Published in final edited form as:

J R Stat Soc Ser A Stat Soc. 2018 June ; 181(3): 869–888. doi:10.1111/rssa.12334.

Bayesian joint modelling of longitudinal data on abstinence, frequency and intensity of drinking in alcoholism trials

Eugenia Buta, Stephanie S. O'Malley, and Ralitza Gueorguieva

Yale University, New Haven, USA

Summary.

In alcoholism research, several complementary outcomes are of interest: abstinence from drinking during a specific time frame, and, when the individual is drinking, frequency of drinking (the proportion of days on which drinking occurs) and intensity of drinking (the average number of drinks per drinking day). The outcomes are often measured repeatedly over time on the same subject and, although they are closely related, they are rarely modelled together. We propose a joint model that allows us to fit these longitudinal outcomes simultaneously, using correlated random effects to model the association between the outcomes and between repeated measurements on the same subject. The model has three parts: a logistic part for sustained abstinence over the period of interest, a truncated binomial part for frequency of drinking and a log-normal model for drinking intensity when drinking occurs. Because of the computational impracticality of fitting models with many random effects by using standard frequentist approaches, we use a Bayesian approach to fit the joint model. We also conduct a simulation study to investigate the gains in parameter estimate bias and mean-squared error associated with joint *versus* separate modelling. We illustrate the approach on data from an alcoholism clinical trial.

Keywords

Bayesian inference; Correlated random effects; Drinking; Hurdle models; Joint modelling; Zero inflation

1. Introduction

In alcoholism clinical trials, subjects' alcohol consumption is often recorded by using timeline methods that capture the number of standard drinks for each day over a particular period with the ultimate goal of assessing the effect of treatment and/or time on drinking. Commonly used summary outcomes in these trials are the abstinence over a particular period of time, frequency of drinking, presented as the proportion of days with any drinking (i.e. drinking days) over that period, and intensity of drinking (when drinking), presented as the average number of drinks per drinking day. These outcomes are analysed separately, thus ignoring important information about drinking in the analyses and ignoring the correlation that exists between the outcomes. In particular, focusing solely on the frequency of any

drinking ignores information on the intensity of drinking and would not be able to distinguish between subjects who, for example, drink lightly every day of the week and subjects who drink heavily every day of the week. Focusing solely on intensity ignores information on the frequency of drinking and would not be able to distinguish between subjects who drink heavily on most days and those who drink heavily only on some days. Modelling the outcomes jointly would use all the information about drinking in the same model, estimate the correlation between different measures and may provide a better insight into and description of the longitudinal alcohol consumption data. At the same time, this type of analysis might better predict health and psychosocial effects of drinking over the long term.

Previous references have considered various statistical methods for modelling alcohol consumption. Wang *et al.* (2002) discussed several basic methods for analysing drinking data (such as reducing the repeated measures of drinking over the whole study period to one summary for each subject, such as the proportion of drinking days or average alcohol consumption per day, and then analysing these univariate summaries, or using survival models to analyse time to the first (any or heavy) drinking day) and argued for the use of more sophisticated statistical methods (specifically, recurrent event survival analyses) that give a more comprehensive description of drinking behaviour over time. Other researchers (DeSantis *et al.*, 2013; Zhu *et al.*, 2017) proposed a longitudinal hurdle or zero-inflated Poisson model to analyse the zero-inflated average number of drinks per day (as opposed to the average number of drinks per drinking day) for each week (rounded to the nearest whole number). This artificially inflates the proportion of zeros by rounding and ignores the number of days in which drinking occurred during the week.

For modelling the longitudinal amount of daily drinking, Liu *et al.* (2008, 2016) proposed two-part models: a logistic model for the zero outcomes and a log-normal or generalized gamma model for the non-zero outcomes. Shirley *et al.* (2010) and Prisciandaro *et al.* (2012) used hidden Markov models for the amount of drinking treated as an ordinal variable with three levels (no, light or heavy drinking) and two levels (heavy or not heavy drinking), whereas Das *et al.* (2010) modelled extreme drinking only (drinking exceeding a prespecified threshold).

In the context of daily dietary intake data with many 0s, Tooze *et al.* (2006) proposed the National Cancer Institute (NCI) method, consisting of a two-part model: a logistic part used to model zero *versus* non-zero values, and a second part to model the Box–Cox-transformed positive amounts, with the two parts correlated via random intercepts (the SAS macro code is available from <https://epi.grants.cancer.gov/diet/usualintakes/macros.html>). The NCI approach macro cannot be easily extended to handle more random effects (e.g. random slopes) as SAS procedure NLMIXED on which it is based uses a quadrature approach, which cannot be used with many random effects.

Although joint models for the analysis of repeatedly measured discrete and continuous outcomes are becoming increasingly popular (see Dunson (2000), Gueorguieva and Agresti (2001), Daniels and Normand (2006) and De Leon and Chough (2013), with De Leon and Chough (2013) containing a comprehensive review of the most commonly used approaches

for joint modelling), to our knowledge these models have not been used before to analyse abstinence, frequency and intensity of drinking jointly in a three-part model. Joint models enable us to assess simultaneously the effect of a variable of interest (e.g. treatment) on multiple outcomes and may be more efficient than separate analysis by borrowing information across the various outcomes. They also allow estimation of the covariances and correlations between the different drinking measures.

In this paper, we propose a joint model that enables us to fit longitudinal data on abstinence, frequency and intensity of drinking during a specific timeframe simultaneously. We use correlated random effects to account for between-subject heterogeneity in drinking patterns and to model the association between the drinking outcomes and between repeated measurements on the same subject over time. The model has three parts: a logistic model for abstinence from drinking over the period of interest, a truncated binomial model for the frequency of drinking and a log-normal model for the skewed continuous distribution of drinking intensity when drinking occurs. We illustrate the approach on data from a subset of subjects enrolled in the 'Combined pharmacotherapies and behavioral interventions' study (called the COMBINE study) (Anton *et al.*, 2006), a large multisite randomized clinical trial of treatments for alcohol dependence. In a simulation study, we investigate how the estimation of model parameters is affected by modelling the outcomes together rather than separately.

Our method extends previous work by accommodating more random effects through a Bayesian approach based on Markov chain Monte Carlo (MCMC) sampling, and also by allowing us to model data aggregated over periods of times such as weeks or months, not just at the daily level. In contrast with models in previous work that consisted of two parts, our model was designed to analyse data over intervals of time larger than a single day. Thus it has three parts: two parts for modelling abstinence over the interval and proportion of drinking days when drinking occurs and one part for intensity of drinking. We defined the intensity of drinking as the average number of drinks per drinking day during the interval, not the average number of drinks per day, which is a subtle but important distinction. The latter dilutes the intensity of drinking by dividing the total number of drinks by the number of days, and not by the number of drinking days, which may be much lower. Consider subject A who drinks lightly (two drinks) every day of the week; for this subject, we jointly model the number of drinking days (7) and the average of two drinks per drinking day. For subject B, who drinks heavily during the weekend (seven drinks on Saturday and seven on Sunday) and abstains for the rest of the week, we model the number of drinking days (2) and the average of seven drinks per drinking day. Previous literature using average drinks per day (DeSantis *et al.*, 2013; Zhu *et al.*, 2017) would solely model an average of 2 in both cases. Distinguishing between the two drinking patterns of subjects A and B is important because these two types of individual could have different future health and psychosocial outcomes, and treatments may be associated with different drinking patterns.

The paper is organized as follows. We first describe the COMBINE clinical trial data that motivated our work (Section 2), and then in Section 3 we present in detail the separate models and the proposed joint models that we used to analyse these data. Section 4 contains the results that were obtained from fitting these models to the COMBINE study data. A

simulation study investigating differences between joint and separate models in terms of parameter estimate bias, mean-squared error MSE and observed coverage rates of the 95% credible intervals (CIs) are presented in Section 5. Lastly, we briefly discuss the proposed joint modelling approach, its benefits and limitations.

2. Motivating data description

The goal of the COMBINE study (Anton *et al.*, 2006) was to assess the efficacy of two medications (naltrexone and acamprosate), behavioural therapy (combined behavioural intervention (CBI)), and their combinations for treatment of alcohol dependence. As part of the study, $N = 1383$ recently abstinent subjects diagnosed with alcohol dependence were randomized to one of nine treatment arms: eight treatment arms ($N = 1226$ subjects) corresponding to a 2 (naltrexone or placebo–naltrexone) $\times$ 2 (acamprosate or placebo–acamprosate) $\times$ 2 (CBI or not-CBI) factorial design, and a ninth treatment arm ($N = 157$, not included in primary analyses) receiving CBI alone (no pills). Self-reported drinking data (the number of ‘standard’ drinks) was recorded daily during the 16 weeks of treatment for all subjects. The effects of interest were the effects of the three therapies: naltrexone, acamprosate and CBI, alone and in combination, on two co-primary drinking outcomes: monthly percentage day abstinent (PDA) from drinking and time to first heavy drinking day during the 16 weeks of treatment. The primary analyses in Anton *et al.* (2006) revealed that naltrexone interacted with CBI in its effect on PDA (naltrexone’s benefit was apparent only in the absence of CBI) and reduced the hazard of a heavy drinking day.

To simplify models, we restrict analyses to the subset of enrolled subjects who did not receive CBI ($N = 607$) and focus on the effect of naltrexone on drinking during the last 8 weeks of treatment (weeks 9–16). A ‘grace period’ (a period at the beginning of alcoholism trials during which drinking is not considered, such as the first 8 weeks) has been suggested for alcoholism trials as treatment effects during this period may not represent the full potential of the drug (De Beaurepaire, 2012; Falk *et al.*, 2010). For each subject $i = 1, \dots, N$, the outcomes of interest, measured at the week level, were

- a. a discrete measure of drinking frequency, the number of drinking days per week j , denoted by Y_{ij} (ranging from 0 to at most 7), with $Y_{ij} = 0$ representing abstinence during week j , and
- b. a continuous measure of drinking intensity, the average number of standard drinks per drinking day during week j , denoted by W_{ij} (which is only defined if any drinking was reported during that week, i.e. if $Y_{ij} \geq 1$).

These two outcomes were computed for each of the 8 weeks under consideration. (Note that there is a simple correspondence between our frequency outcome and PDA, one of the primary outcomes of the COMBINE study, as PDA equal to $100 - \text{percentage drinking days}$.)

3. Methods

3.1. Hurdle binomial model

When binomial data exhibit more 0s than would be expected under the binomial distribution, the hurdle binomial model HUB can be used to accommodate the extra 0s. The hurdle binomial model is a two-part model: the first part models the probability of a zero value (probability of not crossing the ‘hurdle’), whereas the second part models the non-zero values as arising from a zero-truncated binomial distribution (i.e. a binomial distribution whose support excludes zero values). The probability mass function of a random variable Y with a hurdle binomial distribution $\text{HUB}(\pi, n, p)$ can be written as

$$f_{\text{HUB}}(y) = \pi I(y = 0) + (1 - \pi) \frac{f(y)}{1 - f(0)} I(y = 1, \dots, n),$$

where π is the probability of a zero response, $f(y)$ denotes the probability mass function of a binomial distribution $\text{Bin}(n, p)$ with n trials and probability of success p ,

$$f(y) = \binom{n}{y} p^y (1 - p)^{n - y},$$

and $f(y)/\{1 - f(0)\}$ is the corresponding probability mass function of the zero-truncated binomial distribution (denoted by $\text{TBin}(n, p)$).

In the presence of longitudinal data, random effects can be incorporated in the hurdle binomial model to deal with subject heterogeneity in intercept, slope or both. Suppose that, for each subject $i = 1, \dots, N$, we have available repeated hurdle binomial measurements Y_{ij} (with parameters n_{ij} , p_{ij} and π_{ij}), where $j = 1, \dots, t_i$ denotes the observation time, and that measurements on different subjects are independent. Assuming a logit link for both parts of the model as a function of available covariates, the hurdle binomial model becomes

$$\begin{aligned} Y_{ij} | \mathbf{u}_i &\sim \text{HUB}(\pi_{ij}, n_{ij}, p_{ij}), \quad (1) \\ \text{logit}(\pi_{ij}) &= \mathbf{x}_{ij}^T \boldsymbol{\beta}_0 + \mathbf{z}_{ij}^T \mathbf{u}_{i0}, \\ \text{logit}(p_{ij}) &= \mathbf{x}_{ij}^T \boldsymbol{\beta}_1 + \mathbf{z}_{ij}^T \mathbf{u}_{i1}, \\ \mathbf{u}_i &= (\mathbf{u}_{i0}^T, \mathbf{u}_{i1}^T)^T \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_{01}). \end{aligned}$$

The first logit models the probability of a zero response ($\pi_{ij} = P(Y_{ij} = 0)$), conditionally on the random effects $\mathbf{u}_{i0}$ of subject i , as a function of the subject’s covariate vector for fixed effects, $\mathbf{x}_{ij}$, the covariate vector for random effects, $\mathbf{z}_{ij}$, and the regression coefficient vector $\boldsymbol{\beta}_0$. In the case that the hurdle has been passed (i.e. $Y_{ij} > 0$), the second logit models the probability of success, p_{ij} corresponding to the truncated binomial distribution with n_{ij} trials, conditionally on the random effects $\mathbf{u}_{i1}$ of subject i . We assume that responses within a subject i are independent given the random effects $\mathbf{u}_i$, which follow a multivariate normal distribution with unknown variance–covariance matrix $\mathbf{D}_{01}$. The random effects introduce

correlations between repeated measurements on the same subject and between the two parts of the hurdle binomial model. In the simplest case, a random-intercept-only model, $\mathbf{z}_{ij} = 1$, and $\mathbf{u}_{i0}$ and $\mathbf{u}_{i1}$ are scalars. Usually $\mathbf{u}_{i0}$ and $\mathbf{u}_{i1}$ are vectors of random effects.

The likelihood function L of the hurdle binomial model can be written as a product of the likelihood contribution L_i of each subject i , $L = \prod_{i=1}^N L_i$, where

$$L_i = \int \left\{ \prod_{j=1}^{t_i} f_{\text{HUB}}(y_{ij} | \mathbf{u}_i) \right\} f_U(\mathbf{u}_i) d\mathbf{u}_i$$

$$= \int \left[\prod_{j=1}^{t_i} \left\{ \pi_{ij} I(y_{ij} = 0) + (1 - \pi_{ij}) \frac{\binom{n_{ij}}{y_{ij}} p_{ij}^{y_{ij}} (1 - p_{ij})^{n_{ij} - y_{ij}}}{1 - (1 - p_{ij})^{n_{ij}}} I(y_{ij} \geq 1) \right\} \right] f_U(\mathbf{u}_i) d\mathbf{u}_i,$$

where f_U represents the normal density of the random effects $\mathbf{u}_i$.

In the COMBINE data, we assume that Y_{ij} , the number of drinking days per week j for subject i , follows a hurdle binomial distribution (with Y_{ij} ranging from 0 to the number of days in week j for which subject i has provided drinking data, n_{ij}). This distribution has two parts: a logistic part for modelling weekly abstinence, and a truncated binomial part for the frequency of drinking when drinking occurs. Count distributions could have been used instead of the binomial (e.g. negative binomial) distribution. However, since the outcome is a count with an upper bound, the hurdle binomial may be more appropriate than standard count distributions that have an unbounded support.

3.2. Log-normal model

A continuous random variable W is said to have a log-normal distribution with parameters μ and σ^2 (denoted by $\text{LN}(\mu, \sigma^2)$) if $\log(W)$ is normally distributed with mean μ and variance σ^2 . The log-normal distribution is commonly used for modelling outcomes that take only positive values (e.g. costs). The probability density function of the log-normal distribution is

$$f_{\text{LN}}(w) = \frac{1}{w\sigma\sqrt{2\pi}} \exp\left[-\frac{\{\log(w) - \mu\}^2}{2\sigma^2}\right] I(w > 0).$$

With random effects added to accommodate repeated measurements on the same subject, the model for the j th ($j = 1, \dots, t_i$) log-normal observation W_{ij} on subject $i = 1, \dots, N$ can be written as

$$W_{ij} | \mathbf{u}_{i2} \sim \text{LN}(\mu_{ij}, \sigma^2), \quad (2)$$

$$\mu_{ij} = \mathbf{x}_{ij}^T \beta_2 + \mathbf{z}_{ij}^T \mathbf{u}_{i2},$$

$$\mathbf{u}_{i2} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_2).$$

The vector β_2 contains the regression coefficients corresponding to covariate vector $\mathbf{x}_{ij}$ and $\mathbf{u}_{i2}$ represents the normally distributed random-effect vector for subject i .

The likelihood function L of the log-normal model is again a product of the likelihood contributions L_i of all subjects, $L = \prod_{i=1}^N L_i$, with

$$\begin{aligned} L_i &= \int \left\{ \prod_{j=1}^{t_i} f_{\text{LN}}(w_{ij} | \mathbf{u}_{i2}) \right\} f_U(\mathbf{u}_{i2}) d\mathbf{u}_{i2} \\ &= \int \left(\prod_{j=1}^{t_i} \frac{1}{w_{ij} \sigma \sqrt{2\pi}} \exp \left[-\frac{\{\log(w_{ij}) - \mu_{ij}\}^2}{2\sigma^2} \right] \right) f_U(\mathbf{u}_{i2}) d\mathbf{u}_{i2}, \end{aligned}$$

where f_U denotes the density of random effect $\mathbf{u}_{i2}$.

In the COMBINE data, W_{ij} represents the average number of drinks per drinking day in week j for subject i .

3.3. Joint hurdle binomial–log-normal model

We jointly model a longitudinal discrete binomial outcome with an excess of 0s, Y —the weekly number of drinking days in the COMBINE data—and a longitudinal continuous positive outcome W —the weekly average number of drinks per drinking day—using a joint hurdle binomial–log-normal model HUBLN that brings together the two separate models in equations (1) and (2):

$$\begin{aligned} Y_{ij} | \mathbf{u}_i &\sim \text{HUB}(\pi_{ij}, n_{ij}, p_{ij}), \\ \text{logit}(\pi_{ij}) &= \mathbf{x}_{ij}^T \beta_0 + \mathbf{z}_{ij}^T \mathbf{u}_{i0}, \\ \text{logit}(p_{ij}) &= \mathbf{x}_{ij}^T \beta_1 + \mathbf{z}_{ij}^T \mathbf{u}_{i1}, \\ W_{ij} | Y_{ij} \geq 1, \mathbf{u}_i &\sim \text{LN}(\mu_{ij}, \sigma^2), \mu_{ij} = \mathbf{x}_{ij}^T \beta_2 + \mathbf{z}_{ij}^T \mathbf{u}_{i2}, \\ \mathbf{u}_i &= (\mathbf{u}_{i0}^T, \mathbf{u}_{i1}^T, \mathbf{u}_{i2}^T)^T \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \end{aligned} \quad (3)$$

where, as before, $i = 1, \dots, N$ indicates subject and $j = 1, \dots, t_i$ indicates time point. The drinking intensity component W_{ij} is conditional on drinking occurring during week j ($Y_{ij} \geq 1$). We assume that the random effects for subject i , $\mathbf{u}_i$, have a normal distribution with mean $\mathbf{0}$ and variance–covariance matrix $\mathbf{D}$. The elements of the matrix $\mathbf{D}$ represent the association between the random effects corresponding to the three parts of the model (zero, truncated binomial and log-normal). Note that modelling the hurdle binomial and log-normal parts separately would amount to assuming zero correlation between the hurdle–binomial random effects ($\mathbf{u}_{i0}$, $\mathbf{u}_{i1}$) and the log-normal effect, $\mathbf{u}_{i2}$. For ease of exposition, we include the same set of covariates $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ in all three parts, although differing sets can be readily accommodated.

The likelihood of the joint model is $L = \prod_{i=1}^N L_i$, where L_i the likelihood for the i th subject, equals

$$\int \left\{ \prod_{j=1}^{t_i} f_{\text{HUB}}(y_{ij} | \mathbf{u}_i) f_{\text{LN}}(w_{ij} | u_i)^{I(y_{ij} \geq 1)} \right\} f_U(\mathbf{u}_i) d\mathbf{u}_i,$$

with f_U denoting the random-effect distribution (multivariate normal density).

3.4. Model inference

To accommodate the model complexity due to the number of random effects—six random effects would be needed in total in the vector $\mathbf{u}_i$ if a random intercept and slope were used for each of the three parts of the joint model—we use Bayesian inference. In the Bayesian framework, all parameters are considered random (with specified prior distributions) so random effects are easily accommodated. We can use prior knowledge about the parameters, if available, or non-informative priors. Also, by basing inferences on MCMC samples from the posterior distribution, the Bayesian approach avoids the need for asymptotic approximations.

To evaluate and maximize the likelihood function in a frequentist maximum likelihood estimation approach, repeated multi-dimensional integrations over the random effects would be needed. Since these integrals have no closed form expressions, they must be approximated by using numerical, stochastic or analytical approximations. Analytical approximations may provide biased results especially when the data are discrete. Stochastic approximations are possible but would need to be programmed from scratch as they are not readily available in software programs. Numerical techniques (e.g. adaptive Gauss–Hermite quadrature) are computationally unfeasible for complex models, in particular when there are more than four or five random effects. However, for smaller numbers of random effects, numerical approximations such as those implemented in SAS PROC NLMIXED (Littel *et al.*, 1996) can be used for statistical inference. Thus we compared the Bayesian approach with non-informative priors and the frequentist approach based on numerical approximations on a simpler random-intercept-only model (with three random effects in total).

Bayesian models were fitted by using WinBUGS (Lunn *et al.*, 2000) version 1.4.3 and the R package R2WinBUGS (Sturtz *et al.*, 2005). We checked convergence of the MCMC chains by using trace and auto-correlation plots (provided by the coda package in R). For each model fit on the COMBINE data, we computed estimates based on 10 000 iterations (after discarding 10 000 iterations as burn-in). We used non-informative priors for model parameters: independent normal with mean 0 and large variance (1000) for regression coefficients (elements of β_0 , β_1 and β_2) and a gamma(0.001, 0.001) distribution (with mean 1 and variance 1000) for the precision $1/\sigma^2$ of the log-normal distribution. For the variance–covariance matrix of the random effects ($\mathbf{D}$, with dimension $\dim(\mathbf{D})$), we used an inverse Wishart prior

$$\mathbf{D} \sim \text{InvWishart}(\mathbf{R}, \text{df})$$

with scale matrix $\mathbf{R} = \mathbf{I}$ (the $\dim(\mathbf{D}) \times \dim(\mathbf{D})$ identity matrix) and degrees of freedom $df = \dim(\mathbf{D}) + 1$ (Gelman and Hill, 2006), i.e. the precision matrix $\mathbf{E} = \mathbf{D}^{-1}$ was assumed to have a Wishart distribution: $\mathbf{E} \sim \text{Wishart}(\mathbf{R} = \mathbf{I}, df)$, with density

$$f_E(\mathbf{E}) \propto |\mathbf{R}|^{df/2} |\mathbf{E}|^{\{df - \dim(\mathbf{D}) - 1\}/2} \exp\left\{-\frac{1}{2} \text{tr}(\mathbf{R}\mathbf{E})\right\},$$

where tr denotes the trace function.

The simple frequentist model with random intercepts was fitted in SAS PROC NLMIXED (version 9.4) using adaptive Gaussian quadrature with 10 quadrature points for each dimension of integration and relative gradient convergence criterion $GCONV = 1 \times 10^{-12}$.

We used posterior predictive checks (Gelman and Hill, 2006; Gelman *et al.*, 1996) to evaluate the fit of the joint model. We generated 10000 replicated data sets from the posterior predictive distribution (one data set corresponding to the parameter estimates from every 10th MCMC iteration out of the total 100000 iterations). We compared the replicated data with the observed data in terms of four statistics: the mean and standard deviation (SD) of the number of abstinent weeks across subjects (i.e. the mean and SD of

$\sum_{j=1}^t I(Y_{ij} = 0)$, $i = 1, \dots, N$), and mean and SD across subjects of drinking intensity on the

log-scale (the mean and SD of $\left\{1/\sum_{j=1}^t I(Y_{ij} \geq 1)\right\} \sum_{j=1}^t I(Y_{ij} \geq 1) \log(W_{ij})$). If the statistic

based on the observed data, T_{obs} , is not too extreme relative to the distribution of the statistic based on the replicated data, $T_{\text{rep}}, \text{rep}=1, \dots, 10000$ (i.e. if $P = P(T_{\text{rep}} > T_{\text{obs}})$, is not near 0 or 1), then the model adequately fits the aspect of the data represented by that statistic.

Access to the COMBINE data, which is a controlled access data set, can be obtained by filling out the application forms available at <https://www.niaaa.nih.gov/research/major-initiatives/combine-study>. The programs that were used to analyse the data can be obtained from <http://wileyonlinelibrary.com/journal/rss-datasets>

4. Results

Out of 607 COMBINE trial participants who did not receive CBI, 29 (5%) had no data during the last 8 weeks of treatment (weeks 9–16) and were excluded from all analyses. That left 578 subjects for analyses, out of which 287 (50%) received naltrexone and the rest received placebo–naltrexone (which we call ‘placebo’ for brevity). The data on these subjects were fairly complete: they provided a total of 4551 weeks of daily drinking data out of the planned $8 \times 578 = 4624$ weeks, and 99.5% of these weeks had drinking data available for each day of the week.

Fig. 1 displays distribution plots for the two drinking outcomes by treatment group (naltrexone or placebo): the number of drinking days and the average number of drinks per drinking day during the last week of the trial (week 16). Both outcomes have highly skewed distributions, with almost half of subjects (55% in the naltrexone group and 43% in the placebo group) reporting no drinking days during this week. Among those who drank during

the week, the mean number of drinking days was slightly higher in the placebo group (3.96) than in the naltrexone group (3.51). The subject level average number of drinks per drinking day had a median (with interquartile range IQR in parentheses) of 6.1 (3.1–9.6) in the naltrexone group and 6.4 (4.3–9.92) in the placebo group.

Fig. 2 presents, for each week from week 9 to 16 and by treatment group, the proportion of subjects who were abstinent (subjects with zero drinking days), the mean drinking days among those who drank and the mean drinks per drinking day (mean of subject level average drinks per drinking day). Subjects receiving naltrexone appear to have consistently better outcomes over time.

We first fitted a simpler joint hurdle binomial–log-normal model to the COMBINE drinking data that included, in each of the three parts, fixed effects for treatment group and week and a random intercept:

$$\begin{aligned} Y_{ij} | \mathbf{u}_i &\sim \text{HUB}(\pi_{ij}, n_{ij}, p_{ij}), \\ \text{logit}(\pi_{ij}) &= \beta_{00} + u_{i00} + \beta_{01} \text{nalt}_i + \beta_{02} \text{week}_{ij}, \\ \text{logit}(p_{ij}) &= \beta_{10} + u_{i10} + \beta_{11} \text{nalt}_i + \beta_{12} \text{week}_{ij}, \\ W_{ij} | Y_{ij} \geq 1, \mathbf{u}_i &\sim \text{LN}(\mu_{ij}, \sigma^2), \\ \mu_{ij} &= \beta_{20} + u_{i20} + \beta_{21} \text{nalt}_i + \beta_{22} \text{week}_{ij}, \\ \mathbf{u}_i &= (u_{i00}, u_{i10}, u_{i20})^T \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \end{aligned} \quad (4)$$

where Y_{ij} is the number of drinking days for subject i during week ij , W_{ij} is the average number of drinks per drinking day consumed by subject i during week ij , nalt_i is a binary indicator of whether subject i was randomized to receive naltrexone, and u_{ik0} is the random intercept corresponding to subject i and part k of the model ($k = 0, 1, 2$). For the intercept to be more interpretable, the week variable was centred at the first week of interest (week 9). We initially also considered a naltrexone-by-week interaction term, but since there was no indication that it was statistically significant we focused on the simpler model with just main effects.

We then fitted a more complex joint hurdle binomial–log-normal model, with both random intercept and slope in each of the three parts (six random effects in total), i.e.

$$\begin{aligned} Y_{ij} | \mathbf{u}_i &\sim \text{HUB}(\pi_{ij}, n_{ij}, p_{ij}), \\ \text{logit}(\pi_{ij}) &= \beta_{00} + u_{i00} + \beta_{01} \text{nalt}_i + (\beta_{02} + u_{i01}) \text{week}_{ij}, \\ \text{logit}(p_{ij}) &= \beta_{10} + u_{i10} + \beta_{11} \text{nalt}_i + (\beta_{12} + u_{i11}) \text{week}_{ij}, \\ W_{ij} | Y_{ij} \geq 1, \mathbf{u}_i &\sim \text{LN}(\mu_{ij}, \sigma^2), \\ \mu_{ij} &= \beta_{20} + u_{i20} + \beta_{21} \text{nalt}_i + (\beta_{22} + u_{i21}) \text{week}_{ij}, \\ \mathbf{u}_i &= (u_{i00}, u_{i01}, u_{i10}, u_{i11}, u_{i20}, u_{i21})^T \sim \mathcal{N}(\mathbf{0}, \mathbf{D}), \end{aligned} \quad (5)$$

where we used the same notation as in model (4). In addition, u_{ik1} now represents the random slope of subject i for the k th part of the model ($k = 0, 1, 2$).

Table 1 contains results from fitting the random-intercept-only joint model (4) in SAS PROCNLMIXED (frequentist maximum likelihood estimation approach) and WinBUGS (Bayesian approach). The results from the two approaches were very similar, although the computational time was considerably longer for SAS than for WinBUGS (about 10 h by using 10 quadrature points for SAS *versus* 1.5 h for WinBUGS on an Intel central processor unit E5620 with two processors of 2.4 GHz each). Since the two sets of results agreed, we focus below on interpreting regression coefficients from the Bayesian model, noting that odds ratios OR have conditional (subject-specific) interpretations, comparing probabilities conditional on the random effects (holding the random effects constant). Results showed evidence of a naltrexone effect on all three parts: abstinence, frequency and intensity of drinking during the last 8 weeks of treatment. The estimates from the hurdle binomial component showed that naltrexone was associated with increased odds of weekly abstinence from drinking (OR = 3.33 and 95% CI (1.40, 7.58) for the effect of naltrexone in the zero part of the HUB model) and significantly decreased probability of daily drinking among those who drank (OR = 0.42 and 95% CI (0.25, 0.74) in the truncated binomial part of HUB). Additionally, there was evidence from the log-normal model of drinking intensity that naltrexone was associated with 18% (95% CI (6%, 29%)) lower mean weekly average number of drinks per drinking day. At the same time, there was not much evidence of a time (week) effect on either frequency or intensity of drinking. There was a moderate positive correlation (0.37) between the random intercepts corresponding to the truncated binomial part (modelling non-zero drinking frequency) and the log-normal part (modelling drinking intensity), which indicates that, among subjects who drank, those with higher frequency of drinking also showed a higher intensity of drinking.

Estimates obtained from fitting the more complex joint model (5) (both random intercept and slope in each of the three parts of the model) to the COMBINE data are presented in Table 2. The estimates of the fixed effects largely agreed with the estimates from the intercept-only models in Table 2. Naltrexone significantly increased the odds of no weekly drinking (OR = 3.57; 95% CI (1.40, 8.47)) and significantly decreased the probability of daily drinking among those who drank OR = 0.45; 95% CI (0.26, 0.77)). Naltrexone was also associated with 16% (95% CI (4%, 27%)) fewer drinks per drinking day. As before, the slope fixed effect (week) was not significant in any of the three parts of the model. However, there was substantial variability in slopes across subjects in all three parts of the model ($SD(u_{j01}) = 0.43$, $SD(u_{j11}) = 0.23$ and $SD(u_{j21}) = 0.10$).

For the random-intercept and slopes joint model (5), the posterior predictive distributions of the four chosen statistics (mean and SD of the number of abstinent weeks across subjects, and mean and SD of drinking intensity across subjects (average drinks per drinking day on the log-scale)) are displayed in Fig. 3, with the vertical broken lines representing the observed values of the test statistics. The observed values were within the bulk of the distribution, with the proportion of the replicated values falling above the observed value ranging from 0.43 to 0.72. This indicates that these four statistics seem to be captured reasonably well by the model.

5. Simulation study

To examine the benefits of using a joint model, we performed a simulation study to compare results from the joint model with results from fitting separate models to the hurdle binomial and log-normal components. We used the estimates that were obtained from the COMBINE data (estimates from Table 1) to simulate 100 data sets from a joint hurdle binomial–log-normal model with random intercepts only, under three sample size scenarios ($N = 500$, $N = 200$ and $N = 50$ subjects), with outcomes measured over 4 weeks for all subjects.

For models fitted on the simulated data, estimates were based on 50000 MCMC iterations (after discarding 5000 as burn-in). We used normal priors with mean 0 and variance 1000 for the regression coefficients of all models. For both the joint model and the separate log-normal model, $\text{gamma}(0.001, 0.001)$ (with mean 1 and variance 1000) was assumed as a prior for the precision $1/\sigma^2$ of the log-normal distribution. The variance–covariance matrix of the random effects, $\mathbf{D}$, had an inverse Wishart prior with identity scale matrix and degrees of freedom $\text{df} = 4$ for the joint model and $\text{df} = 3$ for the separate HUB model. For the separate log-normal model, the precision of the random effects was assumed to have a $\text{gamma}(0.001, 0.001)$ prior.

We monitored the posterior mean estimates from the joint and separate models (separately fitting the hurdle binomial and log-normal models), mean bias of these estimates, their MSE, and the coverage rate of the 95% CIs (the proportion of intervals that contain the true value) across the 100 simulated data sets. The programs that were used to run the simulation can be obtained from <http://wileyonlinelibrary.com/journal/rss-datasets>. The simulation results are summarized in Table 3.

For the hurdle binomial component, the results from the joint and separate models were similar in terms of bias and MSE, with greater bias and MSE apparent as the sample size decreased. For the log-normal component, the estimate of the intercept term had lower absolute bias in the joint model than in the separate model (e.g. 0.008 *versus* 0.075 for $N = 500$). For the joint model, coverage rates were relatively close to 0.95. In the separate model, there was a noticeable breakdown in the performance of the CI for estimating the intercept of the log-normal component (for $N = 500$, a coverage rate of 0.72 for the separate model *versus* 0.93 for the joint model).

The problems with separately estimating the log-normal (drinking intensity) part arise because the hurdle binomial component modelling frequency of drinking contributes information about the number of repeated measurements (cluster size) that are available for the log-normal component. The log-normal outcome (average drinks per drinking day) is defined only when the frequency of drinking is non-zero, which is information which is being ignored when the hurdle binomial and the log-normal models are modelled separately. This leads to an ‘informative cluster size’ situation (Zhang *et al.*, 2017): subjects who drink more will have more repeated measurements of average drinks per drinking day, and they will be overrepresented in the log-normal part, leading to positive bias.

These simulation results are similar to findings from Su *et al.* (2009), who investigated the bias in analysing separately the two parts of a two-part mixed model for longitudinal

semicontinuous data (data consisting of a mixture of 0s and continuous positive values), and Neelon *et al.* (2013), who investigated the consequences of using a model with correlated spatial effects on a model with separate spatial effects for the two components of a spatial Poisson hurdle model. Both these studies found that the regression coefficient estimates of the second (non-zero) component of models were most affected by separate modelling.

6. Discussion

In this paper we proposed a method for simultaneous modelling longitudinal data on abstinence, frequency and intensity of drinking. The model has three parts: a logistic part for sustained abstinence over the period of interest, a truncated binomial part for drinking frequency and a log-normal model for drinking intensity when drinking occurs. Correlated random effects were used to relate these outcomes and the repeated measurements over time within a subject. We used both a Bayesian and a frequentist approach to fit these joint models by using readily available statistical software (WinBUGS and SAS PROC NLMIXED), but the frequentist approach is for now limited to simpler cases with small numbers of random effects (fewer than 5). We thus could not fit the joint model with both random intercepts and slopes in each part (six random effects in total) by using PROC NLMIXED. Because misspecification of the random-effects structure can lead to bias in the regression parameter estimates in random-effects models for non-normal data (Gueorguieva, 2005; Grigorova and Gueorguieva, 2016), we recommend the Bayesian approach for this model because it can seamlessly incorporate a larger number of random effects.

We used a small simulation to investigate the advantages of using a joint model rather than a separate model and we found that joint modelling is particularly beneficial for estimating the coefficients of the drinking intensity component (especially the intercept) as it resulted in lower bias than separate modelling. Other advantages of the joint modelling approach include the ability to use all the information about drinking in the same model and to estimate the association between the different drinking measures.

We illustrated the proposed modelling approach on data from a clinical trial of alcoholism treatments (COMBINE). We found evidence that the drug naltrexone increased the odds of abstinence and decreased the probability of daily drinking among those who drank. Naltrexone was also associated with fewer drinks per drinking day. Although these results complement the results from the primary analysis of the trial (Anton *et al.*, 2006), the analysis that is presented in this paper is illustrative only and not meant to be definitive, as the models did not adjust for all relevant variables (such as other treatment variables, randomization site and baseline drinking measures). We illustrated the method on the COMBINE data, but this model could be easily applied to other data sets from alcoholism or substance use trials.

Although the data were available at the day level, we decided to conduct the analysis at the week level because we lacked information on day of the week (weekend *versus* weekday), which is an important predictor of drinking (Maggs *et al.*, 2011; Huh *et al.*, 2015). Also, prior literature seems to indicate that day-to-day timeline follow-back values do not correspond as well to realtime recordings (e.g. daily electronic interviews) as composite

measures over an interval (Carney *et al.*, 1998). Rather than focusing on single-day abstinence, our method, by using data aggregated over an interval of time, allows us to assess whether abstinence was sustained over a period of time (such as a week). This is more meaningful in treatment studies as individuals who can maintain abstinence over a time period have better long-term outcomes than individuals who use more frequently (Weisner *et al.*, 2003); for example, in the literature on treatments for cocaine use disorder, attainment of periods of abstinence of 3 or more weeks is associated with better long-term outcomes (Higgins *et al.*, 2000). However, our method can be easily changed to accommodate any time periods of interest besides weeks (e.g. days or months). For analysis at the day level, the hurdle binomial distribution for the number of drinking days would degenerate to a simpler Bernoulli distribution for whether the subject drank each day, i.e. the day level analysis is a special case of our analysis and can be obtained by removing the truncated binomial part from the three-part model. The analysis at the day level would be similar to the NCI method of Tooze *et al.* (2006), except for the fact that the NCI method uses only random intercepts in the two parts (no random slopes) and uses the Box–Cox transformation for the non-zero values which can make results more difficult to interpret. We used a more interpretable log-normal distribution for the non-zero values but, if needed for other applications, this can easily be replaced with more flexible distributions for positive outcomes such as gamma distributions.

An alternative to the hurdle binomial model would be a zero-inflated binomial model. However, we chose the hurdle model since the zero-inflated model has been noted to give unreliable parameter estimates in certain situations (Min and Agresti, 2005; Buu *et al.*, 2012). Furthermore, as opposed to the zero-inflated model, the hurdle model can deal with zero deflation as well as with zero inflation in the data and, by separately modelling the zeros from the non-zeros, provides results that are easier to interpret. Also, in theory, a zero-inflated model assumes that, at each time point, the population can be divided into two distinct latent groups: a group not at risk of having the event and another group at risk (who may or may not have the event); a hurdle model is more appropriate when the entire population is at risk, with the hurdle being crossed when the event occurs (Buu *et al.*, 2012). For the COMBINE sample, consisting of alcohol-dependent subjects, the hurdle assumption of all subjects being at risk of drinking seems more reasonable.

There are related areas of future work that could be considered. More flexible priors could be used for the variance–covariance matrix of the random effects instead of the inverse Wishart prior, such as the scaled inverse Wishart prior (Gelman and Hill, 2006), which is a prior that is not yet built into standard statistical software. Since our study had a large sample size, we do not expect the choice of the uninformative prior distribution to affect the results substantially. However, in small studies one should carefully assess sensitivity of results to changes in the prior distribution. To accommodate possible overdispersion in the binomial component due to the fact that the drinking on a certain day of the week is not independent of drinking on other days of the week, the model could include an additional random effect at the subject–occasion level or a beta–binomial distribution. Because of the computational time demands of fitting these models, our simulation study was small; future work on decreasing computation time and larger simulations studies may be needed. Additional work could explore methods for dealing with missing data under the joint

modelling framework (which is an issue that is especially important to address in case of informative missing data (Albert, 2000), e.g. missing drinking data for periods of time when the subject is heavily drinking). Finally, further work could be done on loosening the assumption of normality of the random effects (Kleinman and Ibrahim, 1998) or log-normality of drinking intensity (Xing *et al.*, 2017).

The joint model that was developed in this paper is useful for modelling data from alcoholism trials where interest lies in estimating abstinence, frequency and intensity of drinking over time and/or their relationship with certain predictors and each other. At the same time, this approach easily extends to analysing regularly collected longitudinal data in other fields of study such as smoking or substance use research.

Funding

This work was supported by the National Institute on Alcohol Abuse and Alcoholism (grant P50 AA012870) and by the National Center for Advancing Translational Sciences (grant 5UL1T R000142–08). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Alcohol Abuse and Alcoholism, the National Center for Advancing Translational Sciences or the National Institutes of Health. The sponsor had no role in the study design, in the collection, analysis and interpretation of data, in the writing of the report or in the decision to submit the paper for publication.

References

- Albert PS (2000) A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics*, 56, 602–608. [PubMed: 10877323]
- Anton RF, O'Malley SS, Ciraulo DA, Cisler RA, Couper D, Donovan DM, Gastfriend DR, Hosking JD, Johnson BA, LoCastro JS, Longabaugh R, Mason BJ, Mattson ME, Miller WR, Pettinati HM, Randall CL, Swift R, Weiss RD, Williams LD, Zweben A and COMBINE Study Research Group (2006) Combined pharmacotherapies and behavioral interventions for alcohol dependence: the COMBINE study; a randomized controlled trial. *J. Am. Med. Ass.*, 295, 2003–2017.
- Buu A, Li R, Tan X and Zucker RA (2012) Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Statist. Med.*, 31, 4074–4086.
- Carney MA, Tennen H, Affleck G, Del Boca FK and Kranzler HR (1998) Levels and patterns of alcohol consumption using timeline follow-back, daily diaries and real-time “electronic interviews”. *J. Stud. Alchl.*, 59, 447–454.
- Daniels MJ and Normand ST (2006) Longitudinal profiling of health care units based on continuous and discrete patient outcomes. *Biostatistics*, 7, 1–15. [PubMed: 15917373]
- Das S, Harel O, Dey DK, Covault J and Kranzler HR (2010) Analysis of extreme drinking in patients with alcohol dependence using Pareto regression. *Statist. Med.*, 29, 1250–1258.
- De Beaurepaire R (2012) Suppression of alcohol dependence using baclofen: a 2-year observational study of 100 patients. *Front. Psychiatr.*, 3, article 103.
- De Leon AR and Chough KC (2013) *Analysis of Mixed Data*. Boca Raton: Chapman and Hall–CRC.
- DeSantis SM, Bandyopadhyay D, Baker NL, Randall PK, Anton RF and Prisciandaro JJ (2013) Modeling longitudinal drinking data in clinical trials: an application to the combine study. *Drug Alchl Depend.*, 132, 244–250.
- Dunson DB (2000) Bayesian latent variable models for clustered mixed outcomes. *J. R. Statist. Soc. B.*, 62, 355–366.
- Falk D, Wang XQ, Liu L, Fertig J, Mattson M, Ryan M, Johnson B, Stout R and Litten RZ (2010) Percentage of subjects with no heavy drinking days: evaluation as an efficacy endpoint for alcohol clinical trials. *Alchlism Clin. Exptl Res.*, 34, 2022–2034.
- Gelman A and Hill J (2006) *Data Analysis using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

- Gelman A, Meng X-L and Stern H (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sin.*, 6, 733–760.
- Grigorova D and Gueorguieva R (2016) Correlated probit analysis of repeatedly measured ordinal and continuous outcomes with application to the health and retirement study. *Statist. Med.*, 35, 4202–4225.
- Gueorguieva RV (2005) Comments about joint modeling of cluster size and binary and continuous subunit-specific outcomes. *Biometrics*, 61, 862–866. [PubMed: 16135040]
- Gueorguieva RV and Agresti A (2001) A correlated probit model for joint modeling of clustered binary and continuous responses. *J. Am. Statist. Ass.*, 96, 1102–1112.
- Higgins ST, Wong CJ, Badger GJ, Haug Ogden D and Dantona RL (2000) Contingent reinforcement increases cocaine abstinence during outpatient treatment and 1 year of follow-up. *J. Consltng Clin. Psychol.*, 68, 64–72.
- Huh D, Kaysen DL and Atkins DC (2015) Modeling cyclical patterns in daily college drinking data with many zeroes. *Multiv. Behav. Res.*, 50, 184–196.
- Kleinman KP and Ibrahim JG (1998) A semi-parametric Bayesian approach to generalized linear mixed models. *Statist. Med.*, 17, 2579–2596.
- Littel R, Milliken G, Stroup W and Wolfinger R (1996) *SAS System for Mixed Models*. Cary: SAS Institute.
- Liu L, Ma JZ and Johnson BA (2008) A multi-level two-part random effects model, with application to an alcohol-dependence study. *Statist. Med.*, 27, 3528–3539.
- Liu L, Strawderman RL, Johnson BA and O’Quigley JM (2016) Analyzing repeated measures semi-continuous data, with application to an alcohol dependence study. *Statist. Meth. Med. Res.*, 25, 133–152.
- Lunn DJ, Thomas A, Best N and Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statist. Comput.*, 10, 325–337.
- Maggs JL, Williams LR and Lee CM (2011) Ups and downs of alcohol use among first-year college students: number of drinks, heavy drinking, and stumble and pass out drinking days. *Addict. Behav.*, 36, 197–202.
- Min Y and Agresti A (2005) Random effect models for repeated measures of zero-inflated count data. *Statist. Modelling*, 5, 1–19.
- Neelon B, Ghosh P and Loebs PF (2013) A spatial Poisson hurdle model for exploring geographic variation in emergency department visits. *J. R. Statist. Soc. A*, 176, 389–413.
- Prisciandaro JJ, DeSantis SM and Bandyopadhyay D (2012) Simultaneous modeling of the impact of treatments on alcohol consumption and quality of life in the COMBINE study: a coupled hidden Markov analysis. *Alchlm Clin. Exptl Res.*, 36, 2141–2149.
- Shirley KE, Small DS, Lynch KG, Maisto SA and Oslin DW (2010) Hidden Markov models for alcoholism treatment trial data. *Ann. Appl. Statist.*, 4, 366–395.
- Sturtz S, Ligges U and Gelman AE (2005) R2WinBUGS: a package for running WinBUGS from R. *J. Statist. Softwr.*, 12, 1–16.
- Su L, Tom BD and Farewell VT (2009) Bias in 2-part mixed models for longitudinal semicontinuous data. *Biostatistics*, 10, 374–389. [PubMed: 19136448]
- Toozee JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, Guenther PM, Carroll RJ and Kipnis V (2006) A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J. Am. Diet. Ass.*, 106, 1575–1587. [PubMed: 17000190]
- Wang S-J, Winchell CJ, McCormick CG, Nevius SE and O’Neill RT (2002) Short of complete abstinence: an analysis exploration of multiple drinking episodes in alcoholism treatment trials. *Alchlm Clin. Exptl Res.*, 26, 1803–1809.
- Weisner C, Ray GT, Mertens JR, Satre DD and Moore C (2003) Short-term alcohol and drug treatment outcomes predict long-term outcome. *Drug Alchl Depend.*, 71, 281–294.
- Xing D, Huang Y, Chen H, Zhu Y, Dagne GA and Baldwin J (2017) Bayesian inference for two-part mixed-effects model using skew distributions, with application to longitudinal semicontinuous alcohol data. *Statist. Meth. Med. Res.*, 26, 1838–1853.

- Zhang B, Liu W, Zhang Z, Qu Y, Chen Z and Albert PS (2017) Modeling of correlated data with informative cluster sizes: an evaluation of joint modeling and within-cluster resampling approaches. *Statist. Meth. Med. Res.* 26, 1881–1895.
- Zhu H, Luo S and DeSantis SM (2017) Zero-inflated count models for longitudinal measurements with heterogeneous random effects. *Statist. Meth. Med. Res.* 26, 1774–1786.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.

Distribution plots of (a), (b) frequency and (c), (d) intensity of drinking during week 16, by treatment group (a), (c) naltrexone and (b), (d) placebo: (a) $N = 279$, proportion zero 0.55, mean (SD) non-zero 3.51 (2.39); (b) $N = 283$, proportion zero 0.43, mean (SD) non-zero 3.96 (2.28); (c) $N = 125$, mean (SD) 7.05 (5.19), median (IQR) 6.1 (3.1–9.6); (d) $N = 160$, mean (SD) 7.88 (6.14), median (IQR) 6.4 (4.3–9.92)

Fig. 2.
(a), (b) Frequency and (c) intensity of drinking over time by treatment group naltrexone (—) or placebo (-----) during weeks 9–16

Fig. 3.

Posterior predictive check of the Bayesian joint model with both random intercepts and slopes (model (5)) for four statistics (a) mean and (b) SD of the number of abstinent weeks across subjects, and (c) mean and (d) SD of drinking intensity across subjects (with drinking intensity for a subject measured as the average drinks per drinking day on the log-scale) ($\square$, distribution of the 10000 replicated values, T_{rep} , from the posterior predictive distribution of the four statistics; $|$, observed value of the statistic T_{obs}): (a) p (proportion of the replicated values exceeding the observed value ($P(T_{\text{rep}} > T_{\text{obs}})) = 0.43$; (b) $p = 0.72$; (c) $p = 0.62$; (d) $p = 0.7$

Table 1.

Parameter estimates and 95% CI (credible or confidence) from fitting the joint model with random intercepts only (model (4)) to the COMBINE data ($N = 578$) in WinBUGS (Bayesian approach) and SAS PROC NLMIXED (frequentist maximum likelihood estimation approach)[†]

Variable	Results for Bayesian approach			Results for frequentist maximum likelihood estimation approach		
	Posterior median	95% CI		Estimate	95% CI	
<i>Hurdle binomial component, zero part</i>						
Intercept ($\exp(\beta_{00})$)	0.63	0.33	1.18	0.62	0.32	1.19
Naltrexone ($\exp(\beta_{01})$)	3.33	1.40	7.58	3.39	1.37	8.37
Week ($\exp(\beta_{02})$)	0.96	0.92	1.01	0.96	0.92	1.01
<i>Truncated binomial part</i>						
Intercept ($\exp(\beta_{10})$)	0.24	0.15	0.37	0.25	0.16	0.39
Naltrexone ($\exp(\beta_{11})$)	0.42	0.25	0.74	0.41	0.23	0.75
Week ($\exp(\beta_{12})$)	1.02	1.00	1.04	1.02	1.00	1.04
<i>Log-normal component</i>						
Intercept ($\exp(\beta_{20})$)	5.60	5.01	6.26	5.61	5.03	6.26
Naltrexone ($\exp(\beta_{21})$)	0.82	0.71	0.94	0.82	0.71	0.94
Week ($\exp(\beta_{22})$)	1.00	0.99	1.01	1.00	0.99	1.01
<i>SD or correlations</i>						
σ	0.36	0.35	0.37	0.36	0.35	0.37
SD(u_{00})	5.00	4.44	5.62	5.00	4.40	5.59
SD(u_{10})	3.20	2.86	3.57	3.19	2.83	3.54
SD(u_{20})	0.69	0.64	0.74	0.68	0.63	0.74
Corr(u_{00}, u_{10})	−0.88	−0.91	−0.84	−0.88	−0.92	−0.84
Corr(u_{00}, u_{20})	−0.22	−0.36	−0.07	−0.22	−0.37	−0.07
Corr(u_{10}, u_{20})	0.37	0.25	0.49	0.38	0.26	0.50

[†]Regression coefficients were exponentiated and represent the OR of abstinence during a week for the zero part, the OR of daily drinking when drinking occurred during the week for the truncated binomial part and mean ratios of drinks per drinking day for the log-normal part.

Table 2.

Parameter estimates and 95% CIs from fitting the Bayesian joint model with both random intercepts and slopes (model (5)) to the COMBINE data ($N = 578$)[†]

Variable	Posterior median	95% CI	
<i>Hurdle binomial component, zero part</i>			
Intercept ($\exp(\beta_{00})$)	0.64	0.30	1.36
Naltrexone ($\exp(\beta_{01})$)	3.57	1.40	8.47
Week ($\exp(\beta_{02})$)	0.95	0.89	1.02
<i>Truncated binomial part</i>			
Intercept ($\exp(\beta_{10})$)	0.21	0.13	0.32
Naltrexone ($\exp(\beta_{11})$)	0.45	0.26	0.77
Week ($\exp(\beta_{12})$)	1.02	0.96	1.09
<i>Log-normal component</i>			
Intercept ($\exp(\beta_{20})$)	5.94	5.25	6.71
Naltrexone ($\exp(\beta_{21})$)	0.84	0.73	0.96
Week ($\exp(\beta_{22})$)	0.99	0.97	1.01
<i>SD or correlations</i>			
σ	0.32	0.31	0.34
SD(u_{00})	6.26	5.27	7.36
SD(u_{10})	3.25	2.85	3.71
SD(u_{20})	0.71	0.65	0.77
SD(u_{01})	0.43	0.34	0.55
SD(u_{11})	0.23	0.19	0.27
SD(u_{21})	0.10	0.09	0.11
Corr(u_{00}, u_{10})	-0.90	-0.94	-0.85
Corr(u_{00}, u_{20})	-0.09	-0.27	0.09
Corr(u_{00}, u_{01})	-0.50	-0.76	-0.04
Corr(u_{00}, u_{11})	-0.02	-0.34	0.31
Corr(u_{00}, u_{21})	-0.17	-0.35	0.02
Corr(u_{10}, u_{20})	0.26	0.11	0.41
Corr(u_{10}, u_{01})	0.37	-0.09	0.66
Corr(u_{10}, u_{11})	0.01	-0.31	0.34
Corr(u_{10}, u_{21})	0.12	-0.05	0.28
Corr(u_{20}, u_{01})	-0.03	-0.23	0.17
Corr(u_{20}, u_{11})	0.03	-0.16	0.21
Corr(u_{20}, u_{21})	-0.26	-0.38	-0.13
Corr(u_{01}, u_{11})	-0.20	-0.44	0.07
Corr(u_{01}, u_{21})	0.14	-0.03	0.30
Corr(u_{11}, u_{21})	0.12	-0.04	0.27

[†]Regression coefficients were exponentiated and represent the OR of abstinence during a week for the zero part, the OR of daily drinking when drinking occurred during the week for the truncated binomial part and mean ratios of drinks per drinking day for the log-normal part.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Simulation results [†]

Parameter	True	Results for joint HUB-log-normal model				Results for separate HUB and log-normal models				
		Mean	Bias	MSE	Cov	Mean	Bias	MSE	Cov	
<i>N = 500 subjects</i>										
HUB zero	β_{00}	-0.5	-0.437	0.063	0.122	0.97	-0.409	0.091	0.126	0.97
	β_{01}	1.2	1.149	-0.051	0.304	0.93	1.137	-0.063	0.306	0.94
	β_{02}	-0.04	-0.042	-0.002	0.005	0.97	-0.045	-0.005	0.005	0.97
HUB TBin	β_{10}	-1.4	-1.464	-0.064	0.081	0.94	-1.479	-0.079	0.085	0.94
	β_{11}	-0.9	-0.889	0.011	0.133	0.92	-0.881	0.019	0.134	0.93
	β_{12}	0.02	0.020	0.000	0.001	0.93	0.019	-0.001	0.001	0.93
Log-normal	β_{20}	1.7	1.692	-0.008	0.007	0.93	1.775	0.075	0.010	0.72
	β_{21}	-0.2	-0.198	0.002	0.008	0.95	-0.183	0.017	0.007	0.95
	β_{22}	0	0.000	0.000	0.000	0.94	-0.001	-0.001	0.000	0.94
SD or correlation	σ	0.4	0.401	0.001	0.000	0.96	0.401	0.001	0.000	0.97
	$SD(u_{00})$	5	4.977	-0.023	0.155	0.96	5.031	0.031	0.157	0.96
	$SD(u_{10})$	3.2	3.254	0.054	0.052	0.97	3.252	0.052	0.052	0.97
	$SD(u_{20})$	0.7	0.706	0.006	0.001	0.93	0.694	-0.006	0.001	0.93
	$Corr(u_{00}, u_{10})$	-0.9	-0.908	-0.008	0.001	0.94	-0.906	-0.006	0.001	0.94
	$Corr(u_{00}, u_{20})$	-0.2	-0.207	-0.007	0.010	0.98				
	$Corr(u_{10}, u_{20})$	0.4	0.396	-0.004	0.006	0.95				
<i>N = 200 subjects</i>										
HUB zero	β_{00}	-0.5	-0.352	0.148	0.430	0.93	-0.312	0.188	0.481	0.92
	β_{01}	1.2	1.106	-0.094	0.672	0.98	1.021	-0.179	0.755	0.97
	β_{02}	-0.04	-0.052	-0.012	0.013	0.96	-0.049	-0.009	0.014	0.94
HUB TBin	β_{10}	-1.4	-1.551	-0.151	0.240	0.91	-1.587	-0.187	0.267	0.90
	β_{11}	-0.9	-0.868	0.032	0.340	0.94	-0.789	0.111	0.361	0.95
	β_{12}	0.02	0.018	-0.002	0.002	0.95	0.020	0.000	0.002	0.96
Log-normal	β_{20}	1.7	1.706	0.006	0.012	0.96	1.792	0.092	0.017	0.87

Parameter	True	Results for joint HUB-log-normal model				Results for separate HUB and log-normal models			
		Mean	Bias	MSE	Cov	Mean	Bias	MSE	Cov
<i>N = 200 subjects</i>	β_{21}	-0.2	-0.001	0.016	0.97	-0.184	0.016	0.015	0.96
	β_{22}	0	-0.004	0.000	0.94	-0.004	-0.004	0.000	0.93
	σ	0.4	0.402	0.002	0.97	0.402	0.002	0.000	0.98
	SD or correlation	5	4.923	-0.077	0.369	5.052	0.052	0.417	0.93
	$SD(u_{80})$	3.2	3.333	0.133	0.166	3.326	0.126	0.165	0.93
	$SD(u_{10})$	0.7	0.711	0.011	0.003	0.692	-0.008	0.003	0.95
	$Corr(u_{80}, u_{10})$	-0.9	-0.915	-0.015	0.002	-0.911	-0.011	0.002	0.86
	$Corr(u_{80}, u_{20})$	-0.2	-0.212	-0.012	0.025				
	$Corr(u_{10}, u_{20})$	0.4	0.389	-0.011	0.014				
<i>N = 50 subjects</i>	β_{10}	-0.5	-0.700	-0.200	1.779	-0.710	-0.210	2.019	0.93
	β_{01}	1.2	1.819	0.619	5.322	1.646	0.446	5.617	0.91
	β_{12}	-0.04	-0.027	0.013	0.060	-0.029	0.011	0.062	0.92
	β_{10}	-1.4	-1.361	0.039	0.899	-1.418	-0.018	0.989	0.94
	β_{11}	-0.9	-1.461	-0.561	2.051	-1.241	-0.341	1.691	0.87
	β_{12}	0.02	0.006	-0.014	0.012	0.004	-0.016	0.013	0.92
	β_{20}	1.7	1.701	0.001	0.051	1.799	0.099	0.050	0.90
	β_{21}	-0.2	-0.254	-0.054	0.081	-0.209	-0.009	0.065	0.95
	β_{21}	0	0.008	0.008	0.001	0.008	0.008	0.001	0.94
	σ	0.4	0.403	0.003	0.001	0.403	0.003	0.001	0.97
SD or correlation	$SD(u_{80})$	5	5.191	0.191	2.222	5.416	0.416	2.907	0.92
	$SD(u_{10})$	3.2	3.450	0.250	0.613	3.419	0.219	0.713	0.95
	$SD(u_{20})$	0.7	0.745	0.045	0.014	0.691	-0.009	0.011	0.95
	$Corr(u_{80}, u_{10})$	-0.9	-0.913	-0.013	0.007	-0.911	-0.011	0.008	0.85
	$Corr(u_{80}, u_{20})$	-0.2	-0.234	-0.034	0.074				
	$Corr(u_{10}, u_{20})$	0.4	0.365	-0.035	0.047				

[†] HUB zero, zero part of the HUB model; HUB TBin, truncated binomial part; Cov, coverage rate of the 95% CI.