Differences in Mortality among Heroin, Cocaine, and Methamphetamine Users: A Hierarchical Bayesian Approach

Li-Jung Liang[1], **David Huang**[2], **Mary-Lynn Brecht**[2], and **Yih-ing Hser**[2]

[1]Department of Medicine Statistics Core, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles

[2]Integrated Substance Abuse Programs, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles

Abstract

Studies examining differences in mortality among long-term drug users have been limited. In this paper, we introduce a Bayesian framework that jointly models survival data using a Weibull proportional hazard model with frailty, and substance and alcohol data using mixed-effects models, to examine differences in mortality among heroin, cocaine, and methamphetamine users from five long-term follow-up studies. The traditional approach to analyzing combined survival data from numerous studies assumes that the studies are homogeneous, thus the estimates may be biased due to unobserved heterogeneity among studies. Our approach allows us to structurally combine the data from different studies while accounting for correlation among subjects within each study. Markov chain Monte Carlo facilitates the implementation of Bayesian analyses. Despite the complexity of the model, our approach is relatively straightforward to implement using WinBUGS. We demonstrate our joint modeling approach to the combined data and discuss the results from both approaches.

Keywords

Long-term substance abusers; Heroin; Cocaine; Methamphetamine; Joint model

Introduction

One of the important topics in substance abuse research is whether mortality is associated with long term use of different substances and alcohol consumption. However, studies examining differences in mortality among substance users over a long period of observation have been limited. Previous studies have discussed the mortality caused by drug overdose among polydrug or multi-drug users (Shah, Lathrop, Reichard, & Landen, 2008; Pereiro, Bermejo, Fernandez, & Tabernaro, 2003). Other studies have examined mortality among long-term heroin addicts (e.g., Smyth, Hoffman, Fan, & Hser, 2007; Sanchez-Carbonell & Seus, 2000; Oppenheimer, Tobutt, Taylor, & Andrew, 1994), cases of cocaine- and opiate-related overdoses (Bernstein et al., 2007; Coffin et al., 2003), and HIV-infected subjects who were heroin or cocaine users with alcohol problems (Walley et al., 2008). Others discuss the mortality rates among subjects with a history of addiction to heroin, cocaine, and/or amphetamines (e.g., Termorshuizen, Krol, Prins, & van Ameijden, 2005). Still, more needs to be known regarding the differences in mortality among long-term heroin, cocaine, and methamphetamine abusers. In addition, the literature seems far from being in concrete agreement on whether mortality is associated with age, education, duration of use, or overdose (Odegard, Amundsen, & Kielland, 2007). For example, one study reported that younger age at first drug use and lower educational status are risk factors associated with overdose death

(Davoli et al., 1993); however, another study reported that duration of abuse, but not aging, is found to be a risk factor of fatal overdose (Odegard et al., 2007).

In this paper, we investigate whether and how characteristics of long-term drug abusers (gender, racial/ethnic background, age at first use substance, etc.) and their overall substance and alcohol use are associated with their survival. Available data from five studies of long-term heroin, cocaine, and methamphetamine addicts obtained from the UCLA Center for Advancing Longitudinal Drug Abuse Research (CALDAR, http://www.caldar.org/) are used. These five long-term follow-up studies, for which addicts were recruited from different resources or agencies, are likely to be heterogeneous.

Standard statistical methods for analyzing mortality and time-to-event data are available and easy to implement in many statistical packages, such as a proportional hazard (PH) model on time-to-event data for finding risk factors. A common problem with most survival models is that covariates may not fully represent the hazard function and as such, estimates may be biased due to unobserved heterogeneity. We wish to model the study-level heterogeneity associated with survival, primary substance use, and alcohol consumption simultaneously. However, the conventional methods, which cannot simultaneously incorporate the study-level of heterogeneity associated with different types of outcome measures, may not be efficient in our situation (Wulfsohn & Tsiatis, 1997).

We propose a Bayesian framework that jointly models the survival data with substance and alcohol data to examine differences in mortality among heroin, cocaine, and methamphetamine users from the five long-term follow-up studies. To model the survival data, the proportional hazard (PH) with a shared frailty model provides a convenient way to account for association in clustered time-to-death data (e.g., Clayton & Cuzick, 1985; Gutierrez, 2002). A shared frailty model is a random effects model where individuals within the same study are assumed to be correlated because they share the same frailty. Thus, we use a Weibull PH with a shared frailty model to examine the clustered time-to-death data (Sahu, Dey, Aslanidou, & Sinha, 1997). Similarly, mixed-effects models with study-level random effects are used to model the substance and alcohol data, which can be linked to the survival model.

Our method has several advantages. First, our joint modeling approach allows us to hierarchically combine the data from different types of studies, accounting for correlation among subjects within each study, as well as to estimate systematic and random components within and across studies. Second, since fitting such models involves high dimensional integration, maximum likelihood approaches are difficult to implement except under normality and linearity assumptions. Recent developments in Markov chain Monte Carlo (MCMC) methodology (Gilks, Richardson & Spiegelhalter, 1996; Tierney, 1994; Smith & Gelfand, 1992) and rapid improvements in computing speed have made biomedical applications using Bayesian methods feasible and increasingly popular. A major advantage of the Bayesian approach is that it provides flexibility in implementing a complex hierarchical model that involves different types of data using MCMC techniques. Our approach can be implemented in a relatively straightforward fashion using the freely available statistical software WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000; http://www.mrc-bsu.cam.ac.uk/bugs/), despite the complexity of the model. Researchers interested in applying Bayesian methods to their data can work through some of the examples provided at the WinBUGS website (Spiegelhalter, Thomas, Best, & Lunn, 2002). Third, a Bayesian approach has another important benefit, which is the ability to incorporate information from previous studies (e.g., estimates of the regression coefficients) through an informative prior distribution (Gelman, Carlin, Stern, & Rubin, 2003). Investigators who wish to avoid incorporating prior information from previous studies, to avoid bias, can certainly choose vague (non-informative) priors for the unknown model parameters.

Methods

Datasets and Samples

Data were analyzed from five long-term follow-up studies in which subjects had reported their primary drug problem as heroin, cocaine, or methamphetamine. All studies were observational and conducted in California. Studies included Heroin Follow-up Study (HFS, N=472; Hser, Hoffman, Grella, & Anglin, 2001), Cocaine Follow-up Study (CFS, N=319; Hser et al., 2006), Methamphetamine Natural History (METH, N=346; Brecht, O'Brien, Mayrhauser, & Anglin, 2004), Treatment Utilization and Effectiveness (TUE, N=265; Hser, Teruya, & Anglin, 2003), and Treatment Process Study (TPS, N=391; Hser, Huang, Teruya, & Anglin, 2004).

Study Description—The HFS study examined male heroin addicts previously committed to the California Civil Addict Program (CAP), which was a compulsory drug treatment program for narcotic-dependent criminal offenders committed under court order. The study was conducted from 1964 to 1997. The CFS study examined cocaine-dependent men followed for approximately 12 years. Participants who met the DSM III-R criteria for cocaine dependence were recruited between 1988 and 1989 as part of a larger study, and admitted to the West Los Angeles Veterans Affairs Medical Center. The METH study examined methamphetamine (meth) users who were admitted to residential or outpatient treatment facilities and enrolled in publicly funded Los Angeles County programs from 1995 to 2002. The TUE study included subjects recruited from non-treatment settings (emergency rooms, sexually transmitted disease clinics, and jails) starting in 1993 and ending in 1996. The TPS study recruited addicts from five major treatment modalities (hospital inpatient, residential therapeutic community, methadone maintenance, outpatient drug-free, and day treatment) starting in 1995 and ending in 1998. The HFS and CFS studies had male addicts only, but the other three studies had both male and female addicts. The primary drug (i.e., drug for which the subject was in treatment at the baseline assessment) was methamphetamine for the Meth study, heroin for the HFS study, and cocaine for the CFS study. The primary drug was self-identified in the other two studies. Note that while many of these subjects reported use of drugs other than their primary drug, a separate analysis shows that use of other drugs was generally at a much lower level than the primary drug (Brecht, Huang, Evans, & Hser, 2008).

Death Records—Death certificates were acquired for subjects determined to have died during their respective studies. Furthermore, in October 2006, we searched the National Death Index for possible death records on the remaining subjects.

Outcome Measures

Length of survivorship was computed as months from first primary substance use to death or to the end of the observation period, October 2006. Subjects were considered as censored if they were alive at the end of the observation period. *Primary substance usage* was a percentage calculated as the number of months using primary substance divided by the number of months from the first primary substance use to the end of the follow-up, excluding any incarceration time. *Alcohol usage* was calculated in a similar way.

Statistical Methods

In this paper, we are interested in (1) estimating survivorship of long-term substance abusers, (2) examining differences in survival curves among three types of primary drugs, and (3) studying risk factors. Commonly used statistical methods that address these interests are described in the next section and the proposed hierarchical models are described in the subsequent sections (beginning with "A Bayesian Joint Hierarchical Model").

Mortality Rates and Kaplan-Meier Survival Curves—To address the first two questions of interest, we calculated the mortality rate for users of each of the three primary drugs within different lengths of substance use since the first primary substance use. The comparison of mortality rates among users of the three primary substances was done using Chi-square tests. The survival curve was estimated using the Kaplan-Meier method, and differences on survival curves among the three types of substance users were assessed by a log-rank test. In order to systematically investigate survival patterns among the three types of drug users over the years since the first primary drug use, these statistical methods were applied separately to deaths that occurred within 20 years, 30 years, and by the end of combined study (October 2006). The statistical analyses described in this subsection were implemented in SAS; PROC FREQ procedure for Chi-square tests and PROC LIFETEST procedure for the Kaplan-Meier estimations and log-rank tests were used.

A Bayesian Joint Hierarchical Model—Since the drug users were recruited from five different studies in diverse calendar years, heterogeneity among studies may bias the estimates on survivor function. To structurally combine the data from different types of studies, accounting for correlation among subjects within the same study, and to estimate systematic and random components within and across studies, we propose a Bayesian framework that hierarchically joins the survival data, using a Weibull proportional hazard (PH) model with frailty, with the substance and alcohol usage data using mixed-effects models. Our approach allows us to fit the three models simultaneously, and Markov chain Monte Carlo (MCMC) methodology facilitates the implementation of Bayesian analyses of complex data containing survival and substance and alcohol use data. WinBUGS was used to implement our method. Our models are described in the next three sub-sections.

<u>**Weibull PH with Frailty Model – Time to Death:**</u> We model the conditional hazard function $h_{ijk}(t)$ by a Weibull proportional hazard model with study-level shape parameter r_i and frailty b_i,

$$h_{ijk}(t) = t^{r_i-1} \times r_i \times b_i \times e^{\beta' X_{ijk}}, \ i=1,\ldots,5, \ j=1,\ldots,3, k=1,\ldots,n_{ij},$$

where t is years since subject k in study i initially used primary drug j, and X_{ijk} is the subject-level covariates. Studies have shown that the Weibull model is appropriate for mortality trajectories (e.g., Manton, Stallard, & Vaupel, 1986). In other words, to control for possible confounding effects, we model the time-to-death outcome t_{ijk} as a Weibull distribution with study-level shape parameter r_i, individual location parameter $\mu_{ijk} = b_i \times m_{ijk} = b_i \times e^{\beta' X_{ijk}}$, and the study-level random effects (frailty) b_i. The subject-level covariates in the proportional hazard model included age of enrollment, gender, ethnicity, and age at first primary substance use. The survival model specification is as follows.

$$
\begin{aligned}
t_{ijk} &\sim Weibull \ (\mu_{ijk}, r_i) \\
\mu_{ijk} &= m_{ijk} \times b_i \\
\log(m_{ijk}) &= \beta_0 + \beta_1 \times Coca_{ijk} + \beta_2 \times Meth_{ijk} \\
&+ \beta_3 \times ageofonset_{ijk} + \beta_4 \times gender_{ijk} + \beta_5 \times DrugFirstUse_{ijk} \\
&+ \beta_6 \times White_{ijk} + \beta_7 \times Black_{ijk} \\
&+ \beta_9 \times U_{ijk} + \beta_{10} \times A_{ijk},
\end{aligned}
\tag{1}
$$

where $U_{ijk} = usage_{ijk} - \eta^u_{ij}$, $A_{ijk} = alcohol_{ijk} - \eta^a_{ij}$, and $b_i \sim Gamma \ (\theta^{-1}, \theta^{-1})$. In equation (1), η^u_{ij} and η^a_{ij} are the mean substance usage and mean alcohol usage, respectively, for the

jth drug-type in the ith study. Frailty b_i follows a Gamma distribution with mean one for purpose of model identifiability and variance θ, which characterizes the heterogeneity across studies and induces dependence within a study. One of the methods quantifying the unconditional dependence in survival time is Kendall's τ, which is $\theta/(2+\theta)$. Larger θ means stronger dependence within a study.

Mixed-effects Regression Models – Substance and Alcohol Usages: The primary substance usage ($usage_{ijk}$) follows a normal distribution with mean η^u_{ij} and precision parameter τ_1. The mean parameter η^u_{ij} is a function of two indicators from three types of substances and the study-level of random effects ξ_i. The regression coefficient α_1 represents the difference in overall substance usage between cocaine and herion users, and α_2 represents the difference in overall substance usage between meth and heroin users. The mixed-effects model can be written as follows.

$$usage_{ijk} \sim Normal\ (\eta^u_{ij}, \tau_1)$$
$$\eta^u_{ij} = \alpha_0 + \alpha_1 \times Coca_{ijk} + \alpha_2 \times Meth_{ijk} + \xi_i$$

(2)

Similarly, we model the alcohol usage ($alcohol_{ijk}$) as a normal distribution with mean η^a_{ij} and precision parameter τ_2. The same setting as the mean parameter for the substance usage is used for the mean parameter η^a_{ij} for the overall alcohol usage. The mixed-effects model is as follows.

$$alcohol_{ijk} \sim Normal\ (\eta^a_{ij}, \tau_2)$$
$$\eta^a_{ij} = \gamma_0 + \gamma_1 \times Coca_{ijk} + \gamma_2 \times Meth_{ijk} + \zeta_i,$$

(3)

where γ_1 is the regression coefficient for the difference in overall alcohol usage between cocaine and heroin users, γ_2 is the regression coefficient for the difference in overall alcohol usage between meth and heroin users.

Prior Specification: We adopt a Bayesian approach, which assumes prior distributions for all unknown parameters. The parameters in the Weibull PH model include the regression coefficients $\beta's$ and the variance parameter θ of the study-level random effects. The intercept follows a normal distribution with mean parameter derived by setting a median survival equal to the pre-specified number of years from first use of substance and precision 1.0E-6. The pre-specified number can be from previous studies, or in our situation, is estimated approximately to be the mean age at enrollment (=32) plus 30 years of observation and minus the mean age of first drug use (=20). As mentioned in the introduction, incorporation of prior information from previous studies or literature is one of many practical advantages to the Bayesian approach. Each of the other regression coefficients is assumed to be distributed as a normal with mean zero and precision 1.0E-6, which is relatively flat. The variance parameter of the frailty follows an exponential distribution with mean one.

Non-informative priors for all the regression coefficients for the substance and alcohol usage models were used in our application. The random effects in Models (2) and (3) have normal priors with mean zero and variances D_1 and D_2, respectively. The normality assumption on the random effects can be relaxed and replaced by non-parametric prior, such as Dirichlet process prior (Bush & MacEachern, 1997). Intraclass correlations (ICCs) for Models (2) and (3) are calculated as $D_1 / (D_1 + 1/\tau_1)$ and $D_2 / (D_2 + 1/\tau_2)$, respectively. ICC provides a measure of the proportion of total variation that is explained by between-study variability. The precision parameters τ_d, d=1, 2, follow Gamma distributions with shape parameters a_d and scale

parameters λ_d. The variance parameters D_d follow inverse Gamma distributions with shape parameters c_d and scale parameters ω_d. The unknown hyper-parameters a_d, λ_d, c_d and ω_d are pre-specified.

Bayesian Computation - Posterior Distributions and Interpretation—According to Bayes' theorem, the joint posterior distribution is proportional to the product of the likelihood function and priors,

$$p_r(\beta_1, \ldots, \beta_p, \alpha_1, \alpha_2, \alpha_3, \gamma_1, \gamma_2, \gamma_3, b_i, \tau_1, \tau_1, \theta | T, Usage, Alcohol) \propto$$
$$\prod_{i,j,k} \{ p_r(t_{ijk}, usage_{ijk}, alcohol_{ijk} | \beta_1, \ldots, \beta_p, \alpha_1, \alpha_2, \alpha_3, \gamma_1, \gamma_2, \gamma_3, b_i, \tau_1, \tau_2, \theta) \times p_r(b_i | \theta) \}$$
$$\times p_r(\beta_1, \ldots, \beta_p) \times p_r(\theta) \times p_r(\alpha_1, \alpha_2, \alpha_3) \times p_r(\gamma_1, \gamma_2, \gamma_3) \times p_r(\tau_1) \times p_r(\tau_2),$$

where the likelihood receives more and more weights as the sample size increases. The posterior distribution summarizes the state of knowledge about all the unknown parameters conditional on the choice of prior, different types of outcome, observed data, and the hierarchical model described above (under "A Bayesian Joint Hierarchical Model").

Next, since there was no closed form for the joint posterior distribution, we simulated the posterior distributions of the parameters using MCMC algorithm, which iteratively generated samples of the parameters in a statistical model. These posterior distributions can be summarized using posterior means, medians, and 95% credible intervals (CI). The interpretation of Bayesian interval estimates, being the interval containing the true parameter with some probability (e.g., 95%), is intuitively appealing. The interval estimates are appropriate in small samples.

An additional advantage of using the Bayesian approach is the use of posterior probability. For example, we can obtain estimates of the posterior distributions of the hierarchical regression coefficients (β's and α's) and functions of these regression coefficients (i.e., hazard ratio e^β). We can also estimate the posterior probability that a regression coefficient is positive (or negative) or equivalently that a hazard ratio is greater (or less) than 1. This posterior probability is more intuitive than the frequentist's p value, which is the chance of observing a value as extreme as the observed value given repeated sampling under the null.

Results

Background characteristics of participants and onset of primary drug use were significantly different by the primary drug type (Table 1). In the combined study, there were 470 (26.2%) meth users, 629 (35.1%) heroin users, and 694 (38.7%) cocaine users. Overall 72.8% of participants were male, and 34.4% were White, 32.2% Black, 29.8% Hispanic, and 3.6% Asian or of other racial/ethical groups. On average, onset of primary drug use occurred when participants were 21 years old.

Most primary heroin users in the sample were male (89.4%), Hispanic (53.4%) or White (35.5%), and started their heroin use at around the age of 19. In contrast, while most of primary cocaine users were also male (79.6%), the majority were Black (66%), and on average, started cocaine use at a much later age (23 at first use). Relative to users of heroin and cocaine, primary meth users were more often women (46%), more than half of them were White (54.3%), and they initiated meth use at around the age of 20. The median length of observation for the heroin addicts was 32 years, which was 13 years longer than that for cocaine and meth addicts; however, the median length of observation without counting incarcerated time for the heroin addicts (23 years) was at least 5 years longer than that for the cocaine and meth addicts. In

terms of primary substance usage, heroin addicts used at least 10% more frequently on average than cocaine or methamphetamine addicts (p < 0.0001), whereas their alcohol use was at least 7% less frequent on average than cocaine or methamphetamine users (p < 0.0001).

Mortality rates by type of primary drugs

Table 2 summarizes mortality rates among the three types of drug users. Mortality in the first 10 years of primary substance use was not significantly different between heroin and cocaine users. However, mortality among heroin users increased dramatically after 20 years of heroin use. Cocaine users had a significantly higher mortality than meth users after 20 years of primary substance use. After 30 years of substance use, the heroin users had the highest mortality (16%), which was 2.5 times higher than cocaine users (6.5%), and slightly more than 10 times higher than meth users (1.5%).

Table 2 also presents mortality by type of substance and by gender. The mortality rates for both male (50.7%) and female (16.4%) heroin users were the highest among the three types of primary drug users (p-values < 0.0001). In contrast to male addicts, female meth users had a higher mortality rate than female cocaine users.

The median age of first drug use (not shown) among the 54 decedents in the cocaine group was 24.5 (range: 11-58), which was older than those for the heroin (N=296; median=18; range: 11-38) and meth (N=11; median=16; range: 13-24) groups. However, the median length of drug career (not shown) among the 296 decedents in the heroin group was 36 years since their first use of substance, which was longer than those for the cocaine (median=19.5; range: 3-47) and meth (median=22; range: 11-38) groups.

Figure 1 presents Kaplan-Meier (KM) survival curves of the three types of drug users. In Figure 1(a), the KM survival curves for primary heroin and cocaine users were closer than that for primary meth users. After 10 or more years of substance use, survivals in Figure 1(b) started to separate, with heroin users having the most drop. The KM survival curves of primary heroin and cocaine users in Figure 1(c) went down significantly faster than that of primary meth users (Log-rank test: p < 0.0001).

Posterior Results from Joint Hierarchical Models

We fit three hierarchical models; Model 1 is an overall model, and Models 2 and 3 are stratified by gender. The first two models involve all the studies, whereas Model 3 only involves three studies (METH, TUE, and TPS). As described in the "A Bayesian Joint Hierarchical Model" section, in the overall model (Model 1), the covariates included in the survival model (EQ 1) were gender, age at enrollment, indicators of primary substance group, ethnicity, age at first substance use, and overall primary substance and alcohol usages. Model 2 excluded gender and Model 3 excluded gender and age at enrollment. Since the regression coefficients for age at enrollment and age at first drug usage for female addicts were highly correlated, the covariate "age at enrollment" was removed from Model 3. The MCMC algorithms were run for 65,000 iterations with the first 5,000 iterations excluded as a burn-in period. We also ran multiple chains with different starting values and graphically examined the convergence.

Table 3 presents the posterior results (posterior median hazard ratio (HR) and 95% credible interval) from the three hierarchical models. The results from Model 1 indicated that primary heroin users had a significantly higher risk of death than primary cocaine and meth users. The posterior median HRs for cocaine versus heroin and meth versus heroin are 0.46 (95% CI: 0.24-0.79) and 0.28 (95% CI: 0.13-0.62), respectively. The respective posterior probabilities (*P*) of a value greater than 1 for the HR are 0.003 and < 0.0001. Risks were not significantly different between cocaine and meth users. Male primary meth users had a significantly lower

risk than male primary heroin and cocaine users (Model 2: posterior median HRs: 0.14 (95% CI: 0.03-0.49) and 0.22 (95% CI: 0.05-0.90); P=0.002 and 0.015, respectively), whereas there was no risk difference between male heroin and cocaine users. A significantly lower risk for female primary cocaine users versus heroin users (Model 3: posterior median HR=0.12; 95% CI: 0.03-0.44; P=0.001) was observed, whereas there was no risk difference between female primary meth users and the other two female groups.

The effects of demographic covariates across the three models are summarized in Table 3. Risk of death increased with older age at first drug use, and Hispanic/others in contrast to Blacks had a significantly higher risk of death. An increasing frequency of primary substance usage for these long-term addicts (posterior median HR=2.69; 95% CI: 1.80-4.16; $P< 0.0001$) was significantly associated with risk of death. Similar results were seen for male users (HR=2.79, 95% CI: 1.82-4.32; $P< 0.0001$). However, for the female addicts, it was not the substance use that was significant but rather the alcohol usage (posterior median HR=3.42, 95%CI: 1.06-13.2; P=0.021). The posterior mean measure of dependence from Model 3 (Kendall's $\tau = 0.17$) is highest among the three models, implying that the female users were relatively more heterogeneous across the three studies (METH, TUE, and TPS). This can be seen in Figure 2, which presents the posterior Kernel Density Estimate (KDE) curves for the Kendall's τ.

The results from the mixed-effects model on overall substance usage (in Model 1) indicated that primary cocaine and meth addicts used 7% and 8% less than heroin users ($P< 0.05$). Male users had slightly larger differences ($P< 0.05$), whereas female users had slightly smaller differences. The posterior median intra-class correlation (ICC) across the studies (5 studies for Models 1 and 2, and 3 studies for Model 3) for all three models ranges from 0.39 to 0.57, where the posterior median ICC was higher for the female (0.57) than male (0.39) model. This agrees with what we found in the survival models (Kendall's τ). No significant difference in overall alcohol usage was observed among the three drug groups, and the posterior mean ICCs for overall alcohol usage were less than those for the overall substance usage.

We also performed sensitivity analyses to evaluate the influence of our prior specification in all three hierarchical regression models. Results were robust when we changed the prior means for the regression coefficients. When we removed the sub-models for the overall substance use and the overall alcohol use from our proposed framework, the findings were in a similar direction, but the estimates of parameter were less precise, indicating the benefit of modeling all three types of outcomes jointly.

Lastly, the conventional approach, which is the overall model (Model 1) that includes the overall substance and alcohol uses as covariates, but excludes the frailty and the sub-models, was performed using the publicly available statistical software R (R Development Core Team, 2008). As expected, the results were in a similar direction. The interesting finding is that the proposed joint modeling approach indicated larger differences in the risk of death for both the primary cocaine versus heroin users and the primary meth versus heroin users, as compared to those from the conventional approach (e.g., posterior median HRs for cocaine versus heroin: 0.46 (95% CIs: 0.24-0.79) from our approach and 0.68 (95% CI: 0.45-1.01) from the conventional approach). By modeling the three types of data (survival, substance, and alcohol) jointly and accounting for the unobserved heterogeneity among the studies, our approach resulted in more precise estimates as compared to the conventional approach.

Discussion

The traditional approach to analyzing combined survival data from, in this case, five different long-term follow-up studies, either assumes that the studies are homogeneous or includes *study* as one of the covariates. However, the survival estimates may be biased due to unobserved

heterogeneity, meaning that the included covariates may not fully represent the hazard function. In this article, we provide drug abuse researchers with our alternative method for analyzing the time-to-event, substance, and alcohol data simultaneously. We take the advantages of Bayesian models, which are flexible to (a) handle the complex structure of data and different types of outcomes (e.g., time-to-event and substance data from different types of studies), (b) account for correlation among subjects within each study, and (c) estimate systematic and random components within and across studies. A Bayesian approach, as implemented by simulation, provides flexibility with which posterior interferences can be summarized even after complicated transformations, such as cumulative hazard. MCMC methodology facilitates the implementation of Bayesian analyses of complex data containing survival, substance use, and alcohol consumption.

Our method, which was implemented using the freely available software WinBUGS, was used to investigate the differences in mortality among long-term heroin, cocaine, and meth users. Our findings suggest that, taking into account the correlations among subjects within the same study for time-to-event, substance, and alcohol data simultaneously, heroin users had a significantly higher risk of death than cocaine and meth users. Participants' risk of death increased with older age at first use of substance or with higher usages of their primary substance. In terms of ethnicity, Hispanics and those from "other" ethnic groups had a higher risk of death than did Blacks. The gender-specific analyses indicated that male heroin users and male cocaine users had a significantly higher risk of death than male meth users, and female heroin users had a significantly higher risk than female cocaine users. We also found that an increase of primary substance use for these long-term addicts was significantly associated with risk of death. However, for the female addicts, it was not their substance use that was significant but rather their alcohol use. In the overall substance usage model, we found that the overall substance use was significantly more frequent for the heroin addicts than for the cocaine or methamphetamine addicts.

Our combined study is limited because the length of observation was different among the three primary drug groups; observations of the long-term heroin addicts were, on average, much longer than those of the other two primary drug groups. Almost 80% of the heroin addicts versus less than 50% of the cocaine or meth addicts had at least 20 years of observations. However, in terms of the length of observation minus incarceration time, the median for the heroin addicts was only 5 years longer than that for the other two drug groups.

Our method can be extended. One possible extension would be to take into account the treatment data. Treatment for substance use is a potentially important factor associated with mortality. Our current models did not take into account the treatment data because (1) subjects may not have received prior treatment for their substance use (e.g., subjects recruited from emergency room), (2) subjects received treatment under sporadic conditions (e.g., irregular treatment time, variable dosage), and (3) treatment data for some subjects were not available. However, there are methods in the literature (Robins, Hernan, Brumback, 2000; Hernan, Brumback, & Robins, 2000) that might be used to address these treatment issues and could be built on top of the framework proposed in this article.

Another possible extension is to model the longitudinal substance and alcohol usage data with growth mixture models (Muthen & Muthen, 2000; Nagin, 1999), mixed-effects models (Guo & Carlin, 2004; Henderson, Diggle & Dobson, 2000; Wulfsohn & Tsiatis, 1997), or Bayesian semi-parametric models (Brown & Ibrahim, 2003; Kleinman & Ibrahim, 1997). The current submodels for the substance and alcohol usage data were simplified by taking everyone's summary, which is the percent usage of primary substance and alcohol over the individual study period. The individual longitudinal trajectories of different substances could be incorporated into the survival model and estimated simultaneously. The joint Bayesian

approach considered in this paper can be extended to accommodate the longitudinal substance data using these parametric and non-parametric methods. This type of approach is becoming increasingly useful in many applications.

Acknowledgments

References

Bernstein KT, Bucciarelli A, Piper TM, Gross C, Tardiff K, Galea S. Cocaine- and opiate-related fatal overdose in New York City, 1990-2000. BMC Public Health 2007;7(1):31. [PubMed: 17349051]

Brecht ML, Huang D, Evans E, Hser YI. Polydrug use and implications for longitudinal research: Ten-year trajectories for heroin, cocaine, and methamphetamine users. Drug and Alcohol Dependence 2008;96(3):193–201. [PubMed: 18329825]

Brecht M, O'Brien A, Mayrhauser CV, Anglin MD. Methamphetamine use behaviors and gender differences. Addictive Behaviors 2004;29:89–106. [PubMed: 14667423]

Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics 2003;59:221–228. [PubMed: 12926706]

Bush CA, MacEachern SN. A semiparametric Bayesian model for randomised block designs. Biometrika 1996;83:275–285.

Clayton DG, Cuzick J. Multivariate generalizations of the proportional hazards model. Journal of Royal Statistical Society: Series B 1985;34:187–220.

Coffin PO, Galea S, Ahern J, Leon AC, Vlahov D, Tardiff K. Opiates, cocaine and alcohol combinations in accidental drug overdose deaths in New York City, 1990-98. Addiction 2003;98(6):739–747. [PubMed: 12780362]

Davoli M, Perucci CA, Forastiere F, Doyle P, Rapiti E, Zaccarelli M, Abeni DD. Risk factors for overdose mortality: A case-control study within a cohort of intravenous drug users. International Journal of Epidemiology 1993;22:273–277. [PubMed: 8505184]

Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian data analysis. 2nd. Boca Raton, FL: Chapman & Hall/CRC; 2003.

Gilks, WR.; Richardson, S.; Spiegelhalter, DJ. Markov Chain Monte Carlo in practice. London: Chapman & Hall/CRC; 1996.

Guo X, Carlin BP. Separate and joint modeling of longitudinal and event time data using standard computer packages. The American Statistician 2004;58:16–24.

Gutierrez RG. Parametric frailty and shared frailty survival models. The Stata Journal 2002;2:22–44.

Henderson R, Diggle P, Dobson A. Joint modeling of longitudinal measurements and event time data. Biostatistics 2000;4:465–480. [PubMed: 12933568]

Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology 2000;11:561–570. [PubMed: 10955409]

Hser YI, Hoffman V, Grella CE, Anglin MD. A 33-year follow-up of narcotics addicts. Archives of General Psychiatry 2001;58:503–508. [PubMed: 11343531]

Hser YI, Huang D, Teruya C, Anglin MD. Gender differences in drug abuse treatment outcomes and predictors. Drug and Alcohol Dependence 2003;72:255–264. [PubMed: 14643942]

Hser YI, Huang D, Teruya C, Anglin MD. Diversity of drug abuse treatment utilization patterns and outcomes. Evaluation and Program Planning 2004;27(3):309–319.

Hser YI, Stark ME, Paredes A, Huang D, Anglin MD, Rawson R. A 12-year follow-up of a treated cocaine-dependent sample. Journal of Substance Abuse Treatment 2006;30:19–226.

Kleinman KP, Ibrahim JG. A semiparametric Bayesian approach to the random effects model. Biometrics 1997;54:921–938. [PubMed: 9750242]

Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modeling framework: concept, structure, and extensibility. Statistics and Computing 2000;10(4):325–327.

Manton KG, Stallard E, Vaupel JW. Alternative models for the heterogeneity of mortality risks among the aged. Journal of American Statistical Association 1986;81(395):635–644.

Muthen B, Muthen L. Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. Alcoholism: Clinical and Experimental Research 2000;24:882–891.

Nagin DS. Analyzing developmental trajectories: A semiparametric, group-based approach. Psychological Methods 1999;4:139–157.

Odegard E, Amundsen E, Kielland K. Fatal overdoses and deaths by other causes in a cohort of Norwegian abusers – A competing risk approach. Drug and Alcohol Dependence 2007;89(2-3):176–182. [PubMed: 17240086]

Oppenheimer E, Tobutt C, Taylor C, Andrew T. Death and survival in a cohort of heroin addicts from London clinics: A 22-year follow-up study. Addiction 1994;89:1299–1308. [PubMed: 7804091]

Pereiro C, Bermejo AM, Fernandez P, Tabernaro MJ. Deaths from drug abuse in northwestern Spain, 1992-97. Addiction Biology 2003;8:85–95.

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008 [September 29, 2009]. from http://www.R-project.org

Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550–560. [PubMed: 10955408]

Sahu SK, Dey DK, Aslanidou H, Sinha D. A Weibull regression model with Gamma frailties for multivariate survival data. Lifetime Data Analysis 1997;3:123–137. [PubMed: 9384618]

Sánchez-Carbonell X, Seus L. Ten-year survival analysis of a cohort of heroin addicts in Catalonia: The EMETYST project. Addiction 2000;95:941–948. [PubMed: 10946442]

Shah NG, Lathrop SL, Reichard RR, Landen MG. Unintentional overdose death trends in New Mexico, USA, 1990-2005: Combinations of heroin, cocaine, prescription opioids and alcohol. Addiction 2008;103:126–136. [PubMed: 18028518]

Smith AFM, Gelfand A. Bayesian statistics without tears: A sampling-resampling perspective. The American Statistician 1992;46:84–88.

Smyth B, Hoffman V, Fan J, Hser YI. Years of potential life lost among heroin addicts 33 years after treatment. Preventive Medicine 2007;44:369–374. [PubMed: 17291577]

Spiegelhalter, D.; Thomas, A.; Best, N.; Lunn, D. WinBUGS User Manual Version 1.4. Cambridge, UK: MRC Biostatistics Unit; 2002.

Termorshuizen F, Krol A, Prins M, van Ameijden EJC. Long-term outcome of chronic drug use the Amsterdam cohort study among drug users. American Journal of Epidemiology 2005;161:271–279. [PubMed: 15671259]

Tierney L. Markov chains for exploring posterior distributions. The Annals of Statistics 1994;22:1701–1762.

Walley AY, Cheng DM, Libman H, Nunes D, Horsburgh CR, Saitz R, Samet JH. Recent drug use, homelessness and increased short-term mortality in HIV-infected persons with alcohol problems. AIDS 2008;22:415–420. [PubMed: 18195568]

Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. Biometrics 1997;53:330–339. [PubMed: 9147598]
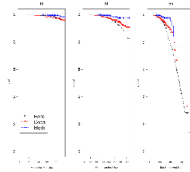
Figure 1.
Kaplan-Meier Curves for (a) within 20 years and (b) 30 years of first use of substance, and (c) from first use of substance to October, 2006. The curves with circle (black), triangular (red), and plus (blue) are for Heroin, Cocaine, and Methamphetamine, respectively. The horizontal dotted lines are 80% and 50% survival.
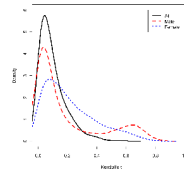
Figure 2.
Posterior Kernel Density Estimates (KDEs) of Kendall's τ for all addicts (solid), male addicts (dash), and females addicts (dot).

Table 1

Substance Abuser's Characteristics by Type of Primary Drug

Parameter	Primary Drug			
	Heroin n=629	Cocaine n=694	Meth n=470	Total N=1793
Male (%) [**]	562 (89.4)	490 (79.6)	254 (54.0)	1306 (72.8)
Age at Enrollment (yr) [**,1]				
30 or younger (%)	424 (67.4)	179 (25.8)	210 (44.7)	813 (45.3)
Mean (SD)	28.5 (9.1)	34.9 (6.8)	32.5 (7.2)	32.0 (8.2)
Age at First Use (yr) [**,1]				
20 or younger (%)	476 (75.7)	297 (42.8)	313 (66.6)	1086 (60.6)
Mean (SD)	18.9 (4.7)	23.0 (6.8)	19.7 (5.5)	20.7 (6.1)
Ethnicity (%) [**,2]				
White	223 (35.5)	138 (19.9)	255 (54.3)	616 (34.4)
Black	58 (9.2)	457 (66.0)	61 (13.0)	576 (32.2)
Hispanic	336 (53.4)	75 (10.8)	123 (26.2)	534 (29.8)
Asian/Other	12 (1.9)	22 (3.2)	31 (6.6)	65 (3.6)
Length of Observation (yr) [**,1]				
w/ incarceration				
≥ 20 years (%)	502 (79.8)	291 (41.9)	202 (43.0)	995 (55.5)
Median (Range)	32 (3-58)	19 (2-55)	19 (3-41)	22 (2-58)
w/o incarceration				
≥ 20 years (%)	383 (60.9)	263 (37.9)	164 (34.9)	810 (45.2)
Median (Range)	23 (3-50)	18 (2-48)	17 (1-41)	19 (1-50)
Primary Substance % Usage [**,1]				
Mean (SD)	61.0 (26.0)	50.8 (23.5)	44.1 (24.0)	52.6 (25.4)
Alcohol % Usage [**,1]				
Mean (SD)	45.3 (33.4)	62.0 (33.8)	52.3 (32.5)	53.6 (34.1)

[**] $p < 0.001$;

[1] p-value from Analysis of Variance;

[2] p-value from Chi-square test.

Table 2

Mortality Rates (Overall and by Gender) by Type of Primary Drug

	Primary Drug N (%)			Total N (%)
	Heroin n=629	Cocaine n=694	Meth n=470	N=1793
No. of deaths within the first				
10 years	4 (0.6)	7 (1.0)	1 (0.2)	12 (0.7)
20 years[*]	30 (4.8)	28 (4.0)	6 (1.3)	64 (3.6)
30 years[**]	101 (16.1)	45 (6.5)	7 (1.5)	153 (8.5)
By October, 2006[**]	296 (47.1)	54 (7.8)	11 (2.3)	361 (20.1)
Male[**]	N=562 285 (50.7)	N=490 51 (10.4)	N=254 4 (1.6)	N=1306 340 (26.0)
Female[**]	N=67 11 (16.4)	N=204 3 (1.5)	N=216 7 (3.2)	N=487 21 (4.3)

[*] $p < 0.05$,

[**] $p < 0.001$; Chi-square tests were used.

Table 3

Posterior Results from Bayesian Joint Hierarchical Models

	Joint Hierarchical Models		
	Model 1: All Subjects (n=1793)	Model 2: Male (n=1306)	Model 3: Female (n=487)
Weibull PH with Frailty Model	Posterior Median Hazard Ratio (HR) (95% Credible Interval)		
Demographicsz			
Gender (Ref: Female)	0.99 (0.60, 1.67)	--	--
Age at Enrollment	0.99 (0.98, 1.01)	1.00 (0.98, 1.01)	--
Age at First Use	1.10 (1.07, 1.12)[***]	1.09 (1.06, 1.11)[***]	1.09 (1.03, 1.17)[***]
Ethnicity (Ref: Hispanic & Others)			
White	1.09 (0.85, 1.37)	1.11 (0.86, 1.40)	1.02 (0.35, 3.26)
Black	0.60 (0.41, 0.87)[*]	0.62 (0.43, 0.90)[*]	0.70 (0.19, 9.23)
Primary Drug Group			
Cocaine to Heroin	0.46 (0.24, 0.79)[*]	0.63 (0.27, 1.23)	0.12 (0.03, 0.44)[**]
Meth to Heroin	0.28 (0.13, 0.62)[***]	0.14 (0.03, 0.49)[*]	0.40 (0.11, 1.46)
Meth to Cocaine	0.62 (0.27, 1.57)	0.22 (0.05, 0.90)[*]	3.37 (0.68, 17.5)
%Usage			
Substance	2.69 (1.80, 4.16)[***]	2.79 (1.82, 4.32)[***]	1.17 (0.17, 9.23)
Alcohol	1.12 (0.81, 1.57)	1.03 (0.74, 1.47)	3.42 (1.06, 13.2)[*]
Kendall's τ	0.07 (0.003, 0.34)	0.08 (0.002, 0.71)	0.17 (0.008, 0.63)
Mixed-Effects Models	Posterior Median (95% Credible Interval)		
Overall % Substance Usage			
Cocaine – Heroin	-6.70 (-11.2, -2.10)[**]	-7.30 (-14.2, -0.60)[*]	-5.70 (-11.6, 0.20)
Methamphetamine – Heroin	-7.60 (-12.8, -2.00)[*]	-8.70 (-16.6, -0.10)[*]	-6.60 (-14.1, 0.70)
Intra-Class Correlation (ICC)	0.40 (0.21, 0.72)	0.39 (0.19, 0.72)	0.57 (0.29, 0.87)
Overall % Alcohol Usage			
Cocaine – Heroin	4.00 (-2.20, 10.2)	3.10 (-5.20, 11.5)	7.20 (-2.30, 17.3)
Methamphetamine – Heroin	5.50 (-1.90, 12.6)	7.80 (-2.20, 18.0)	2.90 (-8.70, 15.6)
Intra-Class Correlation (ICC)	0.33 (0.15, 0.66)	0.34 (0.16, 0.68)	0.36 (0.14, 0.81)

[*] $P < 0.05$,

[**] $P < 0.001$,

[***] $P < 0.0001$.

Note: In the survival models, P represents the posterior probability of a value less than one (or > 1) for the hazard ratio. In the mixed-effects models, P represents the posterior probability of a value less than zero (or > 0) for the regression coefficient. The posterior probability can be used in place of the frequentist's p-value.