Duration Models for Repeated Events

Janet M. Box-Steffensmeier
Ohio State University
Christopher Zorn
Emory University

An important feature of most political events is their repeatability: nearly all political events reoccur, and theories of learning, path dependence, and institutional change all suggest that later events will differ from earlier ones. Yet, most models for event history analysis fail to account for repeated events, a fact that can yield misleading results in practice. We present a class of duration models for analyzing repeated events, discuss their properties and implementation, and offer recommendations for their use by applied researchers. We illustrate these methods through an application to widely used data on international conflict.

Introduction

In recent years, the application of duration (or event history) models in political science has grown dramatically. Increased interest in the temporal dynamics of political processes, coupled with the greater availability of longitudinal data and increases in computing power, has led scholars to adopt such methods with greater frequency.[1] The development of models for event histories, however, has occurred almost entirely outside the scope of political science; as a result, many of the issues prevalent in political science duration data have received only limited attention in work in economics, biostatistics, epidemiology, and other areas.

Thanks to Mario Cleves, John Freeman, Stanley Lemeshow, Melvin Moeschberger, Dan Reiter, Narayan Sastry, Terry Therneau, anonymous reviewers, and the editor for helpful discussions, and to Kevin Sweeney and Andy Tomlinson for research assistance. Box-Steffensmeier thanks the National Science Foundation for support via a Mid-Career Methodological Opportunities Grant, number SES-0083418, and Zorn thanks the John M. Olin Foundation for a Faculty Fellowship supporting this research. An earlier version of this paper was presented at the 1999 Summer Meeting of the Political Methodology Society, Texas A&M University, College Station, and at the 2001 Speaker Series for The Center for Biostatistics and The School of Public Health, Division of Epidemiology and Biometrics, Ohio State University.

[1] See Beck (1998), Box-Steffensmeier and Jones (2002), Hosmer and Lemeshow (1999), and Klein and Moeschberger (1997) for accessible introductions to duration models.

THE JOURNAL OF POLITICS, Vol. 64, No. 4, November 2002, Pp. 1069–1094

One such issue is that of repeated events. Unlike in most clinical trials, in which the event under study may occur only once (e.g., the death of a patient), the majority of events defining durations in political science are repeatable. Such diverse phenomena as international conflicts (Beck, Katz, and Tucker 1998), cabinet transitions (King et al. 1990), presidential nominations for executive branch appointments (McCarty and Razaghian 1999), and EU decision making (Schulz and König 2000) are all examples of processes that may occur repeatedly over time. Moreover, the existence of repeated events raises important methodological and substantive issues about the way in which we conduct duration analyses.

The purpose of this article is to address the issue of repeated-events duration analysis, both in general and in the more specific context of political science research. We begin with a general discussion of the importance, both statistical and substantive, of addressing repeated events when they arise in our data. We then outline a range of methods for modeling such events and discuss how the various approaches differ. We go on to illustrate the application of these techniques on a widely used data set of dyadic conflicts between nations during the post-World War II period. Those data are characterized by large numbers of repeated events, and our reanalysis of the data, using techniques that account for the repeated occurrence of dyadic disputes, uncovers important differences from previous studies that failed to address the issue of repeated events. We conclude with a brief review of our findings and provide a set of suggestions for applied researchers faced with the prospect of analyzing repeated events data.

Repeated Events and the Analysis of Political Data

Models for duration data were initially developed in the health sciences, and the bulk of the innovation in such models continues in such fields as biostatistics and epidemiology. There, the canonical study is one of mortality: researchers use duration analysis to determine the effectiveness of some treatment on prolonging the lifespan of its subjects. As a result, it is only recently that work in such areas has begun to address the issue of repeatable events. By contrast, the nature of most political science phenomena is that they are capable of repetition: temporal changes in individuals, groups, and institutions are often ongoing and may occur many times over a study period. Moreover, scholars have long understood the contingent nature of repeated political events. Whether through learning, path dependence, or other mechanisms, it is almost always the case that our subjects—be they voters, nations, or others—respond differently to reoccurrences of the same phenomena.

The study of international conflict is a prominent and important case in point. Whereas one-time conflicts between nations may reflect discrete events or short-term political forces, repeated disputes are likely to be an indicator of more fundamental, and longer lasting, differences. Moreover, under those circumstances, nations involved in an ongoing series of confrontations will respond differently to internal and external pressures than they would were the dispute centered around an isolated event. Students of international relations are well aware of this aspect of temporal dependence, as indicated by the extensive attention given to enduring rivalries (e.g., Diehl 1998).

More generally, examples of repeated events are the rule rather than the exception in political science and cross all subfields of the discipline. The study of executive branch nominations is just one example in American politics. McCarty and Razaghian (1999) examine the duration of the confirmation process for more than 3,500 executive-branch nominations and recognize the distinctive qualities of being a repeated nominee in their article. They argue that repeated nominations will proceed faster since "if one assumes that the primary purpose of the confirmation process is to acquire information on and evaluate the qualifications of various nominees, we should expect to see individuals who have recently been confirmed by the Senate to have an easier time getting confirmed subsequently" (1999, 1131).

Hammons' (1999) important article on the longevity of state constitutions serves as another example, this time in the area of state politics. He concludes that the "longer and more detailed design of state constitutions actually enhances rather than reduces their longevity" (1999, 837). The data used for his study come from the 50 states. Over half of the states have had more than one constitution; Louisiana has had the most events with eleven constitutions (1999, Table 2, 841). Clearly, one would expect that the durability of these eleven constitutions are correlated and that statistically accounting for the correlation may make a difference in the conclusions drawn.

In the realm of comparative politics, the study of leadership duration (e.g., Bienen and van de Walle 1992; Bueno de Mesquita and Siverson 1995; Londregan and Poole 1990) serves an obvious example to illustrate the potential advantages of taking into account repeated events. Bienen and van de Walle study 2,256 leaders from 167 different countries. To the extent that heads of state may learn from the experiences of that nation's past leaders—avoiding their mistakes and mimicking their successes—a repeated events analysis of the duration of a country's leaders seems promising.

Variance-Correction Models for Repeated Events

In each of these examples, the issue of repeated events is one of dependence: second and subsequent events are likely to be influenced by, and therefore different from, first events. As a result, analyses that treat repeated events as independent, when in fact they are not, run the risk of yielding misleading results for at least two reasons. First, the presence of correlated events presents a problem similar to autocorrelation in conventional regression analysis: by treating such observations as independent, we overstate the amount of information each observation provides, leading to incorrect estimates of standard errors. Second, such models implicitly restrict the influence of covariates to be the same across events when, in fact, there may be varying effects from one event occurrence to the next. Given these potential problems, it is unsurprising to find that "Multiple event studies, where events are of the same type [as opposed to the competing risks framework for different event types], are the most difficult for analysis" (Therneau and Hamilton 1997, 2043) and that different models for repeated events with serial duration times may yield widely different empirical conclusions (e.g., Clayton 1994; Gao and Zhou 1997; Lin 1994; Wei and Glidden 1997).

We consider the issue of repeated events in the context of Cox's (1972) proportional hazards model:

$$h(t) = \lambda_0(t) exp(\mathbf{X}_{it}\beta) \qquad \textbf{(1)}$$

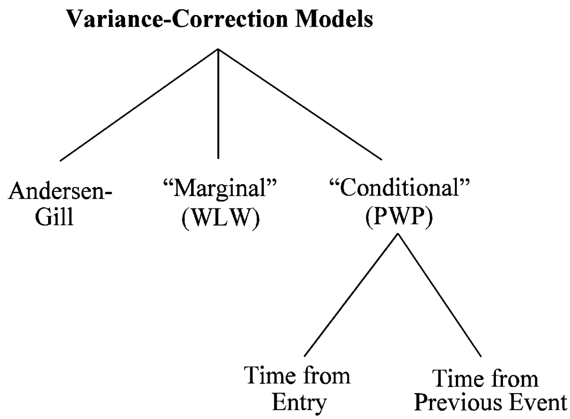
where $h(t)$ is the hazard of the event of interest, $\mathbf{X}_{it}$ is a vector of covariates, and λ_0 is an unspecified baseline hazard.[2] The dominant approach to repeated events duration models is the class of *variance-correction* models. Variance-correction models take advantage of the fact that in the presence of repeated events, standard Cox model estimates for β converge to a well-defined vector (usually termed β^*), which can often be interpreted meaningfully, but the estimated covariance matrix is inappropriate for hypothesis testing (Lin and Wei 1989; Struthers and Kalbfleisch 1984). These models thus estimate a standard Cox model and adjust the variance-covariance matrix to account for the individual- or group-specific effects that remain.[3]

[2] Parametric models, most notably those based on the Weibull distribution, have also seen widespread use in political science. Work addressing repeated events in the Weibull context are limited to the random-effects class, typically with gamma-distributed frailties (e.g., Lancaster 1979, 1985; Larsen and Vaupel 1993; Vaupel, Manton, and Stallard 1979). Because of the strong parametric assumptions required by the Weibull model, it is not as widely used outside political science as is the Cox model. Accordingly, we focus on the Cox model, and generally prefer it to the Weibull due to its less restrictive assumptions about the data-generating process and better developed diagnostics for model checking. Importantly, some scholars have inappropriately used the existence of "tied" observations, that is, coterminous event occurrences, as a reason not to use the Cox model. Advances in the computational methods to address the issue of ties (such as the Efron approximation and exact likelihood methods) have all but alleviated such concerns; see Box-Steffensmeier and Jones 2002 for a discussion.

[3] An alternative approach is to use models that allow for unit-specific effects; such models generally fall into two classes, *fixed-effects* and *random-effects* (or *frailty*) models. Both approaches consider repeated events as a special case of more general unit-level heterogeneity. Fixed-effects approaches treat such effects as a fixed quantity to be estimated; such models have significant consistency problems (e.g., Andersen, Klein, and Zhang 1999) and so are rarely used. Frailty models treat the individual effects as random draws from a specific parametric distribution, the parameters of which are then estimated along with the structural parameters of the model (e.g., Hougaard 1991; Keiding, Andersen, and Klein 1997; Oakes 1992; Sastry 1997). Frailty models have been the target of two primary criticisms. First, neither theory nor data typically provides much guidance for imposing a specific distribution on the frailties, and "parameter estimates *can* be highly sensitive to the assumed parametric form of the error term" (Blossfeld and Rohwer 1995, 255). Second, the random effects are generally required to be independent of the model's covariates; failure to meet this assumption can yield estimates that are badly biased and inconsistent (Hausman 1978).

FIGURE 1

Schematic of Approaches to Repeated Events in Duration Models



We focus on four widely used variance-correction models: the "independent increments" model developed by Anderson and Gill (1982), the marginal risk-set model of Wei, Lin, and Weissfeld (1989), and the conditional risk-set model (Prentice, Williams, and Peterson 1981), which may be estimated in either elapsed or interevent ("gap") time. These four models are illustrated schematically in Figure 1. All are related in that, compared to the standard Cox model, they use the non-independence caused by repeated events to empirically correct the standard error estimates. The key distinction among these models is "the *way that the risk sets are defined* at each failure" (Cleves 1999, 34). The risk set defines which observation may fail at a particular time; as a result of the different risk-set definitions, very different processes are modeled by the four alternatives. Thus the estimated coefficients will vary among these four variance corrected models.

The simplest variance-correction model is that of Andersen and Gill (1982) (hereafter AG). The key 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" (Oakes 1992, 372). That is, multiple events for any particular observation are assumed to be conditionally independent[4]; for this reason, the AG model is often referred to as the "independent increment" model. Therneau and Hamilton note that if events are not independent, robust variance estimates (e.g., White 1980) allowing for clustering within

[4] More specifically, event arrivals follow a nonhomogeneous Poisson process.

units may be used (1997, 2034).[5] They also point out that effects that may change with the number of the event may be modeled explicitly, using time-dependent covariates (e.g., interactions of treatment effects with the number of previous events).

In practical terms, the Cox and AG models are essentially indistinguishable, and in fact the former can be shown to be a special case of the latter (Fleming and Harrington 1991, 164). Thus, while the AG approach is straightforward to estimate, the assumption of independent increments is strong, particularly if the ordering of events may be important. Also, unlike the other models considered here, the AG model restricts the baseline hazard rate for all events to be the same. For many applied problems, the assumption of independent increments will not be acceptable, at least not without empirical testing.

The marginal risk-set approach of Wei, Lin, and Weissfeld (1989) (hereafter WLW) applies the traditional competing risks set-up for multiple events to repeated events. Ordered events data are treated *as if* they presented a typical competing risks problem: each observation is "at risk" for the first, second, third, etc., event from the beginning of the study period.[6] The data are then stratified by event number, and separate baseline hazards are estimated at the first occurrence of the event under study, the second, etc. The approach is thus referred to as the "marginal risk set" model because, within these event-defined strata, marginal data are used, "that is, 'what would result if the data recorder ignored all information except the given event type?'" (Therneau and Hamilton 1997, 2035). As a result, at any point in time, all observations that have not yet experienced k events are assumed to be "at risk" for the kth event. Unlike the AG model, stratification by event allows baseline hazards for each event to differ; as in the AG model, however, covariate effects are assumed to be constant across event ranks.[7]

The signature characteristic of the WLW approach is that all observations are at risk for all events at all times prior to experiencing that event. That is, in the case of repeated events of the same type, the "fifth" event can (in theory) occur at any time, even prior to the "first," "second," etc. events. Whether this is a plausible assumption about one's data-generating process depends on the

[5] Note that, if the independent increments assumption holds, the naive and robust standard errors will be identical.

[6] The difference between this model and a standard competing risks problem is that in the latter the events are of different types.

[7] Recall that stratification is used to permit flexible variation in the baseline hazard, not for an estimate of variable effects. Stratification is vitally important because in repeated event data the hazard rate is likely to be very different for the first failure compared to say, the fifth failure. One can, as Wei, Lin, and Weissfeld (1989) did, include strata-by-covariate interactions to estimate separate effects. Alternatively, one can estimate models for each strata (i.e., each repeated event) separately and obtain strata specific coefficients, which can be illuminating. However, this should only be done if there are theoretical reasons to think that *both* the baseline hazards *and* the variable effects vary across events.

nature of the question at hand. Wei, Lin, and Weissfeld (1989, 1065) give the example of the time until recurrence of tumors among bladder cancer patients: because more than one tumor may develop at a time, the model's assumption that individuals are at risk for the first, second, etc. of several identical events is a reasonable one.

One could imagine similar processes in political data. The occurrence of coups, for example, may be caused by the actions of multiple, different factions in a country. Under these circumstances, a regime may be "at risk" for multiple coups from the beginning of its tenure. Moreover, the occurrence of one coup will undoubtedly influence the incidence and success of subsequent attempts (Londregan and Poole 1990), thus suggesting the WLW model may be appropriate. In most instances involving political data, however, repeated events occur sequentially and assume a natural order. Thus, while some have strongly recommended use of the WLW model for the analysis of repeated events data (Barai and Teoh 1997; Lin 1994; Therneau and Hamilton 1997; Wei and Glidden 1997), we concur with others who have expressed concern about the appropriateness and logic of this signature characteristic in most repeated events contexts (Cook and Lawless 1997; Kelly and Lim 2000; Oakes 1997).

By comparison, in the conditional model of Prentice, Williams, and Peterson (1981) (hereafter PWP), an observation is not at risk for a later event until all prior events have already occurred. Accordingly, the "risk set" at time t for the kth occurrence of an event is limited to those observations under study at time t who have already experienced $k - 1$ events of that type.[8] As in the WLW model, estimates are then stratified by event rank, so that the different events have varying baseline hazards. As in the previous models, however, covariate effects are again assumed to be constant across strata, though as in the WLW model strata-by-covariate interactions may be estimated. An additional feature of the conditional risks model is that the model may be estimated in either *elapsed* time (i.e., time from each unit's entry into the observation set) or in *interevent* time (also referred to as "gap-time"), defined as the duration since the previous event.[9] The PWP model's explicit ordering of sequential events makes it an intuitively appealing choice for the majority of repeated events applications in political science, particularly in instances where the events in question necessarily occur sequentially and where a unit is not "at risk" for a later event until all previous such events have occurred. Moreover, Oakes (1992) notes that because the PWP model takes into account the ordering of events, it

[8] Therneau and Hamilton (1997) clearly illustrate the difference between the WLW and PWP models by pointing out that if events occurred at 100 and 185 days and the subject has been observed for 250 days, then the WLW model treats the subject as "at risk" for the second event from day 0 to day 185. In contrast, since an observation cannot be at risk for event two until event one occurs, the PWP model only places the subject "at risk" for the former from day 101 to day 185.

[9] The latter correspond to "renewal" or "semi-Markov" models (e.g., Lancaster 1990, 85–97), of which the Poisson process is a special case.

provides efficiency gains over the marginal model. The similarities and differences of the variance-corrected models are summarized in Table 1.[10] The partial likelihoods are the same for the WLW, PWP elapsed time, and PWP gap time models. Note that the partial likelihood for the AG model is unrestricted/unstratified, whereas it is stratified for the other three models.[11]

All three models use robust variance estimates (Lin and Wei 1989) to address the potential for interdependence due to repeated events. Robust standard errors assume that observations are independent across units (or "clusters") but not necessarily within those units. The robust variance estimator is then based on a "sandwich" estimate:

$$\mathbf{V_R} = \mathbf{V}^{-1}\mathbf{B}\mathbf{V}^{-1} \tag{2}$$

where $\mathbf{V}^{-1}$ is the usual variance estimate of a Cox model (that is, the inverse of the information matrix $\mathbf{V}$) and $\mathbf{B}$ is a correction factor (see Appendix A for details). There are several ways to motivate this correction, for example, as the proper variance when a likelihood for distribution f is fit, but the data come from g (Huber 1967) or as an approximation to the jackknife estimate (Therneau 1997). Because the unobserved intracase correlations are generally positive, the estimates of the variance-corrected standard errors are typically larger than those from a "naive" estimate based on $\mathbf{V}^{-1}$.

[10] Considerable care should be taken when organizing one's data for the variance-corrected models because the data organization defines the risk set. Therneau (1997), Therneau and Hamilton (1997), and Cleves (1999) present exceptionally clear descriptions of this process. For the AG and PWP models, each subject must contain one row (for time-invariant data) or set of rows (for time-varying data) for each event or censored period. For the WLW model, by contrast, each observation (or set of time-varying observations) appears in the data once for every *possible* event rank. Thus, if the researcher observed a maximum of k events, information for each subject would appear in the data k times, once for each event rank. In all three variance-corrected models, time-varying covariates may be included, at some complication to the data structure. Likewise, because of the differences in data organization, it should also be stressed that the various variance-correction models "may yield parameter estimates of a similar value for the same covariate, yet their interpretation is quite different because entirely different processes are being modeled" (Hosmer and Lemeshow 1999, 312). This is particularly true regarding the differences between elapsed-time and gap-time implementations of the PWP model; we discuss this at greater length below. For interested readers, we have provided an appendix that outlines and illustrates the organization of the data for the models discussed here for both time-constant and time-varying data; this appendix is available at http://www.emory.edu/POLS/zorn/Data/.

[11] The notation for the partial likelihoods and hazards comes directly from Kelly and Lim (2000). Specifically, T_{ik} is the time until the kth event for the ith subject, C_{ik} is the censoring time for the kth event in the ith subject, and X_{ik} is the observed duration (i.e., $X_{ik} = \min(T_{ik}, C_{ik})$). The censoring indicator is $\delta_{ik} = I(T_{ik} \leq C_{ik})$. Interevent times are defined as $G_{ik} = X_{ik} - X_{i,k-1}$, with $X_{i0} = 0$. They also define $\lambda_{ik}(t)$ to be the hazard function for the kth event of the ith subject at time t, $\lambda_0(t)$ represents a common baseline hazard for all events, and $\lambda_{0k}(t)$ is an event-specific baseline hazard for the kth event. Covariates are defined as $Z_{ik} = (Z1_{ik}, \ldots, Z_{pik})'$ for the ith subject with respect to the kth event, with $Z_i = (Z'_{i1}, \ldots, Z'_{iK})$ as the covariate vector for the ith subject, where K is the maximum number of events within a subject, and $\beta = (\beta_1, \ldots, \beta_p)'$ is the $p \times 1$ vector of regression parameters to be estimated (Kelly and Lim 2000, 18).

TABLE 1

Comparison of Variance-Correction Models for Repeated Events

Model Property	Andersen-Gill (AG)	Marginal (WLW)	Conditional (PWP), Elapsed Time	Conditional (PWP), Interevent Time
Risk Set for Event k at Time t	Independent Events	All Subjects That Haven't Experienced Event k at Time t	All Subjects That Have Experienced Event $k-1$, and Haven't Experienced Event k, at Time t	
Time Scale	Duration Since Starting Observation	Duration Since Starting Observation	Duration Since Starting Observation	Duration Since Previous Event
Robust Standard Errors?	Yes	Yes	Yes	
Stratification by Event?	No	Yes	Yes	
Partial Likelihood	$L(\beta)=\prod_{i=1}^{n}\prod_{k=1}^{K}\left(\frac{e^{\beta' Z_{ik}(X_{ik})}}{\sum_{j=1}^{n}\sum_{l=1}^{K} Y_{jl}(X_{ik})e^{\beta' Z_{jl}(X_{ik})}}\right)^{\delta_{ik}}$		$L(\beta)=\prod_{i=1}^{n}\prod_{k=1}^{K}\left(\frac{e^{\beta' Z_{ik}(X_{ik})}}{\sum_{j=1}^{n} Y_{jk}(X_{ik})e^{\beta' Z_{ik}(X_{ik})}}\right)^{\delta_{ik}}$	
Hazard	$\lambda_{ik}(t;Z_{ik})=\lambda_{o}(t)e^{\beta' Z_{ik}(t)}$ $Y_{ik}(t)=I(X_{i,k-1}<t\leq X_{ik})$	$\lambda_{ik}(t;Z_{ik})=\lambda_{ok}(t)e^{\beta' Z_{ik}(t)}$ $Y_{ik}(t)=I(X_{ik}\geq t)$	$\lambda_{ik}(t;Z_{ik})=\lambda_{ok}(t)e^{\beta' Z_{ik}(t)}$ $Y_{ik}(t)=I(X_{i,k-1}<t\leq X_{ik})$	$\lambda_{ik}(t;Z_{ik})=\lambda_{ok}(t-t_{k-1})e^{\beta' Z_{ik}(t)}$ $Z_{jk}(X_{ik})$ *replaced by* $Z_{ik}(X_{i,k-1}-G_{ik})$ $Y_{ik}(t)=I(G_{ik}>t)$

Model Comparisons

In comparing these alternative models, the key factor to be considered is the consistency of the model's assumptions with the data-generating process in question. The primary advantage of the AG model is its simplicity; in addition, the counting process formulation of the AG model has the advantage of preserving the sequence of events in that it only allows subjects to be "at risk" for the kth event if that unit has already experienced $k - 1$ previous ones (see Fleming and Harrington 1991). Its primary drawbacks are its assumption of independence across events within a unit, the restriction of the baseline hazards across all events, and its concomitant inability to estimate separate variable effects for those events except via time-dependent covariates, a technique Lin (1994) suggests may yield misleading results.

By contrast, the WLW model estimates different baseline hazards across events and provides estimates of the possible change in treatment effect over time through the use of strata-by-variable interactions (Therneau and Hamilton 1997, 2044). In this respect, the WLW model may be a better choice when events are nonindependent or when covariate effects may change across events. At the same time, numerous authors have questioned the WLW assumptions with respect to the specification of the risk set (e.g., Cook and Lawless 1997); in many instances, the failure of the WLW model to preserve the *sequence* of events in specification of the risk set will not accurately reflect the data-generating process. Oakes (1992) notes that relative to the conditional model, the marginal model is inefficient because it fails to make use of the information contained in the ordering of events. An equally important issue is that the marginal model may badly strain the proportional hazards assumption (Therneau 1997; see also Box-Steffensmeier and Zorn 2001), particularly in instances where separate, strata-specific covariate effects are not specified.

Intuitively, the PWP model's preservation of the order of sequential events in the creation of the risk set renders it an attractive alternative. The appeal of this model is supported by several recent studies comparing the various variance-corrected models. Kelly and Lim (2000) evaluate four key components (risk intervals, baseline hazards, risk sets, and correlation adjustments) with simulated and actual data in order to systematically identify how variance-corrected models differ from each other and to assess the best modeling approach for repeated events. They recommend the conditional (PWP) gap-time model for repeated events, particularly when within-subject events are conditionally independent. Similarly, Bowman (1996) uses simulation evidence to determine the best approach to the analysis of repeated events data. He considers the level of significance, power, bias, and mean squared error for several estimators and also concludes that the PWP model is preferred. Thus, analysts using variance-corrected approaches have strong statistical evidence favoring the conditional model for repeated events analysis from among the possibilities available.

Within the PWP model, one still needs to consider which risk interval (time from entry or time from previous event) follows most closely from the question being posed by the substantive inquiry. Note that elapsed times "have a 'carry-over' effect . . . the total time of the second risk interval includes the first interval; the third risk interval contains the first and the second intervals, and so on" (Kelly and Lim 2000, 28). That is, elapsed times are typically correlated even if gap times are not (Lipschutz and Snapinn 1997). In some applied contexts, the choice is relatively clear. For example, in a recent study of when and why members of the U.S. House of Representatives change their long-term positions on controversial issues, such as abortion, considering elapsed time clearly makes the most sense. Meinke (2001) contends that there are important limits to the stability that scholars have long observed in congressional voting: members follow their past positions in order to make economical decisions on similar questions over time, but changes in their decision environment can lead them to discount the value of their past decisions and to change their position, even when their actual policy views have not changed. Here, the elapsed time method is superior to gap time "because members can be thought of as developing a risk of switching not only the first time but also a second or third time from the time of their first vote" (2001, 35). Put differently, members develop a risk of position instability, which begins with their first vote on an issue.

In most instances, we expect that the PWP gap-time approach will offer the best combination of characteristics for addressing questions of interest to political scientists, including such diverse studies as those of the duration of state constitutions (see Hammons 1999), confirmation of executive branch nominations (McCarty and Razaghian 1999) and international conflict (e.g., Beck, Katz, and Tucker 1998; Oneal and Russett 1997, 1999). Using elapsed time presumes there are substantive reasons to believe that "clock should restart" after each event; such a model is used to determine the effect of covariates on the kth event since the time from the previous event. In contrast, elapsed time models assess the effect of covariates on the kth event since the time of the start of the study. Using gap time, one could determine, for example, whether an imposed settlement would be effective for delaying the first subsequent conflict but not second and subsequent ones (Werner 1999).

The primary criticism of the conditional model in light of this recent work is that for higher ranked events, the risk set may be very small (since few observations have experienced the events necessary to place them in the risk set for that event), yielding estimates that may be both unstable and imprecise (Wei, Lin, and Weissfeld 1989). This is particularly true when separate variable effects are estimated for second and higher order events. However, this can be relatively easily addressed by combining several higher level risk sets based on theoretical and/or statistical criteria. Alternatively, the analyst can acknowledge that the standard error bounds may be very large for estimates on higher ranked events due to the low numbers of uncensored observations in those categories.

For applied researchers, the key factors to consider when addressing repeated events should be the characteristics of the data-generating process. That is, consideration must be given first and foremost to the process by which repeated events occur. Importantly, researchers need to remember that the interpretation differs across these models because of the different conditioning assumptions.

We discuss the models in the order they appear in our paper. First, despite being one of the most widely used models in the literature, the AG model's unrestricted/unstratified risk set make it an unacceptable choice if the hazards are expected to change after each event. Consequently, we expect that the AG model will rarely be appropriate for political science applications. In fact, if there is any doubt that the baseline hazard *may* change with each event, this restrictive assumption of the AG model needs to be relaxed; thus we expect the AG model to be satisfactory only in unusual cases. In short, the AG model lacks the detail and versatility of event-specific models (Kelly and Lim 2000, 32). If multiple events may develop simultaneously (as in the coup example), marginal approaches are appropriate. However, we expect this also to be rare; in general, the assumption that an observation can be at risk for the kth event even before any or all $k - 1$ events have occurred lead us to expect that the WLW model will not be widely used for the study of repeated events. Purely sequential events, such as intradyadic conflicts, suggest methods that retain the ordering of events. In these cases, a PWP conditional model should be used. Below, we illustrate the use of these various models in practice and show how they lead to significantly different results in the analysis of international conflicts during the post-World War II era.

Repeated Events and International Conflict

While its importance has always been without question, the prevention of international military conflict has taken on particular significance in our increasingly interdependent world. Indeed, Kirby (1999) foresees that with the rise of the global community, there will be more frequent and pronounced tensions between states as they attempt to maintain political and ideological integrity in the face of decreasingly autonomous economic and cultural identities. Thus, interstate "conflict is a critical, core, international political interaction" (Diehl 1999), and despite the controversies between realists and neoliberal international relations scholars the study of conflict remains a central focus of the field.

In recognition of its central role in the creation and maintenance of international security, questions surrounding the causes of international conflict have received a tremendous amount of attention by scholars. While significant debates exist regarding those causes, arguably the most prominent view is that of the "liberal peace" (e.g., Russett 1993). Liberal theory is the locus of the hugely influential "democratic peace" literature, currently one of the most important and active research programs in political science, and one with enormous policy ramifications.[12] The fact that democracies rarely fight one another has come to be regarded as part of the conventional wisdom and has been characterized as the only law-like generalization in the study of international relations (Levy 1988). A host of studies have addressed the origins, existence, and extent and of the democratic peace phenomenon (e.g., Bremer 1992; Bueno de Mesquita et al. 1999; Cederman 2001; Dixon 1994; Fearon 1994; Lake 1992; Maoz 1998; Maoz and Abdolai 1989; Maoz and Russett 1993; Oneal and Russett 1997; Ray 1995; Rousseau et al. 1996; Russett 1993; Ward and Gleditsch 1998). Liberal international relations theorists have also led the way in considering the relationship between conflict and economic interdependence (e.g., Barbieri and Schneider 1999; Beck, Katz, and Tucker 1998; Mansfield 1994; Morrow 1999; Oneal and Russett 1997, 1999), as well as that between conflict and international organizations (e.g., Diehl 1997; Russett, Oneal, and Davis 1998).

At the same time, there is a sharp difference between much of the work on the liberal peace and the theoretical literature that considers international conflict as a repeated phenomenon. Students of international politics have long recognized the importance of repeated conflicts among nations; this is most apparent in the extensive body of work on enduring rivalries (see, e.g., Gartzke and Simon 1999 and citations therein). While estimates of the incidence of such disputes vary, it appears that at least 50% of all post-World War II international disputes represent the second time—or more—that conflict between the states involved has occurred. That is, fully half of all international conflict was presaged by earlier conflict(s) between the same nations. Moreover, a number of studies suggest that nations that have engaged in disputes in the past will behave differently from those facing an opponent for the first time (e.g., Jervis 1976; Levy 1994; Reiter 1994, 1996).

Put simply, we have reason to believe that second (and further) disputes are different. Yet, remarkably, quantitative scholars have largely ignored this fundamental distinction, instead treating the first, second, and even the tenth dispute in a series as equivalent. The result of this oversimplification is to leave us with an incomplete, and in many respects incorrect, picture of the true relationship among democratization, economic growth and trade, institutions and alliances, and the onset of war. Thus, in the case of international conflict studies, we believe there are strong theoretical as well as methodological reasons for considering repeated events.

Models of Repeated Disputes

We reexamine the liberal peace hypotheses using Oneal and Russett's (1997) data on the relationship among economic interdependence, democracy, and peace;

[12] Beck, Katz, and Tucker point out that between 1993 and 1998, over 75 papers on the democratic peace agenda were published or presented at conferences (1998, 1274).

TABLE 2

Variance-Correction Models for Repeated Events

Variable	Cox Regression: Time to First Event	Andersen-Gill (AG)	Wei et al. (WLW)	Prentice et al. (PWP): Elapsed Time	Prentice et al. (PWP): Interevent Time
Democracy	−0.424**	−0.439**	−0.438**	0.162	0.099
	(0.125)	(0.123)	(0.123)	(0.103)	(0.075)
Growth	−2.202	−3.227**	−3.183**	−3.776**	−3.422**
	(1.903)	(1.318)	(1.302)	(1.064)	(1.242)
Alliance	−0.450**	−0.414**	−0.409**	0.144	−0.202*
	(0.164)	(0.170)	(0.168)	(0.108)	(0.094)
Contiguity	1.054**	1.213**	1.203**	0.287**	0.618**
	(0.177)	(0.178)	(0.176)	(0.111)	(0.104)
Capability Ratio	−0.198**	−0.214**	−0.213**	0.059*	0.056*
	(0.079)	(0.082)	(0.081)	(0.029)	(0.025)
Trade	−5.348	−13.162	−13.001	5.997	0.812
	(13.737)	(13.827)	(13.717)	(6.504)	(9.604)
Wald Test	85.12**	92.92**	93.20**	34.54**	51.09**
N	17158	20448	163584	20448	20448

Note: Cell entries are coefficient estimates; standard errors are in parentheses. One asterisk indicates $p < .05$, two indicate $p < .01$ (one-tailed). See text for details.

this approach allows easy comparisons to previous results.[13] The data consist of 20,448 annual observations on 827 "politically relevant" dyads between 1950 and 1985. Following Beck, Katz, and Tucker (1998), we model the hazard of a militarized international conflict as a function of six primary covariates: a score for ***democracy*** (based on the Polity III data), the level of ***economic growth***, the presence of an ***alliance*** in the dyad, the two nations' ***contiguity***, their military ***capability ratio***, and the extent of bilateral ***trade*** in the dyad. Liberal theory suggests that all variables except contiguity ought to decrease the hazard of a dispute, while contiguity should increase it. In addition, we address the issue of repeated conflicts through the application of the variance correction models outlined above; these estimates are presented in Table 2.

For a point of comparison on the question of multiple events, we begin with a model that considers only "first events." This model uses data only on the

[13] Beck, Katz, and Tucker (1998) also use these data; like them, we limit our analysis to observations that are not continuations of conflicts. See Oneal and Russett (1997) for details of the variables and coding. Other studies that have used these data include Gartzke (1998, 2000), Maoz and Russett (1993), Oneal et al. (1996), Oneal and Russett (1997, 1999), Reed (2000), Russett (1993), and Russett, Oneal, and Davis (1998).

time until each dyad's first post-War dispute and thus implicitly presumes that the first event is representative of all events, a questionable assumption here as in most situations.[14] In addition, omitting second and subsequent disputes results in a loss of data and information, with resulting efficiency losses (Barai and Teoh 1997). As seen in Table 2, the first-events model also tends to underestimate covariate effects relative to the others; these differences are most clearly reflected in the variable for growth, where the first-events model is the only one that fails to find a significant effect for that variable. Our results are thus consistent with Beck, Katz, and Tucker (1998), who recommend against considering only first events in such data.

The second column of Table 2 presents the simplest repeated events model, that of Andersen and Gill; because it offers the simplest approach, the AG model provides a useful baseline for comparison with the other models. The AG gives the same parameter estimates as a standard Cox model, but it estimates robust standard errors clustered by dyad to account for repeated measures. Here, the robust standard errors are larger than the naive estimates (not shown), though the differences are not great. Results of the AG model are also similar to those of Beck, Katz, and Tucker's (1998) grouped duration model: all variables save that for trade are statistically significant and in the expected directions, with, for instance, a one-unit increase in the democracy variable corresponding to a $[(exp(-0.439) - 1) \times 100 =]$ 36% decrease in the hazard of conflict at any given time.

Results for the marginal model of Wei, Lin, and Weissfeld (WLW) are presented in column three of Table 2. Recall that the WLW model permits separate baseline hazards for each event number but also allows dyads to be "at risk" for the second, third, etc. dispute from the beginning of the observation period.[15] Here, the AG and WLW results are strikingly similar, likely due to the fact that both models use the same timescale (that is, time from entry). This similarity also seems to suggest that little is to be gained, in this example, from allowing baseline hazards to vary by event. Note, though, that the creation of the "risk set" in the WLW model makes interpretation of the WLW coefficients themselves difficult (Cook and Lawless 1997) since they are in fact a weighted average of the effects across each of the dispute numbers. Moreover, to the extent that dyadic conflict is a necessarily sequential process, it is likely that the results of the WLW model do not accurately reflect the means by which international disputes arise.

[14] The exception is, of course, when one is intrinsically interested in modeling the time to a first event. For example, Box-Steffensmeier (1996) studies the deterrence effect of a war chest on the time until entry of the first high-quality challenger in an incumbent's reelection race.

[15] Graphical tests of the proportional hazards assumption (e.g., Grambsch and Therneau 1994), not shown here, suggest that the assumption is generally valid when the data are appropriately modeled as a repeated events problem; however, it is also clear that hazards grow less proportional in the higher strata as the number of dyads experiencing a high number of events decreases.

The final two columns of Table 2 present results for the PWP models; the first models elapsed time from entry, while the second models the time from the previous event. In both instances, dyads are not considered "at risk" for a kth conflict until after the occurrence of the $(k-1)$th dispute. While there are a few similarities, the PWP models also yield results that are quite different from those already discussed: for example, the democracy and capability ratio variables are now positive but only the capability ratio is statistically significant, while the effect for alliances is significantly smaller than in prior models. In nearly all cases, the effects of the variables are estimated to be smaller than in either the AG or WLW specifications.

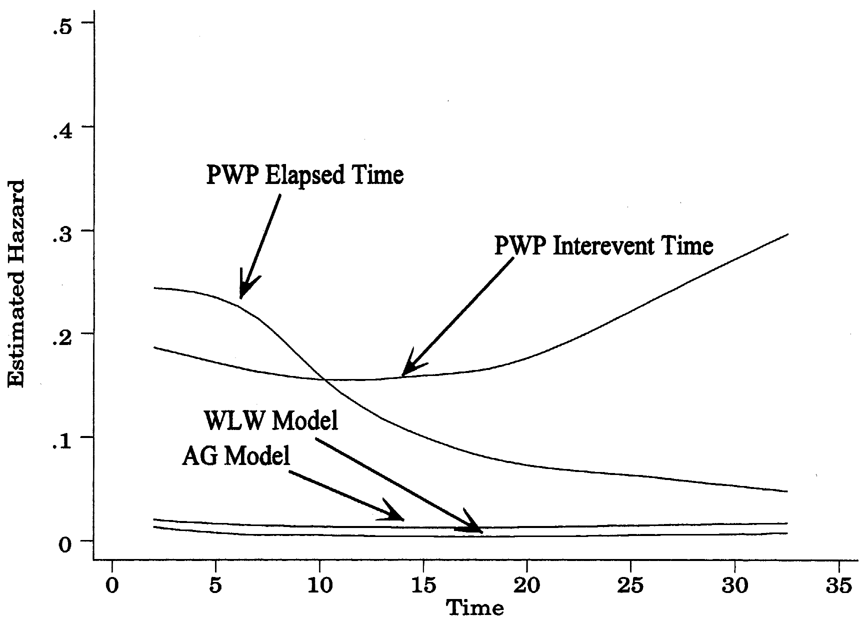
The differences across models highlight the fact that whether one chooses to model repeated events, and if so, how that modeling is done, is vitally important to one's conclusions. A key reason for the differences is the specification of the risk set for a particular event. As noted by Kelly and Lim (2000), among others, the WLW model's "semi-restricted" risk set often results in significant "carry-over" effects, analogous to sample selection. If a particular variable has a negative influence on the hazard, "then treated subjects will have fewer events in the same period. A semi-restricted risk set includes all subjects in each event stratum, and so with each consecutive event the number of treated subjects with a censored observation increases. These censored observations are compared to those untreated subjects who are experiencing an event, and so exaggerate the treatment effect in the later strata" (29). Thus, scholars need to be aware of the different modeling assumptions made by each of the various approaches, in particular, whether the ordering of events is preserved or not. Because of the conditioning assumptions—the choice of different risk sets—care needs to be exercised in the interpretation of the different models. That is, the interpretation of the parameter estimates is different across the models. For example, in the PWP model, the analyst must recognize the contingent influence of the independent variables on the second and higher order events: Pr(first failure|X), Pr(second failure|first failure, X), Pr(third failure|first failure, second failure, X), etc. In contrast, the AG model is simply Pr(any failure|X). This is a subtle but important point that makes cross-model comparisons difficult.

At a substantive level, the absence of a strong negative relationship between democracy and international conflict may be surprising at first; yet, this fact only serves to emphasize the importance of addressing the issue of repeated conflict. A consistent finding in international relations is that in general democracies fight less often than autocracies, but also that democracies tend to select into wars that they believe they will win (e.g., Reed 2000; Reed and Clark 2000; Reiter and Stam 1998). This suggests, among other things, that baseline hazards change from one dyadic conflict to the next, a fact that the AG model is incapable of addressing. Similarly, because democracies self-select into more winnable wars, the WLW model undoubtedly overstates the effect of democracy since there are likely to be increasingly lower proportions of democratic disputants in higher conflict strata. Only the PWP models, which both allow for discrete changes in the baseline hazard from one dispute to the next *and* impose a sequentially structured risk set, can account for both the differences in baseline hazards across conflicts and the "selection effects" over those same conflicts. The same selection phenomenon likely explains the otherwise anomalous result for the effect of relative military capabilities.

A useful way of comparing these models is to examine similarities and differences in their predicted baseline hazards. Figure 2 plots smoothed baseline hazards for the four central repeated events models in Table 2; several interesting characteristics are apparent. First, as we would expect, given the similarity of the coefficient estimates, the baseline hazards for the AG and WLW models are almost identical and highly correlated at 0.91. In addition, both baseline hazards are very low, reflecting the fact that only a small number of observations at risk in these two models experience a conflict. In contrast, the baseline hazards for the two conditional models are substantially higher and also exhibit greater variability. The effect of the decreasing size of the conditional models' risk sets over time is also apparent, both in the higher overall predicted hazards

FIGURE 2

Baseline Hazards for Repeated-Events Models



Note: Figures are lowest-smoothed estimated baseline hazards (bandwidth = 0.8); see text for details.

and in the greater variability and instability of the estimates at later time points. Moreover, the hazard for the elapsed-time model is generally decreasing, while that for the interevent-time model increases slightly. The difference is consistent with the fact that, while the incidence of conflict has, on average, declined over the period studied, the hazard of conflict from one event to the next remained relatively constant. The clear implication is that hazards for first disputes differ significantly from those for second and further conflicts.

Event Counter Corrections and Parameter Stability

In their influential article, Beck, Katz, and Tucker (1998) touch on the issue of recurring events by incorporating into their model a variable that counts the number of previous events that have occurred to that dyad. The effect of this variable is to introduce a rudimentary form of event dependence by allowing the baseline hazard to increase or decrease by a set proportion with each subsequent conflict. This approach has also been recognized by other scholars, who note that, particularly in the context of the AG model, the only way to incorporate information about repeated events into one's analysis is through the introduction of such covariates (e.g., Wei, Lin, and Weissfeld 1989).

This event counter solution to repeated events is appealing in its simplicity, and in some instances, the assumptions that underlie it (i.e., that the odds of an event increase by a factor of proportionality with each subsequent event occurrence) will be reasonable. Here, however, we have reason to believe that a simple monotonic change in the baseline hazard will not fully capture the effects of repeated events.[16] Moreover, as we note above, there are numerous reasons to believe that variable effects will differ across events. Thus, we compare the Beck, Katz, and Tucker event counter approach to the previous results and to the PWP model in which the covariate effects are explicitly allowed to vary by event number. In so doing, we can assess the relative usefulness of the Beck et al. approach and the alternative methods presented here.

Comparing the first column of Table 3 with the results in Table 2, we see that the results for the model of all conflicts—presented in column 1—bear a strong resemblance to those in column 2 of Table 2, albeit with slightly smaller coefficient estimates for most variables.[17] In both models, the effects of all variables save trade are as predicted, statistically significant and relatively large. Furthermore, the estimate for the previous events counter suggests that the ef-

[16] Beck, Katz, and Tucker also note that the event counter solution to repeated events is "primitive" (1998, 1272); nonetheless, event counters have been widely adopted by applied researchers.

[17] Because the Cox model explicitly allows for an unspecified baseline hazard, inclusion of Beck, Katz, and Tucker's (1998) peace-year splines is not necessary here. All models in Table 3 are estimated in interevent time, as in the final column of Table 2; this means that only observations that experienced an event were considered "at risk" for a subsequent event. The maximum number of separate conflicts observed in the data is seven; however, because of the relatively low number of observations experiencing five or more events, we collapsed these events into a single category.

TABLE 3

Event Counter and By-Conflict PWP Models of International Conflict

Variable	PWP with Previous Event Counter	Separate PWP Models, By Conflict: First Conflict	Second Conflict	Third Conflict	Fourth Conflict	Fifth or Higher Conflict
Democracy	−0.322**	−0.424**	−0.143	−0.145	0.397	0.931
	(0.105)	(0.125)	(0.217)	(0.426)	(0.514)	(0.580)
Growth	−3.530**	−2.202	−5.197*	−4.837*	−5.145	12.145
	(1.187)	(1.903)	(2.830)	(2.770)	(4.072)	(8.163)
Alliance	−0.303**	−0.450**	−0.272	−0.154	−0.003	−0.599
	(0.127)	(0.164)	(0.230)	(0.350)	(0.399)	(0.596)
Contiguity	0.877**	1.054**	0.906**	−0.017	1.104**	2.524**
	(0.135)	(0.177)	(0.260)	(0.350)	(0.434)	(1.046)
Capability Ratio	−0.195**	−0.198**	−0.062	0.056	−0.076	0.326
	(0.071)	(0.079)	(0.090)	(0.375)	(0.306)	(0.492)
Trade	−4.075	−5.348	10.940	13.906	−29.276	−315.841
	(10.671)	(13.737)	(17.873)	(36.329)	(35.593)	(216.778)
Previous Conflicts	0.310**	—	—	—	—	—
	(0.040)					
Wald Test	293.06**	85.12**	22.07**	3.44	8.33	34.17**
ln(L)	−2585.36	−1259.03	−450.29	−213.90	−83.18	−36.47
$N_{At\ Risk}$	20448	17158	2084	725	342	134
$N_{Conflicts}$	405	205	98	57	28	17

Note: Cell entries are coefficient estimates; robust standard errors are in parentheses. One asterisk indicates $p > .05$, two indicate $p < .01$ (one-tailed). See text for details. All models are estimated in interevent time.

fect of each additional previous dispute is to increase the baseline probability of a subsequent dispute by 36.3%.

Columns 2 through 6 of Table 3 present the results of separate models for each conflict; these models simply estimate separate coefficients on all observations at risk for the first, second, etc. conflict. Thus, as in the stratified models in Table 2, each event is allowed its own baseline hazard; here, each covariate's effect on the hazard is also allowed to vary for each subsequent conflict. Immediately apparent is the effect of the limited risk sets: estimates for each subsequent conflict are based on successively smaller numbers of observations, and standard error estimates increase almost uniformly across the columns as a result. Similarly, with the exception of the final column, the joint significance of the models also decreases with increasing numbers of conflicts, suggesting that the models do a progressively worse job of explaining the incidence of conflict as greater numbers occur. Taken together, these results are strongly consistent with the warning of earlier authors suggesting the possible efficiency losses accompanying strata-by-covariate interactions (e.g., Wei, Lin, and Weissfeld 1989).

The results are also quite interesting for what they reveal about the nature and effects of repeated conflicts. The effects of some factors (e.g., contiguity) remain relatively constant across repeated events. The effect of growth is also relatively stable, though in only one instance does its estimate attain conventional levels of statistical significance. In other instances, however, the effects of variables change dramatically over successive events. This is most true for the effects of the trade and democracy measures. Interestingly, while the effect of trade fluctuates between positive and negative, at no time are the coefficients estimated precisely enough for us to say with any confidence that the effect is other than a statistical artifact. For democracy, on the other hand, we find that the widely supported findings of a negative effect is driven entirely by its effect on dyads' first conflict; in second and subsequent conflicts, this effect largely disappears.

Perhaps most striking is the fact that after many repeated conflicts, the effect of democracy is to make further conflict even *more* likely ($p = 0.06$, one-tailed). This finding, seemingly unusual at first blush, is in fact consistent with our earlier results concerning selection effects and changing hazards. Moreover, it also comports with a dynamic model of international interaction that takes learning and beliefs into account. For example, Reiter (1994) illustrates that nations adopt a relatively simple mode of learning in their decisions regarding alliances: retain that which works, while modifying or rejecting strategies that fail. To the extent that democracies tend to win wars more often than autocracies (e.g., Bueno de Mesquita et al. 1999; Lake 1992), some democracies may, over time, find conflict to be a "winning strategy," particularly in their relationships with autocratic states (Bennett and Stam 1998). Such effects cannot be uncovered by simply allowing the baseline hazard to change monotonically with each additional conflict.[18]

Conclusions

Repeated events data are common to political science applications, as many of the phenomena in which we are interested can occur repeatedly over time. Here we have outlined a set of methods researchers may confidently use to model repeated events data. While we focus our analysis on one example for

[18] One can also plot the hazard rates for the PWP interevent time models by event number (not shown). Doing so indicates that the hazard rates do vary significantly by event number: hazards both increase and grow steeper as event numbers increase, with the highest hazard rate for events numbering five or more. This provides additional evidence against the AG model, which imposes a common baseline hazard for all events.

coherence, these methods are applicable to a diverse set of topics in political science. Studies of wars (Bennett and Stam 1996), coups (Londregan and Poole 1990), cabinet formations (Warwick 1992), presidential nominations to the executive branch (McCarty and Razaghian 1999), state constitutions (Hammons 1999), and Congressional committee assignments (Katz and Sala 1996) are all examples of the wide range of events to which these methods may be profitably applied. Moreover, the approaches we present are relatively simple to implement because they require only the accurate definition of risk sets and durations and the estimation of simple stratified Cox models, a technique available in a host of widely used software packages (e.g., SPSS®, SAS®, and Stata®).

On the basis of both statistical and theoretical criteria, we recommend the conditional interevent time model of Prentice, Williams, and Peterson (1981) for most instances of repeated events in political science. We emphasize, however, that model selection should be driven by substantive considerations specific to the question and data at hand. In particular, the nature of the means by which repeated events occur (that is, sequentially or simultaneously) and the corresponding construction of the risk set for each datum should provide the primary motivation for selecting one model over another.

Our illustration of the application of these methods, using data on international conflict, highlights the importance of accounting for temporal dependence due to repeated events. We demonstrate that the techniques we suggest, in addition to their statistical advantages, offer the potential for greater insights into the processes under study by accurately accounting for the empirical facts and theoretical understandings of repeated conflicts. Thoughtful application of models that account for repeated events raise and answer interesting questions about the means by which the democratic peace phenomenon occurs, as well as unifying empirically a number of heretofore disparate empirical findings with respect to that phenomenon.

Manuscript submitted 5 March 2001
Final manuscript received 30 January 2002

Appendix A:

Calculation of Robust Standard Errors

All of the variance-corrected models rely on "robust" standard errors to account for interdependence across repeated or otherwise heterogeneous events. The standard maximum-likelihood variance estimate is based on the negative inverse of the Hessian:

$$V = -\left(\frac{-\partial^2 \ln L}{\partial \beta^2}\right)^{-1} \quad \textbf{(A.1)}$$

A more general "robust" estimate (e.g., White 1980) is equal to:

$$V_R = V \sum_{i=1}^{N} (u_i' u_i) V \tag{A.2}$$

where u_i is the contribution of i to the scores $\partial ln \mathrm{L}/\partial \beta$, i.e., $\partial ln \mathrm{L}_i/\partial \beta$, evaluated at the estimated β. This is the familiar "sandwich" estimator of the variance (e.g., Greene 1997, 504–5).

In the case of heterogeneous data, we can account for correlation within "clusters" by summing scores first within clusters, and then across them, before correcting the variance matrix. So, for $\mathrm{N_C}$ clusters $j = \{1,2,\ldots N_C\}$, each consisting of n_j observations $i = \{1,2,\ldots n_j\}$, the "clustered" robust variance-covariance matrix is:

$$V_C = V \sum_{j=1}^{N_C} \left[\left(\sum_{i=1}^{n_j} u_{ij} \right)' \left(\sum_{i=1}^{n_j} u_{ij} \right) \right] V. \tag{A.3}$$

References

Andersen, Per Kragh, and R. D. Gill. 1982. "Cox's Regression Model for Counting Processes: A Large Sample Study." *The Annals of Statistics* 10(4): 1100–20.

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(12): 1489–1500.

Barai, U., and N. Teoh. 1997. "Multiple Statistics for Multiple Events, with Application to Repeated Infections in the Growth Factor Studies." *Statistics in Medicine* 16(8): 941–49.

Barbieri, Kathleen, and Gerald Schneider. 1999. "Globalization and Peace: Assessing the New Directions in the Study of Trade and Conflict." *Journal of Peace Research* 36(4): 387–404.

Beck, Nathaniel. 1998. "Modeling Space and Time: The Event History Approach." In *Research Strategies in the Social Sciences*, eds. E. Scarbrough and E. Tanenbaum. New York: Oxford University Press.

Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4): 1260–88.

Bennett, D. Scott, and Allan C. Stam III. 1996. "The Duration of Interstate Wars." *American Political Science Review* 90(2): 239–57.

Bennett, D. Scott, and Allan C. Stam III. 1998. "The Declining Advantages of Democracy: A Combined Model of War Outcomes and Duration." *Journal of Conflict Resolution* 42(3): 344–66.

Bienen, Henry, and Nicholas van de Walle. 1992. "A Proportional Hazard Model of Leadership Duration." *Journal of Politics* 54(3): 685–717.

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." Ph.D. diss. Ohio State University.

Box-Steffensmeier, Janet M. 1996. "A Dynamic Analysis of the Role of War Chests in Campaign Strategy." *American Journal of Political Science* 40(2): 352–71.

Box-Steffensmeier, Janet M., and Bradford Jones. 2002. "Timing and Political Change: Event History Modeling in Political Science." Typescript.

Box-Steffensmeier, Janet M., and Christopher J. W. Zorn. 2001. “Duration Models and Proportional Hazards in Political Science.” *American Journal of Political Science* 45(4): 951–67.

Bremer, Stuart. 1992. “Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War.” *Journal of Conflict Resolution* 26(3): 309–41.

Bueno de Mesquita, Bruce, and Randolph M. Siverson. 1995. “War and the Survival of Political Leaders: A Comparative Study of Regime Types and Political Accountability.” *American Political Science Review* 89(4): 841–55.

Bueno de Mesquita, Bruce, James D. Morrow, Randolph M. Siverson, and Alastair Smith. 1999. “An Institutional Explanation of the Democratic Peace.” *American Political Science Review* 93(4): 791–807.

Cederman, Lars-Erik. 2001. “Back to Kant: Reinterpreting the Democratic Peace as a Macrohistorical Learning Process.” *American Political Science Review* 95(1): 15–32.

Clayton, David. 1994. “Some Approaches to the Analysis of Recurrent Event Data.” *Statistical Methods in Medical Research* 3(3): 244–62.

Cleves, Mario. 1999. “Analysis of Multiple Failure-Time Data with Stata.” *Stata Technical Bulletin* 49: 30–39.

Cook, Richard J., and J. F. Lawless. 1997. “An Overview of Statistical Methods for Multiple Failure Time Data in Clinical Trials—Discussion.” *Statistics in Medicine* 16(8): 841–43.

Cox, David Roxbee. 1972. “Regression Models and Life Tables.” *Journal of the Royal Statistical Society*, Series B 34(2): 187–220.

Diehl, Paul F., ed. 1997. *The Politics of Global Governance: International Organizations in an Interdependent World*. Boulder: Lynne Rienner.

Diehl, Paul F. 1998. *The Dynamics of Enduring Rivalries*. Urbana: University of Illinois Press.

Diehl, Paul F. 1999. *A Road Map to War: Territorial Dimensions of International Conflict*. Nashville: Vanderbilt University Press.

Dixon, William J. 1994. “Democracy and the Peaceful Settlement of International Conflict.” *American Political Science Review* 88(1): 14–32.

Fearon, James. 1994. “Domestic Political Audiences and the Escalation of International Disputes.” *American Political Science Review* 88(3): 577–92.

Fleming, Thomas R., and David P. Harrington. 1991. *Counting Processes and Survival Analysis*. New York: Wiley.

Gao, S., and X. H. Zhou. 1997. “An Empirical Comparison of Two Semi-parametric Approaches for the Estimation of Covariate Effects from Multivariate Failure Time Data.” *Statistics in Medicine* 16(18): 2049–62.

Gartzke, Erik. 1998. “Kant We All Just Get Along? Opportunity, Willingness, and the Origins of the Democratic Peace.” *American Journal of Political Science* 42(1): 1–27.

Gartzke, Erik. 2000. “Preferences and the Democratic Peace.” *International Studies Quarterly* 44(2): 191–212.

Gartzke, Erik, and Michael W. Simon. 1999. “Hot Hand: A Critical Analysis of Enduring Rivalries.” *Journal of Politics* 61(3): 777–98.

Grambsch, Patricia M., and Terry M. Therneau. 1994. “Proportional Hazards Tests and Diagnostics Based on Weighted Residuals.” *Biometrika* 81(3): 515–26.

Greene, William H. 1997. *Econometric Analysis*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall.

Hammons, Christopher W. 1999. “Was James Madison Wrong? Rethinking the American Preference for Short, Framework-Oriented Constitutions.” *American Political Science Review* 93(4): 837–49.

Hausman, J. A. 1978. “Specification Tests in Econometrics.” *Econometrica* 46(6): 1251–71.

Hosmer, D., and Stanley Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. New York: John Wiley & Sons.

Hougaard, P. 1991. “Modeling Heterogeneity in Survival Data.” *Journal of Applied Probability* 28(3): 695–70.

Huber, Peter J. 1967. “The Behavior of Maximum Likelihood Estimates Under Non-Standard Conditions." In *Proceedings of the Fifth Annual Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, eds. Lucien M. LeCam and Jerzy Neyman. Berkeley: University of California Press.

Jervis, Robert. 1976. *Perceptions and Misperceptions in International Politics*. Princeton: Princeton University Press.

Katz, Jonathan N., and Brian R. Sala. 1996. "Careerism, Committee Assignments, and the Electoral Connection." *American Political Science Review* 90(1): 21–33.

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(2): 214–24.

Kelly, Patrick J., and Lynette L.-Y. Lim. 2000. "Survival Analysis for Recurrent Event Data: An Application to Childhood Infectious Diseases." *Statistics in Medicine* 19(1): 12–33.

King, Gary, James E. Alt, Nancy E. Burns, and Michael Laver. 1990. "A Unified Model of Cabinet Dissolution in Parliamentary Democracies." *American Journal of Political Science* 34(3): 846–71.

Kirby, Andrew. 1999. "Love Thy Neighbor? Economic Integration and Political Friction in a Period of Globalization." In *A Road Map to War: Territorial Dimensions of International Conflict*, ed. Paul F. Diehl. Nashville: Vanderbilt University Press.

Klein, J. P., and Melvin L. Moeschberger. 1997. *Survival Analysis Techniques for Censored and Truncated Data*. New York: Springer-Verlag.

Lake, David A. 1992. "Powerful Pacifists: Democratic States and War." *American Political Science Review* 86(1): 24–37.

Lancaster, Tony. 1979. "Econometric Methods for Duration of Unemployment." *Econometrica* 47(4): 939–55.

Lancaster, Tony. 1985. "Generalized Residuals and Heterogeneous Duration Models—With Applications to the Weibull Model." *Journal of Econometrics* 28(1): 155–69.

Lancaster, Tony. 1990. *The Econometric Analysis of Transition Data*. New York: Cambridge University Press.

Larsen, Ulla, and James W. Vaupel. 1993. "Hutterite Fecundability by Age and Parity: Strategies for Frailty Modeling of Event Histories." *Demography* 30(1): 81–102.

Levy, Jack S. 1988. "Domestic Politics and War." *Journal of Interdisciplinary History* 18(4): 653–73.

Levy, Jack S. 1994. "Learning, and Foreign Policy: Sweeping a Conceptual Minefield." *International Organization* 48(2): 279–312.

Lin, D. Y. 1994. "Cox Regression Analysis of Multivariate Failure Time Data: The Marginal Approach." *Statistics in Medicine* 13(21): 2233–47.

Lin, D. Y., and L. J. Wei. 1989. "The Robust Inference for the Cox Proportional Hazards Model." *Journal of the American Statistical Association* 84(408): 1074–78.

Lipschutz, K. H., and S. M. Snapinn. 1997. "Discussion of Paper by Wei and Glidden." *Statistics in Medicine* 16(8): 846–48.

Londregan, John B., and Keith T. Poole. 1990. "Poverty, the Coup Trap, and the Seizure of Executive Power." *World Politics* 42(2): 151–83.

Mansfield, Edward D. 1994. *Power, Trade, and War*. Princeton: Princeton University Press.

Maoz, Zeev. 1998. "Realist and Cultural Critiques of the Democratic Peace: A Theoretical and Empirical Re-assessment." *International Interactions* 24(1): 3–89.

Maoz, Zeev, and Nazrin Abdolai. 1989. "Regime Types and International Conflict, 1816–1976." *Journal of Conflict Resolution* 33(1): 3–36.

Maoz, Zeev, and Bruce Russett. 1993. "Normative and Structural Causes of the Democratic Peace, 1946–1986." *American Political Science Review* 87(3): 624–38.

McCarty, Nolan, and Rose Razaghian. 1999. "Advice and Consent: Senate Responses to Executive Branch Nominations." *American Journal of Political Science* 43(4): 1122–43.

Meinke, Scott. 2001. "The Dynamics of Position Change in House Voting." Ph.D. diss. Ohio State University.

Morrow, James D. 1999. "How Could Trade Affect Conflict?" *Journal of Peace Research* 36(4): 481–89.

Oakes, D. A. 1992. "Frailty Models for Multiple Event Times." In *Survival Analysis: State of the Art*, eds. J. P. Klein and P. K. Goele. Dordrecht: 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*, eds. D. Y. Lin and T. R. Fleming. New York: Springer.

Oneal, John R., Frances H. Oneal, Zeev Maoz, and Bruce Russett. 1996. "The Liberal Peace: Interdependence, Democracy and International Conflict." *Journal of Peace Research* 33(1): 11–28.

Oneal, John R., and Bruce Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependence, and Conflict, 1950–1985." *International Studies Quarterly* 41(2): 267–94.

Oneal, John R., and Bruce Russett. 1999. "Assessing the Liberal Peace with Alternative Specifications: Trade Still Reduces Conflict." *Journal of Peace Research* 36(4): 423–42.

Prentice, R. L., B. J. Williams, and A. V. Peterson. 1981. "On the Regression Analysis of Multivariate Failure Time Data." *Biometrika* 68(2): 373–79.

Ray, James Lee. 1995. *Democracy and International Conflict*. Columbia: University of South Carolina Press.

Reed, William. 2000. "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science* 44(1): 84–93.

Reed, William, and David H. Clark. 2000. "War Initiators and War Winners: The Consequences of Linking Theories of Democratic War Success." *Journal of Conflict Resolution* 44(3): 378–95.

Reiter, Dan. 1994. "Learning, Realism and Alliances: The Weight of the Shadow of the Past." *World Politics* 46(4): 490–526.

Reiter, Dan. 1996. *Crucible of Beliefs: Learning, Alliances and World Wars*. Ithaca, NY: Cornell University Press.

Reiter, Dan, and Alan Stam III. 1998. "Democracy, War Initiation and Victory." *American Political Science Review* 92(2): 377–89.

Rousseau, David L., Christopher Gelpi, Dan Reiter, and Paul K. Huth. 1996. "Assessing the Dyadic Nature of the Democratic Peace, 1918–88." *American Political Science Review* 90(3): 512–33.

Russett, Bruce. 1993. *Grasping the Democratic Peace: Principles for a Post-Cold War World*. Princeton: Princeton University Press.

Russett, Bruce, James R. Oneal, and David W. Davis. 1998. "The Third Leg of the Kantian Tripod for Peace: International Organizations and Militarized Disputes, 1950–85." *International Organization* 52(3): 441–67.

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(438): 426–35.

Schulz, Heiner, and Thomas König. 2000. "Institutional Decision-Making Efficiency in the European Union." *American Journal of Political Science* 44(4): 653–66.

Struthers, C. A., and J. D. Kalbfleisch. 1984. "Misspecified Proportional Hazard Models." *Biometrica* 73(2): 363–69.

Therneau, Terry M. 1997. "Extending the Cox Model." *Proceedings of the First Seattle Symposium in Biostatistics*. New York: Springer-Verlag.

Therneau, Terry M., and Scott A. Hamilton. 1997. "rhDNase as an Example of Recurrent Event Analysis." *Statistics in Medicine* 16(18): 2029–47.

Vaupel, J. W., K. G. Manton, and E. Stallard. 1979. "The Impact of Heterogeneity in Individual Frailty Models and the Dynamics of Mortality." *Demography* 16(3): 439–54.

Ward, Michael D., and Kristian S. Gleditsch. 1998. "Democratizing for Peace." *American Political Science Review* 92(1): 51–62.

Warwick, Paul. 1992. "Economic Trends and Government Survival in West European Parliamentary Democracies." *American Political Science Review* 86(4): 875–86.

Wei, L. J., and David V. Glidden. 1997. "An Overview of Statistical Methods for Multiple Failure Time Data in Clinical Trials." *Statistics in Medicine* 16(8): 833–39.

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(408): 1065–73.

Werner, Suzanne. 1999. "The Precarious Nature of Peace: Resolving the Issues, Enforcing the Settlement, and Renegotiating the Terms." *American Journal of Political Science* 43(3): 912–34.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity." *Econometrica* 48(4): 817–38.

Janet M. Box-Steffensmeier is associate professor of political science, Ohio State University, Columbus, OH 43210-1373.

Christopher Zorn is associate professor of political science, Emory University, Atlanta, GA 30322.