Comparison of Statistical Models For Estimating Intervention Effects Based On Time-To-Recurrent-Event In Stepped Wedge Cluster Randomized Trial Using Open Cohort Design

Shunsuke Oyamada (shunsuke.oyamada@jortc.jp)
JORTC Data Center
Shih-Wei Chiu
Tohoku University Graduate School of Medicine
Takuhiro Yamaguchi
Tohoku University Graduate School of Medicine

Research Article

Keywords: Stepped-wedge, Cluster randomized trial, Open cohort design, Recurrent event, Time-to-event, Statistical model, Time-dependent covariate, Simulation, Comparison

Posted Date: November 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-999906/v1

Version of Record: A version of this preprint was published at BMC Medical Research Methodology on April 26th, 2022. See the published version at https://doi.org/10.1186/s12874-022-01552-6.

Comparison of statistical models for estimating intervention effects based on time-to-recurrent-event in stepped wedge cluster randomized trial using open cohort design

Shunsuke Oyamada[1, 2*], Shih-Wei Chiu[1], Takuhiro Yamaguchi[1]

Author details

[1] Division of Biostatistics, Tohoku University Graduate School of Medicine, Sendai, Japan.

[2] Departments of Biostatistics, JORTC Data Center, Tokyo, Japan.

[*] Correspondence: shunsuke.oyamada@jortc.jp

Abbreviations

RCT Randomized Controlled Trial

CRT Cluster Randomized Trial

SWCRT Stepped Wedge Cluster Randomized Trial

TTHA Time To Hospital Admission

TTE Time-To-Event

TTFE Time-To-First-Event

TTRE Time-To-Recurrent-Event

TTTE Time-To-Terminal-Event

CoxPH Cox Proportional Hazard

AG Andersen-Gill

PWP-TT Prentice-Williams-Peterson Total-Time

PWP-GT Prentice-Williams-Peterson Gap-time

HR Hazard Ratio

MSE Mean Square Error

CP Coverage Probability

CI Confidence Interval

SE Standard Error

Abstract

Background: There are currently no methodological studies on the performance of the statistical models for estimating intervention effects based on the time-to-recurrent-event (TTRE) in stepped wedge cluster randomised trial (SWCRT) using an open cohort design. This study aims to address this by evaluating the performance of these statistical models using an open cohort design with the Monte Carlo simulation in various settings and their application using an actual example.

Methods: Using Monte Carlo simulations, we evaluated the performance of the existing extended Cox proportional hazard models, i.e., the Andersen-Gill (AG), Prentice-Williams-Peterson Total-Time (PWP-TT), and Prentice-Williams-Peterson Gap-time (PWP-GT) models, using the settings of several event generation models and true intervention effects, with and without stratification by clusters. Unidirectional switching in SWCRT was represented using time-dependent covariates.

Results: Using Monte Carlo simulations with the various described settings, the PWP-GT model with stratification by clusters showed the best performance in most settings and reasonable performance in the others. The only situation in which the performance of the PWP-TT model with stratification by clusters was not inferior to that of the PWP-GT model with stratification by clusters was when there was a certain amount of follow-up period, and the timing of the trial entry was random within the trial period, including the follow-up period. The AG model performed well only in a specific setting. By analysing actual examples, it was found that almost all the statistical models suggested that the risk of events during the intervention condition may be somewhat higher than in the control, although the difference was not statistically significant.

Conclusions: The PWP-GT model with stratification by clusters had the most reasonable performance when estimating intervention effects based on the TTRE in SWCRT in various settings using an open cohort design.

Keywords

Stepped-wedge, Cluster randomized trial, Open cohort design, Recurrent event, Time-to-event, Statistical model, Time-dependent covariate, Simulation, Comparison

Background

A cluster randomised trial (CRT) is a randomised trial design in which a cluster of regions or sites is used when it is not possible or appropriate to assign an intervention to an individual patient, like a randomised controlled trial (RCT) [1, 2]. The stepped wedge CRT (SWCRT) is a type of CRT, in which the order that the interventions are applied to the clusters is randomised, and all clusters are sequentially transferred (unidirectional switch) from the control condition to the intervention condition [3, 4].

There are three main types of SWCRT design: (i) continuous recruitment short exposure, (ii) closed cohort, and (iii) open cohort [5]. In the open cohort design, each subject is assessed repeatedly at a series of measurement points or at a subject-specific time point, such as the occurrence of an event. In this design, each subject may enter and leave the trial at any time during the trial period. Thus, some subjects are exposed to both control and intervention conditions during the trial, while others are only exposed to one.

The INSPIRED trial, which is the actual example used in this study, was a multi-centre SWCRT that

77 examines whether a model of care that provides specialist palliative care interventions in residential care
78 homes (i.e. the intervention condition) leads to fewer (acute care) hospitalisations and shorter lengths of stay
79 in hospital for care home residents, when compared to usual care (i.e. the control condition) [6]. A schematic
80 representation of actual example is presented in Fig. 1. It is an open cohort design as all the residents in each
81 facility at the start of the trial and all-new enrolments to the facility after the start of the trial were included.
82 Many residents were exposed to both the control and intervention conditions, as they remained in their
83 residences continuously unless they died or were discharged from the care home. The primary outcome was
84 the length of the hospital stays, and the secondary outcomes were the number of hospitalisations and the cost.
85 Some residents never experienced hospitalisation, while others were repeatedly hospitalised in the actual
86 example. If the same event occurs repeatedly over time to the same individual it is called a recurrent event
87 [7]. A common way to analyse recurrent event is the recurrence rate (average number of recurrences per unit
88 time), which corresponds to the No. of hospitalisations per facility-month as a secondary outcome in the
89 actual example. This analysis requires the assumption that the incidence of hospitalisation is always constant
90 in the interval per facility-month which is generally a strong assumption. In addition, even if the number of
91 hospitalisations per facility-month are the same, there may be differences in the time it takes for each
92 hospitalization to occur, and this is called the time to hospital admission (TTHA) and may represent the
93 effects of the intervention. Since admission and discharge data are collected for each hospitalisation in the
94 actual example, that is, the TTHA is measured repeatedly, it may be useful to evaluate hospitalisations as
95 recurrent events within the framework of a time-to-event (TTE) analysis.

When assessing the impact of a covariate on the TTE with a hazard ratio (HR), the Cox proportional hazard (CoxPH) model is most often used [8], and it assumes that the event is a one-time terminal event. When the CoxPH model is applied to recurrent events, only time-to-first-event (TTFE) can be included in the analysis. Against this background, the extension of the CoxPH model to recurrent events has been actively pursued, especially in the 1980s [9-11], and it has mainly been used to evaluate the time-to-recurrent-event (TTRE) in RCTs.

Methods to analyse TTE in SWCRT are currently unclear [12]. SWCRT using an open cohort design, by its nature, must deal with subjects who are exposed to both the control and intervention conditions (observed across the unidirectional switch). When estimating the intervention effects based on the TTFE, if the change in the time-dependent covariate is independent of TTE, then the unidirectional switch in the CoxPH model can be explained using the time-dependent covariate [13, 14], and methodological studies on the performance in the context of SWCRT have previously been conducted [15]. In TTRE, the existing extended CoxPH model with time-dependent covariates possibly apply to SWCRT with unidirectional switching [16, 17, 18]. In addition, CRT is known to have a problem with cluster effects when the outcomes of individuals in the same cluster become similar for various reasons, and this is also a concern for SWCRT. When estimating intervention effects based on TTFE, the cluster effect in SWCRT can be treated using the CoxPH model stratified by clusters [15], as it assumes that each cluster's baseline hazard function is different. For TTRE, the existing extended CoxPH model stratified by clusters possibly be used.

However, there are currently no methodological studies on the performance of the statistical models for

estimating intervention effects based on TTRE in SWCRT with an open cohort design, or examples of its application to actual studies. Investigating the performance of the statistical models used to estimate intervention effects based on TTRE in SWCRT using an open cohort design in various settings, may contribute to the selection of statistical models for the actual planning and analysis of SWCRT.

The purpose of this study was to evaluate "which statistical models resulted in better performance estimating intervention effects using TTRE in SWCRT with an open cohort design" with the Monte Carlo simulation (hereafter, simulation) in various settings. We also applied each statistical model to hospital admission data to test the actual example and interpreted the results based on the simulation results.

Methods

Actual example

Details of the trial design, interventions, resident background information, and efficacy results of the INSPIRED trial have been published previously [6]. The trial included 1700 residents from 12 care homes in Australia, of which 1089 (64.1%) were residents at the start of the trial, and the remaining 611 (35.9%) became residents after the start of the trial. There were 1149 hospitalisations during the trial, of which 943 hospitalizations of more than 24 hours (> 24 h) were used for the primary outcome, length of stay in hospital. Of the residents, 377 had only one hospitalization of > 24 h, while 211 had multiple hospitalizations of > 24 h (137 had two, 45 had three, 11 had four, and 18 had four or more). The number of residents who died during the trial period was 534 (31.4%). The secondary outcome, No. of hospitalizations > 24 h per facility-month,

was 5.6 in the control condition and 4.3 in the intervention condition, a decrease of approximately 23% (no adjustment by covariates or comparison by estimation/statistical testing was performed).

Basic notation

The timing of the unidirectional switch (henceforth, switch) in each cluster of the SWCRT is called a step, and here, we consider SWCRT with m clusters and s steps. For simplicity, we assume that the number of clusters to be switched from the control condition to the intervention condition in one step is one ($s = m$). In the ith cluster ($i = 1, \ldots, m$), n_i is the number of subjects observed during the entire trial duration.

Assuming that the start of the test is t_S and the end of the last step period is t_E, the timing of the switch in each cluster is calculated as follows: $W_i = t_S + i * (t_E - t_S) \ / \ (m + 1)$, and the distance between switches is calculated as follows: $W_d = W_{i+1} - W_i = W_i - W_{i-1} = (t_E - t_S) \ / \ (m + 1)$. Let d_{ij} be the time point at which the jth subject ($j = 1, \ldots, n_i$) in the ith cluster entered the trial. The distance w_{ij} to the switch for each subject is defined as follows:

$$\begin{cases} w_{ij} = W_i - d_{ij} & W_i \geqslant d_{ij} \\ w_{ij} = 0 & W_i < d_{ij} \end{cases}.$$

Where $h_{ijk}(t)$ is the hazard function of the kth recurrence of the jth subject in the ith cluster at time t, and $h_{0ik}(t)$ is the baseline hazard function of the kth recurrence of the ith cluster at time t. $Y_{ijk}(t)$ is the indicator variable for the kth recurrence of the jth subject in the ith cluster at time t, and this is 1 if the subject is at risk of recurrence and under observation, and 0 if not. X_{ijk} is a vector of time-independent covariates for the kth recurrence of the jth subject in the ith cluster, and β_{ik} is a vector of

fixed parameters for the time-independent covariates of the kth recurrence of the ith cluster. $Z_{ijk}(t)$ is the intervention indicator as a time-dependent covariate for the kth recurrence of the jth subject in the ith cluster, which is 0 for $t < w_{ij}$ and 1 for $t \geqslant w_{ij}$ (changes before and after the switch). β_{tik} is the parameter for the intervention effect for the kth recurrence of the ith cluster. The subscript i is omitted if it is assumed that each cluster has a common effect. The subscript k is omitted if it is assumed that each recurrence has a common effect.

Statistical models

The first model considered was the CoxPH model [8, 14]. The hazard of the jth subject in the ith cluster at time t is expressed as follows:.

$$h_{ij}(t) = h_{0i}(t)\ \exp(\beta_{ti} Z_{ij}(t) + \beta_i' X_{ij}).$$

As was previously mentioned, applying the CoxPH model to recurrent events would result in a loss of information because only the TTFE of each subject can be included in the analysis, and the second and subsequent events are ignored. Taking recurrent events into account should theoretically improve the efficiency of estimating the effects of interventions [19]. In the following, we present an extended CoxPH model that allows for the inclusion of TTRE in the analysis.

The Andersen and Gill (AG) model assumes a common baseline hazard function for all events, independent of the number of previous recurrences, and it is considered beneficial when investigating the overall intervention effect on the occurrence of recurrent events [9]. The hazard for the jth subject in the ith cluster

172 at time t is expressed as follows:

173 $$h_{ij}(t) = Y_{ij}(t)\ h_{0i}(t)\ \exp(\beta_{ti} Z_{ijk}(t) + \beta_i' X_{ijk}).$$

174 In the usual CoxPH model, a subject who has experienced one event is no longer at risk for that event. In

175 contrast, the AG model assumes that subjects who have experienced at least one event remain at risk unless

176 they drop out of the trial. In the AG model, multiple events that occur within the same subject are considered

177 to be independent. However, because they may not be independent in reality, it is advised that robust variance

178 is used to handle the correlation within the subject when inferring the parameter vector [20, 21].

179 The Prentice-Williams-Peterson (PWP) model assumes a different baseline hazard function for each

180 recurrence and accounts for correlation by stratifying by the number of prior recurrences. Therefore, it is

181 considered beneficial when the risk of recurrence differs between recurrences [17]. The hazard $h_{ijk}(t)$ for

182 the kth recurrence is defined by the history of the covariates and the number of recurrences up to time t.

183 Conditionally, it is assumed that the $(k-1)$th recurrence is independent of the kth recurrence. Furthermore,

184 it assumes that the subject is not at risk for the kth recurrence until the $(k-1)$th recurrence, so that $Y_{ijk}(t)$

185 is 0 until the $(k-1)$th recurrence and 1 after that.

186 The PWP model can be broadly divided into two models depending on the treatment of the time points.

187 First, the PWP total-time (PWP-TT) model uses the time from the start of the follow-up to each recurrence.

188 The hazard of the kth recurrence of the jth subject in the ith cluster at time t is expressed as follows:

189 $$h_{ijk}(t) = Y_{ijk}(t)\ h_{0ik}(t)\ \exp(\beta_{tik} Z_{ijk}(t) + \beta_{ik}' X_{ijk}).$$

190 The second is the PWP gap-time (PWP-GT) model, which uses the time from the occurrence of the

previous recurrence to each recurrence. The hazard of the kth recurrence for the jth subject in the ith cluster at time t is expressed as:

$$h_{ijk}(t) = Y_{ijk}(t)\ h_{0ik}(t - t_{k-1})\ \exp(\beta_{tik} Z_{ijk}(t) + \beta'_{ik} X_{ijk}).$$

As the number of recurrences increases in the PWP model, the number of subjects at risk becomes relatively small. This would make the estimates unstable, so limiting the data to a specific number of recurrences is usually necessary [22]. Due to these characteristics, the PWP model is helpful in situations where the number of recurrences per subject is small [17]. Our study assumes that each recurrence has a common effect when estimating parameters using the PWP model.

For each of the statistical models described so far, there are two analysis policies: (i) with stratification by clusters, which assumes that the baseline hazard function is different for each cluster, and (ii) without stratification by clusters, which assumes that the baseline hazard function is the same for each cluster.

The performance of each statistical model in the simulation was evaluated in terms of bias, mean square error (MSE), and coverage probability (CP). Bias is the difference between the estimates of the parameters of the intervention effect based on each statistical model and the true intervention effect β_t, where a positive value indicates underestimation and a negative value indicates overestimation; MSE is measured by the variance of the estimated intervention effect based on each statistical model and considers both bias and variability, with smaller values indicating better performance. CP is the proportion of the 95% confidence interval (CI) for the HR obtained by each statistical model that includes the HR based on the true intervention effect β_t. The closer the CI is to 0.95, the better the performance.

Data generation process

For the time point d_{ij} of the jth subject in the ith cluster to enter in the trial, we use t_S at the beginning of the trial and t_E at the end of the last step period already mentioned, and generate them randomly within the interval of $t_S + ((t_E - t_S) * e)/E$ or $t_S + ((t_F - t_S) * e)/E$. From this point, the TTFE at least, always occurs starting from d_{ij}. Here, e is a pseudo-random number generated from a uniform distribution, $e \sim U(0,1)$.

t_F indicates the end of the trial and is expressed as $t_F = t_E + (W_d * F)$ using the distance W_d between t_E and the switch at the end of the last step period, as described above. F is a coefficient that specifies the follow-up period that may be set after the end of the last step period. When $F = 0$, there is no follow-up period, and $t_F = t_E$. If $F = X(> 1)$, there is a follow-up period of X step after the end of the last step period. In the actual example, as shown in Fig. 1, each step is set every two months, and there is a follow-up period of 5 months (= 2.5 steps) after the end of the last step period. Based on the purpose and setting of the trial, other SWCRT have adopted a similar design [23-25].

In the actual simulation, three policies are considered: (i) no follow-up period and $d_{ij} = t_S + ((t_E - t_S) * e)/E$; (ii) there is a follow-up period and $d_{ij} = t_S + ((t_F - t_S) * e)/E$ (allow trial entry until the end of the follow-up period; illustrated in Fig. 2a); (iii) there is a follow-up period but $d_{ij} = t_S + ((t_E - t_S) * e)/E$ (terminate trial entry at the end of the last step period; illustrated in Fig. 2b).

In addition, E is a coefficient that specifies the timing of the trial entry. If $E = 1$, the subject enters the

229 trial randomly between t_S and t_E or t_F, which reflects the open cohort design in that the subject may enter
230 in the trial at any time. If E is greater than 1, it reflects a situation where the entry of the trial is concentrated
231 at an earlier stage of the trial (illustrated in Fig. 2c). In the actual example, 64.1% of the residents entered at
232 the start of the trial. Depending on the purpose and setting of the trial, other SWCRT show similar situations
233 [26, 27].

234 In the actual simulation, policies (i) to (iii) above regarding the follow-up period and the time of trial entry
235 can be taken for $E = 1$ and $E > 1$, respectively. Our study adopts only policy (iii) instead of (ii) at $E > 1$
236 (illustrated in Fig. 2d).

237 To compare our results with the secondary outcome of the actual example, No. of hospitalisations > 24 h
238 per facility-month, we decided to treat only hospitalizations > 24 h as a TTE in this study. It was previously
239 published [6] that the number of residents repeatedly hospitalised more than four times was very small.
240 Therefore, in our study, the maximum number of recurrent events generated in the simulation was three.

241 The relative performance of the statistical models used in TTRE, which are based on bias and variability,
242 depend on the event generation model used in the simulation, and it is thus recommended that simulations
243 based on multiple event generation models be considered [28]. Therefore, in this study, three types of event
244 generation model were used.

245 The first is the Poisson process, which generates TTEs based on exponential distributions independent of
246 each other, not only between subjects but also within subjects. The exponential distribution consists only of
247 scale parameters. The starting point of all TTEs is d_{ij} at the time of trial entry, and the hazard of a TTE is

248 always constant, regardless of the time and number of recurrences (illustrated in Fig. 3a).

249 The second is the Weibull model, where the starting point of the first TTE is d_{ij}, as in the Poisson process,

250 but the starting point of the second and subsequent TTEs is the time of the previous event (illustrated in Fig.

251 3b). Then, a Weibull distribution was assumed for the time between events within each subject. In addition

252 to a scale parameter similar to an exponential distribution, the Weibull distribution contains the shape

253 parameter. The Weibull distribution allows the hazard to vary with time depending on the setting of the shape

254 parameter. As this model adopts a Weibull distribution with a common parameter from the first to the third

255 TTE (i.e. the way the hazard changes are common from the first to the third TTE), we refer to it as the Weibull

256 model (constant).

257 The third model uses the same Weibull model as the second one, but adopts the Weibull distribution with

258 different parameters between the "first TTE" and the "second and third TTE" (i.e., the way the hazard changes

259 is different between the first and second and third TTEs), and so it is referred to as the Weibull model (change).

260 In a simple RCT situation where an intervention effect exists, previous studies with time-independent

261 covariates have shown that in the Poisson process, both AG and PWP-TT models perform well, while in the

262 Weibull model (constant), only the PWP-TT model performs well, and the AG model performs poorly [28].

263 To generate TTREs that can account for unidirectional switching, which is assumed to be

264 estimating intervention effects using the CoxPH model and several extended CoxPH models, we use a data

265 generation process for the CoxPH model with time-dependent covariates, based on the three event generation

266 models previously described [29]. If the generated TTRE exceeds t_E or t_F, it is treated as right-censored at

t_E or t_F.

In the generation of TTRE in the Poisson process, three pseudo-random numbers were generated independently from the uniform distribution $U(0,1)$ and sorted in increasing order, u_1, u_2, u_3 in turn $(u_k, k = 1,2,3)$. If the scale parameter of the exponential distribution is λ, the baseline hazard function is λ, which is always constant regardless of the time or number of recurrences. The kth TTRE of the jth subject in the ith cluster, when the starting point is not considered, is as follows:

$$T^*_{ijk} = \begin{cases} \frac{-\log(u_k)}{\lambda \exp(\beta' x + \tau_i)} & -\log(u_k) < \lambda \exp(\beta' x + \tau_i)\, w_{ij} \\ \frac{-\log(u_k) - \lambda \exp(\beta' x + \tau_i) w_{ij} + \lambda \exp(\beta' x + \beta_{tik} + \tau_i) w_{ij}}{\lambda \exp(\beta' x + \beta_{tik} + \tau_i)} & -\log(u_k) \geqslant \lambda \exp(\beta' x + \tau_i)\, w_{ij} \end{cases},$$

where τ_i is the random effect on the variations between clusters, $\tau_i \sim N(0, \sigma^2)$. As already mentioned, β_{tik} is the parameter of the intervention effect on the kth recurrence of the ith cluster, and w_{ij} is the distance to switch for each subject. For simplicity, we omitted the $\beta' x$ for the time-independent covariates in the simulation. The TTRE, which is used in the analysis considering the starting point, is represented by $T_{ijk} = d_{ij} + T^*_{ijk}$.

In the generation of TTRE in the Weibull model, three pseudorandom numbers were generated independently from the uniform distribution $U(0,1)$, u_1, u_2, u_3 in the order in which they are generated $(u_k, k = 1,2,3)$. Let the scale parameter of the Weibull distribution for each recurrence be λ_k, and the shape parameter be ν_k. The baseline hazard function is $\lambda_k \nu_k t^{\nu_k - 1}$ and it is allowed to vary with time. The kth

TTRE of the jth subject in the ith cluster, when the starting point is not considered, is as follows:

$$T_{ijk}^{*} = \begin{cases} \left(\frac{-\log(u_k)}{\lambda_k \exp(\beta' x+\tau_i)}\right)^{1/\nu_k} & -\log(u_k) < \\ \left(\frac{-\log(u_k)-\lambda_k \exp(\beta' x+\tau_i) w_{ij}{}^{\nu_k}+\lambda_k \exp(\beta_{tik}) \exp(\beta' x+\tau_i) w_{ij}{}^{\nu_k}}{\lambda_k \exp(\beta_{tik}) \exp(\beta' x+\tau_i)}\right)^{1/\nu_k} & -\log(u_k) \geqslant \lambda_k \exp(\beta' x+\tau_i) w_{ij}{}^{\nu_k} \end{cases}$$

.

τ_i, β_{tik}, w_{ij}, and $\beta' x$ were explained in the previous sentence. The TTRE that is actually used for the analysis considering the starting point is:

$$\begin{cases} T_{ijk} = d_{ij} + T_{ijk}^{*} & k = 1 \\ T_{ijk} = T_{ijk-1} + T_{ijk}^{*} & k = 2,3 \end{cases}.$$

The parameters are $\lambda_1 = \lambda_2 = \lambda_3, \nu_1 = \nu_2 = \nu_3$ for the Weibull model (constant), and $\lambda_1 \neq \lambda_2 = \lambda_3, \nu_1 \neq \nu_2 = \nu_3$ for the Weibull model (change).

In the actual example, 31.4% of the residents died during the trial period. Therefore, in our simulation, we considered the time-to-terminal-event (TTTE) as independent of the distance to switch and TTRE. If the generated TTTE does not exceed t_E or t_F and it is before the third TTRE, it is treated as mid-trial right-side censoring at the occurrence of the terminal event. The scale parameter of the Weibull distribution for the terminal event is λ_c, and the shape parameter is ν_c. Without considering the starting point, the TTTE of the jth subject in the ih cluster, C_{ij}^{*}, can be expressed using the probability density function as follows:

$$f(x) = \frac{\nu_c}{\lambda_c^{\nu_c}} x^{\nu_c - 1} \ exp\left\{-\left(\frac{x}{\lambda_c}\right)^{\nu_c}\right\}, x > 0.$$

The TTTE used in the actual analysis considering the starting point is expressed as $C_{ij} = d_{ij} + C_{ij}^{*}$.

Parameter settings

The scale parameter for the exponential distribution in the generation of the TTRE by the Poisson process was set to $\lambda = 0.003281$. This parameter was estimated based on the TTHA up to the third of the actual example, with all starting points set to zero.

The scale and shape parameters of the Weibull distribution in the generation of TTRE using the Weibull model (constant) were set to $\lambda_1 = \lambda_2 = \lambda_3 = 0.004703, \nu_1 = \nu_2 = \nu_3 = 1.1219$. These parameters were estimated based on the TTHA, up to the third of the actual example. The starting point of the second and subsequent TTHA was the time of the previous hospitalisation.

The scale and shape parameters of the Weibull distribution in the generation of the TTRE using the Weibull model (change) were set to $\lambda_1 = 0.003599$, $\lambda_2 = \lambda_3 = 0.009910$, $\nu_1 = 1.5122, \nu_2 = \nu_3 = 0.9108$. These parameters were estimated based on the "first TTHA" and the "second and third TTHA" of the actual example, respectively. The starting point of the second and subsequent TTHAs was the time of the occurrence of the previous hospitalisation.

The scale and shape parameters of the Weibull distribution in the generation of TTTE as mid-trial right-side censoring were set to $\lambda_c = 0.003674$ and $\nu_c = 1.7191$. These parameters were estimated based on the time to death in the actual example.

Two parameters were set for the true intervention effect. The first is $\beta_{tik} = \beta_t = -0.264$, which was calculated as $\ln(4.3/5.6)$ based on the secondary outcome of the actual example, No. of hospitalisations per facility month. The second is $\beta_{tik} = \beta_t = 0$, a setting used in previous studies on event generation

models: HR = 1, which indicates that there is no difference in the risk of event occurrence between the control and intervention conditions. In a simple RCT situation where there is no intervention effect, both the AG and PWP-TT models have been shown to perform well, regardless of the type of event generation model.

Simulation set-up

For all simulations, we fixed $t_S = 0$ at the beginning of the trial, $t_E = 360$ at the end of the last step period, and the total sample size per simulation (total number of subjects per trial) $N = 2000$. These settings were based on the fact that the actual example lasts for 12 months from the start of the trial to the end of the final step period; if one month is considered to be approximately 30 days, the trial period can be calculated as 12 × 30 = approximately 360 days, and the total number of subjects was 1700. Unless otherwise noted, the basic settings for each simulation scenario are as follows: the number of simulations is 1000, the event generation model consists of three types (Poisson process, Weibull model (constant), Weibull model (change)), the parameters of the true intervention effect are two ways $(-0.264, 0)$, and $s(= m) = 5, n_i = n = N/m = 400, W_d = (t_E - t_S) \ / \ (m + 1) = 60, \sigma^2 = 0, E = 1, F = 0$. The setting of $s = m = 5$ is in reference to the fact that the number of steps in the actual example is five (Fig. 1).

Each simulation scenario is listed below. Scenario II applied two policies for each statistical model: (i) with stratification by clusters and (ii) without stratification by clusters. In all scenarios, except for scenario II, only (i) was applied.

In Scenario I, the number of steps (clusters) varied as $s(= m) = 2, 4, 5, 8, 10, 20$ to investigate how the

341 performance of each statistical model changed as the number of steps (clusters) increased. As the number of
342 steps changes, it becomes $n = N/m = 1000, 500, 400, 250, 200, 100, W_d = 120, 72, 60, 40, 33, 17$. The
343 results based on $s(= m) = 5, n = 400, W_d = 60$ in this scenario were used as a reference throughout the
344 simulations in our study.

345 In Scenario II, we varied the variance with respect to the random effect τ_i, which represents the
346 variation among clusters, as $\sigma^2 = 0.25, 0.5, 1$, and investigated how the performance of each statistical
347 model changed as the variation between clusters increased.

348 In Scenario III, the follow-up period varied as follows, $F = 1, 2, 3, 4$ to investigate how the
349 performance of each statistical model changed as the follow-up period increased. The setting of F is based
350 on the follow-up period of 2.5 steps in the actual example (Fig. 1). In this scenario, the time point of the trial
351 entry point was $d_{ij} = t_S + ((t_F - t_S) * e)/E$, and the subject was allowed to enter until the end of the
352 follow-up period.

353 In Scenario IV, the follow-up period was changed to $F = 1, 2, 3, 4$ to investigate how the
354 performance of each statistical model changed as the follow-up period increased. In this scenario, the time
355 point of the trial entry point was $d_{ij} = t_S + ((t_E - t_S) * e)/E$, and entry was terminated at the end of the
356 final step period.

357 In Scenario V, we varied the timing of the trial entry as follows, $E = 1.5, 2, 4, 6$ to investigate how
358 the performance of each statistical model changed as trial entry was concentrated at an earlier stage of the
359 trial.

In Scenario VI, the time of trial entry varied as follows, $E = 1.5, 2, 4, 6$, and the follow-up period was changed to $F = 1, 2, 3, 4$, to investigate how the performance of each statistical model changed in a situation where trial entry was concentrated in an earlier stage of the trial, and there was a follow-up period. In this scenario, for convenience, we used $d_{ij} = t_S + ((t_E - t_S) * e)/E$ as the time point for trial entry.

Analysis of an actual example

The time-independent covariates employed in the model analysis for the primary outcome in the actual example (age, sex, medical power of attorney, health directive, advance care plan/statement of choices, primary diagnosis, age-adjusted Charlson comorbidity index, and fidelity) were used for adjustment, when analysing hospitalization > 24 h repeatedly occurred with the TTRE in the actual example using each statistical model.

Two policies were applied to each statistical model: (i) with stratification by clusters and (ii) without stratification by clusters. Fidelity is a per-cluster variable and was employed only with policy (ii), as it is not available for adjustment in (i). The unidirectional switch from the control condition to the intervention condition in each cluster was expressed using the intervention indicator as a time-dependent covariate.

In the usual TTRE analysis, continuous risk intervals were employed. However, in reality, they are not exposed to the risk of further hospitalisation during their hospital stay. Therefore, in this study, we adopted a discrete risk interval [30]. Thus, for example, if a resident was hospitalised, subsequent exposure to the risk of new hospitalisation would be from the day of discharge.

The results of the analysis were evaluated using HR and its 95% CI and p-value. In addition, parameter estimates and standard error (SE) were evaluated for the intervention effects.

Software and code

All statistical analyses, including simulations, were performed using SAS, version 9.4 (SAS Institute, Cary, NC, USA). The PROC PHREG of SAS was used to analyse the TTRE.

Results

Simulation

The results for Scenario I with $s\ (=m)=5,\ n=400,\ W_d=60$ are shown in Table 1.These results were used as a reference for all the other simulations assessed in this study, as the setting $s\ (=m)=5$ references the fact that the number of steps in the actual example is five (Fig. 1).

Table 1 Performance for the reference results throughout the simulations

True intervention effect (β_t)	Statistical model	Poisson process			Weibull model (constant)			Weibull model (change)		
		Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
-0.264	CoxPH	0.0021	0.0054	0.942	0.0016	0.0059	0.940	-0.0040	0.0035	0.947
	AG	0.0266	0.0040	0.933	0.0567	0.0067	0.858	0.1247	0.0168	0.162
	PWP-TT	0.0021	0.0040	0.940	0.0529	0.0065	0.864	0.0328	0.0033	0.896
	PWP-GT	0.0380	0.0054	0.899	0.0032	0.0037	0.949	-0.0004	0.0022	0.955

0	CoxPH	0.0024	0.0049	0.944	0.0010	0.0054	0.946	-0.0039	0.0032	0.946
	AG	0.0023	0.0030	0.955	0.0025	0.0032	0.948	0.0011	0.0011	0.991
	PWP-TT	0.0024	0.0036	0.939	0.0025	0.0034	0.948	-0.0005	0.0020	0.961
	PWP-GT	0.0021	0.0034	0.933	0.0019	0.0031	0.944	-0.0011	0.0020	0.950

Settings: $s(=m)=5, n=400, W_d=60, \sigma^2=0, E=1, F=0$.

From the reference results for $\beta_t=-0.264$, the MSE under the Poisson process was smaller for the AG and PWP-TT models, and slightly larger for the PWP-GT model; the CP performances of the AG and PWP-TT models were similar, but the bias was much smaller for the PWP-TT model. The PWP-GT model performed very well in both the Weibull model (constant) and Weibull model (change). Under the Weibull model (change), the performance of the AG model was found to be very poor. In reference to the results for $\beta_t=0$, the overall performance was higher than that of $\beta_t=-0.264$, and the MSE was smaller in all the extended CoxPH models for TTRE than in the CoxPH model for TTFE only. The AG model under the Weibull model (change) tended to overestimate CP. In all event generation models, the PWP-TT and PWP-GT models showed similar results.

The results for Scenario I when the parameter for the true intervention effect is $\beta_t=-0.264$ are shown in Table 2, and the results when $\beta_t=0$ are shown in Additional File (S.1). Regardless of the setting for β_t, the overall MSE increased slightly as the number of steps (clusters) increased, but this did not substantially impact on the performance comparison between the statistical models.

The results for Scenario II when the parameter for the true intervention effect is $\beta_t=-0.264$ are shown in Table 3, and the results when $\beta_t=0$ are shown in Additional File (S.2). Regardless of the setting of β_t,

411 the performance of policy (ii) without stratification by clusters decreased as inter-cluster variation increased.

412 At $\sigma^2 = 0.25$, the lowest variance in the setting, the decrease in performance was already apparent,

413 especially for CP, as the performance was very poor. The reference results where policy (i) with stratification

414 by clusters was performed in the absence of inter-cluster variation were similar to the results when (i) with

415 stratification by clusters was performed in this scenario where inter-cluster variation was present.

416

417 Table 2 Performance for scenario I with true intervention effect of $\beta_t = -0.264$

Statistical	Number of steps	Poisson process			Weibull model (constant)			Weibull model (change)		
model	(clusters	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
CoxPH	2	<-0.0001	0.0044	0.952	0.0002	0.0050	0.946	-0.0033	0.0027	0.957
	4	0.0022	0.0049	0.954	-0.0007	0.0056	0.942	-0.0045	0.0034	0.948
	5	0.0021	0.0054	0.942	0.0016	0.0059	0.940	-0.0040	0.0035	0.947
	8	0.0019	0.0056	0.957	0.0009	0.0066	0.936	-0.0031	0.0036	0.944
	10	0.0011	0.0056	0.953	0.0001	0.0065	0.940	-0.0026	0.0039	0.941
	20	0.0016	0.0060	0.944	0.0006	0.0070	0.933	-0.0030	0.0041	0.948
AG	2	0.0267	0.0034	0.937	0.0462	0.0052	0.870	0.1229	0.0162	0.100
	4	0.0259	0.0036	0.941	0.0546	0.0065	0.848	0.1221	0.0162	0.157
	5	0.0266	0.0040	0.933	0.0567	0.0067	0.858	0.1247	0.0168	0.162
	8	0.0255	0.0041	0.943	0.0563	0.0071	0.849	0.1240	0.0167	0.171
	10	0.0251	0.0042	0.936	0.0566	0.0072	0.842	0.1250	0.0171	0.191
	20	0.0259	0.0042	0.938	0.0553	0.0070	0.869	0.1237	0.0169	0.222
PWP-TT	2	0.0006	0.0032	0.959	0.0424	0.0050	0.882	0.0253	0.0026	0.897
	4	0.0013	0.0036	0.948	0.0509	0.0062	0.867	0.0299	0.0032	0.892
	5	0.0021	0.0040	0.940	0.0529	0.0065	0.864	0.0328	0.0033	0.896
	8	0.0011	0.0041	0.938	0.0522	0.0068	0.863	0.0330	0.0035	0.902
	10	0.0016	0.0042	0.943	0.0522	0.0069	0.860	0.0348	0.0038	0.888
	20	0.0020	0.0042	0.949	0.0496	0.0067	0.884	0.0334	0.0038	0.901
PWP-GT	2	0.0393	0.0047	0.868	0.0016	0.0032	0.944	-0.0014	0.0019	0.950
	4	0.0386	0.0051	0.894	0.0003	0.0037	0.945	-0.0019	0.0021	0.963
	5	0.0380	0.0054	0.899	0.0032	0.0037	0.949	-0.0004	0.0022	0.955
	8	0.0397	0.0056	0.896	0.0016	0.0043	0.934	-0.0014	0.0024	0.954
	10	0.0376	0.0057	0.897	0.0019	0.0042	0.945	0.0010	0.0025	0.956
	20	0.0363	0.0057	0.912	0.0004	0.0044	0.949	-0.0012	0.0026	0.949

418 CoxPH: Cox Proportional Hazard, AG: Andersen-Gill, PWP-TT: Prentice-Williams-Peterson Total-Time,

419 PWP-GT: Prentice-Williams-Peterson Gap-Time, MSE: Mean square error, CP: Coverage probability

420

421 The results for Scenario III, when the parameter for the true intervention effect is $\beta_t = -0.264$ are shown

422 in Table 4, and the results when $\beta_t = 0$ are shown in Additional File (S.3). When $\beta_t = -0.264$, the

423 performance of the AG and PWP-TT models under the Weibull model (constant) and the PWP-TT model

424 under the Weibull model (change) improved as the follow-up period increased, when the trial entry was

425 allowed until the end of the follow-up period. In particular, for CP, the performance was comparable to that

426 of the PWP-GT model under the respective event generation model.

427

428 Table 4 Performance for scenario III with true intervention effect of $\beta_t = -0.264$

Statistical		Poisson process			Weibull model (constant)			Weibull model (change)		
model	F	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
CoxPH	1	0.0012	0.0047	0.953	0.0005	0.0052	0.950	-0.0034	0.0033	0.952
	2	0.0012	0.0047	0.953	0.0005	0.0050	0.948	-0.0038	0.0032	0.951
	3	0.0017	0.0048	0.946	0.0010	0.0053	0.944	-0.0033	0.0034	0.944
	4	0.0010	0.0051	0.939	0.0011	0.0058	0.944	-0.0025	0.0035	0.952
AG	1	0.0307	0.0037	0.931	0.0453	0.0052	0.874	0.1227	0.0161	0.119
	2	0.0338	0.0038	0.936	0.0377	0.0046	0.916	0.1206	0.0156	0.130
	3	0.0358	0.0039	0.928	0.0320	0.0042	0.932	0.1203	0.0155	0.139
	4	0.0369	0.0041	0.929	0.0280	0.0042	0.938	0.1202	0.0155	0.133
PWP-TT	1	0.0015	0.0034	0.947	0.0406	0.0049	0.894	0.0235	0.0027	0.929
	2	0.0013	0.0034	0.947	0.0326	0.0044	0.918	0.0173	0.0024	0.939
	3	0.0013	0.0034	0.949	0.0264	0.0041	0.934	0.0148	0.0023	0.949
	4	0.0010	0.0036	0.952	0.0219	0.0042	0.935	0.0129	0.0024	0.953
PWP-GT	1	0.0408	0.0050	0.895	0.0028	0.0030	0.953	-0.0002	0.0020	0.951
	2	0.0435	0.0051	0.878	0.0011	0.0029	0.959	-0.0008	0.0020	0.944
	3	0.0447	0.0051	0.863	0.0016	0.0030	0.940	0.0001	0.0019	0.950
	4	0.0445	0.0052	0.873	0.0022	0.0030	0.937	0.0006	0.0020	0.948

429 CoxPH: Cox Proportional Hazard, AG: Andersen-Gill, PWP-TT: Prentice-Williams-Peterson Total-Time,

PWP-GT: Prentice-Williams-Peterson Gap-Time, MSE: Mean square error, CP: Coverage probability

The results for Scenario IV when the parameter for the true intervention effect was $\beta_t = -0.264$ are shown in Table 5, and the results when $\beta_t = 0$ are shown in Additional File (S.4). When $\beta_t = -0.264$, the performance of the AG and PWP-TT models under the Weibull model (constant) and the PWP-TT model under the Weibull model (change), improved as the follow-up period increased, given the policy of terminating trial entry at the end of the final step period. However, none of them reached the same level of performance as the PWP-GT model in their respective event generation models. In contrast, the performance of the PWP-GT model under the Poisson process tended to decrease as the follow-up period increased.

Table 5 Performance for scenario IV with true intervention effect of $\beta_t = -0.264$

Statistical		Poisson process			Weibull model (constant)			Weibull model (change)		
model	F	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
CoxPH	1	0.0018	0.0047	0.945	0.0009	0.0049	0.947	-0.0041	0.0031	0.954
	2	0.0013	0.0045	0.940	0.0001	0.0047	0.950	-0.0042	0.0031	0.953
	3	0.0014	0.0045	0.943	-0.0003	0.0047	0.948	-0.0042	0.0031	0.953
	4	0.0012	0.0045	0.943	-0.0004	0.0046	0.949	-0.0042	0.0031	0.953
AG	1	0.0320	0.0037	0.919	0.0475	0.0051	0.866	0.1282	0.0174	0.062
	2	0.0350	0.0036	0.911	0.0418	0.0044	0.886	0.1264	0.0169	0.065
	3	0.0361	0.0037	0.910	0.0391	0.0042	0.890	0.1249	0.0165	0.068
	4	0.0361	0.0037	0.909	0.0381	0.0041	0.899	0.1246	0.0165	0.069
PWP-TT	1	0.0020	0.0033	0.948	0.0429	0.0048	0.882	0.0264	0.0026	0.908
	2	0.0015	0.0031	0.944	0.0367	0.0042	0.907	0.0233	0.0024	0.922
	3	0.0012	0.0031	0.943	0.0339	0.0039	0.912	0.0222	0.0023	0.930
	4	0.0009	0.0031	0.944	0.0329	0.0039	0.913	0.0219	0.0023	0.930
PWP-GT	1	0.0429	0.0050	0.876	0.0028	0.0028	0.948	0.0001	0.0018	0.950
	2	0.0470	0.0052	0.840	0.0024	0.0025	0.942	-0.0002	0.0017	0.946
	3	0.0497	0.0054	0.825	0.0026	0.0023	0.945	-0.0004	0.0017	0.949
	4	0.0512	0.0055	0.810	0.0022	0.0023	0.947	-0.0004	0.0017	0.949

441 CoxPH: Cox Proportional Hazard, AG: Andersen-Gill, PWP-TT: Prentice-Williams-Peterson Total-Time,

442 PWP-GT: Prentice-Williams-Peterson Gap-Time, MSE: Mean square error, CP: Coverage probability

443

444 The results for Scenario V when the parameter for the true intervention effect is $\beta_t = -0.264$ are shown

445 in Table 6, and the results when $\beta_t = 0$ are shown in Additional File (S.5). Regardless of the setting of β_t,

446 there was a tendency for the overall MSE to increase as the trial entry was more concentrated at the beginning

447 of the trial. When $\beta_t = -0.264$, for the PWP-GT model under the Poisson process, the AG and PWP-TT

448 models under the Weibull model (constant), and the PWP-TT model under the Weibull model (change), CP

449 always performed poorly when compared to the reference results, regardless of the value for E.

450

451 Table 6 Performance for scenario V with true intervention effect of $\beta_t = -0.264$

Statistical		Poisson process			Weibull model (constant)			Weibull model (change)		
model	E	Bias	MSE	CP	Bias	MSE	CP	Bias	MSE	CP
CoxPH	1.5	0.0014	0.0060	0.950	0.0009	0.0059	0.951	-0.0026	0.0043	0.948
	2	<0.0001	0.0073	0.951	-0.0029	0.0080	0.945	-0.0015	0.0058	0.948
	4	0.0018	0.0148	0.947	-0.0079	0.0146	0.954	-0.0056	0.0111	0.951
	6	0.0028	0.0241	0.936	-0.0034	0.0215	0.951	-0.0057	0.0168	0.949
AG	1.5	0.0265	0.0040	0.937	0.0721	0.0084	0.768	0.1363	0.0201	0.122
	2	0.0232	0.0045	0.943	0.0876	0.0115	0.724	0.1391	0.0214	0.213
	4	0.0110	0.0088	0.944	0.1148	0.0203	0.713	0.1437	0.0251	0.521
	6	0.0066	0.0119	0.954	0.1274	0.0264	0.757	0.1447	0.0275	0.629
PWP-TT	1.5	0.0014	0.0041	0.937	0.0689	0.0080	0.793	0.0497	0.0050	0.834
	2	0.0015	0.0046	0.938	0.0855	0.0112	0.734	0.0656	0.0073	0.780
	4	-0.0012	0.0095	0.938	0.1139	0.0202	0.731	0.0961	0.0149	0.769
	6	-0.0021	0.0128	0.948	0.1267	0.0264	0.766	0.1061	0.0193	0.788
PWP-GT	1.5	0.0498	0.0064	0.848	0.0002	0.0034	0.957	-0.0005	0.0025	0.937
	2	0.0603	0.0084	0.842	0.0013	0.0042	0.945	0.0009	0.0031	0.954
	4	0.0786	0.0152	0.856	-0.0044	0.0080	0.944	-0.0016	0.0063	0.949
	6	0.0878	0.0212	0.870	-0.0019	0.0107	0.957	-0.0040	0.0089	0.952

CoxPH: Cox Proportional Hazard, AG: Andersen-Gill, PWP-TT: Prentice-Williams-Peterson Total-Time, PWP-GT: Prentice-Williams-Peterson Gap-Time, MSE: Mean square error, CP: Coverage probability

The results for Scenario VI, when the parameter for the true intervention effect is $\beta_t = -0.264$, are shown in Additional File (S.6), and the results when $\beta_t = 0$ are shown in Additional File (S.7), respectively. The results are similar to those of Scenario V, regardless of the setting of β_t or the value of F.

Actual example

The results summarising only the intervention indicators as time-dependent covariates are shown in Table 7. The overall results, including the time-independent covariates used for adjustment, are shown in Additional File S.8.

Table 7 Analysis results of actual example (intervention indicator only)

Dealing with clusters	Statistical model	Parameter Estimates	Standard Error	HR [95%CI]	p-value
With stratification by clusters	CoxPH	0.045	0.142	1.046 [0.792, 1.382]	0.751
	AG	0.033	0.122	1.034 [0.814, 1.314]	0.785
	PWP-TT	-0.061	0.123	0.941 [0.739, 1.198]	0.621
	PWP-GT	0.054	0.117	1.056 [0.840, 1.327]	0.641
Without stratification by clusters	CoxPH	0.07	0.111	1.073 [0.862, 1.335]	0.528
	AG	0.096	0.094	1.102 [0.915, 1.326]	0.306
	PWP-TT	0.040	0.095	1.041 [0.863, 1.254]	0.676
	PWP-GT	0.087	0.091	1.090 [0.912, 1.304]	0.360

CoxPH: Cox Proportional Hazard, AG: Andersen-Gill, PWP-TT: Prentice-Williams-Peterson Total-Time,

PWP-GT: Prentice-Williams-Peterson Gap-Time, HR: Hazard ratio, CI: Confidence interval

The HR for the intervention indicator shows the relative risk of the intervention condition when compared to the control. Except for the PWP-TT model under policy (i) with stratification by clusters, the overall HR was slightly above 1, suggesting that the risk of events in the intervention condition may be higher than in the control, although the difference was not statistically significant. Reviewing the results of the statistical model, under policy (ii) without stratification by clusters, the HR tended to be larger, and the range of the SE and 95% CI was smaller than under policy (i) with stratification by clusters. By comparing the results across the statistical models, it was found that all extended CoxPH models for TTRE tended to have smaller SEs and 95% CIs than the CoxPH models for TTFE.

The results of the covariates other than the intervention indicator, showed that the primary diagnosis of "dementia and Parkinson's disease", and the age-adjusted Charlson comorbidity index were statistically significant for all statistical models. Residents with dementia and Parkinson's disease had a lower risk of event occurrence than those without dementia and Parkinson's disease, suggesting that the risk of event occurrence may increase with the severity of comorbidities.

Discussion

In this study, we have conducted comparative simulations to identify the statistical model's whose performance for estimating intervention effects based on TTRE in SWCRT using an open cohort design were

485 superior and could effectively be applied to actual clinical trial data.

486 The results of the simulations show that the performance under policy (ii) without stratification by clusters
487 was worse when compared with policy (i) with stratification by clusters, in both the statistical models and
488 settings. As SWCRT is implemented at the cluster level, the existence of cluster effects should be considered
489 in any setting. Furthermore, even if there is no variation among the clusters, there is no difference in
490 performance with and without stratification by clusters, so (i) with stratification by clusters should always be
491 adopted in the estimation of intervention effects based on TTRE in SWCRT when using an open cohort
492 design.

493 The results of the simulations, in a situation where there is no follow-up period, and the timing of the trial
494 entry tends to be random, showed that Poisson processes were similar to those of previous studies in settings
495 that did not include time-dependent covariates [28]. The result that the performance of the PWP-TT model,
496 as well as the AG model, was degraded in the Weibull model (constant) is somewhat different from those of
497 previous studies. This is a tendency that is considered to be specific to SWCRT with an open cohort design.

498 In real-world SWCRT, there may be situations in which a follow-up period is established, or trial entry is
499 concentrated in the early period, due to the nature of the study objectives and target clusters. The simulation
500 results are important because they show that the performance of the statistical models against TTRE depends
501 not only on the true intervention effects and event generation model, but also on the trial design of SWCRT
502 (the presence of a follow-up period and the timing of trial entry).

503 The event generation model used in our study was hypothetical. The primary analysis methods in the

clinical trials usually need to be specified in advance in the study protocol or statistical analysis plan. If the policy is to adopt a statistical model for the primary analysis, and it needs to determine a statistical model in the early phase of trial planning, it would be desirable to adopt one that shows reasonable performance in various settings, rather than one that performs well only in a particular event generation model. In our study, through simulations based on various settings, the PWP-GT model with stratification by clusters showed the best performance in most settings and reasonable performance in other settings. Therefore, if the policy is to adopt a statistical model as the primary analysis, and this needs to be determined in the early phase of the trial planning, the PWP-GT model with stratification by clusters should be adopted.

Under the Weibull model (change), the overall performance of the AG model tended to be very low when intervention effects were present, and the CP of the AG model tended to be excessive when there were no intervention effects. Considering the possibility that the actual event generation model is a Weibull model (change), it is challenging to adopt the AG model during the early phase of trial planning.

The only situation in which the performance of the PWP-TT model with stratification by clusters is not inferior to that of the PWP-GT model with stratification by clusters is when there is a certain amount of follow-up period, and the timing of the trial entry tends to be random within the trial period, including the follow-up period. Therefore, in this situation, it may be acceptable to adopt the PWP-TT model with stratification by clusters during the early phase of the trial planning, instead of the PWP-GT model with stratification by clusters. In our study, the performance of the PWP-TT model was particularly good when the follow-up period was more than three steps. In addition, considering that the original trial period consisted

523 of six steps ($s = m = 5$), it may be possible to think of it as a rough guide that "a certain amount of follow-
524 up period" as "a follow-up period that is more than half the length of the original trial period". The results
525 presented in Additional File (S.9) indicate that it can be assumed that the same is true for different numbers
526 of steps (clusters). The choice of which statistical model to use depends on the nature of the intervention, the
527 characteristics of the subjects, and the clinical interpretability of the analysis results. In our study, for the sake
528 of comparability, we estimated only the overall effects based on the PWP model, assuming that each
529 recurrence had a common effect. However, in an actual analysis, it is possible to estimate event-specific
530 effects. The PWP-TT model is appropriate when one wants to know the effect of each recurrence since the
531 start of the subject's follow-up. On the other hand, the PWP-GT model is suitable for understanding the effect
532 of recurrence, in relation to the previous occurrence.

533 A previous study on the CoxPH model in the context of SWCRT showed a tendency for the MSE to
534 decrease as the number of steps (clusters) increased. However, the simulations in our study showed an
535 opposite trend. This difference is not apparent, but it is thought to be due to the differences in the various
536 settings during the simulation. For example, in the previous study, the true intervention effect was set to 1,
537 whereas in our study, it was set to −0.264 or 0.

538 There is a follow-up period in the actual example, and trial entry is concentrated early in the trial period.
539 Therefore, based on the results of the simulations, the PWP-GT model with stratification by clusters is likely
540 to be the most appropriate method for estimating intervention effects based on TTRE against the actual
541 example, if the policy is to adopt a statistical model as the primary analysis, and this needs to be determined

542 in the early phase of trial planning. However, considering that the parameter estimates are close to zero for
543 any of the statistical models, the PWP-TT model with stratification by clusters may also be adopted for
544 exploratory analysis in terms of performance. The number of hospitalizations per facility-month, was
545 evaluated as a secondary outcome in the actual example and showed an obvious decrease in the intervention
546 condition when compared to the control, which is a substantial deviation from the results from the TTRE
547 analysis of our study. One possible reason for this is that the analysis of the number of hospitalizations per
548 facility-month ignores that residents are exposed to both the control and intervention conditions. The purpose
549 of our study was to provide a different perspective to the existing evaluations. Therefore, it does not negate
550 the conclusions of the actual example, which have previously been published.

551 Our study had several limitations. First, all of the statistical models employed treat a terminal event before
552 the third TTRE as a mid-trial censoring event. However, if a death occurs, for instance, in actual example,
553 the possibility of a subsequent hospitalisation is lost. An event such as a death in such a situation is called a
554 competing risk [31], but in our study, we did not account for terminal events as competing risks.

555 Second, we assumed non-informative censoring for the terminal event, which was treated as mid-trial
556 censoring. This assumes that censoring occurs independently due to causes unrelated to the TTRE. However,
557 if, for example, repeated hospitalisations occur in an actual example, the risk of death is likely to increase. In
558 such situations, it is possible to use an approach that considers the terminal event as informative censoring
559 and corrects for it, but this was not applied [32, 33].

560 Third, the simulation in our study employed continuous risk intervals as it has been adopted in many

previous studies [19, 22, 34]. However, we believe that simulations for discontinuous risk intervals (adopted in the analysis of the data from actual example) should be considered in the future.

Fourth, for simulation simplicity, we assumed that the number of clusters moving from the control condition to the intervention condition in one step was one ($s = m$). However, in actual example, two or three care homes are included in one cluster that transitions in one step. If the intervention effects can be assumed to be common among multiple care homes within a cluster, this is not an issue. If they cannot, they should be considered in the analysis, but we were not able to do this in our study.

Conclusions

The PWP-GT model with stratification by clusters showed the most reasonable performance for estimating the intervention effects based on the TTRE in SWCRT in various settings using an open cohort design.

Declarations

Ethics approval and consent to participate

In the INSPIRED trial (actual example of our study), consent to run the trial was gained at the site, rather than individual resident, level given the impracticalities of gaining informed consent from a large population. This follows national guidelines for Australia from the National Health and Medical Research Council (NHMRC). Since our study is only an analysis based on simulations and pre-collected data, we did not obtain additional consent from the individual resident. For the INSPIRED trial group to provide us with the

electronic data (anonymized) collected in the INSPIRED trial, we obtained the approval of the Ethics Review Committee of the Tohoku University Graduate School of Medicine for the study protocol. (Reception No.: 2020-1-1180) Our study follows Ethical Guidelines for Medical and Biological Research Involving Human Subjects (Japanese).

Consent for publication

Not applicable.

Availability of data and materials

Simulation codes supporting the conclusions of this article are available from a GitHub repository at https://github.com/s-oyamada/TimeToRecurrentEventInSteppedWedge.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

SO, SC and TY participated in the design of the study. SO carried out the simulation study and the statistical analysis of an actual example data, and drafted the manuscript. SC and TY participated in a discussion about statistical aspects. All the authors read and approved the final manuscript.

Acknowledgements

We thank and acknowledge Liz Forbat (University of Stirling) and Wai-Man Liu (Australian National University) for providing electronic data (anonymized) collected in an actual example and advice on the content of our study. We would like to thank Editage (www.editage.com) for English language editing.

References

1. Eldridge S, Kerry S. A practical guide to cluster randomised trials in health services research. 1st ed. US: John Wiley & Sons Inc; 2012.

2. Meurer WJ, Lewis RJ. Cluster randomized trials: evaluating treatments applied to groups. JAMA. 2015;313(20):2068-2069.

3. Ellenberg SS. The stepped-wedge clinical trial: evaluation by rolling deployment. JAMA. 2018;319:607–08.

4. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis and reporting. BMJ. 2015; 350:h391.

5. Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials. 2015;16:352.

6. Forbat L, Liu WM, Koerner J, Lam L, Samara J, Chapman M, et al. Reducing time in acute hospitals: A stepped-wedge randomised control trial of a specialist palliative care intervention in residential care homes. Palliat Med. 2020;34:571-9.

7. Cook RJ, Lawless JF. The Statistical Analysis of Recurrent Events. 2nd ed. NY: Springer; 2010.

8. Cox DR. Regression Models and Life-Tables. J Royal Stat Soc B. 1972;34(2):187-220.

9. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat. 1982;10:1100-1120.

10. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. Biometrika. 1981;68:373-379.

11. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Stat Assoc. 1989;84:1065-1073.

12. Zhan Z, de Bock GH, van den Heuvel ER. Statistical methods for unidirectional switch designs: past, present, and future. Stat Meth Med Res. 2018;27:2872–2882.

13. Cox DR, Oakes D. Analysis of Survival Data, Monographs on statistics and applied probability. 1st ed. London: Chapman & Hall; 1990.

14. Fisher LD, Lin DY. Time-dependent covariates in the cox proportional-hazards regression model. Annu Rev Public Health. 1999;20:145–57.

15. Zhan Z, de Bock GH, Wiggers T, van den Heuvel E. The analysis of terminal endpoint events in stepped wedge designs. Stat Med. 2016;35:4413–4426.

16. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. 2nd. New York: Wiley; 2002.

17. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. Int J Epidemiol. 2015;44:324–33.

18. Lemeshow S, May S, Hosmer SW. Applied Survival Analysis: Regression Modeling of Time-to-Event Data. 2nd ed. US: John Wiley & Sons Inc; 2008.

19. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer-Verlag; 2010.

20. Lin DY, Wei DJ. The robust inference for the Cox proportional hazards model. J Am Stat Assoc. 1989;84:1074–8.

21. Therneau TM, Hamilton SA. RhDNase as an example of recurrent event analysis. Statistics in Medicine. 1997;16:2029–2047.

22. Kelly PJ, Lim L L-Y. Survival analysis for recurrent event data: an application to childhood infectious diseases. Stat Med. 2000;19:13–33.

23. Bouwsma EVA, Huirne JAF, van de Ven PM, Noordegraaf AV, Schaafsma FG, Koops SES, et al. Effectiveness of an internet-based perioperative care programme to enhance postoperative recovery in

655 gynaecological patients: Cluster controlled trial with randomised stepped-wedge implementation. BMJ
656 Open. 2018;8:e017781.
657 24. Kerber KA, Damschroder L, McLaughlin T, Brown DL, Burke JF, Telian SA, et al. Implementation of
658 evidence-based practice for benign paroxysmal positional vertigo in the emergency department: a stepped-
659 wedge randomized trial. Ann Emerg Med. 2020;75(4):459–470.
660 25. Freeman CR, Scott IA, Hemming K, Connelly LB, Kirkpatrick CM, Coombes I, et al. Reducing
661 Medical Admissions and Presentations Into Hospital through Optimising Medicines (REMAIN HOME): a
662 stepped wedge, cluster randomised controlled trial. Med J Aust. 2021;214(5):212-217.
663 26. Leontjevas R, Gerritsen DL, Smalbrugge M, Teerenstra S, Vernooij-Dassen MJ, Koopmans RT. A
664 structural multidisciplinary approach to depression management in nursing-home residents: a multicentre,
665 stepped-wedge cluster-randomised trial. Lancet. 2013;381:2255–64.
666 27. Halek M, Reuther S, Muller-Widmer R, Trutschel D, Holle D. Dealing with the behaviour of residents
667 with dementia that challenges: A stepped-wedge cluster randomized trial of two types of dementia-specific
668 case conferences in nursing homes (FallDem). Int J Nurs Stud. 2020;104:103435.
669 28. Metcalfe C, Thompson SG. The importance of varying the event generation process in simulation
670 studies of statistical methods for recurrent events. Stat Med. 2006;25:165-179.
671 29. Austin PC. Generating survival times to simulate Cox proportional hazards models with time-varying
672 covariates. Stat Med. 2012;3129):3946–3958.

673 30. Braga JR, Tu JV, Austin PC, Sutradhar Rinku, Ross HJ, Lee DS. Recurrent events analysis for
674 examination of hospitalizations in heart failure: insights from the Enhanced Feedback for Effective Cardiac
675 Treatment (EFFECT) trial. Eur Heart J Qual Care Clin Outcomes. 2018;4:18-26.
676 31. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of
677 Competing Risks. Circulation. 2016;133:601–609.
678 32. Wang MC, Qin J, Chiang CT. Analyzing recurrent event data with informative censoring. J Am Stat
679 Assoc. 2001;96(455):1057–1065.
680 33. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event.
681 Biometrics. 2004;60:747–756.
682 34. Twisk JWR, Smidt N, de Vente W. Applied analysis of recurrent events: a practical overview. J
683 Epidemiol Community Health. 2005;59:706-710.
684
685
686 **Figure Legends**
687 **Fig. 1 Schematic representation of the actual example.** White cells correspond to the periods during which
688 the residents received the standard-of-care (control condition), and grey cells correspond to periods during
689 which the residents received new interventions (intervention condition). Each cluster C_1 to C_5 contains two
690 or three facilities, and one cluster moves from the control condition to the intervention condition in Steps 1
691 to 5. The duration for one Step (between one Step and the next) is two months, and the time period T_0 to T_8

692 is also every two months. The duration between T_8 and $T_{8.5}$ is one month. The start of the trial is T_0, and

693 after T_6, which is the end of the last step period, there is a follow-up period for 5 months until $T_{8.5}$

694 (equivalent to 2.5 Steps).

695

696 **Fig. 2 Schematic diagram of the simulation considering the follow-up period and the timing of trial**

697 **entry.** White cells correspond to the control condition and grey cells to the intervention condition. Cross

698 marks show examples of the time points when the five subjects in each cluster entered the trial. F is a

699 coefficient that specifies the follow-up period that may be set after the end of the last step period. When $F =$

700 0, there is no follow-up period, and $t_F = t_E$. If $F = X(> 1)$, there is a follow-up period of X step after the

701 end of the last step period. E is a coefficient that specifies the timing of the trial entry. If $E = 1$, the subject

702 enters the trial randomly between t_S and t_E or t_F, which reflects the open cohort design in that the subject

703 may enter in the trial at any time. If E is greater than 1, it reflects a situation where the entry of the trial is

704 concentrated at an earlier stage of the trial. (a) Example of a case where $F = 3, E = 1$, and trial entry is

705 allowed until the follow-up period. (b) Example of setting $F = 3, E = 1$ and trial entry is terminated in the

706 final step period. (c) Example of setting $F = 0, E = 2$. (d) Example of setting $F = 3, E = 2$.

707

708 **Fig. 3 Visualization of the event generation models.** White cells correspond to the control condition and

709 grey cells to the intervention condition. Cross marks indicate when a subject enters the trial, filled black

710 circles indicate relapse, and filled white circles indicate censoring. (a) Example of a Poisson process: all three

time-to-events occur at the time of trial entry. (b) Example of the Weibull model: the first time-to-event occurs at the time of trial entry, and the second and subsequent time-to-events occur at the time of the previous event.

Table 3 Performance for scenario II with true intervention effect of $\beta_t = -0.264$

Dealing with clusters	Event generation model	Analysis method	σ^2	Bias	MSE	CP
Stratification by clusters	Poisson process	CoxPH	0.25	0.0003	0.0052	0.951
			0.5	0.0016	0.0053	0.953
			1	0.0013	0.0053	0.949
		AG	0.25	0.0250	0.0036	0.943
			0.5	0.0296	0.0040	0.933
			1	0.0358	0.0043	0.899
		PWP-TT	0.25	-0.0001	0.0037	0.957
			0.5	0.0016	0.0037	0.947
			1	0.0001	0.0036	0.951
		PWP-GT	0.25	0.0367	0.0051	0.898
			0.5	0.0387	0.0050	0.901
			1	0.0358	0.0047	0.892
	Weibull model (parameter constant)	CoxPH	0.25	0.0013	0.0058	0.952
			0.5	0.0004	0.0057	0.947
			1	-0.0012	0.0062	0.947
		AG	0.25	0.0572	0.0069	0.844
			0.5	0.0556	0.0067	0.831
			1	0.0510	0.0061	0.836
		PWP-TT	0.25	0.0531	0.0066	0.852
			0.5	0.0512	0.0064	0.857
			1	0.0461	0.0058	0.861
		PWP-GT	0.25	0.0041	0.0038	0.950

			0.5	0.0015	0.0033	0.959
			1	0.0008	0.0033	0.939
	Weibull model (parameter change)	CoxPH	0.25	-0.0031	0.0035	0.951
			0.5	-0.0034	0.0036	0.943
			1	-0.0027	0.0038	0.942
		AG	0.25	0.1237	0.0166	0.158
			0.5	0.1196	0.0157	0.197
			1	0.1087	0.0134	0.275
		PWP-TT	0.25	0.0333	0.0034	0.892
			0.5	0.0323	0.0034	0.889
			1	0.0300	0.0033	0.886
		PWP-GT	0.25	-0.0002	0.0022	0.952
			0.5	-0.0006	0.0022	0.965
			1	0.0005	0.0022	0.954
Non-stratification by clusters	Poisson process	CoxPH	0.25	0.0037	0.0183	0.605
			0.5	0.0135	0.0535	0.369
			1	0.0411	0.1415	0.229
		AG	0.25	0.0362	0.0139	0.560
			0.5	0.0451	0.0438	0.324
			1	0.0666	0.1344	0.172
		PWP-TT	0.25	0.0057	0.0179	0.521
			0.5	0.0228	0.0537	0.289
			1	0.0634	0.1262	0.180
		PWP-GT	0.25	0.0361	0.0200	0.481
			0.5	0.0552	0.0595	0.274
			1	0.1067	0.1530	0.161
	Weibull model (parameter constant)	CoxPH	0.25	0.0031	0.0196	0.595
			0.5	0.0109	0.0582	0.366
			1	0.0373	0.1528	0.238
		AG	0.25	0.0417	0.0195	0.483
			0.5	0.0486	0.0623	0.270
			1	0.0720	0.1760	0.159
		PWP-TT	0.25	0.0405	0.0194	0.482
			0.5	0.0558	0.0567	0.286
			1	0.0950	0.1295	0.182
		PWP-GT	0.25	0.0085	0.0195	0.466
			0.5	0.0264	0.0606	0.260

		1	0.0828	0.1525	0.168
Weibull model (parameter change)	CoxPH	0.25	0.0027	0.0154	0.555
		0.5	0.0156	0.0448	0.338
		1	0.0458	0.1129	0.217
	AG	0.25	0.1227	0.0203	0.254
		0.5	0.1216	0.0338	0.276
		1	0.1227	0.0814	0.186
	PWP-TT	0.25	0.0281	0.0156	0.438
		0.5	0.0445	0.0456	0.258
		1	0.0835	0.1011	0.179
	PWP-GT	0.25	0.0068	0.0162	0.429
		0.5	0.0274	0.0506	0.255
		1	0.0845	0.1259	0.161

717 CoxPH: Cox Proportional Hazard, AG: Andersen-Gill, PWP-TT: Prentice-Williams-Peterson Total-Time,

718 PWP-GT: Prentice-Williams-Peterson Gap-Time, MSE: Mean square error, CP: Coverage probability

719

720

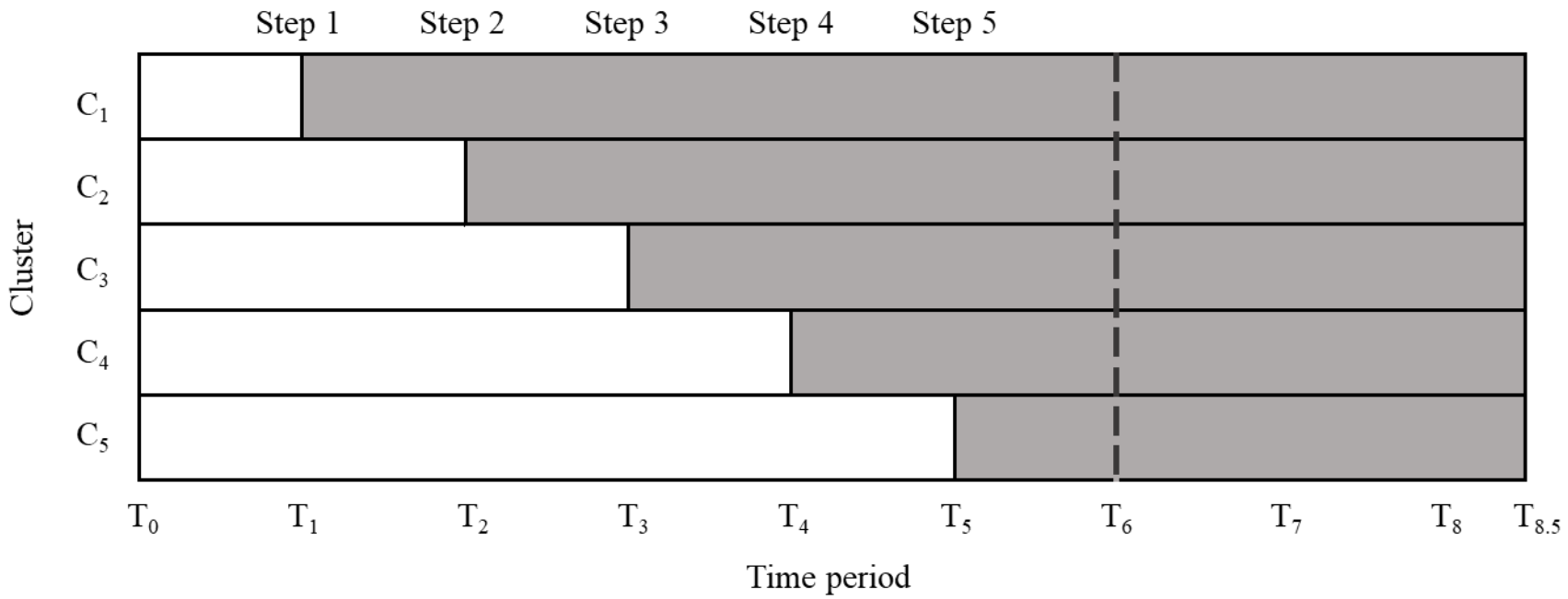

Fig. 1 Schematic representation of the actual example

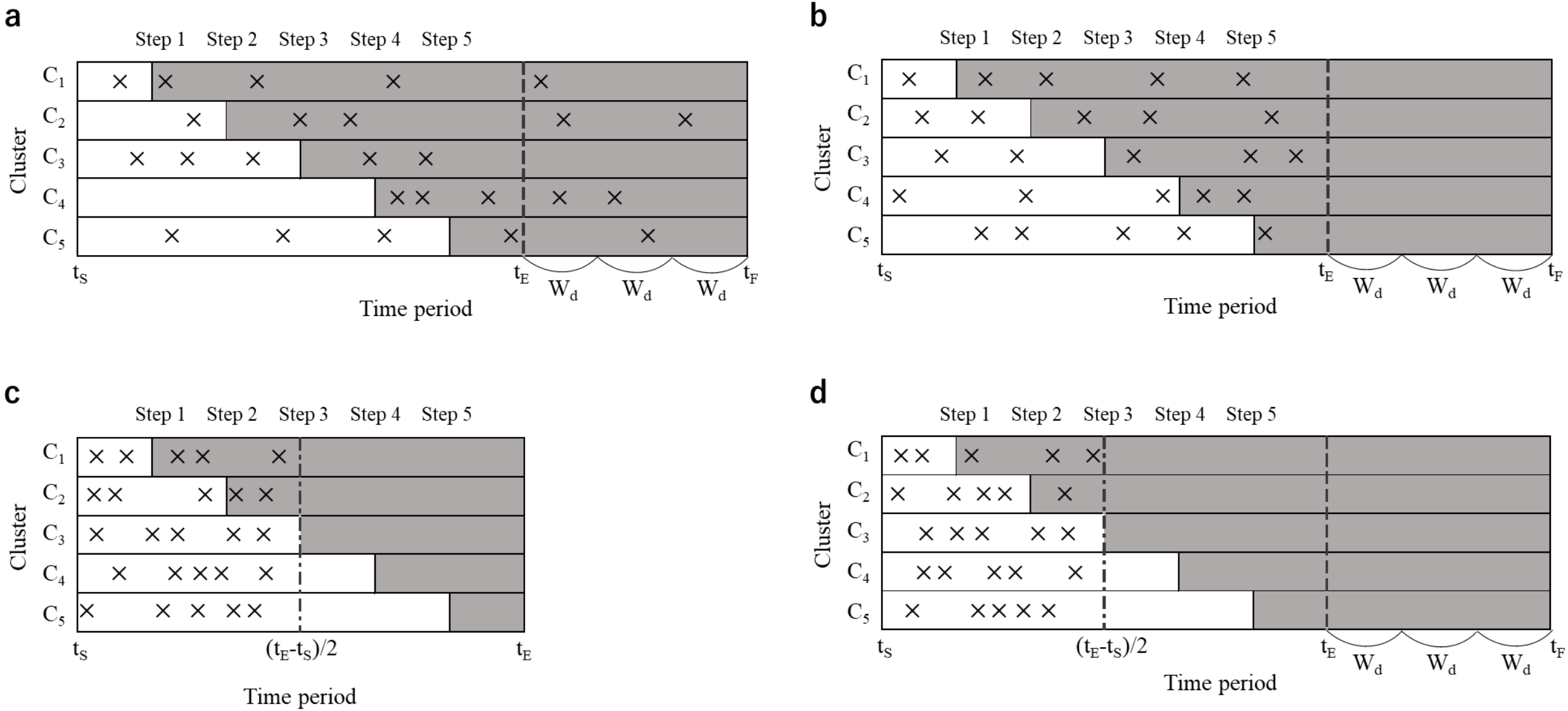

Fig. 2 Schematic diagram of the simulation considering the follow-up period and the timing of trial entry

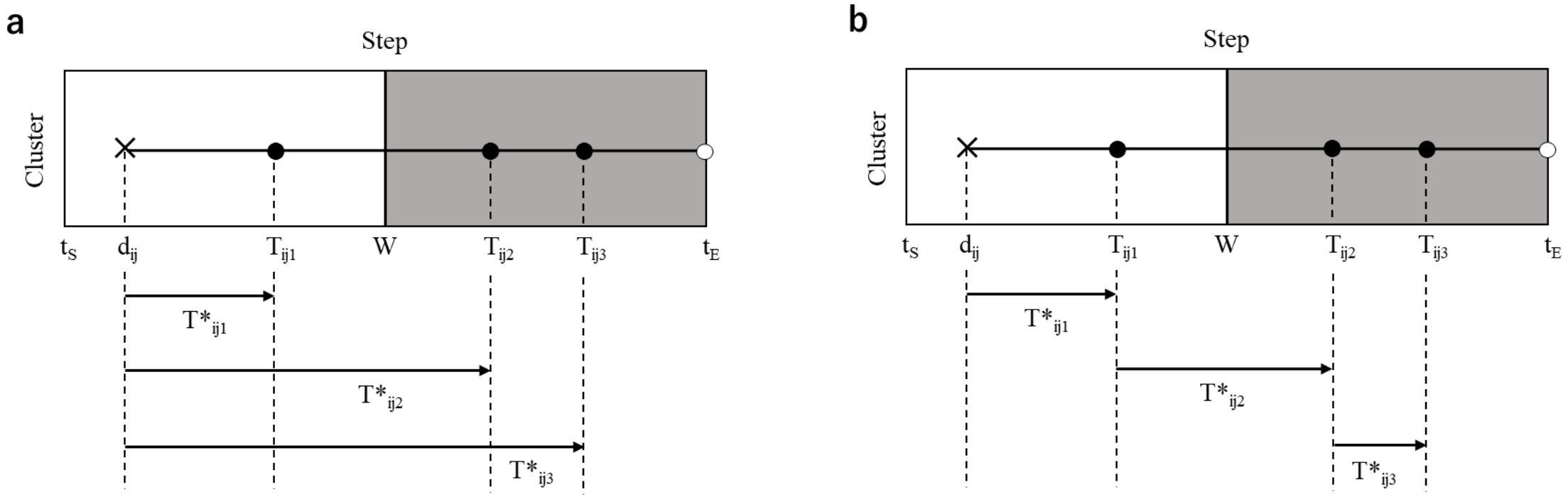

Fig. 3 Visualization of the event generation models

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile.xlsx