Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Self-reporting and screening: Data with right-censored, left-censored, and complete observations

Jonathan Yefenof*[1] | Yair Goldberg[2] | Jennifer Wiler[3] | Avishai Mandelbaum[2] | Ya'acov Ritov[1,4]

[1]Statistics and Data Science, The Hebrew University of Jerusalem, Israel
[2]The Faculty of Industrial Engineering and Management, Technion - Israel Institute of Technology, Israel
[3]School of Medicine, University of Colorado, Colorado, U.S.
[4]Department of Statistics, University of Michigan, Michigan, U.S.

Correspondence
*Corresponding author name, This is sample corresponding address. Email: jonathan.yefenof@mail.huji.ac.il

Present Address
Department of Statistics, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel

Summary

We consider survival data that combine three types of observations: uncensored, right-censored, and left-censored. Such data arises from screening a medical condition, in situations where self-detection arises naturally. Our goal is to estimate the failure-time distribution, based on these three observation types. We propose a novel methodology for distribution estimation using both semiparametric and nonparametric techniques. We then evaluate the performance of these estimators via simulated data. Finally, as a case study, we estimate the patience of patients who arrive at an emergency department and wait for treatment. Three categories of patients are observed: those who leave the system and announce it, and thus their patience time is observed; those who get service and thus their patience time is right-censored by the waiting time; and those who leave the system without announcing it. For this third category, the patients' absence is revealed only when they are called to service, which is after they have already left; formally, their patience time is left-censored. Other applications of our proposed methodology are discussed.

KEYWORDS:
Survival analysis, Left censoring, Right censoring, Current status data, Nonparametric estimation

1 | INTRODUCTION

We study the estimation of failure time distribution where the failure times can be either observed directly, or be right-censored or left-censored. This type of survival data arises, for example, in estimation of time to the appearance of a medical condition where characteristic symptoms may or may not appear when the condition exists. Specific medical settings include relapse in childhood brain tumors, which may be observed due to clinical symptoms, or right-censored due to periodic screening with negative result (no tumor), or left-censored due to periodic screening with a positive result.[1] Another medical setting is melanoma cancer, which is observed if self-detected, or is right censored due to a negative screening (no melanoma), or left-censored if it goes undetected until screening. Additional examples can be found in Whitehead.[2]

The motivating example for this work comes from estimating customer patience in service system which is a challenging problem.[3] In our study, we focus on patients who wait for treatment in an emergency department (ED). Three categories of patients are observed. The first category consists of patients who get service and thus their patience time is right-censored by the waiting time. The second category comprises those who leave the system and announce it, and thus their patience time is observed while the waiting time is right-censored. The third category consists of patients who leave the system without announcing it; their absence is hence revealed only when they are called to service, which is after they have already left; formally, their patience time is left-censored. Note that the data structure is a special case of interval-censored data.[4] Here, interval-censored data is a

general data structure which many popular survival data settings are special cases of, including both right-censored data and left-censored data.[5] The specific setting that is considered here includes both left-censored and right-censored observations as well as complete observations.

Estimating the patience time is of importance as the decision of patients to leave the system before getting served might have a strong effect on their physical well-being. There has been considerable research on the reasons why patients leave an ED before being served,[6–9]. However, these and other authors have not proposed a model by which ED patience time - namely the duration that a potential patient is willing to wait for ED service - can be estimated, and this is our goal here.

We propose novel semiparametric and nonparametric estimators of the unknown survival function for this 3-type survival data. We then study their rates of convergence. The semiparametric estimator is based on both full and partial likelihoods. We provide condition under which the semiparametric estimator is a linear asymptotic normal (LAN) estimator and converges to a normal distribution in a root-n rate. The nonparametric estimator is based on nonparametric kernel estimators for density functions and on a novel estimator of the cumulative probability function that has some similarities to the Nelson–Aalen estimator.[10] We show that, under some regularity conditions, the nonparametric estimator point-wise converges to the normal distribution.

We perform a simulation study and compare the proposed semiparametric and nonparametric estimators. For the semiparametric model, we study both correct and misspecified models and show the different corresponding results. We show how the accuracy changes with sample size. We then carry out a case study that is based on data of patients waiting for treatment in an ED, in the U.S. in 2008. We analyzed separately different severity levels (15106 observations in the emergency group, 43600 in the urgent group, and 26541 in the semi-urgent group). We conclude with a comparison of the semiparametric and nonparametric estimators for the three different severity levels of this dataset.

2 | BRIEF LITERATURE REVIEW

Developing screening methods for medical conditions, such as breast and melanoma cancers, has a long history.[11,12] In the classical setting, the medical condition either already exists at the time of screening and is thus left-censored, or does not exist, and is thus right-censored. The setting in which self-detection is possible, and thus the condition time is observed, has been surprisingly mostly ignored in the literature. For example, Minn et al.[1] treat both self-detection times and screening times as event times, ignoring the censoring. The closest model to the one that we present here appears in Whitehead.[2] It is assumed there that the condition can be detected at screening or before screening due to symptoms. In both cases, the condition already exists at the time of detection. It is also assumed that screenings take place at a sequence of fixed time points. Whitehead[2] recommends to ignore the extra knowledge gained due to self-reporting and to replace these times with the time of the next screening. The survival function is then estimated only at the discrete fixed screening times using standard techniques.[13]

There has been considerable research effort, dedicated to modeling and analysis of customer (im)patience while waiting for service. Here we describe several papers that, together with references therein, provide what is required for a historical background and state-of-art perspective. First, we recommend the recent literature review[9] (Section 3) in Batt and Terwiesch, accompanied by Gans et al.[14] These survey patience-research from an operational/queueing view point (mainly Section 6.3.3 in the latter), while connecting it to the medical literature on patients who are left without being seen (LWBS) (mainly Section 3 in the former); see also Aksin et al.[15] who expand on managerial challenges. Next we mention Mandelbaum and Zeltyn[16], which is an Explanatory Data Analysis of (im)patience in telephone call centers (that appears in a special issue that is devoted to models of queues abandonment). Finally, and the most related to the present study, are the following two studies. Brown et al.[17] applies, in Section 5, the Kaplan–Meier estimator[18] to estimate the survival functions and consequently hazard rates, of both virtual waiting time and impatience; the data is that of a call center, in which times of abandonment are all recorded hence the data is right-censored. Then Wiler et al.,[19] which is also the source of our present ED data case study, estimate LWBS rates as a function of ED patient arrival rates, treatment times, and ED boarding times. There was no attempt in that work to estimate the patience-time distribution.

We conclude this brief survey with the observation that the estimation of customer (im)patience is relevant beyond screening, call centers, and EDs. For example, Nah[20] studies tolerance of Web users (during information retrieval). Yom-Tov et al.[21] analyzes chat services, in which customers abandon at any phase during chat-exchanges with a service center: one expects that such services give rise to the same options as in EDs: some customers receive service, others abandon without letting anyone know, and the rest announce their abandonment time.

3 | THE MODEL

In the standard setting of right-censored data one observes, for each patient, either the failure time or the censoring time. In terms of our motivating example, failure time is patience time while censoring time is the waiting time. Patience time is observed when patients leave the ED while informing the system of their departure; waiting time is observed when a patient is called for service. However, unlike in standard right-censored data and like in current status data, there are also patients who leave without informing; in this case their absence is observed only when they are called for service, and this latter time provides an upper bound for their patience time. In other words, the (virtual) waiting time is observed, and the only information on patience time is that it is less than this observed waiting time. Hence, in this case, the patience time is left-censored.

More formally, let T be the patient's failure time, i.e., the time until the patient loses patience. Let W be the censoring time, i.e., the waiting time until the patient gets (or could have gotten) service. We assume that T has a cumulative distribution function (cdf) F and a probability density function (pdf) f, and that W has cdf G and pdf g. Let Δ be the indicator $\Delta \equiv 1\{T < W\}$; i.e., $\Delta = 1$ if the patient loses patience before being called to service, and $\Delta = 0$ otherwise.

Let Y be the indicator that is 1 for a patient who leaves and informs when leaving, and 0 otherwise. Denote by $q(t)$ the conditional probability that a patient reports leaving given that the waiting time equals to t. In other words, $q(t) = pr(Y = 1 \mid T = t)$. The assumption that patience time T and waiting time W are independent is common in survival analysis, for example, when using the Kaplan-Meier estimator.[18] Since T and W may be dependent, one can use strata to overcome this challenge as was done in the case study in Section 7. The announcement indicator Y depends on the time through the function $q(t)$. In other words, given the patience time $T = t$, the decision on announcement does not depend on actual waiting time W. However, due to censoring, the decision on the announcement is observed only when $t < W$. Summarizing, we assume that the pair (Y, T) is independent of the waiting time W. When this assumption does not hold, different theoretical tools are needed for a valid estimation.

Let U be the recorded time: $U \equiv YT + (1-Y)W$. The observed data consist of the triplets (U_i, Y_i, Δ_i), $i = 1, \ldots, n$, and there are three categories of patients:

$C = 1$**:** The patient gets service, hence the waiting time is observed, which serves as a lower bound on the patience time; thus the patience time is right censored. Formally, $\Delta = 0$, $Y = 0$, and $U = W$.

$C = 2$**:** The patient leaves without being treated and reports departure. The patience time is thus revealed: $Y = 1$, $\Delta = 1$, and $U = T$.

$C = 3$**:** The patient leaves without reporting, hence virtual waiting time (the time that the patient would have waited had he stayed in the ED) is observed, which provides an upper bound for the patience time, thus the patience time is left-censored. Formally, $Y = 0$, $\Delta = 1$, and $U = W$.

A graphical diagram of these categories appears in Figure 1.

Lemma 1. The following equalities hold:

i) $pr(U \le t, C = 1) = \int_0^t g(w)\overline{F}(w)dw$.

ii) $pr(U \le t, C = 2) = \int_0^t q(w)f(w)\overline{G}(w)dw$.

iii) $pr(U \le t, C = 3) = \int_0^t g(w) \int_0^w \{1 - q(x)\} f(x)dxdw$.

Here, $\overline{F}(t) = 1 - F(t)$ and $\overline{G}(t) = 1 - G(t)$ are the survival functions of the patience time and the waiting time, respectively. See the proof in A.1.

For $i = 1, 2, 3$, we introduce the following sub-stochastic density functions

$$h_i(t) := \frac{d}{dt} pr(U \le t, C = i). \tag{1}$$

From Lemma 1 above, we deduce that

$$h_1(t) = g(t)\overline{F}(t), \quad h_2(t) = q(t)f(t)\overline{G}(t), \quad h_3(t) = g(t) \int_0^t \{1 - q(x)\} f(x)dx.$$

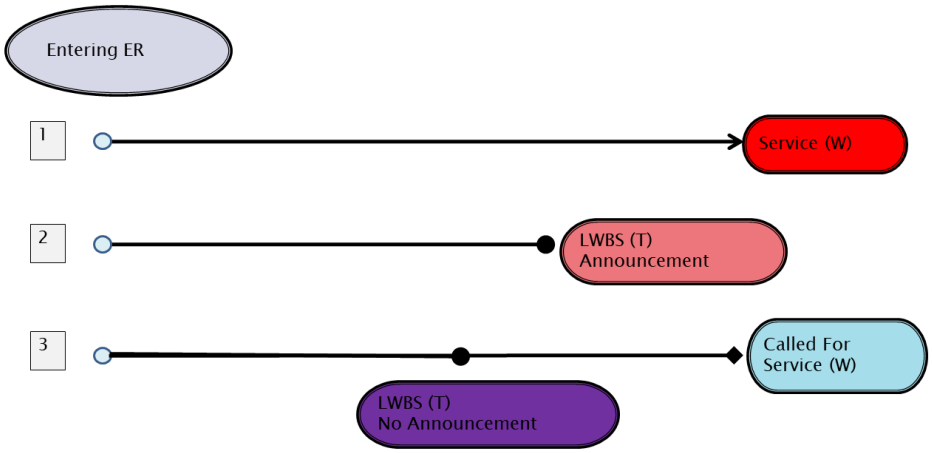

FIGURE 1 The three patient categories. Category 1 includes patients that received service. Category 2 includes patients that left without being seen and announced before leaving. Category 3 includes patients that left without being seen but did not announce leaving.

Define

$$r_1(t) \equiv \frac{h_1(t)}{pr(W \le T)}, \quad r_2(t) \equiv \frac{h_2(t)}{pr(Y = 1, W > T)}, \quad r_3(t) \equiv \frac{h_3(t)}{pr(Y = 0, W > T)}. \tag{2}$$

Then r_i is the density function of the observed time U given $C = i$. Our model assumes that all denominators are positive.

To summarize, what is known and what is to be estimated, there are two unknown distributions in our setting, G and F, and we aim to estimate them using both semiparametric and nonparametric techniques. For each patient, the waiting time is either observed or right censored. If the patient reports and then leaves, the waiting time is longer than the observed patience time. Hence, the waiting time is right-censored. Therefore, semiparametric and nonparametric estimation for the distribution of waiting time W can be done by standard techniques for right-censored data. However, estimation of the distribution of patience time T, is more complicated and is discussed in Sections 4 and 5.

4 | SEMIPARAMETRIC ESTIMATION

Assume now that the distributions of both the patience time and the waiting time belong to some parametric families. More formally, let $\mathcal{F} = \{f(\cdot;\theta), \theta \in \Theta\}$ where $\Theta \subseteq \mathbb{R}^d$, $\mathcal{G} = \{g(\cdot;\gamma), \gamma \in \Gamma\}$ where $\Gamma \subseteq \mathbb{R}^p$. We assume that the density of the patience time can be written as $f(t;\theta_0) \in \mathcal{F}$. We also assume that the density of the waiting time can be written as $g(t;\gamma_0) \in \mathcal{G}$. Write $h_1(t;\theta,\gamma) \equiv g(t;\gamma)\overline{F}(t;\theta)$, and similarly $h_2(t;\theta,\gamma) \equiv q(t)f(t;\theta)\overline{G}(t;\gamma)$ and $h_3(t;\theta,\gamma) \equiv g(t;\gamma)\int_0^t \{1-q(x)\}\, f(x;\theta)dx$.

The likelihood of the observed data $D = \{(U_i, Y_i, \Delta_i), i = 1, \dots, n\}$ can be written in terms of the functions h_1, h_2, and h_3, as follows:

$$L(D;\theta,\gamma) = \prod_{i=1}^{n} \left\{h_1(U_i;\theta,\gamma)\right\}^{1-\Delta_i} \left\{h_2(U_i;\theta,\gamma)\right\}^{\Delta_i Y_i} \left\{h_3(U_i;\theta,\gamma)\right\}^{\Delta_i(1-Y_i)}.$$

Using the explicit representations of h_1, h_2, h_3, we obtain that $L(D;\theta,\gamma)$ is given by

$$\prod_{i=1}^{n}\Bigg(\left\{g(U_i;\gamma)\overline{F}(U_i;\theta)\right\}^{1-\Delta_i} \left\{q(U_i)f(U_i;\theta)\overline{G}(U_i;\gamma)\right\}^{\Delta_i Y_i} \times \left[g(U_i;\gamma)\int_0^{U_i} \{1-q(s)\}\, f(s;\theta)ds\right]^{\Delta_i(1-Y_i)}\Bigg).$$

The value of γ that maximizes this likelihood is independent of θ. Therefore, a maximum likelihood estimator (MLE) $\widehat{\gamma}_n$ to γ_0 can be constructed from this likelihood.

Maximizing the likelihood with respect to θ is difficult. Even if γ_0 is given or estimated, the maximizer of θ depends on the unknown function $q(t)$. To address this challenge, we propose using a partial likelihood approach[22] which avoids the need to estimate $q(t)$. The partial likelihood that we use here is the likelihood calculated only for a specific category while ignoring the data for the other categories. In Theorem 1 below we show that, under standard regularity conditions, the maximizer of the partial likelihood is a consistent and asymptotically normal estimator for θ_0.

We consider the partial likelihood $L_{partial}(D;\theta;\gamma)$ of category $C = 1$,

$$\prod_{i=1}^{n} \left\{ \frac{g(U_i;\gamma)\overline{F}(U_i;\theta)}{\int_0^\infty g(s;\gamma)\overline{F}(s;\theta)ds} \right\}^{1-\Delta_i}.$$

The value of θ that maximizes this partial likelihood depends on γ. We plug the MLE $\widehat{\gamma}_n$ into this partial likelihood. Clearly, the resulting estimator for θ does not depend on the function $q(t)$ and thus no estimation of $q(t)$ is needed. Finding an estimator for the announcement probability function $q(t)$ is an interesting and challenging research question that is beyond the scope of this paper.

We need the following assumptions:

(A1) The derivative $\frac{\partial}{\partial\theta} f(t;\theta)$ is continuous in t for each $\theta \in \Theta$, $\frac{\partial}{\partial\gamma} g(t;\gamma)$ is continuous in t for each $\gamma \in \Gamma$.

(A2) For all $\theta \in \Theta$, $\arg\max_{\gamma\in\Gamma} L(D;\theta,\gamma)$ is unique, hence denote $\widehat{\gamma}(\theta) \equiv \arg\max_{\gamma\in\Gamma} L(D;\theta,\gamma)$. It is assumed as well that for each $\theta \in \Theta$, $\frac{\partial}{\partial\gamma} L\left\{D;\theta,\widehat{\gamma}(\theta)\right\} = 0$.

(A3) For all $\gamma \in \Gamma$, $\arg\max_{\theta\in\Theta} L_{partial}(D;\theta,\gamma)$ is unique, hence denote $\widehat{\theta}(\gamma) \equiv \arg\max_{\theta\in\Theta} L_{partial}(D;\theta,\gamma)$. It is assumed as well that for each $\gamma \in \Gamma$, $\frac{\partial}{\partial\theta} L_{partial}\left\{D;\widehat{\theta}(\gamma),\gamma\right\} = 0$.

Theorem 1. Let $\widehat{\gamma}_n$ be the maximizer of $L(D;\theta;\gamma)$ and let $\widehat{\theta}_n$ be the maximizer of $L_{partial}(D;\theta;\widehat{\gamma}_n)$. Then, as $n \to \infty$,

i) $\widehat{\gamma}_n \to \gamma_0$ in probability.

ii) $\sqrt{n}\left(\widehat{\gamma}_n - \gamma_0\right) \to N\left(0, V_{\gamma_0}\right)$ in distribution.

iii) $\widehat{\theta}_n \to \theta_0$ in probability.

iv) $\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \to N\left(0, S_{\theta_0,\gamma_0}\right)$ in distribution.

Here $V_{\gamma_0}, S_{\theta_0,\gamma_0}$ are covariance matrices as defined in Appendix A.1.

The proof appears in Appendix A.1.

Example 1. Assume that T follows an exponential distribution with rate θ and W follows an exponential distribution with rate γ. Then

$$\widehat{\gamma}_n = \frac{n - \Sigma_{i=1}^n \Delta_i Y_i}{\Sigma_{i=1}^n U_i}$$
$$\widehat{\theta}_n = \frac{\Sigma_{i=1}^n (1-\Delta_i)}{\Sigma_{i=1}^n U_i(1-\Delta_i)} - \widehat{\gamma}_n = \frac{\Sigma_{i=1}^n (1-\Delta_i)}{\Sigma_{i=1}^n U_i(1-\Delta_i)} - \frac{n - \Sigma_{i=1}^n \Delta_i Y_i}{\Sigma_{i=1}^n U_i}..$$

The details of the computation appears in Appendix A.5

5 | NONPARAMETRIC ESTIMATION

In this section we propose nonparametric estimators for the survival function of the patience time $\overline{F}$ and study its theoretical properties. For simplicity, we restrict the estimation to an interval $[0,\tau]$ for some $\tau > 0$, such that the probability of W and T being larger than τ is positive. This is a standard condition in survival estimation[23] (Chapter 4.2).Note that for observations of Categories 1 and 3, the waiting-time is observed. For Category 2, only a lower bound of the waiting time is observed. Hence, the waiting time is either observed or right-censored. Therefore, estimating the waiting time distribution can be done by using standard survival analysis estimators such as the Kaplan–Meyer estimator. On the other hand, estimating the distribution of the

patience time is more challenging since we cannot distinguish between the density function f and the unknown function q. Our goal is thus to estimate the distribution of the patience time F.

Assume that over all positive numbers, the waiting time density function g is strictly positive. Recall that $h_1(t) = g(t)\overline{F}(t)$, $h_3(t) = g(t)\int_0^t \{1 - q(s)\} f(s)ds$, where the functions h_1, h_3 are defined as in (1). Therefore,

$$\frac{h_1(t)}{h_3(t)} = \frac{\overline{F}(t)}{F(t) - \int_0^t q(s)f(s)ds}. \tag{3}$$

which is well defined as $g(t) > 0$. Reordering the terms in (3), we get that

$$\left\{ F(t) - \int_0^t q(s)f(s)ds \right\} \frac{h_1(t)}{h_3(t)} = 1 - F(t).$$

Hence,

$$F(t) = \frac{h_3(t) + h_1(t)\int_0^t q(s)f(s)ds}{h_3(t) + h_1(t)}.$$

From the definitions in (2), it follows that

$$F(t) = \frac{pr(Y = 0, T < W)r_3(t) + pr(W \le T)r_1(t)\int_0^t q(s)f(s)ds}{pr(Y = 0, T < W)r_3(t) + pr(W \le T)r_1(t)}. \tag{4}$$

Therefore, we propose to estimate $F(t)$ by estimating the following terms:

(i) $pr(W \le T)$ and $pr(Y = 0, T < W)$,

(ii) $r_1(t)$ and $r_3(t)$,

(iii) $A(t) \equiv \int_0^t q(s)f(s)ds$.

Estimating the expression in (i) can be done by the empirical estimators: $\widehat{pr}(T \le W) = n^{-1}\Sigma_{i=1}^n(1 - \Delta_i)$, $\widehat{pr}(Y = 0, W < T) = n^{-1}\Sigma_{i=1}^n \Delta_i(1 - Y_i)$. These estimators converge, by the central limit theorem (CLT), to $pr(W \le T)$ and $pr(Y = 0, T < W)$, respectively, at the rate of $n^{1/2}$.

Since r_1 and r_3 are density functions, they can be estimated using a kernel estimator[24] (Chapter 1.2). Let $\widehat{r}_1$ and $\widehat{r}_3$ be kernel estimators of r_1 and r_3, respectively. Assume that both r_1 and r_3 belong to a Sobolev function class of order β. Then for each $t > 0$, both $\widehat{r}_1(t)$ and $\widehat{r}_3(t)$ converge at a rate of $n^{\beta/(2\beta+1)}$. Here, the parameter $\beta \ge 1$ is an integer that represents the smoothness of a function. Specifically, if $\beta > k$ for some integer k, then the function is at least k-time differentiable.[24].

We now turn to estimate the term $A(t) = \int_0^t q(s)f(s)ds$. A nonparametric estimator that we created for this term is defined and proven to be consistent in the following lemma.

Lemma 2. Let

$$\widehat{N}_n(t) \equiv \frac{1}{n}\sum_{i=1}^n Y_i\Delta_i 1\{U_i \le t\}, \quad \widehat{Y}_n(t) \equiv \frac{1}{n}\sum_{i=1}^n 1\{U_i \ge t\}.$$

Define $\widehat{D}_n(t) \equiv \int_0^t \frac{d\widehat{N}_n(s)}{\widehat{Y}_n(s)}$. Then $\widehat{A}(t) \equiv 1 - \exp\left\{-\widehat{D}_n(t)\right\}$ converges pointwise to $A(t)$, at a rate of $n^{1/2}$, for every $t \in [0, \tau]$.

The proof is given in Appendix A.3.

By plugging in the estimators

$$\widehat{pr}(Y = 0, W < T),\ \widehat{pr}(T \le W),\ \widehat{r}_3(t),\ \widehat{r}_1(t),\ \widehat{A}(t),$$

to the equation in (4), we get that

$$\widehat{F}_n(t) = \frac{\widehat{pr}(Y = 0, W < T)\widehat{r}_3(t) + \widehat{pr}(T \le W)\widehat{r}_1(t)\widehat{A}(t)}{\widehat{pr}(Y = 0, W < T)\widehat{r}_3(t) + \widehat{pr}(T \le W)\widehat{r}_1(t)}, \tag{5}$$

is an estimator of $F(t)$.

Theorem 2. The estimator $\hat{F}_n(t)$ converges pointwise to $F(t)$ at a rate of $n^{\beta/(2\beta+1)}$,
for every $t \in [0, \tau]$.

The proof appears in Appendix A.4. Since that $\hat{F}_n$ is based on density estimation, it is not necessarily monotonic, we therefore replace it with a monotonic approximation. The monotonic approximation is by taking the cumulative sup.

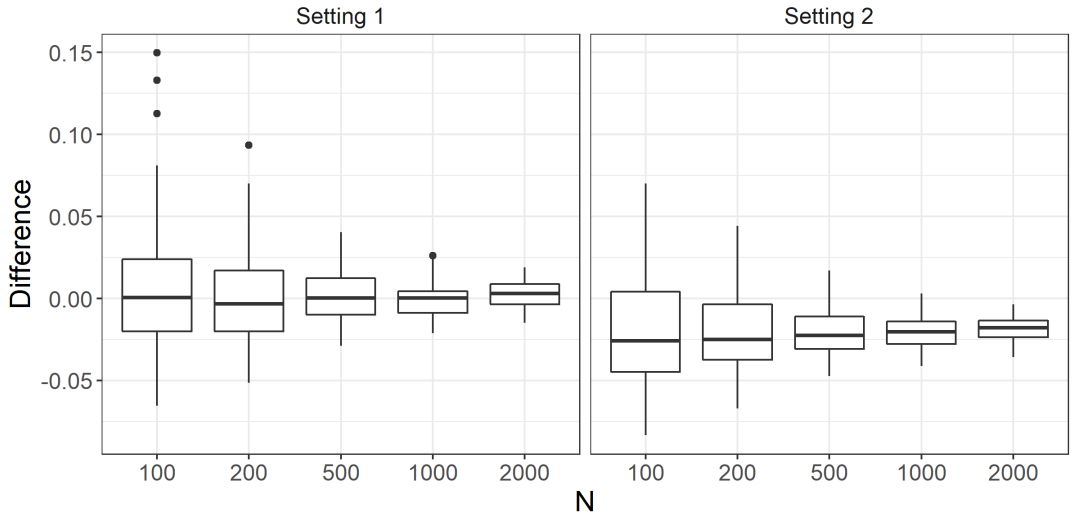

FIGURE 2 The difference between the semiparametric estimator of θ_0 and θ_0. Setting 1: The patience time T follows an exponential distribution with expectation of 16 hours and the waiting time W follows and exponential distribution with expectation of 2 hours. Setting 2: The patience time T follows a Weibull distribution with scale 16 and shape 1.5 while the waiting time W follows an exponential distribution with expectation of 2 hours.

	Setting 1: Exponential						Setting 2: Weibull					
	Semiparametric			Nonparametric			Semiparametric			Nonparametric		
N	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std
100	1.014	0.432	1.44	1.428	0.804	1.812	0.474	0.228	0.618	0.216	0.114	0.264
200	0.51	0.21	0.828	1.062	0.672	1.11	0.414	0.252	0.492	0.072	0.036	0.12
500	0.162	0.084	0.198	0.462	0.288	0.474	0.378	0.294	0.288	0.03	0.018	0.03
1000	0.078	0.042	0.102	0.246	0.186	0.186	0.342	0.312	0.204	0.018	0.012	0.018
2000	0.054	0.03	0.066	0.168	0.132	0.12	0.348	0.318	0.132	0.012	0.0006	0.012

TABLE 1 MSE for Settings 1 and 2. The table summarizes the MSE that was calculated (100 times) for each of the sample sizes. For Setting 1, the patience time T follows an exponential distribution with expectation of 16 hours and the waiting time W follows an exponential distribution with expectation of 2 hours. In Setting 2 the patience time T follows a Weibull distribution with scale 16 and shape 1.5, while the waiting time W follows an exponential distribution with expectation of 2 hours. The estimates are given in minutes. As can be seen the nonparametric estimator responded with a lower MSE.

6 | SIMULATIONS

We study the performance of both the semiparametric and nonparametric estimators that were proposed in Sections 4 and 5, respectively. Based on the setting of the case study discussed in Section 7, we consider two simulation settings. In the case study, both the exponential and Weibull distributions seem to fit well the waiting time and patience time distributions, respectively. Thus, we chose parameters based on the fit for the urgent level, which is the middle severity level.

Specifically, the two simulation settings consist of samples from exponential and Weibull distributions in which the waiting time has a smaller mean then the patience time mean, as was observed in the case study. In the first setting, following the data from the case study, a sample was taken from the model in which the patience time T follows an exponential distribution with expectation of 16 hours, and the waiting time W follows an exponential distribution with expectation of 2 hours. In the second setting a sample was taken from a model in which the patience time T follows a Weibull distribution with scale 16 and shape 1.5, which closely related to the observed data; and where the waiting time W follows an exponential distribution with expectation of 2 hours as before. In both settings, the unknown probability of announcement is $q(t) = \exp(-t)$. Taking the probability of

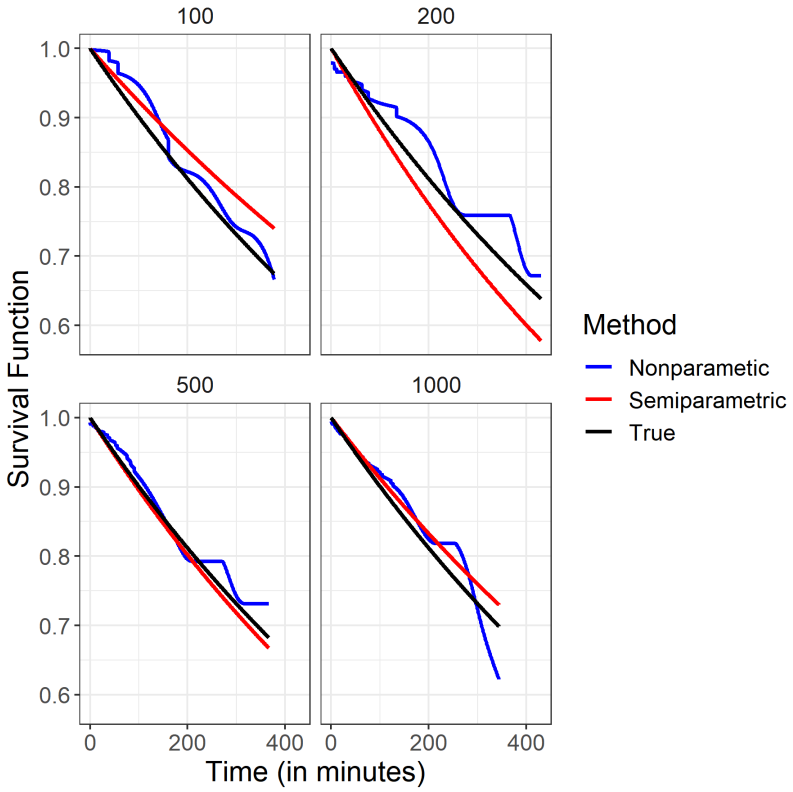

FIGURE 3 Setting 1. The blue, red, and black curves represent the nonparametric, semiparametric, and true survival functions, respectively, for $N = 100, 200, 500$, and 1000.

announcement to be the increasing function $q(t) = 1-\exp(-t)$ or the constant function $q(t) = 0.5$ yields similar results which are omitted. Moreover, we experimented with additional numerical values. The behavior and conclusions, as reported here, remain consistent across these experiments.

In each setting, we calculated the semiparametric estimator for the scale of T for five different sample sizes ($N = 100, 200, 500, 1000, 2000$). For each sample size, we repeated the simulation 100 times. When using the semiparametric method, it was assumed that both T and W follow an exponential distribution with unknown parameters. Note that this assumption holds for the first setting but does not hold for the second one. In other words, the second setting is carried out under a misspecified model. The results are shown in Figure 2.

We compare $\widehat{\overline{F}}_n$, the estimator of the survival function of T, to the true survival function $\overline{F}_0$. For the semiparametric estimation, $\widehat{\overline{F}}_n(t) = \exp(-\hat{\theta}t)$, while for the nonparametric estimator $\widehat{\overline{F}}_n(t)$ is given by (5). The comparison is done using mean square error (MSE), which is defined by

$$MSE(\widehat{\overline{F}}_n, \overline{F}_0) \equiv \int_{-\infty}^{\infty} \left\{ \widehat{\overline{F}}_n(t) - \overline{F}_0(t) \right\}^2 f_0(t)dt\,,$$

where f_0 is the density of T. The semiparametric and nonparametric survival function estimators are demonstrated in Figures 3 and 4. Figure 3 represents the results of the first setting in which T follows an exponential distribution with scale 13 and W follows an exponential distribution with scale 2. Figure 4 represents the results of the second setting in which T follows a Weibull distribution with scale 13 and shape 1.5, and W follows an exponential distribution with scale 2. Summaries of the MSE are given in Table 1. Not surprisingly, for Setting 1, since the semiparametric model is correct, the MSE is smaller for the semiparametric estimator. Similarly, since in Setting 2 the semiparametric model is incorrect, the MSE is smaller for the nonparametric estimator.

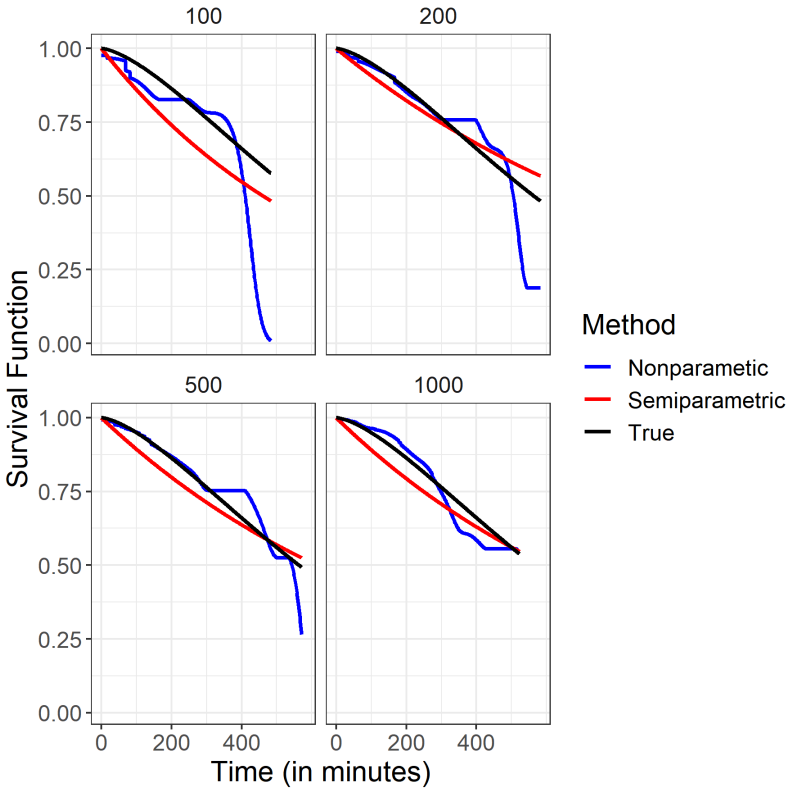

FIGURE 4 Second setting. The blue, red, and black curves represent the nonparametric, semiparametric, and true survival functions, respectively.

7 | CASE STUDY

As leaving without being seen by a physician may have a strong effect on patient well-being and satisfaction, estimating the time that patients are willing to wait in the emergency department (ED) is an important and challenging question.[19,25] While there has been considerable research in this field,[6–9] due to the special structure of the data, the duration that a potential patient is willing to wait for ED service has not been thoroughly investigated.

We analyze data from all patient presentations to triage at an urban, academic, adult-only emergency department (ED) with visits in calendar year 2008. This data was used for the analysis in Wiler et al.[19] The data consist of the waiting time of patients arriving at the emergency room stratified by acuity levels. We focused on the three main levels of acuity: emergency, urgent, and semi-urgent. For each acuity level, we categorized each visit into one of the three categories: received service, left without being seen and announced, and left without being seen and did not announce. We considered only patients that were not served upon arrival, or left without waiting at all. The characteristics of the dataset appear in Table 2. As can be seen, the are considerably fewer emergency visits and only 1.2% of these left without being seen. In comparison, in the urgent and semi-urgent acuity levels, about 10% left without being seen. The distribution of the observed times for each acuity levels, stratified by the patient's category, appears in Figure 5. Overall, the distribution of the three categories is similar in each acuity level.

We analyzed the data using the semiparametric and nonparametric estimators for the distribution of the patience time proposed in Sections 4 and 5. Since our model assumes that all patients follow the same distribution, we calculated the estimators for each level of acuity separately. The data consist of the triple variables (U_i, Δ_i, Y_i) described in Section 3 such that each observation is categorized to one of the three possible categories. The results of these estimators are given in Figures 6 and 7. As can be seen from Figure 6, the results of the semiparametric and nonparametric estimators agree, which suggests that modeling the patience time using the exponential distribution is reasonable. Figure 7 shows that the patience times are stochastically ordered by levels of acuity. In other words, patients at the severe acuity level are less probable to lose patience than patients at the urgent level, who in turn are less prone to lose patience than patients at the semi-urgent level, as expected.

	Emergency	Urgent	Semi-Urgent
n	8579	36249	26036
Category (%)			
Service	8478 (98.8%)	32607 (90.0%)	23788 (91.4%)
LWBS & Announcement	69 (0.8%)	1908 (5.3%)	1019 (3.9%)
LWBS & No Announcement	32 (0.4%)	1734 (4.8%)	1229 (4.7)%
Mean Observed Time (SD)	25.11 (24.26)	111.73 (108.21)	82.18 (72.92)

TABLE 2 The characteristics of the different visits stratified by acuity level.

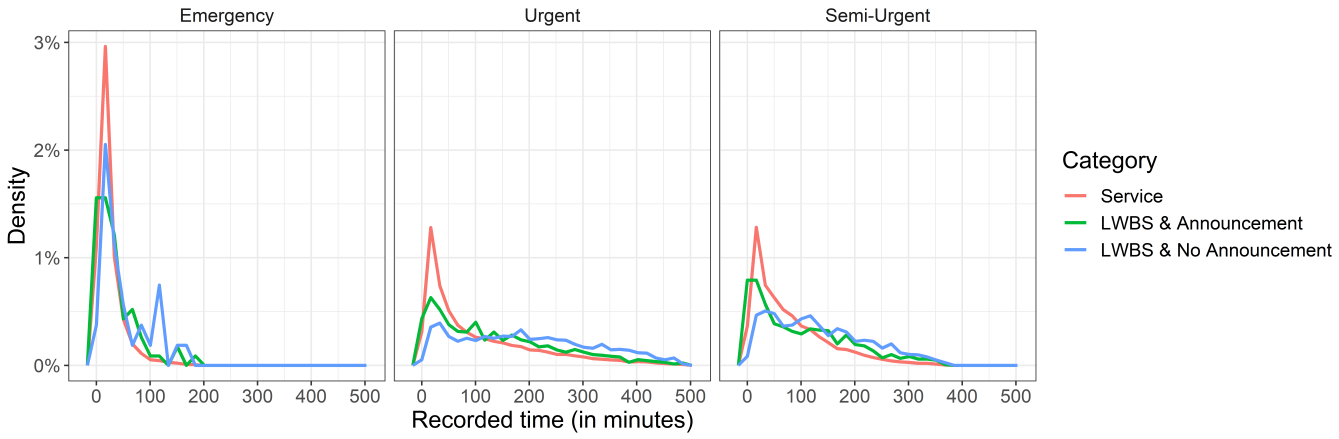

FIGURE 5 The distribution of the observed time stratified by acuity levels and category.

8 | DISCUSSION

In this paper, we consider survival data that combine observed, right-censored, and left-censored data. The setting we analyzed was that of patients who wait for treatment in an emergency department, where some patients may leave without being seen. We proposed both semiparametric and nonparametric estimators for the distribution of the patience time.

Using simulation, we showed that when the semiparametric model holds, the semiparametric estimator estimates the patience time well. However, when the model is misspecified, the nonparametric estimator behaves better. While in our case study, both estimators behave similarly, it is of importance to further investigate when each of these estimators is preferable. So far, no baseline covariates were given. Novel semiparametric and nonparametric estimators are needed for addressing settings that include baseline covariates.

ACKNOWLEDGEMENT

Y. Ritov was partially supported by the Israeli Science Foundation (grant No. 1770/15). Y. Goldberg was partially supported by the Israeli Science Foundation (grant No. 849/17).

DATA AVAILABILITY

Due to privacy and confidentiality issues, research data are not shared.

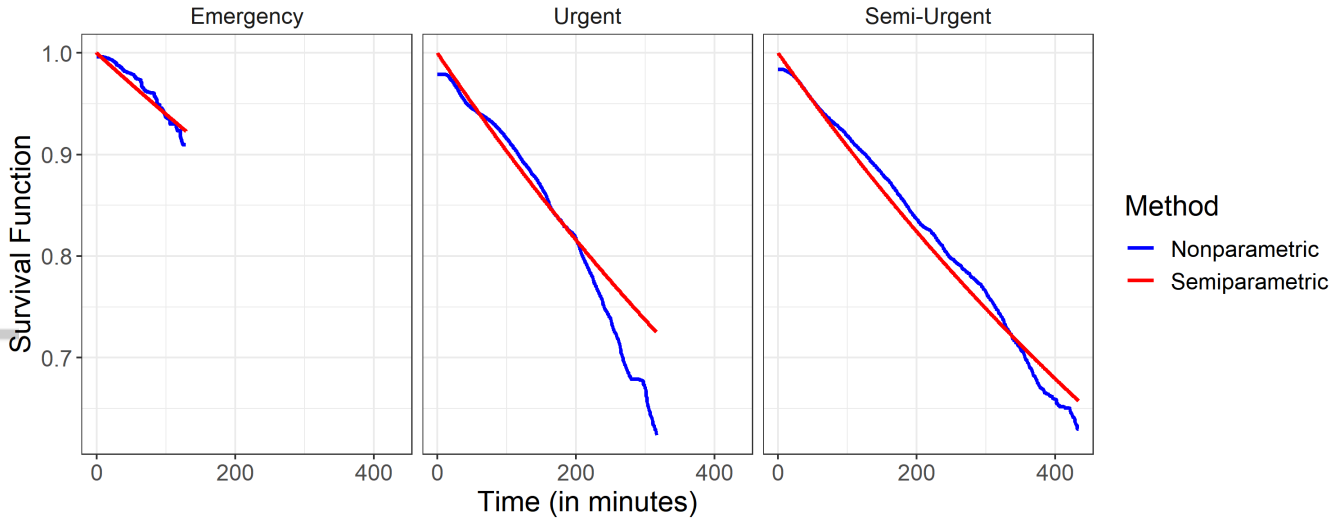

FIGURE 6 Compression of the nonparametric and semiparametric estimators for the survival of the patience time by different levels of acuity.

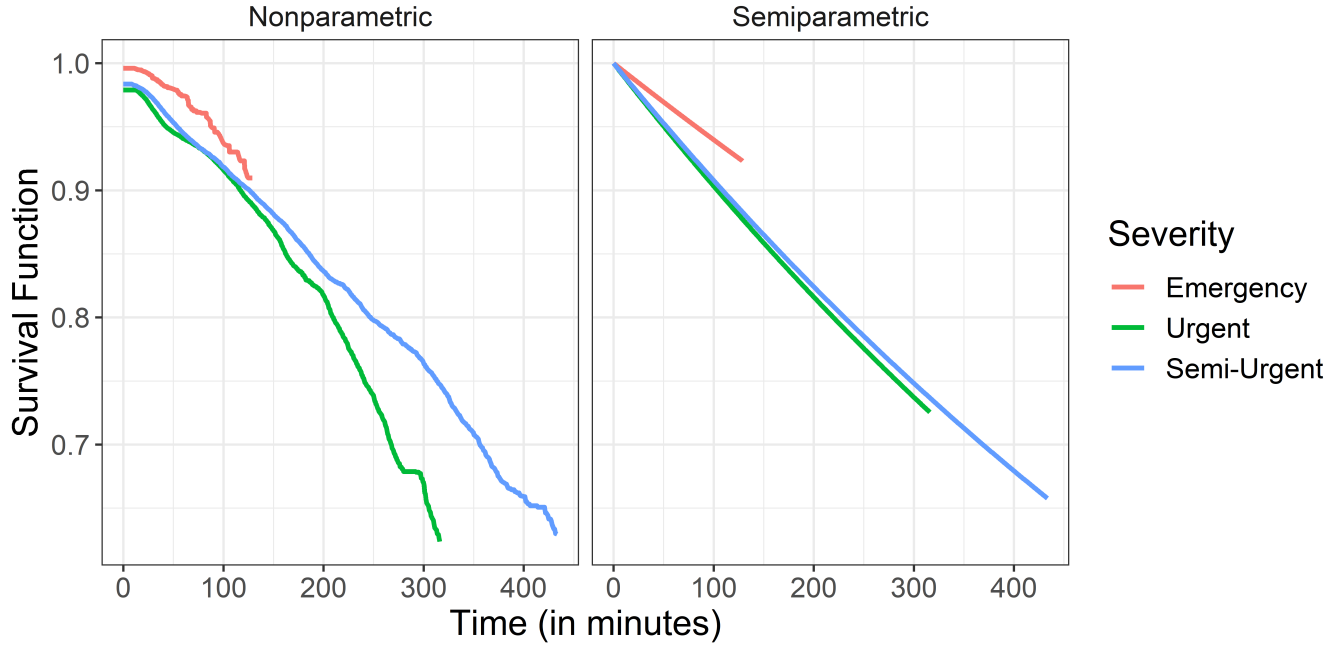

FIGURE 7 Compression of the estimator for the survival function of the patience time at the three different levels of severity.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

APPENDIX

A PROOFS

A.1 Proof of Lemma 1

$$\begin{aligned} pr(U \le t, C = 1) =& pr(U \le t, W \le T) \\ =& pr(W \le t, W \le T) \\ =& \int_0^t pr(W \le T \mid W = s) g(s) ds \\ =& \int_0^t pr(s \le T) g(s) ds \\ =& \int_0^t g(s) \overline{F}(s) ds, \end{aligned}$$

where in the fourth equality we use the independence between W and (Y, T).

This establishes i). For ii), we have

$$\begin{aligned} pr(U \le t, C = 2) =& pr(U \le t, Y = 1, T < W) \\ =& pr(T \le t, Y = 1, T < W) \\ =& \int_0^t pr(s < W \mid Y = 1, T = s) q(s) f(s) ds \\ =& \int_0^t pr(s < W) q(s) f(s) ds \\ =& \int_0^t q(s) f(s) \overline{G}(s) ds, \end{aligned}$$

where in the fourth equality we use the independence between W and (Y, T).

Finally, for iii),

$$\begin{aligned}
P(U \le t, C = 3) =& pr(U \le t, Y = 0, T < W) \\
=& pr(W \le t, Y = 0, T < W) \\
=& \int_0^t pr(Y = 0, T < s \mid W = s) g(s) ds \\
=& \int_0^t g(s) pr(Y = 0, T < s) ds \\
=& \int_0^t g(s) \int_0^s pr(Y = 0 \mid T = x) f(x) dx ds \\
=& \int_0^t g(s) \int_0^s \{1 - q(x)\} f(x) dx ds,
\end{aligned}$$

where in the fourth equality we use the independence between W and (Y, T).

A.2 Proof of Theorem 1

The log of the full likelihood is

$$\begin{aligned}
m(D; \theta, \gamma) = \sum_{i=1}^{n} \Bigg[& \left(1 - \Delta_i\right) \left(\log g(U_i; \gamma) + \log \overline{F}(U_i; \theta) \right) \\
& + \Delta_i Y_i \left(\log q(U_i) + \log f(U_i; \theta) + \log \overline{G}(U_i; \gamma) \right) \\
& + \Delta_i (1 - Y_i) \left\{ \log g(U_i; \gamma) + \log \int_0^{U_i} (1 - q(s)) f(s; \theta) ds \right\} \Bigg].
\end{aligned}$$

Given the data D,

$$\frac{1}{n} l(D; \theta, \gamma) = \mathbb{P}_n \left\{ m_\gamma(U, \Delta, Y) + c(U, \Delta, Y; \theta) \right\} \equiv \frac{1}{n} \sum_{i=1}^{n} \left\{ m_\gamma(U_i, \Delta_i, Y_i) + c(U_i, \Delta_i, Y_i; \theta) \right\} \tag{A1}$$

where $m_\gamma : \mathbb{R}^+ \times \{0, 1\}^2 \to \mathbb{R}$ is defined by

$$m_\gamma(u, \delta, y) \equiv (1 - \delta) \log g(u; \gamma) + \Delta y \log \overline{G}(u; \gamma) + \delta (1 - y) \log g(u; \gamma).$$

and

$$\begin{aligned}
c(u, \delta, y; \theta) \equiv & (1 - \delta) \log \overline{F}(u; \theta) \\
& + \delta y \left(\log q(u) + \log f(u; \theta) \right) + \delta (1 - y) \log \int_0^u \{1 - q(s)\} f(s; \theta) ds.
\end{aligned}$$

From assumption A1 we obtain that, for each $\theta \in \Theta$, $\arg\max_{\gamma \in \Gamma} l(D; \gamma, \theta) = \arg\max_{\gamma \in \Gamma} \mathbb{P}_n(m_\gamma)$. The γ that maximizes $L(D; \theta, \gamma)$ does not depend on the value of θ or the function p. Define $M_n(\gamma) \equiv \mathbb{P}_n m_\gamma$ and $M(\gamma) \equiv P m_\gamma$.If, for a general function h, $Ph \equiv \int h(x) d P(x)$ and $\mathbb{P}_n h \equiv n^{-1} \sum_{i=1}^{n} h(X_i)$ then by Assumptions A1–A3, Theorem 5.7 in van der Vaart[26] can be applied. Therefore $\widehat{\gamma}_n \to \gamma_0$, in probability, which concludes the proof of i).

Given the data D, the term $\frac{\partial l(D; \theta, \gamma)}{\partial \gamma}$ is a function of γ and does not depend on the unknown function p. We also have

$$\frac{1}{n} \frac{\partial l(D; \theta, \gamma)}{\partial \gamma} = \mathbb{P}_n \psi_\gamma(U, \Delta, Y) \equiv \frac{1}{n} \sum_{i=1}^{n} \psi_\gamma(U_i, \Delta_i, Y_i),$$

where $\psi_\gamma : \mathbb{R}^+ \times \{0,1\}^2 \to \mathbb{R}$ is defined as

$$\psi_\gamma(u,\delta,y) \equiv (1-\delta)\frac{\frac{\partial}{\partial\gamma}g(u;\gamma)}{g(u;\gamma)} - \delta y\frac{\frac{\partial}{\partial\gamma}\overline{G}(u;\gamma)}{\overline{G}(u;\gamma)} + \delta(1-y)\frac{\frac{\partial}{\partial\gamma}g(u;\gamma)}{g(u;\gamma)}.$$

By Assumptions (A1)–(A3), ψ_γ satisfies the conditions of Theorem 5.41 in van der Vaart[26] and, therefore,

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) = -(P\dot{\psi}_{\gamma_0})^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\gamma_0}(U_i,\Delta_i,Y_i) + o_p(1),$$

where $\dot{\psi}_\gamma(x) = \frac{\partial}{\partial\gamma}\psi_\gamma(x)$. Hence $\hat{\gamma}_n$ is a linear asymptotically normal (LAN) estimator with influence function $\varphi \equiv -(P\dot{\psi}_{\gamma_0})^{-1}\psi_{\gamma_0}$. From all of the above we get that ii) is proved with $V_{\gamma_0} = (P\dot{\psi}_{\gamma_0})^{-1}P\psi_{\gamma_0}\psi_{\gamma_0}^t(P\dot{\psi}_{\gamma_0})^{-1}$.

To prove iii), note that due to the term $\log\int_0^{U_i}(1-q(s))f(s;\theta)ds$ that appears in $l(D;\theta,\gamma)$, the term $\frac{\partial l(D;\theta,\gamma)}{\partial\theta}$ depends on the unknown function p. We therefore consider a partial likelihood function such that its derivative with respect to θ does not depend on p. The partial likelihood that satisfies this request is the partial likelihood of $C = 1$:

$$\prod_{i=1}^{n}\left\{\frac{g(U_i;\gamma)\overline{F}(U_i;\theta)}{\int_0^\infty g(s;\gamma)\overline{F}(s;\theta)ds}\right\}^{1-\Delta_i}$$

The log of the partial likelihood is

$$\text{"}l_{partial}(D;\theta,\gamma) = \sum_{i=1}^{n}\left(1-\Delta_i\right)\left\{\log g(U_i;\gamma) + \log\overline{F}(U_i;\theta) - \log\int_0^\infty g(s;\gamma)\overline{F}(s;\theta)ds\right\}.$$

Given the data D, the term $l_{partial}(D;\theta,\gamma)$ is a function only of the parameters θ and γ. We also have

$$\frac{1}{n}l_{partial}(D;\theta,\gamma) = \mathbb{P}_n r_{\theta,\gamma}(U,\Delta,Y) \equiv \frac{1}{n}\sum_{i=1}^{n}r_{\theta,\gamma}(U_i,\Delta_i,Y_i),$$

where $r_{\theta,\gamma} : \mathbb{R}^+ \times \{0,1\}^2 \to \mathbb{R}$ is given by

$$r_{\theta,\gamma}(U,\Delta,Y) \equiv (1-\Delta)\left\{\log g(U;\gamma) + \log\overline{F}(U;\theta) - \log\int_0^\infty g(s;\gamma)\overline{F}(s;\theta)ds\right\}.$$

Define $M_n(\theta,\gamma) \equiv \mathbb{P}_n r_{\theta,\gamma}$, and $M(\theta,\gamma) \equiv Pr_{\theta,\gamma}$. Then, Theorem 5.7 in van der Vaart[26] can be applied. Therefore $\left(\widehat{\theta}_n,\widehat{\gamma}_n\right) \to \left(\theta_0,\gamma_0\right)$ in probability, and in particular $\widehat{\theta}_n \to \theta_0$ in probability, and iii) is proven.

In order to prove iv), note that

$$\frac{1}{n}\frac{\partial l_{partial}(D;\theta,\gamma)}{\partial\gamma} = \mathbb{P}_n\phi_{\theta,\gamma}(U,\Delta,Y) \equiv \frac{1}{n}\sum_{i=1}^{n}\phi_{\theta,\gamma}(U_i,\Delta_i,Y_i),$$

where $\phi_{\theta,\gamma} : \mathbb{R}^+ \times \{0,1\}^2 \to \mathbb{R}$ is defined as

$$\phi_{\theta,\gamma}(u,\delta,y) \equiv (1-\delta)\left\{\frac{\frac{\partial}{\partial\theta}\overline{F}(u;\theta)}{\overline{F}(u;\theta)} - \frac{\frac{\partial}{\partial\theta}\int_0^\infty g(s;\gamma)\overline{F}(s;\theta)ds}{\int_0^\infty g(s;\gamma)\overline{F}(s;\theta)ds}\right\}.$$

Using Assumptions A1–A3, together from Theorem 5.41 in van der Vaart[26], we obtain that

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{-(P\dot{\psi}_{\gamma_0})^{-1}\psi_{\gamma_0}(U_i,\Delta_i,Y_i)\right\} + o_p(1).$$

Define $\Phi_n(\theta,\gamma) \equiv \mathbb{P}_n\phi_{\theta,\gamma}$ and note that $\Phi(\theta_0,\gamma_0) \equiv P\phi_{\theta_0,\gamma_0} = 0$ (since under the true parameters $P\phi_{\theta_0,\gamma_0} = \frac{\partial}{\partial\theta}\int dh_0 = \frac{\partial}{\partial\theta}1 = 0$, where dh_0 is the true distribution of category 1).

By Talyor's theorem,

$$
\begin{aligned}
0 = \Phi_n(\widehat{\theta}_n, \widehat{\gamma}_n) &= \Phi_n(\theta_0, \gamma_0) + \left\{ \frac{\partial}{\partial \theta} \Phi_n(\theta_0, \gamma_0) \right\}^T (\widehat{\theta}_n - \theta_0) + \left\{ \frac{\partial}{\partial \phi} \Phi_n(\theta_0, \gamma_0) \right\}^T (\widehat{\gamma}_n - \gamma_0) + o_p\left(n^{-1/2}\right) \\
\Rightarrow 0 &= \sqrt{n}\Phi_n(\theta_0, \gamma_0) + \left\{ \frac{\partial}{\partial \theta} \Phi_n(\theta_0, \gamma_0) \right\}^T \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) + \left\{ \frac{\partial}{\partial \gamma} \Phi_n(\theta_0, \gamma_0) \right\}^T \sqrt{n}\left\{\widehat{\gamma}_n - \gamma_0\right\} + o_p(1). \\
&= \sqrt{n}\Phi_n(\theta_0, \gamma_0) + \left\{ \frac{\partial}{\partial \theta} \Phi_n(\theta_0, \gamma_0) \right\}^T \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \\
&\quad - \left\{ \frac{\partial}{\partial \gamma} \Phi_n(\theta_0, \gamma_0) \right\}^T (P\dot{\psi}_{\gamma_0})^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\gamma_0}(U_i, \Delta_i, Y_i) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\phi_{\theta_0,\gamma_0}(U_i, \Delta_i, Y_i) - \left\{ \frac{\partial}{\partial \gamma} \Phi_n(\theta_0, \gamma_0) \right\}^T (P\dot{\psi}_{\gamma_0})^{-1} \psi_{\gamma_0}(U_i, \Delta_i, Y_i) \right] \\
&\quad + \left\{ \frac{\partial}{\partial \theta} \Phi_n(\theta_0, \gamma_0) \right\}^T \sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1).
\end{aligned}
$$

Elementary arithmetic leads to

$$
\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) = -\left(EE^T\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \phi_{\theta_0,\gamma_0}(U_i, \Delta_i, Y_i) - B^T (P\dot{\psi}_{\gamma_0})^{-1} \psi_{\gamma_0}(U_i, \Delta_i, Y_i) \right\} + o_p(1),
$$

where $B \equiv \frac{\partial}{\partial \gamma}\Phi(\theta_0, \gamma_0)$, and $E \equiv \frac{\partial}{\partial \theta}\Phi(\theta_0, \gamma_0)$. Hence, $\widehat{\theta}_n$ is a LAN estimator with the influence function

$$
\varphi = -\left(EE^T\right)^{-1} \frac{1}{\sqrt{n}} \left\{ \phi_{\theta_0,\gamma_0} - B^T (P\dot{\psi}_{\gamma_0})^{-1} \psi_{\gamma_0} \right\}.
$$

Summarizing,

$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to N(0, S_{\theta_0,\gamma_0})$ in distribution, where $S_{\theta_0,\gamma_0} = P\varphi\varphi^T$, hence, iv) is proven.

A.3 Proof of Lemma 2

Proof. We use similar arguments to those in the proof of the convergence of the Nelson–Aalen estimator to a cumulative hazard function, see Kosorok[23], page 240. Hence we have that

$$
\sqrt{n} \begin{Bmatrix} \widehat{N}_n(t) - N(t) \\ \widehat{Y}_n(t) - Y(t) \end{Bmatrix} = O_p(1),
$$

where $N(t) = pr(Y = 1, T \le W, T \le t)$ and $Y(t) = pr(U \ge t)$.

Since, by Section 3,

$$
\{U > t\} = \{Y = 0, W > t\} \cup \{Y = 1, W > t, T > t\}.
$$

Hence,

$$
\begin{aligned}
pr(U > t) &= pr(Y = 0, W > t) + pr(Y = 1, W > t, T > t) \\
&= pr(W > t)\left\{pr(Y = 0) + pr(Y = 1, T > t)\right\} \\
&= pr(W > t)\big(pr(Y = 0) + pr(Y = 1) - pr(Y = 1, T \le t)\big) \\
&= pr(W > t)\left\{1 - pr(Y = 1, T \le t)\right\} \\
&= \overline{G}(t)\left\{1 - \int_0^t q(s) f(s) ds\right\},
\end{aligned}
$$

where in the second equality we use the independence between W and (Y, T).

By Lemma 1

$$pr(Y = 1, T \leq W, T \leq t) = \int_0^t q(s)f(s)\overline{G}(s)ds.$$

Using the continuity of the derivative operator and of the integral operator, we get that

$$\widehat{D}_n(t) \to \int_0^t \frac{q(s)f(s)\overline{G}(s)}{\overline{G}(s)\left\{1 - \int_0^s q(x)f(x)ds\right\}}ds$$

in probability.

Note that

$$\int_0^t \frac{\overline{G}(s)q(s)f(s)}{\overline{G}(s)\left\{1 - \int_0^s q(x)f(x)dx\right\}}ds = \int_0^t \frac{q(s)f(s)}{\left\{1 - \int_0^s q(x)f(x)dx\right\}}ds$$
$$= -\int_0^t \frac{\partial}{\partial s}\log\left\{1 - \int_0^s q(x)f(x)dx\right\}ds = -\log\left\{1 - \int_0^t q(s)f(s)ds\right\}.$$

Hence, by the delta method, see Kosorok[23] Chapter 12.2.2.2, we get that

$$\widehat{D}_n(t) \to -\log\left\{1 - \int_0^t q(s)f(s)ds\right\}$$

in probability, with convergence at rate $n^{1/2}$.

Since $y = -\log(1-x) \Leftrightarrow x = 1 - \exp(-y)$ and by the continuous mapping theorem, see Theorem 7.7 of Kosorok[23], we get that $\widehat{A}(t) = 1 - \exp(-\widehat{D}_n(t))$ is an estimator of $\int_0^t q(s)f(s)ds$, at the rate of $n^{1/2}$ as desired. □

A.4 Proof of Theorem 2

For the proof of Theorem 2, we need the following lemma, which is elementary hence stated without proof.

Lemma 3. Let $(a_n)_{n=1}^{\infty}, (b_n)_{n=1}^{\infty}$ be positive sequences. If $X_n - X = O_p(a_n)$ and $Y_n - Y = O_p(b_n)$, as well as $P(|X| > l) = 1$ for some $l > 0$. Then we have:

i) $X_n + Y_n - (X + Y) = O_p(a_n \vee b_n)$,

ii) $X_nY_n - XY = O_p(a_n \vee b_n)$,

iii) $\frac{1}{X_n} - \frac{1}{X} = O_p(a_n)$.

Proof of Theorem 2. Recall that

$$\widehat{F}_n(t) = \frac{\widehat{pr}(Y = 0, W < T)\widehat{r}_3(t) + \widehat{pr}(T \leq W)\widehat{r}_1(t)\widehat{A}(t)}{\widehat{pr}(Y = 0, W < T)\widehat{r}_3(t) + \widehat{pr}(T \leq W)\widehat{r}_1(t)}$$

is an estimator of $F(t)$.

For all $t > 0$,

$$\widehat{pr}(Y = 0, W < T) - pr(Y = 0, W < T) = O_p(n^{-1/2}),$$

and

$$\widehat{pr}(T \leq W) - pr(T \leq W) = O_p(n^{-1/2}),$$

as both are empirical distribution estimators. By Chapter 1.7 of Tsybakov[24],
for all $t > 0$, $\widehat{r}_j(t) - r_j(t) = O_p(n^{-\beta/(2\beta+1)})$, for $j = 1, 3$. By Lemma 2, $\widehat{A}(t) - A(t) = O_p(n^{-1/2})$.

By Lemma 3.ii)

$$\widehat{pr}(Y=0,W<T)\widehat{r}_3(t)-pr(Y=0,W<T)\widehat{r}_3(t)=O_p(n^{-\beta/(2\beta+1)}),$$

and

$$\widehat{pr}(T\le W)\widehat{r}_1(t)\widehat{A}(t)-pr(T\le W)\widehat{r}_1(t)A(t)=O_p(n^{-\beta/(2\beta+1)}).$$

Therefore, by Lemma 3.i),

$$\begin{aligned}&\widehat{pr}(Y=0,W<T)\widehat{r}_3(t)+\widehat{pr}(T\le W)\widehat{r}_1(t)\widehat{A}(t)-pr(Y=0,W<T)r_3(t)-pr(T\le W)r_1(t)A(t)\\&=O_p(n^{-\beta/(2\beta+1)}).\end{aligned}$$

Similarly,

$$\begin{aligned}&\widehat{pr}(Y=0,W<T)\widehat{r}_3(t)+\widehat{pr}(T\le W)\widehat{r}_1(t)-pr(Y=0,W<T)r_3(t)-pr(T\le W)r_1(t)\\&=O_p(n^{-\beta/(2\beta+1)}).\end{aligned}$$

By Lemma 3.iii),

$$\begin{aligned}&\frac{1}{\widehat{pr}(Y=0,W<T)\widehat{r}_3(t)+\widehat{pr}(T\le W)\widehat{r}_1(t)}-\frac{1}{pr(Y=0,W<T)r_3(t)+pr(T\le W)r_1(t)}\\&=O_p(n^{-\beta/(2\beta+1)}).\end{aligned}$$

By Lemma 3.i) again,

$$\begin{aligned}&\frac{\widehat{pr}(Y=0,W<T)\widehat{r}_3(t)+\widehat{pr}(T\le W)\widehat{r}_1(t)\widehat{A}(t)}{\widehat{pr}(Y=0,W<T)\widehat{r}_3(t)+\widehat{pr}(T\le W)\widehat{r}_1(t)}-\frac{pr(Y=0,W<T)r_3(t)+pr(T\le W)r_1(t)A(t)}{pr(Y=0,W<T)r_3(t)+pr(T\le W)r_1(t)}\\&=O_p(n^{-\beta/(2\beta+1)}).\end{aligned}$$

In other words, $\widehat{F}_n(t)-F(t)=O_p(n^{-\beta/(2\beta+1)})$, which complete the proof of Theorem 2. □

A.5 Details on Example 1

Assuming that T follows an exponential distribution with rate θ and W follows an exponential distribution with rate *gamma* then the likelihood $L(D;\theta,\gamma)$ is

$$\begin{aligned}\prod_{i=1}^{n}\Big(&\left\{\gamma\exp(-\gamma U_i)\exp(-\theta U_i)\right\}^{1-\Delta_i}\left\{q(U_i)\theta\exp(-\theta U_i)\exp(-\gamma U_i)\right\}^{\Delta_i Y_i}\\&\times\left[\gamma\exp(-\gamma U_i)\int_0^{U_i}\{1-q(s)\}\,\theta\exp(-\theta s)\right]^{\Delta_i(1-Y_i)}\Big).\end{aligned}$$

Hence,

$$\begin{aligned}&\log L(D;\theta,\gamma)\\&=\sum_{i=1}^{n}(1-\Delta_i)(\log\gamma-\gamma U_i)+\sum_{i=1}^{n}\Delta_i Y_i(-\gamma U_i)+\sum_{i=1}^{n}\Delta_i(1-Y_i)(\log\gamma-\gamma U_i)+C(U_1,U_2,...,U_n;\theta)\\&=(n-\sum_{i=1}^{n}\Delta_i Y_i)\log\gamma-\gamma\sum_{i=1}^{n}U_i+C(U_1,U_2,...,U_n;\theta).\end{aligned}$$

Therefore,

$$\frac{\partial\log L(D;\theta,\gamma)}{\partial\gamma}=\frac{n-\sum_{i=1}^{n}\Delta_i Y_i}{\gamma}-\sum_{i=1}^{n}U_i,$$

and hence, the value of γ which maximize $L(D;\theta,\gamma)$ is $\widehat{\gamma}_n=\frac{n-\sum_{i=1}^{n}\Delta_i Y_i}{\sum_{i=1}^{n}U_i}$.

The partial likelihood $L_{partial}(D;\theta;\gamma)$ of category $C=1$,

$$\prod_{i=1}^{n}\left\{\frac{g(U_i;\gamma)\overline{F}(U_i;\theta)}{\int_0^\infty g(s;\gamma)\overline{F}(s;\theta)ds}\right\}^{1-\Delta_i}$$
$$=\prod_{i=1}^{n}\left\{\frac{\gamma\exp(-\gamma U_i)\exp(-\theta U_i)}{\int_0^\infty \gamma\exp(-\gamma s)\exp(-\theta s)ds}\right\}^{1-\Delta_i}=\prod_{i=1}^{n}\left\{\frac{\gamma\exp\left(-U_i(\gamma+\theta)\right)}{\int_0^\infty \gamma\exp(-s(\gamma+\theta))ds}\right\}^{1-\Delta_i}$$
$$=\prod_{i=1}^{n}\left\{\frac{\gamma\exp\left(-U_i(\gamma+\theta)\right)}{\frac{\gamma}{\gamma+\theta}}\right\}^{1-\Delta_i}=\prod_{i=1}^{n}[(\gamma+\theta)\exp\left(-U_i(\gamma+\theta)\right)]^{1-\Delta_i}.$$

Hence,

$$\log L_{partial}(D;\theta;\gamma)=\sum_{i=1}^{n}(1-\Delta_i)\left(\log(\gamma+\theta)-U_i(\gamma+\theta)\right).$$

Therefore,

$$\frac{\partial\log L_{partial}(D;\theta,\gamma)}{\partial\theta}=\sum_{i=1}^{n}(1-\Delta_i)(\frac{1}{\gamma+\theta}-U_i).$$

The semiparametric estimator for θ is the maximizer of $L_{partial}(D;\theta,\widehat{\gamma}_n)$ by θ which is

$$\widehat{\theta}_n=\frac{\sum_{i=1}^{n}(1-\Delta_i)}{\sum_{i=1}^{n}U_i(1-\Delta_i)}-\widehat{\gamma}_n=\frac{\sum_{i=1}^{n}(1-\Delta_i)}{\sum_{i=1}^{n}U_i(1-\Delta_i)}-\frac{n-\sum_{i=1}^{n}\Delta_i Y_i}{\sum_{i=1}^{n}U_i}.$$

References

1. Minn AY, Pollock BH, Garzarella L, et al. Surveillance neuroimaging to detect relapse in childhood brain tumors: a Pediatric Oncology Group study. *Journal of clinical oncology* 2001; 19(21): 4135–4140.
2. Whitehead J. The analysis of relapse clinical trials, with application to a comparison of two ulcer treatments. *Statistics in Medicine* 1989; 8(12): 1439–1454.
3. Mandelbaum A, Zeltyn S. Service engineering in action: The Palm/Erlang-A queue, with applications to call centers. *In Advances in Services Innovations, D. Spath and K. P. Fahnrich (eds.). Springer, New York,* 2007: 17–45.
4. Peto R. Experimental survival curves for interval-censored data. *Journal of the Royal Statistical Society: series C (Applied Statistics)* 1973; 22(1): 86–91.
5. Zhang Z, Sun J. Interval censoring. *Statistical methods in medical research* 2010; 19(1): 53–70.
6. Baker DW, Stevens CD, Brook RH. Patients who leave a public hospital emergency department without being seen by a physician. Causes and consequences. *JAMA* 1991; 266(8): 1085–1090.
7. Hunt KA, Weber EJ, Showstack JA, Colby DC, Callaham ML. Characteristics of frequent users of emergency departments. *Annals of Emergency Medicine* 2006; 48(1): 1–8.
8. Bolandifar E, DeHoratius N, Olsen T, Wiler JL. Modeling the Behavior of Patients Who Leave the ED Without Being Seen. *Chicago Booth Research Paper* 2014(12–14).
9. Batt RJ, Terwiesch C. Waiting patiently: An empirical study of queue abandonment in an emergency department. *Management Science* 2015; 61(1): 39–59.
10. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer . 2013.
11. Wilson JMG, Jungner G, Organization WH, others . Principles and practice of screening for disease. 1968.
12. Zelen M, Feinleib M. On the theory of screening for chronic diseases. *Biometrika* 1969; 56(3): 601–614.

13. Prentice RL, Gloeckler LA. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 1978: 57–67.

14. Gans N, Koole G, Mandelbaum A. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management* 2003; 5(2): 79–141.

15. Aksin Z, Armony M, Mehrotra V. The modern call center: A multi-disciplinary perspective on operations management research. *Production and operations management* 2007; 16(6): 665–688.

16. Mandelbaum A, Zeltyn S. Data-stories about (im) patient customers in tele-queues. *Queueing Systems* 2013; 75(2-4): 115–146.

17. Brown L, Gans N, Mandelbaum A, et al. Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American statistical association* 2005; 100(469): 36–50.

18. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; 53(282): 457–481.

19. Wiler JL, Bolandifar E, Griffey RT, Poirier RF, Olsen T. An emergency department patient flow model based on queueing theory principles. *Academic Emergency Medicine* 2013; 20(9): 939–946.

20. Nah F. A study on tolerable waiting time: how long are web users willing to wait?. *Behaviour & Information Technology* 2004; 23(3): 153–163.

21. Yom-Tov GB, Rafaeli A, Ashtar S, et al. Customer emotion in chat services: Automatic identification and new insights. *Under review* 2018.

22. Wong WH. Theory of partial likelihood. *The Annals of statistics* 1986: 88–123.

23. Kosorok MR. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York . 2008.

24. Tsybakov AB. *Introduction to Nonparametric Estimation*. Springer Science & Business Media, New York . 2008.

25. Shaikh SB, Jerrard DA, Witting MD, Winters ME, Brodeur MN. How long are patients willing to wait in the emergency department before leaving without being seen?. *Western Journal of Emergency Medicine* 2012; 13(6): 463.

26. Vaart v. dAW. *Asymptotic Statistics*. Cambridge University Press, Cambridge . 2000.