Published in final edited form as:

Stat Methods Med Res. 2018 April ; 27(4): 1258–1270. doi:10.1177/0962280216659312.

Joint modeling of longitudinal zero-inflated count and time-to-event data: A Bayesian perspective

Huirong Zhu*, Stacia M. DeSantis[†], and Sheng Luo[‡]

*Ph.D candidate, Department of Biostatistics, The University of Texas Health Science Center at Houston, Houston, Texas 77030, USA

[†]Associate Professor, Department of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, Texas 77030, USA

[‡]Associate Professor, Department of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, Texas 77030, USA

Abstract

Longitudinal zero-inflated count data are encountered frequently in substance-use research when assessing the effects of covariates and risk factors on outcomes. Often, both the time to a terminal event such as death or dropout and repeated measure count responses are collected for each subject. In this setting, the longitudinal counts are censored by the terminal event, and the time to the terminal event may depend on the longitudinal outcomes. In the study described herein, we expand the class of joint models for longitudinal and survival data to accommodate zero-inflated counts and time-to-event data by using a Cox proportional hazards model with piecewise constant baseline hazard. We use a Bayesian framework via Markov chain Monte Carlo simulations implemented in the BUGS programming language. Via an extensive simulation study, we apply the joint model and obtain estimates that are more accurate than those of the corresponding independence model. We apply the proposed method to an alpha-tocopherol, beta-carotene lung cancer prevention study.

Keywords

Joint model; Zero-inflation generalized Poisson; Count data; Markov chain Monte Carlo; Mixed model

1 Introduction

Longitudinal-count-outcome data with excess zeros are encountered frequently in clinical and public health studies. Examples include skin cancer counts,¹ blood product units transfused after surgery,² the number of complications experienced after a procedure,³ and alcohol or tobacco consumption.^{4,5} Counts provide useful information when observing the natural history of the target disease and/or evaluating the long-term effects of interventions against the disease. Zero-inflated count models [e.g., the zero-inflated Poisson (ZIP),⁶ zero-

inflated negative binomial (ZINB)^{7]} are commonly used to analyze such data; these distinguish between sources of zeros by using a two-part mixture of a point mass at zero and a count distribution. Compared with the ZIP model, the ZINB model has an additional parameter that captures variability due to over-dispersion.⁸

Longitudinal studies, however, are quite prone to incomplete records owing to terminal events such as death and dropout during the study.⁹ Importantly, longitudinal measurements can indicate such terminal events, which then censor the sequence of repeated measures. For example, in an alcohol- or tobacco-consumption study, the dropout process may be significantly informative of the drinking or smoking process because heavy drinkers or smokers are more likely to drop out of the trial early. Thus, heavy substance consumption is expected to provide information about the dropout time. Censoring the longitudinal measures due to the dropout events also may be informative because dropouts may exhibit heavier consumption during the study. In the literature, a dependent terminal event is often referred to as “dependent censoring” or “informative censoring,” and ignoring it has been shown to cause biased estimation and even invalid inference.^{10,11} However, a joint analysis of these two processes can aid simultaneous investigations of the effect of covariates on both the longitudinal and survival processes.¹² When informative censoring is present, incorporating traditional zero-inflated count models into the class of joint models with time-to-event data can provide accurate inference by considering the existing informative censoring feature contained in the data.

The statistical literature on joint longitudinal and survival models for many different types of longitudinal outcomes has expanded over the past decade.^{13–15} Joint modeling is a flexible dynamic tool used to capture the relationship between longitudinal predictors and survival time. It makes estimation more efficient via simultaneous estimation of shared and unshared parameters that characterize both the longitudinal and survival processes. Tsiatis and Davidian¹³ and Wu *et al.*¹⁴ have published thorough reviews on the topic. Specifically, subject-specific random effects are typically used to link two submodels for the longitudinal and event-time data; such submodels usually include a mixed-effects model for longitudinal data and a semiparametric Cox or accelerated failure time model for event-time data.^{16,17} Other authors have proposed extensions of the joint model, such as including non-Gaussian random effects,¹⁸ assuming a latent stochastic Gaussian process instead of a random effects model,¹⁹ and incorporating the multidimensional longitudinal and/or survival components into the joint model framework.^{20,21}

The vast majority of research in the joint modeling field assumes continuous or multivariate continuous outcomes in the longitudinal submodel. Recently, Rizopoulos proposed a two-part shared parameter framework to illustrate the association between a dichotomous variable and the time to renal graft failure.²² Hatfield *et al.*¹² developed a zero-augmented joint model for longitudinal patient-reported proportional outcomes with many zeros and survival events. However, our study is concerned with the impact of dependent terminal events on a longitudinal vector of counts subject to zero inflation, and how those characteristics influence data-inferential procedures.

As described herein, we propose a comprehensive joint model to simultaneously accommodate the association among zero-inflated longitudinal counts with dispersion and the presence of informative censoring. We construct the zero-inflated count submodel by using a mixture of a degenerate point mass at zero and a count distribution such as the Poisson or generalized Poisson distribution.²³ We use the Cox proportional hazards model with piecewise constant baseline hazard to model the corresponding time-to-event data and link the submodels via two shared random effects. In addition, we perform simulation and data analysis by using a Bayesian framework to allow for full and exact posterior inference for all parameters. We specify noninformative prior distributions for all parameters and assess sensitivity to these priors.

The paper is organized as follows: In Section 2 we present the motivation from the alpha-tocopherol, beta-carotene (ATBC) lung cancer prevention study. In Section 3, we describe the proposed joint model structure. In Section 4, we report a simulation study that assesses the performance of the proposed approach and compare its performance with that of the corresponding independence approach. In Section 5, we apply competing models to the ATBC study. Finally, in Section 6, we conclude with remarks and an outline of future work.

2 Motivating Data

The methodological development for this study is motivated by the longitudinal ATBC study, which was sponsored by the National Cancer Institute. We refer the reader to the original report from the ATBC Study Group (2003)²⁴ for more details. Briefly, 29 133 eligible smoking-dependent males aged 50 to 69 years were included under the randomized primary prevention study to determine whether alpha-tocopherol or beta-carotene reduces the incidence of lung cancer. All individuals in the study smoked at least five cigarettes per day at baseline. Extensive medical histories and examination data for the participants were collected at baseline, and the participants were observed for five to eight years with scheduled follow-up visits every four months. In each follow-up visit, participants were queried about their health and smoking status since their last visit. Dropout events were also collected during the study. In this paper, we define the response variable as the average daily cigarette consumption over the last four months; it has excessive zeros compared to the Poisson distribution with a comparable mean. The terminal event under consideration is the dropout rate since study onset (24% dropout rate in the ATBC study).

In the current study, we are interested in how risk factors, e.g., age, years of smoking, and baseline smoking consumption, influence the severity of smoking addiction and the probability of smoking abstinence accounting for all relevant correlation and subject-specific random effects.

In the preliminary analysis, we observe that the cigarette counts are associated with the dropout time. To demonstrate this correlation, we split the dataset by the observed mean of 15 cigarettes per day. The 2 Kaplan–Meier curves for the mean-split data in Figure 1 demonstrate a significant difference in time to dropout between the two consumption groups (log-rank test p -value < 0.001), whereby high cigarette consumption increases the likelihood

of dropout. The association between cigarette counts and dropout events requires simultaneous consideration of these two correlated processes.

3 Methods and estimation

3.1 Zero-inflated longitudinal submodel

3.1.1 Zero-inflated Poisson model—Assuming independence between subjects, let Y_{ij} denote the longitudinal response (average daily cigarette consumption) for subject $i = 1, \dots, n$ at time j for $j = 1, \dots, J_i$. The distribution of Y_{ij} is

$$Y_{ij} \sim \begin{cases} 0 & \text{with probability } \phi_{ij} \\ \text{Poisson}(\lambda_{ij}) & \text{with probability } 1 - \phi_{ij}, \end{cases}$$

where ϕ_{ij} denotes the probability of the observation arising from the degenerate distribution at zero and λ_{ij} represents the mean of the Poisson distribution. The probability distribution function is

$$\begin{aligned} P(Y_{ij}=0|\mathbf{Z}_{ij}, \mathbf{X}_{ij}) &= \phi_{ij} + (1 - \phi_{ij})e^{-\lambda_{ij}}, \\ P(Y_{ij}=y_{ij}|\mathbf{X}_{ij}) &= (1 - \phi_{ij}) \frac{\lambda_{ij}^{y_{ij}} e^{-\lambda_{ij}}}{y_{ij}!} \text{ if } y_{ij}=1, 2, 3, \dots, \end{aligned}$$

where $0 < \phi_{ij} < 1$ and $0 < \lambda_{ij} < \infty$. This formulation incorporates more zeros than permitted under the Poisson assumption (i.e., where $\phi_{ij} = 0$). The vector $\mathbf{Z}_{ij}$ is the set of covariates predicting the probability of abstaining from smoking and $\mathbf{X}_{ij}$ is the set of covariates predicting the “intensity” of smoking. The vectors $\mathbf{X}_{ij}$ and $\mathbf{Z}_{ij}$ may overlap.

3.1.2 Zero-inflated generalized Poisson model—The flexible zero-inflated generalized Poisson (ZIGP) model is used to adjust for dispersion in an otherwise ZIP model with an additional dispersion parameter ω , which is positive for over-dispersed and negative for under-dispersed data. Let Y_{ij} denote the average daily cigarette consumption for subjects $i = 1, \dots, n$ and at times $j = 1, \dots, J_i$. Then Y_{ij} has the following distribution:

$$Y_{ij} = \begin{cases} 0 & \text{w. p. } \phi_{ij} \\ \text{general Poisson}(\lambda_{ij}, \omega) & \text{w. p. } 1 - \phi_{ij}, \end{cases}$$

where λ_{ij} and the zero-inflation parameter ϕ_{ij} are the same as described in Section 3.1.1. Typically, the range of ω is $\max\{-1, -\frac{\lambda_{ij}}{N}\} \leq \omega < 1$ corresponding to counts $y_{ij} = 0, 1, 2, \dots, N$, where $\omega = 0$ reduces to the ZIP model. Note that a positive count with $N = \infty$ leads to $\omega \in [0, 1)$.²⁵

The probability distribution of the longitudinal ZIGP model is

$$\begin{aligned}
 P(Y_{ij}=0|\mathbf{Z}_{ij}) &= \phi_{ij} + (1-\phi_{ij})e^{-(1-\omega)\lambda_{ij}}, \\
 P(Y_{ij}=y_{ij}|\mathbf{X}_{ij}) &= (1-\phi_{ij})(1-\omega)\lambda_{ij} \frac{[(1-\omega)\lambda_{ij} + \omega y_{ij}]^{(y_{ij}-1)}}{y_{ij}!} e^{-[(1-\omega)\lambda_{ij} + \omega y_{ij}]} \text{ if } y_{ij}=1, 2, 3, \dots,
 \end{aligned}$$

where $0 < \phi_{ij} < 1$ and $\lambda_{ij} > 0$.

For both the ZIP and ZIGP distributions, $\text{logit}(\phi_{ij})$ and $\log(\lambda_{ij})$ are the natural link functions for the Bernoulli probability of success (abstinence) and the mean of the Poisson part, respectively. To account for within-subject dependence in the counting process, we let

$$\text{logit}(\phi_{ij}) = \boldsymbol{\alpha} \mathbf{Z}_{ij} + u_i, \quad (1)$$

$$\log(\lambda_{ij}) = \boldsymbol{\beta} \mathbf{X}_{ij} + v_i, \quad (2)$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are coefficient row vectors, $\mathbf{Z}_{ij}$ and $\mathbf{X}_{ij}$ are covariate vectors, and u_i and v_i are subject-specific random effects accounting for within-subject correlation in each model part.

Owing to the flexibility and straightforward mathematical properties of the Gaussian distribution, we assume that the random effects are independent and identically distributed bivariate normal random vectors for all subjects $i = 1, \dots, n$, i.e.,

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{u,v}),$$

where

$$\boldsymbol{\Sigma}_{u,v} = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}.$$

σ_u^2 and σ_v^2 are variance components in the Bernoulli and Poisson or GP part of the model, respectively, and ρ captures the correlation between those two parts. Intuitively, a positive u_i increases the probability of smoking abstinence; a positive v_i increases average daily cigarette consumption, while ρ reflects the correlation between the probability of smoking abstinence and average daily cigarette consumption.

3.2 Cox proportional hazard submodel

We use the Cox proportional hazards model to model time to terminal events (i.e., dropout). Let t_i be the observed dropout time since study onset for patient i and δ_i be the censoring

indicator ($\delta_i = 1$ if dropout is observed, and $\delta_i = 0$ if it is censored). The terminal event intensity (or hazard) at time t_i is

$$h(t_i) = h_0(t_i) \exp\{\boldsymbol{\gamma} \mathbf{W}_i + \nu_1 u_i + \nu_2 v_i\}, \quad (3)$$

where $h_0(\cdot)$ is the baseline hazard function and $\boldsymbol{\gamma}$ represents coefficients for the risk factors in the fixed covariate matrix $\mathbf{W}_i$, which can share part of or all covariates in $\mathbf{Z}_i$ and $\mathbf{X}_i$ defined in Section 3.1.2. The u_i , v_i are shared random effects accounting for the correlation between the longitudinal and survival processes and $\boldsymbol{\nu} = (\nu_1, \nu_2)'$ reflects the strength of association between the counts and the survival submodels.

We adopt a piecewise constant baseline hazard to approximate the baseline hazard function $h_0(t)$, which yields good estimators of both the fixed and random effects.^{26–28} To specify a piecewise exponential baseline hazard, we set a series of fixed cut points $0 = \tau_0 < \tau_1 < \dots < \tau_m$ with τ_m equal to the maximum observation time in the dataset and assume that the baseline hazard in each interval is constant with $\mathbf{h}_0 = (h_{0,1}, h_{0,2}, \dots, h_{0,m-1})$. We select the number of cut points to allow at least ten dropout events per interval.

We link the two submodels (1), (2), and (3) via the random effects $\mathbf{u}$ and $\mathbf{v}$ and assume that the longitudinal and survival processes are independent given $\mathbf{u}$ and $\mathbf{v}$. The joint model reduces to an independent model if the corresponding association coefficients $\nu_1 = \nu_2 = 0$. The joint modeling framework accounts for three correlated processes: abstinence from smoking, amount of smoking by those at risk of smoking over the study period, and time to terminal events.

Let the unknown-parameter vector be $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \Sigma_{\mathbf{u}, \mathbf{v}}, \boldsymbol{\gamma}, \boldsymbol{\nu}, \mathbf{h}_0)$, then the conditional likelihood of the observed data, $\mathbf{y}_i$, is

$$L_y(\mathbf{y}_i | u_i, v_i) = \prod_{j=1}^{n_i} P(Y_{ij}=0 | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \Sigma_{\mathbf{u}, \mathbf{v}})^{I(Y_{ij}=0)} P(Y_{ij}=y_{ij} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\omega}, \Sigma_{\mathbf{u}, \mathbf{v}})^{I(Y_{ij}>0)}.$$

The conditional likelihood of event outcomes t_i and δ_i for patient i is

$$L_s(t_i, \delta_i | u_i, v_i) = h(t_i)^{\delta_i} S(t_i),$$

where $S(t_i) = \exp[-\int_0^{t_i} h(s) ds]$ denotes the survival function. Finally, the full likelihood of the joint model for patient i is

$$L(y_i, t_i, \delta_i, u_i, v_i) = L_y(y_i | u_i, v_i) L_s(t_i, \delta_i | u_i, v_i) p(u_i, v_i), \quad (4)$$

where $p(u_i, v_i)$ is the joint density function of random effects u_i and v_i .

3.3 Bayesian inference

We use Bayesian inference based on Markov chain Monte Carlo (MCMC) simulations to infer the unknown parameters, summarizing them using posterior means and 95% credible intervals (CI). We select noninformative priors for all the parameters. Specifically, we use the normal distribution $N(0, 100)$ for all components in α , β , γ , and ν . For the covariance matrix $\Sigma_{u,v}$, a commonly used conjugate prior is an inverse-Wishart distribution; however, the standard “noninformative” inverse-Wishart prior cannot guarantee noninformativeness with a single parameter to control the precision of all entries in the covariance matrix.²⁹ Thus, we use the Cholesky decomposition to set the prior distribution and rewrite the covariance matrix as $\Sigma_{u,v} = LL'$, where L is the corresponding lower triangular matrix. We specify a Uniform(0, 10) prior distribution for the diagonal elements in L to ensure positivity and a Normal(0, 100) prior distribution for the off-diagonal elements, allowing for possible negative correlations. This approach yields satisfactory results and has been widely used in longitudinal and survival analysis.^{9,30} We obtain initial values for the model parameters by separately modeling the longitudinal and survival data with a zero-inflated count model and a Cox proportional hazards model.

Model fitting is performed by using the BUGS programming language³¹ by specifying the full likelihood and prior distributions of all unknown parameters. To analyze the real data, we run multiple chains with initial values suggested by results of fitting the independent model. As evidence of convergence, we use the Gelman–Rubin diagnostic to ensure that the scale reduction $\hat{R}$ of all parameters is smaller than 1.1.³² We also examine trace and autocorrelation plots to determine the burn-in length and ensure chain convergence. Finally, we explore other choices of priors to ensure that the results are robust to prior specifications.

3.4 Model selection

From a wide variety of model-selection criteria in the Bayesian toolbox, we adopt the deviance information criterion (DIC), the extended Akaike information criterion (EAIC), the expected Bayesian (or Schwarz) information criterion (EBIC), and log-pseudo-marginal likelihood (LPML).^{33,34}

The DIC provides an assessment of model fitting and a penalty for model complexity. Let $f(\mathbf{y}|\boldsymbol{\theta})$ represent the likelihood function for the observed data $\mathbf{y}$ given the parameter vector $\boldsymbol{\theta}$, and let $h(\mathbf{y})$ be a standardizing function of the data alone. The deviance is defined as $D(\boldsymbol{\theta}) = -2 \log(f(\mathbf{y}|\boldsymbol{\theta})) + 2 \log(h(\mathbf{y}))$. The DIC is computed as $DIC = \bar{D} + p_D$, where $\bar{D} = E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})]$ is the posterior expectation of the deviance, $p_D = \bar{D} - D(\bar{\boldsymbol{\theta}})$ is the effective number of parameters to capture model complexity, and $D(\bar{\boldsymbol{\theta}}) = D(E_{\boldsymbol{\theta}|\mathbf{y}}[\boldsymbol{\theta}])$ is the deviance evaluated at the posterior mean of the parameters. A smaller value of DIC indicates a better model fit. We estimate EAIC and EBIC by $\widehat{EAIC} = -2\bar{D} + 2\nu$ and $\widehat{EBIC} = -2\bar{D} + \nu \log(n)$, where ν is number of parameters in the model and n is number of subjects. The model with the smaller EAIC and EBIC is preferred.

We also compute the conditional predictive ordinate (CPO) statistics for model selection.³⁴ CPO is derived from the posterior distribution and can be written as $CPO_i = \int f(y_i|\theta)p(\theta|D^{(-i)})d\theta$, where $D^{(-i)}$ is the full data with observation i deleted. The LPML, defined as

$LPML = \sum_{i=1}^n \log(\widehat{CPO}_i)$, is used to summarize the CPO_{*i*} values.³⁵ Larger values of LPML indicate a better fit.

4 Simulation Study

We perform extensive simulation studies to compare the performance of the proposed joint models with that of separate independent models. We design the first simulation study to have no correlation between the event time and the zero-inflated longitudinal counts (i.e., $\nu_1 = 0$, $\nu_2 = 0$) and the second simulation study to have a strong correlation (i.e., $\nu_1 = -0.6$, $\nu_2 = 0.4$ for the ZIP joint model and $\nu_1 = -0.3$, $\nu_2 = 1.2$ for the ZIGP joint model). We consider 500 datasets, each with sample size $N = 1600$. We generate the longitudinal zero-inflated counts under both ZIP and ZIGP distributions. The simulated data structure is similar to that of the ATBC study.

The joint model for the simulation study is

$$\begin{aligned}\text{logit}(\phi_{ij}) &= \alpha_0 + \alpha_1 I(trt_i) + \alpha_2 \text{Time}_j + u_i, \\ \log(\lambda_{ij}) &= \beta_0 + \beta_1 I(trt_i) + \beta_2 \text{Time}_j + v_i, \\ h(t_i) &= h_0(t_i) \exp\{\gamma I(trt_i) + \nu_1 u_i + \nu_2 v_i\},\end{aligned}$$

where $I(trt_j)$ is a binary treatment indicator for subject i and $j = 1, \dots, 6$ time points. We set $\alpha = (0.8, -0.3, 0.2)'$, $\beta = (0.5, 0.4, -0.2)'$, and $\gamma = -1.5$. For the random effects covariance matrix, we choose $\rho = -0.5$, $\sigma_u^2 = 1.2$, and $\sigma_v^2 = 0.6$ for ZIP and $\sigma_u^2 = 1.5$ and $\sigma_v^2 = 0.7$ for ZIGP. We set the dispersion parameter in the ZIGP distribution to 0.3.

The time to terminal event is simulated by using the Cox submodel with constant baseline hazard submodel, i.e., $h_0 = 1$. The independent censoring time is generated from a Uniform(2, 6) distribution resulting in 30% censoring, which is similar to that in the ATBC study. We set $\delta_j = 1$ if the event time t_j is not greater than the censoring time and 0 otherwise.

We apply the Bayesian approach described in Section 3.3 to obtain samples from the posterior distributions of the parameters of interest. The OpenBUGS code for the ZIP joint model is included in the online appendix as an example. For each dataset in the simulation study, we run two parallel MCMC chains with over-dispersed initial values. We run both chains for 60 000 iterations. The first 40 000 iterations are discarded as burn-in; the remaining 20 000 samples are used to determine the posterior distributions of the parameters of interest. Trace plots indicate good mixing properties of the chains for the model parameters, and the scale reduction $\hat{R}$ of all parameters is smaller than 1.1. Thus, convergence is achieved.

Through the simulation studies, we want to assess accuracy of both the joint and independent models and determine how parameter estimates, standard errors, and coverage probabilities are affected by ignoring the nonignorable censoring owing to dropout. To that end, we compute the bias (the average of the posterior means minus the true values), standard error (SE: the square root of the average of the posterior variance), standard deviation (SD: the standard deviation of the posterior means), and coverage probabilities (CPs) of 95% equal-tail credible intervals, all displayed in Tables 1 and 2.

In simulation study 1, the data are generated by using the independence model, i.e., the longitudinal counts are not correlated with the terminal event. Table 1 presents the model estimates for the data under both the ZIP and ZIGP distributions for the zero-inflated longitudinal counts. The independent and joint models result in negligible bias, SEs close to the SD, and nominal CPs reasonably close to 95%. We notice that the bias of ρ and σ_u^2 are slightly larger than the bias for other parameters and that the CPs are slightly less than or greater than 0.95, indicating difficulty in distinguishing the random effects.¹¹ Of note, the estimates of the shared random effects parameters ν_1 and ν_2 in the joint models are correctly near 0, and the CP is close to 95%. The fact that over-parameterized joint models still provide unbiased and efficient estimates of the parameters indicates that the joint approach is robust to model over-parameterization.

In simulation study 2, the data are generated by using the joint model with zero-inflated longitudinal counts where the terminal events are designed to have a correlation with $\nu_1 = -0.6$, $\nu_2 = 0.4$ for the ZIP joint model and $\nu_1 = -0.3$, $\nu_2 = 1.2$ for the ZIGP joint model. Table 2 lists the model estimates for the data under both the ZIP and ZIGP distributions for the zero-inflated longitudinal counts. In general, use of a joint model results in negligible bias and small SEs close to the SD, and nominal CPs reasonably close to 95%. However, fitting the independent model to the correlated data results in biased estimates of most parameters and CPs far from the nominal value. These results demonstrate that the joint model can generally recover the true values of parameters in the presence of nonignorable censoring.

5 Data Analysis

We apply the proposed joint model and independent model to the motivating ATBC study. The primary outcome of interest is daily cigarette count, which is zero-inflated with a mean of 15 and range [0, 90] over the study period. The participants are scheduled to visit every four months after the baseline visit and the median follow-up duration is 2.7 years. In the joint model, fixed effects in the longitudinal submodel include age (centered at 57, denoted by Age), years of smoking (centered at 35, denoted by SmokYrs), baseline cigarette counts (centered at 20, denoted by Bsmokcount), and year from baseline (denoted by Visit, not mean centered). Frailty terms in the survival submodel are shared random effects from the two corresponding parts in the zero-inflated longitudinal submodel; the coefficients of those random effects govern the strength of the correlation between the longitudinal and survival submodels. The survival submodel only includes baseline smoking count because it is the only significant predictor.

The longitudinal zero-inflated cigarette count is modeled by using the ZIP and ZIGP models described in sections 3.1 and 3.2 with the following linear predictors:

$$\text{logit}(\phi_{ij}) = \alpha_0 + \alpha_1 \text{Smokyr}_i + \alpha_2 \text{Age}_i + \alpha_3 \text{Visit}_{ij} + \alpha_4 \text{Bsmokcount}_i + u_i, \quad (5)$$

$$\log(\lambda_{ij}) = \beta_0 + \beta_1 \text{Smokyr}_i + \beta_2 \text{Age}_i + \beta_3 \text{Visit}_{ij} + \beta_4 \text{Bsmokcount}_i + v_i. \quad (6)$$

The “survival” time is time to dropout and the model for the dropout hazard is

$$h(t_i) = h_0(t_i) \exp\{\gamma \text{Bsmokcount}_i + \nu_1 u_i + \nu_2 v_i\}. \quad (7)$$

First, we fit the ZIGP model for the longitudinal smoking counts in the ATBC data, and the results suggest that the data do not display strong dispersion since the dispersion parameter ω approaches 0. As described in Section 3.1.2, the ZIGP model reduces to the ZIP model in this circumstance. Therefore, for the data analysis, we only consider joint and independent models under the ZIP framework. Table 3 compares those models by using the model-selection criteria described in Section 3.4. The joint model performs better than the corresponding independence model with smaller DIC, EAIC, and EBIC values and larger LPML values, suggesting that the joint model is preferable.

Table 4 summarizes the posterior mean, SD, and 95 % equal-tail credible interval of the parameters from both the joint and independent models. In the joint model, the results of the longitudinal submodel suggest that older participants have higher probability of smoking abstinence (OR = 1.139, 95% CI [1.118, 1.161]) and lower daily cigarette consumption (RR = 0.991, 95% CI [0.990, 0.992]). Moreover, smoking abstinence is less likely for those participants with greater prior smoking years (OR = 0.905, 95% CI [0.894, 0.914]) while daily cigarette consumption is higher (OR = 0.907, 95% CI [0.898, 0.915]). Participants are more likely to be abstinent and have lower cigarette consumption further from baseline. Importantly, the results of the survival submodel suggest that higher baseline cigarette consumption increases the possibility of dropout.

The coefficients of random effects are $\nu_1 = -0.009$ with 95% CI $[-0.015, -0.003]$ and $\nu_2 = -0.038$ with 95% CI $[-0.147, 0.080]$. The significant coefficient ν_1 in the zero part indicates that participants with a lower probability of abstinence are more likely to drop out. The association between the longitudinal and survival submodels confirms our observations in Section 2. Table 4 also shows a negative correlation coefficient ($\rho = -0.237$, 95% CI = $[0.256, 0.013]$) between random effects in the zero and Poisson parts, confirming the more obvious result that participants who smoke more cigarettes have a lower probability of abstinence.

We also give insight into our model’s ability to provide subject-specific estimates of predictions conditional on the observed covariates for the ATBC study. We randomly choose

two groups of participants in the ATBC study: Group 1 includes participants with varying follow-up visits and Group 2 includes all subjects with 20 follow-up visits with three visits per year. The participants in each group are randomly selected from those with different numbers of average daily cigarettes and distinct smoking abstinence patterns. We save the estimated subject-specific random effects from the MCMC samples and compute the estimated average daily cigarette counts by rounding the value of $(1 - \hat{\phi}_{ij}) \hat{\lambda}_{ij}$ for subject i at visit j . The $\hat{\phi}_{ij}$ and $\hat{\lambda}_{ij}$ are computed based on Equations (5) and (6) using the estimated parameters. This is an informal method to evaluate the performance of the final model.

Table 5 summarizes the predicted average daily number of cigarettes, while the heatmaps of the observed and estimated average daily cigarette counts for both chosen groups are displayed in Figure 2. The observed and estimated average daily cigarette counts are very similar and have consistent trends. These results show that our model has a good capacity to reproduce the smoking patterns for randomly chosen individuals.

6 Discussion

In this article, we propose a joint modeling framework to analyze longitudinal zero-inflated counts censored by terminal events. The joint model consists of a zero-inflated count submodel and a Cox survival submodel linked via shared random effects. This joint model accommodates the co-existent features including zero inflation, over-dispersion, and dependent censoring, within the longitudinal counts data. It results in better inference for covariate effects on longitudinal counts with informative dropout. Our simulation study demonstrates that the proposed joint models improve the accuracy of the parameter estimates compared with independent models in the presence of informative censoring. Conversely, our joint models are robust to misspecification, producing results comparable to the independent model under the independent-terminal-event scenarios.

The results from the present ATBC study are consistent with previous findings. In a paper modeling smoking patterns via a latent transition model, Luo³⁶ showed that older people have a higher probability to make quit attempts, conditional on the corresponding random effects. In the current study, age similarly has a positive effect on the probability of smoking abstinence and a negative effect on average daily cigarette consumption. Our study adds to previous findings by modeling the cigarette counts directly, and by investigating the actual reported daily cigarette consumption with respect to demographic and baseline covariates.

The joint model described herein is a shared random effects model for accommodating monotone missingness of longitudinal data, with the merit of easy implementation by using the OpenBUGS software package. Because a mixture of non-monotone missingness and monotone missingness may exist, future work may extend the joint model to account for non-monotone missingness. In addition, dynamic prediction of the time to terminal event has recently become a research topic with significant clinical interest; future research could consider dynamic prediction under the joint model framework. In this paper, we use the shared random effect method to construct the joint model, which assumes that two submodels are dependent only via their shared random effects. Alternative approaches, such as applying selection and pattern mixture models, can also be accommodate missingness. In

the future, these alternative methods can be investigated and compared to the current findings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the National Institute of Health and the National Institute of Alcohol Abuse and Addiction, Grant No. R03AA020648. Sheng Luo's research was supported in part by the National Institute of Neurological Disorders and Stroke under Award No. R01NS091307 and by the National Center for Advancing Translational Sciences under Award No. KL2-TR000370. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high-performing computing resources that have contributed to the research results reported in this article (<http://www.tacc.utexas.edu>). The authors also thank the UTSPH information technology staff for their technical support.

References

1. Maruotti A. A two-part mixed-effects pattern-mixture model to handle zero-inflation and incompleteness in a longitudinal setting. *Biometrical Journal*. 2011; 53(5):716–734. [PubMed: 21887792]
2. DeSantis SM, Lazaridis C, Ji S, Spinale FG. Analyzing propensity matched zero-inflated count outcomes in observational studies. *Journal of Applied Statistics*. 2013; 127:1–15.
3. Hur, K. PhD dissertation. University of Illinois; Chicago: 1999. A random effects zero-inflated Poisson regression model for clustered extra-zero Count Data.
4. Buu A, Li R, Tan X, Zucker RA. Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Statistics in Medicine*. 2012; 31:4074–4086. [PubMed: 22826194]
5. Wang H, Heitjan DF. Modeling heaping in self-reported cigarette counts. *Statistics in Medicine*. 2008; 27(19):3789–3804. [PubMed: 18407584]
6. Lambert D. Zero-Inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*. 1992; 34:1–14.
7. Ridout M, Hinde J, DemeAtrio CG. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative Binomial alternatives. *Biometrics*. 2001; 57(1):219–223. [PubMed: 11252601]
8. Zhu H, Luo S, DeSantis SM. Zero-inflated count models for longitudinal measurements with heterogeneous random effects. *Statistical Methods in Medical Research*. 2015; 23(1):1–16.
9. Luo S. A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statistics in Medicine*. 2014; 33(4):580–594. [PubMed: 24009073]
10. Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine*. 1996; 15(15):1663–1685. [PubMed: 8858789]
11. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000; 1(4):465–480. [PubMed: 12933568]
12. Hatfield LA, Boye ME, Hackshaw MD, Carlin BP. Multilevel Bayesian models for survival times and longitudinal patient-reported outcomes with many zeros. *Journal of the American Statistical Association*. 2012; 107(499):875–885.
13. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*. 2004; 14(3):809–834.
14. Wu L, Liu W, Yi GY, Huang Y. Analysis of longitudinal and survival data: Joint modeling, inference methods, and issues. *Journal of Probability and Statistics*. 2012
15. Yu M, Law NJ, Taylor JM, Sandler HM. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*. 2004; 14(3):835–862.

16. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997; 53(1):330–339. [PubMed: 9147598]
17. Tseng YK, Hsieh F, Wang JL. Joint modelling of accelerated failure time and longitudinal data. *Biometrika*. 2005; 92(3):587–603.
18. Brown ER, Ibrahim JG. Bayesian Approaches to Joint Cure-Rate and Longitudinal Models with Applications to Cancer Vaccine Trials. *Biometrics*. 2003; 59(3):686–693. [PubMed: 14601770]
19. Xu J, Zeger SL. Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2001; 50(3):375–387.
20. Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*. 2006; 62(2):432–445. [PubMed: 16918907]
21. Elashoff RM, Li G, Li N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Statistics in Medicine*. 2007; 26(14):2813–2835. [PubMed: 17124698]
22. Rizopoulos D, Verbeke G, Lesaffre E, Vanrenterghem Y. A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*. 2008; 64(2):611–619. [PubMed: 17725808]
23. Xie FC, Wei BC, Lin JG. Score tests for zero-inflated generalized Poisson mixed regression models. *Computational Statistics and Data Analysis*. 2009; 53(9):3478–3489.
24. Virtamo J, Pietinen P, Huttunen J, Korhonen P, Malila N, Virtanen M, et al. Incidence of cancer and mortality following alpha-tocopherol and beta-carotene supplementation: a postintervention follow-up. *JAMA: Journal of the American Medical Association*. 2003; 290(4):476–485. [PubMed: 12876090]
25. Ntzoufras, I. Bayesian modeling using WinBUGS. Vol. 698. John Wiley & Sons; 2011.
26. Lawless J, Zhan M. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics*. 1998; 26(4):549–565.
27. Feng S, Wolfe RA, Port FK. Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *Journal of the American Statistical Association*. 2005; 100(471)
28. Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2009; 58(1):65–81.
29. O'Malley AJ, Zaslavsky AM. Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association*. 2008; 103(484):1405–1418.
30. Chen BE, Cook RJ, Lawless JF, Zhan M. Statistical methods for multivariate interval-censored recurrent events. *Statistics in medicine*. 2005; 24(5):671–691. [PubMed: 15558580]
31. Thomas A, O'Hara B, Ligges U, Sturtz S. Making BUGS open. *R news*. 2006; 6(1):12–17.
32. Gelman, A., Carlin, JB., Stern, HS., Dunson, DB., Vehtari, A., Rubin, DB. *Bayesian Data Analysis*. CRC press; 2013.
33. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002; 64(4): 583–639.
34. Carlin, BP., Louis, TA. *Bayesian Methods for Data Analysis*. CRC Press; 2011.
35. Dey DK, Chen MH, Chang H. Bayesian approach for nonlinear random effects models. *Biometrics*. 1997; 53(4):1239–1252.
36. Luo S. Joint analysis of stochastic processes with application to smoking patterns and insomnia. *Statistics in Medicine*. 2013; 32(29):5133–5144. [PubMed: 23913574]

Figure 1.

The Kaplan–Meier curves demonstrate a difference in time to dropout (log-rank test p -value < 0.001) for participants in high versus low consumption groups (≤ 15 , solid line; > 15 , dashed line).

Figure 2. Heatmaps of the observed and estimated average daily smoking counts in two groups of six subjects in the ATBC study.

Table 1

Simulation results for the independent and joint models when the terminal event is **independent** of the longitudinal outcomes under both the ZIP and ZIGP distributions.

	True	Independent model				Joint model				
		Bias	SD	SE	CP	Bias	SD	SE	CP	
For zero-inflated outcomes										
ZIP model	α_0	0.800	-0.006	0.102	0.132	0.956	-0.012	0.110	0.129	0.954
	α_1	-0.300	0.001	0.108	0.134	0.940	0.001	0.112	0.125	0.926
	α_2	0.200	0.001	0.025	0.032	0.946	0.001	0.025	0.031	0.960
	β_0	0.500	-0.011	0.133	0.177	0.948	-0.022	0.135	0.166	0.948
	β_1	0.400	0.010	0.170	0.225	0.956	-0.004	0.216	0.237	0.938
	β_2	-0.200	-0.001	0.063	0.082	0.954	0.013	0.172	0.211	0.964
	ρ	-0.500	0.029	0.153	0.201	0.950	0.008	0.166	0.192	0.942
	σ_u^2	1.200	0.157	0.370	0.491	0.942	0.108	0.366	0.455	0.936
	σ_v^2	0.600	0.027	0.082	0.110	0.944	0.030	0.085	0.104	0.944
	γ	-1.500	-0.004	0.083	0.111	0.950	-0.020	0.085	0.107	0.952
For survival										
h_0	1.000	0.005	0.050	0.067	0.956	0.014	0.051	0.065	0.954	
ν_1	0.000					-0.013	0.144	0.268	0.956	
ν_2	0.000					-0.020	0.164	0.269	0.952	
For zero-inflated outcomes										
ZIGP model	α_0	0.800	0.004	0.155	0.145	0.926	0.003	0.151	0.166	0.938
	α_1	-0.300	0.008	0.140	0.135	0.940	0.004	0.139	0.152	0.944
	α_2	0.200	0.002	0.037	0.039	0.948	0.001	0.039	0.045	0.948
	β_0	0.500	-0.011	0.183	0.189	0.946	-0.011	0.177	0.211	0.960
	β_1	0.400	0.009	0.213	0.218	0.946	-0.003	0.205	0.243	0.952
	β_2	-0.200	-0.012	0.084	0.083	0.948	-0.007	0.082	0.097	0.962
	ω	0.300	-0.000	0.049	0.047	0.948	-0.003	0.051	0.053	0.930

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Independent model					Joint model				
	True	Bias	SD	SE	CP	Bias	SD	SE	CP	
ρ	-0.500	0.074	0.194	0.199	0.932	0.048	0.221	0.225	0.902	0.902
σ_u^2	1.500	0.343	0.616	0.658	0.912	0.275	0.646	0.727	0.904	0.904
σ_v^2	0.700	0.039	0.140	0.135	0.932	0.035	0.136	0.149	0.948	0.948
For survival										
γ	-1.500	-0.006	0.088	0.091	0.956	-0.023	0.091	0.104	0.942	0.942
h_0	1.000	0.003	0.052	0.053	0.958	0.012	0.053	0.062	0.954	0.954
ν_1	0.000					-0.016	0.142	0.283	0.950	0.950
ν_2	0.000					-0.023	0.171	0.301	0.964	0.964

Table 2

Simulation results using the independent and joint models when the terminal event **depends** on the longitudinal outcomes under both ZIP and ZIGP distributions.

	True	Independent model				Joint model				
		Bias	SD	SE	CP	Bias	SD	SE	CP	
For zero-inflated outcomes										
ZIP model	α_0	0.800	-0.069	0.064	0.070	0.772	-0.006	0.061	0.068	0.940
	α_1	-0.300	0.057	0.065	0.074	0.862	-0.005	0.066	0.075	0.946
	α_2	0.200	-0.030	0.017	0.019	0.554	0.001	0.018	0.020	0.956
	β_0	0.500	0.058	0.088	0.097	0.890	-0.005	0.088	0.097	0.922
	β_1	0.400	-0.060	0.104	0.119	0.908	0.001	0.104	0.120	0.954
	β_2	-0.200	0.084	0.036	0.041	0.358	0.001	0.038	0.043	0.958
	ρ	-0.500	-0.033	0.099	0.112	0.900	-0.006	0.098	0.109	0.924
	σ_u^2	1.200	-0.080	0.203	0.227	0.906	0.014	0.212	0.237	0.934
	σ_v^2	0.600	0.016	0.058	0.063	0.922	0.013	0.054	0.061	0.934
	γ	-1.500	0.198	0.083	0.101	0.382	-0.013	0.097	0.122	0.972
For survival	h_0	0.300	0.013	0.015	0.017	0.846	0.001	0.018	0.021	0.960
	v_1	-0.600					-0.045	0.146	0.177	0.946
	v_2	0.400					-0.038	0.174	0.202	0.938
For zero-inflated outcomes										
ZIGP model	α_0	0.800	-0.095	0.097	0.091	0.782	-0.009	0.081	0.083	0.940
	α_1	-0.300	0.131	0.083	0.083	0.626	0.005	0.083	0.083	0.944
	α_2	0.200	-0.088	0.026	0.024	0.048	-0.001	0.024	0.024	0.944
	β_0	0.500	0.042	0.118	0.118	0.936	-0.011	0.105	0.114	0.974
	β_1	0.400	-0.058	0.125	0.130	0.936	-0.001	0.121	0.132	0.962
	β_2	-0.200	0.051	0.038	0.040	0.734	0.001	0.044	0.043	0.958
	ω	0.300	0.020	0.031	0.032	0.882	0.001	0.029	0.030	0.958

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Independent model					Joint model				
	True	Bias	SD	SE	CP	Bias	SD	SE	CP	CP
ρ	-0.500	0.050	0.126	0.128	0.904	0.006	0.129	0.123	0.930	0.930
σ_u^2	1.500	0.030	0.335	0.314	0.916	0.093	0.335	0.335	0.944	0.944
σ_v^2	0.700	-0.059	0.092	0.087	0.866	0.011	0.078	0.085	0.960	0.960
For survival										
γ	-1.500	0.343	0.097	0.088	0.052	-0.006	0.117	0.119	0.956	0.956
h_0	0.300	0.010	0.022	0.015	0.812	0.001	0.023	0.022	0.944	0.944
ν_1	-0.300					-0.002	0.145	0.191	0.938	0.938
ν_2	1.200					0.003	0.189	0.228	0.932	0.932

Table 3

Model comparison statistics for the ATBC dataset.

Criterion	Model	
	ZIP independent model	ZIP joint model
DIC	3 941 316	3 937 350
EAIC	1 971 899	1 969 921
EBIC	1 972 031	1 970 069
LPML	−1 002 022	−999 228.6

DIC: deviance information criterion; EAIC: expected AIC; EBIC: expected BIC; LPML: log-pseudo-marginal likelihood.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Results of fitting both the ZIP joint and independent models in the ATBC dataset.

Variables	Joint model				Independent model			
	Mean	SD	Median	95% CI	Mean	SD	Median	95% CI
Probability of smoking abstinence								
Intercept	-9.172	0.068	-9.172 [	-9.305, -9.047]	-8.969	0.340	-9.060 [	-9.277, -8.006]
Smoke years	-0.100	0.006	-0.100 [	-0.112, -0.090]	-0.098	0.007	-0.097 [	-0.112, -0.085]
Age	0.131	0.010	0.131	[0.112, 0.149]	0.126	0.011	0.123	[0.109, 0.149]
Time (year)	0.974	0.007	0.974	[0.960, 0.988]	0.962	0.025	0.970	[0.903, 0.985]
Baseline cigarettes	-0.098	0.005	-0.098 [	-0.107, -0.089]	-0.102	0.010	-0.103 [	-0.117, -0.077]
Expected number of cigarettes per day								
Intercept	2.882	0.002	2.882	[2.879, 2.885]	2.883	0.006	2.882	[2.877, 2.904]
Smoke years	0.003	0.000	0.003	[0.003, 0.004]	0.003	0.000	0.003	[0.002, 0.004]
Age	-0.009	0.000	-0.009 [	-0.010, -0.008]	-0.009	0.000	-0.009	[-0.010, -0.008]
Time (year)	-0.015	0.000	-0.015 [	-0.016, -0.015]	-0.015	0.000	-0.015	[-0.016, -0.014]
Baseline cigarettes	0.041	0.000	0.041	[0.041, 0.042]	0.041	0.000	0.041	[0.040, 0.042]
Time to dropout								
Baseline cigarettes	0.011	0.001	0.011	[0.009, 0.013]	0.011	0.001	0.011	[0.009, 0.013]
ν_1	-0.009	0.003	-0.009	[-0.015, -0.003]				
ν_2	-0.038	0.059	-0.038	[-0.147, 0.080]				
Random eFFects								
ρ	-0.237	0.009	-0.237	[-0.256, -0.219]	-0.126	0.249	-0.217	[-0.442, 0.522]
σ_u	5.136	0.048	5.137	[5.038, 5.227]	5.055	0.171	5.095	[4.561, 5.254]
σ_v	0.243	0.001	0.243	[0.241, 0.246]	0.250	0.010	0.246	[0.241, 0.283]

Table 5

The observed and estimated average daily smoking counts and the estimated random effects of two groups of six subjects in the ATBC study.

Subjects (Group 1)		(a): Observed average daily smoking counts at each visit (1–24).																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1		40	30	30	30																				
2		25	30	30	30	30	30	30	30	30	30	30	35	30											
3		40	35	40	40	30	30	30	0	0	0	0	0	0											
4		25	15	18	18	15	15	15	15	15	15	15	20	15	15										
5		15	15	15	10	15	10	10	13	15	15	15	15	0	0	0	0	0	0	0	0	0	0	0	0
6		10	10	10	10	10	10	10	10	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
(b): Estimated average daily smoking counts at each visit (1–24).																									
1		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
34		34	34	34																					
2		31	31	31	30	30	30	30	30	30	29	29	29	29											
3		29	27	26	23	21	18	15	12	10	8	6	5	3	3										
4		17	17	17	16	16	16	16	16	16	16	16	16	16	16	16									
5		11	11	11	11	11	10	10	10	10	9	8	8	7	6	5	4	3	2	2	1	1	1	1	1
6		9	9	9	8	8	7	7	6	5	5	4	3	3	2	2	1	1	1	1	0	0	0	0	0
Subjects (Group 2)		(c): Observed average daily smoking counts at each visit (1–20).																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20				
1		6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6					
2		20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20				
3		17	17	15	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	5				
4		20	20	20	20	15	20	20	20	20	20	20	0	0	0	0	0	0	0	0	0				
5		20	20	20	20	20	20	15	15	15	15	20	17	15	15	20	15	20	0	0	0				
6		20	20	20	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
		(d): Estimated average daily smoking counts at each visit (1–20).																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20				

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Subjects (Group 1)	(a):Observed average daily smoking counts at each visit (1–24).																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	7	7	7	7	7	7	7	7	7	7	6	6	6	6	6	6	6	6	6	6				
2	22	22	21	21	21	21	21	21	21	21	21	20	20	20	20	20	20	20	20	20				
3	19	19	19	19	19	19	19	19	18	18	18	18	18	18	18	18	17	17	17	16				
4	19	19	18	18	17	17	16	15	14	12	11	9	8	6	5	4	3	2	2	1				
5	17	17	17	17	17	17	16	16	16	16	16	16	15	15	15	14	14	13	12	12				
6	13	12	10	8	7	6	4	3	3	2	1	1	1	1	0	0	0	0	0	0				