Supplementary materials for "Analysis of gap times based on panel count data with informative observation times and unknown start time"

S1 Additional simulation results

We present in the following subsections the simulation studies, 1) under Scenario I with known time zero; 2) under Scenario II with half-Normal admission time; 3) under Scenario II with fewer events and observations; 4) under Scenario II for sensitivity on model misspecification of admission time distribution.

S1.1 Scenario I with Known Time Zero

For simulation situation (1) without unmeasured admission time, repeated events with intermittent examination times were generated similar to the dilation data but with the assumption that everybody enters observation period at the start of the event process, in other words with no unknown admission time. Specifically, assume that $N(\cdot)$ is a nonhomogeneous Poisson process with $\lambda_0(t) = (t+1)/(\tau/2+1)$, $t \in [0, \tau]$ where $\tau = 15$. Then, given

x and z, the total number of events J occurred by time τ follows the Poisson distribution with mean

$$\Lambda(\tau|x,z) = z\Lambda_0(\tau)\exp(x\beta) = \frac{z(\tau^2/2+\tau)\exp(x\beta)}{\tau/2+1} = z\tau\exp(x\beta),$$

and the event times are the order statistics of a random sample of size J from the distribution with cumulative distribution function

$$\frac{t^2/2+t}{\tau^2/2+\tau} I(0 \le t \le \tau).$$

The observation process is then assumed to be a homogeneous Poisson process with $\lambda_{H0}(t) = 1.0$ for $t \in [0,\tau]$. Then, given x and z, K follows a Poisson distribution with mean

$$\Lambda_H(\tau|x,z) = z\Lambda_{H0}(\tau)\exp(x\gamma) = z\tau\exp(x\gamma),$$

and $(t_1,\ldots,t_K)$ are the order statistics of a random sample of size K from the uniform distribution over $(0,\tau)$. The true values of covariate effect were set to be $\beta = 1.5$ and $\gamma = 2.0$ for $n = 250$ and $n = 500$.

The simulation results of the regression coefficients associated with event process are given in Table S1. The survival times of the first four gap times at the 25th, 50th, and 75th percentiles are shown in Table S2. Results shown in the tables include the estimation bias (Bias) given by the average of the estimates minus the true value, the sample standard deviation of the estimates (SSD), the mean of the bootstrap standard deviation (BSD) and the 95% bootstrap empirical coverage probability (CP). As can be seen in Tables S1–S2, the performance of the estimates are very good with coverage probabilities close to the nominal level of 95%. The bandwidth used for kernel smoothing of f_z was about 0.18 for $n = 250$ and about 0.14 for $n = 500$.

Table S1: Estimation results of regression coefficients under Scenario I with no unmeasured admission times.

		$n = 250$				$n = 500$			
Par	True	Bias	SSD	BSD	CP	Bias	SSD	BSD	CP
β	1.5	-0.004	0.168	0.169	0.944	-0.012	0.124	0.120	0.941
γ	2.0	-0.005	0.176	0.176	0.952	-0.014	0.131	0.125	0.932

S1.2 Scenario II with Half-Normal Admission Time

We also performed simulation studies with half-normally distributed admission times. Table S3 presents the estimation results of the covariate effects under both Scenario I and Scenario II in presence of unmeasured admission times. The bandwidth for kernel smoothing of f_z was about 0.18 for $n = 250$ and 0.13 for $n = 500$ under Scenario I, and was about 0.18 for $n = 250$ and 0.15 for $n = 500$ under Scenario II. The estimation results of the marginal survival functions of the third and fourth gap times under Scenario I, and the first four gap times under Scenario II with unmeasured admission times are in Table S4.

S1.3 Scenario II with Fewer Events and Observations

We conducted additional simulation studies under Scenario II with Gamma admission time, but reduce the number of events and observations so that the median number of events is 5 and the median number of observations is 6, about half of the numbers used in Section 4.1 in the paper. The estimation results with sample size of 250 of regression parameters and gap time survival functions are shown in Table S5 and Table S6, respectively. The estimation

Table S2: Estimation results of marginal survival functions for the first four gap times under Scenario I at the 25th, 50th and 75th percentiles with no unmeasured admission times.

			$n = 250$				$n = 500$			
	p	w	Bias	SSD	BSD	CP	Bias	SSD	BSD	CP
	0.25	0.86	−0.002	0.022	0.022	0.941	−0.002	0.016	0.015	0.934
$S_{M_1}(w)$	0.50	1.76	−0.004	0.023	0.022	0.952	−0.002	0.016	0.016	0.930
	0.75	3.13	−0.003	0.017	0.017	0.937	0.000	0.012	0.012	0.952
	0.25	0.41	−0.004	0.012	0.012	0.937	−0.002	0.009	0.008	0.936
$S_{M_2}(w)$	0.50	1.01	−0.005	0.019	0.018	0.937	−0.003	0.013	0.013	0.945
	0.75	2.10	−0.003	0.017	0.016	0.933	0.000	0.012	0.012	0.952
	0.25	0.31	−0.004	0.013	0.013	0.933	−0.003	0.010	0.009	0.941
$S_{M_3}(w)$	0.50	0.79	−0.006	0.019	0.018	0.930	−0.004	0.013	0.013	0.942
	0.75	1.72	−0.003	0.016	0.015	0.919	−0.000	0.011	0.011	0.949
	0.25	0.26	−0.004	0.011	0.011	0.932	−0.003	0.008	0.008	0.935
$S_{M_4}(w)$	0.50	0.67	−0.006	0.017	0.016	0.924	−0.003	0.012	0.011	0.941
	0.75	1.50	−0.003	0.015	0.015	0.925	0.000	0.011	0.011	0.946

of shape parameter for the admission time seem slightly sensitive but final estimation of the gap time survival functions seem reasonably robust.

Table S3: Estimation results for regression coefficients in presence of unmeasured half-normal admission times.

(a) Scenario I: admitted with 2 events									
		n=250				n=500			
Par	True	Bias	SSD	BSD	CP	Bias	SSD	BSD	CP
β	1.5	−0.641	0.161	0.159	0.031	−0.636	0.114	0.113	0.001
γ	1.0	−0.631	0.164	0.158	0.040	−0.629	0.114	0.112	0.001
(b) Scenario II									
		n=250				n=500			
Par	True	Bias	SSD	BSD	CP	Bias	SSD	BSD	CP
β	1.5	−0.020	0.185	0.180	0.940	−0.025	0.132	0.129	0.931
γ	1.0	0.007	0.168	0.169	0.950	−0.002	0.123	0.120	0.938
η	0.9	0.078	0.170	0.156	0.909	0.078	0.127	0.115	0.905

Table S4: Estimation results of marginal survival functions for the third and fourth gap times under Scenario I and the first four gap times under Scenario II, at 25th, 50th and 75th percentiles in presence of unmeasured half-normal admission times.

(a) Scenario I: admitted with 2 events

			$n = 250$				$n = 500$			
	p	w	Bias	SSD	BSD	CP	Bias	SSD	BSD	CP
$S_{M_3}(w)$	0.25	0.18	−0.011	0.017	0.017	0.908	−0.010	0.012	0.012	0.853
	0.50	0.47	−0.021	0.022	0.022	0.837	−0.020	0.016	0.016	0.749
	0.75	1.12	−0.026	0.015	0.015	0.608	−0.025	0.011	0.011	0.412
$S_{M_4}(w)$	0.25	0.17	−0.010	0.012	0.012	0.870	−0.009	0.009	0.008	0.802
	0.50	0.46	−0.020	0.016	0.017	0.774	−0.019	0.012	0.012	0.621
	0.75	1.11	−0.025	0.013	0.014	0.554	−0.024	0.010	0.010	0.322

(b) Scenario II

			$n = 250$				$n = 500$			
	p	w	Bias	SSD	BSD	CP	Bias	SSD	BSD	CP
$S_{M_1}(w)$	0.25	0.18	−0.006	0.027	0.026	0.926	−0.005	0.020	0.019	0.920
	0.50	0.49	−0.005	0.036	0.036	0.937	−0.002	0.027	0.025	0.931
	0.75	1.16	0.003	0.027	0.026	0.947	0.006	0.020	0.019	0.948
$S_{M_2}(w)$	0.25	0.18	−0.005	0.021	0.021	0.926	−0.004	0.016	0.015	0.919
	0.50	0.48	−0.004	0.029	0.028	0.931	−0.002	0.021	0.020	0.937
	0.75	1.14	0.004	0.021	0.021	0.941	0.006	0.016	0.015	0.929
$S_{M_3}(w)$	0.25	0.18	−0.005	0.017	0.016	0.920	−0.004	0.012	0.011	0.925
	0.50	0.47	−0.003	0.023	0.022	0.933	−0.002	0.016	0.015	0.938
	0.75	1.12	0.004	0.018	0.018	0.946	0.006	0.013	0.013	0.921
$S_{M_4}(w)$	0.25	0.17	−0.004	0.014	0.013	0.920	−0.003	0.010	0.009	0.928
	0.50	0.46	−0.002	0.020	0.019	0.934	−0.001	0.014	0.013	0.941
	0.75	1.11	0.006	0.018	0.018	0.939	0.007	0.013	0.013	0.913

Table S5: Estimation results for regression coefficients in presence of unmeasured Gamma admission times with relatively small numbers of events and observations.

		n=250			
Par	True	Bias	SSD	BSD	CP
β	1.5	0.049	0.216	0.217	0.946
γ	1.0	0.009	0.184	0.179	0.944
η_1	1.0	0.111	0.171	0.174	0.897
η_2	-2.0	0.016	0.324	0.312	0.935

S1.4 Sensitivity on Model Misspecification of Admission Time

To check the sensitivity on model misspecification of distribution of admission time A, we performed the following simulation studies. We generated A from a mixed distribution of Gamma and half-Normal

$$(1-\delta) \times \text{Gamma}\big\{ \exp(\eta_1), \exp(x\eta_2)\big\} + \delta \times \text{half-Normal}(0, \sigma^2),$$

where $P(\delta = 1) = 0.1$, but analyze the data assuming A follows $\text{Gamma}\big\{ \exp(\eta_1), \exp(x\eta_2)\big\}$. The estimation results of regression parameters and gap time survival functions are shown in Table S7 and Table S8, respectively. The estimation of scale parameter for the admission time seem sensitive to model misspecification but the final estimation of gap time distributions seem reasonably robust.

Table S6: Estimation results of marginal survival functions for the first four gap times under Scenario II, at 25th, 50th and 75th percentiles in presence of unmeasured Gamma admission times with relatively small numbers of events and observations.

	$n = 250$					
	p	w	Bias	SSD	BSD	CP
$S_{M_1}(w)$	0.25	0.24	0.020	0.063	0.057	0.924
	0.50	0.65	0.031	0.075	0.069	0.914
	0.75	1.54	0.024	0.035	0.035	0.925
$S_{M_2}(w)$	0.25	0.24	0.005	0.026	0.025	0.956
	0.50	0.63	0.007	0.031	0.032	0.974
	0.75	1.50	0.014	0.021	0.022	0.939
$S_{M_3}(w)$	0.25	0.23	−0.004	0.023	0.024	0.950
	0.50	0.62	−0.000	0.024	0.026	0.960
	0.75	1.48	0.012	0.020	0.020	0.913
$S_{M_4}(w)$	0.25	0.23	−0.003	0.021	0.021	0.942
	0.50	0.61	0.002	0.033	0.034	0.947
	0.75	1.45	0.013	0.020	0.020	0.910

Table S7: Estimation results for regression coefficients when distribution of admission time is misspecified.

		n=250			
Par	True	Bias	SSD	BSD	CP
β	1.5	0.017	0.178	0.177	0.949
γ	1.0	0.005	0.166	0.169	0.946
η_1	0.5	−0.005	0.174	0.174	0.943
η_2	−1.0	0.260	0.263	0.251	0.809
σ	0.9	–	–	–	–

S2 More about CPP and CSL data analysis

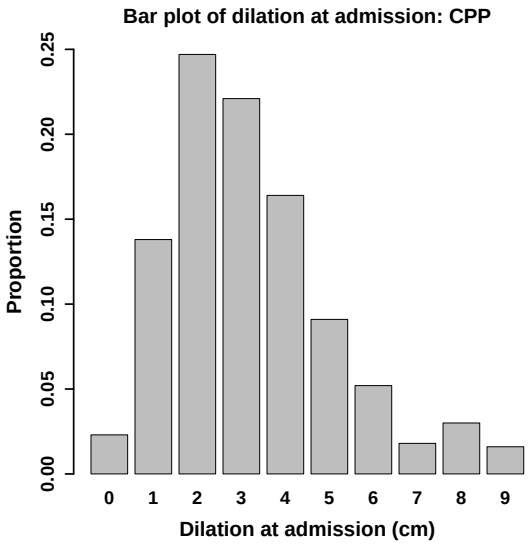
Figure S1 presents the bar graphs of distribution of the dilation at admission as well as the dilations across all observations.

S2.1 Adjusting for Dilation at Admission in Admission Time Distribution

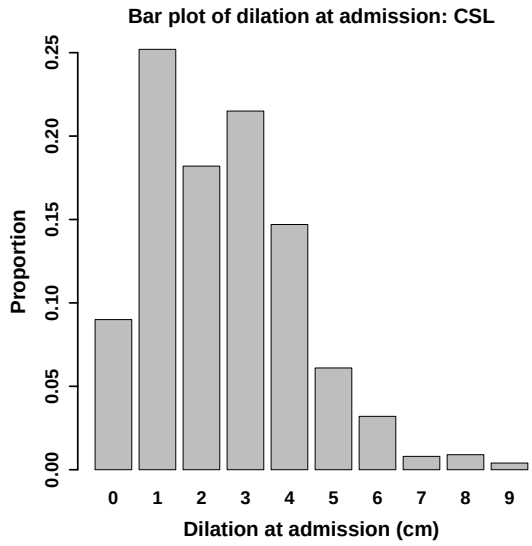
One of the reviewers suggested that we included the observed dilation at admission, m_0 as a predictor for admission time A. In this subsection, we present the analysis results which adjusts for m_0 in the distribution of A. The estimation results of regression parameters for dilation and observation processes, and for admission time are summarized in Table S9 and Table S10 below, respectively. Age effect remains significant with consistent directions and BMI effect remains non-significant, in dilation process, observation process and admission

Table S8: Estimation results of marginal survival functions for the first four gap times under Scenario II, at 25th, 50th and 75th percentiles when distribution of admission time is misspecified.

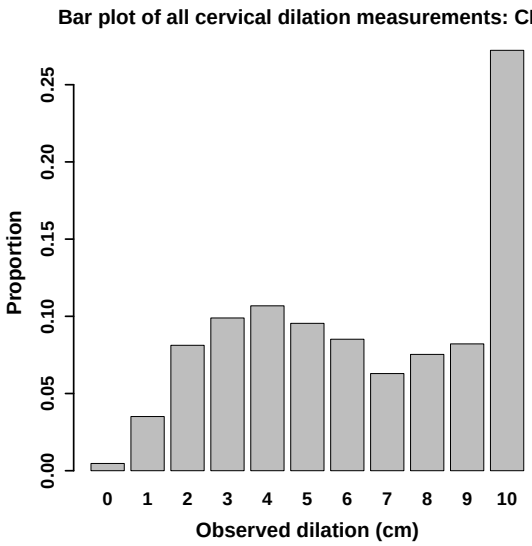
	$n = 250$					
	p	w	Bias	SSD	BSD	CP
$S_{M_1}(w)$	0.25	0.18	0.010	0.044	0.043	0.941
	0.50	0.49	0.018	0.062	0.063	0.947
	0.75	1.16	0.019	0.043	0.045	0.946
$S_{M_2}(w)$	0.25	0.18	0.006	0.031	0.031	0.939
	0.50	0.48	0.010	0.041	0.042	0.940
	0.75	1.14	0.012	0.027	0.028	0.950
$S_{M_3}(w)$	0.25	0.18	0.002	0.020	0.020	0.953
	0.50	0.47	0.004	0.025	0.026	0.959
	0.75	1.12	0.007	0.018	0.018	0.952
$S_{M_4}(w)$	0.25	0.17	−0.002	0.013	0.013	0.955
	0.50	0.46	−0.002	0.019	0.019	0.962
	0.75	1.11	0.004	0.018	0.018	0.943



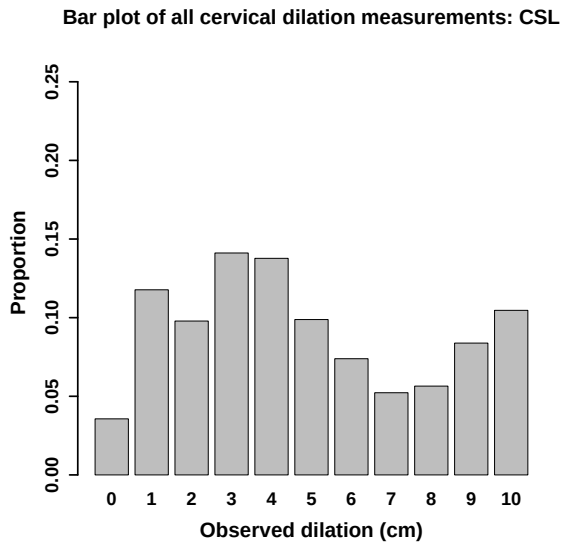
(a)



(b)



(c)



(d)

Figure S1: Bar graphs of cervical dilation at admission to hospital: (a) and (b), and of all cervical dilation measurements: (c) and (d), for subsamples of size 1000 from CPP and CSL data.

Table S9: Covariate effect estimates and 95% confidence intervals for dilation and observation processes under Scenario II adjusting for dilation at admission for admission time distribution.

	Parameter	Dilation Process	Observation Process
CPP	AGE	−0.85 (−1.65, −0.05)	0.56 (0.16, 0.96)
	UW	−0.07 (−0.37, 0.24)	−0.09 (−0.21, 0.03)
	OW	−0.18 (−0.56, 0.19)	0.03 (−0.13, 0.18)
	OB	0.18 (−0.41, 0.77)	−0.11 (−0.32, 0.11)
CSL	AGE	0.24 (−0.30, 0.77)	−0.08 (−0.34, 0.18)
	OW	0.14 (−0.08, 0.36)	0.06 (−0.03, 0.14)
	OB1	−0.14 (−0.40, 0.12)	0.12 (0.03, 0.21)
	OB2	−0.39 (−0.68, −0.10)	0.25 (0.11, 0.38)
	OB3	−0.64 (−0.97, −0.31)	0.28 (0.15, 0.40)

time for the CPP data. In the CSL data, age effects on the dilation process and admission time disappear and OB1 women no longer have significantly longer dilation than normal-weighted women. The observed dilation at admission ,m_0, was found to have significant positive effect on time to admission for both the CPP and CSL datasets.

The estimated density curves of admission time and survival curves of gap times for the two datasets are given in the following Figure S2 and Figure S3, respectively. Figure S2 indicates that women got admitted earlier in the labor process in CSL than in CPP which is consistent to our previous finding without adjusting for m_0 in admission time. The overall trend in Figure S3 is also consistent with the trend in Figure 2 of Section 4.2 which

Table S10: Covariate effect estimates and 95% confidence intervals for admission time distribution adjusting for dilation at admission.

Parameter	CPP	Parameter	CSL
Shape: η_1	0.68 (0.40, 0.96)	Shape: η_1	0.95 (0.63, 1.26)
AGE: η_2	0.78 (0.10, 1.46)	AGE: η_2	-0.54 (-1.23, 0.15)
UW: η_3	0.04 (-0.26, 0.35)	OW: η_3	-0.22 (-0.62, 0.18)
OW: η_4	0.17 (-0.16,0.49)	OB1: η_4	0.02 (-0.42, 0.45)
OB: η_5	-0.22 (-0.75, 0.31)	OB2: η_5	0.04 (-0.46, 0.54)
m_0: η_6	0.15 (0.10, 0.21)	OB3: η_6	0.29 (-0.23, 0.80)
		m_0: η_7	0.34 (0.15, 0.54)

didn't adjusting for m_0 in admission time. Overall, the analysis results are consistent with our previous analysis which didn't adjust for dilation at admission in admission time distribution.

S2.2 Graphical Model-Checking for Admission Time

We also performed overall model checking of the distribution of admission time using graphical exploratory diagnostic tool by Verbeke and Molenberghs (2013). Specifically, the fit of a specific distribution function $\hat{F}_A$ can be checked graphically by inspecting its gradient function $\Delta(\hat{F}_A, A)$, where

$$\Delta(\hat{F}_A, A) = \frac{1}{n}\sum_{i=1}^{n} \frac{f_i(\tilde{\boldsymbol{O}}_i|A)}{\int f_i(\tilde{\boldsymbol{O}}_i|A)d\hat{F}_A(A)},$$

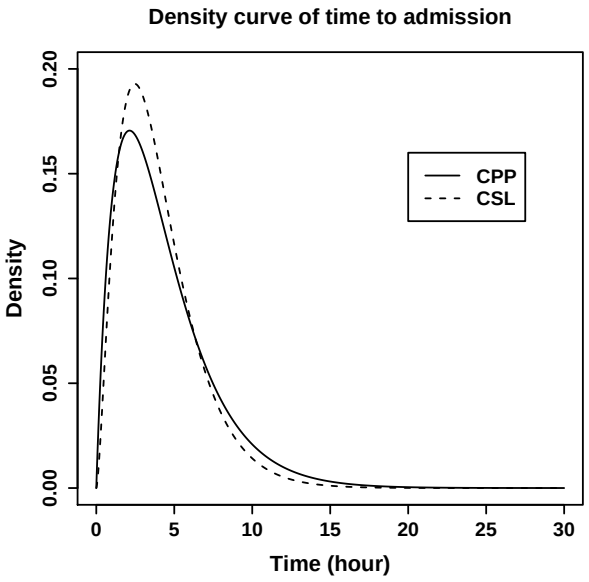


Figure S2: Estimated density curve of time to admission adjusting for dilation at admission for CPP and CSL data.

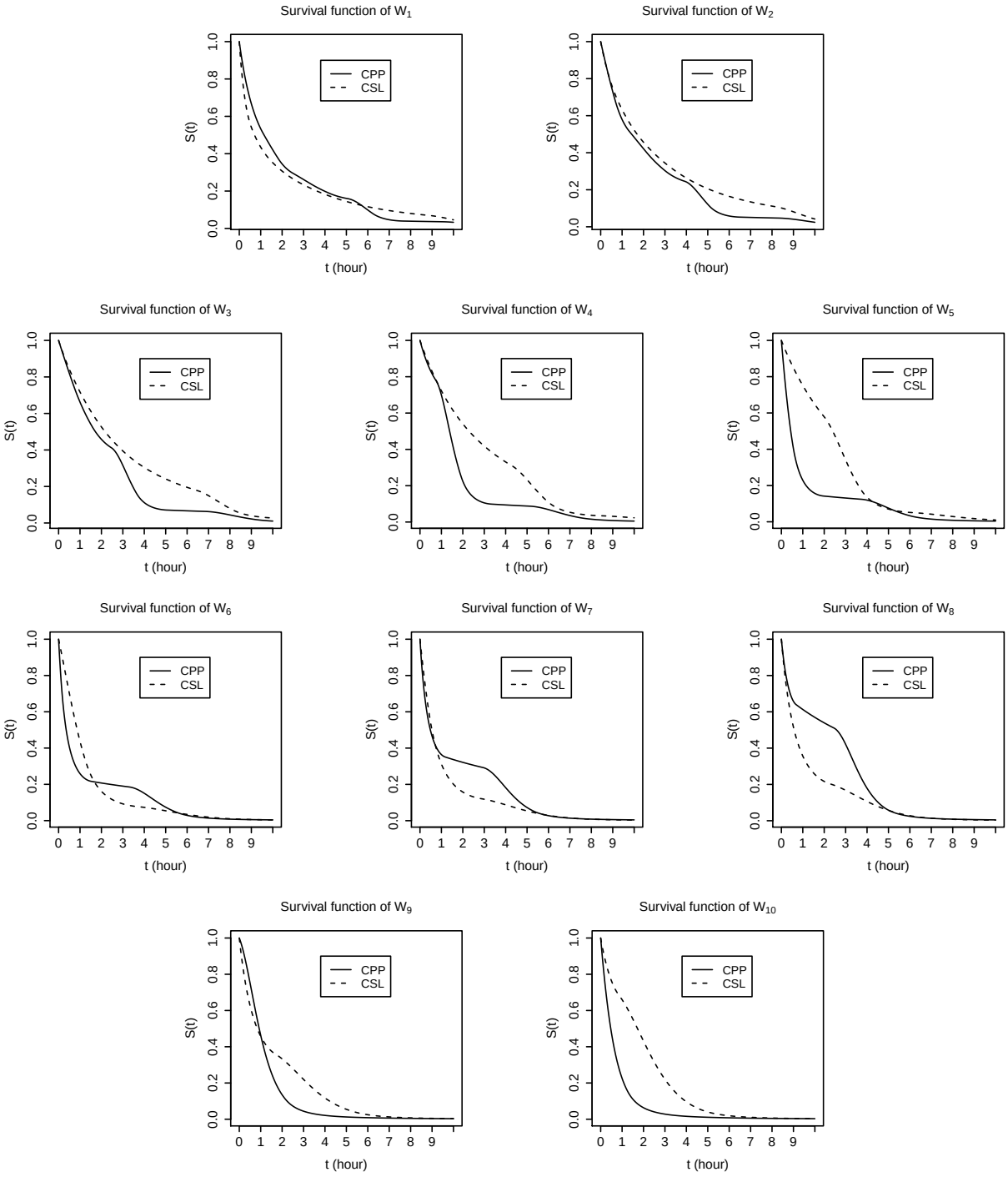


Figure S3: Estimated survival curves of gap times for CPP data (solid line) and CSL data (dashed line) after adjusting for dilation at admission in admission time distribution.

where

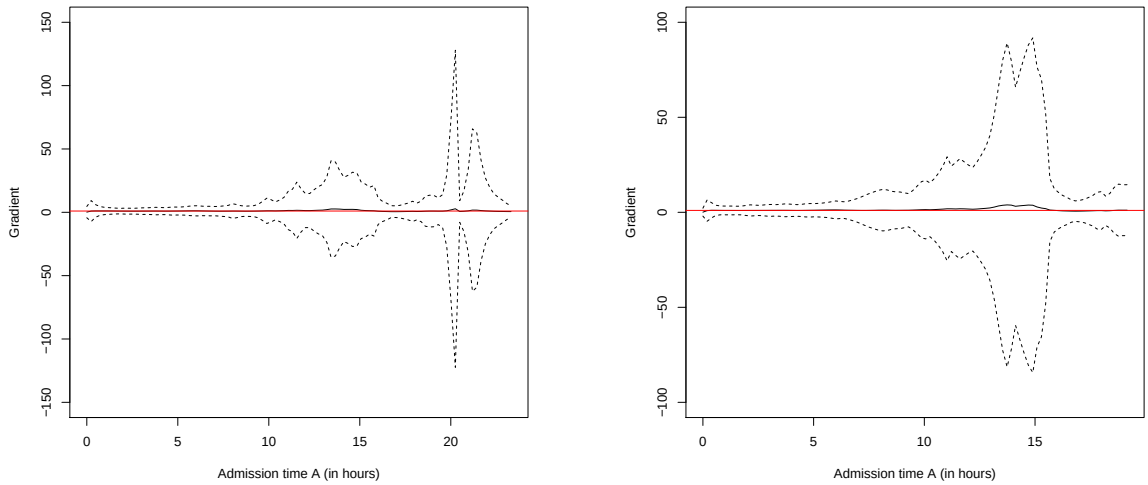
$$f_i(\tilde{\boldsymbol{O}}_i|A) = \prod_{j=1}^{K_i} \left\{ \frac{\Lambda_0(\tilde{t}_{ij-1} + A) - \Lambda_0(\tilde{t}_{ij-2} + A)}{\Lambda_0(\tilde{t}_{i\tilde{K}_i} + A)} \right\}^{m_{ij}}.$$

If no other distribution can yield a substantially better fit than the fitted distribution $\hat{F}_A$, we expect the gradient function $\Delta(\hat{F}_A, A)$ to be close to 1. In order to distinguish true deviations from 1, we also constructed a pointwise confidence band around $\Delta(\hat{F}_A, A)$ following Verbeke and Molenberghs (2013). Figures S4(a) and S4(b) give the gradient function with 95% pointwise confidence bands for the CPP data with Gamma admission time and half-Normal admission time, respectively. Both confidence bands cover 1 throughout the support for admission time, so that graphically we can not reject either the Gamma distribution or the half-Normal distribution. Figures S4(c) and S4(d) compare the Gamma distribution with the Half-normal distribution and we found that the gradient function of Gamma distribution is overall closer to 1. This indicates that Gamma distribution could be a better fit than Half-Normal for the admission time of CPP data.

Similarly for the CSL data, the gradient function in Figure S5 is close to 1, indicating that the Gamma distribution seem to be a good fit for the data.

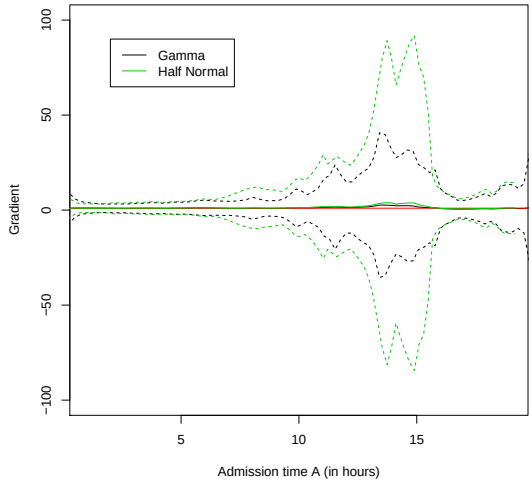
References

Verbeke, G. and Molenberghs, G. (2013). The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. *Biostatistics*, 14(3):477–490.

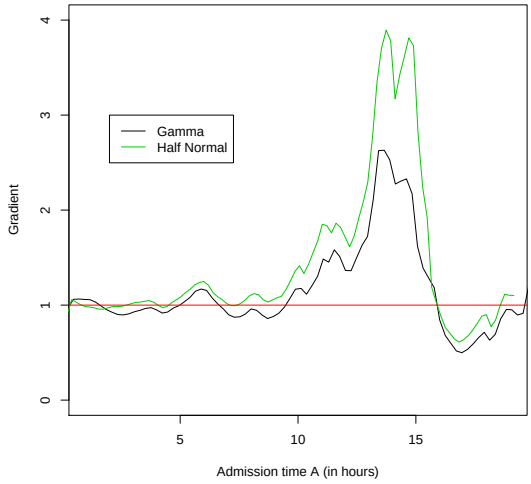


(a)

(b)

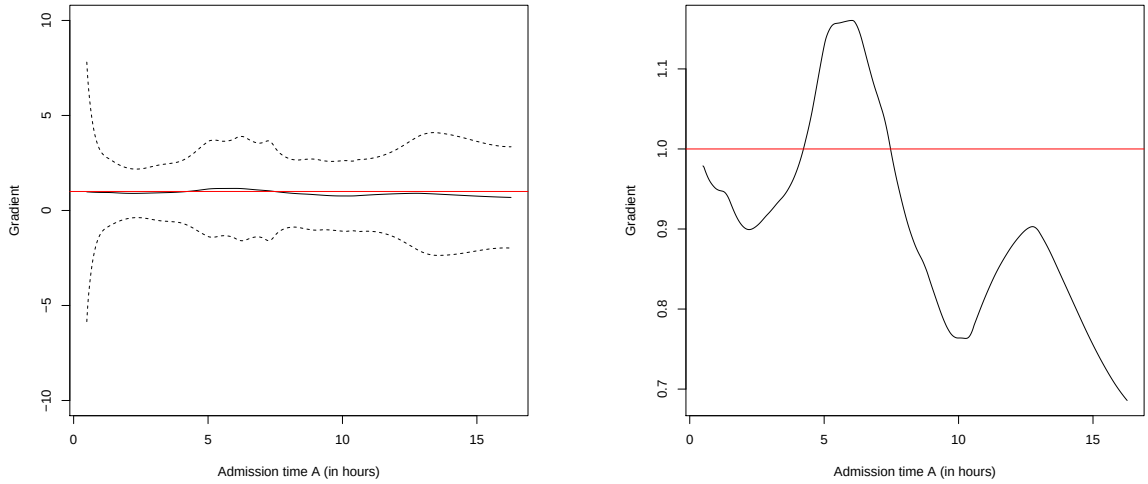


(c)



(d)

Figure S4: Gradient function and 95% pointwise confidence bands for CPP data with (a): Gamma, (b) half Normal, (c) Gamma vs. half Normal , (d) Gamma vs. Half Normal (estimate only), admission time with red line denoting value 1.



(a) (b)

Figure S5: Gradient function for CSL data with Gamma admission time (a) with 95% pointwise confidence bands, (b) estimate only.