Statistics and Computing (2026) 36:172
https://doi.org/10.1007/s11222-026-10921-w

ORIGINAL PAPER

Nonparametric estimation of the joint and conditional survival functions of the time to an event of interest and associated integrated covariate processes

Ashwini Joshi[1] · **Dario Gasbarra**[2] · **Sangita Kulathinal**[1]

Received: 13 November 2025 / Accepted: 10 June 2026 / Published online: 10 July 2026

Abstract

During the treatment of chronic diseases, a covariate process may exhibit a response over the treatment period and may also be associated with the event of interest. The integrated covariate process commonly known as the area under the response curve, over a specific treatment interval is a widely used measure of a cumulative treatment response in medical research. Our interest is in the survival probability that the time to the event is larger than a fixed time t and the cumulative response up to that time is larger than a fixed value y. Because the cumulative response grows with time, the latter event of the survival probability may realise at the random time when the integrated process crosses y. A common censoring mechanism may censor both these random times. In fact, our setting gives rise to a different type of censoring mechanism of the bivariate event times. In order to handle the censoring and to use all available information, we study inverse-probability of censoring weighted estimators of the survival probability. The estimators and their efficiency differ according to the use of the information in estimation of the censoring distribution. We also give a pooled estimator of the bivariate survival function in a stratified analysis. Further, we study the conditional survival function of the time to the event given the history of the covariate process and discuss its use in the medical decision making. We suggest jackknife method for estimating variances for all the proposed estimators. The proposed estimators can be easily generalised to more than one covariate processes. We illustrate our methodology using two sets of data on the age-related macular degeneration of eye disease; one from a clinical trial and the other from a real-world study.

Keywords Bivariate survival function · Integrated covariate process · Area under the curve · Inverse probability of censoring weights · Medical decision making · Treatment regime · Stratification

1 Introduction

During the treatment of chronic diseases, maintenance of a treatment response or responses to certain levels throughout the follow-up is necessary. In addition, events related to the disease condition or the treatment status are of interest. An integrated treatment response or associated covariate process is a good measure to gauge the changes in the time-dependent covariate over time until the event of interest occurs. Based on the covariate process studied, it might represent the integrated response, the quality-adjusted survival time, or medical cost among other possible aspects. The integrated covariate process is also known as the area under the curve or the cumulative response, and the cumulative response up to the time of the event of interest is known as a mark in the literature (Huang et al. 1998).

The motivation for this work comes from clinical visit data of the wet age-related macular degeneration of eye disease (wet-AMD) where a covariate process is the visual acuity (VA, a treatment response), measured over a treatment period and is expected to be associated with the discontinuation of

Dario Gasbarra and Sangita Kulathinal contributed equally to this work.

✉ Ashwini Joshi
ashwini.joshi@helsinki.fi

Dario Gasbarra
dario.gasbarra@uwasa.fi

Sangita Kulathinal
sangita.kulathinal@helsinki.fi

[1] Department of Mathematics and Statistics, University of Helsinki, Pietari Kalmin katu 5, Helsinki 00560, Finland

[2] Department of Mathematics and Statistics, University of Vaasa, Yliopistonranta 10, Vaasa 65200, Finland

Table 1 General notations

τ	Time to event of interest
C	Censoring time
T	Observed time $\tau \wedge C$
δ	Event indicator $\mathbf{1}(\tau \leq \mathbf{C})$ (1: event and 0: censoring)
$Z(.)$	Covariate process
$\Psi(t) = \int_0^t Z(u)du$	Integrated covariate process until time t
$\rho(y) = \inf\{t : \Psi(t) \geq y\}$	The first time $\Psi(t)$ reaches y
δ_y	Event indicator $\mathbf{1}(\rho(\mathbf{y}) \leq \tau \wedge \mathbf{C})$
$\widetilde{\tau}(t, y)$	Time to event of interest in the reconstructed data
$\widetilde{\delta}(t, y)$	Event indicator in the reconstructed data $\mathbf{1}(\widetilde{\tau}(\mathbf{t}, \mathbf{y}) \leq \mathbf{C})$
$S(t, y) = P(\tau > t, \Psi(\tau) > y)$	Bivariate survival probability of $(\tau, \Psi(\tau))$ for fixed (t, y)
$G(t)$	Censoring survival function
$\Lambda^C(t)$	Cumulative hazard function of C
IPCW	Inverse probability of censoring weight
n	number of observations
$S_n(t, y)$	IPCW estimator of $S(t, y)$ for given $G(.)$
$\widehat{G}(t)$	Kaplan-Meier estimator of $G(t)$ obtained by using data $(T, 1 - \delta)$
$\widehat{S}_n(t, y)$	Plug-in estimator of $S(t, y)$ obtained by replacing $G(.)$ by $\widehat{G}(.)$
$\widetilde{G}(t)$	Kaplan-Meier estimator of $G(t)$ obtained by using data $(T, 1 - \delta)$ as well as $\Psi(t), t \leq T$
$\widetilde{S}_n(t, y)$	Plug-in estimator of $S(t, y)$ obtained by replacing $G(.)$ by $\widetilde{G}(.)$.
$\widehat{\Lambda}(t \mid \Psi(t) \leq y)$	Nelson-Aalen type estimator of the conditional hazard of τ at time t given $\Psi(t) \leq y$

Table 2 Contribution of observations and determination of the value of the event $\mathbf{1}(\tau > t, \tau > \rho(y))$. Columns (i)-(iii) give three different ways of defining the indicator of the censoring event for estimation of the censoring survival function described in section 3.1

Quadrant[2]	(δ, δ_y)[3]	$\mathbf{1}(\tau > t, \tau > \rho(y))$[1]	(i)	(ii)	(iii)
(1)	(1, 1)	1	-	0	0
	(0, 1)		-	1	0
(2)	(1, 1)	0	-	0	0
	(0, 1)		-	1	1
(3)	(1, 0)	0	-	0	0
	(0, 0)		-	1	1
(4)	(1, 0)	0	-	0	0
	(0, 0)		-	1	1

[1] Here $(\tau, C, \rho(y))$ are times to event, censoring and stopping time, respectively.
[2]The four quadrants are shown in Fig. 1.
[3]$\delta = \mathbf{1}(\tau \leq C)$ and $\delta_y = \mathbf{1}(\rho(y) \leq \tau \wedge C)$

the treatment. Higher VA means better vision, and therefore, is considered as a good response to the treatment. The event of interest in this example is the discontinuation of the treatment. The idea behind integrating the covariate (treatment response or disease marker) process comes from the use of area under the response trajectory often used in the medical field. The reason behind this is that the area is a monotonic summary measure of the longitudinally measured covariate process. The area under the VA trajectory is commonly used for assessing the treatment response as well as for decision making for the treatment of AMD (Wilding et al. 2012; Gross et al. 2015). The computation of the area is difficult in the presence of death or censoring due to other reasons. In this paper, we study the event of discontinuation of the treatment and longitudinally measured VA response together by incorporating censoring, with inverse probability of censoring weighted method.

Time-dependent covariates and time to an event of interest have been studied extensively in the literature (Faucett and Thomas 1996; Ibrahim et al. 2010). Joint modeling (Tsiatis et al. 1995; Wang and Zhong 2025) and landmark modeling (Van Houwelingen 2007; Van Houwelingen and Putter 2011) are commonly used for modeling longitudinal trajectories and related event of interest together. The success of joint modelling relies on correctly specifying two sub-models; the longitudinal model and the survival model; and the link between them. Generally, a landmark model is used for prediction of future value using the marker value or slope of the trajectory at the landmark time. The use of only the latest value ignores the complete history of longitudinal measurements. The area under the trajectory gives an overall response (accounting for the history of the process) over a period of time as compared to a single measurement (ignores the history of the process). It is clear from Fig. 7 (from supplementary figures) that even if the VA values at 24 weeks were exactly the same for four patients (left panel, Fig. 7), the corresponding areas under the VA trajectory up to 24 weeks were quite different (right panel, Fig. 7). This is because the four paths leading up to week 24 differed.

For a clear comparison and guidance on when to use each of the approaches, we refer to Putter and Van Houwelingen (2022). In this article, we adopt a nonparametric approach for estimation of the joint survival probabilities as well as the conditional hazard functions which avoids specification of a particular model for the longitudinal trajectory of the covariate process.

Our interest is in the survival probability of the type $S(t, y) = P(\tau > t, \Psi(\tau) > y)$, for some $t, y > 0$, where τ

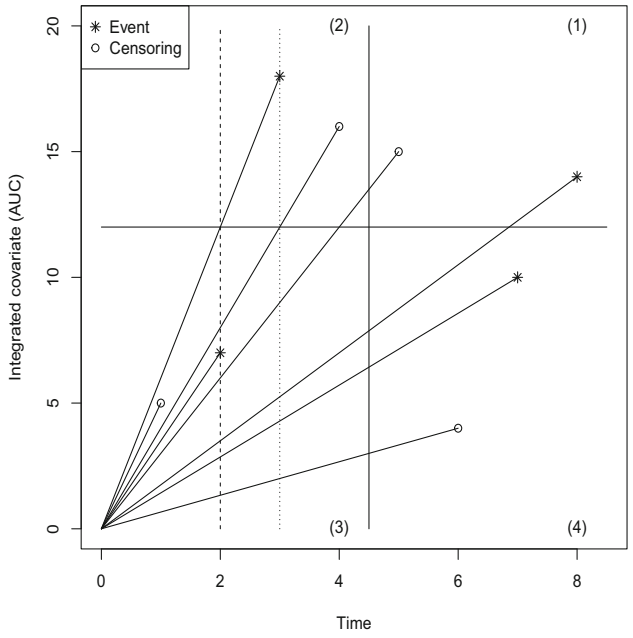


Fig. 1 Trajectory for eight hypothetical patients are plotted in this graph of transformed time (AUC) vs. original time. Time is on the X-axis and the integrated covariate value (the area under the covariate trajectory) is on the Y-axis. The solid vertical line near $x = 5$ indicates a fixed value of time t and the horizontal line near $y = 12$ indicates the fixed value of the integrated covariate y. The two dimensional plot area is divided into four quadrants. One event time and one censoring time are plotted in each quadrant for illustration purposes. The dotted and dashed vertical lines are shown to illustrate the risk sets for the estimators of the cumulative hazards with respect to the events times observed in quadrants (2) and (3).

is the time to the event of interest and $\Psi(\tau)$ is the integrated response up to τ. The probability $S(t, y)$ combines the information about the event and the treatment response over time. In practice, a common censoring mechanism may result in the censoring of both times. Also, the event $(\Psi(\tau) > y)$ may be observed even before the actual event time τ. This is because the cumulative response grows with time, and it may cross the value specified in the survival probability before the event occurs. The time when the cumulative response or the integrated covariate process crosses a specified value for the first time is random and is known as the stopping time. The time to the event of interest may also censor the stopping time. An estimator of the survival probability must make use of all available information over time. In order to facilitate medical decision making, the conditional cumulative hazard of τ given $\Psi(t)$ is of relevance. The problem of estimation of $S(t, y)$ bears several similarities to the widely studied field of estimation of the bivariate survival or distribution functions. However, there are subtle differences between the two. We first provide a brief review of the nonparametric estimators of the bivariate distributions based on the literature that come close to our setting. The empirical distribution function is the commonly used naive estimator for estimating the bivariate distribution when there is no censoring. Various estimators are studied in the presence of a univariate or bivariate censoring mechanism. Some are direct extensions of the Kaplan-Meier estimator, for example see Dabrowska (1988). These estimators depend on the censoring scheme in the bivariate setting; univariate independent censoring (Huang et al. 1998; Lin 1993) where both event times are censored by a common censoring variable, bivariate independent censoring where each event time is censored by own independent censoring variable and the two censoring variables can be modelled jointly (Janssen et al. 2023) or a combination of estimators of the conditional distribution or cumulative hazard and the marginal distribution (Akritas and Keilegom 2003).

The nonparametric estimation of the joint distribution function of $(\tau, \Psi(\tau))$ using the data $(\tau \wedge C, I(\tau \leq C), \Psi(\tau) I(\tau \leq C))$ has been studied in Huang et al. (1998). In their setting, the mark variable $\Psi(\tau)$ is observed only if the event is observed and τ is subjected to an independent right censoring by C. The authors have ignored the history of the associated covariate process in their estimator because of consistent estimation of $S(t, y)$ over the entire range. On the other hand, a series of consistent and efficient estimators of the survival function of the quality-adjusted life time $\Psi(\tau)$ have been suggested where the covariate process history have been used in inverse-probability of censoring weighted (IPCW) estimators (Zhao and Tsiatis 1997, 1999). On similar lines, various estimators of cumulative medical costs, in the presence of censoring, have also been studied using the history of the cost of the treatment (Zhao et al. 2007).

Our second problem related to the estimation of the conditional cumulative hazard of τ given that the integrated covariate process $\Psi(t)$ has crossed (or not crossed) a specific value y has similarities to the estimation of the cumulative hazard in nonparametric survival regression with censored data. Again, there are subtle differences because of the special setting that we consider. The nonparametric estimation problems of the cumulative hazard discussed in Akritas and Keilegom (2003); Van Keilegom et al. (2001); Keilegom and Veraverbeke (2001) are the most relevant for our setting.

In this paper, we propose nonparametric estimators of the survival probability $S(t, y)$ and clearly exhibit how each observation contributes to the estimators. We also provide asymptotic theory for the proposed estimators. We compare the true variances with jackknife estimators. We also provide a stratified estimation of the survival probability for different subgroups created based on baseline covariate values. Finally, we propose estimators of the conditional survival functions of the time to the event of interest, given specific events related to the integrated covariate process. We then illustrate their applications in medical decision making prob-

lems for AMD. With this approach, at each patient visit, the possibility of discontinuation (or continuation) of the treatment can be examined conditioned on the VA history of that patient until that visit.

The paper is structured as follows. In sect. 2, we describe the notations and formulate the problem. Sect. 3 gives an IPCW estimator of $S(t, y)$ when the censoring survival function is known and then discusses alternatives when it needs to be estimated. Sect. 4 extends the method to stratified analysis, sect. 5 discusses estimation of the conditional survival functions and sect. 6 illustrates our proposed estimation approaches with one dataset from a clinical trial and one real-world dataset of AMD. The paper ends with a general discussion presented in Sect. 7.

2 Notations

Let τ be the time to an event of interest and $\{Z(t), t \geq 0\}$ be an associated covariate process. We define the integrated covariate process as $\Psi(t)$ and the corresponding random time (stopping time) $\rho(y)$ such that $\Psi(t)$ crosses y for the first time as:

$$\Psi(t) = \int_0^t Z(s)ds, \text{ and } \quad \rho(y) = \inf\{t : \Psi(t) \geq y\}. \quad (1)$$

Note that $\Psi(t)$ is interpreted as the area under the covariate process over the time interval $[0, t]$. Further, $\Psi(t)$ can also be interpreted as the transformed time to the event with respect to the covariate process. Table 1 lists general notations used in this paper.

As mentioned earlier, we are interested in the estimation of the survival probability of $(\tau, \Psi(\tau))$ for fixed $t, y > 0$ defined as:

$$\begin{aligned} S(t, y) &= P(\tau > t, \Psi(\tau) > y) = P(\tau > t, \tau > \rho(y)) \\ &= P(\tau > t \vee \rho(y)). \end{aligned}$$

We emphasise that the bivariate survival function of $(\tau, \Psi(\tau))$ for fixed (t, y) can be studied using the bivariate random times $(\tau, \rho(y))$ or more precisely, using the events $\mathbf{1}(\tau > t, \tau > \rho(y))$. In practice, both these random times are subjected to censoring. We denote the censoring time as C, the observed time as $T = \tau \wedge C$ and the event indicator $\delta = \mathbf{1}(\tau \leq C)$. The stopping time $\rho(y)$ is censored by T and we define $\delta_y = \mathbf{1}(\rho(y) \leq T)$ as an indicator function assuming value 1 when $\rho(y)$ is observed and 0 otherwise. In case when $\delta_y = 0$, we use the convention that $\rho(y) = \infty$. This is just to indicate that the integrated covariate process has not crossed the level y in a finite interval $[0, T]$. Further, we assume that the censoring is independent of $(\tau, Z(t), t \geq 0)$.

The process $\Psi(t)$ increases as the follow-up time increases, thereby making it highly dependent on the time of follow-up for each patient. To remove this dependency we use $\Psi(t)$ until a fixed time, say L. The event times are understood as $\tau = \tau \wedge L$ and $\Psi(\tau) = \int_0^{\tau \wedge L} Z(s)ds$. Such artificial fixed end of the study is suggested in the literature for various reasons such as for reducing the bias in Kaplan-Meier estimation of transformed time (Gelber et al. 1989) or for satisfying the condition of the uncensored largest observation for the inverse probability of censoring weighted methods (Zhao and Tsiatis 1997).

3 Estimation of $S(t, y)$

In this section, we consider estimation of $S(t, y)$ using observations $\{T_i, \delta_i, \rho_i(y) \wedge T_i, \delta_{yi}\}$ on $i = 1, \ldots, n$ independent patients. We develop our estimator for fixed (t, y) such that $P(\tau \wedge C > t \vee \rho(y)) \neq 0$. Let the survival function of C be G. An inverse probability of censoring weighted (IPCW) estimator of $S(t, y)$ can be defined as follows.

$$S_n(t, y) = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\mathbf{1}(T_i > t \vee \rho_i(y))}{G(t \vee \rho_i(y))} \right]. \quad (2)$$

Note that observation i contributes to the sum only when $(T_i > t \vee \rho_i(y))$. Hence, $(T_i > \rho_i(y))$ i.e., $\delta_{yi} = 1$ and $\rho_i(y)$ is observed. The estimator (2) is simple to compute and reduces to the empirical probability in the absence of censoring (taking $C \equiv \infty$). Further, it may seem a simple consequence of the independence of τ and C since $P(T > s) = P(\tau > s)P(C > s)$ but the heuristic idea behind it lies in the literature on IPCW estimators of survival functions (Satten et al. 2001; Robins and Rotnitzky 1992). The estimator (2) is unbiased provided $(\tau, \Psi(\tau))$ are independent of C. The independent censoring assumption implies that the censoring may depend on the measured covariates but not on the unmeasured ones. Even though the estimator (2) is constructed without any covariates, when the additional measured covariates are present and censoring depends on the covariates, then the estimator (2) needs to be adjusted. One way to do the adjustment is to carry out the stratified analysis presented in sect. 4 where the strata are defined according to the covariate values. It is to be noted that the positivity of the censoring survival function for each stratum, that is, $G_k(t) > 0$ for all t of interest is required. Another approach is to adjust (2) using, for example, an additive model where the covariates are included. In order to ensure the monotonicity of the estimator some adjustment as used in isotonic regression can be employed. For more detailed discussion on the adjustment under dependent censoring, please refer to (sec-

tions 4.2.6 and 4.2.7, Aalen et al. (2008)) and (section 3.4.2, Cook and Lawless (2018)).

In Table 2 and Fig. 1, we exhibit how each observation contributes to the estimator. The patients in quadrant (1) of Fig. 1 contribute to the numerator. The contribution of patients to the denominator depends on the way the censoring survival function is estimated (see section 3.1).

3.1 Handling of the censoring survival function *G* in (2)

The censoring distribution in (2) may not be known in practice and needs to be estimated. We replace the unknown censoring distribution by its consistent estimator, and define the plug-in estimators of $S(t, y)$. The efficiency of the plug-in estimator depends on the information used in estimating the censoring survival function. In our setting, the censoring survival function can be estimated consistently in different ways and the efficiency of the plug-in estimator would depend on the estimator used. The standard Kaplan-Meier estimator of the censoring survival function considers $\delta_i = 0$ as a censoring event (that is $C_i < \tau_i$), but it ignores the information contained in $\rho_i(y)$. Here we consider three different ways of handling the censoring survival function, and denote the corresponding estimators of $S(t, y)$ as $S_n(t, y)$, $\widehat{S}_n(t, y)$ and $\widetilde{S}_n(t, y)$.

(i) **Using external data.** If censoring distribution of the underlying bigger population, from which the analysis data is sampled, is known, then it can be used directly, avoiding the estimation altogether. This is particularly of interest when the censoring is due to death and can be taken as an independent censoring with regards to the event of interest. In this case, the population mortality distribution can be used and $S(t, y)$ can be estimated using equation (2). Since the censoring distribution is known and need not be estimated from the observed data, the contribution of the observed data is not applicable as shown in the column (i) of Table 2.

(ii) **Using the information only from** (T, δ). The standard Kaplan-Meier (KM) estimator $\widehat{G}$ using censoring as an event of interest is the estimator of G in this case.

$$\widehat{G}(t) = \prod_{s \le t}(1 - \Delta\widehat{\Lambda}_n^C(s))$$

is the Kaplan-Meier estimator of the survival function $G(t)$ and

$$\widehat{\Lambda}_n^C(t) = \sum_{i=1}^{n} \frac{\mathbf{1}(T_i \le t)(1-\delta_i)}{\sum_{j=1}^{n}\mathbf{1}(T_j \ge T_i)}$$

is the Nelson-Aalen estimator of the cumulative hazard of C. This estimator is consistent and computationally efficient. Please note that the information regarding censoring from $C_i < \rho_i(y)$ is ignored. The indicator of the censoring event for each observation in the Fig. 1 is given in the column (ii) of Table 2. The bivariate survival estimator by using this censoring survival estimator is given as

$$\widehat{S}_n(t, y) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathbf{1}(T_i > t \vee \rho_i(y))}{\widehat{G}(t \vee \rho_i(y))}\right]. \tag{3}$$

The marginal survival estimator of the estimator (3) for $y = 0$, $\widehat{S}_n(t, 0)$ is equivalent to the Kaplan-Meier estimator (Appendix A). The idea behind the estimator (3) appears similar to the one behind a simple estimator of the bivariate survival function proposed in Lin and Ying (1993). They considered a pair of failure times (X, Y) distributed according to $S(x, y) = P(X > x, Y > y)$, the pair was subjected to censoring by an independent random variable C, and the observations were $(X_i \wedge C_i, \mathbf{1}(X_i \le C_i), Y_i \wedge C_i, \mathbf{1}(Y_i \le C_i))$, $i = 1, \ldots, n$. Observations with either X or Y or both censored contributed to the censoring event and were used in the Kaplan-Meier estimator of the censoring survival function. In our case, the stopping time $\rho(y)$ can be censored by $\tau \wedge C$. This would then mean that the random variable X (from the pair in Lin and Ying (1993)) is subjected to censoring by C while Y is subjected to censoring by $X \wedge C$. In this sense, the censoring mechanism in our case is different from the commonly studied univariate censoring schemes in the literature. Estimator (3) is asymptotically efficient when the entire survival curve is of interest. When the interest is in the survival probability for specific (t, y), this may not be the case. A more efficient estimator may be obtained by regenerating the data as discussed in (iii) below. In general, the efficiency would depend on the percentage of censoring and the association between the event of interest and the covariate process.

(iii) **Using information from** (T, δ) **as well as from** $\Psi(t), 0 \le t \le T$. Here we use the information about the observed stopping time $\rho_i(y)$ along with (T_i, δ_i). This can be done by recognising the censoring events only from those who were censored before (t, y), that is the patients who satisfy the condition that either $T \le t$ or $\psi(T) \le y$ and $\delta = 0$. This can be achieved by marking all other patients not satisfying these conditions, that is, with $T > t$ and $\psi(T) > y$ as the event. We reconstruct the data for each fixed pair (t, y) where the event indicator and observation times are redefined as

follows.

$$\widetilde{\tau}_i(t,y) = \tau_i \wedge (t \vee \rho_i(y)),\ \widetilde{T}_i(t,y) = (\widetilde{\tau}_i(t,y) \wedge C_i),$$
$$\widetilde{\delta}_i(t,y) = \begin{cases} 1 & \text{when } C_i \geq \widetilde{\tau}_i(t,y) \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

We denote the Kaplan-Meier estimator of G obtained using the reconstructed data as $\widetilde{G}$ where $(\widetilde{\delta}_i(t,y) = 0)$ is the event in estimation of G and give the corresponding estimator of the joint survival function below.

$$\widetilde{S}_n(t,y) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\mathbf{1}(T_i > t \vee \rho_i(y))}{\widetilde{G}(t \vee \rho_i(y))}\right]. \quad (5)$$

It is to be noted that $\widetilde{G}$ depends on $(t,\ y)$ unlike known G and $\widehat{G}$. We show the event indicator of the bivariate survival for each patient in Fig. 1 in the column $\mathbf{1}(\tau > t, \tau > \rho(y))$ of Table 2 and the indicator of the censoring event for estimation of the censoring survival function in the column (iii). Those with $\mathbf{1}(\tau > t, \tau > \rho(y)) = 1$ contribute to the numerator of the bivariate survival function and those with indicator value 1 in column (iii) are considered as censoring events while estimating the censoring survival distribution of the reconstructed data. The estimator $\widetilde{S}_n(t,y)$ obtained using $\widetilde{G}$ is expected to give a point-wise efficient estimator but at the cost of high computation time, due to the reconstruction for each choice of (t,y), especially for large data.

Zhao and Tsiatis (1997) suggested similar reconstruction of data for the estimation of the survival function of the quality-adjusted survival time (in our case, the marginal survival function of $\Psi(\tau)$). The reconstruction of the data used in their paper was done for fixed y only. The marginal estimator of survival function of $\Psi(\tau)$ is $S_n(0,y)$ and its version $\widetilde{S}_n(0,y)$ is equivalent to the estimator of the survival function of the quality adjusted survival time in Zhao and Tsiatis (1997). For brevity, the proof is given in Appendix B. It is to be noted that our proposed estimator (2) is defined in terms of the observed quantities while the estimator of the survival function of $\Psi(\tau)$ in Zhao and Tsiatis (1997) involves the original times, which are not directly observed, although the value of the event is observed. In principle, we can estimate G using the reconstructed data as in Zhao and Tsiatis (1997) to obtain yet another estimator of $S(t,y)$. However, we will not use it in our illustrative examples.

When computing (3) and (5), we have adopted a model without any covarites. When the censoring depends on the measured covariates, any of the two approaches, stratification or regression, mentioned earlier in connection with (2), can also be adopted here. It is important to note that in either of these cases, the marginal estimator of $\Psi(\tau)$ is not necessarily monotonic and thus, useful only in the point-wise estimation for a specific value of y. The univariate Kaplan-Meier estimator of $\Psi(\tau)$ or bivariate nonparametric estimators like Huang and Louis (1998) use information from the uncensored cases $\delta = 1$. The marginal estimators of $\Psi(\tau)$ obtained from these estimators are monotonic but not efficient, since they do not use the information from the censored observations that reach the integrated covariate value y.

3.2 Asymptotic distributions

The asymptotic distributions of the proposed estimators depend on the plug-in estimator of the censoring survival function. When the censoring distribution is known, the asymptotic result follows from the theory of empirical processes (Fleming and Harrington 1991). The following theorem gives the asymptotic normality of the estimator $S_n(t,y)$.

Theorem 1 *For (t,y) such that $P(\tau \wedge C > t \vee \rho(y)) \neq 0$, as $n \to \infty$, $\sqrt{n}\{S_n(t,y) - S(t,y)\}$ converges in law to a zero-mean normal variate with variance*

$$\sigma_1^2(t,y) = E\left[\frac{\mathbf{1}(\tau > t \vee \rho(y))}{G(t \vee \rho(y))}\right] - \{S(t,y)\}^2.$$

Proof Please see Appendix C. □

In the absence of censoring, the variance corresponds to the variance of the empirical estimator of the survival function, that is $S(t,y)\ (1 - S(t,y))$.

The variance $\sigma_1^2(t,y)$ involves unknown survival function and it can be estimated as follows.

$$\widehat{\sigma}_1^2(t,y) = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(T_i > t \vee \rho_i(y))}{\{G(t \vee \rho_i(y))\}^2} - \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(T_i > t \vee \rho_i(y))}{G(t \vee \rho_i(y))}\right)^2.$$

Alternatively, a resampling method such as the jackknife method can be used to estimate the variance. The jackknife variance estimator of $S_n(t,y)$ is given as follows.

$$\sigma_{1JK}^2(t,y) = \frac{n-1}{n}\sum_{i=1}^{n}\left(S_{[n]\setminus i}(t,y) - \bar{S}_n(t,y)\right)^2, \quad (6)$$

where $S_{[n]\setminus i}(t, y)$ is as in (2) but computed by removing i^{th} patient and

$$\bar{S}_n(t, y) = \frac{1}{n}\sum_{i=1}^{n} S_{[n]\setminus i}(t, y).$$

When the censoring distribution is estimated using the standard KM estimator, we can establish the asymptotic distribution of (3) for fixed (t, y) using the theory of counting processes and martingale (Andersen et al. 1992).

Theorem 2 *As $n \to \infty$, for fixed (t, y) such that $P(\tau \wedge C > t \vee \rho(y)) \neq 0$, $\sqrt{n}\{\widehat{S}_n(t, y) - S(t, y)\}$ converges to a zero-mean normal variate with variance*

$$\sigma_2^2(t, y) = var\left(\frac{\mathbf{1}(T > t \vee \rho(y))}{G(t \vee \rho(y))} + \int_0^{\infty} w(u) dM^C(u)\right),$$

where $M^C(t) = \mathbf{1}(T \leq t, \delta = 0) - \int_0^t \mathbf{1}(T \geq u) d\Lambda^C(u)$ and

$$w(u) = \frac{1}{r(u)} E\left[\mathbf{1}(u \leq t \vee \rho(y)) \frac{\mathbf{1}(T > t \vee \rho(y))}{G(t \vee \rho(y))}\right]. \quad (7)$$

Proof For fixed (t, y), we sketch the proof in Appendix D, following the theory of counting processes, martingale, and the usual regularity conditions that are applied to establish asymptotic behaviour of the Kaplan-Meier or the Nelson-Aalen estimator, in Andersen et al. (1992). □

We also give an estimator of $\sigma_2^2(t, y)$ in Appendix D. Our proof is similar to the proof given in connection to the U-statistics for censored data in Appendix A.2, Datta et al., (2010).

The asymptotic normality of (5) follows exactly on the same lines as in Theorem 2 by defining the censoring counting process, the appropriate filtration and the corresponding martingale for the reconstructed data.

Theorem 3 *As $n \to \infty$, for fixed (t, y) such that $P(\tau \wedge C > t \vee \rho(y)) \neq 0$, $\sqrt{n}\{\widetilde{S}_n(t, y) - S(t, y)\}$ converges to a zero-mean normal variate with variance*

$$\sigma_3^2(t, y) = var\left(\frac{\mathbf{1}(T > t \vee \rho(y))}{G(t \vee \rho(y))} + \int_0^{\infty} w(u) d\widetilde{M}^C(u)\right),$$

where $\widetilde{M}^C(u) = \mathbf{1}(\widetilde{T}(t, y) \leq u, \widetilde{\delta}(t, y) = 0) - \int_0^u \mathbf{1}(\widetilde{T}(t, y) \geq s) d\Lambda^C(s)$ and $w(u)$ is as defined in (7).

The jackknife variance estimates $\sigma_{2JK}^2(t, y)$ and $\sigma_{3JK}^2(t, y)$ of $\sigma_2^2(t, y)$ and $\sigma_3^2(t, y)$ can be obtained as in (6).

4 Stratified analysis

As mentioned above, $\Psi(t)$ may be highly influenced by the baseline measurements of the covariate process (initial level of the covariate). Those who start with a high baseline covariate value, have a higher integrated value compared to others. Without any loss of generality, we assume that higher values of the covariate indicate better prognosis and good response. In such cases, separate analyses of strata or subgroups, defined based on the underlying medical condition or prognosis or baseline values may be more meaningful.

The stratified analyses can help in understanding the differences among the strata and the effect of each stratum on the overall result. We propose a pooled estimator of the joint survival function $S(t, y)$ as the weighted average of the stratum-specific estimator of $S(t, y)$ where the weight for each stratum is the inverse of the variance of the stratum-specific estimate.

For K strata with n_k number of patients in each stratum, $k = 1, \ldots, K$, such that $\sum_k n_k = n$, let

$$\widehat{S}_{n_1}^1(t, y), \widehat{S}_{n_2}^2(t, y), \ldots, \widehat{S}_{n_K}^K(t, y),$$

be the estimators of the stratum-specific joint survival functions (see equation (3)) and the corresponding jackknife variance estimates be

$$\sigma_{2JK,1}^2(t, y), \sigma_{2JK,2}^2(t, y), \ldots, \sigma_{2JK,K}^2(t, y).$$

The weights assigned to each stratum are

$$\hat{w}_k = (1/\sigma_{2JK,k}^2(t, y))/(\sum_{l=1}^{K} 1/\sigma_{2JK,l}^2(t, y)), k = 1, \ldots, K.$$

The stratified estimate of the survival probability at (t, y) can then be obtained as

$$\widehat{S}^{pooled}(t, y) = \sum_{k=1}^{K} \hat{w}_k \widehat{S}_{n_k}^k(t, y). \quad (8)$$

The above pooled estimator need not be monotone and some adjustments can be done to achieve monotonicity (section 4.2.6, Aalen et al. (2008)). The estimator (8) is a consistent estimator of $S^{pooled}(t, y) = \sum_{k=1}^{K} w_k S^k(t, y)$, where w_k is the limit in probability of $\hat{w}_k$. The pooled survival function $S^{pooled}(t, y)$ may be useful in capturing the heterogeneity of the treatment response and discontinuation between strata. The precision with which each of $\widehat{S}_{n_k}^k(t, y)$ is estimated depends on the size of the stratum and the proportion of censoring in the stratum. In practice, such pooling should be avoided, especially when the censoring mechanism may vary between the strata and each stratum specific survival

function should be studied separately. A more stable estimator can be obtained by first obtaining the pooled estimator using $\widehat{\Lambda}^{k}_{n_k}(t, y) = \log(-\log(\widehat{S}^{k}_{n_k}(t, y)))$ and weights corresponding to its variance, and then transforming it to obtain an estimator of $\widehat{S}^{pooled}(t, y) = \exp(-\exp(\widehat{\Lambda}^{pooled}(t, y)))$. This will preserve the monotonicity of the survival curve. Survival curves from different strata can be compared using methods suggested in the literature (Giurcanu and Karrison 2025). Also, the unstratified estimate from the full data and the stratified estimate can be compared to see the effect of stratification. In principle, any estimator of $S(t, y)$ along with the corresponding variance can be used in (8) for the stratified analysis. When the censoring depends on the measured covariates, an alternative to the stratification is to compute pooled estimator simply by using a weighting procedure where individual counting and at-risk processes are weighted according to the censoring survival estimate and their covariate values. The censoring hazard rate estimate is obtained by using, for example, an additive model (provided the fit is good) and then the corresponding survival function is estimated. The pooled estimator thus obtained will always be monotone. We refer to (section 4.2.7, Aalen et al. (2008)) for more details.

5 Estimation of the conditional survival function

In previous sections, we discussed the estimation of the bivariate survival function of $(\tau, \Psi(\tau))$. For estimating $S(t, y)$, we used $\rho(y)$, the first time when the integrated covariate process crossed y and compared it with the observation times τ or C. In this sense, we used the history of the covariate process in the estimation.

In a real-world setting, medical practitioners have to make decisions regarding the discontinuation (or the continuation or the change) of a treatment at any time t during the treatment follow-up based on the treatment response up to that time point $\psi(t)$. For this purpose, the quantity of interest is the conditional survival function of τ given a conditioning event in terms of $\Psi(t)$. In principle, we can condition on any predictable event and study the survival function of τ given the history $\mathcal{F}_t$ generated by the data available up to time t. Note that $\Psi(t)$ is a (continuous) $\mathbb{F}$-predictable non-decreasing process.

We first consider the conditioning event $\Psi(t) \leq y$ or equivalently $t \leq \rho(y)$ and note that the three stopping times, the time to the event of interest τ, the censoring time C and $\rho(y)$ the first time when the integrated covariate process crosses y, are all $\mathbb{F}-$stopping times, where $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$. If there are ties in these stopping times, we always consider the order $\rho(y) < \tau < C$. We refer to Andersen et al. (1992) for the general technical details and also for the theory behind the estimation of a cumulative hazard in the presence of right censoring.

The Nelson-Aalen type estimator of the conditional hazard of τ given $\Psi(t) \leq y$ using the data $(T_i, \delta_i, \Psi_i(t), t \leq T_i)$, $i = 1, \ldots, n$ is given as follows.

$$\widehat{\Lambda}(t \mid \Psi(t) \leq y) = \sum_{T_i \leq t} \frac{\mathbf{1}(T_i \leq \rho_i(y))\delta_i}{\mathbb{R}_1(T_i)}, \tag{9}$$

where the risk set $\mathbb{R}_1(s) = \sum_j \mathbf{1}(T_j \geq s)\mathbf{1}(s \leq \rho_j(y))$. The heuristic justification behind the estimator (9) is the following. Given that $\rho_i(y) \geq u$, we are interested in the probability of observing an event $\tau_i \in [u, u + du)$. This conditional event process jumps only at τ_i satisfying the condition $\mathbf{1}(\Psi_i(\tau_i) \leq y)$ and the event indicator at time T_i is defined as $\mathbf{1}(T_i \leq \rho_i(y))\mathbf{1}(\delta_i = 1)$ and the corresponding risk set includes all those patients who have not observed any of the stopping times before the event time. From Fig. 1, we can see the contribution of each observation clearly. Events in the numerator of estimator (9) are events in quadrant (3) of Fig. 1 and the risk set at each event time in the denominator has to be visualised by shifting the vertical line to that event time, as shown by a dashed line at $t = 2$ in the figure. Then from Fig. 1 where the four quadrants are defined with respect to the shifted vertical line, patients whose response trajectory traverses from quadrant (3) to quadrant (4) and may further traverse to quadrant (1) constitute the risk set in the denominator for that event time. Any patient whose trajectory traverses from quadrant (3) to quadrant (2) will not contribute to the risk set for that event time.

Analogously, the Nelson-Aalen type estimator of the cumulative hazard of τ given that $\Psi(t) > y$, that is the stopping time $\rho(y) < t$ is

$$\widehat{\Lambda}(t \mid \Psi(t) > y) = \sum_{T_i \leq t} \frac{\mathbf{1}(T_i > \rho_i(y))\delta_i}{\mathbb{R}_2(T_i)}, \tag{10}$$

where the risk set $\mathbb{R}_2(s) = \sum_j \mathbf{1}(T_j \geq s)\mathbf{1}(s > \rho_j(y))$. Events in the numerator of estimator (10) are events in quadrant (2) of Fig. 1 and the risk set at each event time in the denominator has to be visualised by shifting the vertical line to that event time, as shown by the dotted line at $t = 3$ in the figure. Then from Fig. 1 where the four quadrants are defined with respect to the shifted vertical line, patients whose response trajectory traverses from quadrant (3) to quadrant (2) and further traverse to quadrant (1) constitute the risk set in the denominator for that event time. Any patient whose trajectory traverses from quadrant (3) to quadrant (4) will not contribute to the risk set for that event time. The corresponding conditional survival functions can be estimated by taking exponential of minus of the estimator of the conditional cumulative hazard function. For example,

$\widehat{S}(t \mid \Psi(t) \leq y) = \exp(-\widehat{\Lambda}(t \mid \Psi(t) \leq y))$. The proofs of the consistency and asymptotic normality for these estimators follow the proof of Nelson-Aalen estimator (Andersen et al. 1992) and are not included in this paper. We compute the variances of these estimators using the jackknife technique.

For a fixed (t, y), comparison of $\widehat{\Lambda}(t \mid \Psi(t) > y)$ and $\widehat{\Lambda}(t \mid \Psi(t) \leq y)$ helps in deciding which of these conditions is more favorable for discontinuation (or continuation) of the treatment. In our AMD example, higher value of VA is considered as a good treatment response. Hence, we expect that $\widehat{\Lambda}(t \mid \Psi(t) \leq y) \geq \widehat{\Lambda}(t \mid \Psi(t) > y)$ and the decision may rely on this inequality.

6 Applications

We illustrate the proposed estimation methods using data on age-related macular degeneration (AMD) of eye in this section. AMD especially its variant wet-AMD, is a chronic disease, which is one of the leading causes of visual handicap in Finland (Tuuminen et al. 2017). A patient with AMD starts losing eyesight from the central region of the retina. The most severe and long-standing handicap in AMD is mediated by the secondary growth of new vessels under the retina (wet-AMD). The anti-vascular endothelial growth factor (anti-VEGF) therapy, in which anti-VEGF agents are injected intra-ocularly and repeatedly for years, inhibits new vessel growth under the retina (Chakravarthy et al. 2010). Visual acuity (VA), generally measured as the number of letters or symbols one can read on eye chart, is the main outcome of interest. VA above 70 is considered as good eyesight while VA below 35 is the worst condition of the disease. There are other markers like the central foveal thickness and other lesions such as intra-retinal fluid (IRF), sub-retinal fluids (SRF). Fluctuations (appear/disappear, increase/decrease) in these other markers are suggestive of higher disease activity. Real-world AMD patient data consist of activity markers, treatment intervals and medications, all varying. The response to the treatment varies between patients and over time within a patient, and due to such heterogeneity it is difficult to decide about discontinuation of the treatment.

Visual acuity gives a direct measure of the vision in the measured eye and is easy to measure. Hence, we use VA, as the treatment response, measured repeatedly over the treatment period. Until now, most reports in this field have been based on mean visual acuity values at fixed, often annual, time points and mean number of injections. Impact of censoring, for example due to death, has commonly not been dealt with in many reported studies. For demonstrating the application of our proposed methods for wet-AMD, we consider the discontinuation of treatment by clinician's decision due to insufficient treatment response as the event of interest. Visual acuity measurements as longitudinal covariate process, and death or administrative end of the study or dropout for other unrelated reasons as the reasons for censoring. The censoring mechanism is assumed to be independent of the event as well as the covariate process. The reason behind assuming independent censoring due to death is that AMD is never likely to be the primary cause of death.

We apply our methods to different treatment protocols and subgroups of patients and exhibit thier use in medical decision making. We analyse two types of data from wet type of AMD (wet-AMD) patients; one from a clinical trial where generally the treatment protocol is either *intent-to-treat* (injection at predecided times) or *PRN* (injection when disease activity is seen), and the other one from a real-world clinical setting where the protocol is either *PRN* or *treat-and-extend* (a treatment interval that minimises the disease activity is determined). Table 3 shows the analyses specifications for these two studies according to the specifications for survival analysis suggested in Bull and Spiegelhalter (1997).

6.1 Computation of the integrated covariate process

The integrated covariate process is the area under the covariate trajectory, a summary measure frequently used by medical practitioners to assess disease progression or treatment effect. To make use of the integrated covariate process, it is necessary that all patients have measurements with the

Table 3 Estimation of survival probabilities from diagnosis (Observational real-world study) and from randomisation (CATT clinical trial)

Analysis specification	Observational real-world study	CATT clinical trial
Inclusion criteria	all patients	patients with the same protocol throughout the trial
Outcome	discontinuation of the treatment	discontinuation of the treatment
Time origin	day of diagnosis of the disease	randomisation
Entry time	0 (week)	0 (week)
Censoring rule	death, dropout or end of study (2-years)	death, dropout or end of study (2-years)
Survival time	time from 0 until discontinuation of treatment or censoring	time from 0 until discontinuation of treatment or censoring

same frequency and that there be no sudden changes in the covariate value between two consecutive measurements (Bonate 1998) otherwise frequent measurements would be required. Area under the curve methods become biased if there are missing data (Wilding et al. 2012; Bell et al. 2014). The clinical trial data in our illustration has measurements at every four weeks and the real-world observational study data has measurements approximately every month, not exactly on the same day of each month but a day near enough. The visual acuity of eye is a slowly changing response and it is not expected to change suddenly within a few days, hence the observations made approximately every month are acceptable. We used the actual day of the measurements and converted all time measurements into weeks from the day of entry of the person into the study. Since both our datasets had almost regular measurements, we connected consecutive measurements by lines and then computed the area under such piecewise linear curve using the trapezoid rule of integration. More elaborate smoothing techniques, if required, can be used. If such regular measurements are not available, some imputation techniques for missing data need to be employed.

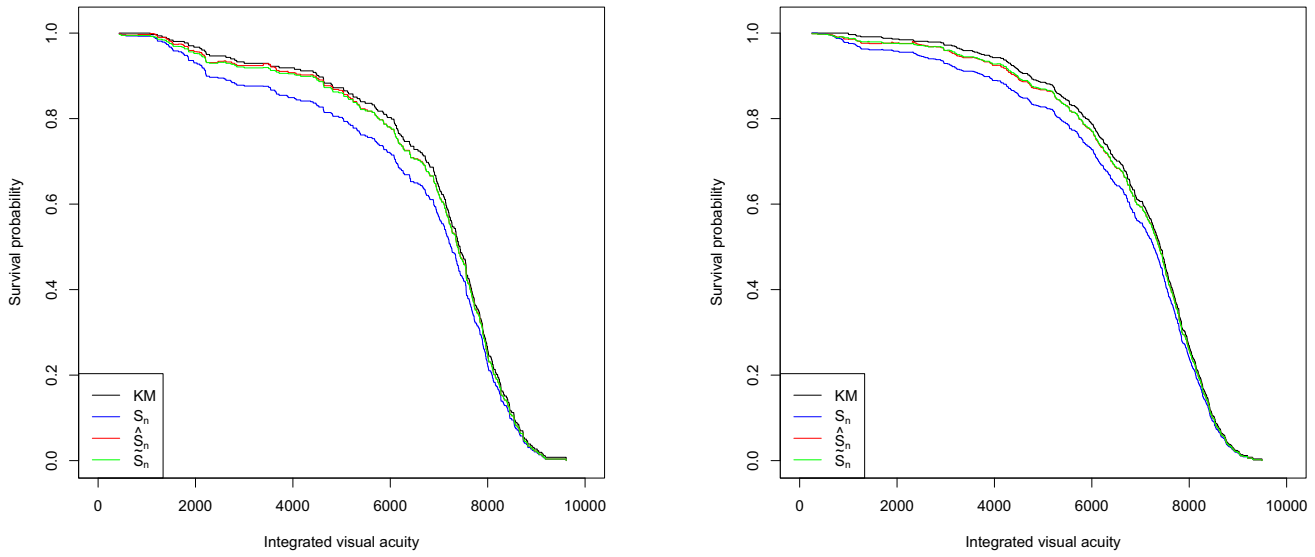
6.2 Clinical trial data for wet-AMD

We analysed data from the CATT (Comparison of AMD Treatment Trials) clinical trial study (https://www.med.upenn.edu/cpob/catt.html). CATT data consist of a 2-year

Table 4 Summary of the number of patients and the number of injections received over the two year period are shown in the table.

		CATT study		Real-world study
		Regular[1]	Variable[2]	
Patients	Total	313	595	110
	Dead	18	39	28
	Discontinued	8	4	21
Injections	Range	(2, 25)	(0, 25)	(0, 22)
	Median (Q1, Q3)	22 (19, 24)	11 (6, 17)	11 (6, 15)
	Mean	19.6	11	11.06

Patients were followed-up for two years after randomization.
[1]Regular: patients in this group received the treatment every four weeks throughout the follow-up.
[2]Variable: patients in this group received treatment as and when needed based on the disease activity.



(a) Regular group

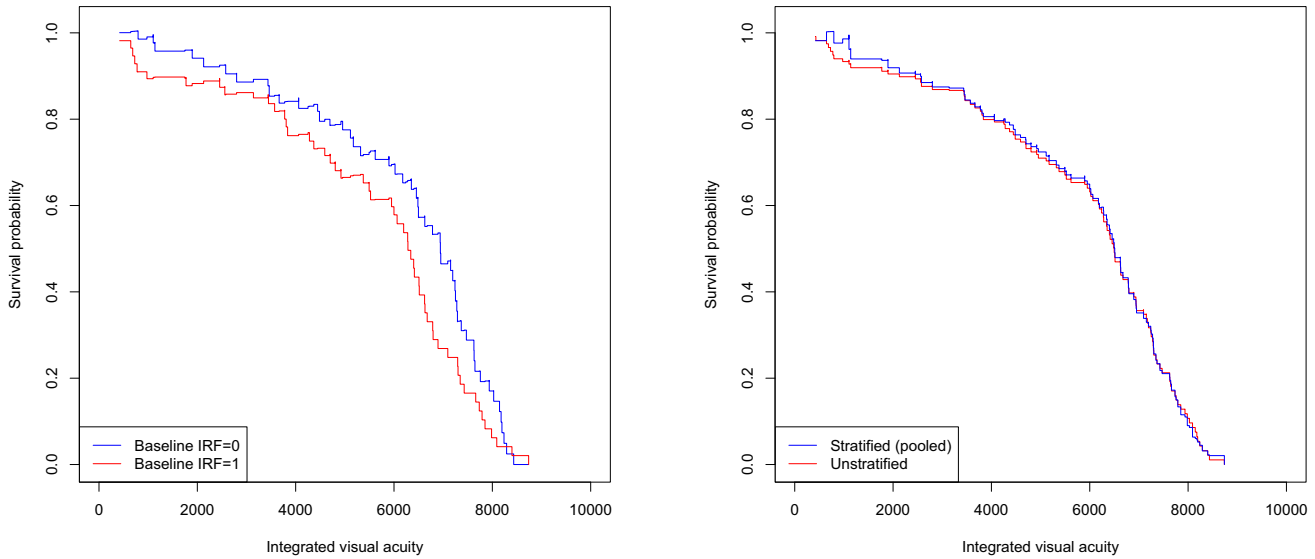
(b) Variable group

Fig. 2 Estimates of the survival function $P(\Psi(\tau) > y)$ of the integrated visual acuity for regular and variable groups of patients using two year follow-up data from the CATT study. The plot shows three estimates S_n, $\widehat{S}_n$ and $\widetilde{S}_n$ (obtained by setting $t = 0$ in (2), (3) and (5), respectively. KM corresponds to the naive Kaplan-Meier estimator of the survival function of $\Psi(\tau)$ when $\Psi(\tau)$ is censored whenever τ (time to discontinuation of the treatment) is censored by a censoring variable C. Regular group consists of patients who received the treatment every four weeks throughout the follow-up, while variable group consists of patients who received treatment as per *PRN* protocol

Table 5 Estimates of the survival probabilities (upto 2 decimal places) and estimates of their variances for the regular group (CATT study data)

(t, y)	$S_n(t, y)$ Survival Prob.	Variance Exact	JK	$\widehat{S}_n(t, y)$ Survival Prob.	Variance Exact	JK	$\widetilde{S}_n(t, y)$ Survival Prob.	Variance Exact	JK
(26, 1000)	0.98	0.0001	0.0001	0.99*	0.0000	0.0000	0.99	0.0000	0.0000
(26, 3000)	0.87	0.0004	0.0004	0.92	0.0002	0.0003	0.92	0.0003	0.0003
(26, 6000)	0.72	0.0007	0.0007	0.78	0.0006	0.0006	0.78	0.0006	0.0006
(52, 1000)	0.91	0.0003	0.0004	0.96	0.0001	0.0001	0.96	0.0001	0.0001
(52, 3000)	0.86	0.0004	0.0004	0.92	0.0002	0.0003	0.91	0.0003	0.0003
(52, 6000)	0.72	0.0007	0.0007	0.78	0.0006	0.0006	0.78	0.0006	0.0006
(78, 1000)	0.90	0.0004	0.0005	0.97	0.0001	0.0001	0.97	0.0001	0.0001
(78, 3000)	0.85	0.0005	0.0005	0.92	0.0003	0.0003	0.92	0.0003	0.0003
(78, 6000)	0.72	0.0007	0.0007	0.78	0.0006	0.0006	0.78	0.0006	0.0006

[1]Three estimators $S_n(t, y)$, $\widehat{S}_n(t, y)$ and $\widetilde{S}_n(t, y)$ are given in (2), (3) and (5), respectively.
[2]The corresponding variance estimates using the analytical formula (Exact) and jackknife method (JK) are shown under the columns Exact and JK, rounded to the 4^{th} decimal place.
[3]Regular group refers to patients of the CATT study who received the treatment every four weeks throughout the follow-up.
[1]$*$ indicates that the third digit is 5



(a) IRF subgroups

(b) Pooled estimate

Fig. 3 Estimates of $P(\Psi(\tau) > y)$ with stratification on the baseline IRF status in the real-world study. The plots are obtained by setting $t = 0$ in (3). Figure (a) shows marginal survival estimates for integrated VA for two subgroups of patients with and without baseline IRF. Figure (b) shows marginal survival curves for integrated VA from the full data (unstratified) and pooled estimate from two strata defined based on presence or absence of baseline IRF

follow-up of 1185 patients suffering from wet-AMD. Only for illustration purposes, we considered two subgroups of the data based on the treatment protocol: regular or variable injection intervals throughout the study period.

At the start of the trial, the first injection was administered, to all patients, after which, the respective protocol was applied.

- Regular group: Patients from this group received either drug Lucentis or Avastin with *intent-to-treat* protocol for two years. In this protocol, injections of the assigned drug were administered every four weeks independent of the disease activity and VA was also measured.
- Variable group: Patients from this group received either drug Lucentis or Avastin with *PRN* protocol for two years. In this protocol, an evaluation of lesion activities

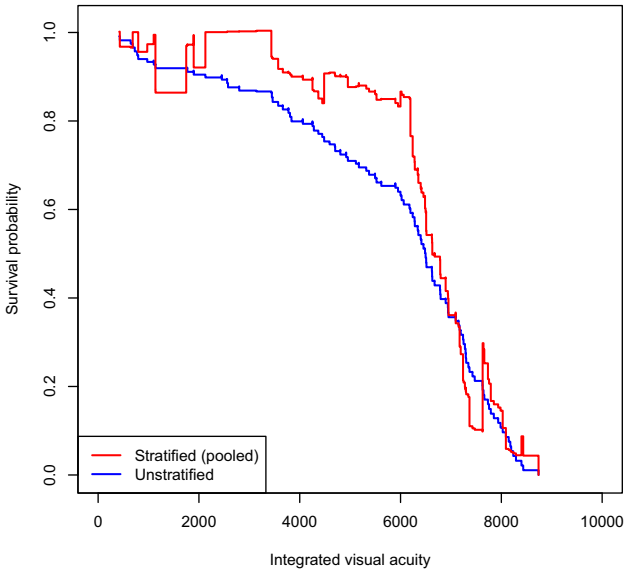


Fig. 4 Estimates of the survival function $P(\Psi(\tau) > y)$ of the integrated visual acuity with stratification on the baseline VA in the real-world study. The plots are obtained by setting $t = 0$ in (3). The marginal survival estimates for the full data and weighted average of stratum specific estimates from four strata defined based on baseline visual acuity are displayed.

after every four weeks was performed and the decision to administer an injection was taken based on lesion activity at that time. VA was measured for all patients roughly around every four weeks.

Since VA is of interest when assessing the response to AMD treatment, time-dependent VA measurements define the covariate process in our illustration. The inclusion criteria of this clinical trial required the best corrected E-ETDRS visual acuity score to be more than 23 letters (20/320 or better) and less than 82 letters (20/25 or worse). However, during the follow-up period of two years, a VA score was observed between 0 to 97. Each group had a small number of dropouts due to death or withdrawal (discontinuation of treatment). The discontinuation of treatment was due to an insufficient treatment response. So, we considered it as the event of interest. The patients suffering from AMD are generally elderly. Thereby, a common cause of censoring is death.

We were interested in an integrated covariate process until 2 years from the start of the trial. Those who had the complete 2 years follow-up were considered as complete cases. Hence, a complete follow-up until 2 years or a discontinuation of the treatment were considered as events of interest. The main time scale was number of weeks since the randomization.

The time to this event was τ and the integrated VA at that time was $\Psi(\tau)$. Further, τ was subjected to independent censoring by death (C). More details of the data are available in Table 4.

We applied the estimators of $S(t, y)$ described in section 3 to the two groups of patients separately. The censoring distribution was obtained in three different ways:

(i) Estimation using (2) S_n: When the censoring distribution is known. Here we used all 1185 patients data as the bigger population and estimated the censoring survival with death in this bigger group during the 2 years of follow-up. One can use the population registry data for deaths, if it is available.
(ii) Estimation using (3) $\widehat{S}_n$: When the censoring distribution is estimated using Kaplan-Meier estimator with censoring as an event of interest from the observed data. Here the data under analysis were used to get the censoring distribution.
(iii) Estimation using (5) $\widetilde{S}_n$: When the censoring distribution is estimated from the reconstructed data for each pair of time and integrated VA value, (t, y), as described in (4).

Marginal survival curves for the integrated VA, $\Psi(\tau)$, (with $t = 0$ in the bivariate survival function) for the two groups are shown in Fig. 2. In this figure, the survival curves of $\widehat{S}_n$ and $\widetilde{S}_n$ are close to each other suggesting that the reconstruction of the data is not advantageous. Hence, the use of $\widehat{S}_n$ may be sufficient. The KM curve is also close to these two curves but slightly upward. The S_n curve with known censoring distribution is lower than all other curves indicating that there are differences in the censoring distributions used for estimation of survival probabilities.

From Fig. 2 (a) and (b), it can be seen that the survival curves of $\Psi(\tau)$ in the regular and variable groups do not differ much, even though the number of injections is much less in the variable group as compared to the one in the regular group (Table 4). This shows that due to the *intent-to-treat* protocol used for the regular group, more injections than needed might have been administered.

For this large data, the computation times for these estimation methods were of great concern. The method involving the reconstruction of the data took about six times longer than other methods.

For each fixed pair (t, y), the variance was computed using the formula obtained in Appendix C, Appendix D and also with the jackknife method. These two variance estimates are close to each other as seen from the Table 5 and Figures 8 and 9 in supplementary figures Table 5 shows that estimators $\widehat{S}_n$ and $\widetilde{S}_n$ have similar amount of variance which is lesser than variance of the estimator S_n.

6.3 Real-world data for wet-AMD

Clinical trials are designed to study the treatment effects typically over a short period of follow-up, which is one of the

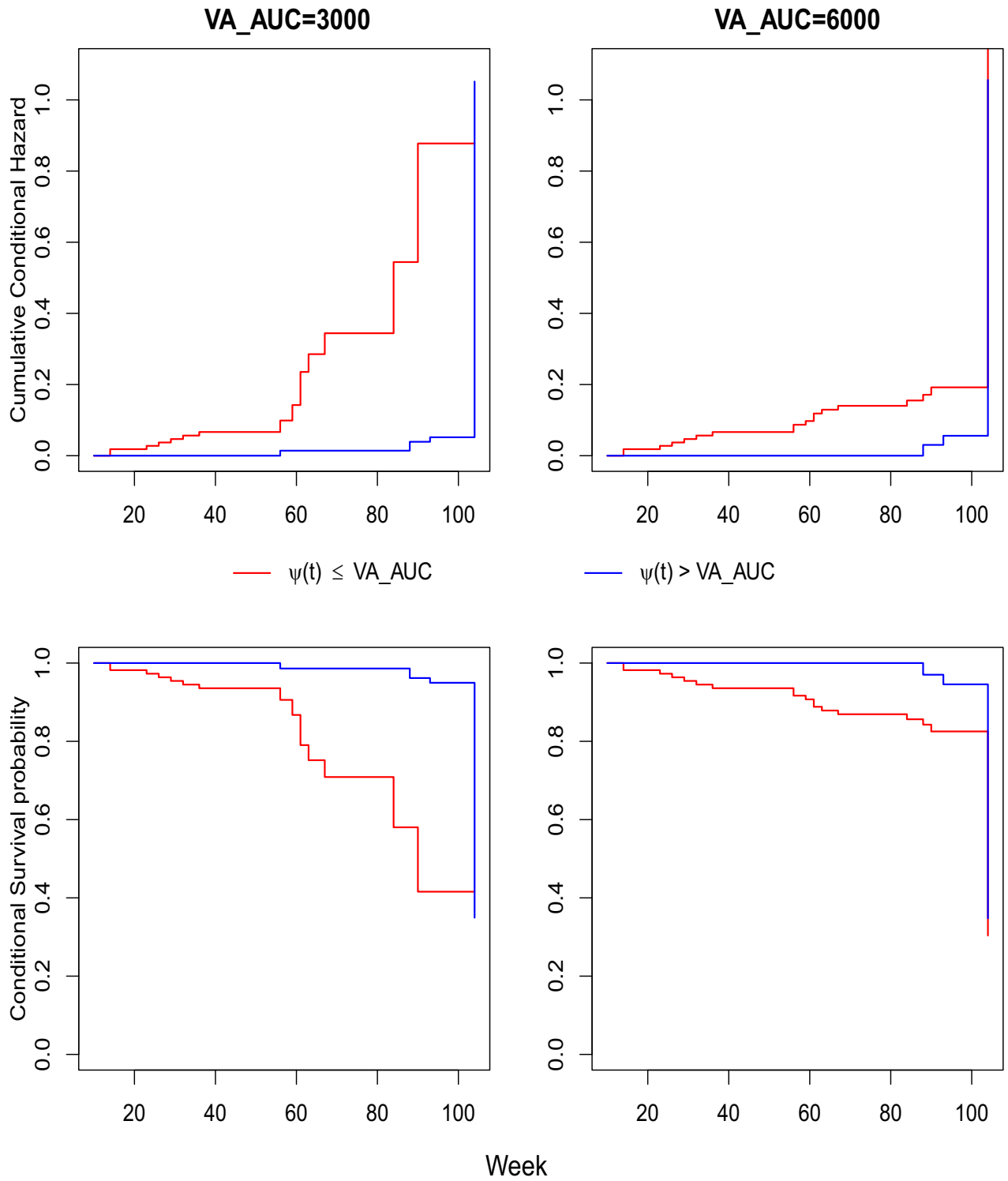


Fig. 5 In the top row, two curves of cumulative hazard of treatment discontinuation conditioned on comparison of cumulative VA till time t in weeks and specified fixed integrated VA value are displayed for VA_AUC= 3000 and 6000. The corresponding conditional survival function curves are displayed in the second row

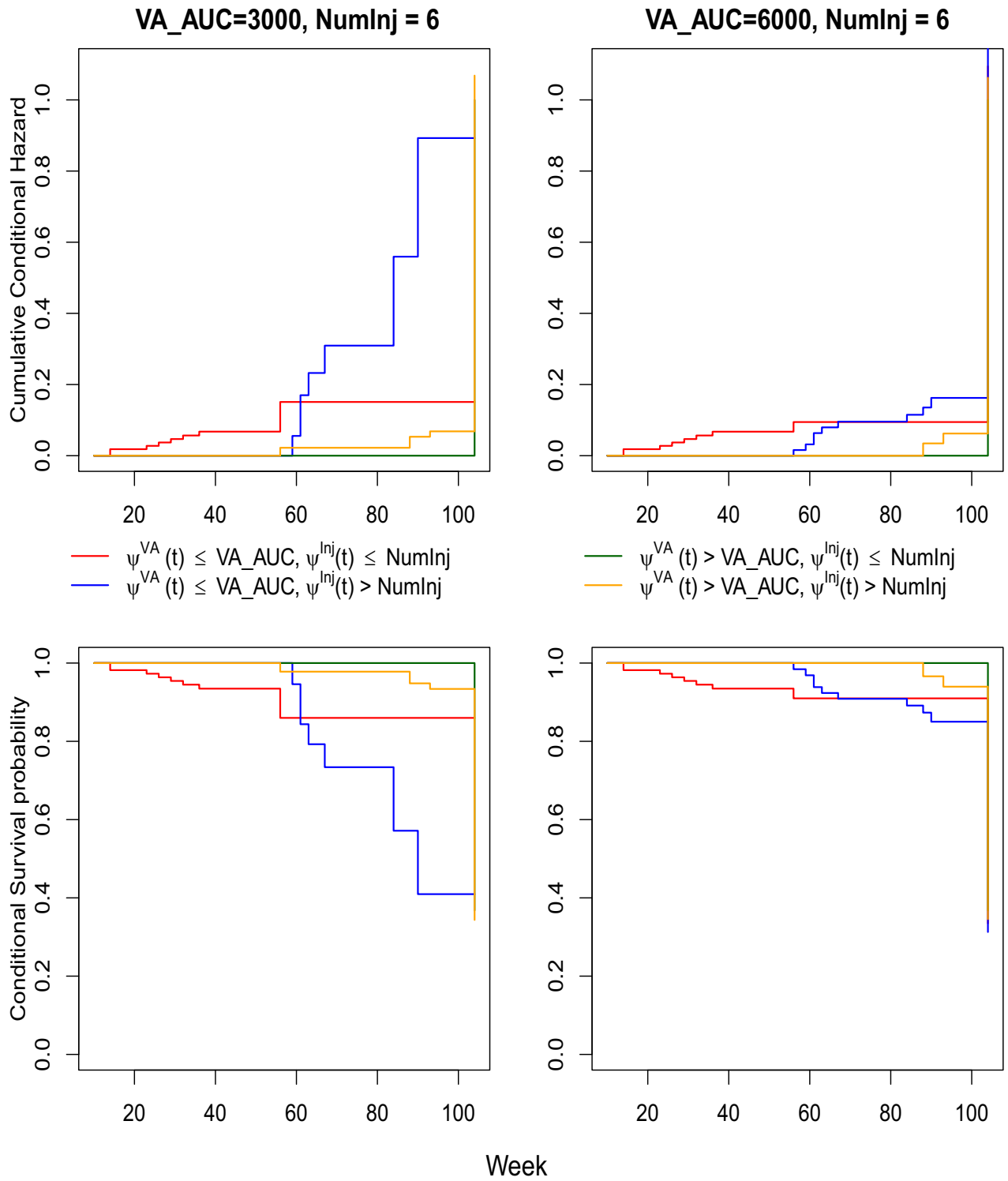


Fig. 6 In the top row, cumulative conditional hazard of treatment discontinuation is plotted for four conditions. Four combinations of condition on benefit (integrated VA) and cost (total number of injections) are plotted for 2 values of integrated VA and number of injections = 6. The corresponding conditional survival function curves are displayed in the second row

reasons for fewer treatment discontinuations as compared to the real-world data. Also the number of injections can be much more in the former than required as shown in Section 6.2. To understand the use of our proposed methods for making the decision of continuation (or discontinuation) of treatment, we considered real-world wet-AMD data with either *PRN* or *treat-and-extend* protocol which allows medical practitioners to determine the longest injection interval that keeps activity markers absent and to inject with long interval even when there is no sign of disease activity. We considered two-year follow-up data of 110 wet-AMD patients. The measurements included the VA, the presence of abnormal fluid within and around the retina and the injection timings. All patients underwent the same treatment with a slight variation in the duration between injections. A general description of this cohort and data are available in Ollila et al. (2022). All graphs of point-wise survival probabilities and their variance estimates are presented in the section F.2 of supplementary figures.

In the following two sections, we analyse subsets of these data and apply the methods described in sections 4 and 5.

6.3.1 Stratification on baseline values

Baseline IRF

The intra-retinal fluid is one of the main lesions indicating disease activity and VA is affected by the presence or absence of the intra-retinal fluid. For this real-world data, we studied two subsets of patients separately; one set of patients with intra-retinal fluid at the start of the study (Baseline IRF = 1) and the other set without intra-retinal fluid at the start of the study (Baseline IRF = 0). We estimated the bivariate survival function of $(\tau, \Psi(\tau))$ separately for these two groups, and then computed the pooled estimate as given in (8).

In Fig. 3, we have shown marginal survival curves for $\rho(y)$ for $t = 0$. The group with baseline IRF had lower survival probabilities as compared to the group without baseline IRF as shown in Fig. 3(a). It can be seen from Fig. 3(b) that the stratified estimate and the unstratified full data estimate are not far apart in the latter part of the time axis. However, in the initial part, the unstratified estimate is closer to the (Baseline IRF = 1) curve and the stratified estimate is closer to the (Baseline IRF = 0) curve. One possible explanation for this could be the lower variance of the survival probability for the group (Baseline IRF = 0), resulting in the higher weight in the pooled estimate.

Baseline VA

Since the VA is bounded above, the patients starting with a very high VA value may not experience much increase in VA, as opposed to the patients starting at lower VA values, who have more potential to experience a growth in their VA value. However, the integrated response tends to be higher among those who start out at a higher VA value compared to those with a lower VA value. For this reason, we also performed a stratified analysis in which the strata are defined based on the baseline VA values. In this real data, we observed VA values in the range 0 and 85. These subgroups, decided in consultation with clinicians, indicate specific stages of wet-AMD.

- VA ≥ 70: required for a driving license,
- $55 \leq$ VA < 70: moderate visual handicap level,
- $35 \leq$ VA < 55: above moderate visual handicap level,
- VA < 35: severe visual handicap level.

The pooled estimate of survival probabilities is clearly different than the unstratified estimate, which is indicative of the differences in the subgroups, as shown in Fig. 4.

The survival curves for $\Psi(\tau)$ and the corresponding variance graphs for each stratum are available in section F.2 in supplementary figures. As expected, the survival curve of the integrated VA for the lowest VA group (VA < 35) is much lower as compared to the one for the other three groups. The top two groups do not show much difference among themselves. From section F.2 in supplementary figures, it can be seen that the stratum-specific survival probabilities show large differences at integrated VA near 5500. The survival curve of stratum VA< 35 has already reached zero around 5500, whereas the group $55 \leq$ VA < 70, shows a survival probability of around 80% and the group $35 \leq$ VA < 55 has a survival probability of around 50% at the integrated VA near 5500. The major contribution in the pooled estimate comes from the group VA ≥ 70, which has an almost horizontal survival curve above 80% probability with very low variance. Due to this low variance, group VA ≥ 70 gets the highest weight when the integrated VA is near 5500. This explains the difference in the stratified and the unstratified estimates. After the integrated VA 2000, a small jump can be seen in the pooled estimate. This jump can be explained as being due to the increased contribution from the groups VA ≥ 70 and $55 \leq$ VA < 70, with a high survival probability and a low variance, as against the minimal contribution from the lowest group VA < 35 with a low survival and a high variance.

6.3.2 Decision using conditional survival probabilities

We analyse the real-world data to illustrate the use of conditional cumulative hazard for the decision making of continuation (or discontinuation) of treatment as described in Section 5. The conditional cumulative hazard is computed for some fixed values of integrated VA (VA_AUC). At any time t, if we are interested in the conditional probability of treatment discontinuation based on whether VA_AUC at time t is above or below 3000, we use $y = 3000$ in estimators (9) and (10).

Fig. 5 shows the cumulative hazard plots and the corresponding survival function conditioned on the VA_AUC value. In both the graphs in the top row, the cumulative hazard of the discontinuation of treatment is higher if the cumulative VA has not reached the specified VA_AUC value. The cumulative hazard of treatment discontinuation, when the integrated VA has not reached the specified VA_AUC, increases much faster for VA_AUC = 3000, as compared to VA_AUC = 6000. The survival curves corresponding to these cumulative hazard functions are shown in the second row of Fig. 5. The estimate of the conditional survival probability and the 95% confidence interval obtained using the corresponding jackknife variance estimate at 52 weeks for VA_AUC $\leq$ 3000 is 0.93 (0.89, 0.98), as against the one for VA_AUC $>$ 3000, which is near 1 (1, 1). The conditional survival probabilities at 80 weeks for these conditions are 0.7 (0.54, 0.87) and 0.98 (0.95, 1), respectively. The graph of the conditional survival probabilities with their confidence intervals is available in Fig. 16 in supplementary figures.

We can further put a condition on the number of injections. In this real-world study, one measurement visit was planned every month. Based on the patient's eye measurements, the decision to administer an injection was taken. In order to use the observed data in decision making, one can define another integrated process as the number of injections until time t weeks, and consider the injection process as covariate process. As discussed in section 5, the integrated VA can be viewed as the benefit of the treatment and the number of injections as the cost of the treatment. A more elaborate computation of the benefit and the cost can be obtained if more details are available from a study. However for illustration purposes, here we have used the integrated VA and the number of injections as the benefit and the cost respectively.

Fig. 6 shows the cumulative conditional hazard plots and the corresponding survival curves for the fixed integrated VA and the fixed number of injections. Four combinations are possible based on the values of these two integrated processes at time t. The columns represent the fixed combination of the specified VA_AUC and the specified Num_Inj. The top row shows the cumulative hazard conditioned on VA_AUC and Num_Inj and the second row shows the corresponding conditional survival. The top left graph represents the cumulative conditional hazard with the condition on two comparisons, (I) the comparison between the integrated VA at time t and VA_AUC=3000, and (II) the comparison between the number of injections at time t and Num_Inj=6. In this figure, for the first 52 weeks (1 year), the cumulative hazard of treatment discontinuation is the highest for those who received less than 6 injections till that time and did not reach the integrated VA of 3000. However, after 52 weeks, that is in the second year of follow-up, the treatment discontinuation hazard is higher for the patients receiving more than 6 injections but not achieving the integrated VA of 3000.

In both graphs in the top row, the conditional cumulative hazard of treatment discontinuation is lower when the observed integrated VA is more than the specified fixed VA_AUC, and the conditional cumulative hazard of treatment discontinuation is higher when the observed integrated VA is less than the specified fixed VA_AUC. It can be seen that those who receive less number of injections compared to the fixed Num_Inj are more likely to discontinue the treatment. The cumulative hazard for those who receive more injections than fixed Num_Inj but do not reach the specified VA_AUC is low initially, but in the later part of the follow-up if more number of injections do not help in improving the integrated VA then the hazard increases for the patients with $\psi^{VA}(t) \leq$ VA_AUC, $\psi^{Inj}(t) >$ Num_Inj.

This kind of observations can be useful in medical decision making. For example, a clinician may increase the number of injections for a patient who has not reached certain response for VA_AUC with smaller number of injections at the end of the first year while reducing the number of injections for those who have not achieved the response in spite of larger number of injections.

This methodology of comparing the estimates of the cumulative hazard conditional on the benefit and the cost allows one to take a decision of discontinuation (or continuation) of the treatment at any time during the follow-up. The main advantage is that we do not need the final observation times for making this decision.

7 Discussion

In this article, we have proposed estimator of the joint survival function of the time to event of interest and the integrated covariate process by using all available information on covariate process over the observation period. Our proposed estimators for fixed (t, y) have been developed for the integrated covariate process, however they can also be used for the estimation of the survival function of, for example, the quality-adjusted survival time or the total medical cost.

In our estimation methodology, we assumed that the non-zero contributions come from the observations for which the time when the integrated covariate process crosses the fixed point y is observed. This assumption is justifiable because we observe the covariate process over time. We note that the estimators (3) and (5) are IPCW estimators and the weights are the Kaplan-Meier estimates of the censoring survival functions to compensate for those who get censored. Higher weights are given to the events occurring late during the follow-up especially when most of the censoring happens towards the end. Moreover, the Kaplan-Meier estimate remains constant after the last observed censoring time and the events later than that get the same weights. The two estimates are very close as seen (Fig. 2) in most cases except

in the beginning of the study for some small sample subsets in Figures 11 and 13 in supplementary figures. We observe that for smaller values of y, the weights in (3) can be higher compared to the weights in (5) because weights in (3) get a contribution from all the data whereas the weights in (5) get a contribution from the reconstructed data for fixed (t, y). We recommend the use of the estimator with known censoring distribution whenever possible. If the censoring distribution is not known, then we recommend to use both (3 and 5) because their performances would depend on the positioning of the censoring times.

Our experience, based on the empirical analyses presented in Section 6, suggest that the computation time required for the estimation of the survival function using (5) is 5 to 6 times more, as compared to (3), even for the small sample size of 25 and the variance estimation takes even longer. In our case, the censoring mechanism is independent of τ and the covariate process, and so, the data reconstruction method has a little advantage over other methods. When the censoring mechanism depends on the covariate process, we think that the data reconstruction method will bring in the visible effect of y through the estimate of the censoring distribution, making it preferable over other methods.

If the censoring distribution is known from the underlying bigger population and if the assumption of independent censoring is valid, the proposed estimator with known censoring distribution gives an unbiased estimation of the bivariate survival function $S(t, y)$. In settings where the censoring is due to reasons unrelated to the event of interest and the covariate process, for example, death, in our dataset, we can use external data from the general population from which the patient population comes from, in order to obtain the censoring distribution. We have discussed adjustments to the proposed estimators when the censoring depends on the measured covariates. Such adjustments are expected to improve the performance of IPCW estimators. We have also presented alternative ways of computing the pooled estimate of the survival function. For a stratified analysis, the overall performance of the pooled estimator will depend on the stratum-specific estimator and its variance. These quantities depend on the stratum size as well as the proportion of censoring in each stratum. The thumb rule for these when computing Kaplan-Meier estimator apply here as well.

In all our datasets and the subgroup analysis, the variance estimates using an analytical formula and the jackknife method are very close to each other. We have chosen the jackknife method for the estimation of the variances throughout this paper. In principle, it is possible to give the exact expression for (D8) by using the martingale theory. However, we have chosen to provide the estimates of the asymptotic variances for our proposed estimators (D9) only for practical reasons. We recommend the use of (3) and the jackknife estimate of the variance in practice. Other resampling methods such as bootstrap can also be used. Choice of the resampling method depends on various aspects like the size of the data under analysis, computational power, availability of resources for parallel processing.

Our proposed estimators of the joint survival function and the conditional survival functions can easily be extended to multivariate settings. When two covariate processes are of interest, we define two integrated processes $\Psi^{(1)}(t)$ and $\Psi^{(2)}(t)$, and extend the estimation methodology. We have already demonstrated this for the cumulative conditional hazard estimation using the visual acuity and the number of injections in the AMD case study. The conditional survival probabilities are especially helpful in the decision making regarding the change in the treatment regime. For a new patient or for an ongoing study, the decision of continuation (or discontinuation) of the ongoing treatment regime at any time during the follow-up can be made using the existing joint or conditional distribution from the existing studies where the patient population is the same as the new study, and the covariate measurement settings are also similar in the two studies. The two data sets that we have used for illustration come from two different geographical regions and they represent two different patient populations. For this reason, the two sets of analyses were kept separate. If the interest is in the computation of the integrated response over a fixed interval, then one may employ the natural progression of the disease as described in Appendix E.

Although, the motivation of this research came from the medical field, the formulation and methods proposed can be applied to other fields such as engineering, social sciences.

Appendix A Equivalence of the proposed estimator (3) and Kaplan-Meier estimator for original time t

The equivalence between Kaplan-Meier and the marginal estimator

$$\widehat{S}_n(t, 0) = \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbf{1}(\tau_i \wedge C_i > t)}{\widehat{G}(t)}$$

is clear by using the following result proven in Satten et al. (2001),

$$\widehat{S}^{KM}(t) \times \widehat{G}^{KM}(t) = \sum_{i=1}^{n} \frac{I(\tau_i \wedge C_i > t)}{n}.$$

where $\widehat{S}^{KM}(t)$ denotes Kaplan-Meier survival estimate and $\widehat{G}^{KM}(t)$ denotes Kaplan-Meier censoring survival estimate.

Appendix B Equivalence of the proposed estimator (2) and Zhao-Tsiasis estimator

The Zhao-Tsiatis (ZT) estimator given in Zhao and Tsiatis (1997) is for the quality-adjusted survival time. For a fixed value of y, the observation times, and the corresponding event indicator is redefined as follows.

$$\begin{aligned}\widetilde{\tau}_i(y) &= \tau_i \wedge \rho_i(y)\\ \widetilde{T}_i(y) &= (\widetilde{\tau}_i(y) \wedge C_i)\\ \widetilde{\delta}_i(y) &= \begin{cases} 1 & \text{when } C_i \geq \widetilde{\tau}_i(y)\\ 0 & \text{otherwise.}\end{cases}\end{aligned} \tag{B1}$$

The ZT estimator of survival function of the quality-adjusted survival time, for the censoring survival function G, is given by

$$S^{ZT}(y) = \frac{1}{n}\sum_i \frac{\widetilde{\delta}_i(y)}{G(\widetilde{\tau}_i(y))}\mathbf{1}(\Psi_i(\tau_i) > y). \tag{B2}$$

Here we establish the equivalence of our marginal estimator $S_n(0, y)$, and the estimator $S^{ZT}(y)$ proposed in Zhao and Tsiatis (1997) for estimating the survival function of $\Psi(\tau)$, the quality-adjusted survival time or integrated covariate process, at y.
Consider the estimator (B2) for fixed y.

$$\begin{aligned}S^{ZT}(y) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbf{1}(\tau_i > \rho_i(y))\mathbf{1}(\tau_i \wedge \rho_i(y) \leq C_i)}{G(\tau_i \wedge \rho_i(y))}\right),\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbf{1}(\tau_i > \rho_i(y))\mathbf{1}(\rho_i(y) < C_i)}{G(\rho_i(y))}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbf{1}(\tau_i \wedge C_i > \rho_i(y))}{G(\rho_i(y))}\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbf{1}(T_i > 0 \vee \rho_i(y))}{G(0 \vee \rho_i(y))}\right) = S_n(0, y),\end{aligned}$$

the second equality follows because the i^{th} term is nonzero only when $\tau_i \wedge \rho_i(y) = \rho_i(y)$. It is to be noted that the reconstructed data in (4) for $t = 0$ are exactly the same as (B1). The ZT estimator when G is estimated using the Kaplan-Meier estimator using (B1) also coincides with our estimator (5) for $t = 0$.

Appendix C Proof of Theorem 1

Consider

$$\begin{aligned}&E\left[\frac{\mathbf{1}(T_i > t, \Psi(T_i) > y)}{G(t \vee \rho_i(y))}\right]\\ &= E\left[\frac{\mathbf{1}(\tau_i > t, \Psi_i(\tau_i) > y)\mathbf{1}(C_i > t, \Psi_i(C_i) > y)}{G(t \vee \rho_i(y))}\right]\\ &= E\left[\frac{\mathbf{1}(\tau_i > t \vee \rho_i(y))\mathbf{1}(C_i > t \vee \rho_i(y))}{G(t \vee \rho_i(y))}\right]\\ &= E\left(E\left[\frac{\mathbf{1}(\tau_i > t \vee \rho_i(y))\mathbf{1}(C_i > t \vee \rho_i(y))}{G(t \vee \rho_i(y))}\middle|\tau_i, \rho_i(y)\right]\right)\\ &= E\left[\frac{\mathbf{1}(\tau_i > t \vee \rho_i(y))G(t \vee \rho_i(y))}{G(t \vee \rho_i(y))}\right]\\ &\text{(because of independence of } C_i \text{ and } (\tau_i, Z_i(\cdot)))\\ &= E\left[\mathbf{1}(\tau_i > t \vee \rho_i(y))\right] = S(t, y).\end{aligned} \tag{C3}$$

Hence, the estimator (2) is an unbiased estimator of $S(t, y)$. For the second moment, we have

$$\begin{aligned}&E\left[\frac{\mathbf{1}(T_i > t, \Psi(T_i) > y)}{G(t \vee \rho_i(y))^2}\right]\\ &= E\left[\frac{\mathbf{1}(\tau_i > t \vee \rho_i(y))G(t \vee \rho_i(y))}{G(t \vee \rho_i(y))^2}\right]\\ &= E\left[\frac{\mathbf{1}(\tau_i > t \vee \rho_i(y))}{G(t \vee \rho_i(y))}\right].\end{aligned}$$

Hence, the result.

Appendix D Proof of Theorem 2

By adding and subtracting $S_n(t, y)$, we obtain

$$\begin{aligned}\sqrt{n}(\widehat{S}_n(t, y) - S(t, y)) &= \sqrt{n}(S_n(t, y) - S(t, y))\\ &- \sum_i \frac{\sqrt{n}}{n}\left[\frac{(\widehat{G}(t \vee \rho_i(y)) - G(t \vee \rho_i(y)))\mathbf{1}(T_i > t \vee \rho_i(y))}{\widehat{G}(t \vee \rho_i(y))G(t \vee \rho_i(y))}\right].\end{aligned} \tag{D4}$$

The first term $\sqrt{n}(S_n(t, y) - S(t, y))$ converges to $\mathbb{N}(0, \sigma_1^2)$ in distribution as shown in earlier section. Due to the weak convergence of Kaplan-Meier estimator $\sqrt{n}(\widehat{G}(.) - G(.))$, the second term converges asymptotically to zero mean Gaussian random variable for fixed (t, y). Hence, $\widehat{S}_n(t, y)$ is asymptotically unbiased estimator of $S(t, y)$ and $\sqrt{n}(\widehat{S}_n(t, y) - S(t, y)) \longrightarrow \mathbb{N}(0, \sigma_2^2)$.

Using the weak convergence of the Kaplan-Meier estimator $\sqrt{n}(\widehat{G}(.) - G(.))$, we can replace $\widehat{G}$ in the denominator (D4) by G, with the additional $o_p(1)$ error term

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_i\left[\frac{(\widehat{G}(t \vee \rho_i(y)) - G(t \vee \rho_i(y)))\mathbf{1}(T_i > t \vee \rho_i(y))}{\widehat{G}(t \vee \rho_i(y))G(t \vee \rho_i(y))}\right]\\ &- \frac{1}{\sqrt{n}}\sum_i\left[\frac{(\widehat{G}(t \vee \rho_i(y)) - G(t \vee \rho_i(y)))\mathbf{1}(T_i > t \vee \rho_i(y))}{G^2(t \vee \rho_i(y))}\right]\\ &\longrightarrow 0\end{aligned}$$

in probability as $n \to \infty$.

Also, note that for any u

$$\begin{aligned}\sqrt{n}(\widehat{G}(u) - G(u)) &= \sqrt{n}(e^{-\widehat{\Lambda}^C(u)} - e^{-\Lambda^C(u)}) + o_p(1),\\ &= -\sqrt{n}G(u)(\widehat{\Lambda}^C(u) - \Lambda^C(u)) + o_p(1),\\ &\quad \text{by delta method}\end{aligned}$$

where $\Lambda^C(u)$ is the cumulative hazard of the censoring variable C and $\widehat{\Lambda}^C(u)$ is its Nelson-Aalen estimator. We refer to Andersen et al. (1992) for technical details. Thus, using the above in (D4) we get

$$\begin{aligned}\sqrt{n}(\widehat{S}_n(t,y) - S(t,y)) &= \sqrt{n}(S_n(t,y) - S(t,y))\\ &+ \sum_i \frac{\sqrt{n}}{n}\left[\frac{(\widehat{\Lambda^C}(t \vee \rho_i(y)) - \Lambda^C(t \vee \rho_i(y)))\mathbf{1}(T_i > t \vee \rho_i(y))}{G(t \vee \rho_i(y))}\right]\\ &+ o_p(1). \qquad \text{(D5)}\end{aligned}$$

By martingale representation of the Nelson-Aalen estimator, we can show that the expression in (D5) equals

$$\begin{aligned}\sqrt{n}(\widehat{S}_n(t,y) - S(t,y)) &= \sqrt{n}(S_n(t,y) - S(t,y))\\ &+ \frac{1}{\sqrt{n}}\sum_i \frac{\mathbf{1}(T_i > t \vee \rho_i(y))}{G(t \vee \rho_i(y))}\int_0^{t \vee \rho_i(y)} \frac{d\widetilde{M}_n^C(u)}{r(u)} + o_p(1), \qquad \text{(D6)}\end{aligned}$$

where $M_i^C(t) = N_i^C(t) - \int_0^t R_i(u)d\Lambda^C(u)$ is the zero-mean martingale of the counting process $N_i^C(t) = \mathbf{1}(T_i \le t, \delta_i = 0)$ defined with respect to the appropriate filtration, and $R_i(t) = \mathbf{1}(T_i \ge t)$ is the at-risk process predictable with respect to the filtration, $\widetilde{M}_n^C(t) = n^{-1/2}\sum_i M_i^C(t)$ and $r(t) = E(R_i(t)) = P(T \ge t)$. The equation (D6) can be expressed as:

$$\begin{aligned}\sqrt{n}(\widehat{S}_n(t,y) - S(t,y)) &= \sqrt{n}(S_n(t,y) - S(t,y))\\ &+ \sum_i\left(\int_0^\infty w_n(u)dM_i^C(u)\right) + o_p(1), \qquad \text{(D7)}\end{aligned}$$

where

$$w_n(u) = \frac{1}{n}\sum_{j=1}^n \mathbf{1}(u \le t \vee \rho_j(y))\frac{\mathbf{1}(T_j > t \vee \rho_i(y))}{G(t \vee \rho_j(y))r(u)}.$$

The asymptotic variance of $\sqrt{n}(\widehat{S}_n(t,y) - S(t,y))$ can be given by

$$\sigma_2^2 = \text{var}\left(\frac{\mathbf{1}(T_i > t \vee \rho_i(y))}{G(t \vee \rho_i(y))} + \int_0^\infty w(u)dM_i^C(u)\right), \qquad \text{(D8)}$$

where

$$w(u) = \frac{1}{r(u)}E\left[\mathbf{1}(u \le t \vee \rho_j(y))\frac{\mathbf{1}(T_j > t \vee \rho_j(y))}{G(t \vee \rho_j(y))}\right],$$

and $w_n(u)$ converges to $w(u)$ in probability.

Estimation of σ_2^2. In practice, we need appropriate estimators for various terms in (D8). Estimator of the first term in (D8) is straightforward. For the second term, we use results from (Datta et al. (2010), Appendix A.2). Define

$$\widehat{M}_i^C(t) = N_i^C(t) - \int_0^t R_i(u)d\widehat{\Lambda}^C(u),$$

where $d\widehat{\Lambda}^C(t) = dN^C(t)/R(t)$ and $R(t) = \sum_i R_i(t)$. We can now obtain the following result in order to estimate the second term of (D8).

$$\begin{aligned}\int_0^\infty w(u)d\widehat{M}_i^C(u) &= w(T_i)(1-\delta_i) - \int_0^\infty \frac{w(u)R_i(u)}{R(u)}dN^C(u)\\ &= w(T_i)(1-\delta_i)\\ &\quad - \sum_{j=1}^n \int_0^\infty \frac{w(u)\mathbf{1}(T_i \ge u)}{R(u)}dN_j^C(u)\\ &= w(T_i)(1-\delta_i)\\ &\quad - \sum_{j=1}^n \frac{w(T_j)\mathbf{1}(T_i \ge T_j)(1-\delta_j)}{R(T_j)}\end{aligned}$$

and

$$\widehat{w}(u) = \frac{1}{R(u)}\sum_{j=1}^n \mathbf{1}(u \le t \vee \rho_j(y))\frac{\mathbf{1}(T_j > t \vee \rho_j(y))}{\widehat{G}(t \vee \rho_j(y))}.$$

The variance (D8) can be estimated as

$$\widehat{\sigma}_2^2 = \frac{1}{n-1}\sum_{i=1}^n\left(V_i - \bar{V}\right)^2, \qquad \text{(D9)}$$

where

$$\begin{aligned}V_i &= \frac{\mathbf{1}(T_i > t \vee \rho_i(y))}{\widehat{G}(t \vee \rho_i(y))} + \widehat{w}(T_i)(1-\delta_i)\\ &\quad - \sum_{j=1}^n \frac{\widehat{w}(T_j)\mathbf{1}(T_j \le T_i)(1-\delta_j)}{R(T_j)},\end{aligned}$$

and $\bar{V} = n^{-1}\sum_i V_i$

Appendix E Use natural progression for area under the trajectory

To compare two patients or two groups of patients with respect to the integrated covariate process, we need to estimate their integrated covariate processes over a fixed interval of time. Let C_A be the administrative end of follow-up time. To estimate the integrated covariate process or area under the curve over the interval $[0, C_A]$, the covariate process need to be observed until C_A. In the case of wet-AMD, if censoring C due to death is observed before C_A, VA becomes zero after the death of a patient, so the required area under the curve is to be computed until C. The VA trajectory is observed up to $\tau \wedge (C_A \wedge C)$. If the event τ of treatment discontinuation is observed during the follow-up $[0, C_A \wedge C]$, soon after the treatment has stopped, the natural progression of the disease can be assumed to obtain the area after the event time.

$$\int_0^{C_A \wedge C_i} Z_i(s)ds \simeq \Psi_i(\tau_i) + \int_{(\tau_i+\epsilon_i)}^{C_A \wedge C_i} g(t; a, b)dt \text{ for } \delta_i = 1,$$

where $g(t; a, b)$ is a carefully chosen known function, essentially capturing the natural progression of the disease with regard to $Z_i(.)$, for example one such function suggested by Shah and Del Priore (2007) Shah and Priore (2007); Elshout et al. (2017) based on the meta-analysis of many wet-AMD studies is

$$g(t; a, b) = Z_i(\tau_i) - \frac{t - \tau_i}{a(t - \tau_i) + b}, \text{ with } a = 0.0133,$$
$$b = 0.1448.$$

Appendix F Supplementary figures

F.1 Clinical trial data

Many existing methods use only the latest covariate value for the prediction of the future trajectory or the time to an event of interest. However not accounting for the full history may give misleading inference. To emphasize our point, in Fig. 7. we give an example from CATT study data. Four patients were selected by having the same VA value of 40 at 24 weeks. Their visual acuity is same at six months but their trajectories are different and hence their area under the trajectories.

Survival curve and corresponding variance curve for the regular group receiving injections every 4 weeks are given in Fig. 8

Survival curve and corresponding variance curve for the variable group receiving injections as and when needed are given in Fig. 9.

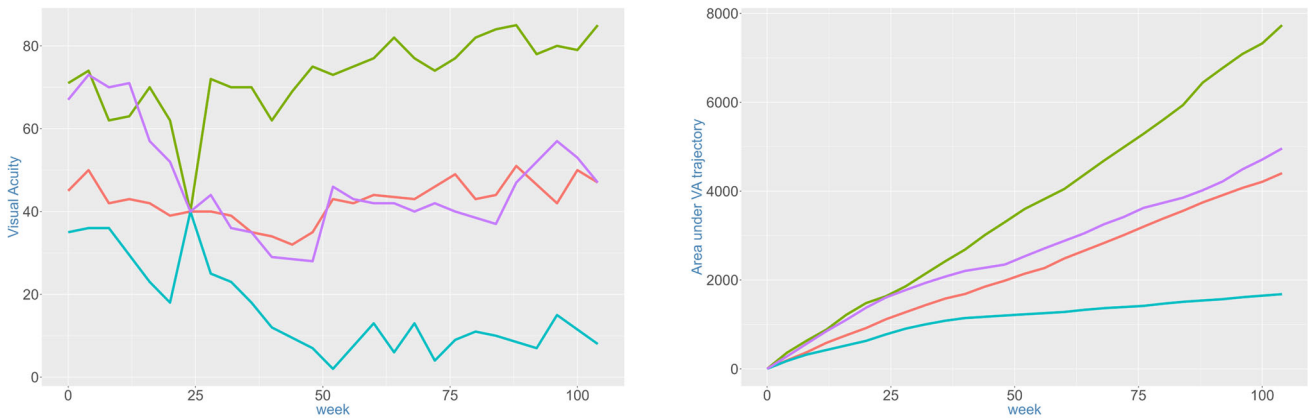
F.2 Real-world data

F.2.1 Full data

Survival curve and corresponding variance curve for the full data are given in Fig. 10.

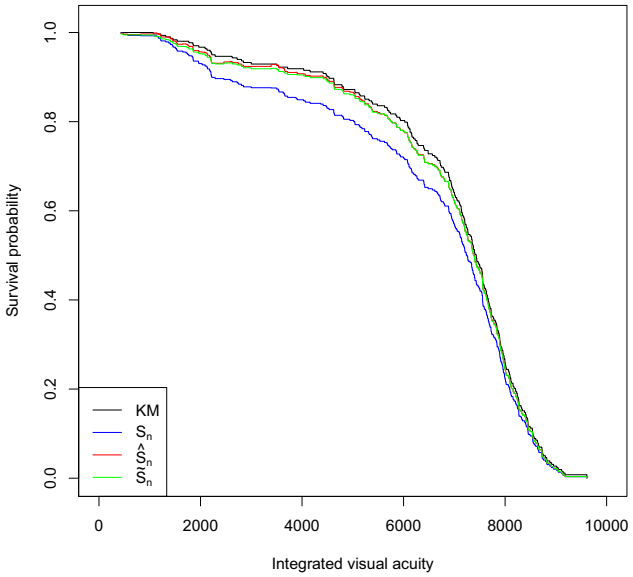
F.2.2 Strata based on baseline IRF

Full data from real-world study is divided into 2 subgroups based on the presence (number of patients = 55) or absence (number of patients = 51) of IRF at baseline. For four patients

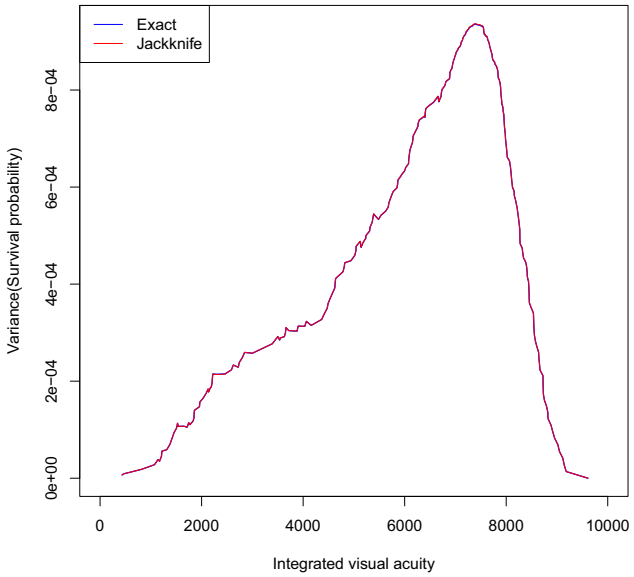


(a) VA trajectory (b) Area under the VA trajectory

Fig. 7 Observed visual acuity trajectories and the corresponding area under the curve. (a) Visual acuity trajectories for four patients. (b) The corresponding area under the visual acuity trajectories. All four patients had the same VA measurement VA(24) of 40 at 24 weeks but different area under the trajectory $\psi(24)$

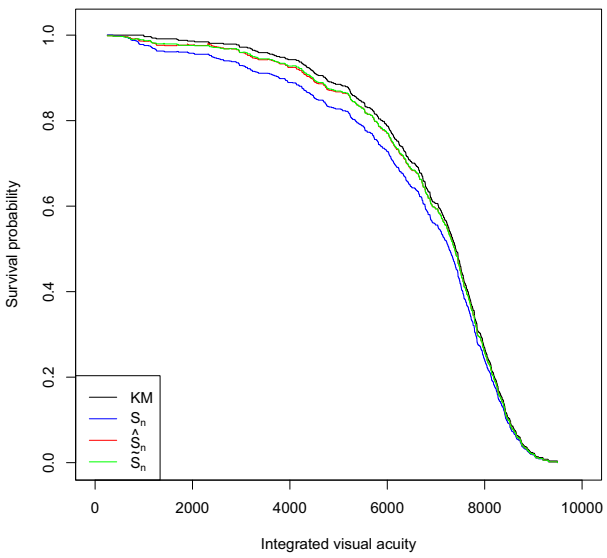


(a) Survival curve

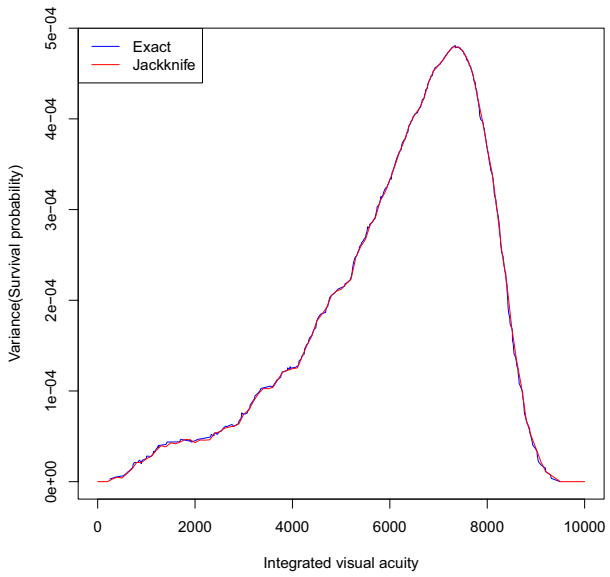


(b) Variance comparison

Fig. 8 Estimates of the survival probabilities of the integrated VA for the regular group of CATT study data (a) and the corresponding variances computed by using the analytical formula and jackknife method for $\widehat{S}_n$ method (b) are plotted. Three estimates S_n, $\widehat{S}_n$ and $\widetilde{S}_n$ of the survival function $P(\Psi(\tau) > y)$ of the integrated visual acuity for regular group of patients from the CATT study (obtained by setting $t = 0$ in (2), (3) and (5), respectively. KM corresponds to the naive Kaplan-Meier estimator of the survival function of $\Psi(\tau)$ when $\Psi(\tau)$ is censored whenever τ (time to discontinuation of the treatment) is censored by a censoring variable C. Regular group consists of patients who received the treatment every four weeks throughout the follow-up

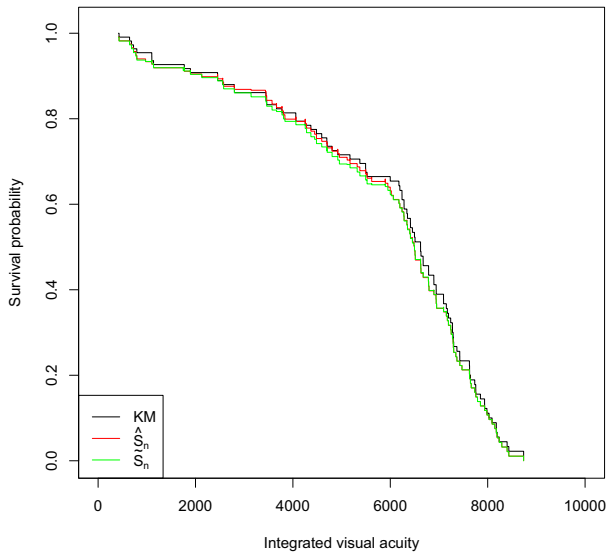


(a) Survival curve

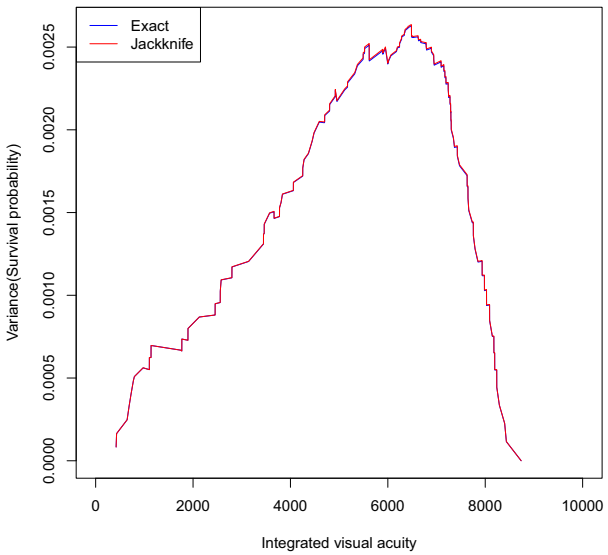


(b) Variance comparison

Fig. 9 Estimates of the survival probabilities of the integrated VA for the variable group of CATT study data (a) and the corresponding variances computed by using the analytical formula and jackknife method for $\widehat{S}_n$ method (b) are plotted. Three estimates S_n, $\widehat{S}_n$ and $\widetilde{S}_n$ of the survival function $P(\Psi(\tau) > y)$ of the integrated visual acuity for variable group of patients from the CATT study (obtained by setting $t = 0$ in (2), (3) and (5), respectively. KM corresponds to the naive Kaplan-Meier estimator of the survival function of $\Psi(\tau)$ when $\Psi(\tau)$ is censored whenever τ (time to discontinuation of the treatment) is censored by a censoring variable C. Variable group consists of patients who received the treatment as per *PRN* protocol

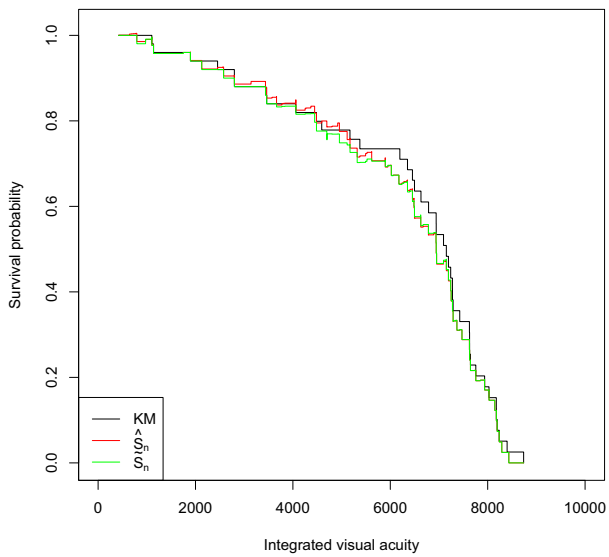


(a) Survival curve

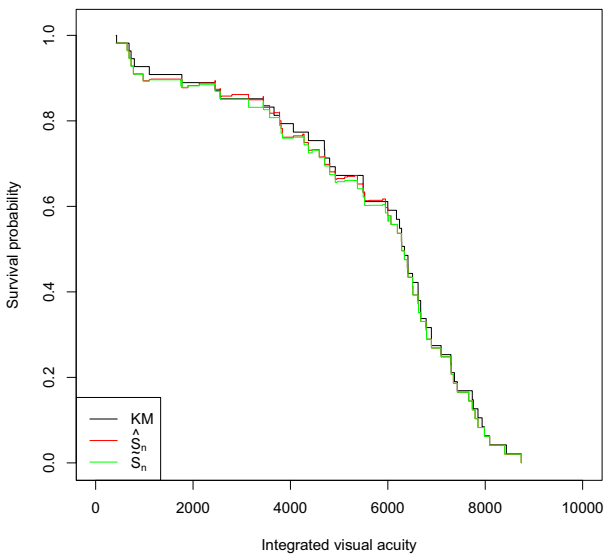


(b) Variance comparison

Fig. 10 Estimates of the survival probabilities of the integrated VA for 2-year follow-up of the real-world data (a) and the corresponding variance computed by using the analytical formula and jackknife method for $\widehat{S}_n$ method (b). Two estimates $\widehat{S}_n$ and $\widetilde{S}_n$ of the survival function $P(\Psi(\tau) > y)$ of the integrated visual acuity for variable group of patients from the CATT study (obtained by setting $t = 0$ in (3) and (5), respectively. KM corresponds to the naive Kaplan-Meier estimator of the survival function of $\Psi(\tau)$ when $\Psi(\tau)$ is censored whenever τ (time to discontinuation of the treatment) is censored by a censoring variable C

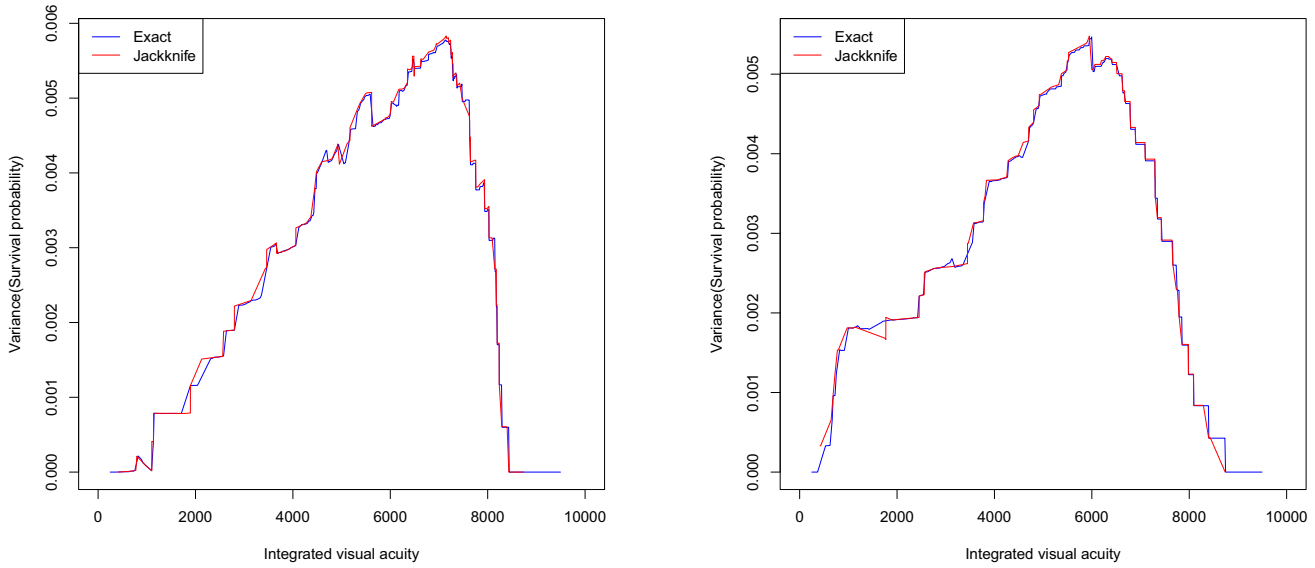


(a) Baseline IRF = 0



(b) Baseline IRF = 1

Fig. 11 Estimates of the survival probabilities of the integrated VA at $t = 0$ for the subset of the real-world data with IRF absent in the baseline measurement (a) and for the subset with IRF present in the baseline measurement (b)



(a) Baseline IRF = 0 (b) Baseline IRF = 1

Fig. 12 Variance in survival probability estimates for integrated VA at $t = 0$ for the subset of the real-world data with IRF absent in the baseline measurement (a) and for the subset with IRF present in the baseline measurement (b)

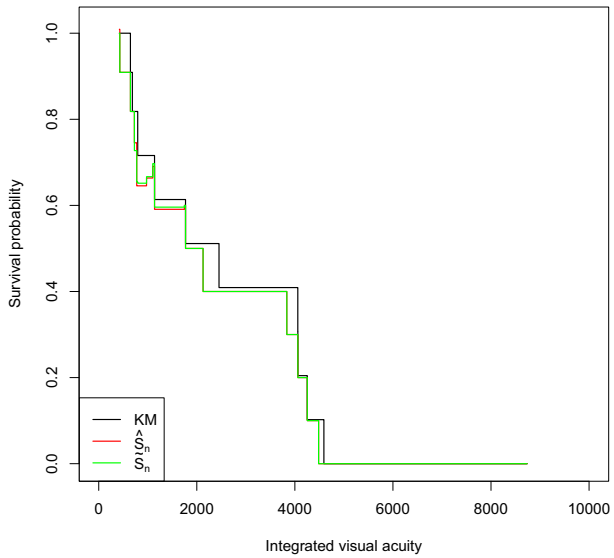
the baseline IRF status was not available and they were not included in this analysis. Survival curves and corresponding variance curves for these subgroups are shown in Fig. 11 and Fig. 12.

F.2.3 Strata based on baseline VA

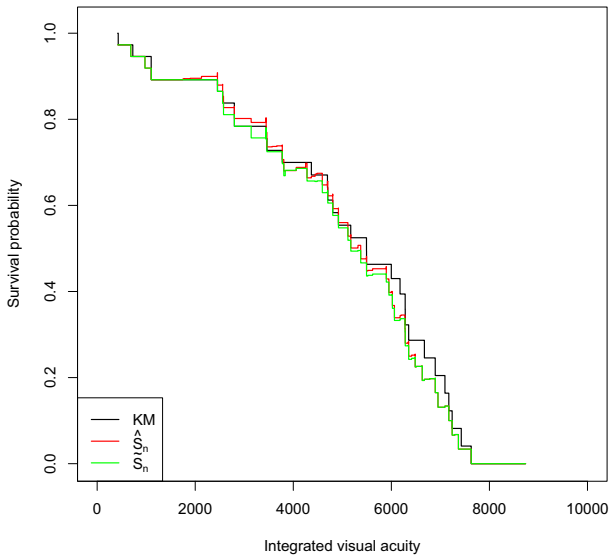
Full data from real-world study is divided into 4 subgroups based on the visual acuity at baseline. The survival curves and corresponding variance curves for these subgroups are shown in Fig. 13 and Fig. 14.

F.2.4 Conditional cumulative hazard and conditional survival

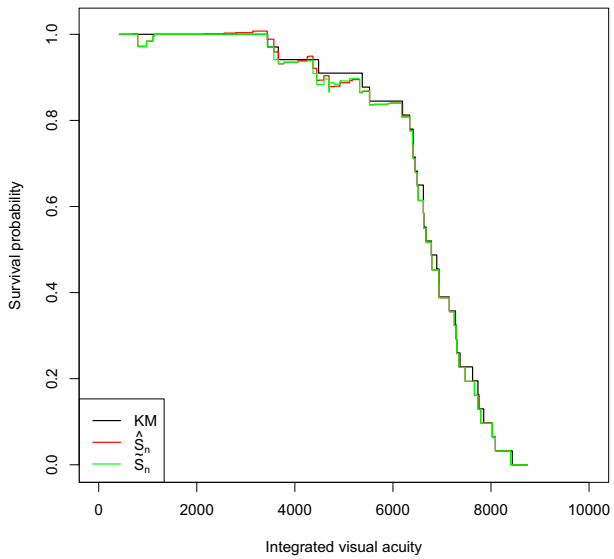
In conditional cumulative hazard plots in Fig. 5 of the article, it is difficult to view all four curves due to overlapping. In Fig. 15, the same data is reproduced but up to week 80 only for better display of four conditions. In Fig. 16 and Fig. 17, conditional survival probabilities along with their confidence limits are plotted for specific condition on covariates, VA_AUC and number of injections.



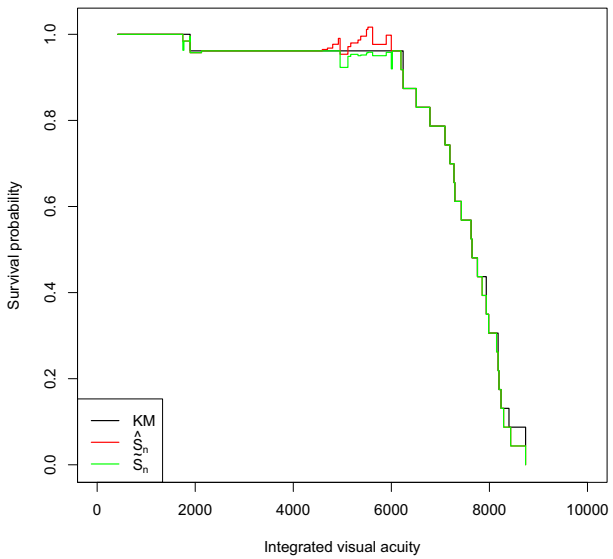
(a) Baseline VA < 35



(b) Baseline $35 \leq$ VA < 55

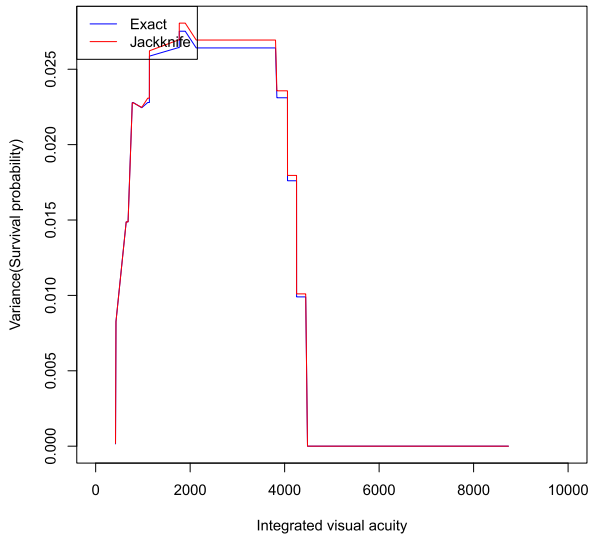


(c) Baseline $55 \leq$ VA < 70

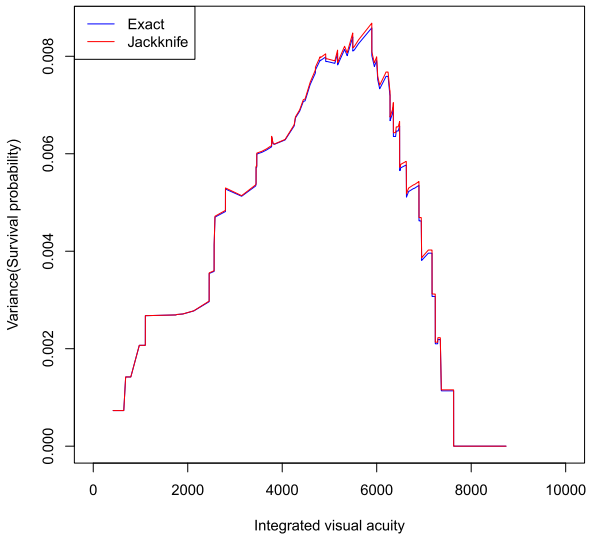


(d) Baseline VA ≥ 70

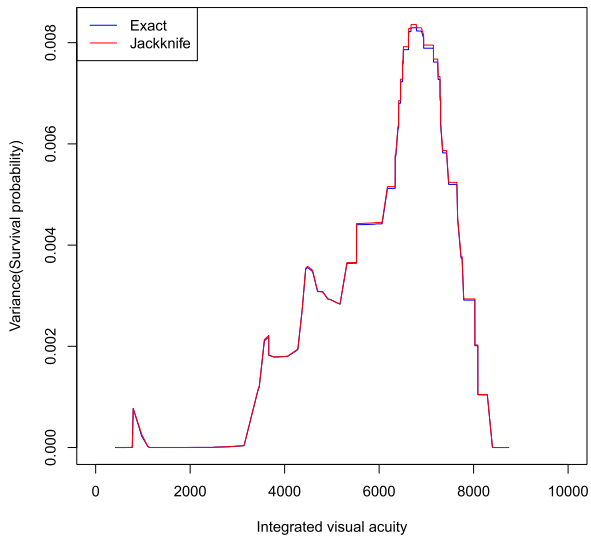
Fig. 13 Estimates of the survival probabilities of integrated VA at $t = 0$ when the real-world data are stratified according to the baseline visual acuity – (a) VA < 35 (number of patients $n = 11$), (b) $35 \leq$ VA < 55 ($n = 37$), (c) $55 \leq$ VA < 70 ($n = 35$) and (d) VA ≥ 70 ($n = 27$)



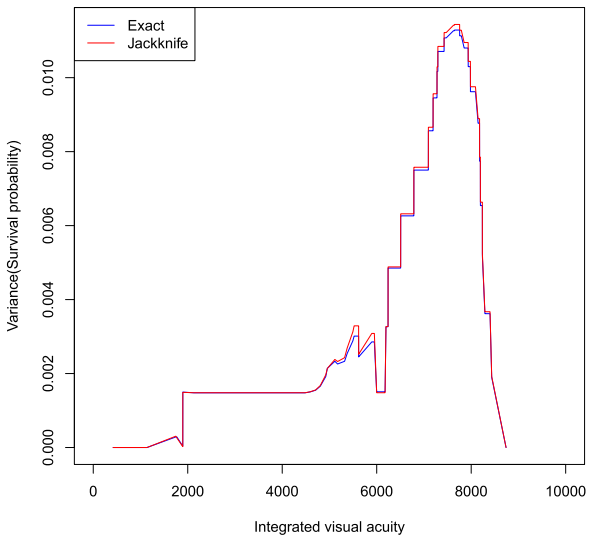
(a) Baseline VA $<$ 35



(b) Baseline VA $>=$ 35 and $<$ 55



(c) Baseline VA $>=$ 55 and $<$ 70



(d) Baseline VA $>=$ 70

Fig. 14 Variance in survival probability estimates of integrated VA at $t = 0$ for the subset of the real-world data with - (a) VA < 35 , (b) $35 \leq$ VA < 55, (c) $55 \leq$ VA < 70 and (d) VA ≥ 70

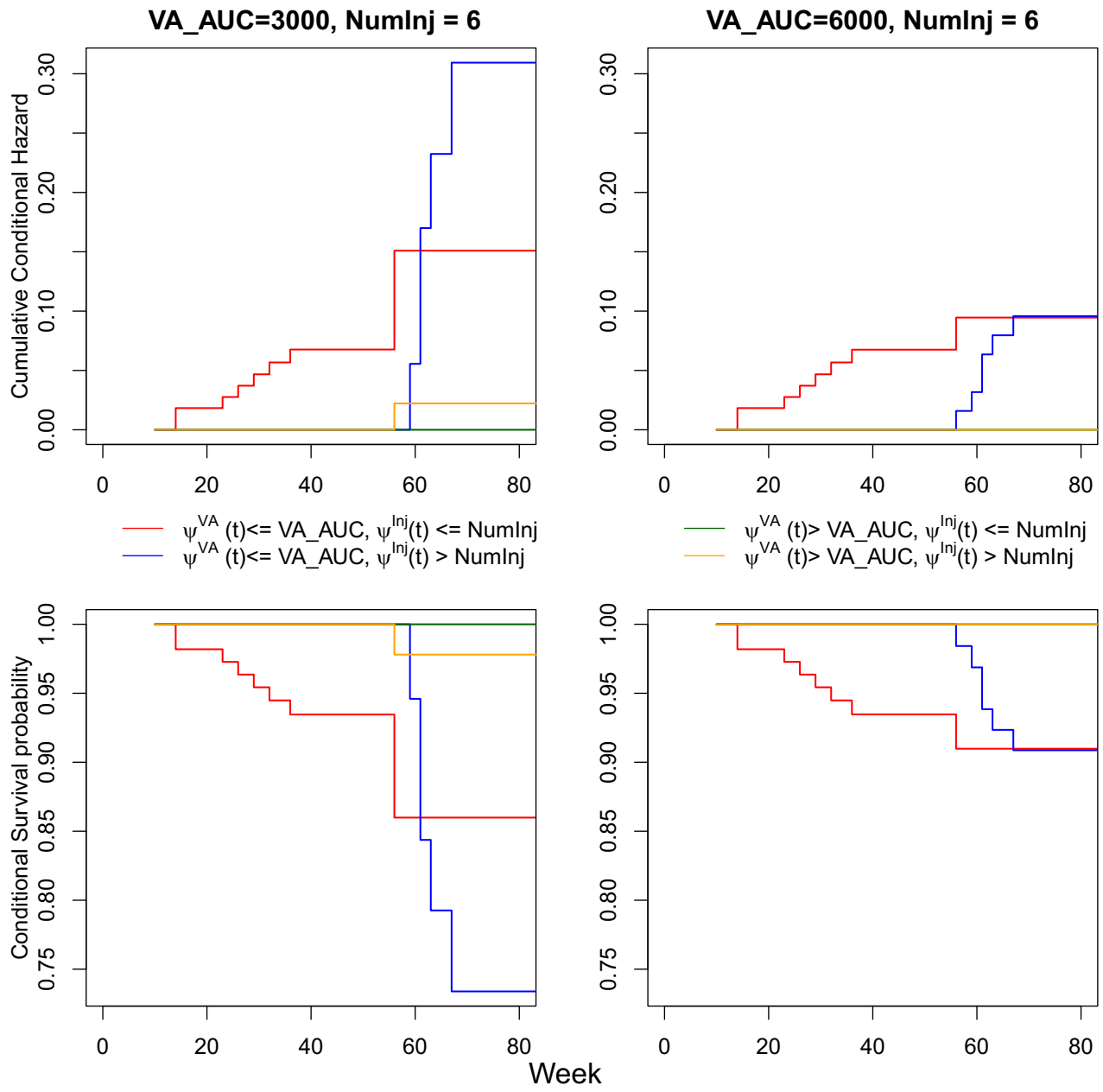


Fig. 15 Conditional cumulative hazard and conditional survival function of four combinations of conditions on benefit (integrated VA) and cost (total number of injections) are plotted for 2 values of integrated VA and number of injections = 6

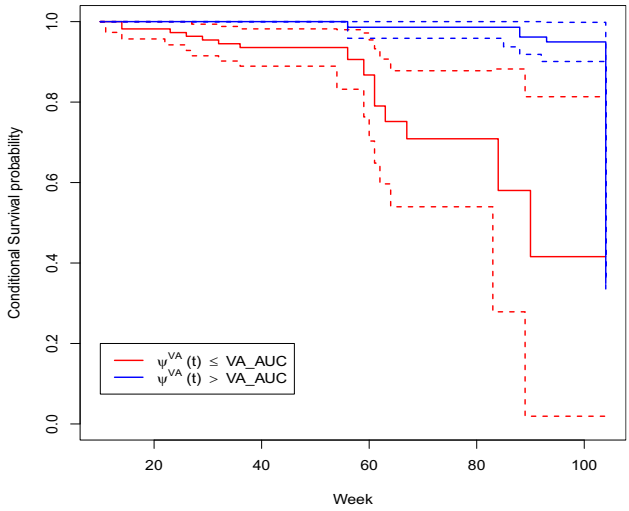


Fig. 16 Conditional survival probabilities along with their confidence limits computed using jackknife variance are plotted for VA_AUC=3000

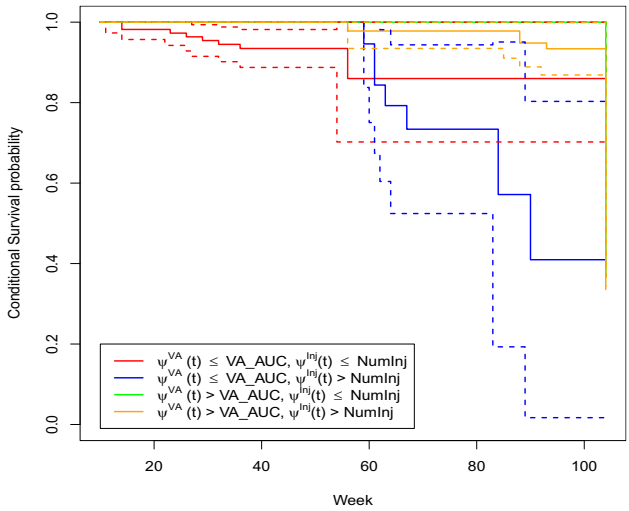


Fig. 17 Conditional survival probabilities along with their confidence limits computed using jackknife variance are plotted for VA_AUC=3000 and number of injections = 6

Acknowledgements We thank Dr. Ilkka Immonen and Dr. Terhi Ollila (Helsinki University Hospital) for providing the knowledge on wet-AMD disease and making the real-world data available for illustration. We are grateful to Prof. Isha Dewan (Stat. Math. Unit, Indian Statistical Institute, New Delhi, India) for useful discussions and a critical review of the manuscript. We are grateful to Center for Preventive Ophthalmology and Biostatistics (CPOB) (Philadelphia, PA 19104) for making the CATT data available for research purposes. We would like to thank the associate editor and the reviewers for their constructive comments.

Author Contributions A.J. and S.K. started developing the statistical method for joint survival function and writing the initial draft. D.G. introduced the conditional hazard function theory and provided guidance on theoretical accuracy of results. A.J. performed all computations, graphical representation and wrote all drafts. S.K. provided overall statistical guidance in all aspects of developing new methods, writing and choosing appropriate datasets. All authors reviewed and approved the manuscript.

Funding Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital). The first author received grant from ALFRED KORDELIN FOUNDATION (Code: b48699) for this research work.

Data Availability Data sources are properly cited in the manuscript. The real-world data are not openly available while CATT study data can be requested for research from the website https://www.med.upenn.edu/cpob/catt.html.

Declarations

Conflict of interest/Competing interests The authors have no competing interests to declare.

Ethics approval and consent to participate 'Not applicable'.

Materials availability 'Not applicable'.

Code availability On request.

Competing interests The authors declare no competing interests.

References

Aalen, O.O., Borgan, Ø., Gjessing, H.K.: Survival and Event History Analysis. Springer Science+Business Media LLC, A process point of view (2008)

Akritas, M., Keilegom, I.: Estimation of bivariate and marginal distributions with censored data. J. R. Stat. Soc. Ser. B Stat Methodol. **65**(2), 457–471 (2003). https://doi.org/10.1111/1467-9868.00396

Andersen, P., Borgan, Ø., Gill, R., Keiding, N.: Statistical models based on counting processes. In Springer series in statistics (p. 779). Springer (1992)

Bell, M.L., King, M.T., Fairclough, D.L.: Bias in area under the curve for longitudinal clinical trials with missing patient reported outcome data, (2014). (SAGE open, 4.)

Bonate, P.L.: Coverage and precision of confidence intervals for area under the curve using parametric and nonparametric methods in a toxicokinetic experimental design. Pharm. Res. **15**(3), 405–410 (1998). https://doi.org/10.1023/a:1011968129921

Bull, K., Spiegelhalter, D.J.: Tutorial in biostatistics survival analysis in observational studies. Stat. Med. **16**, 1041–1074 (1997)

Chakravarthy, U., Evans, J., Rosenfeld, P.: Clinical review, age related macular degeneration. British Medical **340**, c981 (2010)

Cook, R.J., Lawless, J.F.: Multistate Models for the Analysis of Life History Data, Chapman and Hall CRC (2018). https://doi.org/10.1201/9781315119731. (1st ed.)

Dabrowska, D.M.: Kaplan-meier estimate on the plane. Ann. Stat. **16**(4), 1475–1489 (1988). (https://www.jstor.org/stable/2241775)

Datta, S., Bandyopadhyay, D., Satten, G.: Inverse probability of censoring weighted u statistics for right censored data with an application to testing hypotheses. Scand. J. Stat. **37**(4), 680–700 (2010). https://doi.org/10.1111/j.1467-9469.2010.00697.x

Elshout, M., Webers, C.A., van der Reis, M.I., de Jong-Hesse, Y., Schouten, J.S.: Tracing the natural course of visual acuity and quality of life in neovascular age-related macular degeneration: A systematic review and quality of life study. BMC Ophthalmol. **17**(1), 120 (2017). https://doi.org/10.1186/s12886-017-0514-3

Faucett, C.L., Thomas, D.C.: SIMULTANEOUSLY MODELLING CENSORED SURVIVAL DATA AND REPEATEDLY MEASURED COVARIATES: A GIBBS SAMPLING APPROACH. Statist. Med. **15**, 1663–1685 (1996). https://doi.org/10.1002/(SICI)1097-0258(19960815)15:15<1663::AID-SIM294>3.0.CO;2-1

Fleming, T., Harrington, D.: Counting processes and survival analysis. In Counting processes and survival analysis. Wiley Series in Probability; Mathematical Statistics: Applied Probability; Statistics (1991)

Gelber, R., Gelman, R., Goldhirsch, A.: A quality-of-life-oriented endpoint for comparing therapies. Biometrics **45**(3), 781–795 (1989). https://doi.org/10.2307/2531683

Giurcanu, M., Karrison, T.: Nonparametric estimators of hazard ratios for comparing two survival curves. Biometrics **81**(2), ujaf072 (2025). https://doi.org/10.1093/biomtc/ujaf072

Gross, J.G., Glassman, A.R., Jampol, L.M., Inusah, S., Aiello, L.P., Antoszyk, A.N., Baker, C.W., Berger, B.B., Bressler, N.M., Browning, D., Elman, M.J., Ferris, 3, F. L., Friedman, S. M., Marcus, D. M., Melia, M., Stockdale, C. R., Sun, J. K., Beck, R. W., Network, D.R. C. R.: Panretinal photocoagulation vs intravitreous ranibizumab for proliferative diabetic retinopathy: A randomized clinical trial. JAMA **314**(20), 2137–2146 (2015) https://doi.org/10.1001/jama.2015.15217

Huang, Y., Louis, T.: Nonparametric estimation of the joint distribution of survival time and mark variables. Biometrika **85**(4), 785–798 (1998). https://doi.org/10.1093/biomet/85.4.785

Ibrahim, J.G., Chu, H., Chen, L.M.: Basic concepts and methods for joint models of longitudinal and survival data. J. Clin. Oncol. **28**(16), 2796–801 (2010) https://doi.org/10.1200/JCO.2009.25.0654. Epub 2010 May 3. PMID: 20439643; PMCID: PMC4503792

Janssen, P., Veraverbeke, N.: Nonparametric estimation of univariate and bivariate survival functions under right censoring: A survey. Metrika **87**(3), 211–245 (2023). https://doi.org/10.1007/s00184-023-00911-7

Keilegom, V.I., Veraverbeke, N.: Hazard rate estimation in nonparametric regression with censored data. Ann. Inst. Stat. Math. **53**(4), 730–745 (2001). https://doi.org/10.1023/a:1014696717644

Lin, D.Y., Y., Z.: A simple nonparametric estimator of the bivariate. survival function under univariate censoring. Biometrika **80**(3), 573–581 (1993)

Ollila, T., Silvennoinen, J., Joshi, A., Liu, J., Kulathinal, S., Immonen, I.: Analysing subgroups and treatment discontinuation in a finnish cohort of patients with neovascular amd. Ophthalmologica. International journal of ophthalmology. **245**(4), 358–367 (2022). https://doi.org/10.1159/000524848

Putter, H., van Houwelingen, H.C.: Landmarking 2.0:Bridging the gap between joint models and landmarking. Stat. Med. **41**(11), 1901–1917 (2022). https://doi.org/10.1002/sim.9336

Robins, J. M., Rotnitzky, A.: Recovery of information and adjustment for dependent censoring using surrogate markers. In J. NP, K. D,& F. VT. (Eds.), Aids epidemiology: Methodological issues (pp. 297–331). Birkhäuser Boston. (1992) https://doi.org/10.1007/978-1-4757-1229-214

Satten, G., Datta, S.: The kaplan-meier estimator as an inverse-probability-of-censoring weighted average. Am. Stat. **55**(3), 207–210 (2001). https://doi.org/10.1198/000313001317098185

Shah, A.R., Del Priore, L.V.: Progressive visual loss in subfoveal exudation in age-related macular degeneration: A meta-analysis using lineweaver burke plots. Am. J. Ophthalmol. **143**(1), 83–89 (2007). https://doi.org/10.1016/j.ajo.2006.09.043

Tsiatis, A.A., Degruttola, V., Wulfsohn, M.S.: Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS. J. Am. Stat. Assoc. **90**(429), 27–37 (1995). https://doi.org/10.1080/01621459.1995.10476485

Tuuminen, R., Uusitalo-Järvinen, H., Aaltonen, V., Hautala, N., Kaipiainen, S., Laitamäki, N., Ollila, M., Rantanen, J., Välimäki, S., Sipilä, R., Laukkala, T., Komulainen, J., Tommila, P., Immonen, I., Tuulonen, A., Kaarni-ranta, K.: The finnish national guideline for diagnosis, treatment and follow-up of patients with wet age-related macular degeneration. Acta Ophthalmol. **95**(A105 Suppl), 1–9 (2017). https://doi.org/10.1111/aos.13501

van Houwelingen, H., Putter, H.: Dynamic Prediction in Clinical Survival Analysis (1st ed.). CRC Press. (2011) https://doi.org/10.1201/b11311

Van Houwelingen, H.C.: Dynamic prediction by landmarking in event history analysis. Scand. J. Stat. **34**, 70–85 (2007)

Van Keilegom, I., Akritas, M.G., Veraverbeke, N.: Estimation of the conditional distribution in regression with censored data: A comparative study. Computational statistics & data analysis **35**(4), 487–500 (2001)

Wang, J.-L., Zhong, Q.: Joint Modeling of Longitudinal and Survival Data. Annual Review Statistics and Its Application. **12**, 449–476 (2025). https://doi.org/10.1146/annurev-statistics-112723-034334

Wilding, G.E., Chandrasekhar, R., Hutson, A.D.: A new linear model-based approach for inferences about the mean area under the curve. Stat. Med. **31**(28), 3563–3578 (2012). https://doi.org/10.1002/sim.5387

Zhao, H., Tsiatis, A.: A consistent estimator for the distribution of quality adjusted survival time. Biometrika **84**(2), 339–348 (1997). https://doi.org/10.1093/biomet/84.2.339

Zhao, H., Tsiatis, A.: Efficient estimation of the distribution of quality-adjusted survival time. Biometrics **55**(4), 1101–1107 (1999). https://doi.org/10.1111/j.0006-341x.1999.01101.x

Zhao, H., Bang, H., Wang, H., Pfeifer, P.: On the equivalence of some medical cost estimators with censored data. Stat. Med. **26**, 4520–4530 (2007). https://doi.org/10.1002/sim.288240