Author/s:
Degeling, K;IJzerman, MJ;Groothuis-Oudshoorn, CGM;Franken, MD;Koopman, M;Clements, MS;Koffijberg, H

Title:
Comparing Modeling Approaches for Discrete Event Simulations With Competing Risks Based on Censored Individual Patient Data: A Simulation Study and Illustration in Colorectal Cancer

Date:
2022-01-01

Citation:
Degeling, K., IJzerman, M. J., Groothuis-Oudshoorn, C. G. M., Franken, M. D., Koopman, M., Clements, M. S. & Koffijberg, H. (2022). Comparing Modeling Approaches for Discrete Event Simulations With Competing Risks Based on Censored Individual Patient Data: A Simulation Study and Illustration in Colorectal Cancer. Value in Health, 25 (1), pp.104-115. https://doi.org/10.1016/j.jval.2021.07.016.

Persistent Link:
https://hdl.handle.net/11343/289776

ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Methodology

Comparing Modeling Approaches for Discrete Event Simulations With Competing Risks Based on Censored Individual Patient Data: A Simulation Study and Illustration in Colorectal Cancer

Koen Degeling, PhD, Maarten J. IJzerman, PhD, Catharina G.M. Groothuis-Oudshoorn, PhD, Mira D. Franken, MD, Miriam Koopman, MD, PhD, Mark S. Clements, PhD, Hendrik Koffijberg, PhD

ABSTRACT

Objectives: This study aimed to provide detailed guidance on modeling approaches for implementing competing events in discrete event simulations based on censored individual patient data (IPD).

Methods: The event-specific distributions (ESDs) approach sampled times from event-specific time-to-event distributions and simulated the first event to occur. The unimodal distribution and regression approach sampled a time from a combined unimodal time-to-event distribution, representing all events, and used a (multinomial) logistic regression model to select the event to be simulated. A simulation study assessed performance in terms of relative absolute event incidence difference and relative entropy of time-to-event distributions for different types and levels of right censoring, numbers of events, distribution overlap, and sample sizes. Differences in cost-effectiveness estimates were illustrated in a colorectal cancer case study.

Results: Increased levels of censoring negatively affected the modeling approaches' performance. A lower number of competing events and higher overlap of distributions improved performance. When IPD were censored at random times, ESD performed best. When censoring occurred owing to a maximum follow-up time for 2 events, ESD performed better for a low level of censoring (ie, 10%). For 3 or 4 competing events, ESD better represented the probabilities of events, whereas unimodal distribution and regression better represented the time to events. Differences in cost-effectiveness estimates, both compared with no censoring and between approaches, increased with increasing censoring levels.

Conclusions: Modelers should be aware of the different modeling approaches available and that selection between approaches may be informed by data characteristics. Performing and reporting extensive validation efforts remains essential to ensure IPD are appropriately represented.

Keywords: censoring, competing events, competing risks, discrete event simulation, modeling, survival analysis

VALUE HEALTH. 2021; ■(■):■–■

Introduction

Decision analytic modelers increasingly use alternatives to the commonly used cohort-level state-transition modeling (STM) technique to reflect the complex dynamics of clinical pathways.[1-3] Modeling techniques, such as discrete-time agent-based modeling and continuous-time discrete event simulation (DES), are able to incorporate patient-level characteristics and clinical histories, multiple timescales, competing resources, and interactions among different actors, such as physicians and patients. Nevertheless, these techniques are more demanding than cohort-level STM, mainly in terms of computational complexity and required analytical skills to implement them.[4,5] Therefore, evidence-based methodological guidance is essential to inform the design choices that need to be made and reported on when applying these techniques to support the development of high-quality models.

DES is a useful alternative to STM and can be used to model personalized treatment processes because of its ability to reflect dynamic pathways based on patient-level histories and characteristics.[6,7] Although aggregated data can be used to develop DES models, individual patient data (IPD) are the preferred source of evidence to account for stochastic uncertainty and patient heterogeneity. An important design choice in developing DES models based on IPD is how competing events are implemented.[8] Competing events are those that prevent other events of interest from occurring or change the probability of their occurrence.[9] Four strategies have been identified to implement competing events in DES models: (1) sampling times for all competing events and selecting the event that is the first to occur, (2) sampling the event

to occur first and sampling the corresponding time-to-event second, (3) sampling the time-to-event first and sampling the corresponding event second, and (4) using discretized cyclic probabilities to resemble discrete-time STM.[10]

Modeling approaches corresponding to these strategies have been proposed and compared when informed by "uncensored" IPD generated according to mixtures of event-specific distributions (ESDs).[11] A general recommendation was made to sample the time to event based on a multimodal time-to-event distribution and then to determine the corresponding event second. An approach that first selects the event to occur and then the corresponding time to event also showed good performance with an easier implementation. Unfortunately, this guidance cannot be generalized to DES models informed by "censored" IPD. For "uncensored" IPD, it is known which competing event occurred for each patient, allowing the data to be analyzed conditional on which competing event occurred. For "censored" IPD, there are patients for whom it is unknown which event would eventually occur and extrapolation is required. Methods for extrapolation differ between modeling approaches, and hence, performance of the approaches may differ between uncensored and censored data. It is important to appropriately account for censoring in analyses involving competing events, because neglecting to do so may affect model outcomes.[12] Hence, there is a need for guidance on how to implement competing events in DES models informed by censored IPD.

This study aims to describe, illustrate, and compare modeling approaches for implementing competing events in DES models informed by censored IPD. Modeling approaches are compared in a simulation study, and potential differences in terms of health economic outcomes are analyzed for a case study in colorectal cancer.

Methods

Censoring can be classified as informative or noninformative and according to whether individuals are interval, left, or right censored.[13] Here, focus is on noninformative, right-censored data, because this type of censoring is most common in a clinical and health economic context. Two modeling approaches for implementing competing events in DES models informed by censored IPD were defined: (1) using event-specific time-to-event distributions (ESD) and (2) using a unimodal time-to-event distribution and regression (UDR) model. The approaches are defined using statistical notation in Supplemental Material 1 found at https://doi.org/10.1016/j.jval.2021.07.016. The code for implementation in R[14] is provided online: https://personex.nl/research/competing-risks/. Modeling approaches using event-specific probabilities and distributions (ESPD) or a multimodal distribution and regression model (MDR), which have been recommended for uncensored IPD,[11] are not presented here because their implementation was considered too cumbersome for censored IPD (see Discussion for a more detailed discussion).

Event-Specific Time-to-Event Distributions

The ESD modeling approach uses event-specific time-to-event distributions to sample a time-to-event for each competing event and subsequently selects the first event to occur to be simulated. See Box 1 for a pseudoalgorithm.

The ESD are estimated according to a cause-specific hazards model, which assumes that the risk set only includes individuals who are event free at the respective point in time.[15] When fitting the ESDs, both censored individuals and individuals who experienced a competing event are considered to be censored. This is a strong assumption that was believed to negatively affect the performance of the ESD approach for uncensored IPD,[11] given that it is unlikely that censoring that arises from an occurring competing event would be noninformative.

In a simulation according to the ESD approach, a time for each competing event needs to be sampled from each corresponding time-to-event distribution. Subsequently, the event that is the first to occur, that is, the event corresponding to the lowest sampled time to event, is selected and will be simulated.

Unimodal Joint Time-to-Event Distribution and Regression Model

The UDR modeling approach samples the time at which an event will occur using a single time-to-event distribution that represents the minimum of all competing time to events, that is, the time-to-event for the event that is observed. A regression model is subsequently used to predict to which event the selected time corresponds. See Box 2 for a pseudoalgorithm.

Here, we assume that the single time-to-event distribution is unimodal. In theory, this distribution could also be multimodal, such as a mixture or flexible parametric distribution, but estimation of such distributions may prove challenging for censored IPD (see Discussion). Because the time-to-event distribution represents all competing events, only 1 hazard function needs to be modeled. In doing so, all competing events are treated as 1 event, and consequently, an individual is either censored or experienced any type of event that is not censored.

The event corresponding to a time to event is modeled using a (multinomial) logistic regression model. A logistic regression model can be used to model binary data, that is, 2 competing events, whereas a multinomial logistic regression model is required to model more than 2 competing events. The time to event is included as a predictor variable in this regression, which may involve the use of transformations or splines to accurately model the relation between the time to event and type of event (response variable). Transformations of the time to event were not considered in the current simulation or case study, because this process is hard to automate (see Simulation Study).

A simulation according to the UDR approach is performed by first sampling a time from the time-to-event distribution and subsequently an event using the (multimodal) logistic regression model conditional on the sampled time.

Simulation Study to Compare the Performance of the Approaches

To compare the modeling approaches' performance for different data and disease pathway characteristics, different hypothetical datasets were simulated by first sampling which event would occur based on event-specific probabilities and then conditionally on the sampled event, sampling a time from the corresponding event-specific time-to-event distribution (see Supplemental Material 2 found at https://doi.org/10.1016/j.jval.2021.07.016 for the ESPD parameters). The conceptual model structure used for this simulation study is presented in Appendix Figure 1 (see Supplemental Material 3 found at https://doi.org/10.1016/j.jval.2021.07.016). As illustrated in Figure 1, simulations were performed for scenarios including different numbers of competing events n_{event} ($n_{event} = 2, 3, 4$), level of overlap of the corresponding competing time-to-event distributions $p_{overlap}$ ($p_{overlap} \approx 10\%, 50\%, 90\%$, ie low, medium, or high; see Supplemental Material 4 found at https://doi.org/10.1016/j.jval.2021.07.016), sample sizes n_{sample} ($n_{sample} = 50, 150, 500$), and levels of censoring $p_{censoring}$ ($p_{censoring} \approx 0\%, 10\%, 30\%, 60\%$). Two types of noninformative right censoring were considered: random

BOX 1. Pseudoalgorithm for the ESD approach to model the time to event *T* for *k* mutually exclusive competing events.

Data Analysis:
- For each competing event j, $j = 1, \ldots, k$, fit a time-to-event distribution D_j:
  - For each individual i, $i = 1, \ldots, n$, define observed censoring indicator c_{ij} that indicates whether individual i experienced event j ($c_{ij} = 1$) or not ($c_{ij} = 0$).
  - Parameterize cause-specific likelihood function $L_j(t_1, \ldots, t_n|\varphi_j)$ according to a specific distribution type.
  - Estimate parameters φ_j that define distribution D_j by maximum likelihood.

Simulation:
- Obtain time to events for each competing event:
  - Draw a time t_j for each competing event j, $j = 1, \ldots, k$ by performing a random draw from the corresponding distribution D_j.
- Select the competing event to occur:
  - Select event j with the lowest time to event (ie, the first event to occur).
- Simulate the selected event j at the corresponding time t_j.

Note. ESD indicates event-specific distribution.

censoring and follow-up censoring. In the "random censoring" approach, individuals were censored at a random point before their event would have happened. In the follow-up "censoring" approach, individuals were censored if their time to event exceeded a certain threshold, representing the scenario in which there is a maximum follow-up time per individual. For the follow-up censoring approach, $p_{censoring} \approx 60\%$ could not be applied, because that would censor all observations of some events. Similarly, $p_{censoring} \approx 30\%$ could not be applied for $p_{overlap} \approx 90\%$ because of convergence issues.

For each unique combination of the censoring approach, n_{event}, $p_{overlap}$, n_{sample}, and $p_{censoring}$, a total of 10 000 simulation runs were performed. In each simulation run, an uncensored sample $s_{uncensored}$ of corresponding sample size n_{sample} was sampled based on ESPD parameters. This sample $s_{uncensored}$ was right censored to censoring level $p_{censoring}$ according to the censoring approach to obtain censored sample $s_{censored}$. Next, both modeling approaches were applied to analyze sample $s_{censored}$ and, subsequently, to simulate event incidences and time to events for 100 000 new individuals to obtain a simulation sample $s_{simulation}$ for each modeling approach. Finally, the modeling approaches' performance was assessed by comparing event incidences and time-to-event distributions in these newly simulated samples $s_{simulation}$ with those in the corresponding uncensored sample $s_{uncensored}$.

Bias in terms of relative absolute incidence difference (RAID) was used as performance measure for the event incidence:

$$RAID = \frac{|Incidence_{simulated} - Incidence_{observed}|}{Incidence_{observed}} \times 100\% \quad (1)$$

The bias with regard to simulated time-to-event distributions was based on the relative entropy, that is, the Kullback–Leibler divergence (KLD), which is a measure of the difference between probability distribution $f(t)$ and $g(t)$, for which lower values indicate a better performance[16]:

$$KLD(f|g) = \int_0^\infty f(t) \log \frac{f(t)}{g(t)} dt = \int_0^\infty f(t) \times (\log(f(t)) - \log(g(t))) dt \quad (2)$$

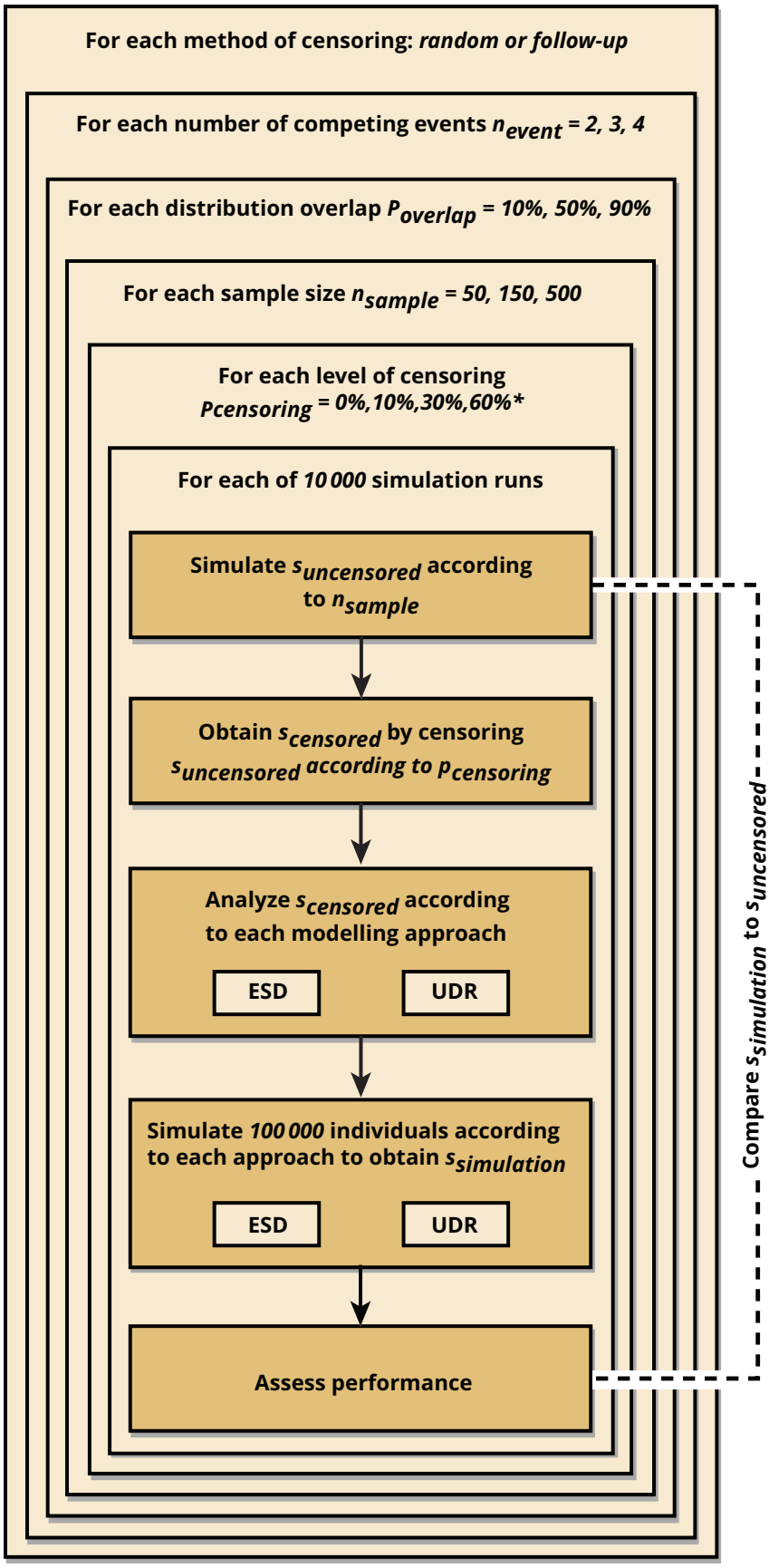
BOX 2. Pseudoalgorithm for the UDR approach to model the time to event *T* for *k* mutually exclusive competing events.

Data Analysis:
- For all competing events combined, fit a single unimodal time-to-event distribution D:
  - For each individual i, $i = 1, \ldots, n$, define observed censoring indicator c_i so that individual i experienced any event ($c_i = 1$) or is censored ($c_i = 0$).
  - Parameterize likelihood function $L(t_1, \ldots, t_n|\varphi)$ according to a specific unimodal distribution type.
  - Estimate parameters φ that define distribution D by maximum likelihood.
- Fit a (multinomial) logistic regression model to predict the competing event to occur:
  - Fit (multinomial) logistic regression model $r(t)$ that predicts probabilities $P(C_j = 1|T = t)$ of each competing event j, $j = 1, \ldots, k$ to occur (dependent variable) based on time t (independent variable).

Simulation:
- Obtain a time to event for an event to occur:
  - Draw a time t for the event to occur by performing a random draw from unimodal distribution D.
- Select the competing event to occur:
  - Obtain probabilities $P(C_j = 1|T = t)$ for each competing event j, $j = 1, \ldots, k$ to occur based on time to event t, using (multinomial) logistic regression model $r(t)$.
  - Draw a random number from a Uniform distribution $U[0, 1]$.
  - Select event j to occur using event probabilities $P(C_j = 1|T = t)$ and the random number.
- Simulate selected event j at time t.

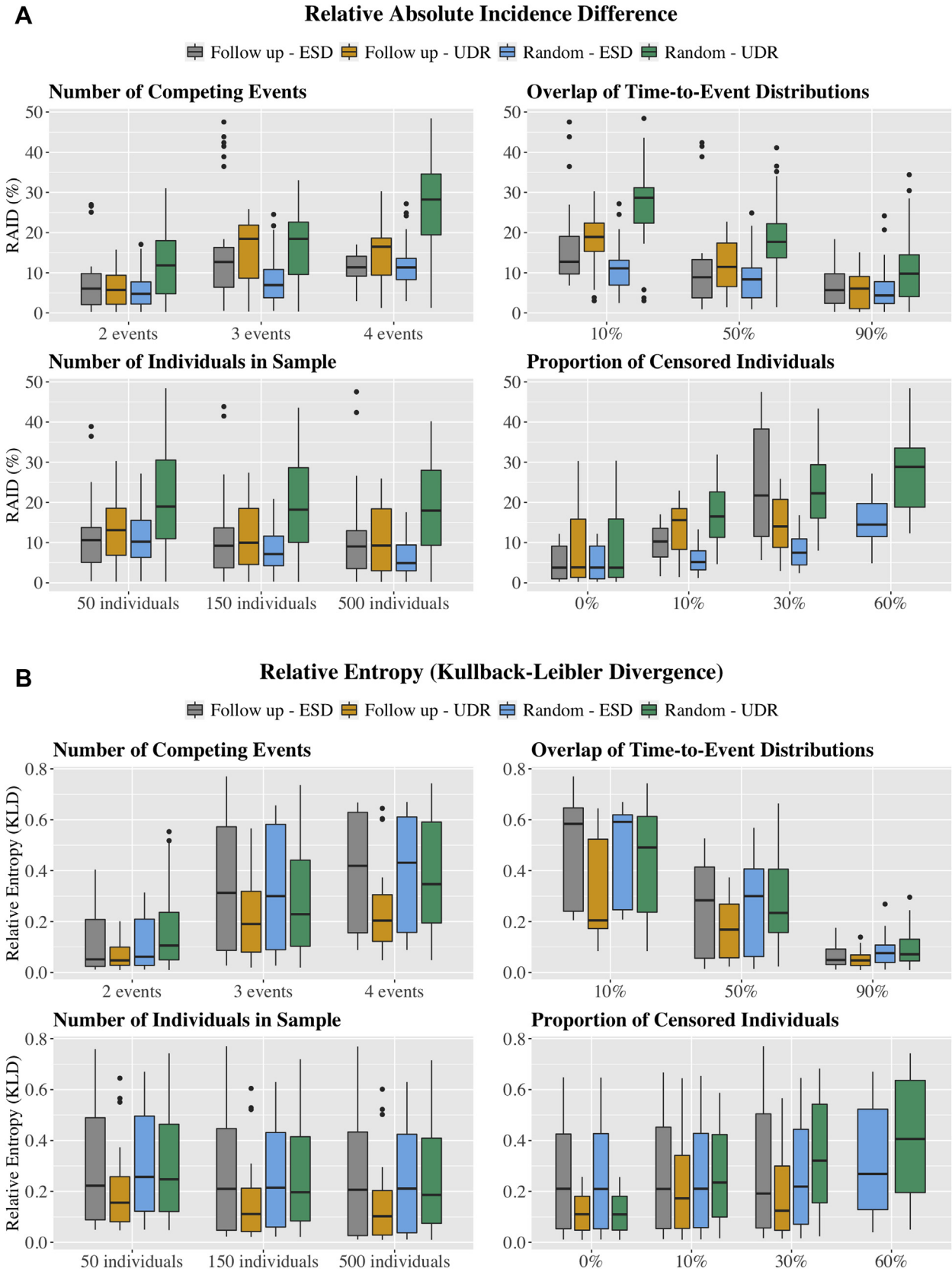
Note. UDR indicates unimodal distribution and regression.

Figure 1. Overview of the simulation study. *Not all levels of censoring were applied for censoring occurring because of a maximum follow-up time (see Methods).



ESD indicates event-specific distribution; UDR, unimodal distribution and regression.

Figure 2. Boxplots summarizing the bias of the modeling approaches (lower is better) for the different censoring mechanisms (random vs follow-up) in the simulation study with regard to (A) the RAID and (B) the relative entropy in terms of the KLD. For each boxplot, the data points are the mean outcomes of the 10 000 simulation runs for the applicable scenarios. For example, in the upper left panel (number of competing events), the left, gray boxplot is generated based on the bias of ESD under follow-up censoring in all scenarios in which the number of events was 2.



ESD indicates event-specific distribution; KLD, Kullback–Leibler divergence; RAID, relative absolute incidence difference; UDR, unimodal distribution and regression.

Table 1. Mean weighted bias (95% CI) of the modeling approaches for selected scenarios in the simulation study (lower is better).

Number of events	Distribution overlap (%)	Censored proportion (%)	Sample size	Relative absolute incidence difference (%)	
				Random censoring	
				ESD	UDR
2	10	10	50	6.3 (0.3, 16.3)	17.8 (5.7, 28.3)
2	10	10	500	5.0 (1.9, 8.3)*	17.3 (13.7, 20.7)
2	10	30	50	7.9 (0.3, 22.1)*	28.9 (11.7, 46.9)
2	10	30	500	2.5 (0.1, 7.0)*	28.4 (22.9, 33.8)
2	10	60	50	17.1 (0.7, 44.9)	31.1 (6.6, 58.5)
2	10	60	500	12.2 (1.7, 23.0)*	30.4 (22.9, 38.4)
2	90	10	50	3.9 (0.2, 10.0)	5.8 (0.3, 14.9)
2	90	10	500	1.2 (0.0, 3.4)*	4.6 (1.4, 7.8)
2	90	30	50	7.5 (0.3, 20.8)	9.8 (0.4, 25.8)
2	90	30	500	2.4 (0.1, 6.7)*	8.0 (2.6, 13.3)
2	90	60	50	14.5 (0.5, 40.0)	14.6 (0.6, 38.2)
2	90	60	500	4.8 (0.2, 13.3)*	12.3 (4.8, 20.0)
4	10	10	50	13.3 (7.9, 19.9)	31.9 (13.2, 51.4)
4	10	10	500	11.0 (9.4, 12.7)*	27.9 (21.7, 34.4)
4	10	30	50	16.8 (5.9, 30.6)*	43.3 (22.6, 73.2)
4	10	30	500	11.2 (7.0, 16.4)*	36.6 (29.5, 44.7)
4	10	60	50	27.2 (7.4, 52.8)	48.4 (25.3, 84.3)
4	10	60	500	17.6 (8.1, 27.1)*	40.2 (32.8, 48.8)
4	90	10	50	7.6 (2.1, 14.9)	16.5 (7.3, 27.5)
4	90	10	500	3.3 (1.4, 5.6)*	13.3 (10.4, 16.5)
4	90	30	50	13.2 (3.7, 25.6)	25.5 (11.0, 44.6)
4	90	30	500	4.4 (1.3, 8.5)*	20.4 (15.7, 25.4)
4	90	60	50	24.2 (6.6, 47.6)	34.4 (15.0, 62.4)
4	90	60	500	8.2 (2.4, 16.6)*	28.5 (21.6, 36.3)

CI indicates confidence interval; ESD, event-specific distribution; UDR, unimodal distribution and regression.
*Results for the modeling approach that performed best with regard to that respective outcome (ie, lower bias) if the corresponding mean bias was outside the CI of the other approach.

continued on next page

Both RAID and KLD were calculated separately for each competing event. To summarize the RAID and KLD across competing events per scenario, event-specific performance outcomes were weighted according to theoretical event incidences (see Supplemental Material 2 found at https://doi.org/10.1016/j.jval.2021.07.016).

The simulation study was performed in R version 3.4.1.[14] Time-to-event data were simulated and analyzed using Weibull distributions to rule out potential bias because of mismatching distributions. Weibull distributions were selected because these distributions are commonly used in survival analysis and showed to accurately represent the distribution of the time-to-event data in the case study (see Case Study). The nnet package was used to estimate (multinomial) logistic regression models[17] and the flexmix package to calculate KLD.[18-20]

Case Study to Illustrate the Potential Impact on Health Economic Outcomes

A case study based on data from the randomized phase III CAIRO3 study of the Dutch Colorectal Cancer Group (NCT00442637) was performed to illustrate the potential impact of the different modeling approaches on health economic outcomes in a real-world scenario. The CAIRO3 study randomized 558 metastatic colorectal cancer patients with stable disease or better after 6 cycles of capecitabine, oxaliplatin, and bevacizumab induction therapy to either capecitabine and bevacizumab maintenance treatment (intervention) or observation (control) until progression of disease.[21] For both the maintenance and observation strategy, capecitabine, oxaliplatin, and bevacizumab treatment was to be reintroduced upon progression and continued until second progression, which was the primary endpoint of the study. For the case study, we adapted a DES that was previously developed for comparison with a discrete-time cohort STM in a health economic evaluation of the CAIRO3,[5] to allow for simulations according to the 2 modeling approaches. The DES model was structured according to the treatment stages used in the CAIRO3 study: postinduction, reintroduction, salvage, and death (see Appendix Fig. 1 in Supplemental Material 3 found at https://doi.org/10.1016/j.jval.2021.07.016).

In addition to an analysis based on the complete CAIRO3 patient cohort, clinically relevant subgroup analyses were performed to

Table 1. Continued

Relative absolute incidence difference (%)		Relative entropy (Kullback-Leibler divergence)			
Censoring owing to maximum follow-up		Random censoring		Censoring owing to maximum follow-up	
ESD	UDR	ESD	UDR	ESD	UDR
10.3 (5.6, 17.4)*	15.8 (12.0, 21.2)	0.231 (0.126, 0.371)	0.257 (0.126, 0.424)	0.226 (0.118, 0.367)	0.190 (0.085, 0.336)
9.7 (8.0, 11.9)*	15.5 (13.7, 17.5)	0.208 (0.173, 0.245)	0.237 (0.196, 0.282)	0.206 (0.171, 0.244)	0.173 (0.137, 0.213)
25.1 (9.4, 53.5)	14.7 (2.1, 25.3)	0.245 (0.135, 0.389)*	0.419 (0.268, 0.613)	0.404 (0.143, 1.266)	0.201 (0.054, 0.454)
26.6 (19.1, 37.1)	14.1 (9.7, 17.7)*	0.214 (0.180, 0.252)*	0.395 (0.348, 0.446)	0.245 (0.181, 0.416)	0.140 (0.081, 0.211)*
-	-	0.315 (0.169, 0.539)*	0.553 (0.378, 0.763)	-	-
-	-	0.247 (0.208, 0.288)*	0.515 (0.463, 0.568)	-	-
5.3 (0.2, 14.5)	4.3 (0.2, 12.3)	0.054 (0.020, 0.120)	0.058 (0.022, 0.126)	0.050 (0.015, 0.121)	0.047 (0.015, 0.113)
3.0 (0.2, 6.6)	1.4 (0.1, 4.0)	0.012 (0.005, 0.023)	0.016 (0.008, 0.027)	0.013 (0.005, 0.026)	0.011 (0.004, 0.021)
11.6 (0.5, 30.9)	10.4 (0.4, 29.4)	0.067 (0.022, 0.151)	0.074 (0.025, 0.160)	0.074 (0.015, 0.214)	0.065 (0.017, 0.173)
6.1 (0.3, 14.0)	7.9 (2.2, 14.1)	0.016 (0.008, 0.028)	0.024 (0.013, 0.039)	0.017 (0.005, 0.038)	0.014 (0.006, 0.031)
-	-	0.127 (0.034, 0.318)	0.118 (0.036, 0.262)	-	-
-	-	0.042 (0.022, 0.067)	0.050 (0.028, 0.077)	-	-
15.4 (10.8, 18.3)	18.4 (14.5, 24.5)	0.654 (0.508, 0.830)	0.587 (0.415, 0.795)	0.651 (0.498, 0.864)	0.645 (0.499, 0.823)
17.0 (11.4, 19.4)*	19.2 (17.1, 20.7)	0.622 (0.579, 0.667)	0.550 (0.500, 0.602)*	0.667 (0.585, 0.773)	0.601 (0.552, 0.660)
-	-	0.646 (0.501, 0.830)	0.682 (0.525, 0.875)	-	-
-	-	0.605 (0.559, 0.652)	0.649 (0.600, 0.699)	-	-
-	-	0.670 (0.518, 0.868)	0.742 (0.583, 0.938)	-	-
-	-	0.609 (0.563, 0.656)*	0.715 (0.658, 0.773)	-	-
10.9 (4.0, 18.4)	11.4 (3.7, 22.5)	0.156 (0.062, 0.322)	0.205 (0.090, 0.415)	0.176 (0.065, 0.394)	0.139 (0.054, 0.286)
9.7 (6.9, 12.7)	7.8 (4.7, 10.9)	0.089 (0.057, 0.124)*	0.127 (0.094, 0.165)	0.100 (0.064, 0.145)	0.067 (0.041, 0.100)
-	-	0.179 (0.074, 0.358)	0.245 (0.115, 0.455)	-	-
-	-	0.095 (0.062, 0.132)*	0.161 (0.122, 0.206)	-	-
-	-	0.269 (0.113, 0.544)	0.296 (0.144, 0.502)	-	-
-	-	0.130 (0.090, 0.175)*	0.201 (0.153, 0.252)	-	-

illustrate potential sample size impact on modeling outcomes. A total of 8 subgroups with sample sizes ranging from 50 to 410 were defined according to patient characteristics that were found relevant in the evaluation of the CAIRO3 study,[21] that is, treatment response (stable disease vs complete or partial response) and stage of disease (synchronous vs metachronous). See Supplemental Material 5 found at https://doi.org/10.1016/j.jval.2021.07.016 for an overview of these subgroups and their event incidences. Different levels of censoring $p_{censoring}$ ($p_{censoring} \approx 0\%$, 10%, 30%, 60%) were applied for each subgroup analysis to assess the impact of this data characteristic on the modeling approaches' performance and health economic outcomes. Censoring was performed according to the 2 censoring approaches as for the simulation study, with $p_{censoring} \approx 60\%$ not considered for follow-up censoring (discussed earlier). For each subgroup, censoring approach, and $p_{censoring}$ combination, a probabilistic analysis was performed based on 5000 runs of 10 000 patients per treatment strategy in each run. Uncertainty in time-to-event parameters was reflected using a nonparametric bootstrap approach,[22] and uncertainty in other parameters was reflected using standard parametric distributions according standard practice, for example, beta distributions for health utility values (see Degeling et al[5] for details).

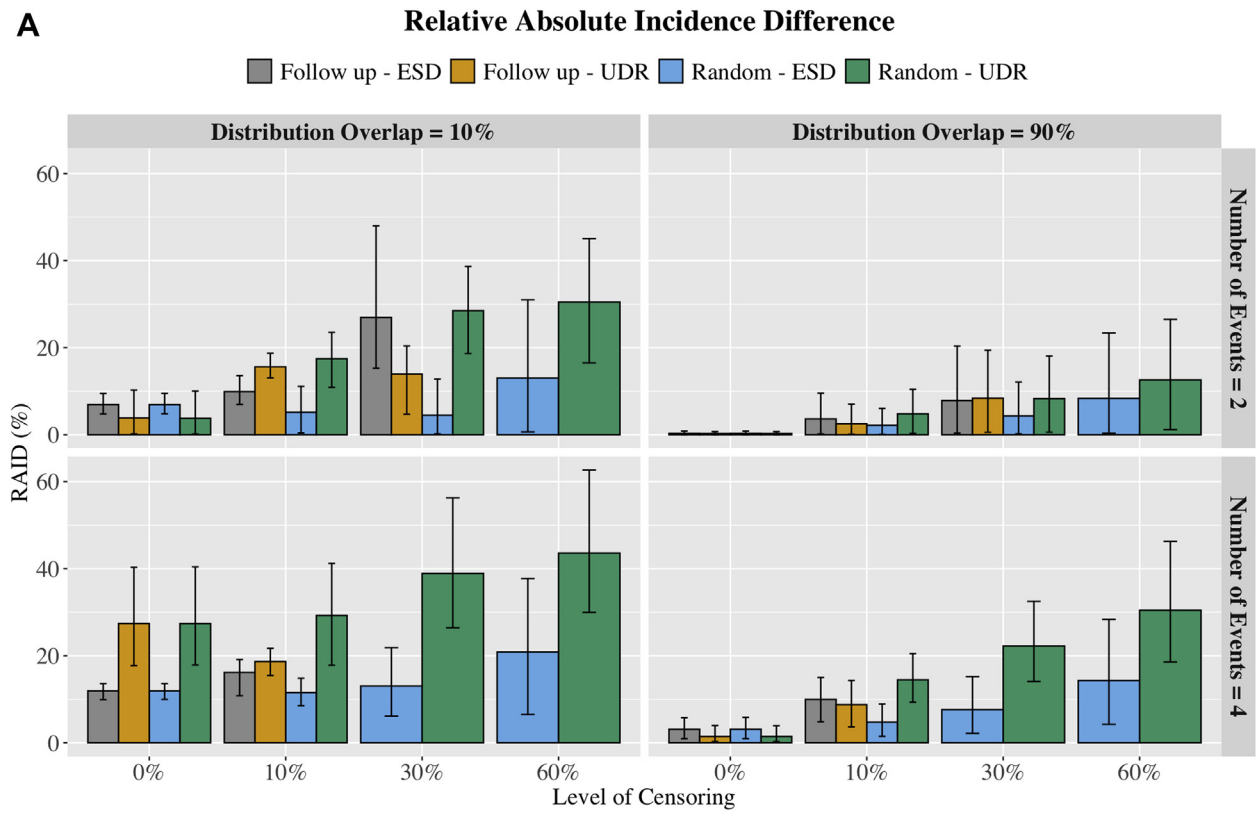
Results

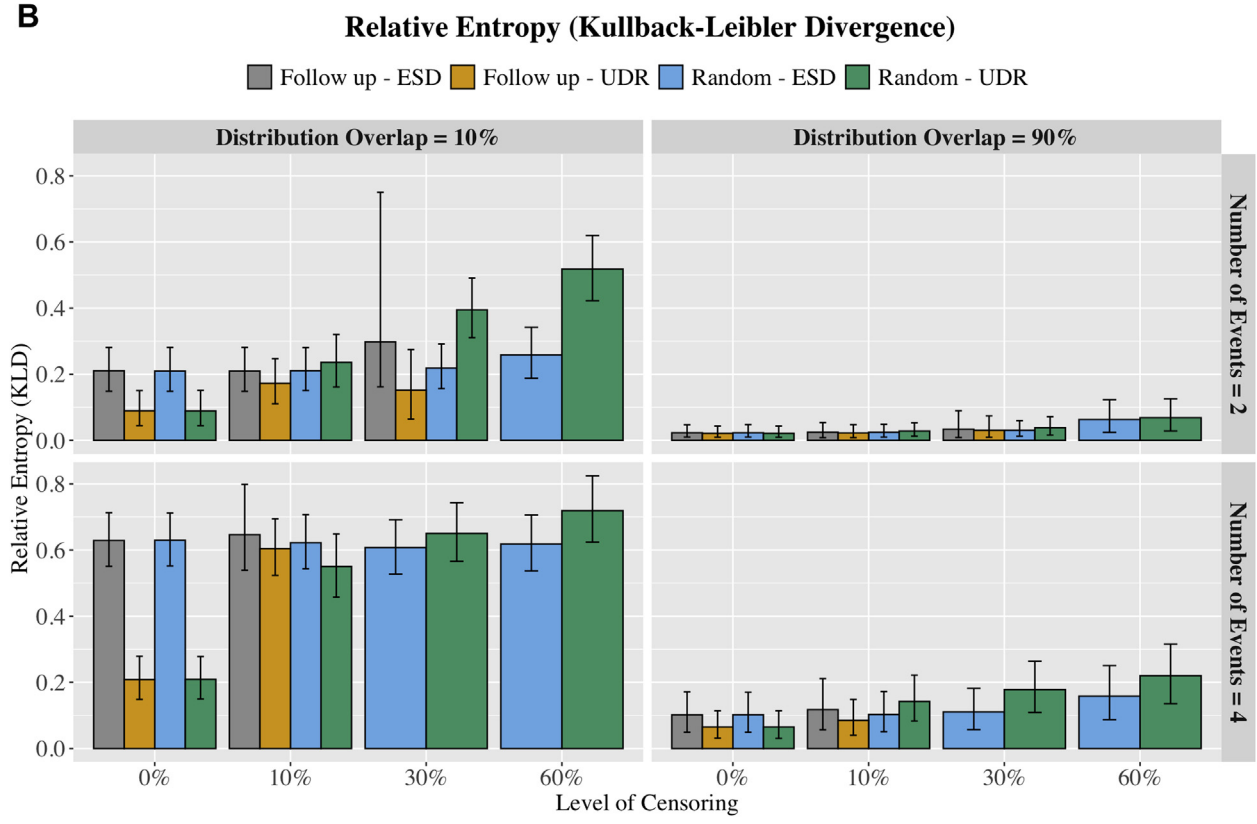
The results of the simulation study are summarized in Figure 2, which visualizes trends in the bias of the modeling approaches according to the data characteristics. Results for selected scenarios are presented in Table 1 and Figure 3, whereas results for all scenarios of the simulation study are presented in Supplemental Material 6 found at https://doi.org/10.1016/j.jval.2021.07.016. For the case study, cost-effectiveness outcomes are presented for selected subgroup analyses in Figure 4 and for all subgroup analyses in Supplemental Materials 7 and 8 found at https://doi.org/10.1016/j.jval.2021.07.016. In summarizing these results, the following section refers to censoring levels of 10%, 30%, and 60% as low, moderate, and high, respectively.

Bias of the ESD and UDR Approaches in the Simulation Study

A higher number of competing events resulted in worsened (ie, higher) RAID of 8.4%, 13.9%, and 17.1% on average over all other scenario variables for 2, 3, and 4 events, respectively. Higher overlap between time-to-event distributions resulted in improved (ie,

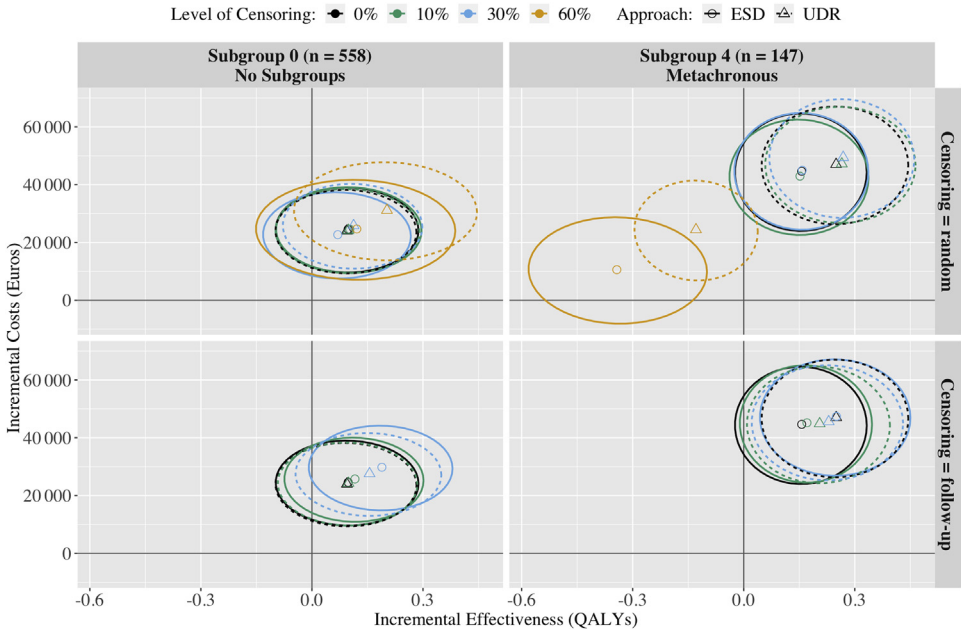
Figure 3. Bar charts of the bias of the modeling approaches (lower is better) for the different censoring mechanisms (random vs follow-up) for selected scenarios of the simulation study with regard to (A) the RAID and (B) the relative entropy in terms of the KLD. The height of the bar charts indicates the mean bias (RAID or KLD) for that specific scenario over the 10 000 simulation runs, and the error bars show the 95% confidence interval around that mean bias.





ESD indicates event-specific distribution; KLD, Kullback–Leibler divergence; RAID, relative absolute incidence difference; UDR, unimodal distribution and regression.

Figure 4. Incremental cost-effectiveness planes for selected levels of censoring and clinical subgroups of the case study and different censoring mechanisms (ie, random censoring and follow-up censoring). The ellipses represent the multivariate 95% confidence intervals around the cost-effectiveness point estimates, which are represented by the different point shapes (ie, small open circle and triangle). The different colors distinguish between different levels of censoring, whereas the different point shapes (ie, small open circle or triangle) and line types (ie, solid of dashed) distinguish between the different modeling approaches. Note that a level of 60% censoring is not included in the bottom graphs because this level of censoring could not be achieved for the follow-up censoring mechanism (see Methods).



ESD indicates event-specific distribution; UDR, unimodal distribution and regression.

lower) RAID of 18.6%, 12.7%, and 7.5% on average for 10%, 50%, and 90% overlap, respectively. Although the impact was lower than the other scenario variables, higher sample sizes resulted in improved RAID of 14.5%, 12.6%, and 11.7% on average for a sample size of 50, 150, and 500 individuals, respectively. Higher levels of censoring resulted in worsened RAID of 7.0%, 11.5%, 16.9%, and 21.2% on average for levels of 0%, 10%, 30%, and 60% censoring, respectively.

On average, the ESD modeling approach (9.8%) performed better in terms of RAID than the UDR approach (16.1%). Nevertheless, this was mainly because of a substantial difference in performance under random censoring (ESD 8.6%, UDR 18.8%), whereas under follow-up censoring the difference was limited (ESD 11.6%, UDR 11.9%). In the scenarios including random censoring, the ESD approach clearly performed better, with significantly better performance in many experiments. Under follow-up censoring, the ESD approach performed better for low censoring and the UDR approach for moderate censoring, unless when overlap was high.

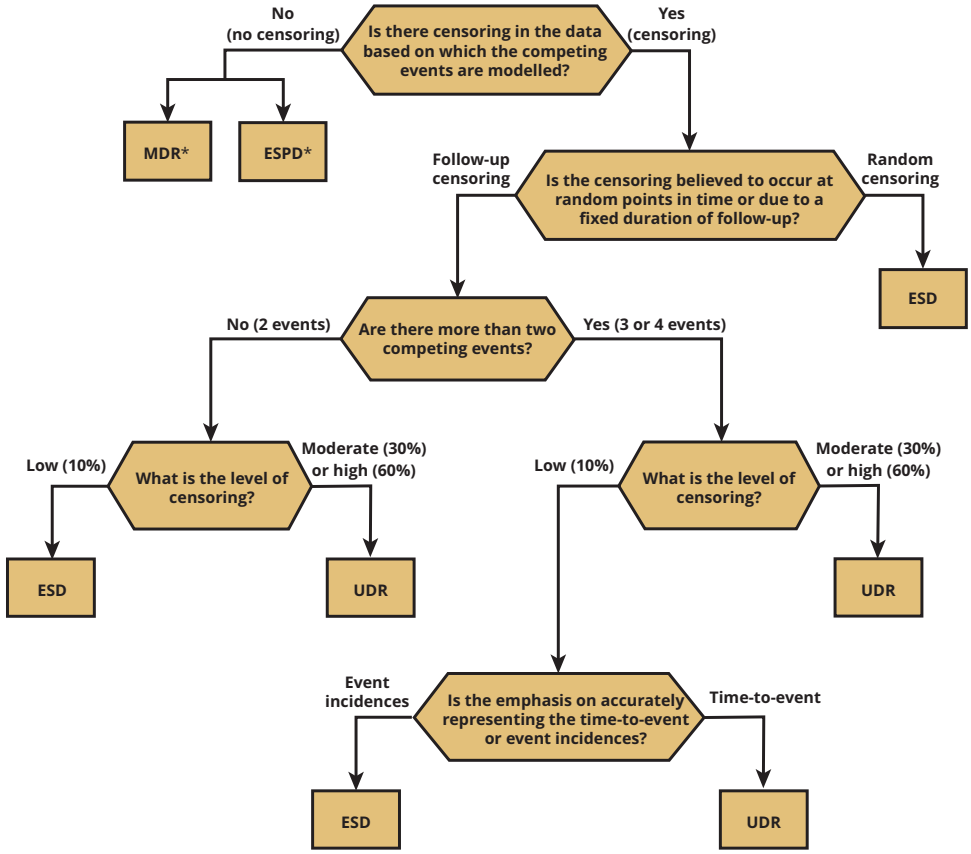
Trends in performance for KLD were in line with those for RAID. A higher number of competing events resulted in worsened (ie, higher) KLD of 0.117, 0.300, and 0.371 on average over all other scenario variables for 2, 3, and 4 events, respectively. Higher overlap between time-to-event distributions resulted in improved (ie, lower) KLD of 0.443, 0.251, and 0.078 on average for 10%, 50%, and 90% overlap, respectively. Although the impact was lower than the other scenario variables, higher sample sizes resulted in improved KLD of 0.288, 0.247, and 0.237 on average for a sample size of 50, 150, and 500 individuals, respectively. Higher levels of censoring resulted in worsened KLD of 0.189, 0.255, 0.279, and 0.360 on average for levels of 0%, 10%, 30%, and 60% censoring, respectively.

On average, the UDR modeling approach (0.238) performed better in terms of KLD than the ESD approach (0.277). Nevertheless, this was mainly because of a substantial difference in performance under follow-up censoring (UDR 0.179, ESD 0.274), whereas under random censoring the difference was limited (UDR 0.277, ESD 0.278). Under follow-up censoring, the UDR approach overall performed better, with significantly better performance in one-third of the experiments. In the scenarios including random censoring, the UDR approach performed better than the ESD approach when there was no censoring. The ESD approach performed better in scenarios including moderate or high levels of random censoring. For low levels of random censoring, the ESD approach overall performed better for scenarios including 2 competing events, whereas for 3 or 4 events no overall best-performing approach could be identified.

Impact of Health Economic Outcomes in the Case Study

The case study demonstrated that censoring had an impact on cost-effectiveness point estimates (Fig. 4 and Supplemental Material 7 and 8 found at https://doi.org/10.1016/j.jval.2021.07.016). For example, for the full cohort (n = 558), the net monetary benefit at a willingness to pay of €20 000 per quality-adjusted life-year gained changed from −€22 665 and −€22 130 without censoring to −€22 246 and −€27 078 at 60% random censoring, for ESD and UDR, respectively. This impact overall was larger for clinical subgroups with smaller sample sizes. No consistent direction in the bias introduced by increased censoring could be identified. High levels of censoring consistently resulted in increased uncertainty surrounding cost-effectiveness point estimates. For most clinical subgroups, the impact of censoring was comparable between modeling approaches.

Figure 5. Flowchart summarizing the high-level guidance for selecting a modeling approach based on the type and level of censoring, number of competing events, and importance of time-to-event predictions versus event incidences. Note that the guidance for uncensored data is based on Degeling et al,[11] which found that the ESPD and MDR approaches were preferable over the ESD and UDR approaches, and that this summary figure does not capture all nuances, such as the overlap between competing time-to-event distributions. *The ESPD and MDR were found to have similar performance and both were recommended for use, but the ESPD approach was considered more straightforward to implement.



ESD indicates event-specific distribution; ESPD, event-specific probabilities and distributions; MDR, multimodal distribution and regression; UDR, unimodal distribution and regression.

Discussion

It is widely known that censoring needs to be accounted for in health economic analyses, and our results substantiate this for DES models including competing events. Previous research on modeling approaches for implementing competing events in DES models argued a limited generalizability of recommendations for scenarios "without censoring" to those "with censoring" because of anticipated differences in performance following different implementations of the modeling approaches.[11] Our study indeed demonstrates that increased levels of censoring affected the performance of modeling approaches in the simulation study and cost-effectiveness outcomes in the case study. Additionally, moderate (30%) or higher levels of censoring inflated uncertainty in cost-effectiveness estimates and, hence, the presence of censoring may affect decision uncertainty and the value of collecting further information.

Which modeling approach is preferable based on the simulation study results depends on the type of noninformative right censoring and whether accuracy in event incidences or time-to-event distributions is considered more important. Guidance for selecting a modeling approach for uncensored IPD based on Degeling et al[11] and for censored IPD based on the current study is presented in Figure 5. It is important to realize that this summary cannot possibly capture all nuances, such as the interaction between the overlap of competing time-to-event distributions and level of censoring, but it may serve as a general guide for identifying an appropriate modeling approach. As can be seen from the provided R code, both approaches are straightforward to implement, although the ESD may be considered slightly more practical because it only involves time-to-event modeling, whereas the UDR approach also involves (multinomial) logistic regression modeling conditional on the time to event. Importantly, modelers should be aware of the different approaches that are available and perform extensive internal validation to inform which approach will be used or, at least, verify that the chosen approach extrapolates the event incidences and time-to-events appropriately.

These findings are partly in line with the Professional Society for Health Economics and Outcomes Research and Society for Medical Decision Making modeling good research practices guidelines[23] to first sample the time to event from a joint time-to-event distribution and then sample the corresponding event, which corresponds to the UDR approach. Nevertheless, it is important to realize this recommendation did not discuss the impact of using censored IPD. Furthermore, the performance of this strategy heavily depends on the absence of multimodality in

the combined time-to-event distribution, which is why mixture distributions were used in the study providing guidance for uncensored IPD.[11] Because it uses event-specific time-to-event distributions, the ESD approach is better able to represent distributions with low overlap.

The results of the simulation study demonstrate that decision analytic modelers should be aware of the alternative modeling approaches available and that these might result in different levels of performance and health economic outcomes. Extensive (reporting of) validation efforts is essential to assess whether IPD are appropriately represented. Nevertheless, we acknowledge that validation in the presence of censoring can be challenging, because common measures to assess performance of the modeling approaches, such as the RAID and KLD, cannot be used for censored data. Nevertheless, these measures could be used in this study, because the simulated event incidences and time-to-event distributions could be compared with the "uncensored truth" (Fig. 1), which will not be available for studies using censored data in practice. Although less straightforward to interpret and challenging to combine into 1 performance measure for modeling approaches as a whole, alternative measures are available to assess discrimination and calibration of single survival models, while accounting for censoring.[24] An example is Demler et al's Greenwood-Nam-D'Agostino statistic,[25] which is a modification of Hosmer-Lemeshow's statistic by Nam and D'Agostino.[26]

Another point of consideration is whether multivariable models are being developed to reflect patient heterogeneity, so that patient characteristics or treatment histories influence the occurrence and timing of events. Multivariable models are most easily implemented for the UDR approach, especially for scenarios including more than 2 competing events, because the variable selection procedures, for example, only need to be performed for 1 survival model and 1 (multinomial) logistic regression model. Nevertheless, further research is needed to assess potential differences between approaches when used to reflect such heterogeneity.

In this article, we did not include an ESPD modeling approach based on the strategy of sampling an event first and the corresponding time-to-event second, nor did we include a MDR modeling approach based on the same strategy as UDR. Although both approaches were included in the study considering uncensored IPD[11] and we have successfully implemented them for censored IPD, we found their implementation unproportionally cumbersome for censored IPD without clear benefits in terms of performance compared with ESD and UDR. The ESPD approach requires event-specific probabilities and time-to-event distributions to be estimated, for which there is no support in standard statistical software packages, requiring modelers to define and optimize custom likelihood functions themselves, preferably for different parametric distribution types. Although mixture distributions can theoretically be used to implement the multimodal time-to-event distribution in an MDR approach,[27,28] we believe this approach will unlikely be convenient in practice. It required the ESPD approach to be applied to generate start values to increase the probability of convergence in the maximum likelihood estimation. We found convergence and performance were poor compared with the other modeling approaches. Another approach that could be explored in further research is to use flexible parametric survival distributions, such as survival splines,[29] to model multimodal time-to-event distributions.

This study has certain limitations. First, we did not consider transformations of time in the (multinomial) logistic regression model for the UDR approach, which may have negatively affected its performance with regard to RAID because a linear relation does not necessarily best describe the relation between the time to event and event probabilities. Second, only noninformative and right-censored data were considered. Hence, results and recommendations cannot be generalized to scenarios involving informative, interval-, or left-censored data. Third, generalizability is also limited by the deliberate decision to use Weibull distributions in the simulation study. Although this allowed for an unbiased comparison of the modeling approaches, underlying distributions are unknown in practice and performance of the approaches may vary depending on the flexibility of selected distributions to describe the data. Finally, we generated datasets by first sampling an event based on event-specific probabilities and conditionally sampling times from event-specific time-to-event distributions, which may have benefited one of the approaches. Alternatively, datasets can be generated according mechanisms that are consistent with the ESD or UDR approach. Further research is warranted into all these aspects.

Conclusions

Censoring has an impact on the performance of modeling approaches to implement competing events in DES models and, thereby, affects cost-effectiveness point estimates. When IPD are censored at random times, the ESD modeling approach performed best in the simulation study. When censoring occurs because of a maximum follow-up time for 2 competing events, the ESD approach performed best for low levels of censoring (ie, 10%) and the UDR approach for moderate levels of censoring (ie, 30%). For scenarios including 3 or 4 competing events and follow-up censoring, the UDR approach represented the time-to-event distributions more accurately, whereas the ESD approach performed better in terms of the event incidences. Nevertheless, substantial differences in performance between the modeling approaches for different scenarios highlighted the need for extensive validation efforts by modelers to assure IPD are appropriately represented.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2021.07.016.

Article and Author Information

Accepted for Publication: July 29, 2021

Published Online: xxxx

doi: https://doi.org/10.1016/j.jval.2021.07.016

Author Affiliations: Department of Health Technology and Services Research, Faculty of Behavioural, Management, and Social Sciences, Technical Medical Centre, University of Twente, Enschede, The Netherlands (Degeling, IJzerman, Groothuis-Oudshoorn, Koffijberg); Cancer Health Services Research, Centre for Cancer Research, Faculty of Medicine, Dentistry, and Health Sciences (Degeling, IJzerman) and Cancer Health Services Research, Centre for Health Policy, Melbourne School of Population and Global Health (Degeling, IJzerman), University of Melbourne, Melbourne, Australia; Department of Cancer Research, Peter MacCallum Cancer Centre, Melbourne, Australia (IJzerman); Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (Clements); Department of Medical Oncology, University Medical Centre, Utrecht University, Utrecht, The Netherlands (Franken, Koopman).

Correspondence: Koen Degeling, PhD, Cancer Health Services Research, University of Melbourne, VCCC, Level 10, 305 Grattan St, Melbourne, Victoria, Australia 3000. Email: koen.degeling@unimelb.edu.au

Author contributions:
Concept and design: Degeling, Koffijberg

Acquisition of data: Degeling, Koopman
Analysis and interpretation of data: Degeling, IJzerman, Koopman, Clements, Koffijberg
Drafting of the manuscript: Degeling, Groothuis-Oudshoorn, Franken, Koffijberg
Critical revision of the paper for important intellectual content: Degeling, IJzerman, Groothuis-Oudshoorn, Franken, Koopman, Clements, Koffijberg
Statistical analysis: Degeling, Groothuis-Oudshoorn, Clements, Koffijberg
Provision of study materials or patients: Franken
Obtaining funding: IJzerman
Supervision: IJzerman

Conflict of Interest Disclosures: Dr IJzerman reported receiving grants, nonfinancial support, and an advisory board honorarium paid to his institution from Illumina. Dr Koopman reported receiving payments to her institution from Bayer, Bristol Myers Squibb, Merck, Roche, Servier, and Pierre Fabre outside the submitted work. Dr IJzerman is an editor for *Value in Health* and had no role in the peer review process. No other disclosures were reported.

Funding/Support: The authors received no financial support for this research.

REFERENCES

1. Annemans L, Redekop K, Payne K. Current methodological issues in the economic assessment of personalized medicine. *Value Health*. 2013;16(6 Suppl):S20–S26.
2. Caro JJ, Möller J, Getsios D. Discrete event simulation: the preferred technique for health economic evaluations? *Value Health*. 2010;13(8):1056–1060.
3. Miller JD, Foley KA, Russell MW. Current challenges in health economic modeling of cancer therapies: a research inquiry. *Am Health Drug Benefits*. 2014;7(3):153–162.
4. Caro JJ, Briggs AH, Siebert U, Kuntz KM. ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices—overview: a report of the ISPOR-SMDM modeling good research practices task Force-1. *Value Health*. 2012;15(6):796–803.
5. Degeling K, Franken MD, May AM, et al. Matching the model with the evidence: comparing discrete event simulation and state-transition modeling for time-to-event predictions in a cost-effectiveness analysis of treatment in metastatic colorectal cancer patients. *Cancer Epidemiol*. 2018;57:60–67.
6. Karnon J, Haji Ali Afzali H. When to use discrete event simulation (DES) for the economic evaluation of health technologies? A review and critique of the costs and benefits of DES. *Pharmacoeconomics*. 2014;32(6):547–558.
7. Standfield L, Comans T, Scuffham P. Markov modeling and discrete event simulation in health care: a systematic comparison. *Int J Technol Assess Health Care*. 2014;30(2):165–172.
8. Caro JJ, Möller J. Decision-analytic models: current methodological challenges. *Pharmacoeconomics*. 2014;32(10):943–950.
9. Pintilie M. *Competing Risks: A Practical Perspective*. Hoboken, NJ: Wiley; 2006.
10. Barton P, Jobanputra P, Wilson J, Bryan S, Burls A. The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. *Health Technol Assess*. 2004;8(11):iii-91.
11. Degeling K, Koffijberg H, Franken MD, Koopman M, IJzerman MJ. Comparing strategies for modeling competing risks in discrete-event simulations: a simulation study and illustration in colorectal cancer. *Med Decis Making*. 2019;39(1):57–73.
12. Donoghoe MW, Gebski V. The importance of censoring in competing risks analysis of the subdistribution hazard. *BMC Med Res Methodol*. 2017;17(1):52.
13. Leung KM, Elashoff RM, Afifi AA. Censoring issues in survival analysis. *Annu Rev Public Health*. 1997;18(1):83–104.
14. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019.
15. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. *Circulation*. 2016;133(6):601–609.
16. Cover TM, Thomas JA. *Elements of Information Theory*. 2nd ed. Hoboken, NJ: John Wiley; 2006.
17. Venables WN, Ripley BD. *Modern Applied Statistics With S*. 4th ed. New York, NY: Springer; 2002.
18. Grün B, Leisch F. Fitting finite mixtures of generalized linear regressions in R. *Comput Stat Data Anal*. 2007;51(11):5247–5252.
19. Grün B, Leisch F. FlexMix version 2: finite mixtures with concomitant variables and varying and constant parameters. *J Stat Softw*. 2008;28(4):35.
20. Leisch F. FlexMix a general framework for finite mixture models and latent class regression in R. *J Stat Softw*. 2004;11(8):18.
21. Simkens LHJ, van Tinteren H, May A, et al. Maintenance treatment with capecitabine and bevacizumab in metastatic colorectal cancer (CAIRO3): a phase 3 randomised controlled trial of the Dutch Colorectal Cancer Group. *Lancet*. 2015;385(9980):1843–1852.
22. Degeling K, IJzerman MJ, Koopman M, Koffijberg H. Accounting for parameter uncertainty in the definition of parametric distributions used to describe individual patient variation in health economic models. *BMC Med Res Methodol*. 2017;17(1):170.
23. Karnon J, Stahl J, Brennan A, et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM modeling good research practices task Force-4. *Value Health*. 2012;15(6):821–827.
24. Rahman MS, Ambler G, Choodari-Oskooei B, Omar RZ. Review and evaluation of performance measures for survival prediction models in external validation settings. *BMC Med Res Methodol*. 2017;17(1):60.
25. Demler OV, Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. *Stat Med*. 2015;34(10):1659–1680.
26. D'Agostino RB, Nam B-H. Evaluation of the performance of survival analysis models: discrimination and calibration measures. *Handb Stat*. 2003;23:1–25.
27. Bordes L, Chauveau D. Stochastic EM algorithms for parametric and semiparametric mixture models for right-censored lifetime data. *Comput Stat*. 2016;31(4):1513–1538.
28. Davenport JW, Bezdek JC, Hathaway RJ. Parameter estimation for finite mixture distributions. *Comput Math Appl*. 1988;15(10):819–828.
29. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*. 2002;21(15):2175–2197.