Advance Access publication June 30, 2007

Political Analysis (2007) 15:237–256
doi:10.1093/pan/mpm013

Event Dependence and Heterogeneity in Duration Models: The Conditional Frailty Model

Janet M. Box-Steffensmeier
Department of Political Science, 2140 Derby Hall, 154 North Oval Mall, The Ohio State University, Columbus, OH 43210
e-mail: jboxstef+@osu.edu (corresponding author)

Suzanna De Boef and Kyle A. Joyce
Department of Political Science, 219 Pond Laboratory, The Pennsylvania State University, University Park, PA 16802
e-mail: sdeboef@psu.edu
e-mail: kjoyce@psu.edu

We introduce the conditional frailty model, an event history model that separates and accounts for both event dependence and heterogeneity in repeated events processes. Event dependence and heterogeneity create within-subject correlation in event times thereby violating the assumptions of standard event history models. Simulations show the advantage of the conditional frailty model. Specifically they demonstrate the model's ability to disentangle the sources of within-subject correlation as well as the gains in both efficiency and bias of the model when compared to the widely used alternatives, which often produce conflicting conclusions. Two substantive political science problems illustrate the usefulness and interpretation of the model: state policy adoption and terrorist attacks.

1 Introduction

Political scientists are often interested in research questions that involve the time until a subject experiences a repeatable event. We wish, for example, to understand the dynamics of repeated international conflicts, executive branch appointments, cabinet transitions, and terrorist attacks. The analysis of repeatable events is complicated by the unique features of these processes. First, political actors learn from history: countries and their leaders have experiences that teach them the utility of policies, programs, and actions. Prior events set actors down a path where options become more or less limited or certain choices become more or less easy to implement. Second, subjects are heterogeneous: actors learn more or less well, institutions adopt distinct rules, leaders take differential risks, and cultures diverge. Both these features of repeated events processes create dependencies in the data: if and when dyads fight, appointments are successfully made, cabinets dissolve, and terrorists attack depend on past experience and the underlying tendency of subjects to act.

Authors' note: Three anonymous reviewers gave valuable advice. Replication materials and an online appendix are available on the *Political Analysis* Web site. Any errors are our own responsibility.

© The Author 2007. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

Technically, these twin features of repeated events processes are known as subject event dependence and heterogeneity. In the first case, when a subject experiences multiple failures, the timing of these failures is likely to be correlated within subject, that is, whether a subject experiences a second event is influenced by having experienced a first event so that risks change with the number of events experienced. In the case of heterogeneity, some subjects are more prone to experience multiple events but we are unable to specify why. This produces within-subject correlation in the timing and occurrence of events as well. Correlated event times violate the independence assumption of traditional event history models. As a result, in most cases of repeated events, standard parametric and semiparametric models will provide an incomplete and biased picture of the dynamics.

In this article, we introduce the conditional frailty model for repeated events processes. The conditional frailty model captures within-subject correlation due to both event dependence and heterogeneity. Although event history models have previously been developed for the analysis of repeated events, simulation evidence indicates that the conditional frailty model is superior to alternative models when repeated event processes are characterized by both event dependence and heterogeneity. Importantly, analysts are typically unaware whether either or both event dependence and heterogeneity produce correlated event times. The conditional frailty model performs well regardless of the source of correlation making this model of great potential use to political scientists.

We begin with a review of commonly used models for repeated events. Next, we introduce the conditional frailty model and summarize the simulation evidence found in Box-Steffensmeier and De Boef (2006) that compares these models. We illustrate the advantage of the conditional frailty model using two political science applications. We conclude with some suggestions to applied researchers for analyzing repeated events.

2 Repeated Events Models

One of the most widely used event history models is the Cox proportional hazards model (Cox 1972). The Cox model assumes that event times, conditional on the covariates included in the model, are independent. Any correlation among event times, produced either by event dependence or heterogeneity or jointly by event dependence and heterogeneity, violates this assumption. As a result, the Cox model produces estimates that are both biased and inefficient in most repeated events contexts (Aalen 1988; Lawless and Nadeau 1995; Kelly and Lim 2000; Nelson 2003). Two variations of the Cox model have been proposed to account for the correlation in event times caused by repeated events: variance-corrected and frailty (or random effects) models.[1] Repeated events models differ with respect to three characteristics.

1. Risk set: Variance-corrected models vary in "the way that the risk sets are defined at each failure" (Cleves 1999, 34). The risk set defines how subjects are considered to be at risk for a given event k at a particular time t. When defining the risk set, it is important to know whether the repeated events follow a sequential order or whether they can occur simultaneously. If a subject cannot experience a subsequent event unless it has experienced a prior event, then it is important to define a conditional risk set to preserve the ordering of events. In contrast, if a subject can simultaneously experience any of the k events at any time point, then the risk set must be defined unconditionally and a marginal model estimated.

[1]For a detailed review of variance-corrected models, see Box-Steffensmeier and Zorn (2002).

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

2. Clock time: Variance-corrected models differ in how time at risk for an event is counted, that is, does the clock begin when a subject enters the observation period and end when a subject is censored or does the clock restart at zero after each event. In elapsed time, subjects are at risk since the beginning of the observation period so that models estimated in elapsed time examine the effect of a covariate on the hazard of an event since the beginning of the observation period. In models estimated in gap (or interevent) time, the time at risk is renewed after each event so that models estimated in gap time examine the effect of a covariate on the hazard of an event since the previous event. For example, in elapsed time a subject with failure events in months 2, 8, and 12 would have start and stop times of 0-2, 0-8, and 0-12. In gap time, the subject would have start and stop times of 0-2, 0-6, and 0-4.
3. Stratification: Variance-corrected models differ in whether or not they allow for event-specific baseline hazards. If the hazard rate is likely to be different for the first failure compared to the third failure—as when event dependence exists—then the ability to stratify the data to allow for the estimation of a separate baseline hazard for each event is important.[2] In contrast, if the risk associated with all events is the same, a single baseline hazard will fit the data and stratification is unnecessary.

2.1 *Variance-Corrected Models*

Variance-corrected models account for correlated event times due to unobserved, or unaccounted for, heterogeneity by adjusting the variance-covariance matrix to account for the subject-specific effects that are explicitly specified in the model. If event times are correlated due to heterogeneity, the covariate effects estimated by the Cox model are biased but still consistent whereas the estimate of the variance-covariance matrix is both biased and inconsistent.

Depending on the definition of the risk set, the clock time, and whether the model stratifies, we estimate very different dynamic processes. Thus, each of the variance-corrected models leads to different empirical findings so that care must be taken when selecting a model (Gail, Santner, and Brown 1980; Prentice, Williams, and Peterson 1981; Andersen and Gill 1982; Wei, Lin, and Weissfeld 1989; Clayton 1994; Lin 1994; Gao and Zhow 1997; Klein and Moeschberger 1997; Cook et al. 1999; Nelson 2003).[3] The Andersen-Gill (AG) model (Andersen and Gill 1982), the conditional risk-set models (Prentice, Williams, and Peterson 1981), and the marginal risk-set model (Wei, Lin, and Weissfeld 1989) are the most popular variance-corrected models.[4] A conditional risk-set model estimated in gap time with stratification seems the most natural variance-corrected model for repeated events political processes as it explicitly orders sequential events, subjects are not at risk for a later event until they have experienced all previous events, and baseline hazards are allowed to vary by the number of events experienced.

The primary characteristic of the AG model is the "assumption that the risk of an event for a given subject is unaffected by any earlier events that occurred to the same subject, unless terms that capture such dependence are included explicitly in the model as covariates"

[2] Although a separate baseline hazard is estimated for each stratum the covariate effects are assumed to be constant across the strata.

[3] See Kelly and Lim (2000) for the hazards and likelihoods for each of the variance-corrected models.

[4] Previous research has largely discredited the marginal risk-set model for the study of repeated events (Cook and Lawless 1997; Oakes 1997; Kelly and Lim 2000). Importantly, many political processes occur sequentially and have a natural order thus rendering this model inappropriate because it does not preserve the sequence of events. We do not consider it further.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

(Oakes 1992, 374). Thus, the AG model specifies an unconditional risk set in elapsed time and without stratification. Practically speaking, the Cox model and the AG model are identical, except that the AG model uses robust standard errors or clustering within units. The AG model assumption that event times are conditionally independent is an especially strong assumption and one that is likely to be violated in most repeated event contexts, particularly if the ordering of events is important.

Conditional risk-set models are characterized by the fact that an observation is not at risk for a subsequent event until it has experienced all previous events, that is, failure times are conditional on experiencing a prior event. The conditional risk-set model may be estimated in either (1) elapsed time, where estimates are provided for the effect of a covariate on the hazard of the kth event since the beginning of the observation period or (2) gap time, where effects are estimated for the effect on the kth event since the occurrence of the $(k-1)$th event. The conditional risk-set models stratify by event rank so that different events have their own baseline hazard. This controls for within-subject correlation due to event dependence and produces consistent estimates of the effects of the included covariates.

Variance-corrected models can be used to account for correlation in event times that is due to heterogeneity; the inefficiency caused by correlated events due to heterogeneity is accounted for by adjusting the variance-covariance matrix. A subset of these models—specifically those that stratify by event, allowing the baseline hazards to vary with the number of events a subject experiences—also accounts for correlation in event times due to event dependence. Yet in all cases, because the variance-corrected models do not incorporate the heterogeneity directly at the estimation stage, the estimates remain biased. Simulation research examining the estimated covariate effects for a subset of variance-corrected models, assuming normally and uniformly distributed random effects, has found that heterogeneity produces negative biases in the estimated effects such that the covariate effects are underestimated (Kelly and Lim 2000; Therneau and Grambsch 2000). Therneau and Grambsch (2000) suggest that the magnitude of the bias may be tolerable in some circumstances, but such a conclusion is unsatisfying. Additional work comparing variance-corrected models in the bivariate case with both event dependence and heterogeneity shows significant bias in estimated effects, poor size and power of hypothesis tests, and large model mean-squared errors (Bowman 1996).

2.2 *Frailty Models*

In contrast to variance-corrected models, frailty or random effects models treat repeated events as a special case of more general unit-level heterogeneity. The logic of the frailty model is that some subjects may be inherently more or less prone to experiencing repeated events than other subjects. The frailties are unobserved effects across subjects and are constant over time.

Frailty models incorporate heterogeneity into the estimator by treating individual effects as random draws from a specific parametric distribution. These random draws are then estimated along with the structural parameters of the model to explicitly model the correlation between failure times (Hougaard 1991; Oakes 1992; Keiding, Andersen, and Klein 1997; Sastry 1997).[5] Event times in a frailty model are assumed to be independent conditional on the chosen parametric distribution, and so inference is made as for the Cox or AG models. The standard frailty model specifies an unconditional risk set in

[5]See Hougaard (2000) for the hazard and likelihood of the basic frailty model.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

elapsed time. However, like the variance-corrected models, frailty models can be estimated in elapsed time or gap time and with conditionally or unconditionally defined risk sets.

Frailty models have been criticized on two points. First, neither theory nor data typically provides much guidance for choosing a specific distribution from which to draw the frailty term, and the "parameter estimates can be highly sensitive to the assumed parametric form of the error term" (Blossfeld and Rohwer 1995, 255). Current research is investigating how to best choose a distribution for the random effects and the potential downsides of misspecifying the distribution (Kosorok, Lee, and Fine 2004; Gao, Manatunga, and Chen 2007), while other researchers are focusing on how to relax the parametric assumption (Andersen, Klein, and Zhang 1999; Li and Lin 2003; Ohman-Strickland 2003). Second, the frailty terms are generally required to be independent of the model's covariates, and failure to meet this assumption can produce estimates that are badly biased and inconsistent (Hausman 1978).

The baseline hazard rate for the standard frailty model, unlike a subset of the variance-corrected models, does not vary by event number, k, and thus does not control for event dependence. However, unlike the variance-corrected models, frailty models directly incorporate heterogeneity via a random effect. Simulations examining the covariate effects in frailty models with gamma-distributed random effects find that frailty models produce unbiased estimates of covariate effects when the variance of the random effect is known (Therneau and Grambsch 2000). However, these simulations do not consider data-generating processes (DGPs) that, in addition to exhibiting heterogeneity, also contain event dependence. Additionally, simulations designed to examine DGPs with both event dependence and heterogeneity have been limited to two events and do not include frailty models (Bowman 1996).

2.3 *Concerns with Both Classes of Estimators*

In conclusion, both the variance-corrected and frailty classes of estimators acknowledge the problems associated with violating the Cox model's assumption of independent event times. However, neither the variance-corrected nor frailty models are consistent with typical repeated events processes that may exhibit both event dependence and heterogeneity. Additionally, these models do not provide a dependable solution when researchers are unsure which source of correlation is producing the correlated event times. Existing research does not examine how well existing models account for the two different sources of correlation. Without a general modeling strategy for repeated events processes that contain event dependence and heterogeneity, we cannot accurately determine the influence of the conditions that influence whether a subject will experience multiple events either because the act of experiencing the event itself influenced whether the subject would experience the event again (i.e., event dependence) or because of some unknown or unmeasured factor (i.e., heterogeneity). Box-Steffensmeier and De Boef (2006) present the conditional frailty model, a modeling solution that performs well regardless of whether the DGP features event dependence and/or heterogeneity. In the next section, we examine this model.

3 The Conditional Frailty Model

The conditional frailty model controls for both event dependence through event-based stratification and unobserved heterogeneity via a random effect. The model is formulated in gap time so that parameter estimates are interpreted as an estimate of the risk for the

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

kth event since the $(k-1)$th event (for right censored failures only).[6] The hazard of a particular event k occurring to a specific subject i (λ_{ik}) for the conditional frailty gap time model follows from Therneau and Grambsch (2000) and Kelly and Lim (2000):

$$\lambda_{ik}(t) = \lambda_{0k}(t - t_{k-1})e^{X_{ik}\beta+\omega_i}, \tag{1}$$

where k denotes event number, λ_{0k} is the baseline hazard rate and varies by event number, $(t - t_{k-1})$ incorporates the gap time structure so that the hazard gives the risk for event k since the $(k-1)$th event, X is a vector of independent variables (which may be time varying), and β gives the effect parameters. The remaining portion of the hazard incorporates the random effect. Each subject i has a random effect that is shared across events (i.e., constant over time) and ω is a vector containing the unknown random effects or frailties (Hougaard 2000).

The partial likelihood for this model, conditional on the frailties, is given by

$$L(\beta) = \prod_{i=1}^{n}\prod_{k=1}^{K}\left(\frac{e^{X_{ik}\beta+\omega_i}}{\sum_{i=1}^{n}\sum_{k=1}^{K}Y_{ik}e^{X_{ik}\beta+\omega_i}}\right)^{\delta_{ik}}, \tag{2}$$

where k refers to the event number, δ is a censoring variable equal to 1 if observed and 0 if censored, and Y is an at-risk indicator equal to 1 when the subject is at risk for the current event k and 0 otherwise.

The conditional frailty model allows for the possibility that both event dependence and heterogeneity make important contributions to the hazard rate of a subject's risk for the recurrence of a particular event. Box-Steffensmeier and De Boef (2006) compare the conditional frailty model to other commonly used repeated event models. Their evidence using clinical trials data suggests that the model is particularly promising for addressing both event dependence and heterogeneity because it allows for both event-specific baseline hazards and a random effect.

4 Simulations of Four Conditions

Box-Steffensmeier and De Boef (2006) conducted simulations that compare the conditional frailty model with the four most commonly used variance-corrected models: the AG model, the conditional risk-set model estimated in elapsed time, the conditional risk-set model estimated in gap time, and the standard frailty model with a gamma-distributed random effect. These simulations consistently show the superiority of the conditional frailty model. These models were estimated under four conditions: when the DGP exhibited *either* heterogeneity or event dependence, *both* heterogeneity and event dependence, and *neither* heterogeneity nor event dependence.[7]

We summarize the simulation results in Fig. 1, which presents the estimated densities for the effect of a negative treatment (covariate) effect and sample size = 100. The vertical line represents the true parameter value (here -1.0), whereas the spread and location of the densities illustrate the bias and relative efficiency associated with each model. For details

[6] We refer to the conditional frailty gap time model simply as the conditional frailty model for ease of presentation, but all references to the conditional frailty model refer to the gap time formulation. A conditional frailty elapsed time model is possible as well; however, in most political science applications, researchers are interested in risk since the last event rather than from the initial observation.

[7] The simulations in Box-Steffensmeier and De Boef (2006) extended previous repeated events simulations by examining DGPs with event dependence and both event dependence and heterogeneity. See Therneau and Grambsch (2000) and Kelly and Lim (2000) for simulations focused on heterogeneity.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

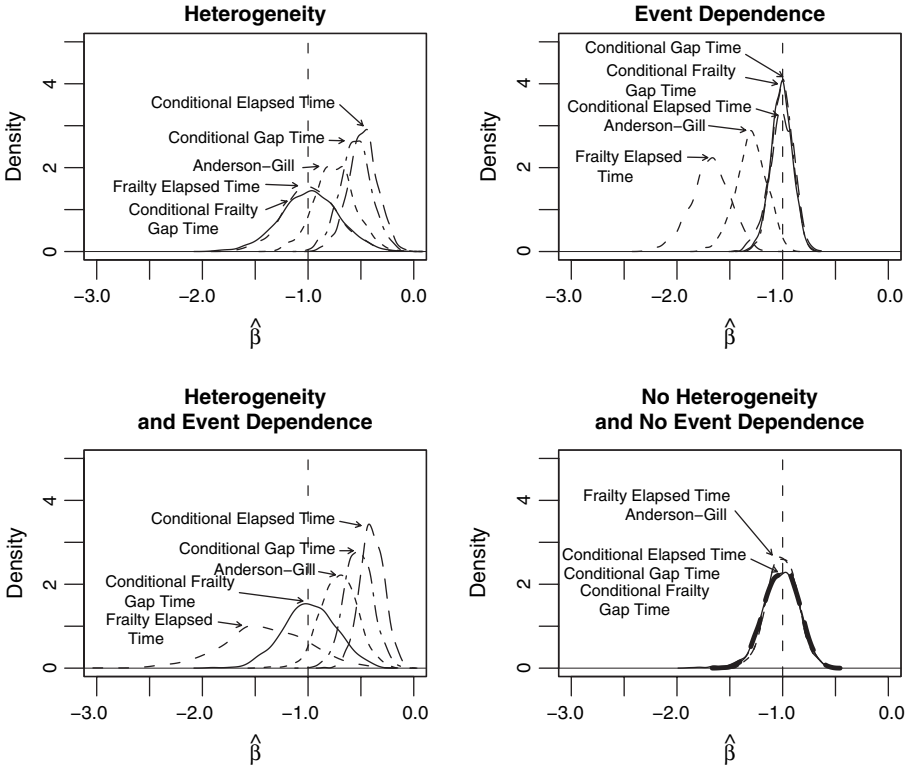


Fig. 1 Densities of the estimated β ($N = 100$) under four conditions. The densities of the estimated effects, $\hat{\beta}$, are calculated for conditions under which the $\lambda_0 = 1.0$, $\theta = 1.0$, $\beta = -1$, $\lambda_{0k} = k\lambda_0$, the maximum follow up time is 4. The maximum number of events, which is determined by the baseline hazard and the follow-up time, is 6.

on the estimated variance of the random effect and the bias in standard errors, see Table 1 in Box-Steffensmeier and De Boef (2006).

The conditional frailty model performed at least as well as, or better than, the variance-corrected and frailty alternatives under all conditions. The conditional frailty model simultaneously estimated the effects of both heterogeneity and event dependence and recovered the parameters of the true DGP better under all four DGPs. Importantly, in the presence of both heterogeneity and event dependence, only the conditional frailty model performed well. When there is a possibility that both heterogeneity and event dependence are present—and often we cannot rule out either—the conditional frailty model is the best choice.

4.1 *Heterogeneity*

Heterogeneity refers to the condition wherein some subjects are more prone to experiencing a larger number of events than other subjects for some unknown, unmeasured, or unmeasurable reason. The estimated effects under heterogeneity revealed five general conclusions (see Fig. 1, panel 1). First, in the presence of heterogeneity, correcting the standard errors (as in the variance-corrected models) does not correct the bias that results from the imbalance in data. The means of the empirical densities of the variance-corrected models are all closer to 0 than the conditional frailty and the frailty model. Second,

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

Table 1 Repeated events event history models of U.S. state (re)adoption of obscenity legislation, 1991–1998

	Conditional gap time	*Frailty (gamma; EM)*	*Conditional frailty (gamma; EM)*
Fundamentalist	0.15** (0.07)	0.13 (0.10)	0.15* (0.08)
South	0.08 (0.30)	0.17 (0.33)	0.08 (0.30)
College educated	0.09** (0.04)	0.09** (0.04)	0.09** (0.04)
Legislative ideology	−0.02** (0.01)	−0.02*** (0.01)	−0.02** (0.01)
Citizen ideology	0.02** (0.01)	0.03*** (0.01)	0.02** (0.01)
Women legislature	0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
Murder rate	0.10*** (0.03)	0.12*** (0.04)	0.10*** (0.03)
θ		0.17 (0,0.64)	0.00 (0,0.29)
N	400	400	400
Number of failures	94	94	94
Likelihood ratio $\chi^2_{(.,1,1)}$ for θ		1.26	0.00
I-likelihood		−340.95	−261.40
Log-likelihood for model	−261.40	−329.18	−261.40
Wald $\chi^2_{(7,14.9,7)}$	29.5***	21.8	17.1**

Note. Significance (two tailed): *.10; **.05; ***.01.

estimating a random effect to account for heterogeneity reduces the bias in the estimated β, substantially improving the performance of the model; the mean of the estimated effects for both the conditional frailty and frailty model are very near the true value of −1.0. Third, unless a random effect is also estimated, stratifying the data by event is problematic: The two conditional risk-set models exhibited the largest downward biases because the stratification attempts to account for any heterogeneity. Since the conditional frailty model already incorporates the heterogeneity, adding the stratification to this model does not induce bias. Fourth, the robust standard errors, which improve as *N* increases, are smaller than the standard deviations of the estimated β in all the models, though the standard deviations are larger in the frailty and conditional frailty models because more parameters are estimated. Fifth and finally, the statistically significant estimate of the frailty term was more accurate in the conditional frailty model than in the standard frailty model.

4.2 *Event Dependence*

Event dependence refers to a condition where the probability that a subject experiences subsequent events is related to whether the subject has experienced a prior event. In this case, the true DGP has baseline hazard rates that vary for each event.

The true DGP examined in the simulations has baseline hazard rates that vary as a positive function of event number such that risk increases with higher strata. Thus, all the models that stratify by event number should outperform the models that do not stratify. By omitting strata, the AG and standard frailty model essentially average the effects across all events. Generally, these averages are biased, and this bias increases the larger the separation that exists among the hazard rates across the number of events. The direction of bias is predictable: if a treatment reduces event rates, as here $\beta < 0$, event dependence means that the control group will have more subjects with a larger number of events relative to the treatment group. The result is that the treatment, which is designed to reduce event rates, looks more effective than it actually is. This can be seen in panel 2 of Fig. 1, where most of the empirical density mass is well to the left of −1.0 for the two models that

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

do not stratify: AG and frailty models. The same is also true for a treatment with a positive effect. More cases in the control group have more events than in the treatment group, making the positive effect look even bigger. Because more events occur in higher strata, the effects look larger unless we correctly incorporate the shift in the baseline hazard. Because the conditional frailty model and conditional gap and elapsed time models stratify, this bias does not occur (see Fig. 1, panel 2).

In addition to the effect on bias, the results of the simulations for the DGP with event dependence show that the conditional frailty model (e.g., frailty with stratification) accurately recovers the true variance of the random effect ($\theta = 0$) whereas the standard frailty model (e.g., frailty without stratification) does not because it picks up the correlation in event times. Thus, in the presence of event dependence, stratification is essential. Importantly, the addition of the random effect in the conditional frailty model does not come at any additional cost. The conditional frailty model correctly estimates the variance of the random effect and produces a relatively unbiased estimate of the treatment effect.[8]

4.3 *Event Dependence and Heterogeneity*

The presence of both event dependence and heterogeneity is of primary concern for two reasons. First, it is likely that both sources of correlation may simultaneously describe the DGP in which analysts are often interested. Second, even if only one source of correlation exists, we do not typically know a priori which source produces the correlated event times and therefore we may not know which model to use (Hougaard 2000; Fong et al. 2001; Cook and Lawless 2002; Lawless 2002).

The simulations suggest that two things occur in the presence of both (positive) event dependence and heterogeneity, which together work to adversely affect the estimates from all but the conditional frailty model. First, event dependence exacerbates the problems with the conditional risk-set models that occur due to heterogeneity. The conditional risk-set models do not do well with heterogeneity; the downward bias in the estimated density of the conditional gap and elapsed time models is evident. Second, models that do not stratify by event number average the varying baseline hazards, which, in general, produces biased estimates. For the frailty model, the bias is away from zero, whereas for the AG model, the bias is toward zero.

The conditional frailty model performs the best in the presence of both heterogeneity and event dependence, and the evidence shows that the correlation produced by both effects can be simultaneously and distinctly well controlled. The conditional frailty model controls for event dependence by stratifying by event and controls for heterogeneity by allowing for the estimation of a random effect. Importantly, the conditional frailty model exhibits very little bias in either the estimated β (see Fig. 1, panel 3) or θ, the variance of the random effect.

4.4 *No Event Dependence or Unobserved Heterogeneity*

The DGP without heterogeneity and event dependence is consistent with the Cox model's assumptions. In this case, all the estimated models should be unbiased and the standard errors should be close to the standard deviation of the estimated densities for $\hat{\beta}$. Estimators that allow for varying baseline hazards or that include a random effect should be less efficient than estimators that do not, given that these models are estimating an (some) additional parameter(s). Thus, the conditional frailty model, by allowing for both event

[8]These results caution against the current endorsement by Therneau and Grambsch (2000) for the AG model.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

dependence and heterogeneity, should be the least efficient of the models. The simulations largely conform to these expectations. Importantly, the only cost associated with using the conditional frailty model when there is no heterogeneity or event dependence is a slight loss of efficiency (see Fig. 1, panel 4).

4.5 *Summary*

The conditional frailty model performed well in these simulations, but it is worth asking under what general conditions will the model provide good estimates? In addition to the simulation results reported here, Box-Steffensmeier and De Boef (2006) considered the case of rare events and a variety of covariate effects. In each case, the findings support the conclusions we have drawn above. These simulations, along with our general understanding of the model, lead us to make a general point that estimation of both heterogeneity and event dependence requires variation across subjects in the timing of events and the number of events experienced. In order to distinguish between the case where a subject is simply more frail and therefore prone to experience the event and the case where a subject's risk for events increases due to having experienced a number of events in the past, there needs to be some subjects that experience multiple events and that vary in the timing of those events. Information (variation) on the number of events experienced informs us about subjects' frailty. Similarly, as long as we have multiple events experienced by some subjects and the timing of the repeated events vary, we can distinguish event dependence; information (variation) on the timing of multiple events informs us about event dependence. This means that in the ideal case, the distribution of events in the data is such that large numbers of subjects experience each number of events in the data and vary in the timing of those events. In the extreme case where only one subject experiences an event, the estimator will have trouble converging as the two sources of within-subject correlation cannot be parsed from one observation. When the number of cases in a particular stratum is relatively small, higher strata should be collapsed. Where few subjects experience a given number of events, the baseline hazard associated with each event number may be poorly estimated, but our evidence suggests that covariate effects will continue to be well estimated. In short, when there is a possibility that both heterogeneity and event dependence are present—and often we cannot rule out either—the conditional frailty model should be used.

5 Applications

We present two applications of the conditional frailty model. In the first application, we examine U.S. state (re)adoption of obscenity legislation. The second application considers the (re)occurrence of transnational terrorist incidents.[9]

5.1 *State Policy Adoption As Repeated Events*

States frequently adopt policies in a given policy domain multiple times. Typically, scholars ignore the repeated nature of these events, focusing instead on the time until a state adopts a single policy (see for example Berry and Berry 1990; Mooney and Lee 1995; Mintrom and Vergari 1998). An important exception to this norm, Jones and Branton (2005) take into account the repeated nature of state adoption of obscenity legislation, estimating a conditional gap time model.

[9]The statistical analysis was conducted using R 2.4.1 (R Development Core Team 2007).

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

The data consist of 400 observations for 50 states between 1991 and 1998.[10] The unit of analysis is a state-year (i.e., legislative session). Each observation records whether the state adopted the legislation or not in a particular year. A state could adopt several different kinds of obscenity legislation, such as regulation of the adult entertainment industry or pornography, multiple times. An event (i.e., failure) occurs when a state adopts some type of obscenity legislation. There are a total of 94 events with each state experiencing an average of $94/50 \approx 2$ events. The event strata in this data set range from 1 to 6, with 13 states passing no pieces of obscenity legislation during the 8 years in the data set, 11 states passing exactly 1 piece, 8 states passing exactly 2, 8 states passing exactly 3, 7 states passing exactly 4, and 3 states passing exactly 5.

Jones and Branton (2005) examine the effect of seven variables on the timing of state adoption of obscenity legislation. *Fundamentalist* is the percentage of the state's population who claim to be fundamentalist protestants. *South* is a binary variable indicating whether a state is in the South. *College educated* is the percentage of college-educated adults in the state. *Legislative ideology* is a state's legislative ideology. *Citizen ideology* is a state's citizen ideology. *Women legislature* is the percentage of women in a state's legislature. *Murder rate* is a state's murder rate.

We build on Jones and Branton (2005), replicating their original results and comparing them to those from the standard frailty and conditional frailty models. The results are found in Table 1.

Estimating a conditional model in gap time (see Table 1, column one), Jones and Branton (2005) find that the percentage of a state's population claiming to be fundamentalist protestants, the percentage of citizens who are college educated, the degree of conservatism of the state's citizens, and the state's murder rate each significantly increases the risk for adoption of obscenity legislation. In contrast, they find that the degree of conservatism of the state's legislature significantly reduces the risk of legislation. Neither Southern state nor the percentage of female legislators exerts a significant effect on the likelihood of adopting obscenity legislation.[11] The conditional gap time model stratifies on the risk sequence allowing the baseline hazard functions to vary across the risks by event number. The cumulative hazards provide evidence for the presence of event dependence; the cumulative hazards for each stratum are distinct.[12]

The second column in Table 1 presents the results from estimating a frailty model. The frailty model does not estimate a statistically significant effect for the percentage of self-identified fundamentalist protestants. The remaining effects identified by Jones and Branton (2005) continue to be statistically significant, but the magnitude of the effects changes. The estimated effects are generally larger using the frailty model as we would expect. The frailty model includes a frailty term, which allows us to determine whether there are any unaccounted for risk factors that increase the hazard for more failure-prone states.[13] The frailty estimates by state are shown in Fig. 2. The estimated frailty equals

[10]This data set was obtained from Brad Jones' Web site and was used in Jones and Branton (2005).

[11]Our results for the conditional gap time model differ from Jones and Branton (2005) because of an error in their gap time data set up. However, this mattered very little in the substantive results; the statistical significance and direction of the estimated coefficients are the same, though the size of the estimated coefficients differs.

[12]The cumulative hazards for the conditional gap time and conditional frailty models are graphically indistinguishable, and so we present the cumulative hazards for the conditional frailty model. The cumulative hazards for the conditional gap time model can be found in the supplementary materials available on the *Political Analysis* Web site.

[13]The cumulative hazard for the frailty model can be found in the supplementary materials available on the *Political Analysis* Web site.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

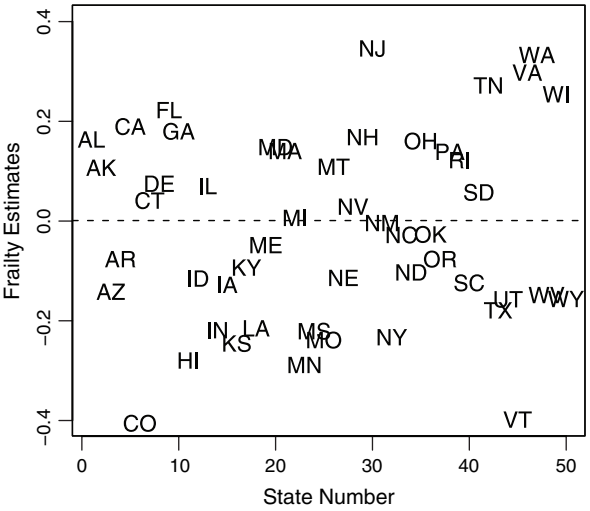


Fig. 2 Estimated frailties by state for the frailty model of U.S. state (re)adoption of obscenity legislation.

0.17 and is not statistically significant, indicating that unobserved heterogeneity is not affecting our model.[14]

The third column in Table 1 presents the results for the conditional frailty model. The estimated effects from the conditional frailty model are very similar to the findings of Jones and Branton (2005).[15] The estimated frailty is smaller (0.00) and still statistically insignificant.[16] The frailty estimates by state are shown in Fig. 3. The range of the frailty estimates is quite small suggesting that the underlying DGP is characterized primarily by event dependence. Figure 4, which plots the cumulative hazards by event number, shows that the cumulative hazards vary by event number. Risks for the adoption of further legislation increase over time, that is, the risk of having a second event increases after experiencing the first event and the risk of having a third event increases after a state has a second event. After a state has a third event, the risk of experiencing a fourth and fifth

[14]There are two test statistics that can be used to evaluate $\hat{\theta}$, an approximate Wald test and a likelihood ratio test. Therneau and Grambsch (2000) state that the "likelihood ratio test is preferred for formal assessment of θ" (236). For this model, a likelihood ratio test for $\hat{\theta}$ is twice the difference between the log likelihood of a model without a frailty term (−341.58) and the log partial likelihood with the frailty terms integrated out (−340.95), 2 (341.58 − 340.95) = 1.26, which is not statistically significant based on a χ^2 test with 1 degree of freedom. Additionally, we can calculate the uncertainty around the estimated frailty parameter by fixing θ, then fitting the model to the data maximizing over the β's, and plotting the profile likelihood (Therneau and Grambsch 2000, 236–7). We fit the frailty model to the data 100 separate times, varying θ from 0.01 to 1, and then plotted the profile likelihood. The 95% confidence interval for $\hat{\theta}$ for this model is (0,0.64). A plot of the profile likelihood for the frailty model can be found in the supplementary materials available on the *Political Analysis* Web site. As a result of the asymmetry in the profile likelihood, Therneau and Grambsch (2000, 237) state that "$\hat{\theta} \pm 1.96\text{se}(\hat{\theta})$will not be an accurate 95% confidence interval."

[15]An unsettled question in the literature is which computational algorithm to use for estimation of the model with random effects (see Hurvich, Simonoff, and Tsai 1998, Therneau 2000, and Therneau and Grambsch 2000): the EM for the gamma distribution, REML for Gaussian, or the corrected Akaike's information criterion for either. We use the EM.

[16]The likelihood ratio test for the frailty for this model is 2 (261.40 − 261.40) = 0, which is not statistically significant based on a χ^2 test with 1 degree of freedom. The 95% confidence interval for $\hat{\theta}$, calculated using the method mentioned above, was (0,0.29). A plot of the profile likelihood can be found in the supplementary materials available on the *Political Analysis* Web site.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

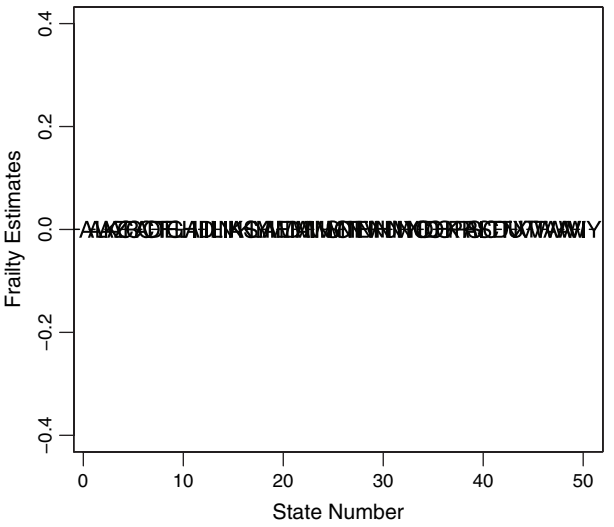


Fig. 3 Estimated frailties by state for the conditional frailty model of U.S. state (re)adoption of obscenity legislation.

event decreases. The main source of within-subject correlation in the state (re)adoption of obscenity legislation is event dependence.

In sum, although the inferences drawn from the conditional gap time and the conditional frailty model are very similar, we learn something important by estimating the conditional frailty model. The conditional frailty model arbitrates between the frailty and conditional gap time models. This example illustrates how a frailty term will attempt to soak up event dependence and masquerade as heterogeneity when stratification is not allowed. If we compare the state-level frailty estimates from the frailty model shown in Fig. 2 to those from the conditional frailty model shown in Fig. 3, we can see that the inferences drawn from the frailty model would indicate that heterogeneity is present when in fact the process is driven by event dependence, which is what the conditional frailty model shows. This is why the estimated variance of the frailty term is larger in the frailty model than the conditional frailty model. Here we can say that the main source of within-subject correlation is event dependence.

5.2 *Terrorist Incidents As Repeated Events*

Terrorist attacks have long been of interest to International Relations scholars and are of increased public interest in the United States since 9-11. The reality of the repeated nature of terrorist attacks needs to be accounted for in our modeling. However, the focus in the scholarly literature is often on the causes of terrorist attacks while the timing and repeated nature of such attacks are ignored.

The data consist of 2576 observations for 112 countries between 1975 and 1997.[17] The unit of analysis is the country-year. Each observation records whether a country

[17]This data set is from Li and Schaub (2004) and was obtained from the *Journal of Conflict Resolution* replication archive. The estimation sample is smaller than the size of the data set because of missing values on several covariates.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

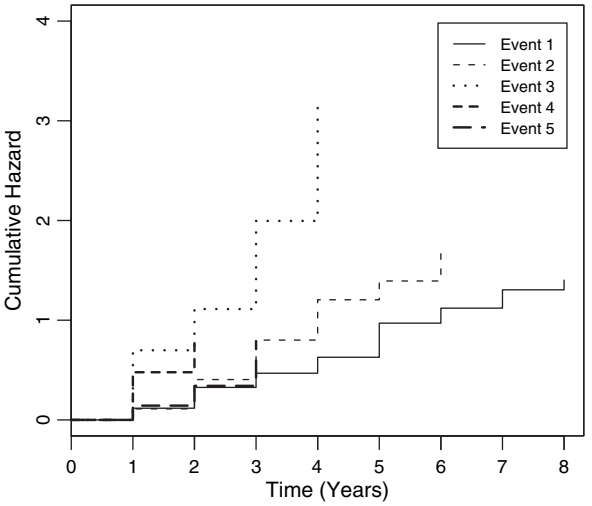


Fig. 4 Estimated cumulative hazards for each event number for the conditional frailty model of U.S. state (re)adoption of obscenity legislation.

experiences a transnational terrorist incident or not in a particular year.[18] The data focus on the country in which the incident began and does not distinguish between the types of attack. An event occurs when a country experiences an attack. There are a total of 1257 terrorist events with each country experiencing an average of $1257/112 \approx 11$ attacks. The event strata in this data set range from 1 to 23. However, when strata have a small number of events, it is generally recommended that strata be collapsed. Convergence problems can occur when there are a large number of strata and/or a small number of events per stratum. We collapsed the event strata such that strata greater than 11 were set equal to 11.

Li and Schaub (2004) are interested in the relationship between economic globalization and transnational terrorist attacks, so many of their covariates are economic variables. *Trade* is measured as the annual sum of the value of a country's imports and exports of goods and services as a percentage of its gross domestic product (GDP). *Foreign direct investment* is measured as the sum of the absolute values of inflows and outflows of a country's foreign direct investments as a percentage of its GDP. *Financial capital flows* (*Portfolio*) is measured as the sum of the financial flows in equity securities and debt securities of a country as a percentage of its GDP. *Economic development* is measured as a country's yearly GDP per capita (logged) in purchasing power parity (PPP) adjusted to international prices. *Partner's economic development* is measured as the yearly GDP per capita (logged, in PPP) average of a country's exports to its largest eight destination countries. *Size* is measured as the log of a country's total population. *Democracy* is measured using the composite polity score of a country from the Polity IV data set. *Government capability* is measured as the annual composite percentage index of a country's portion of the world's (1) total population, (2) GDP per capita, (3) GDP per unit of energy, (4) military manpower, and (5) military expenditures. *Conflict* is a binary indicator that is coded 1 if a country is involved in an interstate military conflict or war and 0 otherwise.

[18]Transnational terrorist incidents are "initiated by foreign terrorists against some domestic target within a country or committed by domestic terrorists against some foreign target within a country" (Li and Schaub 2004, 231).

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

Table 2 Repeated events event history models of (re)occurrence of terrorist attacks, 1975–1997

	Conditional gap time	*Frailty (gamma; EM)*	*Conditional frailty (gamma; EM)*
Trade	−0.0001 (0.001)	−0.003 (0.002)	−0.0003 (0.001)
FDI	−0.02 (0.01)	−0.02 (0.02)	−0.03 (0.02)
Portfolio	−0.002 (0.01)	0.01 (0.01)	−0.001 (0.01)
Economic development	−0.04 (0.05)	0.32*** (0.10)	0.01 (0.06)
Partner's development	−0.77*** (0.11)	−0.70** (0.30)	−0.80*** (0.13)
Size	0.12*** (0.03)	0.25*** (0.07)	0.13*** (0.03)
Democracy	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Government capability	0.46*** (0.11)	0.13 (0.16)	0.39*** (0.12)
Conflict	0.17* (0.09)	−0.05 (0.17)	0.14 (0.15)
θ		0.51*** (0.35,0.71)	0.04 (0,0.13)
N	1996	1996	1996
Number of failures	1074	1074	1074
Likelihood ratio $\chi^2_{(.,1,1)}$ for θ		235.5***	2.54
I-likelihood		−4127.47	−4295.09
Log-likelihood for model	−4296.36	−3993.63	−4265.48
Wald $\chi^2_{(9,92.5,32.8)}$	133***	73.4	118***

Note. Significance (two tailed): *.1; **.05; ***.01.

For purposes of comparison and to triangulate on the existence of event dependence, heterogeneity, or both, we estimate a conditional gap time model and a frailty model in addition to the conditional frailty model. The results are found in Table 2.

The results from the conditional gap time model are presented in column one in Table 2. This model stratifies on the risk sequence, allowing the baseline hazards to vary across the risks. This allows us to control for the possibility that countries that have experienced a terrorist incident are likely to experience more of such attacks and "countries that experience long periods without terrorist activities may continue to enjoy an absence of such activity" (Li and Schaub 2004, 242). The cumulative hazards provide evidence for the presence of event dependence.[19]

The second column in Table 2 presents the results from estimating a frailty model. This model allows us to incorporate any systematic variation across countries in their propensity to experience terrorist incidents that are unaccounted for with the covariates in the model. This heterogeneity may result in some countries being more likely to experience a terrorist attack, that is, they are more frail because of an unknown, unmeasured, or unmeasurable reason. The model includes a frailty parameter to absorb the unmeasured risk factors causing the country heterogeneity. The inclusion of a frailty term allows us to determine whether there are any unmeasured risk factors that increase the hazard for more failure-prone countries.[20] We can do so by testing whether the variance of the frailty term

[19] As in the obscenity legislation example, the cumulative hazards for the conditional gap time and conditional frailty models are graphically indistinguishable, and so we present the cumulative hazards for the conditional frailty model. The cumulative hazards for the conditional gap time model can be found in the supplementary materials available on the *Political Analysis* Web site.

[20] The cumulative hazard and the frailty estimates by country for the frailty model can be found in the supplementary materials available on the *Political Analysis* Web site.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

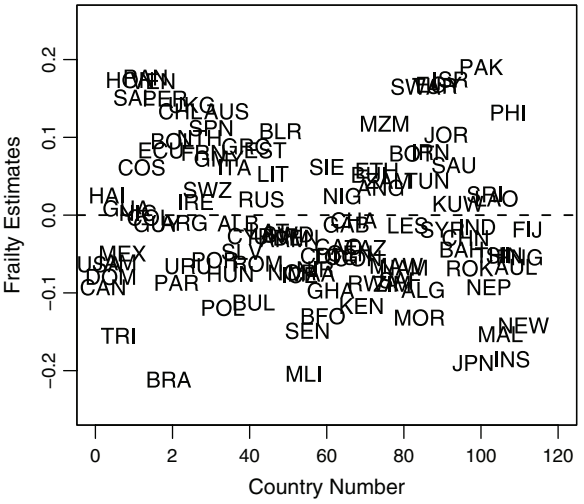


Fig. 5 Estimated frailties by country for the conditional frailty model of (re)occurrence of terrorist attacks.

is statistically different from zero.[21] A likelihood ratio test for θ finds evidence for a very significant frailty effect suggesting that unobserved heterogeneity is affecting our model. The variance of the statistically significant frailty term is 0.51. Thus, the interpretation of the estimated parameters is now conditional on inclusion of the frailty term.

The third column in Table 2 presents the results of the conditional frailty model. In contrast to the frailty model, the estimate of the variance of the frailty (0.04) is not statistically significant, suggesting that unobserved heterogeneity is not affecting our model when we account for event dependence.[22] The frailty estimates by country are depicted in Fig. 5 and the country names are provided in Table 3. Countries above the 0 reference line have higher frailties and thus are more failure prone and countries below 0 have lower frailties and are less failure prone. Less frail countries include Brazil, Mali, and Japan and more frail include Pakistan, Panama, and Israel. Figure 6, which plots the cumulative hazards by event number, reveals that the cumulative hazards vary by event number suggesting that event dependence is present. In general, risks increase more sharply as the number of events experienced increases. This suggests that the underlying DGP is characterized by event dependence.

In sum, the conditional gap time model suggests that there is event dependence and the statistically significant frailty term in the frailty model suggests that there is heterogeneity. However, the conditional frailty model reveals that event dependence is the main source of within-subject correlation. Drawing inference from a standard frailty model without stratification would have produced misleading conclusions because event dependence will masquerade as heterogeneity if separate baseline hazards are not also estimated.

[21]The likelihood ratio test for the frailty for this model is 2 (4245.22 − 4127.47) = 235.5, which is statistically significant based on a χ^2 test with 1 degree of freedom. The 95% confidence interval for $\hat{\theta}$, calculated using the same method as in the obscenity legislation example, was (0.35,0.71). A plot of the profile likelihood can be found in the supplementary materials available on the *Political Analysis* Web site.

[22]The likelihood ratio test for the frailty for this model is 2 (4296.36 − 4295.09) = 2.54, which is not statistically significant based on a χ^2 test with 1 degree of freedom. The 95% confidence interval for $\hat{\theta}$, calculated using the method mentioned above, was (0,0.13). A plot of the profile likelihood can be found in the supplementary materials available on the *Political Analysis* Web site.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

Table 3 Country names

Country	*Name*	*Country*	*Name*	*Country*	*Name*	*Country*	*Name*	*Country*	*Name*
USA	United States of America	IRE	Ireland	NOR	Norway	ZAM	Zambia	SRI	Sri Lanka
CAN	Canada	NTH	Netherlands	DEN	Denmark	ZIM	Zimbabwe	NEP	Nepal
HAI	Haiti	FRN	France	ICE	Iceland	MAW	Malawi	THI	Thailand
DOM	Dominican Republic	SWZ	Switzerland	MLI	Mali	SAF	South Africa	LAO	Laos
JAM	Jamaica	SPN	Spain	SEN	Senegal	NAM	Namibia	MAL	Malaysia
TRI	Trinidad & Tobago	POR	Portugal	MAA	Mauritania	LES	Lesotho	SIN	Singapore
MEX	Mexico	GMY	Germany	NIR	Niger	BOT	Botswana	PHI	Philippines
GUA	Guatemala	POL	Poland	CDI	Ivory Coast	SWA	Swaziland	INS	Indonesia
HON	Honduras	AUS	Austria	BFO	Burkina Faso	MOR	Morocco	AUL	Australia
SAL	El Salvador	HUN	Hungary	SIE	Sierra Leone	ALG	Algeria	PNG	Papua New Guinea
NIC	Nicaragua	ITA	Italy	GHA	Ghana	TUN	Tunisia	NEW	New Zealand
COS	Costa Rica	ALB	Albania	TOG	Togo	IRN	Iran	FIJ	Fiji
PAN	Panama	SLV	Slovenia	CAO	Cameroon	TUR	Turkey		
COL	Colombia	GRC	Greece	NIG	Nigeria	EGY	Egypt		
VEN	Venezuela	CYP	Cyprus	GAB	Gabon	SYR	Syria		
GUY	Guyana	BUL	Bulgaria	CEN	Central African Republic	JOR	Jordan		
ECU	Ecuador	ROM	Romania	CHA	Chad	ISR	Israel		
PER	Peru	RUS	Russia	CON	Congo	SAU	Saudi Arabia		
BRA	Brazil	EST	Estonia	KEN	Kenya	KUW	Kuwait		
BOL	Bolivia	LAT	Latvia	TAZ	Tanzania	BAH	Bahrain		
PAR	Paraguay	LIT	Lithuania	BUI	Burundi	CHN	China		
CHL	Chile	UKR	Ukraine	RWA	Rwanda	ROK	South Korea		
ARG	Argentina	BLR	Belarus	ETH	Ethiopia	JPN	Japan		
URU	Uruguay	ARM	Armenia	ANG	Angola	IND	India		
UKG	United Kingdom	SWD	Sweden	MZM	Mozambique	PAK	Pakistan		

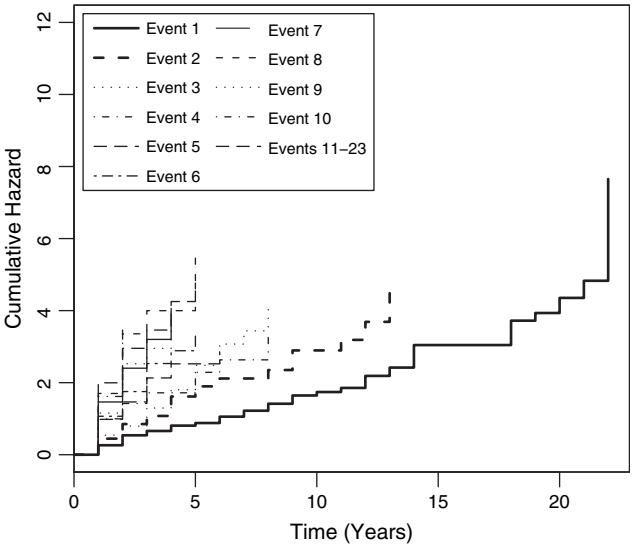


Fig. 6 Estimated cumulative hazards for each event number for the conditional frailty model of (re)occurrence of terrorist attacks.

6 Discussion and Conclusions

Analysts working with repeated events and interested in the timing of their occurrence need to be sensitive to the causes of within-subject correlation of event times. Both theoretically and empirically the evidence suggests that incorrectly assuming that within-subject correlation is a result of event dependence when it is due to heterogeneity in the data will result in biased estimates. Similarly, assuming that the correlation in event times is due to heterogeneity when in fact the source is event dependence will also result in biased inferences. Often these biases will be quite large. The conditional frailty model offers a way to simultaneously account for and determine the effects of both heterogeneity and event dependence, essentially arbitrating between commonly used models or rejecting them while explaining why their findings are biased. The cost associated with incorporating both stratification and random effects when only event dependence or heterogeneity, or neither, is present is a minimal loss of efficiency. Given that repeated events processes are likely to exhibit both heterogeneity and event dependence and typically neither can be ruled out a priori, the bias incurred when the analyst guesses wrong is too high a cost to pay. It can also be easily avoided by using the conditional frailty model.

Importantly, it is not just the covariate effects that should be considered when comparing across alternative models. The variance of the random effect and the plots of the cumulative (or baseline) hazard rates are vitally important for drawing correct substantive inferences.

The conditional frailty model offers the analyst theoretical insights. Evidence of event dependence in the event-specific cumulative hazards reveals patterns in the risk associated with different numbers of events. From this we can infer whether the risk of experiencing an event increases, decreases, or stays the same as more events are experienced. This is often of interest to analysts and policy makers alike as they try to understand the effects of covariates on reducing event rates. Additionally, a significant frailty term allows us to infer that the covariates do not on their own provide enough information about the timing and

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

occurrence of events across the units of analysis. As a plot of the frailties illustrates, for reasons that the model covariates do not distinguish, some units are more or less likely to experience events. Information about the distribution of estimated frailties, particularly which units have large and which small frailties, may help us to identify additional covariates to include in the model. Noting the units with large and small frailties may also tell us something intrinsic about the proneness of these units to experience more events than the rest of the sample. This information can be highly valuable to analysts.

Many processes of interest to social scientists involve the occurrence of repeated events over time. Understanding both the timing and occurrence of these events requires us to consider the possibility that conditional on the covariates in the model the occurrence of prior events may affect event rates and that, for unknown, unmeasured, or unmeasurable reasons, included covariates cannot account for the frailty of some units. The conditional frailty model offers analysts a model that does just that.

Funding

Ohio State University's Initiative in Population Research (1 R21 HK047943-01); National Institute of Child Care and Human Development.

References

Aalen, Odd O. 1988. Heterogeneity in survival anlaysis. *Statistics in Medicine* 7:1121–37.

Andersen, Per Kragh, John P. Klein, and Mei-Jie Zhang. 1999. Testing for centre effects in multi-centre survival studies: A monte carlo comparison of fixed and random effects tests. *Statistics in Medicine* 18:1489–1500.

Andersen, P. K., and R. D. Gill. 1982. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* 10:1100–20.

Berry, Francis Stokes, and William D. Berry. 1990. State lottery adoptions as policy innovations: An event history analysis. *American Political Science Review* 84:395–415.

Blossfeld, Hans-Peter, and Götz Rohwer. 1995. *Techniques of event history modeling.* Mahwah, NJ: Lawrence Erlbaum.

Bowman, Michael E. 1996. An evaluation of statistical models for the analysis of recurrent events data: With application to needlestick injuries among a cohort of female veterinarians. PhD diss., Ohio State Univ.

Box-Steffensmeier, Janet M., and Christopher Zorn. 2002. Duration models for repeated events. *Journal of Politics* 64:1069–94.

Box-Steffensmeier, Janet M., and Suzanna De Boef. 2006. Repeated events survival models: The conditional frailty model. *Statistics in Medicine* 25:3518–33.

Clayton, David. 1994. Some approaches to the analysis of recurrent event data. *Statistical Methods in Medical Research* 3:227–62.

Cleves, Mario. 1999. Analysis of multiple failure-time data using Stata. *Stata Technical Bulletin* 49:30–9.

Cook, Richard J., Edmund T. M. Ng, Jayanti Mukherjee, and David Vaughn. 1999. Two-state mixed renewal processes for chronic disease. *Statistics in Medicine* 18:175–88.

Cook, Richard J., and Jerald F. Lawless. 1997. Marginal analysis of recurrent events and a terminating event. *Statistics in Medicine* 16:841–51.

Cook, Richard J., and Jerald F. Lawless. 2002. Analysis of repeated events. *Statistical Methods in Medical Research* 11:141–66.

Cox, D. R. 1972. Regression models and life tables. *Journal of the Royal Statistical Society* B34:187–220.

Fong, Daniel Y. T., K. F. Lam, J. F. Lawless, and Y. W. Lee. 2001. Dynamic random effects models for times between repeated events. *Lifetime Data Analysis* 7:345–62.

Gail, M. H., T. J. Santner, and C. C. Brown. 1980. An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* 36:255–66.

Gao, Feng, Amita K. Manatunga, and Shande Chen. 2007. Non-parametric estimation for baseline hazards function and covariate effects with time-dependent covariates. *Statistics in Medicine* 26:857–68.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press

Gao, Sujuan, and Xiao-Hua Zhow. 1997. An empirical comparison of two semi-parametric approaches for the estimation of covariate effects from multivariate failure time data. *Statistics in Medicine* 16:2049–62.

Hausman, Jerry A. 1978. Specification tests in econometrics. *Econometrica* 46:1251–71.

Hougaard, Philip. 1991. Modelling heterogeneity in survival data. *Journal of Applied Probability* 28:695–701.

Hougaard, Philip. 2000. *Analysis of multivariate survival data.* New York: Springer-Verlag.

Hurvich, Clifford M., Jeffrey S. Simonoff, and Chih-Ling Tsai. 1998. Smoothing parameter selection in non-parametric regression using an improved Akaike information criterion. *Journal of Royal Statistical Society* B60:271–93.

Jones, Bradford S., and Regina P. Branton. 2005. Beyond logit and probit: Cox duration models of single, repeating, and competing events for state policy adoption. *State Politics and Policy Quarterly* 5:420–43.

Keiding, Niels, Per Kragh Andersen, and John P. Klein. 1997. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* 16:215–24.

Kelly, Patrick J., and Lynette L.-Y. Lim. 2000. Survival analysis for recurrent event data: An application to childhood infectious disease. *Statistics in Medicine* 19:13–33.

Klein, John P., and Melvin L. Moeschberger. 1997. *Survival analysis: Techniques for censored and truncated data.* New York: Springer-Verlag.

Kosorok, Michael R., Bee Leng Lee, and Jason P. Fine. 2004. Robust inference for univariate proportional hazards frailty regression models. *Annals of Statistics* 32:1448–91.

Lawless, Jerald F. 2002. *Statistical models and methods for lifetime data.* 2nd ed. New York: Wiley.

Lawless, J. F., and C. Nadeau. 1995. Some simple robust methods for the analysis of recurrent events. *Technometrics* 37:158–68.

Li, Quan, and Drew Schaub. 2004. Economic globalization and transnational terrorism: A pooled time-series analysis. *Journal of Conflict Resolution* 48:230–58.

Li, Yi, and Xihong Lin. 2003. Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach. *Journal of the American Statistical Association* 98:191–203.

Lin, D. Y. 1994. Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine* 13:2233–47.

Mintrom, Michael, and Sandra Vergari. 1998. Policy networks and innovation diffusion: The case of state education reforms. *Journal of Politics* 60:126–48.

Mooney, Christopher Z., and Mei-Hsien Lee. 1995. Legislative morality in the American states: The case of pre-Roe abortion regulation reform. *American Journal of Political Science* 39:599–627.

Nelson, Wayne B. 2003. *Recurrent events data analysis for product repairs, disease recurrences, and other applications.* Philadelphia, PA: Society for Industrial and Applied Mathematics.

Oakes, D. 1992. Frailty models for multiple event times. In *Survival analysis: State of the art*, ed. John P. Klein and Prem K. Goele, 371–9. Dordrecht, Netherlands: Kluwer Academic Publishers.

Oakes, D. A. 1997. Model based and or marginal analysis for multiple event time data. In *Proceedings of the first Seattle symposium in biostatistics: Survival analysis*, ed. D. Y. Lin and T. R. Fleming, 85–98. New York: Springer-Verlag.

Ohman-Strickland, Pamela. 2003. Fitting the frailty distribution using empirical Bayes. International Conference on Reliability and Survival Analysis, Columbia, SC.

Prentice, R. L., B. J. Williams, and A. V. Peterson. 1981. On the regression analysis of multivariate failure time data. *Biometrika* 68:373–9.

R Development Core Team. 2007. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Sastry, Narayan. 1997. A nested frailty model for survival data, with an application to the study of child survival in Northeast Brazil. *Journal of the American Statistical Association* 92:426–35.

Therneau, Terry M. 2000. *Practical frailty models. Technical report.* Mayo Clinic.

Therneau, Terry M., and Patricia M. Grambsch. 2000. *Modeling survival data: Extending the Cox model.* New York: Springer-Verlag.

Wei, L. J., D. Y. Lin, and L. Weissfeld. 1989. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 84:1065–73.

https://doi.org/10.1093/pan/mpm013 Published online by Cambridge University Press