Latent class modeling with a time-to-event distal outcome: A comparison of one, two and three-step approaches

Daniel T. Lythgoe[ab*], Marta Garcia-Fiñana[b] and Trevor F. Cox[bc]

[a]PHASTAR, Chiswick, United Kingdom; [b]Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; [c]Liverpool Cancer Trials Unit, University of Liverpool, Liverpool, United Kingdom.

*dlythgoe@liverpool.ac.uk

Acknowledgements

We would like to thank two reviewers for their helpful comments on an earlier draft. We would also like Prof. John Neoptolemos for permission to use the ESPAC-3 data set and Dr Ian Smith for help and assistance with use of the high throughput Condor system at the University of Liverpool (http://condor.liv.ac.uk/).

Latent class modeling with a time-to-event distal outcome: A comparison of one, two and three-step approaches

Latent class methods can be used to identify unobserved subgroups which differ in their observed data. Researchers are often interested in outcomes for the identified subgroups and in some disciplines time-to-event outcome measures are common, e.g. overall survival in oncology. In this study Monte Carlo simulation is used to evaluate the empirical properties of latent class effect estimates on a time-to-event distal outcome using one, two and three-step approaches. Both standard and inclusive bias-corrected three-step approaches are considered. One-step latent class effect estimates are shown to be superior to the evaluated alternatives. Both the two-step approach and a standard three-step approach, where subjects are partially assigned to latent classes, produced unbiased estimates with nominal confidence interval coverage when latent classes were well separated, but not otherwise.

Latent class methods encompass a broad range of models which can be used to identify and characterise unobserved subgroups which differ in their observed or 'manifest' data. These models have been widely applied in many scientific disciplines including medicine (e.g. Downing et al., 2010; Rahbar et al., 2015), social and behavioural science (e.g. Chung et al., 2006; Stapinski et al., 2016) and education (e.g. Denson and Ing, 2014; Auer et al., 2016). For example, Stapinski et al. (2016) used latent class analysis (LCA) of a large cohort study to identify four groups of adolescents who differed in their motives for alcohol use.

A common objective of LCA is to assess the relationship between the identified latent classes and a distal outcome variable. In some disciplines, time-to-event variables are common outcome measures, for example, overall and progression-free survival times in oncology. Time-to-event data differs from other data types since it is typically highly right-skewed and subject to censoring (see e.g. Collett, 2015). Applications of various latent class models with a time-to-event distal outcome can be found in Snuderl et al. (2008), Muthén et al. (2009), Zhang and Wang (2010), Desantis et al. (2012) and Leigh et al. (2015).

When model assumptions are met a preferred statistical approach is to jointly model the latent classes and distal outcome in one step (Bakk et al., 2013). Larsen (2004) introduced a one-step latent class model with a time-to-event distal outcome variable and a framework for continuous time latent class models was set out by Asparouhov et al. (2006).

A general criticism of one-step approaches, however, is that the distal outcome variable can influence the composition of the latent classes (Vermunt, 2010; Asparouhov and Muthén, 2014). Moreover, one-step approaches may be impractical if there are many distal outcome variables, or if the outcome data is collected at a different stage of a trial and/or by different researchers (Vermunt, 2010).

A simple and frequently applied alternative approach to incorporating a distal outcome variable into a latent class model is the 'Classify-Analyze' (Clogg, 2013) or 'standard three-step approach': Step 1) a latent class model is fitted, Step 2) subjects are assigned to a latent class, and Step 3) the distal outcome is regressed on the assigned class. Whilst intuitive, the standard three-step approach has two important drawbacks. Firstly, estimates of the relationship between latent class and the distal outcome variable can be attenuated due to misclassification in Step 2 (Bolck et al., 2004). Secondly, standard errors in Step 3 can be underestimated since class is treated as known in the regression model, potentially misleading statistical inference (Clark and Muthén, 2009). Bray et al. (2015) identified that non-inclusion of the distal outcome variable in the classification model (Step 1) as a further cause of bias in Step 3 and proposed an 'inclusive' approach to correct for this bias, where the distal *outcome* variable is included as a latent class *predictor* variable in Step 1, along with other covariates.

Bakk and Kuha (2017) proposed a two-step alternative to address the aforementioned issues with one and standard three-step approaches. In this approach, a latent class model is fitted in Step 1, as in the three-step approach. Then, in Step 2, the full joint latent class and distal outcome model is fitted, as in the one-step model, but the parameters for the latent class part of the model are held fixed at their estimates from Step 1. A correction is then applied to account for additional uncertainty in the second step.

Research into estimating the effect of latent class on distal outcomes has so far been restricted to categorical or continuous outcome variables (Clark and Muthén, 2009; Bakk et al., 2013; Lanza et al., 2013; Asparouhov and Muthén, 2014; Bray et al., 2015; Bakk and Vermunt, 2016; Collier and Leite, 2017; Bakk and Kuha, 2017). In this study Monte Carlo simulation is used to compare one, two and three-step approaches to latent class modeling with a time-to-event distal outcome. For the one and two-step approaches joint latent class models with piecewise constant baseline hazard functions are used (Asparouhov et al., 2006; Muthén et al., 2009). For the three-step models, four approaches to class assignment are compared and the impact of the inclusive approach for bias-correction (Bray et al., 2015) with a time-to-event distal outcome variable is assessed.

LATENT CLASS MODELING WITH A TIME-TO-EVENT DISTAL OUTCOME

The one-step approach

In this section latent class models are introduced and a full one-step latent class model with a time-to-event distal outcome, as introduced by Larsen (2004), is developed.

The latent class model

LCA was introduced by Lazarsfeld in 1950 and is used to identify and characterise unobserved and mutually exclusive subgroups using multiple imperfect indicators known as manifest variables. The basic latent class model is depicted in Figure 1(a). The latent class variable, C, is assumed to consist of J categories with prevalences

$$P(C = j) = \eta_j,$$

for j $= 1, \ldots, J$ and $\sum_{j=1}^{J} \eta_j = 1$. Let $\mathbf{Y} = (Y_1 \ldots, Y_M)^\intercal$ denote a vector of manifest variables with observed values $\mathbf{y} = (y_1, \ldots, y_M)^\intercal$ for a given subject. Typically each Y_m

$(m = 1, \ldots, M)$ is categorical with $g = 1, \ldots, G_m$ categories, so that the probability of observing category g on the mth manifest variable for subjects in the jth class is given by

$$P(Y_m = g|C = j) = \pi_{mgj} = \prod_{g=1}^{G_m} \pi_{mgj}^{I\{y_m = g\}},$$

where $\sum_{g=1}^{G_m} \pi_{mgj} = 1$, and $I\{y_m = g\}$ is an indicator function which equals 1 if y_m takes the value g and 0 otherwise, for a given subject. Normal, Poisson, binomial, gamma and ordinal categorical are some other possibilities for the conditional distribution of the manifest variables (Moustaki, 1996; Bartholomew et al., 2011). The distribution of the responses for an individual is given by

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \sum_{j=1}^{J} P(C = j) f_{\mathbf{Y}|C}(\mathbf{y}|j) \\ &= \sum_{j=1}^{J} \eta_j \prod_{m=1}^{M} f_{Y_m|C}(y_m|j), \end{aligned} \tag{1}$$

where the manifest variables are assumed to be independent conditional on class and $f(.)$ is used to denote a probability density or mass function as required. Some options for introducing dependencies between manifest variables are discussed in Hunt and Jorgensen (1999) and Desantis et al. (2012). The posterior probability that a subject belongs to class j given $\mathbf{Y} = \mathbf{y}$ is obtained using Bayes theorem, so that

$$P(C = j|\mathbf{Y} = \mathbf{y}) = \frac{\eta_j f_{\mathbf{Y}|C}(\mathbf{y}|j)}{\sum_{k=1}^{J} \eta_k f_{\mathbf{Y}|C}(\mathbf{y}|k)}. \tag{2}$$

Latent class regression

A natural extension to the latent class model (Equation 1) is the concomitant-variable or 'latent class regression' model (Dayton and Macready, 1988; Formann, 1992; van der

Heijden et al., 1996; Bandeen-Roche et al., 1997; Chung et al., 2006), as depicted in Figure 1(b). In the LCR model the class prevalences, η_j, are allowed to vary as a function of a vector of 'latent class predictors' $\mathbf{X}$, with observed values $\mathbf{x}$. Following on from Equation 1 the distribution function for a given subject is

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \sum_{j=1}^{J} \eta_j(\mathbf{x}) \prod_{m=1}^{M} f_{Y_m|C}(y_m|j), \tag{3}$$

where the latent class predictors and manifest variables are assumed to be conditionally independent given latent class. Huang and Bandeen-Roche (2004) showed how dependencies between latent class predictors and categorical manifest variables can be added to the model. A generalised linear model with a logit link function is used to model the relationship between the latent class predictors and class prevalences, so that the inverse of the logit link function is

$$P(C = j|\mathbf{X} = \mathbf{x}) = \eta_j(\mathbf{x}) = \frac{\exp(\mathbf{x}^\intercal \boldsymbol{\kappa}_j)}{\sum_{k=1}^{J} \exp(\mathbf{x}^\intercal \boldsymbol{\kappa}_k)}, \tag{4}$$

for $j = 1, \ldots, J$ and where $\boldsymbol{\kappa}_j$ is a vector of log odds ratios for the jth class, $\boldsymbol{\kappa}_J = \mathbf{0}$ for identifiability and the first element of $\mathbf{x}$ is set to 1 in order to include an intercept. An intercept only LCR model is equivalent to the latent class model. Other suitable link functions can be used. To obtain posterior probabilities Equation 2 is updated to $P(C = j|\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{X} = \boldsymbol{x})$.

A time-to-event distal outcome model

Time-to-event (survival) data is typically highly right-skewed and subject to censoring, since the event of interest is not always observed (see e.g. Collett, 2015). Note that we assume throughout that censoring is non-informative. Important functions for time-to-event data are the hazard function, $\alpha(t)$, the instantaneous event rate, cumulative hazard function, $H(t) = \int_0^t \alpha(v)dv$, and survival function, $S(t) = \exp[-H(t)] = P(T > t)$.

Larsen (2004) extended the latent class model to include a time-to-event submodel in order to model latent class and a time-to-event distal outcome in one step, as depicted in Figure 1(d). An extensive framework for these continuous time one-step or 'joint' models is presented in Asparouhov et al. (2006) and software functionality is available in M-Plus (Muthén and Muthén, 2011). One option for the time-to-event submodel is a proportional hazards model extended to include a latent class effect

$$\alpha(t|\mathbf{Z} = \mathbf{z}, C = j) = \alpha_0(t) \exp(\mathbf{z}^\intercal \boldsymbol{\beta} + \gamma_j), \quad (5)$$

for $j = 1, \ldots, J$ where $\alpha(t| \ldots)$ represents the hazard for a given subject at time t, $\alpha_0(t)$ is the baseline hazard at time t, $\boldsymbol{\beta}$ is a vector of log hazard ratios for the corresponding covariates $\mathbf{z}$ and γ_j represents the log hazard ratio for the effect of latent class j on the baseline hazard, with $\gamma_J = 0$ for identifiability. In this model both the covariate and class effects are assumed to act proportionally on the baseline hazard and independently of time. Options for assessing the suitability of the proportionality assumption are discussed later.

A useful approach to modeling the baseline hazard function is the the piecewise exponential model (Friedman, 1982), where the baseline hazard function is assumed to be piecewise constant. For a piecewise exponential time-to-event submodel, let time be partitioned into $s = 1, \ldots, S$ intervals and let $\boldsymbol{\alpha_0} = (\alpha_{01}, \ldots, \alpha_{0S})$ denote a vector of baseline hazard parameters. To complete the required notation, let T and Δ represent the event-time and censoring indicator with observed values t and δ and where δ equals

1 if the event is observed and 0 otherwise. The required density function of the event time for a given subject is

$$f_{T,\Delta|\mathbf{Z},C}(t,\delta|\mathbf{z},j) = \prod_{s=1}^{S} [\alpha_{0s} \exp(\mathbf{z}^{\intercal}\boldsymbol{\beta} + \gamma_j)]^{\delta\psi_s} \times$$
$$\exp\Bigg\{ - \psi_s \left[\alpha_{0s}(t - a_{s-1}) + \sum_{h=1}^{s-1} \alpha_{0h}(a_h - a_{h-1}) \right] \exp(\mathbf{z}^{\intercal}\boldsymbol{\beta} + \gamma_j) \Bigg\},$$

for $j = 1, \ldots, J$ and where ψ_s denotes an indicator variable which equals 1 if the event occurs in the sth interval and 0 otherwise, a_s denotes the upper boundary for the sth interval on the time grid and a_0 equals 0. The joint density for the manifest variables and time-to-event distal outcome for a given subject is then

$$f_{\mathbf{Y},T,\Delta|\mathbf{Z}}(\mathbf{y},t,\delta|\mathbf{z}) = \sum_{j=1}^{J} \eta_j \prod_{m=1}^{M} f_{Y_m|C}(y_m|j) f_{T,\Delta|\mathbf{Z},C}(t,\delta|\mathbf{z},j), \tag{6}$$

where the distributions of the manifest variables and time-to-event distal outcome are assumed to be conditionally independent given class. Latent class predictors can also be included, as in Equation 4 and detailed in Larsen (2004), but this feature is not considered for one-step models in this article. The log likelihood of the observed data is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \left[f_{\mathbf{Y}_i,T_i,\Delta_i|\mathbf{Z}_i}(\mathbf{y}_i,t_i,\delta_i|\mathbf{z}_i) \right], \tag{7}$$

where N is the total number of subjects indexed by i and $\boldsymbol{\theta} = (\boldsymbol{\eta}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ denotes the full vector of parameters to be estimated.

The two-step approach

For the two-step approach of Bakk and Kuha (2017), the required parameters from the one-step model are partitioned into those to be estimated in Steps 1 and 2 so that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. In the first step a latent class model is fitted (Equation 1) so that $\boldsymbol{\theta}_1 = (\boldsymbol{\eta}, \boldsymbol{\pi})$ and therefore $\boldsymbol{\theta}_2 = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$. Let $\tilde{\boldsymbol{\theta}}_1$ denote the estimates from Step 1 and then in Step 2 maximise the log-likelihood of the observed data conditional on the Step 1 estimates, i.e. $\ell(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1 = \tilde{\boldsymbol{\theta}}_1)$. The required log-likelihood is given in Equation 7.

Clearly the uncertainty of the estimates obtained in Step 2, $\tilde{\boldsymbol{\theta}}_2$ will be underestimated since the Step 1 parameters have been held fixed during estimation. Xue and Bandeen-Roche (2004) and Bakk and Kuha (2017) demonstrate how to obtain corrected standard errors in the two-step approach.

Standard and inclusive three-step approaches

Three-step approaches proceed as follows: Step 1) a latent class model is fitted, Step 2) an assignment rule is used to classify subjects according to their class conditional posterior probabilities, Step 3) the assigned classes are used as a covariate in a regression model to estimate the relationship between the latent classes and the external variable. These steps are now considered in more detail in the context of modeling with a time-to-event distal outcome.

Step 1: Fit the latent class model

For Step 1 in a standard three-step approach a latent class model is simply fitted, as in Equation 1 and depicted in Figure 1(a). In an inclusive three-step approach an LCR model is fitted with the distal *outcome* variable as a latent class *predictor* variable, along with other covariates related to the outcome, as depicted in Figure 1(c). How then might we incorporate an event time subject to censoring as a latent class predictor? For the purposes of multiple imputation of baseline covariate data in proportional hazards models, White and Royston (2009) recommended using the estimated cumulative hazard function (notably in preference to the observed survival time or its natural logarithm), the event indicator and other covariates related to the event time in the model. Expressing Equation 4 in logit form and replacing $\mathbf{x}$ with the required elements the inclusive model is given by

$$\operatorname{logit} P(C = j|H(t), \Delta = \delta, \mathbf{Z} = \mathbf{z}) = \kappa_{j0} + \kappa_{j1}H(t) + \kappa_{j2}\delta + \kappa_{j3}z,$$

for $j = 1, \ldots, J$ and where $H(t)$ is the (non-parametric) Nelson-Aalen estimate of the unconditional cumulative hazard, which is estimated separately. For illustration purposes only a single covariate, z, has been included but additional covariates can be incorporated easily.

Step 2: Class assignment

In Step 2, subjects are assigned to a latent class according to an assignment rule. The simplest and most commonly used assignment rule is modal assignment (MA) in which each subject is assigned to the latent class for which they have the highest posterior probability. MA ensures that all subjects with the same response pattern are allocated to the same class.

Another commonly used method is random assignment, also known as the 'pseudo class' method (PC). For PC, class is imputed once for each subject by randomly drawing from a multinomial distribution with probabilities equal to the subject's posterior probabilities from the latent class model (Bolck et al., 2004; Bandeen-Roche et al., 1997). Consequently not all subjects with the same response pattern are guaranteed to be assigned to the same class. Wang et al. (2005) introduced multiple pseudo class draws (mPC) to improve estimation efficiency over a single random draw. With mPC, class is imputed multiple times for each subject, with the authors recommending at least 20 random draws. Note that mPC is distinct from multiple imputation since the estimated posterior probabilities are effectively treated as known (Wang et al., 2005).

Finally, we highlight partial assignment (PA) and proportional assignment (PrA). In these methods, each subject is assigned partially rather than absolutely to a latent class. In PA, no assignment is made and posterior probabilities are used in further analyses. In PrA each subject is assigned to all classes simultaneously with case-weights equal to their corresponding class-specific probabilities, and as a result each subject will enter any further analyses J times.

Step 3: Estimate the effect of latent class on the distal outcome

In Step 3, the distal outcome variable is regressed on the assigned class from Step 2, possibly in addition to other relevant covariates. For a time-to-event distal outcome the Cox proportional hazards model (Cox, 1972) is a natural choice and is utilised in the subsequent simulation study.

For MA and PC, $J-1$ dummy variables are used to represent the assigned/imputed class in the regression model. For mPC this process is repeated for each class imputation and parameter estimates are combined across regression models using Rubin's rules (Rubin, 2004). For PA, $J-1$ posterior probabilities are included as covariates in the regression model. For PrA each subject is included in the regression model J times with case-weights equal to the posterior probabilities from the latent class model. One consequence of PrA in a time-to-event setting is that tied event times are introduced.

Entropy

The extent to which latent classes can be distinguished by the data and the latent class model can be assessed using the principle of entropy (Muthén and Muthén, 2004; Bakk et al., 2013). The Ramaswamy entropy statistic (Ramaswamy et al., 1993; Muthén and Muthén, 2004; Dziak et al., 2014) is defined as

$$E = 1 - \frac{\sum_{i=1}^{N} \sum_{j=1}^{J} -\hat{p}_{ij} \log(\hat{p}_{ij})}{N \log(J)},$$

for a sample of $i = 1, \ldots, N$ subjects and where $\hat{p}_{ij}$ is the estimated posterior probability of the ith subject belonging to class j from a latent class model. E can take values between 0 and 1, where 0 indicates that the model contains no information on class assignment and 1 indicates that all subjects are estimated to belong to a class with 100% probability. Lower entropy implies that classes are less well distinguished and corresponds to greater classification error being introduced in Step 2 for three-step methods. Note that this is contrary to classical entropy measures for which low values correspond to better classification (Dziak et al., 2014).

MONTE CARLO SIMULATION STUDY

Aims

The purpose of this Monte Carlo simulation study was to investigate the empirical properties of latent class effect estimates on a time-to-event distal outcome using a number of different models and simulated scenarios. In particular, we aimed to compare one, two, standard three and inclusive three-step approaches. For both the standard and inclusive three-step approaches subjects were assigned to classes using MA, mPC, PA and PrA.

Software

Data were simulated using R (R Core Team, 2017). Step 1 latent class models for the three-step approaches were fitted using R package *poLCA* (Linzer and Lewis, 2011) and Step 3 Cox regression models were fitted using the *coxph*() function in the *survival* package (Therneau, 2015), with the default of Efron's method for tied survival times. Standard errors and Wald 95% confidence intervals are those returned from these packages. Robust standard errors were used for three-step models with PrA to account for observations entering the analysis model twice. One and two-step models were fitted using an author-written R function, *LCSM*(), which uses an adapted version of the estimation routine detailed in Larsen (2004) to include a piecewise exponential time-to-event submodel. Both *LCSM*() and *poLCA* use the expectation-maximisation algorithm (Dempster et al., 1977) with Newton-Raphson steps to obtain maximum likelihood estimates (Larsen, 2004; Linzer and Lewis, 2011). Standard errors for one and two-step models were obtained using Louis's method (Louis, 1982). Standard errors in the two-step models were corrected to account for Step 1 parameter fixing as described previously (Xue and Bandeen-Roche, 2004; Bakk and Kuha, 2017). The R code for *LCSM*() is available on request from the corresponding author.

Data simulation

Two-class models with equal prevalences and ten independent Bernoulli distributed manifest variables were simulated. The factors manipulated were (a) sample size, $N \in \{500, 1000\}$, (b) approximate entropy statistic values, $E \in \{0.35, 0.50, 0.70\}$ and (c) the hazard ratio for the latent class effect, $\exp(\gamma_1) \in \{1, 1.5, 2, 3\}$ (note that $\exp(\gamma_2) = 1$, i.e. no effect, for identifiability), giving 24 simulation scenarios in total.

The entropy statistic values are similar to those used previously (Clark and Muthén, 2009) and correspond to low, medium and high class separation respectively. The manifest variables were simulated from independent Bernoulli distributions according to a crossed-profile plot, as depicted in Figure 2 and used previously by Clark and Muthén (2009). The entropy settings were obtained by varying the class conditional response probabilities, $\pi_{(1)} \in \{0.60, 0.65, 0.70\}$ and $\pi_{(2)} = 1 - \pi_{(1)}$.

Simulated event times were based on observed data from the ESPAC3v2 trial (Neoptolemos et al., 2010). The ESPAC3v2 trial was an open-label randomised controlled trial in patients with pancreatic ductal adenocarcinoma who had undergone cancer resection. Patients were randomised to either fluorouracil plus folinic acid or gemcitabine (the standard of care). Survival times were generated using the Kaplan-Meier estimate of the overall survival curve from the gemcitabine arm as described in the supplementary materials.

The hazard ratio values for the latent class effect, $\exp(\gamma_1)$, were chosen to represent no effect and approximate small, medium and large effect sizes respectively (Azuero, 2016). In addition to the latent class effect, an independent Bernoulli distributed time-to-event covariate, z, with a probability of 0.5 was simulated for each subject. This covariate was included to mimic randomised treatment in a clinical trial setting and the effect on survival was fixed across simulations as $\exp(\beta) = 0.75$. Administrative censoring was applied at 60 months and uniform censoring was added by generating censoring times from an exponential distribution such that overall approximately 50% of survival times were right-censored in each scenario.

Model fitting

The ten fitted model types (1 $\times$ one-step, 1 $\times$ two-step, 4 $\times$ standard three-step and 4 $\times$ inclusive three-step) are detailed in Table 1. For each model type and simulation scenario, 50 sets of random starting values were used and the best fitting model was selected in order to avoid obtaining local maximum solutions. A tolerance of 10^{-9} was used for convergence and a maximum of 1000 iterations were permitted. For each scenario there were 2000 replications. Class labelling was evaluated using the algorithm described in the supplementary materials. Parameter estimates were to be evaluated in terms of bias, percentage bias, 95% confidence interval coverage and 95% confidence interval length (Burton et al., 2006, Table I).

RESULTS

Simulation results for the estimated hazard ratios and corresponding performance measures are presented by true latent class effect and can be found in Tables 2 to 5. For simplicity, results aggregated over the small, medium and large effect sizes are presented in Table 1 (where the high and medium entropy scenarios have also been aggregated). As an illustrative example, histograms of parameter estimates from Scenario 23 (low entropy, large effect, $N = 500$) are presented in Figure 3.

One and two-step approaches

Latent class effect estimates for the one-step model exhibited no or low bias and approximately nominal coverage in most scenarios, although in the low entropy and low sample size scenarios confidence interval coverage was slightly below the nominal level at 93% on aggregate (Table 1).

In the medium and high entropy scenarios two-step estimates were unbiased with nominal coverage. In the low entropy scenarios the two-step models exhibited some parameter attenuation, although this was less pronounced for the larger sample size (18% and 8% for small and large sample sizes respectively, on aggregate, Table 1). Confidence interval coverage was typically similar to the one-step approach.

Standard three-step approaches

Estimates from the standard three-step models were approximately unbiased in the no effect scenarios (Table 2). In these scenarios *MA* and *PA* exhibited nominal coverage, but coverage was generally too high for *mPC* and *PrA* at approximately 97-99%. For the small, medium and large effect scenarios (Tables 3 to 5), *MA*, *mPC* and *PrA* estimates exhibited considerable bias towards the null and poor coverage. Even in the high entropy and larger sample size scenarios these methods exhibited attenuation in the latent class effect of >19% and this became considerably worse in the low entropy scenarios.

PA estimates were approximately unbiased with nominal coverage in all of the high and medium entropy scenarios, irrespective of effect size. With $N = 500$ in the low entropy scenarios for the small, medium and large effect sizes (Tables 3 to 5) *PA* exhibited considerable attenuation (20% on aggregate, Table 1) and poor coverage, but the bias was far less than the other standard three-step procedures which exhibited >57% bias on aggregate. With $N = 1000$ in the low entropy scenarios, attenuation was improved (10% on aggregate) but coverage was below the nominal level (90% on aggregate).

Inclusive three-step approaches

Latent class effect estimates from the inclusive three-step approaches were further from the null than their counterpart standard three-step approaches, suggesting some reversal of attenuation as intended. This effect is illustrated in Figure 3. *Incl-PrA* and *Incl-mPC* produced no or low bias in all scenarios, and *Incl-MA* exhibited improved bias compared with the counterpart standard three-step approaches. *Incl-PA* estimates, however, considerably exceeded the true values for small, medium and large effects for all entropy levels (Tables 3 to 5).

Coverage was below the nominal value for estimates from all of the inclusive three-step approaches in all scenarios. In the no effect scenarios the low coverage of the inclusive estimates generally resulted in more than double the nominal Type I error rates, and this became far worse as the entropy decreased (Table 2).

DISCUSSION

In this article, one, two and three-step approaches to latent class modeling with a time-to-event distal outcome were presented and the empirical properties of latent class effect estimates were compared using Monte Carlo simulation. To our knowledge this is the first study to investigate various approaches to latent class modeling when the distal outcome is a time-to-event variable. Moreover this is the first study to demonstrate and implement two-step (Bakk and Kuha, 2017) and inclusive bias-correction approaches (Bray et al., 2015) with a time-to-event distal outcome. This study contributes to the emerging body of literature on latent class modeling with a distal outcome variable.

Latent class effect estimates for the one-step model exhibited no or low bias and approximately nominal coverage in most scenarios, although confidence interval coverage was slightly below the nominal value in the low entropy and low sample size scenarios. The lack of bias is consistent with studies with a continuous distal outcome (Clark and Muthén, 2009; Asparouhov and Muthén, 2014; Bakk and Kuha, 2017). Interestingly, standard errors (which determine confidence interval coverage when an estimate is unbiased) in one-step models with a continuous distal outcome have been shown to be both overestimated and underestimated previously when both the entropy and sample size are low (Bakk et al., 2013; Bakk and Kuha, 2017).

The two-step approach resulted in low bias and approximately nominal coverage in the medium and high entropy scenarios. However this approach did exhibit some bias towards the null in the low entropy scenarios, which is consistent with previous research using this approach with continuous distal outcome variables (Bakk and Kuha, 2017). Confidence interval coverage was typically similar to the one-step approach.

Generally standard three-step approaches resulted in attenuated estimates of the latent class effect with underestimated standard errors, resulting in poor confidence interval coverage. This result is consistent with the research literature in this area (Clark and Muthén, 2009; Bakk et al., 2013; Asparouhov and Muthén, 2014). A surprising result however was that a standard three-step approach using partial assignment produced unbiased estimates and nominal coverage in medium and high entropy scenarios. In the low entropy scenarios partial assignment exhibited similar levels of bias to the two-step approach, however confidence interval coverage was typically poorer with partial assignment.

The inclusive approach to bias-correction proposed by Bray et al. (2015) was adapted here to include a time-to-event variable as a latent class predictor. As intended the inclusive approach produced estimates further from the null than their non-inclusive counterpart models and in general improved bias. Proportional and multiple pseudo-class assignment approaches benefited considerably from the inclusive approach with no or low bias in all scenarios. Results for partial assignment (which performed well as a standard approach), however, were worse when combined with the inclusive approach. Despite the improvements in bias, confidence interval coverage of inclusive approaches was too low, notably producing increased Type I errors compared with standard three-step approaches when simulating under the null hypothesis. An alternative approach to obtaining standard errors with inclusive approaches, such as bootstrapping, may help resolve these issues, as has been suggested previously for a closely related approach (Bakk and Vermunt, 2016).

Despite the superior performance of one-step approaches demonstrated in this study, one-step approaches have a few disadvantages as explicated by Vermunt (2010). The main criticism is that the distal outcome variable can influence latent class composition, possibly affecting the characteristics or even the number of latent classes (Vermunt, 2010; Bakk et al., 2016). Asparouhov and Muthén (2014) give an example of a one-step approach 'failing' where class composition is determined solely by the distal outcome variable, which in that case was simulated from a two component normal mixture distribution. The extent to which a time-to-event submodel could influence latent class composition is not clear and this is a relevant topic for further research. Inclusive three-step methods may also be subject to the same limitation. When fitting one-step latent class models we support the recommendations of Larsen (2004) and Asparouhov and Muthén (2014) in fitting latent class models without the distal outcome variable in model building and/or sensitivity analyses.

In this study the performance of the two-step (Bakk and Kuha, 2017) and inclusive bias-correction approaches (Bray et al., 2015) were assessed. A number of other correction methods have been proposed (Bolck et al., 2004; Vermunt, 2010; Petersen et al., 2012; Bakk et al., 2013; Lanza et al., 2013), although not all are suitable for modeling with a time-to-event distal outcome. Investigation of the bias-corrected three-step methods described in Vermunt (2010) and Bakk et al. (2013) with a time-to-event distal outcome would be a valuable addition to the research described here. However, these methods are based upon introducing and subsequently correcting classification error and a key advantage of the two-step approach studied here is that this step is avoided (Bakk and Kuha, 2017). Moreover, estimates from the two-step approach were previously found to have better statistical properties than corrected three-step approaches with a continuous distal outcome (Bakk and Kuha, 2017). An interesting additional feature of the two-step approach is that different observations can be used for the latent classification and distal outcome models (Xue and Bandeen-Roche, 2004; Bakk and Kuha, 2017).

In this study the various models used different hazard functions for the time-to-event outcome variable and, as identified by a reviewer, this feature warrants special attention. To model the distal outcome in both the standard and inclusive three-step approaches a Cox model was used where the baseline hazard function is not estimated. This is the most common model used in practice and does not disadvantage the standard or inclusive three-step models in any way, as demonstrated in a supportive analysis in the supplementary materials. For the inclusive three-step approaches the hazard function used for the distal outcome model should not to be confused with that in latent class prediction in Step 1, see Figure 1(c). In this approach, a non-parametric estimate of the unconditional cumulative hazard is used as a latent class predictor, as recommended for multiple imputation, and notably in preference to the observed event time or its logarithm (White and Royston, 2009).

For the one and two-step models we used piecewise exponential baseline hazard models, see Figure S1(b). Setting the partitions of the time-grid for a piecewise exponential model to the observed event times is equivalent to using a non-parametric baseline hazard, as in Larsen (2004). A non-parametric baseline hazard model is, in turn, equivalent to a Cox model (Breslow, 1974). Whilst the piecewise exponential model can offer improved parameter efficiency (it can also result in bias if the time grid is poorly specified), see Han et al. (2014), the main purpose here was to simplify the calculation of standard errors by reducing the number of parameters required to estimate the baseline hazard function. For the one and two-step models we used Louis's method (Louis, 1982) to obtain standard errors, which requires the inversion of the negative Hessian matrix and is not feasible when the number of parameters is large (Larsen, 2004). Bootstrapping has been recommended (Hsieh et al., 2006) but fitting one and two-step models to bootstrap resamples from each simulated data set was found to be overly computationally burdensome.

In this study time-to-event data were simulated and analysed using a proportional hazards model. In practice the suitability of the proportional hazards assumption should be investigated. Standard residual analyses for time-to-event data (see e.g Collett, 2015) can be used with one-step latent class models by calculating class-specific fitted values and averaging over classes (Proust-Lima et al., 2014). Other possible options are to include a time-dependent latent class effect in the one-step model (Muthén et al., 2009), to estimate separate hazard functions for latent classes (Asparouhov et al., 2006), or by investigating the tenability of the proportional hazards assumption using a pseudo class draw approach to the log-log cumulative hazard plot (Larsen, 2004).

CONCLUSIONS

In this study the empirical properties of various latent class effect estimates on a time-to-event distal outcome were compared. The study contributes to the emerging body of literature on latent class modeling with a distal outcome variable.

One-step models performed very well in general, whilst two-step approaches performed well when classes were well separated. A surprising result was that a standard three-step approach using partial assignment also performed well when classes were well-separated. Although inclusive bias-correction approaches were generally shown to decrease attenuation of the latent class effect estimate, partial assignment was overall the best performing three-step approach. However, when the entropy was low this approach was found to be inferior to one-step approaches and confidence interval coverage was generally worse than the two-step approach.

For the applied researcher, a one-step approach is recommended where possible, although excluding the distal outcome variable in model building and\or accompanying sensitivity analyses is recommended. The suitability of assuming proportional hazards should also be assessed.

Table 1: Details of models used in the simulation study and aggregated results for the estimated latent class effect.

						Absolute % Bias				% CI Coverage			
						N=500		N=1000		N=500		N=1000	
Model No.	Name	Steps	LC predictors	Ass. method	Base haz.	High/Med.	Low	High/Med.	Low	High/Med.	Low	High/Med.	Low
1	1step	1	None	-	Piecewise constant	2	6	1	3	95	93	95	94
2	2step	2	None	-	Piecewise constant	1	18	1	8	95	93	95	94
3	MA	3	None	MA	Unspecified	28	57	28	55	62	30	44	13
4	mPC	3	None	mPC	Unspecified	38	69	38	70	50	22	30	2
5	PA	3	None	PA	Unspecified	3	20	2	10	95	87	94	90
6	PrA	3	None	PrA	Unspecified	38	69	38	70	40	10	23	0
7	Incl-MA	3	$H_0(t)$,δ,z	MA	Unspecified	13	43	14	45	77	41	73	32
8	Incl-mPC	3	$H_0(t)$,δ,z	mPC	Unspecified	1	8	0	3	89	67	88	68
9	Incl-PA	3	$H_0(t)$,δ,z	PA	Unspecified	56	196	56	186	43	16	28	6
10	Incl-PrA	3	$H_0(t)$,δ,z	PrA	Unspecified	1	7	0	2	82	52	82	54

Table 1: MA modal assignment, mPC multiple pseudo class draws, PA partial assignment, PrA proportional assignment, Incl inclusive. Absolute % bias and 95% confidence interval (CI) coverage are aggregate results over the small, medium and large effect sizes. High and medium entropy results have been aggregated. One-step models (1) exhibit little or no bias and approximately nominal coverage, two-step models (2) exhibit no bias and nominal coverage in High/Med entropy scenarios but are biased in Low entropy scenarios, standard three-step models (3-6) generally exhibit bias and poor coverage (excepting PA in High/Med entropy scenarios), inclusive three-step models (7-10) offer improved bias over standard three-step models (excepting Incl-PA) but coverage is generally poor.

Table 2: Simulation results for the effect of latent class in scenarios with a hazard ratio of 1.

Scenario	HR	log(HR)	Entropy	$\pi_{(1)}$	$\pi_{(2)}$	N	Measure	1step	2step	MA	mPC	PA	PrA	Incl-MA	Incl-mPC	Incl-PA	Incl-PrA
1	1	0	High	0.30	0.70	500	Estimate	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
							Bias	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
							Bias (%)	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
							SE	0.15	0.15	0.13	0.13	0.15	0.12	0.13	0.13	0.15	0.12
							CI Coverage (%)	95.0	95.2	95.0	98.5	95.0	96.8	85.2	91.0	84.4	87.8
							CI Length	0.59	0.59	0.51	0.52	0.59	0.47	0.51	0.52	0.59	0.47
2	1	0	High	0.30	0.70	1000	Estimate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
							Bias	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
							Bias (%)	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
							SE	0.11	0.11	0.09	0.09	0.11	0.08	0.09	0.09	0.11	0.08
							CI Coverage (%)	94.9	94.9	94.7	98.0	95.2	96.5	81.5	91.2	85.2	87.7
							CI Length	0.42	0.42	0.36	0.37	0.42	0.33	0.36	0.37	0.42	0.33
3	1	0	Medium	0.35	0.65	500	Estimate	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
							Bias	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
							Bias (%)	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
							SE	0.18	0.18	0.13	0.14	0.17	0.11	0.13	0.14	0.17	0.11
							CI Coverage (%)	94.0	94.4	94.1	99.2	94.3	97.4	75.0	85.5	69.2	77.4
							CI Length	0.70	0.70	0.51	0.54	0.68	0.45	0.51	0.54	0.68	0.45
4	1	0	Medium	0.35	0.65	1000	Estimate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
							Bias	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
							Bias (%)	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
							SE	0.12	0.13	0.09	0.10	0.12	0.08	0.09	0.10	0.12	0.08
							CI Coverage (%)	95.0	95.4	95.0	99.3	95.2	97.8	75.3	85.9	71.4	79.0
							CI Length	0.49	0.49	0.36	0.37	0.48	0.31	0.36	0.37	0.48	0.31
5	1	0	Low	0.40	0.60	500	Estimate	-0.01	-0.01	-0.00	-0.00	-0.01	-0.00	-0.01	0.00	-0.22	0.00
							Bias	-0.01	-0.01	-0.00	-0.00	-0.01	-0.00	-0.01	0.00	-0.22	0.00
							Bias (%)	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
							SE	0.25	0.25	0.14	0.15	0.22	0.11	0.14	0.15	0.22	0.11
							CI Coverage (%)	91.8	96.6	94.9	99.9	93.9	98.4	52.6	68.7	45.9	55.2
							CI Length	0.95	0.96	0.54	0.58	0.87	0.43	0.56	0.59	0.88	0.44
6	1	0	Low	0.40	0.60	1000	Estimate	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.01	-0.00
							Bias	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.01	-0.00
							Bias (%)	NC	NC	NC	NC	NC	NC	NC	NC	NC	NC
							SE	0.17	0.17	0.09	0.10	0.16	0.08	0.10	0.10	0.16	0.08
							CI Coverage (%)	93.1	96.0	94.9	99.9	94.4	99.2	53.1	71.5	44.9	58.7
							CI Length	0.69	0.69	0.37	0.39	0.63	0.29	0.37	0.39	0.63	0.30

Table 2: Estimates are presented on the log scale. NC not calculable, MA modal assignment, mPC multiple pseudo class draws, PA partial assignment, PrA proportional assignment, Incl inclusive. All models are unbiased. One-step, two-step, MA and PA models exhibit approximately nominal coverage. mPC and PrA coverage is too high and for all inclusive methods coverage is too low, implying increased Type I error rates.

Table 3: Simulation results for the effect of latent class in scenarios with a hazard ratio of 1.5.

Scenario	HR	log(HR)	Entropy	$\pi_{(1)}$	$\pi_{(2)}$	N	Measure	1step	2step	MA	mPC	PA	PrA	Incl-MA	Incl-mPC	Incl-PA	Incl-PrA
7	1.5	0.41	High	0.30	0.70	500	Estimate	0.41	0.41	0.33	0.30	0.40	0.30	0.45	0.41	0.55	0.41
							Bias	0.00	-0.00	-0.08	-0.11	-0.00	-0.11	0.04	0.01	0.15	0.00
							Bias (%)	0.98	-0.09	-19.33	-26.61	-0.53	-26.71	10.73	1.29	36.31	1.07
							SE	0.15	0.15	0.13	0.13	0.15	0.12	0.13	0.13	0.15	0.12
							CI Coverage (%)	94.8	95.2	90.1	91.2	95.1	86.7	85.4	91.6	73.6	87.5
							CI Length	0.59	0.59	0.50	0.52	0.58	0.46	0.50	0.52	0.58	0.47
8	1.5	0.41	High	0.30	0.70	1000	Estimate	0.41	0.41	0.33	0.30	0.40	0.30	0.47	0.41	0.55	0.41
							Bias	0.00	-0.00	-0.08	-0.11	-0.00	-0.11	0.07	0.00	0.15	0.00
							Bias (%)	0.44	-0.08	-19.13	-26.70	-0.18	-26.75	17.05	0.53	36.19	0.54
							SE	0.11	0.11	0.09	0.09	0.10	0.08	0.09	0.09	0.11	0.08
							CI Coverage (%)	94.7	94.9	86.7	82.3	94.5	76.0	80.2	91.5	66.8	88.5
							CI Length	0.41	0.41	0.35	0.36	0.41	0.33	0.36	0.36	0.41	0.33
9	1.5	0.41	Medium	0.35	0.65	500	Estimate	0.41	0.40	0.27	0.22	0.40	0.22	0.48	0.42	0.74	0.41
							Bias	0.01	-0.01	-0.14	-0.19	-0.01	-0.19	0.07	0.01	0.33	0.01
							Bias (%)	1.98	-1.62	-33.69	-45.84	-2.48	-46.11	18.25	2.49	82.42	2.17
							SE	0.18	0.18	0.13	0.14	0.17	0.11	0.13	0.14	0.17	0.11
							CI Coverage (%)	94.5	95.0	80.7	78.5	94.7	62.4	72.3	84.5	49.0	75.9
							CI Length	0.69	0.69	0.50	0.53	0.67	0.44	0.51	0.54	0.68	0.44
10	1.5	0.41	Medium	0.35	0.65	1000	Estimate	0.41	0.40	0.27	0.22	0.40	0.22	0.49	0.41	0.74	0.41
							Bias	0.00	-0.01	-0.14	-0.19	-0.01	-0.19	0.08	0.00	0.34	0.00
							Bias (%)	0.53	-1.33	-33.80	-46.49	-1.50	-46.48	19.97	1.07	83.07	0.71
							SE	0.12	0.12	0.09	0.09	0.12	0.08	0.09	0.09	0.12	0.08
							CI Coverage (%)	94.6	95.2	66.8	46.8	94.8	30.6	70.5	85.7	34.0	77.3
							CI Length	0.49	0.49	0.35	0.37	0.48	0.31	0.36	0.37	0.48	0.31
11	1.5	0.41	Low	0.40	0.60	500	Estimate	0.43	0.33	0.17	0.13	0.33	0.13	0.54	0.41	0.97	0.40
							Bias	0.02	-0.08	-0.23	-0.28	-0.08	-0.28	0.14	0.00	0.56	-0.00
							Bias (%)	4.94	-18.91	-56.94	-68.80	-19.29	-68.98	33.63	0.69	138.42	-1.18
							SE	0.24	0.24	0.14	0.15	0.22	0.11	0.14	0.15	0.22	0.11
							CI Coverage (%)	92.8	94.8	63.9	57.2	91.8	26.1	47.3	67.0	30.1	51.1
							CI Length	0.96	0.96	0.54	0.57	0.86	0.42	0.55	0.58	0.86	0.43
12	1.5	0.41	Low	0.40	0.60	1000	Estimate	0.42	0.37	0.18	0.12	0.37	0.12	0.60	0.42	1.23	0.42
							Bias	0.01	-0.03	-0.22	-0.28	-0.03	-0.28	0.20	0.02	0.83	0.02
							Bias (%)	3.04	-8.23	-54.49	-69.28	-8.34	-69.39	48.93	4.37	203.54	3.97
							SE	0.17	0.17	0.09	0.10	0.16	0.07	0.09	0.10	0.16	0.07
							CI Coverage (%)	94.2	95.6	35.6	7.4	94.0	1.1	43.2	69.8	16.0	57.0
							CI Length	0.62	0.62	0.36	0.38	0.62	0.29	0.37	0.38	0.62	0.29

Table 3: Estimates are presented on the log scale. MA modal assignment, mPC multiple pseudo class draws, PA partial assignment, PrA proportional assignment, Incl inclusive. One-step models exhibit no or low bias and approximately nominal coverage. Two-step and PA are unbiased with nominal coverage in medium and high entropy scenarios but both exhibit bias when the entropy is low. Incl-mPC and Incl-PrA exhibit no or low bias but poor coverage.

Table 4: Simulation results for the effect of latent class in scenarios with a hazard ratio of 2.

Scenario	HR	log(HR)	Entropy	$\pi_{(1)}$	$\pi_{(2)}$	N	Measure	1step	2step	MA	mPC	PA	PrA	Incl-MA	Incl-mPC	Incl-PA	Incl-PrA
13	2	0.69	High	0.30	0.70	500	Estimate	0.70	0.70	0.56	0.50	0.68	0.50	0.75	0.70	0.94	0.70
							Bias	0.01	0.00	-0.14	-0.19	-0.01	-0.19	0.06	0.01	0.24	0.01
							Bias (%)	1.34	0.39	-19.87	-27.42	-1.20	-27.54	8.08	0.95	34.96	0.75
							SE	0.15	0.15	0.13	0.13	0.15	0.12	0.13	0.13	0.15	0.12
							CI Coverage (%)	95.2	95.5	80.5	72.9	95.5	63.8	85.9	92.2	60.0	87.9
							CI Length	0.59	0.59	0.50	0.51	0.58	0.46	0.50	0.52	0.58	0.46
14	2	0.69	High	0.30	0.70	1000	Estimate	0.69	0.69	0.55	0.50	0.68	0.50	0.76	0.69	0.93	0.69
							Bias	-0.00	-0.00	-0.14	-0.19	-0.01	-0.20	0.06	-0.00	0.24	-0.00
							Bias (%)	-0.07	-0.57	-20.65	-28.06	-1.63	-28.21	9.11	-0.40	34.13	-0.45
							SE	0.11	0.11	0.09	0.09	0.10	0.08	0.09	0.09	0.11	0.08
							CI Coverage (%)	94.8	94.8	63.4	42.8	94.9	32.2	84.0	90.8	41.9	86.4
							CI Length	0.42	0.42	0.35	0.36	0.41	0.32	0.36	0.36	0.41	0.33
15	2	0.69	Medium	0.35	0.65	500	Estimate	0.71	0.68	0.45	0.37	0.66	0.37	0.81	0.70	1.24	0.70
							Bias	0.01	-0.01	-0.24	-0.33	-0.03	-0.33	0.12	0.01	0.54	0.01
							Bias (%)	1.99	-1.57	-34.97	-47.00	-4.07	-47.28	17.31	1.59	78.23	1.21
							SE	0.18	0.18	0.13	0.13	0.17	0.11	0.13	0.14	0.17	0.11
							CI Coverage (%)	95.2	95.1	52.4	25.1	95.0	10.8	69.2	85.2	25.9	76.7
							CI Length	0.70	0.70	0.50	0.53	0.67	0.43	0.51	0.53	0.67	0.43
16	2	0.69	Medium	0.35	0.65	1000	Estimate	0.69	0.68	0.45	0.36	0.67	0.36	0.81	0.69	1.24	0.69
							Bias	0.00	-0.01	-0.24	-0.33	-0.02	-0.33	0.11	-0.00	0.54	-0.00
							Bias (%)	0.24	-1.53	-35.17	-47.68	-3.26	-47.76	16.48	-0.07	78.47	-0.27
							SE	0.13	0.13	0.09	0.09	0.12	0.08	0.09	0.09	0.12	0.08
							CI Coverage (%)	94.7	94.9	23.1	1.5	94.7	0.2	64.3	84.2	9.3	76.1
							CI Length	0.49	0.49	0.35	0.36	0.47	0.30	0.36	0.37	0.48	0.30
17	2	0.69	Low	0.40	0.60	500	Estimate	0.73	0.57	0.30	0.22	0.57	0.22	1.03	0.77	2.13	0.76
							Bias	0.04	-0.12	-0.39	-0.47	-0.13	-0.47	0.34	0.08	1.44	0.06
							Bias (%)	5.83	-17.37	-56.69	-67.66	-18.12	-67.85	48.95	11.23	207.82	9.08
							SE	0.25	0.25	0.14	0.14	0.22	0.11	0.14	0.15	0.22	0.11
							CI Coverage (%)	93.3	94.1	23.4	8.8	90.4	2.4	41.4	66.7	14.1	51.8
							CI Length	0.91	0.91	0.53	0.56	0.85	0.42	0.55	0.57	0.85	0.41
18	2	0.69	Low	0.40	0.60	1000	Estimate	0.71	0.64	0.31	0.21	0.63	0.21	1.00	0.71	1.99	0.71
							Bias	0.02	-0.06	-0.38	-0.48	-0.06	-0.48	0.31	0.02	1.30	0.02
							Bias (%)	2.34	-8.10	-55.07	-69.58	-9.29	-69.78	44.27	2.73	187.78	2.27
							SE	0.17	0.17	0.09	0.10	0.16	0.07	0.09	0.10	0.16	0.07
							CI Coverage (%)	94.4	94.7	2.3	0.1	92.4	0.1	32.8	68.4	2.0	54.1
							CI Length	0.77	0.78	0.36	0.38	0.61	0.28	0.37	0.38	0.61	0.28

Table 4: Estimates are presented on the log scale. MA modal assignment, mPC multiple pseudo class draws, PA partial assignment, PrA proportional assignment, Incl inclusive. One-step models exhibit no or low bias and approximately nominal coverage. Two-step and PA are unbiased with nominal coverage in medium and high entropy scenarios but both exhibit bias when the entropy is low. Incl-mPC and Incl-PrA exhibit no or low bias but poor coverage.

Table 5: Simulation results for the effect of latent class in scenarios with a hazard ratio of 3.

Scenario	HR	log(HR)	Entropy	$\pi_{(1)}$	$\pi_{(2)}$	N	Measure	1step	2step	MA	mPC	PA	PrA	Incl-MA	Incl-mPC	Incl-PA	Incl-PrA
19	3	1.1	High	0.30	0.70	500	Estimate	1.12	1.11	0.86	0.77	1.07	0.77	1.17	1.11	1.46	1.10
							Bias	0.02	0.01	-0.24	-0.33	-0.03	-0.33	0.07	0.01	0.36	0.00
							Bias (%)	1.97	0.68	-21.75	-29.59	-2.65	-29.75	6.39	0.62	33.15	0.31
							SE	0.16	0.16	0.13	0.14	0.15	0.12	0.14	0.14	0.16	0.12
							CI Coverage (%)	95.0	95.5	53.6	30.0	94.9	19.0	86.2	92.0	39.9	87.1
							CI Length	0.63	0.63	0.51	0.53	0.60	0.46	0.53	0.55	0.61	0.48
20	3	1.1	High	0.30	0.70	1000	Estimate	1.11	1.11	0.86	0.77	1.07	0.77	1.17	1.10	1.46	1.10
							Bias	0.01	0.01	-0.24	-0.33	-0.03	-0.33	0.07	-0.00	0.36	-0.00
							Bias (%)	1.28	0.68	-21.96	-29.71	-2.47	-29.87	6.37	-0.05	32.98	-0.16
							SE	0.11	0.11	0.09	0.09	0.11	0.08	0.10	0.10	0.11	0.09
							CI Coverage (%)	94.8	95.1	26.3	3.6	94.5	1.6	84.4	91.1	15.7	87.3
							CI Length	0.44	0.44	0.36	0.37	0.42	0.32	0.37	0.38	0.43	0.34
21	3	1.1	Medium	0.35	0.65	500	Estimate	1.13	1.09	0.69	0.56	1.03	0.56	1.28	1.11	1.89	1.10
							Bias	0.03	-0.01	-0.41	-0.54	-0.07	-0.54	0.18	0.01	0.80	0.01
							Bias (%)	3.05	-1.17	-37.30	-49.18	-6.36	-49.43	16.14	1.03	72.39	0.48
							SE	0.19	0.19	0.13	0.14	0.17	0.11	0.14	0.14	0.18	0.11
							CI Coverage (%)	94.4	95.4	12.0	0.4	92.8	0.0	64.3	85.8	8.1	75.8
							CI Length	0.76	0.76	0.51	0.54	0.69	0.43	0.54	0.56	0.70	0.44
22	3	1.1	Medium	0.35	0.65	1000	Estimate	1.12	1.10	0.69	0.55	1.04	0.55	1.27	1.10	1.90	1.10
							Bias	0.02	-0.00	-0.41	-0.54	-0.06	-0.55	0.17	-0.00	0.80	-0.00
							Bias (%)	1.91	-0.15	-37.04	-49.51	-5.09	-49.64	15.32	-0.08	73.20	-0.32
							SE	0.14	0.14	0.09	0.09	0.12	0.08	0.10	0.10	0.13	0.08
							CI Coverage (%)	94.5	95.6	0.5	0.0	93.0	0.0	55.9	85.4	0.6	75.9
							CI Length	0.53	0.54	0.36	0.37	0.49	0.30	0.38	0.39	0.50	0.31
23	3	1.1	Low	0.40	0.60	500	Estimate	1.16	0.89	0.46	0.33	0.86	0.33	1.62	1.24	3.75	1.21
							Bias	0.06	-0.21	-0.64	-0.76	-0.24	-0.77	0.52	0.14	2.65	0.11
							Bias (%)	5.90	-18.68	-58.55	-69.63	-21.80	-69.90	47.51	12.45	241.27	9.86
							SE	0.28	0.27	0.14	0.15	0.22	0.11	0.15	0.15	0.23	0.11
							CI Coverage (%)	94.2	88.9	3.3	0.9	79.5	0.2	35.3	66.9	4.0	51.6
							CI Length	1.31	1.32	0.54	0.57	0.87	0.42	0.58	0.59	0.89	0.41
24	3	1.1	Low	0.40	0.60	1000	Estimate	1.15	1.01	0.47	0.32	0.96	0.32	1.55	1.11	2.93	1.11
							Bias	0.05	-0.08	-0.63	-0.78	-0.14	-0.78	0.45	0.02	1.83	0.01
							Bias (%)	4.87	-7.72	-56.91	-70.97	-12.47	-71.11	40.78	1.46	166.26	1.02
							SE	0.20	0.19	0.09	0.10	0.16	0.07	0.10	0.10	0.16	0.07
							CI Coverage (%)	94.4	92.8	0.0	0.0	84.0	0.0	21.3	66.7	0.0	51.0
							CI Length	0.70	0.66	0.36	0.38	0.62	0.28	0.40	0.40	0.63	0.28

Table 5: Estimates are presented on the log scale. MA modal assignment, mPC multiple pseudo class draws, PA partial assignment, PrA proportional assignment, Incl inclusive. One-step models exhibit no or low bias and approximately nominal coverage. Two-step and PA are unbiased with nominal coverage in medium and high entropy scenarios but both exhibit bias when the entropy is low. Incl-mPC and Incl-PrA exhibit no or low bias but poor coverage.

FIGURES AND TABLES

Figure 1: Schematics for the latent class models discussed in this article.

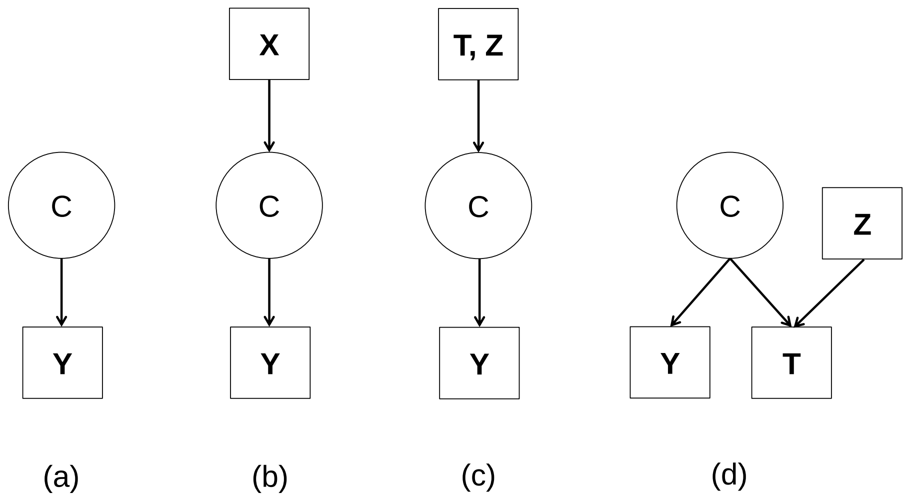

Figure 1: (a) latent class model, (b) latent class regression model, (c) inclusive latent class regression model and (d) one-step latent class model with a distal outcome. Circles and squares are used to identify unobserved (i.e. latent class) and observed variables respectively. C latent class variable, $\mathbf{Y}$ manifest variables, $\mathbf{X}$ latent class predictors, $\mathbf{T}$ distal outcome(s) and $\mathbf{Z}$ covariates possibly related to $\mathbf{T}$.

Figure 2: Class-conditional response probabilities used in the simulation study.

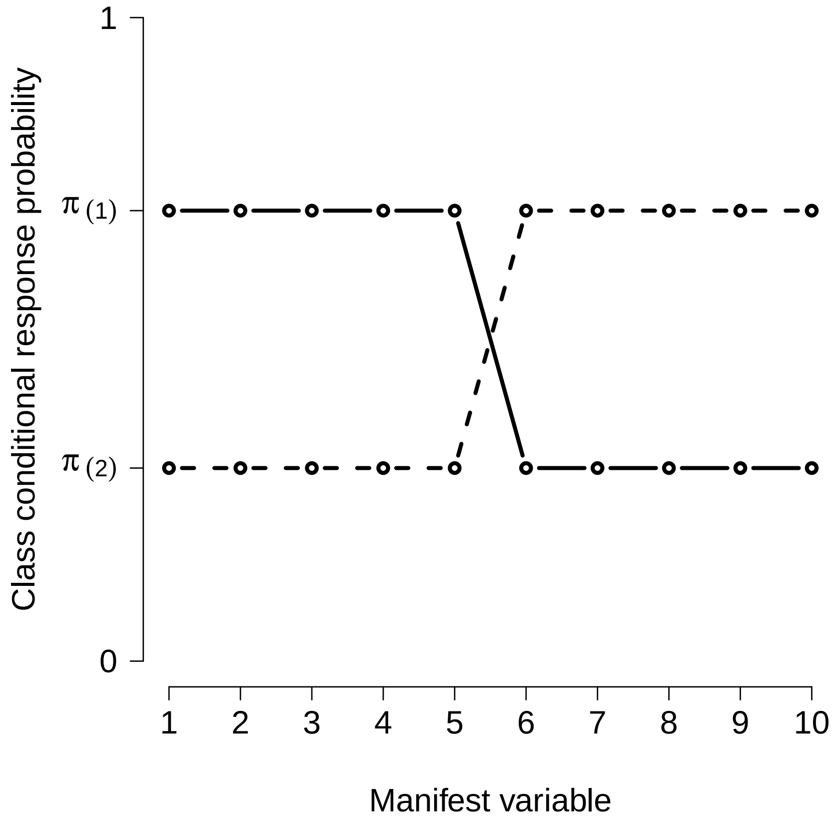

Figure 2: Ten independent Bernoulli distributed manifest variables were simulated according to a crossed profile plot for the two latent classes, where $\pi_{(1)} \in \{0.60, 0.65, 0.70\}$ and $\pi_{(2)} = 1 - \pi_{(1)}$.

References

Asparouhov, T., Masyn, K., and Muthén, B. (2006). Continuous time survival in latent variable models. In *Proceedings of the Joint Statistical Meeting in Seattle*, pages 180–187.

Asparouhov, T. and Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3):329–341.

Auer, M. F. F., Hickendorff, M., Putten, C. M. V., Béguin, A. A., and Heiser, W. J. (2016). Multilevel latent class analysis for large-scale educational assessment data: Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education*, 29(2):144–159.

Azuero, A. (2016). A note on the magnitude of hazard ratios. *Cancer*, 122(8):1298–1299.

Bakk, Z. and Kuha, J. (2017). Two-step estimation of models between latent classes and external variables. *Psychometrika*.

Bakk, Z., Oberski, D. L., and Vermunt, J. K. (2016). Relating latent class membership to continuous distal outcomes: Improving the LTB approach and a modified three-step implementation. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(2):278–289.

Bakk, Z., Tekle, F. B., and Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, 43(1):272–311.

Figure 3: Histograms of simulation results taken from low entropy Scenario 23 ($N = 500$, $\pi_{(1)} = 0.4$, $\pi_{(2)} = 0.6$).

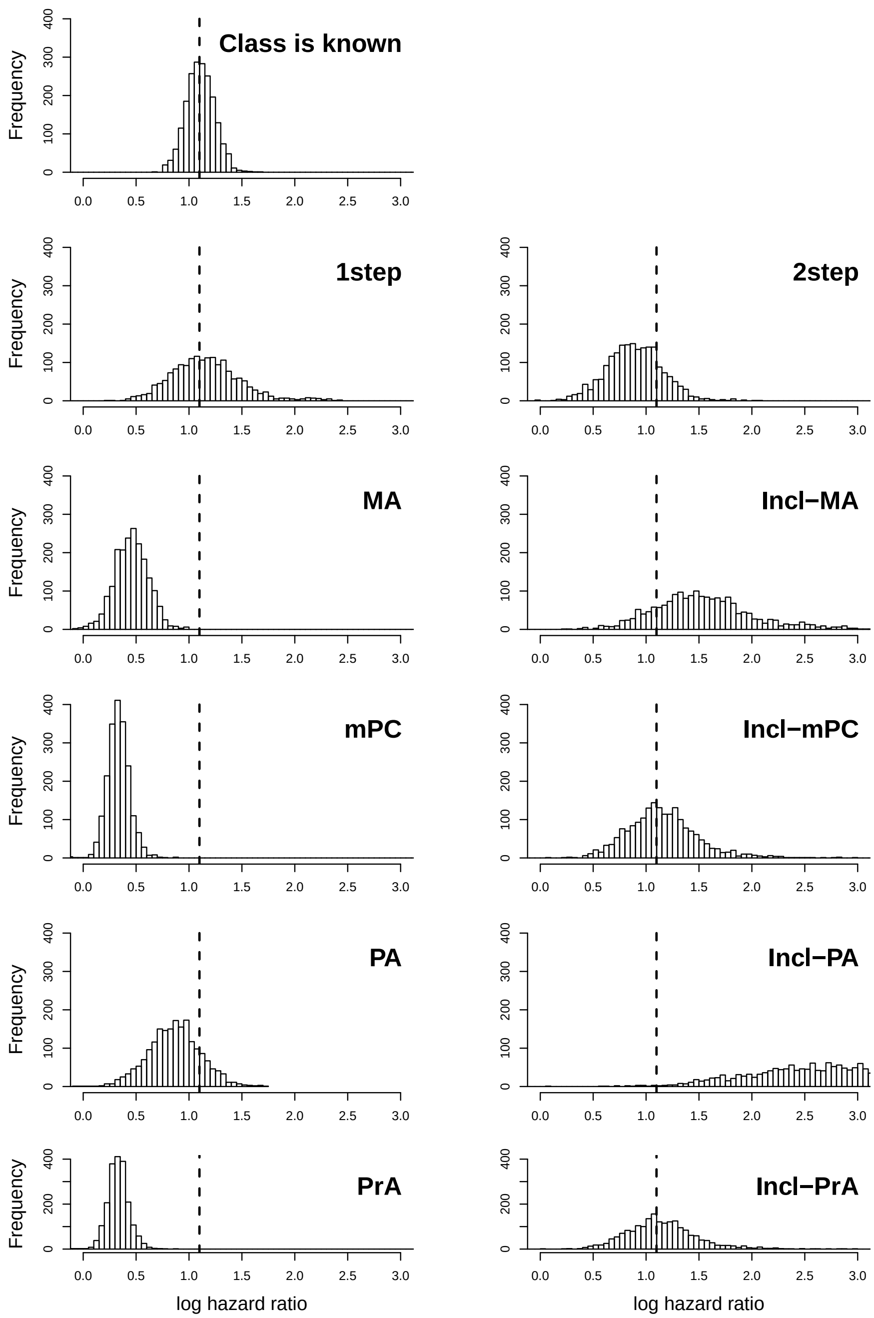

Figure 3: The dashed vertical lines represent the true latent class effect, in this case $\log(3) \approx 1.10$, and deviations of the empirical distributions from the true value indicate bias. For corresponding confidence interval coverage and length see Table 5. 'Class is known' refers to results from a Cox regression model including the known underlying class and is included for demonstration purposes only. MA modal assignment, mPC multiple pseudo class draws, PA partial assignment, PrA proportional assignment, Incl inclusive.

Bakk, Z. and Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(1):20–31.

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., and Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92(440):1375–1386.

Bartholomew, D. J., Knott, M., and Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach*, volume 904. John Wiley & Sons.

Bolck, A., Croon, M., and Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, 12(1):3–27.

Bray, B. C., Lanza, S. T., and Tan, X. (2015). Eliminating bias in classify-analyze approaches for latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(1):1–11.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30(1):89–99.

Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24):4279–4292.

Chung, H., Flaherty, B. P., and Schafer, J. L. (2006). Latent class logistic regression: Application to marijuana use and attitudes among high school seniors. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(4):723–743.

Clark, S. L. and Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. https://www.statmodel.com/download/relatinglca.pdf [Online; accessed March-2017].

Clogg, C. (2013). Latent class models. In Arminger, G., Clogg, C. C., and Sobel, M. E., editors, *Handbook of statistical modeling for the social and behavioral sciences*, chapter 6. Springer Science & Business Media.

Collett, D. (2015). *Modelling survival data in medical research.* CRC press.

Collier, Z. K. and Leite, W. L. (2017). A comparison of three-step approaches for auxiliary variables in latent class and latent profile analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(6):819–830.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Dayton, C. M. and Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, 83(401):173–178.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

Denson, N. and Ing, M. (2014). Latent class analysis in higher education: An illustrative example of pluralistic orientation. *Research in Higher Education*, 55(5):508–526.

Desantis, S. M., Andrés Houseman, E., Coull, B. A., Nutt, C. L., and Betensky, R. A. (2012). Supervised Bayesian latent class models for high-dimensional data. *Statistics in Medicine*, 31(13):1342–1360.

Downing, A., Harrison, W. J., West, R. M., Forman, D., and Gilthorpe, M. S. (2010).

Latent class modelling of the association between socioeconomic background and breast cancer survival status at 5 years incorporating stage of disease. *Journal of Epidemiology & Community Health*, 64(9):772–776.

Dziak, J. J., Lanza, S. T., and Tan, X. (2014). Effect size, statistical power, and sample size requirements for the bootstrap likelihood ratio test in latent class analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4):534–552.

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, 87(418):476–486.

Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics*, 10(1):101–113.

Han, G., Schell, M. J., and Kim, J. (2014). Improved survival modeling in cancer research using a reduced piecewise exponential approach. *Statistics in Medicine*, 33(1):59–73.

Hsieh, F., Tseng, Y. K., and Wang, J. L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics*, 62(4):1037–1043.

Huang, G. H. and Bandeen-Roche, K. (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika*, 69(1):5–32.

Hunt, L. and Jorgensen, M. (1999). Theory & methods: Mixture model clustering using the MULTIMIX program. *Australian & New Zealand Journal of Statistics*, 41(2):154–171.

Lanza, S. T., Tan, X., and Bray, B. C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(1):1–26.

Larsen, K. (2004). Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics*, 60(1):85–92.

Leigh, L., Hudson, I. L., and Byles, J. E. (2015). Sleeping difficulty, disease and mortality in older women: A latent class analysis and distal survival analysis. *Journal of sleep research*, 24(6):648–657.

Linzer, D. and Lewis, J. (2011). poLCA: An R package for polytomous variable latent class analysis. *Journal of Statistical Software, Articles*, 42(10):1–29.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2):226–233.

Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology*, 49(2):313–334.

Muthén, B., Asparouhov, T., Boye, M., Hackshaw, M., and Naegeli, A. (2009). Applications of continuous-time survival in latent variable models for the analysis of oncology randomized clinical trial data using Mplus. http://www.statmodel2.com/download/lilyFinalReportV6.pdf [Online; accessed February-2018].

Muthén, L. K. and Muthén, B. O. (2004). *Mplus Technical Appendices.* Los Angeles, CA.

Muthén, L. K. and Muthén, B. O. (2011). *Mplus User's Guide. Sixth Edition.* Los Angeles, CA.

Neoptolemos, J. P., Stocken, D. D., Bassi, C., Ghaneh, P., Cunningham, D., Goldstein, D., Padbury, R., Moore, M. J., Gallinger, S., Mariette, C., et al. (2010). Adjuvant chemotherapy with fluorouracil plus folinic acid vs gemcitabine following pancreatic cancer resection: A randomized controlled trial. *JAMA*, 304(10):1073–1081.

Petersen, J., Bandeen-Roche, K., Budtz-Jørgensen, E., and Larsen, K. G. (2012). Predicting latent class scores for subsequent analysis. *Psychometrika*, 77(2):244–262.

Proust-Lima, C., Séne, M., Taylor, J. M., and Jacqmin-Gadda, H. (2014). Joint latent class models for longitudinal and time-to-event data: A review. *Statistical Methods in Medical Research*, 23(1):74–90. PMID: 22517270.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rahbar, M. H., Ning, J., Choi, S., Piao, J., Hong, C., Huang, H., Del Junco, D. J., Fox, E. E., Rahbar, E., and Holcomb, J. B. (2015). A joint latent class model for classifying severely hemorrhaging trauma patients. *BMC research notes*, 8(1):602.

Ramaswamy, V., Desarbo, W. S., Reibstein, D. J., and Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, 12(1):103–124.

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

Snuderl, M., Chi, S. N., DeSantis, S., Stemmer-Rachamimov, A., Betensky, R. A., DeGirolami, U., and Kieran, M. W. (2008). Prognostic value of tumor microinvasion and metalloproteinases expression in intracranial pediatric ependymomas. *The FASEB Journal*, 22(1 Supplement):706.8.

Stapinski, L. A., Edwards, A. C., Hickman, M., Araya, R., Teesson, M., Newton, N. C., Kendler, K. S., and Heron, J. (2016). Drinking to cope: a latent class analysis of coping motives for alcohol use in a large cohort of adolescents. *Prevention Science*, 17(5):584–594.

Therneau, T. M. (2015). *A package for survival analysis in S.* version 2.38.

van der Heijden, P. G. M., Dessens, J., and Bockenholt, U. (1996). Estimating the concomitant-variable latent-class model with the EM algorithm. *Journal of Educational and Behavioral Statistics*, 21(3):215–229.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18(4):450–469.

Wang, C. P., Brown, C. H., and Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models. *Journal of the American Statistical Association*, 100(471):1054–1076.

White, I. R. and Royston, P. (2009). Imputing missing covariate values for the Cox model. *Statistics in Medicine*, 28(15):1982–1998.

Xue, Q. and Bandeen-Roche, K. (2004). Combining complete multivariate outcomes with incomplete covariate information: A latent class approach. *Biometrics*, 58(1):110–120.

Zhang, J. J. and Wang, M. (2010). Latent class joint model of ovarian function suppression and DFS for premenopausal breast cancer patients. *Statistics in Medicine*, 29(22):2310–2324.

SUPPLEMENTARY MATERIAL

Simulation of time-to-event data

Survival times were simulated in a similar way to that described by Bender et al. (2005). Two classes, 1 and 2, and two treatment groups, A and B, were simulated in each scenario. The reference survival curve used was based on the Kaplan-Meier estimate of the gemcitabine arm in the ESPAC2 trial (Neoptolemos et al., 2010), Figure S1(a) and (b). Let S_0 represent this reference survival curve for subjects belonging to *true* latent class 2 and treatment group B (i.e. $c_i = 0$ and $z_i = 0$), so that survival probabilities corresponding to proportional hazard effects can be obtained using

$$S_i = S_0^{\exp(\beta z_i + \gamma_j c_i)}, \tag{8}$$

in this case for z_i and $c_i \in \{0, 1\}$. As described previously, the hazard ratio for the effect of Treatment A relative to Treatment B, $\exp(\beta)$, was fixed at 0.75 and the hazard ratio for the effect of latent class 1 relative to latent class 2 was varied, $\exp(\gamma_1) \in \{1, 1.5, 2, 3\}$. 'True' survival probabilities were obtained for each of the four permutations of class and treatment over a sequence of 0 to 60 months in steps of 0.1 months. High-dimensional spline fits were used to approximate these survival curves, as shown for the reference survival curve in Figure S1(a). The splines were fitted separately to each of the four survival curves by regressing the time sequence on polynomials of the survival probabilities. A survival probability was then simulated for each subject from $Uniform(0, 1)$ and a corresponding survival time is obtained from the relevant spline fit. Administrative censoring was applied at 60 months and uniform censoring was added by generating censoring times from an exponential distribution such that overall approximately 50%

of survival times were right-censored.

A label switching algorithm

Latent class models are only identifiable up to a permutation of class labels (McLachlan and Peel, 2004). Whilst this is not an issue in standalone applications, it is a problem for simulation studies since it is not always straight-forward to establish, for a particular simulated data set, the class label that corresponds to the true class. A useful discussion of this issue in latent variable models is given in Tueller et al. (2011), and the same labelling problem arises in Bayesian Monte Carlo Markov Chain simulations (Celeux et al., 2000; Grün and Leisch, 2009; Sperrin et al., 2010).

A number of solutions have been proposed (e.g. Tueller et al., 2011; Yao, 2015; Celeux et al., 2000). In this study, we used a clustering and relabelling strategy based on Euclidean distances, as in Celeux et al. (2000), where the distances between the true parameter values and their estimates is calculated for each simulated data set.

Assume that data are simulated according to a particular latent class model with P 'true' parameter values $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)$. There are $J!$ possible permutations of the class labels, $l = 1, \ldots, J!$, and we let the last permutation represent the correct labelling. We simulate $d = 1, \ldots, D$ data sets according to the true model. For each data set we fit a latent class model of the same form as the true model. Let $\hat{\boldsymbol{\theta}}_d = \left(\hat{\theta}_{d1}, \ldots, \hat{\theta}_{dP}\right)$ be a vector of parameter estimates from the latent class model fitted to the d^{th} data set. We assume that $\hat{\boldsymbol{\theta}}_d$ are unbiased estimates of the true values but possibly labelled incorrectly. If $\hat{\boldsymbol{\theta}}_d$ are labelled 'correctly', then

$$\frac{1}{se(\hat{\theta}_{dp})} \left(\hat{\theta}_{dp} - \theta_p\right) \sim N(0,1)$$

and

$$\left[\frac{1}{se(\hat{\theta}_{dp})}\left(\hat{\theta}_{dp}-\theta_p\right)\right]^2 \sim \chi^2_{(1)},$$

for $d = 1, \ldots, D$ and where $1 \leq p \leq P$. We then assume that parameter estimates are independent, which in practice will be determined by the form of the model fitted; this issue is discussed further below. Summing over P,

$$\sum_{p=1}^{P}\left[\frac{1}{se(\hat{\theta}_{dp})}\left(\hat{\theta}_{dp}-\theta_p\right)\right]^2 \sim \chi^2_{(P)},$$

with mean P, for $d = 1, \ldots, D$. The standardised Euclidean distance, δ_d, between the estimates from a model fitted to the dth data set and the vector of true parameter values is

$$\delta_d = \left\{\sum_{p=1}^{P}\left[\frac{1}{se(\hat{\theta}_{dp})}(\hat{\theta}_{dp}-\theta_p)\right]^2\right\}^{1/2} \sim \chi_{(P)},$$

i.e. a central χ distribution. If $\hat{\boldsymbol{\theta}}_d$ are labelled 'incorrectly', then

$$\frac{1}{se(\hat{\theta}_{dp})}\left(\hat{\theta}_p-\theta_p\right) \sim N\left(\mu_p, 1\right),$$

$$\sum_{p=1}^{P}\left[\frac{1}{se(\hat{\theta}_{dp})}\left(\hat{\theta}_{dp}-\theta_p\right)\right]^2 \sim \chi^2_{(P)}(\lambda),$$

i.e. a non-central χ^2 distribution with non-centrality parameter $\lambda = \sum_{p=1}^{P}\mu_p^2$ and mean $P + \lambda$. It therefore follows that

$$\delta_d \sim \chi_{(P)}\left(\lambda\right).$$

Letting $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_D)$, and assuming that the random starting values for the parameter estimates do not favour one label permutation over another,

$$\boldsymbol{\delta} \sim \frac{1}{J!} \sum_{l=1}^{J!} f_l(\boldsymbol{\delta}),$$

i.e. a $J!$ component mixture distribution with one central χ distribution, $f_{J!} \sim \chi_{(P)}$, and $J!-1$ non-central χ distributions, $f_l \sim \chi_{(P)}(\lambda_l)$, for $l = 1, \ldots, J!-1$. A histogram of $\boldsymbol{\delta}$ should therefore yield a mixture distribution of $J!$ (hopefully distinct) probability distributions for which the component with the lowest mean is labelled correctly. Larger differences in the true parameter values for the latent classes and greater numbers of class distinct parameters to estimate will result in clearer separation of the mixture components, making clustering and relabelling easier. An example of such a histogram for 2000 simulations from a latent class model with $J = 2$ and $P = 21$ parameters is depicted in Figure S2. Assuming sufficient separation between components, either by introducing some threshold or by clustering $\boldsymbol{\delta}$ (e.g. K-means clustering), estimates that have been labelled incorrectly can be easily identified and relabelled accordingly.

A label switching algorithm is therefore as follows

1. Fit latent class models to each of $d = 1, \ldots, D$ data sets.

2. For each of the $d = 1, \ldots, D$ data sets calculate the standardised Euclidean distances, δ_d, between each set of parameter estimates, $\hat{\boldsymbol{\theta}}_d$, and the true parameter values, $\boldsymbol{\theta}$.

3. Inspect a histogram of $\boldsymbol{\delta}$ for distinct component densities.

4. Use K-means clustering to assign each δ_d (and hence $\hat{\boldsymbol{\theta}}_d$) to a cluster (/component density). The cluster with the lowest mean corresponds to the cluster of correctly labelled parameter estimates.

5. Relabel those $\hat{\boldsymbol{\theta}}_d$ which do not belong to the correctly labelled component density.

As a check, the first three steps can be repeated using the relabelled estimates and the new histogram should reveal a unimodal (and central) χ distribution. If $J > 2$ it may be necessary to repeat this process using a permutation of the true parameter values in the place of the true values in order to distinguish between two or more incorrectly labelled clusters.

The histogram of $\boldsymbol{\delta}$ also serves as a useful diagnostic tool, since any outlying values, perhaps exceeding a selected critical threshold, can be identified and investigated further. These may represent local maximum and/or boundary solutions.

In practice, whilst we have provided theoretical justification for the distribution of the standardised Euclidean distances in the case of independent parameters, if dependencies are included in the model, then the algorithm can still be used. In this case, these parameters can be included or excluded in the algorithm, as long as the histogram of the standardised Euclidean distances reveals distinct clusters.

Comparison of the use of different hazard functions in three-step models

In this study one and two-step models used a piecewise exponential baseline hazard function whilst the three-step models used a Cox model in which the baseline hazard is left unspecified. The choice of a piecewise exponential model for the one and two-step models was primarily motivated by the fact that standard errors are easier to obtain when there are few baseline hazard parameters (see Discussion). To illustrate that the three-step methods are not disadvantaged by the choice of hazard function the table below contains results from a small simulation study for Scenario 17 (Low entropy, $N = 500$, HR=2). The results demonstrate that the results are practically unaffected by the choice of baseline hazard function.

[Table S1 about here]

SUPPLEMENTARY TABLE

Model	Estimate	Bias	CI Coverage (%)
MA using Cox model	-0.30	0.39	23.4
MA using PE model	-0.30	0.39	23.8
PA using Cox model	-0.57	0.13	90.4
PA using PE model	-0.57	0.13	90.7

Table S1: Comparison of simulation results for modal assignment and partial assignment when using unspecified (Cox) an piecewise exponential baseline hazard functions. MA Modal assignment, PA Partial assignment. The results demonstrate that the statistical properties of the latent class effect estimates are practically unaffected by the different hazard functions compared here.

SUPPLEMENTARY FIGURES

Figure S1: Kaplan-Meier estimate of overall survival for the gemcitabine arm from the ESPAC3v2 study and overlaid fitted models.

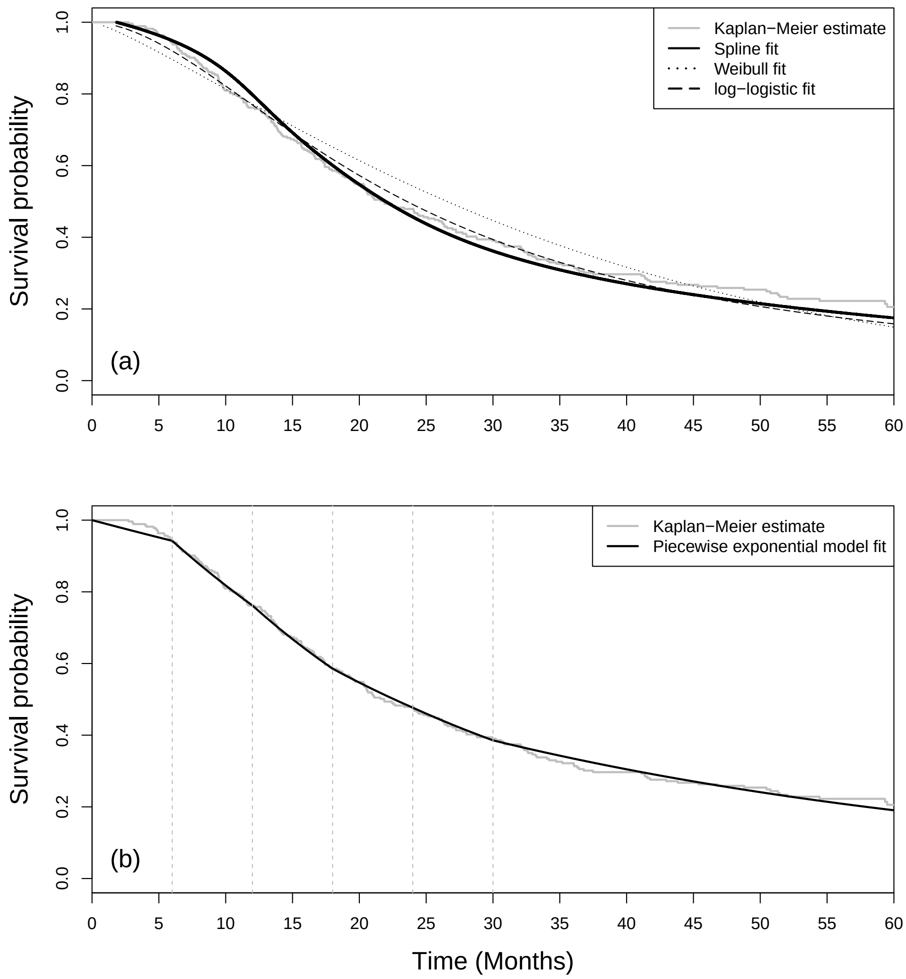

Figure S1: (a) Fitted polynomial spline, Weibull and log-logistic (parametric) models. (b) A piecewise exponential survival model with five partitions approximates the Kaplan-Meier estimate well.

Figure S2: Example of Euclidean distances between true and estimated parameters.

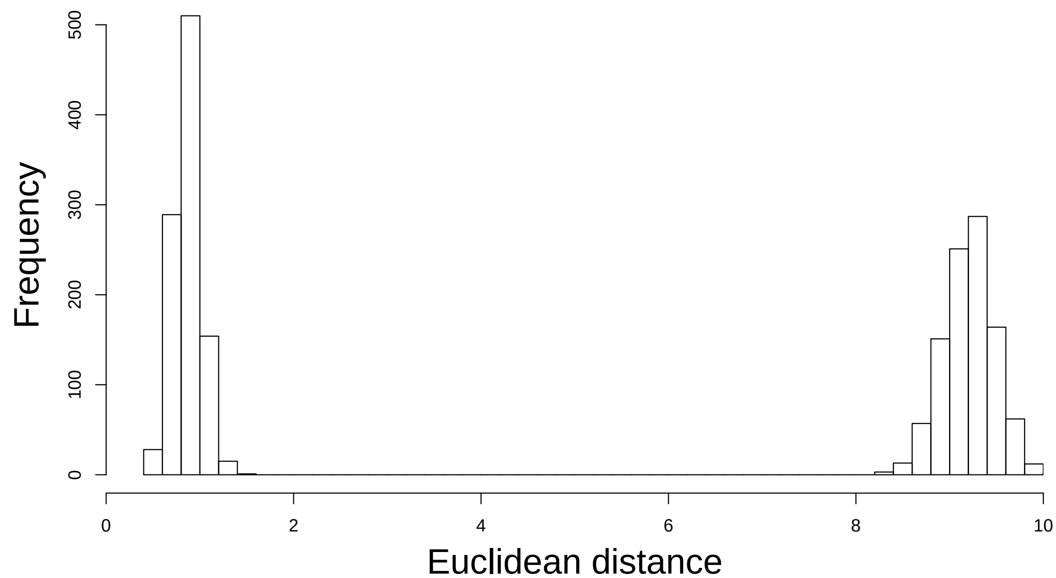

Figure S2: Example of Euclidean distances for 2000 simulations from a latent class model with 2 classes, before relabelling. The distribution on the left contains the models for which the class is correctly labelled.

Supplementary References

Bender, R., Augustin, T., and Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, 24(11):1713–1723.

Celeux, G., Hurn, M., and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970.

Grün, B. and Leisch, F. (2009). Dealing with label switching in mixture models under genuine multimodality. *J. Multivar. Anal.*, 100(5):851–861.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Neoptolemos, J. P., Stocken, D. D., Bassi, C., Ghaneh, P., Cunningham, D., Goldstein, D., Padbury, R., Moore, M. J., Gallinger, S., Mariette, C., et al. (2010). Adjuvant chemotherapy with fluorouracil plus folinic acid vs gemcitabine following pancreatic cancer resection: A randomized controlled trial. *JAMA*, 304(10):1073–1081.

Sperrin, M., Jaki, T., and Wit, E. (2010). Probabilistic relabelling strategies for the label switching problem in Bayesian mixture models. *Statistics and Computing*, 20(3):357–366.

Tueller, S. J., Drotar, S., and Lubke, G. H. (2011). Addressing the problem of switched class labels in latent variable mixture model simulation studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(1):110–131.

Yao, W. (2015). Label switching and its solutions for frequentist mixture models. *Journal of Statistical Computation and Simulation*, 85(5):1000–1012.