Journal of Choice Modelling 41 (2021) 100308

Contents lists available at ScienceDirect

Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Joint analysis of preferences and drop out data in discrete choice experiments

Leonard Maaya *, Michel Meulders, Martina Vandebroek

Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium

ARTICLE INFO

Keywords:
Discrete choice models
Joint models
Dropouts
Shared effects

ABSTRACT

Choice data appear together with drop out data indicating if respondents completed the exercise. In case of non-completion, the choice sequences end at the tasks where the respondents exited the study. In the analysis of choice data, the focus is always on choices made while the drop out behavior is completely ignored. However, the choice making and the drop out process could be latently related. For instance, respondents who are more likely to drop out of the exercise could give less consistent choices throughout or just before they exit. In such cases, ignoring the drop out dimension could lead to biased or inefficient results. In this paper, we use shared random effects and covariate effects to model the association between a scaled multinomial logit model for the choices and two different models for the drop out component. Through simulations, we show that a joint model provides less biased and more precise estimates and its 95% credible intervals have better coverage for true parameter values.

1. Introduction

A Discrete Choice Experiment (DCE) is an attribute-based method used to investigate individuals' preferences among alternatives. DCEs involve hypothetical scenarios (choice sets or choice tasks) composed of several competing alternatives. The alternatives in a choice set are designed to vary along their attribute levels (see Rose and Bliemer, 2009 for a review of design of choice experiments). Respondents are then asked to choose a preferred alternative in each scenario. The choice experiments can be done offline through mailing or by providing choice cards to respondents. With the near-universal availability of technology however, more and more studies are being conducted online. Often, respondents are presented with several choice scenarios and therefore provide a sequence of choices. The sequence of scenarios can additionally be considered as having a discrete *time* feature arising from the serialization of choice tasks. The repeated nature of choice data therefore calls for a special analysis to handle the correlation in responses and the underlying *time* dimension.

Most of the focus in the literature has been directed to accounting for the correlation arising from respondents providing repeated choices. In this regard, the mixed multinomial logit (MMNL) (Train, 2009; Hess and Train, 2017) has been popular and is a highly flexible model that allows parameters to be person-specific and correlated. Restrictions to the MMNL model have also been proposed in cases where estimating a full covariance matrix is uninteresting or to minimize the computational burden. For example, the scaled multinomial logit (SMNL) (Fiebig et al., 2010; Sarrias and Daziano, 2017) model introduces a person-specific scale that multiplies the fixed utility coefficients. This multiplication ensures that there is a proportional change to the affected coefficients for each respondent. Whereas ways to handle the correlation feature in the analysis of choice data have largely been addressed, research into the time-like feature has only been limited to respondents' fatigue and learning effects (Hess et al., 2012; Czajkowski et al.,

* Corresponding author.
E-mail address: lwafulae@gmail.com (L. Maaya).

https://doi.org/10.1016/j.jocm.2021.100308
Received 23 November 2020; Received in revised form 7 May 2021; Accepted 29 July 2021
Available online 7 September 2021

2014). However, as we note below, long and complex choice experiments amplify fatigue and learning effects which can lead to respondents dropping out.

Long surveys can be tiring or could be associated with respondent learning. The influence of these effects on responses has been assumed to be limited to the model scale (Hess et al., 2012; Czajkowski et al., 2014). First, if respondents become fatigued, their involvement in the survey reduces and laxity creeps in. As a result, precision (scale) in their responses decreases (increases). Second, learning implies that as respondents go through the choice exercise, their understanding improves and they therefore provide more precise choices. Studies have also shown that precision in responses can vary with the complexity of choice tasks (Danthurebandara et al., 2015). This is because more complex tasks require more effort and time to make a choice, which can compromise the precision in the responses, compared to simple tasks. Of more relevance for this research, however, is that the effects of lengthy and complex surveys are not limited to the precision in responses. Such surveys can also result in non-completion of choice exercises by some participants. This article focusses on the effects of the time-like feature on the respondents' drop out behavior.

To investigate the time feature in choice data, we juxtapose it with longitudinal data with which it shares some important characteristics. First, responses (choices in choice data) from the same respondent are more likely to be correlated compared to those from other respondents. Second, there is usually a pre-specified number of responses that a respondent is expected to provide by design. Third, while all respondents are expected to provide a complete sequence of responses, it is common that some respondents do not provide responses in some instances or drop out entirely before the end of the study. Sequences for which no choice is observed after the first missing response are said to exhibit monotone missingness. Any respondent so affected is called a dropout (Molenberghs and Verbeke, 2005). Some sequences may also have a missing response preceding or sandwiched between observed choices. These are usually termed as intermittent missing responses and they are said to exhibit non-monotone missingness. The presence of missing responses gives rise to unbalanced data which needs to be incorporated in analyses to provide reliable estimates for decision making (Molenberghs and Verbeke, 2005).

In DCE analyses, interest in missingness has mostly centered on covariates (Qian and Xie, 2011; Sanko et al., 2014; Varotto et al., 2017; Irannezhad et al., 2019). Qian and Xie (2011) proposed a Bayesian approach to include respondents with missing covariate values in analyses. The Bayesian method samples missing values from their conditional distributions through a data augmentation technique. Missing covariate values have also been modeled as latent variables in hybrid models (Sanko et al., 2014; Varotto et al., 2017; Irannezhad et al., 2019). The latent approach assumes that stated values are subject to measurement errors and attempts to correct the errors while replacing missing values. The twofold correction is theoretically better compared to an imputation approach which assumes that stated values are error-free measures of real values (Sanko et al., 2014; Molenberghs and Verbeke, 2005). We add to these studies by modeling a specific form of missingness in responses.

We restrict our investigation to surveys where missingness in choices is monotone. We view that the drop out patterns provide extra insights on choice making behavior which can enrich available information for estimation of parameters. To explore this, we model the *time* to a drop out event as an additional component to the usual choice behavior. Here, *time* is a borrowed term from survival analysis and it is used to refer to the number of tasks respondents attempt before dropping out. In a typical sense, however, time would refer to the time spent by a respondent in a survey as a whole or the time spent at a choice task to make a choice if such data is recorded. Jointly modeling the choices and the risk of stopping the experiment can be advantageous in correcting biases introduced by the occurrence of the drop out event (Tsiatis and Davidian, 2004; Wu et al., 2012; Proust-Lima et al., 2014). The joint model is also useful when there is interest in investigating if the choice making process and the risk to stop the experiment are governed by the same latent process. This could be the case where a latent factor that leads to a respondent stopping to provide choices also compromises their evaluation of tasks leading to less precise choices. Further, incorporating the drop out model can help in investigating reasons for respondents not finishing the choice experiment. Possible reasons may include lengthy surveys, increasing choice complexity, fatigue and boredom or respondents' protests to a combination of the mentioned reasons.

Choice and drop out data appear together frequently in surveys. Fig. 1 shows a sequence of choices for some respondents in an online study on university choices. For this study, 14 choice scenarios were presented to participants. Each scenario comprised of three alternatives (A, B and C) and an optout option (O). The figure shows that some participants completed the exercise (e.g., ID 5, 40 and 165 on the y-axis). These are completers. Some other participants did not provide responses in all the choice sets (e.g., IDs 55, 90, 186, 202 etc.). They are the non-completers. We observe that once a missing value occurs, all succeeding choice sets also have missing responses. Therefore, all non-completers in this case were dropouts.

Analyses on choice data similar to that in Fig. 1 often take one of two forms. First, some analyses exclusively focus on the completers and discard all information from dropouts. This means that out of the fifteen sampled profiles in Fig. 1, only five would be analyzed. This approach can have far-reaching consequences because it reduces the sample size, can result in datasets that do not reflect original populations and can be endogenous due to self-selection (Qian and Xie, 2011; Sanko et al., 2014). Results from analyzing completers can also be biased and inefficient if the drop out mechanism is non-random or if the drop out mechanism is random but the proportion of completers is low.

Second, some analyses include all the choices up to when respondents drop out of the exercise. In this case, all the choices from the fifteen profiles in Fig. 1 are included in the analysis. This approach uses a richer dataset and thus mitigates some of the shortcomings of relying on completers. However, it is possible that the responses provided are associated with the risk of dropping out. For instance, respondents could exhibit fatigue effects before dropping out. In such cases, dropouts are likely to provide unreliable responses towards the end of the study or prior to stopping. For example, respondent ID 77 in Fig. 1 selects the optout option in all the last eight choice tasks. Choosing the optout at this rate could imply that the respondent was no longer trading alternatives in these choice sets. This was particularly possible in this survey since a respondent had to make a choice in a task before accessing subsequent tasks. Respondent 15 could similarly be considered as no longer making choices after the twelfth

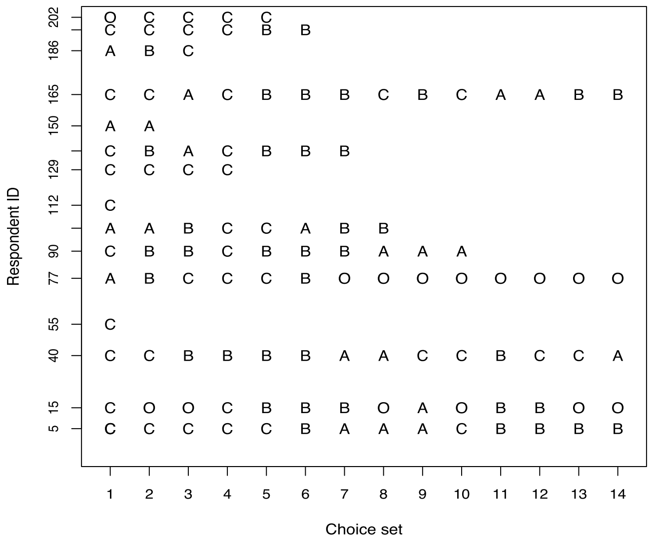

Fig. 1. Sequence of choices for a sample of respondents.

choice task. More dropouts may also be observed around complex choice sets, as the choice complexity builds up or in lengthy surveys. In these cases, analyzing choice data without examining the drop out dimension could bias or lead to inefficient results since it masks the existing behavioral intricacies. Conversely, jointly analyzing preferences and the risk to drop out incorporates all the information at once and provides valid and efficient estimates. Further, no extra effort is needed on the part of the respondents to observe dropouts since they are by-products of the available data. Depending on how extensive the drop out behavior is to be investigated, a researcher may need to record data on time spent by a respondent in the whole survey or at respective choice tasks. This can be done by specifying timestamps at the beginning and at the end of the survey or on every click made at each choice task.

Typically, a joint model combines two or more submodels (or parts). With respect to discrete choice experiments in this article, a joint model will comprise of two submodels: one for the choice data and another for the drop out data. The two submodels are then linked by sharing some random effects or variables. Joint models are commonly estimated using either a two-stage or a unified likelihood maximization approach (Tsiatis and Davidian, 2004; Wu et al., 2012). In a nutshell involving shared random effects, a simple two-stage approach involves first fitting one of the submodels with random effects. Then, the second submodel is fit separately with the shared random effects replaced by their estimates from the first step as if they were observed values. The advantage of the two-stage approach is that it is simple and can be easily estimated using available software. However, this approach can lead to biased estimates since it does not incorporate the uncertainty in the first step into the second step. This can lead to underestimation of standard errors for the submodel estimated in the second step. Further, the approach does not utilize the combined information from the choice and drop out data to provide the most efficient estimates. The unified likelihood maximization approach on the other hand attempts to overcome these shortcomings by simultaneously estimating both submodels. We discuss a standard formulation of a joint model tailored for choice experiments and its estimation in Section 2. We then illustrate and evaluate the method in Section 3 using simulation studies. Section 4 concludes the article with a discussion.

2. Joint modeling

The primary objective in DCEs is to model the choice behavior by estimating the attributes' part-worth utilities. We augment this by modeling the risk of respondents stopping the choice exercise earlier than planned. We introduce a general joint model in Section 2.2. This is then followed by two different joint models in Sections 2.2.1 and 2.2.2. In both joint model formulations, preferences are modeled using an SMNL model (Fiebig et al., 2010; Sarrias and Daziano, 2017; Hess and Train, 2017). The discrete time hazard model (Singer et al., 2003) is first considered for the drop out dimension owing to the discrete nature of choice experiments. We then extend the discrete time hazard model using a piece-wise exponential drop out model (Holford, 1980; Whitehead, 1980; Laird and Olivier, 1981; Rodríguez, 2007) where time is assumed to be continuous but split into small intervals at the choice sets. In each of these time intervals (i.e. at each choice set), it is further assumed that the risk of dropping out is constant (i.e. piece-wise constant). In both cases, we analyze the drop out behavior using equivalent generalized linear models: a logistic model for the discrete-time hazard and a specific Poisson formulation for the piece-wise proportional hazard model. The two parts of the respective joint models are then linked through shared random effects (in the first joint model) and shared covariates (in the second joint model). Notations used and the model descriptions are provided next.

2.1. Notation

We consider a survey with I respondents each faced with S choice scenarios. Each of the scenarios has M alternatives. Let S_i ($\mathrm{S}_i \leq$ S) be the total number of available responses for i. Let $\mathbf{y}_i = (y_{im1}, \ldots, y_{im\mathrm{S}_i})'$ where y_{ims} equals 1 if person i selects alternative m in choice situation s and 0 otherwise. The index i in y_{ims} equals $1, 2, \ldots,$ to I while $s = 1, \ldots, \mathrm{S}_i$ and $m = 1, \ldots, \mathrm{M}$. Let T_i be the total time spent answering all the S_i choice tasks, t_{is} the time spent at choice set s such that $\sum_{s=1}^{\mathrm{S}_i} \mathrm{t}_{is} = \mathrm{T}_i$ and $\mathbf{t}_i$ the vector containing all t_{is} for a respondent i. Let $\mathbf{x}_{ims}$ be the collection of attribute values for alternative m in choice set s faced by person i. Let $\mathbf{d}_i = (d_{i1}, \ldots, d_{iS_i})'$ be a vector with the first $\mathrm{S}_i - 1$ items equal to 0 and d_{iS_i} equal to 1 if person i is a dropout. If person i is a completer, $\mathbf{d}_i$ is a vector of 0's and is of length S. We confine our discussion to monotone missing responses where once there is a missing response, all succeeding tasks also have missing choice responses. For completers, we never know if they would have dropped out had the choice exercise been longer or more complex. They are said to be *right censored*. We assume that the censoring mechanism is independent of the risk to drop out. This means that a person's censoring does not provide any more information regarding the person's risk of dropping out of the survey beyond the censoring time. It is normally referred to as random censoring or *non-informative* censoring. We also assume that for every respondent, $d_{i1} = 0$. This implies that at minimum, every person provides a response for the first task.

2.2. Models for choices and drop out data

A general specification of a joint model for choices and drop out behavior is shown in Eq. (1) where the choices for person i are shown by the vector $\mathbf{y}_i$ and information on time to a drop out event is represented by a pair of vectors $\mathbf{d}_i$ and $\mathbf{t}_i$. Often, instead of working with the joint model directly, researchers factorize the joint model into combinations of marginal and conditional submodels resulting in what are commonly known as selection, pattern mixture and shared parameter models (Molenberghs and Verbeke, 2005). For this article and in line with common specifications of joint models involving drop out events (Tsiatis and Davidian, 2004; Wu et al., 2012), we use the shared parameter model where the joint density factorizes into two conditional densities:

$$f_0(\mathbf{y}_i, \mathbf{d}_i, \mathbf{t}_i | \mathrm{u}_i, \mathbf{z}_i, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = f_{01}(\mathbf{d}_i, \mathbf{t}_i | \boldsymbol{\theta}, \boldsymbol{\gamma}_d, \mathrm{u}_i, \mathbf{z}_i) f_{02}(\mathbf{y}_i | \boldsymbol{\beta}, \boldsymbol{\gamma}_y, \mathrm{u}_i, \mathbf{z}_i). \tag{1}$$

The two conditionals (annotated by functions $f_{01}(.)$ and $f_{02}(.)$ in Eq. (1)) can share either the random effects (u_i), person-specific characteristics ($\mathbf{z}_i$) or both random effects and the individual effects. Conditional on the shared effects, the two submodels are assumed to be independent of each other. The vector $\boldsymbol{\beta}$ contains parameters related to the part-worth utilities in the choice submodel while the vector $\boldsymbol{\theta}$ contains parameters specific to the drop out submodel. The vector $\boldsymbol{\gamma}$ contains parameters related to shared individual covariates between the submodels. In this general specification, we assume different effects (and possibly dimensions) for $\boldsymbol{\gamma}$ in the choices ($\boldsymbol{\gamma}_y$) and the drop out ($\boldsymbol{\gamma}_d$) parts. In the joint models described below, however, we constrain vectors $\boldsymbol{\gamma}_y$ and $\boldsymbol{\gamma}_d$ to be equal for illustration purposes. Details on the two joint models considered in this article, both of which use the SMNL model for the choice submodel but differ on the submodel used for the drop out dimension are provided next. The first associates an SMNL and a discrete time hazard model through random effects (u_i) while the second combines an SMNL and piece-wise proportional hazard model using individual characteristics ($\mathbf{z}_i$).

2.2.1. Discrete-time hazard and the scaled multinomial logit joint model

First, we suppose that respondents make choices at discrete times corresponding to the choice sets. In this case, the first time point is at the first choice set, the second time point at the second choice set and so on until the S^{th} time point at the last choice set. The implication of this specification is that every person takes the same amount of time to make a choice at each of the different tasks. For such a setting, a discrete-time proportional hazards model can be used to estimate the risk of dropping out at the respective choice tasks. Cox (1972) showed that by working with the conditional odds of dropping out at a choice set, the discrete-time proportional hazard model can be fit using a logistic model (Efron, 1988; Singer et al., 2003; Rodríguez, 2007). In this case, the drop out indicators d_{ij} are treated as if they were independent Bernoulli outcomes with a probability equal to the hazard. The hazard (denoted by λ_{is}) is the conditional probability that respondent i drops out in choice set s given that s(he) has not dropped out in an earlier choice set. We note that the drop out indicators d_{is} are not assumed to be independent Bernoulli observations because they are not. Person i can only drop out at choice set s if s/he has not dropped out at choice set k for $k < s$. Therefore, there is interdependence. Rather, it is the equivalence in the likelihood function of a discrete-time survival model under non-informative censoring and the binomial likelihood obtained by treating the drop out indicators as if they were independent Bernoulli outcomes that is utilized. For illustration purposes, we use the logistic model shown in Eq. (2) where the logit of the hazard is modeled as a linear function of *time*. To provide a better understanding of the drop out behavior, the hazard model can be enriched with choice set complexity measures (Danthurebandara et al., 2015) and respondents' socio-demographic information if it is available.

$$\mathrm{logit}(\lambda_{is}) = \mathrm{u}_i + \theta_1 + \theta_2(s-1), \qquad i = 1, \ldots, \mathrm{I}, \quad s = 1, \ldots, \mathrm{S}_i \tag{2}$$

where,

$$\lambda_{is} = P(d_{is} = 1 | d_{ik} = 0, 1 \leq k < s; \mathrm{u}_i).$$

In Eq. (2), we assume that the hazard function (λ_{is}) depends on respondent-specific random effects u_i and unknown parameters θ_1 and θ_2. The random effects u_i are described below because they are uniquely defined to enable a smooth implementation of the SMNL model. Eq. (2) implies that the probability of respondents dropping out after the first choice set varies. Those with higher u_i values are more likely to drop out while those with lower u_i values are less likely to drop out. The unknown parameters θ_1 and θ_2

capture the logit of the probability to drop out in the first and later choice sets respectively given u_i. A positive θ_2 indicates that the logit of the hazard is higher in later choice tasks. This situation can be interpreted as an increased risk for respondents to drop out due to fatigue effects. On the other hand, a negative θ_2 shows that the logit of the hazard is lower in later choice tasks. Therefore, a negative θ_2 can be viewed as a reduced risk to drop out due to learning.

The contribution of respondent i to the drop out component in the joint model is given by:

$$\begin{aligned} f_1(\mathbf{d}_i, \mathbf{t}_i | \mathrm{u}_i, \theta_1, \theta_2) &= \prod_{s=1}^{\mathrm{S}_i - 1} (1 - \lambda_{is}) \lambda_{i\mathrm{S}_i}, \quad \text{if person } i \text{ is a dropout} \\ &\text{or} \\ &= \prod_{s=1}^{\mathrm{S}} (1 - \lambda_{is}), \quad \text{if person } i \text{ is a completer} \end{aligned} \tag{3}$$

where the time component in this joint model can be ignored because it is assumed to be the same for everyone at every choice set.

For the repeated choices in the joint model, we consider the SMNL model which is based on the random utility theory. The theory assumes that the utility that respondent i derives from selecting alternative m in choice situation s can be divided into two summable parts: a deterministic and a random component. The deterministic component (denoted by V_{ims}) is assumed to be linear in the attribute parameters while the random component is assumed to be identically and independently distributed following a type I extreme value distribution. The expression for the deterministic component in an SMNL model is shown in Eq. (4).

$$V_{ims} = \sigma_i \mathbf{x}'_{ims} \boldsymbol{\beta} \tag{4}$$

where σ_i is a person-specific scaling factor. Adapting SMNL specifications in Fiebig et al. (2010) and Sarrias and Daziano (2017) to the joint model in this article, we assume that the scales are log-normally distributed with mean $\bar{\sigma}$ and standard deviation τ:

$$\sigma_i = \exp(\bar{\sigma} - \tau \mathrm{u}_i) \tag{5}$$

with,

$$\bar{\sigma} = -\log\left(\frac{1}{I} \sum_{i=1}^{I} \exp(-\tau \mathrm{u}_i) \right).$$

The u_i's are drawn from a truncated normal distribution at ± 1.96 to avoid numerical problems when τ is large (Fiebig et al., 2010; Sarrias and Daziano, 2017). Further, $\bar{\sigma}$ is specified as a function of the u_i's so that $\bar{\sigma}$, τ and $\boldsymbol{\beta}$ are all separately identified. The presence of the u_i term in Eqs. (2) and (4) (through σ_i) links the drop out and preference components of the joint model. We specify inverted roles for the random effects in the two submodels so that persons with high u_i values are more likely to both drop out and provide less precise choices. The random effect u_i and the random component of the utility are assumed to be independent. Following McFadden (1974), the conditional probability that respondent i selects alternative m in set s is given by Eq. (6):

$$p_{ims} = \frac{e^{V_{ims}}}{\sum_{j=1}^{M} e^{V_{ijs}}}. \tag{6}$$

Assuming that a respondent's choices are independent given the random effects, the probability of observing a sequence of choices is given by:

$$f_2(\mathbf{y}_i | \boldsymbol{\beta}, \mathrm{u}_i) = \prod_{s=1}^{\mathrm{S}_i} \prod_{m=1}^{\mathrm{M}} p_{ims}^{y_{ims}} \tag{7}$$

with $\mathrm{S}_i = \mathrm{S}$ if person i is a completer.

We further assume that the sequence of responses $\mathbf{y}_i$ and vector of indicators for missing responses $\mathbf{d}_i$ are independent conditional on the u_i. Thus, the dependence between the observed responses and the drop out behavior is induced solely by the random effects. With this, we have the joint model:

$$f(\mathbf{d}_i, \mathbf{t}_i, \mathbf{y}_i | \mathrm{u}_i, \boldsymbol{\beta}, \theta_1, \theta_2) = f_1(\mathbf{d}_i, \mathbf{t}_i | \mathrm{u}_i, \theta_1, \theta_2) f_2(\mathbf{y}_i | \mathrm{u}_i, \boldsymbol{\beta}) \tag{8}$$

The joint likelihood for all the observed data is given by

$$L(\theta_1, \theta_2, \boldsymbol{\beta}) = \prod_{i=1}^{\mathrm{I}} \int f_1(\mathbf{d}_i, \mathbf{t}_i | \mathrm{u}_i, \theta_1, \theta_2) f_2(\mathbf{y}_i | \mathrm{u}_i, \boldsymbol{\beta}) f_3(\mathrm{u}_i | \tau) d\mathrm{u}_i \tag{9}$$

where $f_3(\mathrm{u}_i | \tau)$ is a distribution of u_i random effects. Whereas the joint model in this section assumes that time is discrete corresponding to the choice sets, the next section provides a more general setting where time is assumed to be continuous but split into small intervals spent at the different choice sets. We also create an association between the drop out and choice submodels using shared covariates in the next section.

2.2.2. Piece-wise exponential drop out model and the scaled multinomial logit joint model

In the second joint model, we combine a piece-wise proportional hazards model and an SMNL model with shared covariates between them. This drop out model relaxes the unrealistic assumption in the discrete time hazard where everyone was assumed to spend the same amount of time at a task. The equal time assumption disregards differences in the evaluation of tasks by individuals and variation in choice task complexity. Specifically, we use the piece-wise exponential event model where the cumulative continuous time T_i by person i in the survey is split into S_i intervals each representing the time spent at a choice set i.e. t_{is}. Within each interval (corresponding to total time at a choice set), the risk of dropping out of the survey at that choice set is assumed to be constant and is denoted by λ_s, for $\mathrm{s} = 1, 2, .., \mathrm{S}$. This creates S parameters (pieces), that are constant at respective choice sets, which are used to model the overall baseline risk of dropping out of the survey. In this case, we assume that the hazard for person i in choice set s (denoted by λ_{is}) can be modeled as in Eq. (10).

$$\lambda_{is} = \lambda_s \exp(\mathbf{z}_i' \boldsymbol{\gamma}) \tag{10}$$

where λ_s is the baseline risk of dropping out in choice set s and $\exp(\mathbf{z}_i'\boldsymbol{\gamma})$ is the relative risk for a person with covariate vector values $\mathbf{z}_i$ compared to a reference person in any choice set. The $\boldsymbol{\gamma}$ is a vector of parameters to be estimated. When logarithms are taken on both sides of Eq. (10), we have a standard linear model where the log of the hazards at the choice sets are treated as factors:

$$\log(\lambda_{is}) = \log(\lambda_s) + \mathbf{z}_i'\boldsymbol{\gamma}. \tag{11}$$

This piece-wise exponential event model's likelihood is equivalent to a Poisson model where the drop out event indicators d_{is} are treated as if they were independent observations with means μ_{is} given by Eq. (12) (Holford, 1980; Whitehead, 1980; Laird and Olivier, 1981; Rodríguez, 2007).

$$\mu_{is} = \mathrm{t}_{is}\lambda_{is} \tag{12}$$

The coincidence in the likelihoods can be seen by taking logarithms for Eq. (12) and expanding λ_{is} using Eq. (11) to give Eq. (13). Eq. (13) is a Poisson model where the drop out indicators are the responses and the log of time spent at a choice set enters as an offset. Eq. (13) differs from Eq. (11) only in having $\log(\mathrm{t}_{is})$ which depends on the data but is independent of the λ_s parameter.

$$\log(\mu_{is}) = \log(\mathrm{t}_{is}) + \log(\lambda_s) + \mathbf{z}_i'\boldsymbol{\gamma} \tag{13}$$

Similar to fitting a discrete time hazard using the logistic model, we note that the drop out indicators d_{is} are neither assumed to be independent nor Poisson distributed per se. First, if person i drops out at choice set s, s/he must have been making choices at choice set k for $k < s$ so the drop out indicators cannot be independent. Second, the drop out indicators take only values zero or one whereas the Poisson distribution assigns probabilities to values greater than one. Therefore, the drop out indicators cannot be Poisson distributed. It is the similarity in the likelihood of the piece-wise exponential event model and that of the Poisson model that is utilized. Exploiting the equivalence in the likelihoods between the piece-wise proportional hazard and the Poisson models, person i's contribution to the drop out submodel is given by:

$$f_4(\mathbf{d}_i, \mathbf{t}_i | \lambda_s, \boldsymbol{\gamma}, \mathbf{z}_i) = \sum_{s=1}^{S_i} \mathrm{d}_{is}\log(\mu_{is}) - \sum_{s=1}^{S_i} \mu_{is} \tag{14}$$

where $\log(\mu_{is})$ is given by Eq. (13). The Poisson formulation suits better the discrete choice experiments for several reasons. First, the different choice tasks could have different levels of complexity and therefore assuming equal times spent at each of them by respondents may be unrealistic. Second, the Poisson model reduces to a discrete time hazard in an unlikely scenario where the time spent at each of the S choice sets by each respondent is the same. An illustration showing the equivalence of a discrete time hazard and Poisson model with exposure time for everyone equal to one is shown in Eqs. (A.1) and (A.2) in Appendix. Third, when recorded, the time spent answering the choice experiment self-splits at the choice tasks providing an ideal setting for fitting a Poisson regression. Fourth, at an aggregate level, the Poisson model can be used to model the risk of failing to provide a response at different choice sets even when the missingness is non-monotone.

For the SMNL model for the repeated choices where some or all terms in $\mathbf{z}_i'\boldsymbol{\gamma}$ in Eq. (13) are shared between the submodels, the expression for the deterministic part is given by

$$V_{ims} = \sigma_i \mathbf{x}_{ims}' \boldsymbol{\beta} \tag{15}$$

with,

$$\sigma_i = \exp(\bar{\sigma} + \mathbf{z}_i'\boldsymbol{\gamma})$$

$$\bar{\sigma} = -\log\left(\frac{1}{I}\sum_{i=1}^{I}\exp(\mathbf{z}_i'\boldsymbol{\gamma})\right).$$

The σ_i parameter is now an expression of shared covariate effects from $\mathbf{z}_i'\boldsymbol{\gamma}$. The probability of selecting an alternative in a choice set is similar in expression to Eq. (6) as is the probability of observing a sequence of choices to Eq. (7). We denote the latter's logarithm by $f_5(\mathbf{y}_i|\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{z}_i)$ to distinguish it from $f_2(\mathbf{y}_i|\boldsymbol{\beta},u_i)$. Assuming again that the sequence of choices and the risk of dropping out are independent conditional on the shared covariates, the joint log likelihood is now given by:

$$LL_2(\lambda_s,\boldsymbol{\gamma},\boldsymbol{\beta}) = \sum_{i=1}^{I}\Big(f_4(\mathbf{d}_i,\mathbf{t}_i|\lambda_s,\boldsymbol{\gamma},\mathbf{z}_i) + f_5(\mathbf{y}_i|\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{z}_i)\Big) \tag{16}$$

It is possible that the $\boldsymbol{\gamma}$ vector in $f_4(\mathbf{d}_i,\mathbf{t}_i|\lambda_s,\boldsymbol{\gamma},\mathbf{z}_i)$ is different from that in $f_5(\mathbf{y}_i|\boldsymbol{\beta},\boldsymbol{\gamma},\mathbf{z}_i)$. However, for illustration purposes, we assume that the effects of the covariates are the same in both submodels.

To make comparisons in each joint model, we also estimated two SMNL models without modeling the drop out component: one for all choices provided (denoted by SMNL_A) and another for completers (SMNL_C). We estimated model SMNL_A to show the inferential differences that can be associated with overlooking the drop out component. On the other hand, SMNL_C was included to show differences when analyses focus on completers. This may be interesting when a large proportion of respondents fail to finish the experiment. The expressions for V_{ims} and p_{ims} for SMNL_A are similar to those in Eqs. (4) (Eq. (15) for the second joint model) and Eq. (6). The only difference with the joint model in Eq. (8) is that the drop out process in the SMNL_A is not concurrently modeled. In addition to ignoring the drop out component, the SMNL_C model is limited to completers.

We estimated the parameters using Bayesian methods implemented in the $R2jags$ package (Su and Yajima, 2020). Vague priors were used. Each element in vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and the θs followed a $N(0,10^6)$ while the τ parameter followed a half-Cauchy distribution (Gelman, 2006) with a scale of 5. To check the sensitivity of τ to the prior used, we estimated the models with the half-Cauchy prior for τ and the Inverse Gamma$(10^{-3},10^{-3})$ prior for τ^2 on the example in Fig. 1. There were slight changes in coefficient values that were not accompanied by changes in conclusions. In using these vague priors, we provide results that are comparable to those via the likelihood method while also exploiting the flexibility offered by Bayesian methods. Three chains with random starting values were run for each model. An R-hat (Gelman and Rubin, 1992) value of less than 1.1 signified convergence. We ran 5000 draws with burn-ins of 2500 for each model. Where needed, the models were updated using the autojags function until the convergence criterion was achieved.

3. Simulation and data analysis

We conducted two simulation studies to compare the joint model to analyses that concentrate on choices without modeling the drop out process. The simulations followed the same models introduced in Section 2. A D-efficient design for website preferences, which we first described in Maaya et al. (2020), was used. We also present results from the analysis of the data presented in Fig. 1 using the joint model introduced in Section 2.2.1. We did not fit the joint model from Section 2.2.2 on this data because the time taken by respondents at the different choice sets were not recorded. Further, as the socio-demographic characteristics were collected in the last section of the survey, all participants who did not finish the choice section of the survey did not also provide this information.

3.1. Simulation results for a joint model with a discrete-time hazard and an SMNL model

The first simulation was based on the joint model combining the discrete-time hazard and the SMNL models described in Section 2.2.1. The true values for the model parameters were $\boldsymbol{\beta}$ = (−0.07, −0.15, −0.70, 0.15, 0.2, 0.80, 0.50, −0.03), τ = 0.5, θ_1 = −1.5 and θ_2 = -0.25. This simulation generates an example where the risk of dropping out of the study decreases in later choice tasks (a negative θ_2 value). There were 150 respondents each faced with 12 choice sets. Every task contained 3 alternatives described by 8 attributes. We repeated the simulation 100 times, each generating roughly 41% completers, and then averaged the resulting estimates.

We report the simulation results of the means of posterior means (Est) and means of within simulation standard errors (ASE), and the empirical standard errors (ESE) for the parameters in Table 1. The empirical error of a parameter is the standard error of its posterior means from all the simulation datasets. The two standard errors are supposed to be close when estimates are unbiased and can therefore provide a quick bias check if they are very different. Table 1 shows that estimates from the joint model followed the true values most closely. Larger discrepancies were observed for some parameters (e.g., β_3, β_6 and τ) in models where the drop out process was ignored. These results indicate that failure to account for association between the choice and drop out processes, when an association exists, can lead to unreliable estimates. The standard errors (both ASE and ESE) in the joint model were always smaller than or equal to those from models ignoring the drop out for corresponding parameters. The smaller errors show that the joint model estimated the parameters with better precision.

Table 2 shows the bias, mean squared errors (MSE) and coverage rates (CR) of the 95% credible intervals for the parameters. For corresponding parameters, the joint model outperforms the other models in terms of bias and MSEs. The higher bias for models analyzing choices only compared to the joint model reflects the benefit of maximizing information from the data to estimate the parameters. The joint model simultaneously incorporates all information from the choices and drop out to give less biased results. On the other hand, analyzing choices only deprives the estimation procedures of information from the drop out behavior resulting in more biased estimates. The coverage rates for parameters in the joint model were high, save for β_2 and β_5 which had coverage rates close to 90%. For models ignoring the drop out, the coverage rates for some parameters were very low. For instance for β_5, the coverage was 50% when using all responses and 26% when analyzing completers. Tables 1 and 2 results show that the joint model provided less biased estimates, more reliable standard errors and had better coverage compared to models that ignored the drop out process. Overall, the recovery of true parameter values improved when more information available in the data was used.

Table 1
Estimates and standard errors for the first simulation study.

		Joint model			SMNL without drop out process					
					All choices			Completers		
	True value	Est	ASE	ESE	Est	ASE	ESE	Est	ASE	ESE
β_1	−0.07	−0.07	0.03	0.03	−0.07	0.03	0.03	−0.09	0.04	0.04
β_2	−0.15	−0.16	0.02	0.02	−0.18	0.03	0.03	−0.21	0.04	0.04
β_3	−0.70	−0.72	0.14	0.14	−0.89	0.20	0.20	−0.93	0.24	0.25
β_4	0.15	0.15	0.04	0.04	0.19	0.05	0.05	0.20	0.06	0.06
β_5	0.20	0.21	0.01	0.01	0.24	0.03	0.03	0.26	0.03	0.03
β_6	0.80	0.83	0.14	0.14	0.97	0.20	0.20	1.06	0.25	0.25
β_7	0.50	0.49	0.13	0.12	0.54	0.16	0.15	0.61	0.20	0.20
β_8	−0.03	−0.03	0.01	0.01	−0.04	0.02	0.02	−0.04	0.02	0.02
τ	0.50	0.53	0.10	0.10	0.56	0.12	0.13	0.58	0.14	0.14
θ_1	−1.50	−1.53	0.16	0.17						
θ_2	−0.25	−0.25	0.04	0.05						

Note: SMNL — Scaled multinomial logit. Est — Estimate.
ASE (ESE) — means of posterior (empirical) standard error.

Table 2
Bias, MSE and coverage rates for the first simulation.

	Joint model			SMNL without drop out process					
				All choices			Completers		
	Bias	MSE	CR	Bias	MSE	CR	Bias	MSE	CR
β_1	−0.00	0.00	0.95	0.00	0.00	0.96	−0.02	0.00	0.93
β_2	−0.01	0.00	0.88	−0.03	0.00	0.84	−0.06	0.00	0.65
β_3	−0.02	0.02	0.96	−0.19	0.07	0.85	−0.23	0.11	0.82
β_4	0.00	0.00	0.98	0.04	0.00	0.93	0.05	0.01	0.91
β_5	0.01	0.00	0.89	0.04	0.00	0.50	0.06	0.00	0.26
β_6	0.03	0.02	0.96	0.17	0.07	0.86	0.26	0.13	0.83
β_7	−0.01	0.02	0.95	0.04	0.03	0.96	0.11	0.05	0.94
β_8	−0.00	0.00	0.95	−0.01	0.00	0.93	−0.01	0.00	0.97
τ	0.03	0.01	0.95	0.06	0.02	0.88	0.08	0.03	0.88
θ_1	−0.03	0.03	0.95						
θ_2	−0.00	0.00	0.94						

SMNL — Scaled multinomial logit. MSE — Mean squared error. CR — Coverage rate.

Table 3
Dropout patterns: 213 respondents.

Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Count	24	20	14	8	3	4	1	1	0	2	0	8	0	128

3.2. Analysis of the University choices data

We used the models introduced in Section 2.2.1 to analyze the university choices study mentioned in Fig. 1. This data was collected as part of a Master of Business Administration thesis by a student at the university (Arciniegas Duran, 2016). In total, there were 322 participants of which 217 participants provided at least one response in the choice experiment. Four out of the 217 respondents either chose the optout from the first to the last set or chose the optout throughout before dropping out. These were excluded from the analyses on suspicion that they did not trade alternatives in the tasks. Of the remaining 213 respondents, 128 (60%) completed the choice exercise while 85 (40%) were dropouts. The drop out patterns are shown in Table 3 where 24 persons provided a choice in the first set before dropping out, 20 persons gave two responses while 8 made 12 choices. The drop out patterns in Table 3 suggest that most dropouts were observed in the early choice sets while dropouts in the later choice sets were generally low.

Table 4 shows that there were small differences in the magnitudes of some estimates across the models. These differences did not lead to different conclusions at 5% significance level. However for β_5, an analysis at 10% significance level would be significant for completers (p-value = 0.087) but remain non-significant when all choices are used. The overall similarity in results suggests that an association between the drop out and choice behaviors was either low in this sample or a more complicated association structure was required. Further, the negative and significant θ_2 estimate reflects the patterns in Table 3 where fewer dropouts occurred towards the end of the survey.

Table 4
SMNL for University choices.

	Joint model		SMNL without drop out process			
			All choices		Completers	
	Est	SE	Est	SE	Est	SE
Optout	−4.74	0.21	−4.75	0.19	−5.09	0.20
β_1	1.04	0.04	1.10	0.04	1.15	0.05
β_2	0.63	0.04	0.67	0.04	0.68	0.04
β_3	0.52	0.06	0.57	0.07	0.42	0.07
β_4	0.22	0.06	0.24	0.07	0.20	0.07
β_5	−0.05	0.06	−0.06	0.07	−0.13	0.08
β_6	0.57	0.06	0.59	0.06	0.64	0.07
τ	0.62	0.05	0.59	0.05	0.59	0.05
θ_1	−2.15	0.15				
θ_2	−0.27	0.04				
Resp.		213		213		128

Note: SMNL — Scaled multinomial logit. Est — Estimate. SE — standard error.
Grayed out estimates have their 95% credible intervals contain zero.

Table 5
Estimates and standard errors for the second simulation study.

		Joint model			SMNL without drop out process					
					All choices			Completers		
	True value	Est	ASE	ESE	Est	ASE	ESE	Est	ASE	ESE
β_1	−0.07	−0.07	0.02	0.02	−0.08	0.03	0.03	−0.02	0.01	0.01
β_2	−0.15	−0.16	0.02	0.02	−0.18	0.03	0.03	−0.04	0.00	0.01
β_3	−0.70	−0.76	0.13	0.13	−0.84	0.18	0.20	−0.18	0.03	0.04
β_4	0.15	0.16	0.04	0.03	0.18	0.05	0.05	0.04	0.01	0.01
β_5	0.20	0.21	0.02	0.02	0.23	0.03	0.04	0.05	0.00	0.01
β_6	0.80	0.85	0.13	0.13	0.94	0.18	0.19	0.21	0.03	0.05
β_7	0.50	0.53	0.12	0.12	0.57	0.14	0.14	0.13	0.03	0.04
β_8	−0.03	−0.03	0.01	0.01	−0.04	0.01	0.02	−0.01	0.00	0.00
λ	−4.05	−4.53	0.08	0.09						
γ_1	1.50	1.55	0.06	0.06	1.60	0.09	0.11	1.53	0.12	0.13
γ_2	−1.00	−1.04	0.10	0.12	−1.08	0.13	0.15	−1.05	0.16	0.18

Note: SMNL — Scaled multinomial logit. Est — Estimate.
ASE (ESE) — means of posterior (empirical) standard Error.

3.3. Simulation results for a joint model with a piece-wise exponential survival and an SMNL model

The second simulation was based on the joint model with a piece-wise exponential event submodel and an SMNL submodel described in Section 2.2.2. Two covariate values were generated for every person consisting of a continuous variable z_1 from a standard normal distribution and a two-level categorical variable z_2. The total time, T_i for person i, taken to drop out of the survey was generated from an exponential distribution with the rate expressed as a function of the z_1 and z_2 covariates. Each individual was assumed to be under observation from the first choice set until they dropped out or had made a choice in the last task, whichever came first. The time taken between the first choice task and the last choice set answered was then split into smaller S_i intervals (i.e. the t_{is}'s) each representing duration at a choice set. For this illustration, the first S_i - 1 intervals were fixed and equal for everyone (i.e. $t_{i1} = t_{i2}... = t_{i,S_i-1} \quad \forall i$). Therefore, the total time by two persons (say i and j, $i \neq j$) who answered the same number of choice sets before dropping out only differed in their t_{iS_i} and t_{jS_j} values, for S_i equal to S_j. The true values for the model parameters were $\boldsymbol{\beta} = (-0.07, -0.15, -0.70, 0.15, 0.2, 0.80, 0.50, -0.03)$, $\lambda_s = -4.5 \ \forall s$ and $\boldsymbol{\gamma} = (1.5, -1.0)$ for effects of z_1 and z_2. There were 500 respondents each faced with 12 choice tasks. Each task had 3 alternatives described by 8 attributes as before. We generated 200 datasets, resulting in an average of 58% completers.

Table 5 shows the estimates and associated standard errors for the parameters in this simulation. In agreement with results from the first simulation, the joint model provided estimates that followed the true values the closest and were most precise. It is evident that the part-worth utilities from analyzing completers, i.e. items in the vector $\boldsymbol{\beta}$ which are primary targets in choice experiments, were poor representations of the true values. These results provide further evidence of the need to maximize information from the association between the submodels to provide more reliable results.

Table 6 shows the bias, MSE and the 95% coverage rates for the parameters in this simulation. Analogous to results in Table 2, models that ignored the drop out feature resulted in more bias, higher MSE and poorer coverage for all parameters compared to the joint model. Of greater concern here is that an analysis based on completers only completely failed to provide any substantial coverage for most of the parameters. A similar near-zero coverage for some parameters was observed in a simulation of the type

Table 6
Bias, MSE and coverage rates for the second simulation study.

	Joint model			SMNL without drop out process					
				All choices			Completers		
	Bias	MSE	CR	Bias	MSE	CR	Bias	MSE	CR
β_1	0.00	0.00	0.94	−0.01	0.00	0.92	0.05	0.00	0.00
β_2	−0.01	0.00	0.90	−0.03	0.00	0.86	0.11	0.01	0.00
β_3	−0.06	0.02	0.94	−0.14	0.05	0.86	0.52	0.27	0.00
β_4	0.01	0.00	0.96	0.03	0.00	0.93	−0.11	0.01	0.00
β_5	0.01	0.00	0.88	0.03	0.00	0.80	−0.15	0.02	0.00
β_6	0.05	0.02	0.95	0.14	0.05	0.90	−0.59	0.35	0.00
β_7	0.03	0.01	0.95	0.07	0.03	0.93	−0.37	0.14	0.00
β_8	0.00	0.00	0.95	−0.01	0.00	0.91	0.02	0.00	0.00
λ	−0.03	0.01	0.92						
γ_1	0.05	0.01	0.88	0.10	0.02	0.80	0.03	0.02	0.92
γ_2	−0.04	0.01	0.91	−0.08	0.02	0.86	−0.05	0.03	0.92

SMNL — Scaled multinomial logit. MSE — Mean squared error. CR — Coverage rate.

in Section 3.1 where θ_2 was positive (i.e. presence of fatiguing in samples) as shown in Tables 7 and 8 in Appendix. These results show that an analysis of completers only when the association structure is stronger and more complicated can be misleading and should be handled with a lot of caution.

4. Discussion and conclusion

This article focuses on two behaviors that co-occur in online surveys: the utility-driven choices and the intrinsic behavior to or not to complete a choice exercise. Whereas choices are collected directly, drop out data is a spin-off from respondents' observed choices. Joint analysis of choice and drop out behaviors may be necessary when an association is suspected. For instance, the precision of choices may be compromised in lengthy and complex surveys since they require more time and effort to make choices. In addition to affecting choices, long and complex surveys may encourage participants to stop earlier than designed due to fatigue. In such cases, there is a need to account for this association when analyzing the data. In this article, we assumed in one instance involving analyses based on a discrete-time hazard that the random effects which influenced respondents' drop out behavior also affected the scale of their choices in an SMNL model. In particular, we assumed that respondents who had higher random effect values were more likely to drop out and also provide less precise choice sequences. In a second case involving a piece-wise proportional hazard model, we assumed that the drop out and the choice submodels shared some covariate effects between them. We fit three models for each case. The first was a joint model where an association between the SMNL and drop out submodels was modeled. The second model was an SMNL on all choices provided by a respondent regardless of whether the person completed the choice exercise. Lastly, we fit an SMNL model to only the participants that completed the survey.

Simulation results showed that when an association between the drop out and choice making processes exists, a joint model provides less biased and more precise estimates. The coverages of the 95% credible intervals for the true parameter values were always higher compared to models that ignored the drop out component. We also observed that when the association structure becomes more complicated, analyzing completers only leads to unreliable estimates and very low coverage for some of the parameters. These findings agree with results from other settings like biostatistics (e.g., Tsiatis and Davidian (2004), Wu et al. (2012) and Proust-Lima et al. (2014)) where joint modeling has been associated with better coverage, higher precision and less biased estimates. Therefore, whenever an association is suspected to exist, the joint model should be prioritized because it is more beneficial.

When assessed on the university choices data, the results from the joint model based on the discrete-time hazard and models ignoring the drop out and/or limited to completers did not differ much. This may suggest that the association between the drop out and choice behaviors was low in this sample or a more refined association structure was required. The latter is especially likely given that an analysis of the drop out behavior based on the discrete time hazard implicitly assumes that all respondents took the same amount of time to answer every task. Such an assumption renders the variability in the time to a drop out event redundant and is unlikely to hold in real-life choice experiments.

To conclude, we highlight some areas related to this work that would benefit from further research. First, in this article, we used shared random effects and covariates to induce an association between the drop out and choice behaviors. While sharing covariates and random effects is popular in the joint modeling literature, there exist alternative ways to specify associations. One such is the joint latent class model which assumes that the association between the submodels is wholly captured using a latent class structure. In this case, each latent class is characterized by a class-specific risk of dropping out and class-specific utility weights for the attributes. The risk of dropping out and the choices are then assumed to be independent given the latent class. A joint latent class model may help to improve the analysis of the university choices in Table 3 by assuming, for instance, that the dropouts fall into two groups. In this case, persons in the first group can be assumed to drop out early in the survey (say up to the seventh task) while those in the other group drop out late in the survey or are completers. For more insights on latent classes in joint models, interested readers can refer to Proust-Lima et al. (2014).

Whereas the simulation study showed that modeling an existing association between the choice and drop out behaviors leads to less biased and more precise estimates, our data example did not benefit much from these advantages. This could have been due to a very low association between the choices and the drop out behavior in the sample or the data lacked key elements for investigating the two behaviors together. First, we encourage researchers with data where an association is highly suspected to consider modeling the behaviors jointly so that more evidence on this topic is available. Second, we encourage researchers to consider recording time taken by respondents at every choice set in their data collection. This can be done by specifying timestamps in popular data collection tools like (Qualtrics, 2018) and Mechanical Turk (Mason and Suri, 2012). Such a richer dataset can allow extensive modeling of the drop out behavior using approaches like the piece-wise exponential event model described in Section 2.2.2. Third, we encourage researchers to creatively collect data such that there is more opportunity for respondents who do not complete the choice survey to provide their socio-demographic data. In many studies, socio-demographic data is collected at the end of surveys and therefore those who drop out do not have these data captured. If recorded, demographic data can be key to understanding the profiles of dropouts and the association between the preference and drop out behavior.

Similar to usual choice models, the joint model can be extended to accommodate more heterogeneity using a mixed multinomial logit in place of the SMNL model. The modeler, however, needs to carefully specify an association between the drop out and choice submodels. This should be done bearing in mind that separating coefficient and scale heterogeneity in mixed logit models may be infeasible (Hess and Rose, 2012; Hess and Train, 2017; Mariel and Artabe, 2020). It is noteworthy that the use of joint models in choice experiments, like in other research areas, requires researchers to write their own functions since they are yet to be implemented in available packages.

Further extensions to the joint model to handle non-monotone missingness in the choices, informative censoring and correct measurement errors in recorded time for the drop out submodel may be interesting. Non-monotone missingness occurs in experiments where respondents do not have to provide a choice before accessing succeeding tasks. In this case, there will be multiple tasks with missing choices. One option to handle this type of missingness is to use the piece-wise exponential event model via its equivalent aggregate-level Poisson model to model the rate of missing choices at every task. Another approach could be to assume that every person can have multiple drop out events occurring at tasks where s/he does not provide a choice. These multiple events can then be modeled as recurrent drop out events (see Amorim and Cai, 2015 for related literature).

We note that the risk of dropping out in a survey is not always independent of the censoring mechanism. This can be the case in online surveys where some respondents are lost because of technical reasons like time outs. In Mechanical Turk (Mason and Suri, 2012) for instance, *requesters* can create timers for their tasks so that *workers* respond to them within a specified period of time. If the allotted time is exhausted before a choice is made by a *worker*, there is a missing response for that task and possibly in all succeeding tasks. Qualtrics (2018) can also automatically log out users after an hour or a specified period of inactivity. In these cases, an incomplete sequence of responses by participants could be because they actually dropped out of the study or were preoccupied and exhausted their expected time to complete the choice exercise. Time outs in such settings hinder the observation of drop out events and can be considered as competing risks. Any person that experiences such a competing risk is censored in an informative way. Extensions to the joint models are then required so that an additional parameter is included to capture the association between the time to censoring and the time to a drop out event. Alternatively, an extra submodel can be introduced in the joint model to cater for the time to censoring (Leung et al., 1997).

CRediT authorship contribution statement

Leonard Maaya: Contributed in the brooding of the idea, Analyses, Writing and reviewing of the manuscript, Taken a leading role. **Michel Meulders:** Contributed in the brooding of the idea, Analyses, Writing and reviewing of the manuscript, Supervision. **Martina Vandebroek:** Contributed in the brooding of the idea, Analyses, Writing and reviewing of the manuscript, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Alvaro Andres Arciniegas Duran who, under the supervision of Michel Meulders, designed and collected the data for the university career choices data for his M.Sc. thesis.

Leonard Maaya was funded by project G0C7317N of the Flemish Research Foundation (FWO Flanders), Belgium.

Appendix

Equivalent estimates from a discrete time hazard and a piece-wise proportional hazard

In this section, we provide a simplified procedure showing that the hazard estimate from a piece-wise proportional drop out model and that from a discrete-time hazard will be equal when it is assumed that the time spent at a choice set is the same for every person. We show for the case where the risk is assumed to be the same for everyone (i.e $\lambda_{is} = \lambda$ forall i and s) and thus suppress the effects of individual characteristics.

From the likelihood contribution for person i in the discrete-time hazard formulation in Eq. (3), we have:

$$
\begin{aligned}
f_1(\mathbf{d}_i, \mathbf{t}_i | \lambda_{is}) &= \prod_{s=1}^{\mathrm{S}_i} \lambda_{is}^{\mathrm{d}_{is}} (1-\lambda_{is})^{1-\mathrm{d}_{is}} \\
&= \prod_{s=1}^{\mathrm{S}_i} \lambda^{\mathrm{d}_{is}} (1-\lambda)^{1-\mathrm{d}_{is}}
\end{aligned}
$$

where S_i equal to S for completers. We transform to the logarithm scale and recognize that for person i, all the d_{is} indicators equal zero for all s if s/he is a completer. For a dropout, the first S_i-1 d_{is} indicators equal zero and only d_{iS_i} equals 1. We therefore have,

$$
\begin{aligned}
l_{i1} &= \sum_{s=1}^{\mathrm{S}_i} \Big(\mathrm{d}_{is} log(\lambda) + (1-\mathrm{d}_{is}) log(1-\lambda) \Big) \\
&= \mathrm{d}_{iS_i} log(\lambda) + (1-\mathrm{d}_{iS_i}) log(1-\lambda) + (\mathrm{S}_i - 1) log(1-\lambda).
\end{aligned}
$$

With this, the total loglikelihood is given by;

$$
\begin{aligned}
l_1 &= \sum_{i=1}^{\mathrm{I}} \Big(\mathrm{d}_{iS_i} log(\lambda) + (1-\mathrm{d}_{iS_i}) log(1-\lambda) + (\mathrm{S}_i - 1) log(1-\lambda) \Big) \\
&= log(\lambda) \sum_{i=1}^{\mathrm{I}} \mathrm{d}_{iS_i} + log(1-\lambda) \sum_{i=1}^{\mathrm{I}} (1-\mathrm{d}_{iS_i}) + log(1-\lambda) \sum_{i=1}^{\mathrm{I}} (\mathrm{S}_i - 1).
\end{aligned}
$$

Through a few differentiation steps with respect to λ, its maximum likelihood estimate ($\hat{\lambda}$) can be shown to be equal to the ratio of the total number of dropouts and the total number of answered tasks, i.e.

$$
\hat{\lambda} = \frac{\sum_{i=1}^{\mathrm{I}} \mathrm{d}_{iS_i}}{\sum_{i=1}^{\mathrm{I}} \mathrm{S}_i}. \tag{A.1}
$$

We turn to the piece-wise proportional hazard model which was re-expressed as a Poisson formulation in Eq. (14) and is reproduced below,

$$
f_4(\mathbf{d}_i, \mathbf{t}_i | \lambda_s, \boldsymbol{\gamma}, \mathbf{z}_i) = \sum_{s=1}^{S_i} \mathrm{d}_{is} \log(\mu_{is}) - \sum_{s=1}^{S_i} \mu_{is}
$$

where,

$$
\log(\mu_{is}) = log(\mathrm{t}_{is}) + log(\lambda_{is})
$$

we expand Eq. (14), so that we have

$$
l_{i2} = f_4(\mathbf{d}_i, \mathbf{t}_i | \lambda_s, \boldsymbol{\gamma}, \mathbf{z}_i) = \sum_{s=1}^{S_i} \mathrm{d}_{is} \log(t_{is}) + \sum_{s=1}^{S_i} \mathrm{d}_{is} \log(\lambda_{is}) - \sum_{s=1}^{S_i} \mathrm{t}_{is} \lambda_{is}.
$$

With $\lambda_{is} = \lambda$ and $\mathrm{t}_{is} = 1 \ \forall \ (i, s)$ so that $\mathrm{T}_i = \mathrm{S}_i$, the first term drops out because log(1) is zero. Further, only the d_{iS_i} indicators that equal one will contribute to the likelihood function. We therefore have;

$$
\begin{aligned}
l_{i2} &= \log(\lambda) \sum_{s=1}^{S_i} \mathrm{d}_{is} - \lambda S_i \\
&= \log(\lambda) \mathrm{d}_{iS_i} - \lambda S_i \\
l_2 &= \log(\lambda) \sum_{i=1}^{I} \mathrm{d}_{iS_i} - \lambda \sum_{i=1}^{I} S_i.
\end{aligned}
$$

By solving the above equation for λ, the maximum likelihood estimate is found to be same as in Eq. (A.1);

$$
\hat{\lambda} = \frac{\sum_{i=1}^{\mathrm{I}} \mathrm{d}_{iS_i}}{\sum_{i=1}^{\mathrm{I}} \mathrm{S}_i}. \tag{A.2}
$$

Table 7
Estimates and standard errors for the simulation study with a positive θ_2.

		Joint model			SMNL without drop out process					
					All choices			Completers		
	True value	Est	ASE	ESE	Est	ASE	ESE	Est	ASE	ESE
β_1	−0.07	−0.07	0.03	0.03	−0.07	0.03	0.04	−0.55	0.48	0.68
β_2	−0.15	−0.16	0.02	0.02	−0.18	0.03	0.04	−1.68	0.60	1.22
β_3	−0.70	−0.75	0.15	0.16	−0.92	0.22	0.27	−7.18	3.04	5.44
β_4	0.15	0.16	0.04	0.04	0.19	0.05	0.06	1.36	0.70	1.08
β_5	0.20	0.21	0.02	0.02	0.24	0.03	0.04	1.82	0.57	1.19
β_6	0.80	0.83	0.15	0.15	0.96	0.21	0.23	7.23	3.08	5.44
β_7	0.50	0.52	0.14	0.14	0.56	0.17	0.18	4.24	2.37	3.56
β_8	−0.03	−0.03	0.01	0.02	−0.04	0.02	0.02	−0.23	0.23	0.36
τ	0.80	0.87	0.11	0.11	0.90	0.13	0.15	1.61	0.27	0.60
θ_1	−4.00	−4.09	0.26	0.28						
θ_2	0.30	0.31	0.04	0.04						

Note: SMNL — Scaled multinomial logit. Est — Estimate.
ASE (ESE) — means of posterior (empirical) Std Error.

Table 8
Bias, MSE and coverage rates for a simulation study with a positive θ_2.

	Joint model			SMNL without drop out process					
				All choices			Completers		
	Bias	MSE	CR	Bias	MSE	CR	Bias	MSE	CR
β_1	−0.00	0.00	0.92	0.00	0.00	0.92	−0.48	0.47	0.81
β_2	−0.01	0.00	0.94	−0.03	0.00	0.86	−1.53	2.69	0.18
β_3	−0.05	0.03	0.93	−0.22	0.10	0.80	−6.48	51.19	0.32
β_4	0.01	0.00	0.94	0.04	0.00	0.89	1.21	1.94	0.58
β_5	0.01	0.00	0.88	0.04	0.00	0.55	1.62	2.93	0.00
β_6	0.03	0.02	0.95	0.16	0.07	0.89	6.43	50.81	0.32
β_7	0.02	0.02	0.95	0.06	0.03	0.95	3.74	19.63	0.60
β_8	0.00	0.00	0.92	−0.01	0.00	0.91	−0.20	0.09	0.84
τ	0.07	0.02	0.90	0.10	0.03	0.89	0.81	0.72	0.39
θ_1	−0.09	0.08	0.92						
θ_2	0.01	0.00	0.95						

SMNL — Scaled multinomial logit. MSE — Mean squared error. CR — Coverage rate.

Simulation with a positive θ_2

See Tables 7 and 8

References

Amorim, L.D., Cai, J., 2015. Modelling recurrent events: A tutorial for analysis in epidemiology. Int. J. Epidemiol. 44 (1), 324–333.
Arciniegas Duran, A., 2016. Assessing the Factors that Influence Students' University Choice within the Flemish Higher Education in Belgium. KU Leuven. Faculteit Economie en Bedrijfswetenschappen, Leuven.
Cox, D.R., 1972. Regression models and life-tables. J. R. Stat. Soc. B 34 (2), 187–220.
Czajkowski, M., Giergiczny, M., Greene, W.H., 2014. Learning and fatigue effects revisited: Investigating the effects of accounting for unobservable preference and scale heterogeneity. Land Econom. 90 (2), 324–351.
Danthurebandara, V.M., Vandebroek, M., Yu, J., 2015. Designing choice experiments by optimizing the complexity level to individual abilities. Quant. Mark. Econ. 13 (1), 1–26.
Efron, B., 1988. Logistic regression, survival analysis, and the kaplan-meier curve. J. Amer. Statist. Assoc. 83 (402), 414–425.
Fiebig, D.G., Keane, M.P., Louviere, J., Wasi, N., 2010. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. Mark. Sci. 29 (3), 393–421.
Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. 1 (3), 515–534.
Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. Statist. Sci. 7 (4), 457–472.
Hess, S., Hensher, D.A., Daly, A., 2012. Not bored yet - Revisiting respondent fatigue in stated choice experiments. Transp. Res. A 46 (3), 626–644.
Hess, S., Rose, J., 2012. Can scale and coefficient heterogeneity be separated in random coefficients models? Transportation 39 (6), 1225–1239.
Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. J. Choice Model. 23, 1–8.
Holford, T.R., 1980. The analysis of rates and of survivorship using log-linear models. Biometrics 36 (2), 299–305.
Irannezhad, E., Prato, C., Hickman, M., 2019. A joint hybrid model of the choices of container terminals and of dwell time. Transp. Res. E 121, 119–133.
Laird, N., Olivier, D., 1981. Covariance analysis of censored survival data using log-linear analysis techniques. J. Amer. Statist. Assoc. 76 (374), 231–240.
Leung, K.-M., Elashoff, R.M., Afifi, A.A., 1997. Censoring issues in survival analysis. Annu. Rev. Public Health 18 (1), 83–104, PMID: 9143713.

Maaya, L., Meulders, M., Vandebroek, M., 2020. Online consumers' attribute non-attendance behavior: Effects of information provision. Int. J. Electron. Commer. 24 (03), 338–365.

Mariel, P., Artabe, A., 2020. Interpreting correlated random parameters in choice experiments. J. Environ. Econ. Manage. 103, 102363.

Mason, W., Suri, S., 2012. Conducting behavioral research on amazon's mechanical turk. Behav. Res. Methods 44 (1), 1–23.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zare, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

Molenberghs, G., Verbeke, G., 2005. Models for Discrete Longitudinal Data. In: Springer Series in Statistics, Springer New York, New York, NY.

Proust-Lima, C., Sene, M., Taylor, J.M., Jacqmin-Gadda, H., 2014. Joint latent class models for longitudinal and time-to-event data: A review. Stat. Methods Med. Res. 23 (1), 74–90.

Qian, Y., Xie, H., 2011. No customer left behind: A distribution-free bayesian approach to accounting for missing xs in marketing models. Mark. Sci. 30 (4), 717–736.

Qualtrics, 2018. Qualtrics. https://www.qualtrics.com/uk/.

Rodríguez, G., 2007. Lecture notes on generalized linear models. https://data.princeton.edu/wws509/notes/. Accessed: 2021-04-28.

Rose, J.M., Bliemer, M.C.J., 2009. Constructing efficient stated choice experimental designs. Transp. Rev.: Obs. Complex Choice Behav. Stated Prefer. Exp.: Innov. Des. 29 (5), 587–617.

Sanko, N., Hess, S., Dumont, J., Daly, A., 2014. Contrasting imputation with a latent variable approach to dealing with missing income in choice models. J. Choice Model. 12, 47–57.

Sarrias, M., Daziano, R., 2017. Multinomial logit models with continuous and discrete individual heterogeneity in R: The gmnl package. J. Stat. Softw. 79 (2), 1–46.

Singer, J.D., Willett, J.B., Willett, J.B., et al., 2003. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press.

Su, Y.-S., Yajima, M., 2020. Using R to Run 'JAGS'. R package version 0.6-1.

Train, K.E., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press.

Tsiatis, A.A., Davidian, M., 2004. Joint modeling of longitudinal and time-to-event data: An overview. Statist. Sinica 14 (3), 809–834.

Varotto, S.F., Glerum, A., Stathopoulos, A., Bierlaire, M., Longo, G., 2017. Mitigating the impact of errors in travel time reporting on mode choice modelling. J. Transp. Geogr. 62, 236–246.

Whitehead, J., 1980. Fitting cox's regression model to survival data using GLIM. J. R. Stat. Soc. C 29 (3), 268–275.

Wu, L., Liu, W., Yi, G.Y., Huang, Y., 2012. Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. J. Probab. Stat. 2012, 1–17.