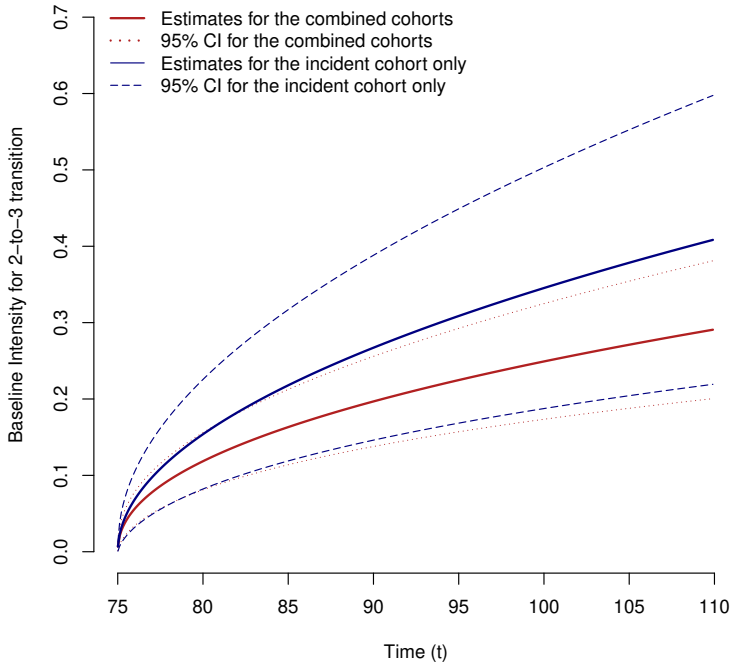

Figure S2: The estimated baseline intensities for the 2-3 transitions with the pointwise 95% confidence intervals, obtained from the combined cohort data (red) and the incident cohort data (blue) of the Nun Study.