Advance Access publication September 25, 2016

Political Analysis (2016) 24:457–477
doi:10.1093/pan/mpw025

Surviving Phases: Introducing Multistate Survival Models

Shawna K. Metzger
University Scholars Programme 18 College Avenue East #02-03
Cinnamon West Learn Lobe Singapore 138593 Singapore
e-mail: smetzger@nus.edu.sg (corresponding author)

Benjamin T. Jones
Department of Political Science P.O. Box 1848
University of Mississippi University, MS 38677, USA
e-mail: btjones1@olemiss.edu

Edited by Jonathan Katz

Many political processes consist of a series of theoretically meaningful transitions across discrete phases that occur through time. Yet political scientists are often theoretically interested in studying not just individual transitions between phases, but also the duration that subjects spend within phases, as well as the effect of covariates on subjects' trajectories through the process's multiple phases. We introduce the multistate survival model to political scientists, which is capable of modeling precisely this type of situation. The model is appealing because of its ability to accommodate multiple forms of causal complexity that unfold over time. In particular, we highlight three attractive features of multistate models: transition-specific baseline hazards, transition-specific covariate effects, and the ability to estimate transition probabilities. We provide two applications to illustrate these features.

The notion of "change over time" is a prominent part of many political science research agendas. What influences democratization (Epstein et al. 2006; Maeda 2010)? What factors influence whether U.S. District Court judges leave the District bench (Hansford, Savchak, and Songer 2010), or whether U.S. House incumbents stay in office (Box-Steffensmeier and Jones 2004), or whether legislators stay in the European Parliament (EP) (Daniel 2015)? How do countries resolve their territorial disputes (Huth and Allee 2002; Jones and Metzger forthcoming)? In all these examples, the passage of time provides an opportunity for some outcome to exhibit variation, particularly within a specific case. Researchers then exploit this variation to help evaluate their theories, often using longitudinal and time-series methods.

However, each of these questions pertains to only a narrow element of a much larger process. For instance, once democratization occurs, how long will the resulting democratic regime survive? If a democracy backslides into nondemocracy, how long until democratization occurs again? For District Court judges, U.S. House incumbents, and EP legislators: After officeholders leave their current positions, where do they go to next? How long do they stay in these new positions? Do they eventually return to the District Court, U.S. House, or EP? For territorial disputes, do disputes take multiple settlement attempts to resolve? Are there negotiations or militarized behavior? How long do these settlement attempts last? Political scientists have a range of methodological tools to help address each of these individual questions, but our ability to address the larger process at play remains limited. Yet, we should fundamentally care about modeling as much of the process as possible, because our theoretical arguments are motivated by a desire to better understand these larger processes. Further, our theories are often capable of generating hypotheses beyond the first

Authors' note: The authors' names appear in reverse alphabetical order. This paper was presented at the 2015 Visions in Methodology conference. We thank Jan Box-Steffensmeier, Sarah Cormack-Patton, Luke Keele, Diana O'Brien, Steve Oliver, and Doug Rice for feedback on earlier drafts. We bear sole responsibility for any remaining errors and shortcomings. All analyses are performed using *R 3.3.1*. Replication material is available at http://dx.doi.org/10.7910/DVN/OZ7YZ1. Supplementary materials for this article are available on the *Political Analysis* Web site.

event a subject experiences within a process. Focusing only on this first event deprives us of an opportunity to test the additional implications of our theories.

Beyond the theoretical motivations, there are methodological reasons we should be concerned with modeling an entire process. In brief, failure to do so can result in inaccurate, misleading inferences about a covariate's effect. The methodological problem's root is the nature of the process itself. Specifically, political processes are causally complex (Braumoeller 2003). They are composed of different stages through which subjects can transition[1]—for example, countries transition between democratic versus nondemocratic regimes. In this situation, causal complexity manifests in two primary ways. The first way is conditional covariate effects: a covariate's effect may differ depending on the stage a subject currently occupies. Second, complexity can arise from the varied ways in which subjects may transition through a process's stages. Subjects may exhibit recursive transitions, in which they re-enter the same stage again after previously exiting it, such as a former democracy returning to democracy following an authoritarian backslide. Subjects may also experience sequential transitions, in which they can experience a particular transition only *after* experiencing a previous event. For instance, a country must first transition into civil conflict before it can transition into a post-civil conflict peace. It is possible for a process to have both recursive and sequential transitions. We show that *even if* we are only interested in one specific transition within a process, our inferences about x's effect may be biased if researchers eschew modeling these facets of causal complexity.

Most of our standard empirical tests are ill equipped to simultaneously handle conditional covariate effects, different transition events, different transition sequences, and transitions' timing.[2] Regime-switching models, as a class of models, are superb at modeling transition-specific covariate effects and can handle very complicated transition sequences and different transition events, but usually do not focus on how processes evolve over time. Standard survival models can speak to questions of time, but are less adroit at handling many transition-specific covariate effects. They also cannot handle complicated transition sequences like recursiveness, because they only focus on a single transition event. More complicated survival models, like competing risks (CR) and models for repeated events, can handle more than one transition event, but focus on either transitions out of a single stage only (CR) or the same event occurring repeatedly (repeated events) (Box-Steffensmeier and Zorn 2002). Other, more complex transition sequences are beyond the respective models' reach. A similar truth holds for logit/probit and their respective multinomial variants. These models have a limited ability to capture more complex transition sequences, and they can be unwieldy when handling many different events and many transition-specific covariate effects. However, the models can accommodate time by adding time counters as regressors (Beck, Katz, and Tucker 1998).[3]

How, then, should we investigate claims about causally complex processes that occur across time? We make the novel suggestion that survival models *are* capable of investigating such claims. We introduce the multistate survival model to political science (Therneau and Grambsch 2000). Multistate survival models are an extension of the familiar Cox model, and allow researchers to model the trajectory of an individual subject across a series of transitions between various stages. For instance, they permit researchers to model the regime transition process suggested above: the time until democratization occurs, the time until democracy breaks down, and the time until democracy is restored again, all within a single model. Multistate models are capable of modeling many more complex political processes, including those with recursive and sequential transition sequences. As a result, multistate models greatly expand researchers' ability to model *complete* processes of theoretical interest.

[1]Formally, transitions describe instances in which a subject moves *from* one stage *into* another one. Notice how they consist of directed "from-to" stage pairings. Transitions and events are synonyms, from this perspective.
[2]For some intriguing partial exceptions, see Chiba, Metternich, and Ward (2015).
[3]The use of time counters, however, introduces additional complications with respect to determining the nature of duration dependence, specifically, whether to apply a smoothing function and what precise form this function ought to take (see Box-Steffensmeier and Jones 2004, 88; Carter and Signorino 2010).

We argue that multistate models should be attractive to political scientists for three specific reasons. First, they allow different events within a process to have different underlying rates of occurrence, by permitting each event to have its own baseline hazard rate. Work on parametric duration models makes strikingly clear the perils of improper assumptions about a single event's baseline hazard rate—the resulting model estimates can be biased (Box-Steffensmeier and Jones 2004). The bias persists when we expand our analysis to multiple events. Multistate models permit *different* events to occur at differing rates. This extends work on repeated events, which principally focuses on the same event occurring on multiple occasions.

Second, multistate models allow covariate effects to differ, also based on the transition in question. The majority of hypothesis testing in political science pertains to x's effect on an outcome of interest, y. The same current practices acknowledge that x's effect on y could be conditioned on some third factor, z. Acknowledging conditional effects is important to accurately recovering an estimate of x's effect on y for hypothesis testing. Failing to model conditional effects can result in an estimate that averages x's effect across all the different conditioning scenarios, potentially yielding a biased estimate. Multistate models actively encourage researchers to assume x's effect is conditional on the transition in question. By doing so, researchers can then easily assess whether x's effect does indeed differ across transitions. It becomes a matter to resolve empirically, rather than one resolved by theory alone.

Different baseline hazards and conditional covariate effects can be useful in their own right, but their true power is realized through multistate models' third and final attractive property: the models' adeptness at calculating transition probabilities. These express the probability that subject i will be in Stage B at time t, given that i began the analysis in Stage A at time s. For a causally complex process, there are many ways that i could move from Stage A to Stage B, in the time frame s to t. Subject i could have transitioned directly from A to B, but it may have also taken an indirect route—for instance, transitioning from A to C to B, or A to C to A to B. Transition probabilities aggregate over all the possible sequences that begin with subject i in State A at time s and end with i in B by time t. In so doing, they incorporate information about the unique underlying rates of each transition. More importantly, transition probabilities also account for a covariate's potentially unique effect on each transition.

Therefore, transition probabilities allow us to evaluate a covariate's *net* effect within the entire process, as opposed to evaluating its direct effect on only one of the process's transitions. If the covariate's effect is indeed conditional on transition, the end result will be x's net effect on A$\rightarrow$B being very different from what its direct effect would imply (e.g., Jones and Metzger forthcoming). This added insight opens up new doors for understanding the process of interest and how x plays a role, and also provides new avenues for assessing our arguments.

Our discussion proceeds in four parts. We begin by introducing the model itself. Second, we highlight the model's attractive features. Third, we provide some illustrative applications to show how the model works and the types of inferences we can draw from it. The fourth and final section concludes.

1 What Are Multistate Models?

A multistate survival model is an econometric estimator capable of modeling a duration process composed of multiple stages (Therneau and Grambsch 2000). Stages are defined based on the failure events that a subject is at risk of experiencing.[4] These failure events therefore amount to transitions between stages. Our interest is often in understanding (a) when transitions between stages will occur, (b) the probability of the transitions, and (c) what covariates increase or decrease these transition probabilities. Multistate models have been used to explore causes of death among Norwegian citizens (Vollset, Tverdal, and Gjessing 2006), bone marrow recipients' health (Putter, Fiocco, and Geskus 2007, 2417–22), and individuals' cohabitation patterns (Mills 2004). However, their use in political science has been rare.[5]

[4]For more on understanding stages as opposed to events, see Online Appendix B.
[5]Exceptions include Jones (2013), Jones and Metzger (forthcoming), and Mattiacci and Jones (2016).

Despite their infrequent use in political science, multistate models are built from methodological pieces that *are* familiar to political scientists. Accordingly, we introduce multistate models from the ground up using these pieces—we begin with simple survival models, move to competing risks models, and then finally arrive at multistate models.

1.1 *Basic Survival Models*

There are situations in which researchers are interested in the occurrence of an event, with a specific interest in how long it takes subject i to experience this event. Take the legislative careers of U.S. House representatives as a running example (Box-Steffensmeier and Jones 2004)—any incumbent House representative's tenure will end, eventually. Political scientists have an interest in exploring how long an incumbent remains in office before leaving. Survival models, also known as duration models or event history models, are well suited to answering questions of this form. They are interested in modeling the event's hazard rate, which (loosely) expresses the probability of i experiencing the event in t, contingent upon i still being at risk of experiencing the event in t (Allison 1984).[6] The hazard itself is unobserved, but we suppose it is a function of i's "time at risk" for the event, permitting us to model the hazard using the observed duration. This duration is typically defined as how much time passes between the first period in which i could have experienced the event and the period in which i did experience the event. If we begin counting from 0 in the first period, t represents the total amount of time a subject has been at risk of experiencing the event.

Semiparametric survival models do not make any parametric assumptions about the baseline hazard rate, where the baseline hazard rate expresses the event's hazard rate when the covariates are equal to zero.[7] Instead, semiparametric models parameterize only the covariates' relationship with the hazard and estimate these coefficient values using partial-likelihood methods (Box-Steffensmeier and Jones 2004, chap. 4). The Cox proportional hazards model is the quintessential semiparametric survival model and is the building block for our more advanced multistate model.[8] The Cox's hazard rate is expressed as (Box-Steffensmeier and Jones 2004, 48):

$$\alpha(t) = \alpha_0(t)e^{\beta' X}, \tag{1}$$

where $\alpha(t)$ is the event's hazard of occurring in time t, $\alpha_0(t)$ represents the baseline hazard rate in t, X is a vector of covariates, β is the vector of coefficients, and $'$ is the vector's transpose.

1.2 *CR Models*

Basic survival models assume that subjects are only at risk of experiencing one event. What happens if subjects are at risk of experiencing *multiple* events? Competing risks (CR) models can accommodate this situation. CR models are a special type of multistate model, making the former useful for beginning to explain the latter's features.

Formally, CR models extend standard Cox survival models; they still model how long it takes for subject i to experience an event. The key difference between the two models is that subject i is at risk of experiencing *two or more such events* in a CR scenario. These multiple events are referred to as "transitions" in the parlance of multistate models. The implication is that there are multiple ways in which i's time at risk can end. Figure 1b visually depicts the CR scenario, while Fig. 1a depicts the standard Cox scenario. In both panels, Stage 1 represents a subject's initial time at risk. In Fig. 1a, there is only one way for a subject to exit Stage 1—a transition into Stage 2. By contrast, in Fig.

[6]In truth, hazards are not unconditional probabilities. They represent the instantaneous risk of failure. A hazard rate can be larger than one, for instance (Cleves et al. 2010, 7–8). For discrete time durations, they are conditional probabilities (Aalen, Borgan, and Gjessing 2008, 5–6).

[7]Parametric survival models assume a specific functional form for both the covariates' relationships with the hazard and the baseline hazard rate (Box-Steffensmeier and Jones 2004, chap. 3).

[8]Multistate models can also be estimated nonparametrically and parametrically; see Online Appendix J for further discussion.

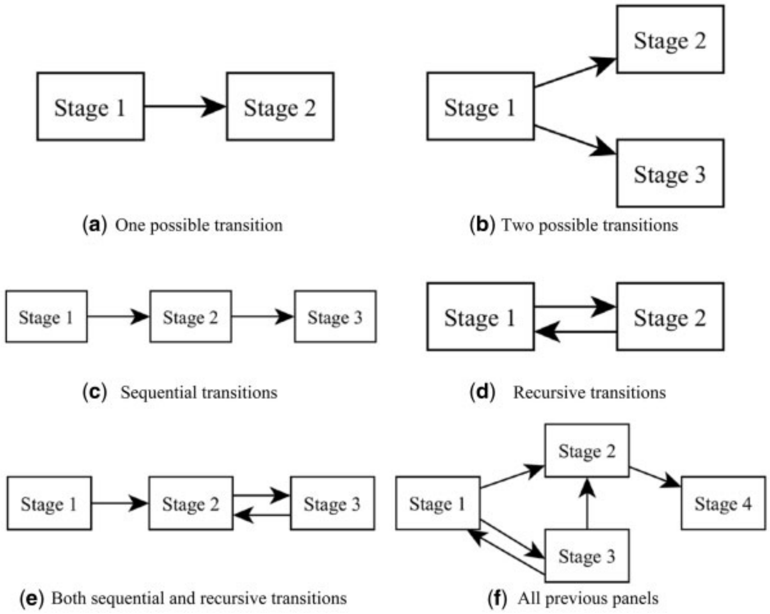

Fig. 1 Illustrative processes. (a) One possible transition, (b) two possible transitions, (c) sequential transitions, (d) recursive transitions, (e) both sequential and recursive transitions, and (f) all previous panels. *Notes*. Arrows denote possible transitions for each panel's process.

1b, there are two possible ways for a subject to exit Stage 1—a transition into Stage 2 or a transition into Stage 3.

Recognizing there are multiple transitions out of a specific stage is important. If we pool all the stages' exiting transitions together, we are implicitly assuming that each transition's data-generating process (DGP) is identical. A covariate would therefore have the same effect on every transition. If the transitions have different DGPs, though, a pooled-transition model would produce biased estimates. The estimates would equal the covariate's average effect across all the transitions. To return to Fig. 1b, x's effect could be -1.5 for transitions from Stage 1 to Stage 2, but $+1.7$ for transitions from Stage 1 to Stage 3. If we erroneously pooled all of Stage 1's exiting transitions together, x's estimated effect would be around zero.[9] As another example, our second application shows that more economically developed countries have a decreased probability of transitioning to nondemocracy via coups, but the same conditions have no effect on the country's probability of transitioning to nondemocracy via self-coup.

In a classic CR setup, all observations (1) begin in the same stage, (2) are simultaneously at risk of experiencing two or more transitions, and (3) after experiencing one transition, are no longer at risk of experiencing *any* transitions (Box-Steffensmeier and Jones 2004).[10] A CR setup for our earlier U.S. House example gives us a stage diagram identical to Fig. 1b, only with two additional stages on the right. Specifically, we are no longer interested in simply *whether* the incumbent leaves office, but *how* s/he leaves office. Several possibilities exist (Box-Steffensmeier and Jones 2004, 169–72). The representative could: (1) be defeated in a primary election, (2) be defeated in the general election, (3) choose to retire, or (4) seek alternative office (e.g., Senate, gubernatorial, cabinet appointment). Once a representative has experienced one of these four events, s/he has "exited the risk set," and is no longer at risk of experiencing the other three. A representative who retires, for instance, would no longer be at risk of exiting the House via electoral defeat.

[9]We show this using simulations in Online Appendix E.I.

[10]A classic CR setup also assumes that the different events are independent of one another. Multistate models make the same assumption. For models that explore dependent competing risks, see Gordon (2002) and Fukumoto (2009).

A classic CR model recognizes the different possible transitions out of a risk set and estimates an equation for each transition. For semiparametric survival models, Cox models (Cox 1972) and Fine-Gray subhazard models (Fine and Gray 1999) are the most common estimators.[11] CR's major modeling strength is that it permits a covariate's effect to vary across transitions. Doing so guards against the biased estimates that would potentially result from pooling all the transitions. For example, a covariate that *appreciably* increases the probability of primary election defeat may only *slightly* increase the probability of general election defeat. A CR model would detect this difference, whereas a standard Cox model with pooled transitions would not.

Yet, a classic CR model is limited in its ability to model more complex situations. It is primarily focused on transitions out of the starting stage. The model does not address what happens to subjects after they transition out of Stage 1 and into Stage 2 or Stage 3 (Fig. 1b). We can imagine situations in which this information would be substantively useful. Additionally, CR models cannot handle situations in which, after subject i experiences one event, i is still at risk of experiencing other events. An ongoing territorial dispute that militarizes, for instance, may still experience peaceful negotiations (Jones and Metzger forthcoming).

1.3 *Multistate Models*

Multistate models take a holistic approach to a process. They "extend the analysis to what happens after the first [transition] event," allowing researchers to model how a subject moves through several stages (Putter, Fiocco, and Geskus 2007, 2390). This means that multistate models permit *multiple* risk sets, as opposed to the single risk set inherent in CR models. Consequently, they are sufficiently flexible to model any number of possible transition sequences using a single framework. They can capture situations in which transition events occur sequentially, recursively, or any combination thereof (Therneau and Grambsch 2000; Putter, Fiocco, and Geskus 2007). In short, we can use multistate models to estimate a process with *any* of the stage structures depicted in Fig. 1, whereas classic CR models can only handle the first two panels, and a standard Cox model could only handle the first.

The premise of multistate models is simple: a subject transitioning *out* of one stage must be transitioning *into* another. Rather than dropping the subject after this first transition (like classic CR does), multistate models consider what new transitions the subject is at risk of experiencing from the new stage.[12] For a concrete example, take a complex process like Fig. 1f. A subject in Stage 1 is at risk of experiencing two transitions: one into Stage 2 and one into Stage 3. A subject that transitions into Stage 2 would then be at risk of experiencing one transition, into Stage 4. For our U.S. House example, a multistate model of politicians' legislative careers would allow us to recognize that former House incumbents sometimes return to office in nonconsecutive terms.[13]

Multistate models can be estimated as stratified Cox models, which differ from a standard Cox model in two key respects.[14] First, the underlying *baseline hazard is stratified* for every transition within the process. A separate baseline hazard, $\alpha_{q_0}(t)$, is estimated for each possible transition q, where t continues to refer to the duration.[15] By contrast, standard Cox models only estimate one baseline hazard, $\alpha_0(t)$; there are no q subscripts.

Second, multistate models include *transition-specific covariates*, X_q, where q, as above, indexes every possible transition in a process. Including X_q allows each variable to have a unique effect,

[11]All parametric models can also handle competing risks. Semiparametric approaches are simply more common because of their more flexible assumption regarding the baseline hazard. See Online Appendix J for a detailed discussion.

[12]More formally, multistate models use stages to define different risk sets, since subject i's current stage determines which transitions i is at risk of experiencing. This is how multistate models are comprised of multiple risk sets. For this reason, some describe certain multistate models as a number of nested competing risks models (e.g., Beyersmann, Allignol, and Schumacher 2011, 28–29; Geskus 2015, 216).

[13]Samuel Cox (OH-D/NY-D, 1857–65, 1869–91) and Lindley Beckworth (TX-D, 1939–53, 1957–67) are two such examples.

[14]Classic CR models are different from standard Cox models in the same two ways, since classic CR is just a specific example of a (simple) multistate model.

[15]However, these baseline hazards need not be modeled separately, should theory or statistical tests indicate that two or more of them are equal. We discuss this further in a later section.

depending on the transition in question. For example, x may decrease the risk of transitioning from Stage 1 to Stage 2, but increase the risk of transitioning from Stage 2 to Stage 3. Empirically allowing for such differences is important, because it permits researchers to test for transition-specific covariate effects and protects against possible biased estimates from improperly pooling transitions. By contrast, a standard Cox model can only accommodate one transition, making the transition-specific designation irrelevant.

Thus, a multistate model's hazard rate is (Wreede, Fiocco, and Putter 2010, 262–63):

$$\alpha_q(t) = \alpha_{q_0}(t)e^{\beta' X_q}. \tag{2}$$

Given this hazard rate, cumulative transition hazards may be estimated as

$$A_q(t) = \int_0^t \alpha_q(u)\mathrm{d}u. \tag{3}$$

The cumulative transition hazards can be aggregated into an $S \times S$ matrix, $\mathbf{A}(t)$, where S is the number of possible stages within the multistate model and u denotes all times at which we observe any transition within some time interval $(s,t]$.[16] For the U.S. House legislator example, S would be equal to 5—(1) in office, (2) primary election defeat, (3) general election defeat, (4) retirement, and (5) assuming an alternative office.

Cumulative transition hazards are relevant because they permit us to calculate transition probabilities. Specifically, we can estimate a transition probability matrix, $P(s,t)$, as

$$P(s, t) = \prod_{u \in (s,t]} (\mathbf{I} + \Delta\mathbf{A}(u)), \tag{4}$$

where $(s,t]$ denotes the time interval. In Online Appendix A, we walk through a detailed example of how to calculate these quantities using our U.S. House example. Each row of P will sum to one because each subject must be in one of the S stages. The individual elements of $P(s,t)$ represent the probability of transitioning from each stage to every other stage within the time interval $(s,t]$.[17] For example, element $P_{1,2}(s,t)$ would denote the probability of a legislator transitioning from Stage 1 (in office) in time period s to Stage 2 (defeat in the primaries) by time period t. Importantly, these transition probabilities will vary over time, because the hazards on which they are based vary as well. As a result, holding all else constant, the probability of a particular transition occurring may be substantially different at time t than it is at $t+5$.[18,19]

Like basic survival models and CR models, t can be defined in one of two ways in a multistate model. In an *elapsed-time* formulation, t begins counting from 0 once a subject enters the data set and never resets back to 0.[20] Thus, time indicates the total amount of time that a subject has been in the process. By contrast, in a *gap-time* formulation, t begins counting from 0 but resets when the subject experiences a transition. Here, time indicates how long the subject has spent in the stage it currently occupies. Which formulation is appropriate depends on the substantive application and the theoretical underpinnings of the process under examination.[21] The key is whether it makes more theoretical sense to say that a subject's risk of transitioning is dependent upon (a) how long it has

[16]The stages are numbered purely for organizational purposes.

[17]Transitions that are impossible, either realistically or theoretically, are fixed at zero.

[18]This implies that the process is time-inhomogeneous in nature: the transition hazards' value (and therefore the transition probabilities' values) can change across t (Hougaard 2000, 143; Beyersmann, Allignol, and Schumacher 2011, 172). By contrast, the transition hazards for a time-homogenous process are constant across time (e.g., exponential survival model). The time-inhomogeneity property is appealing for the same reason that survival models are appealing—we believe that time matters, in that the probability of observing our event of interest changes as time passes.

[19]This section's discussion makes clear that multistate models are an example of a regime-switching model. For details, see Online Appendix F. Multistate models and empirical models of strategic interactions also have some similarities, but have important differences; see Online Appendix K for details.

[20]Elapsed time is synonymously referred to as "total time" and "clock time" by some.

[21]The only ramification for multistate models is computing transition probabilities. Elapsed-time formulations make the model Markovian in nature, allowing us to estimate analytic transition probabilities. Gap-time formulations make the model semi-Markovian, which makes analytic estimation inappropriate. We must use simulation to estimate transition probabilities (see Online Appendix H). For more on Markovian versus semi-Markovian, see Online Appendix C.

been at risk, in general (elapsed time), or (b) how long it has spent in its current stage (gap time). For instance, if the process in question is a territorial dispute, an elapsed-time formulation might be more appropriate. The total amount of time a dispute has been ongoing is likely to be more relevant for determining the risk of subsequent transitions than how long a particular stage of the dispute has endured. Conversely, if we were to consider judicial careers, a gap-time formulation is potentially more appropriate. The amount of time a judge has served on an appellate court may be more important for determining her risk of being nominated to the Supreme Court than the total amount of time that she has served on the bench.

2 Why Are Multistate Models Useful?

Multistate models share many of the beneficial properties of Cox models and other Cox extensions familiar to political scientists, such as CR models. However, a multistate modeling framework extends these familiar tools and introduces three innovations to political scientists' use of survival models.[22] These innovations collectively give multistate models their flexibility and allow them to accommodate causally complex political processes unfolding across time. The same innovations also allow researchers to avoid biased estimates from improperly assuming x's effect is equivalent across transitions. We consider each of these innovations—transition-specific baseline hazards, transition-specific covariate effects, and transition probabilities—in turn.

2.1 *Transition-Specific Baseline Hazards*

One of the primary advantages of using multistate models is the flexibility they afford researchers to model any number and sequence of events that are deemed to be theoretically or substantively meaningful. In order to accommodate these varied transition event sequences, a multistate modeling strategy allows the researcher to stratify the baseline hazard for each of the different transitions in the model. In practice, this means researchers may allow the underlying rate at which one type of event occurs to vary from the underlying rate at which an event of a different type occurs. For instance, the underlying rate at which incumbents leave the House due to general election defeats is likely different from the rate at which they leave due to retirement, which may differ from the rate at which former incumbents regain office. This type of stratification is familiar to researchers that employ CR models (Crowder 2012; Box-Steffensmeier and Jones 2004, chap. 10), depicted visually in Fig. 1b. There, the baseline hazard is stratified by each possible transition to reflect the possibility that the underlying rate at which transitions occur from Stage 1 to Stage 2 may vary from the rate at which transitions occur from Stage 1 to Stage 3. Analogously, using Fig. 1e, it may be that the rate at which subjects transition into the *same* stage (Stage 2) may vary, depending on whether the subject is currently in Stage 1 or Stage 3. For instance, the underlying rate at which peaceful protests escalate to civil war is likely to be different from the rate at which violent protests escalate to civil war.

Stratification is also a prevalent strategy when dealing with repeated events, depicted visually in Fig. 1c. The underlying rate at which subjects experience the first event may differ from the rate at which subjects experience a second or third event, if, for example, experiencing a first event makes subjects more likely to experience subsequent events (Box-Steffensmeier and Zorn 2002; Box-Steffensmeier, Linn, and Smidt 2014). Stratification in the context of repeated events also underscores a related issue, which is that not all subjects are necessarily at risk for all transitions simultaneously. Rather, some transitions may only occur sequentially, such that subjects only become at risk for a particular transition *after* experiencing a previous event. In the context of repeated events of the same type, this is straightforward. A subject is only at risk of experiencing a second event after it has experienced a first event (e.g., conditional models of repeated events; see Prentice, Williams, and Peterson 1981). This same principle can generalize to situations in which different events occur in a sequence. Figure 1c illustrates this more general situation, where all

[22]Some innovations build upon current practices that, though used often, have not been applied more generally.

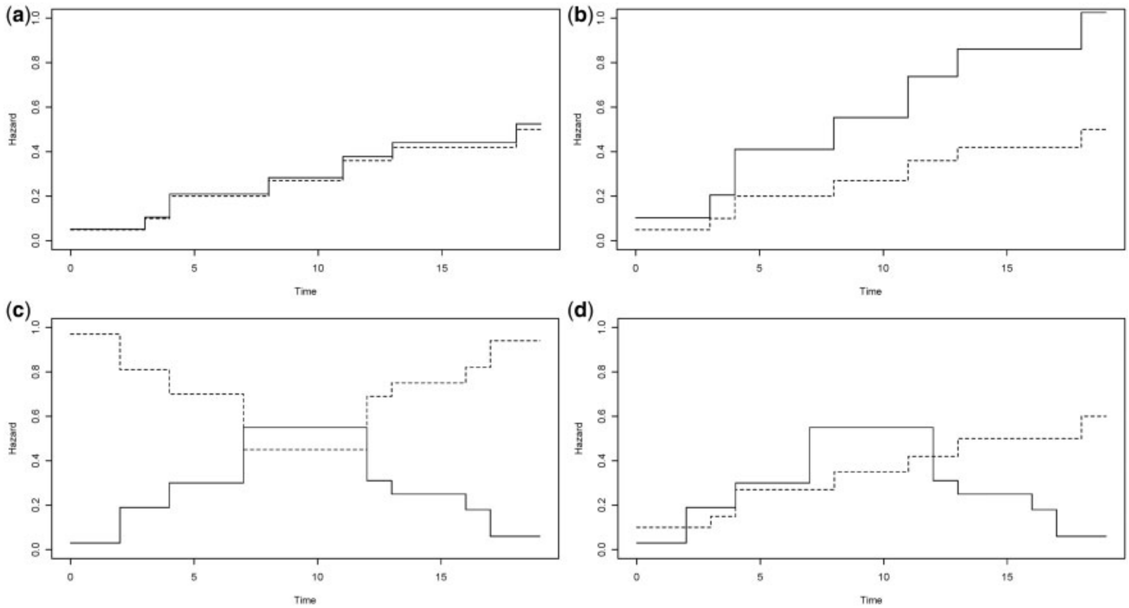

Fig. 2 Baseline hazard comparisons. (a) Proportional, not statistically different, (b) proportional, statistically different, (c) not proportional, not statistically different, (d) not proportional, statistically different.

subjects begin in Stage 1 and are at risk of a transition to Stage 2. However, subjects *only* become at risk of a transition to Stage 3 after they transition into Stage 2.

By employing precisely this stratification approach, multistate models are capable of modeling complex event sequences that may combine the elements of CR and repeated events models we discussed above. As Fig. 1d–f reflect, permitting each transition to have a different baseline hazard allows researchers to differentiate between many different types of event sequences that may arise in their data. The determination of the "appropriate" number of transitions and their sequence is largely a matter of theoretical concern, depending on the particular situation to which multistate models are being applied.

Nevertheless, we can assess whether stratification is statistically appropriate. There are two possible reasons to stratify in a Cox model. First, stratifying is a possible solution to violating the proportional hazards assumption (Box-Steffensmeier and Jones 2004, 132). It allows the baseline hazard to vary for subgroups of the data, rather than assuming the baseline hazard is the same across subgroups. Figure 2a and 2b depict two baseline hazards that are proportional over time, whereas Fig. 2c and 2d depict scenarios in which the baseline hazards are clearly not proportional. Stratification permits these baseline hazard estimates to vary. Second, stratification may be additionally appropriate if the baseline hazards are proportional but statistically different from one another, as in Fig. 2b. Testing whether stratification is appropriate resembles our usual tests for violations of the proportional hazards assumption, which we demonstrate in a later section.

2.2 *Transition-Specific Covariate Effects*

A second, but related, advantage afforded by multistate models is the ability to estimate unique coefficient effects across each of the specified transitions in the model, allowing a researcher to determine whether the same covariate exerts a different effect at different stages of a larger process. Again, this advantage is similar, in a limited sense, to a classic CR model in which the determinants of transitioning into one stage are allowed to vary from the determinants of transitioning into another stage. For the U.S. House example, for instance, an open gubernatorial seat increases the probability of a House incumbent leaving to seek alternative office, but open gubernatorial seats do not affect the probability of primary election defeats, general election defeats, or retirement (Box-Steffensmeier and Jones 2004, 170). More broadly, consider the initial transitions in Fig. 1f, in which subjects located in Stage 1 are simultaneously at risk of two transitions: a transition into Stage 2 and a transition into Stage 3. As in a CR model, the baseline hazard of each transition is permitted to vary, but so too are the effects of the independent variables, such that the same covariate may exert different effects on the timing of each transition.

However, the use of *transition-specific covariates* is not limited to a classic CR situation. It can be extended to each of the specified transitions in the process. For example, it is possible to examine whether the occurrence of intermediate transitions in a process alters the determinants of transitioning into the same stage. Consider the possible transitions into Stage 2 depicted in Fig. 1f. Classic CR would permit only one transition into Stage 2. By contrast, Fig. 1f depicts two separate transitions into Stage 2, depending on whether a subject is directly transitioning into Stage 2 from Stage 1 or whether the subject has experienced an intermediate transition into Stage 3.[23] In this context, multistate models can estimate a unique covariate effect for x equal to $\beta_{(1\to 2)x}$ if a subject is currently in Stage 1 and $\beta_{(3\to 2)x}$ if a subject is instead currently located in Stage 3, as it may be the case that by experiencing an intermediate event in the form of Stage 3 the determinants of transitioning into Stage 2 have fundamentally changed.[24] The use of transition-specific covariates to estimate these unique covariate effects is possible across *each* of the five transitions depicted in Fig. 1f, for all covariates in the model. Including transition-specific covariates helps defend against the biased estimates that would result from inappropriately assuming x's effect is the same across all transitions when, in truth, it is not.

Multistate models are exceptionally flexible in the specification of unique covariate effects. As such, the decision regarding how many unique covariate effects ought to be estimated in any given context is largely a matter of theory. It is entirely possible to estimate a unique coefficient for each covariate in the model across each of the transitions, provided that there is an adequate number of observed transitions of each type (Box-Steffensmeier and Jones 2004, 162; Therneau and Grambsch 2000, 61–64).[25] However, in many contexts, it may be inappropriate to estimate a unique coefficient across each of the transitions in a model, as a covariate may exert a similar effect across one or more transitions. If researchers suspect that x exerts the same effect on two or more transitions, a single coefficient for x may be estimated by constraining x's effect to be equal for those transitions. Wald tests for coefficient estimates' equivalence can aid in determining whether and how many unique coefficient estimates are appropriate in a particular application (Greene 2012, 113–21). This can be done either by conducting pairwise comparisons of coefficient estimates across transitions—for example, testing whether $\beta_{(1\to 2)x}$ is significantly different from $\beta_{(3\to 2)x}$—or by conducting joint tests of significance for whether all $\boldsymbol{\beta}_{(1\to 2)}$ are significantly different from $\boldsymbol{\beta}_{(3\to 2)}$ (Therneau and Grambsch 2000, 226).

2.3 *Transition Probabilities*

Transition-specific baseline hazards and covariate effects allow more precise modeling of the distinct transitions that constitute a larger process. Nevertheless, they are ill suited, on their own, to making more systematic inferences about a political process as a whole. For example, if we consider Stage 4 in Fig. 1f as the final outcome of interest, focusing solely on direct transitions from Stage 3 to Stage 4 would limit our understanding of how subjects arrive in Stage 4. Multiple transition sequences may begin with i in Stage 3 and end with i in Stage 4—what some call

[23] In many ways, this is similar to a probit or logit, which would allow for the estimation of transition-specific coefficients via interactions (Brambor, Clark, and Golder 2006). However, multistate models have an advantage because they are more flexible in terms of the number of estimated transitions, the order in which they are experienced, and the number of subsequent events for which a subject is at risk.

[24] In Online Appendix E.II, we use simulations to show that our estimates of β will indeed be biased when $\beta_{(1\to 2)x}$ and $\beta_{(3\to 2)x}$ are different, but erroneously pooled together as a single transition.

[25] Usually, five to ten observed transitions is sufficient for each additional transition-specific covariate, but not always. See Online Appendix L for more details.

"plurisectality" (Jones and Metzger forthcoming). For example, a subject could transition from Stage 3 to Stage 2 and then to Stage 4, or it could move from Stage 3 to Stage 1 to Stage 2 and then to Stage 4, along with a number of other possible paths given Fig. 1f's recursive nature. Plurisectality has particular import if, for theoretical or substantive reasons, we are interested in how subjects move from the process's initial stage, Stage 1, to its final stage, Stage 4. For instance, U.S. House incumbents can leave office and be reelected in a later Congress.[13] Such an individual would be observationally equivalent in the data to someone who has served consecutive House terms—they would both be coded as incumbents, even though the "history" of their incumbency is different, in ways that may be of theoretical interest. Focusing on the risk of each individual transition in isolation would provide, at best, a piecemeal understanding of how the political process unfolds. In order to understand the process in its entirety, and make inferences about it as a whole, it is necessary to aggregate the risk of each individual transition in the process—for instance, both how a representative exits the U.S. House *and* how she returns to the Congress.

Transition probabilities address this concern by estimating the probability of a subject occupying each stage in the model at time t, $t+1$, $t+2$, etc. (Wreede, Fiocco, and Putter 2010, 2011). Transition probabilities help overcome concerns about plurisectality by using a product integral to account for the probability of both direct and indirect transitions into the stage of interest. In other words, if we are interested in the probability of a subject moving from Stage 3 to Stage 4 over some period of time, the transition probability estimate would take into account *each* of the possible paths through which a subject could arrive at Stage 4, given that it occupies Stage 3 in the present: $3\rightarrow4$, $3\rightarrow2\rightarrow4$, $3\rightarrow1\rightarrow2\rightarrow4$, $3\rightarrow1\rightarrow3\rightarrow2\rightarrow4$, and so on.

As this example suggests, transition probability estimates involve three key pieces of information:

1. *The stage a subject currently occupies*. A subject's current stage may impact the probability of arriving at a subsequent stage of interest. It may be the case that, for example, transitions from Stage 1 to Stage 2 in Fig. 1f occur quickly, whereas transitions from Stage 3 to Stage 2 take a while. If true, the probability of a subject occupying Stage 2 would differ dramatically depending on the subject's stage in the present.
2. *The time frame* for which the transition probabilities are to be estimated. We referred to these quantities as s and t in equation (4) and Online Appendix A. For instance, are the transition probability estimates to begin at the initial time under study ($s=0$), or only after some time has elapsed ($s>0$)?
3. *A covariate profile of interest* by fixing each of the model's covariates at a particular value, similar to estimating predicted quantities of interest from other estimators. Doing so makes it possible to evaluate the effect of a covariate on the process as a whole, and not just on a particular transition. Estimating transition probabilities permits us to evaluate the *net* effect of a particular covariate on the process as a whole, which becomes especially important if the covariate exerts opposite effects on different transitions (e.g., exerting a positive effect on one transition and a negative effect on another).

As we mentioned, calculating transition probabilities is paramount when scholars are interested in assessing a covariate's *net* effect across all transitions, particularly when scholars suspect the covariate's effect differs across transitions. Work on democracy's effect in interstate conflict is an excellent example. Existing work suggests that democracy's effect varies, depending on the outcome of interest—for example, the likelihood of peaceful settlement attempts, militarized attempts, or favorable militarized attempt outcomes (e.g., Reed 2000; Hensel 2001; Slantchev 2004; Hansen, Mitchell, and Nemeth 2008). Recognizing this, Jones and Metzger (forthcoming) extend Huth and Allee's (2002) foundational work on territorial disputes by assessing democracy's *net* effect on territorial dispute resolution using a multistate model. They find that, under certain conditions, democracies take *longer* than autocracies to resolve their disputes after a militarized settlement attempt. This result is counter to common wisdom, which would suggest that democracy's effect "nets to a positive"—that is, democracies would resolve their disputes faster. The insight would go undiscovered without modeling the various stages within a territorial dispute and calculating

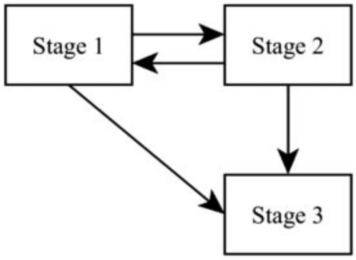

Fig. 3 Stage diagram—simulated.

democracy's net effect across all these transitions. Put differently, our inferences about democracy's net effect are inaccurate and misleading when we model only one piece of the dispute process.

3 Applications

Estimating multistate models is straightforward because they extend semiparametric Cox models. As such, they may be readily estimated using widely used statistical software packages such as *Stata* and *R* once the data are structured properly.[26] The `mstate` package in *R* (Wreede, Fiocco, and Putter 2010, 2011) is specifically designed to facilitate estimating and interpreting multistate models. `mstate` also has a number of utilities that help with data manipulation and, most importantly, provide the ability to directly estimate transition probabilities.[27] We rely on this package in each of the applications below. We use simulated data to demonstrate the first attractive feature of multistate models (transition-specific baseline hazards) and reexamine Maeda's (2010) study on democratic reversals to showcase the models' other two attractive features (transition-specific covariate effects and transition probabilities) (Metzger and Jones 2016).

3.1 *Simulated Data*

Figure 3 displays our simulated data's stage structure.[28] There are three stages, and four transitions among them. All our subjects begin in Stage 1, and all eventually end up in Stage 3. This stage structure is commonly referred to as an "illness-death" model in biostatistics, with Stage 1 corresponding to healthy individuals, Stage 2 to sick individuals, and Stage 3 to deceased individuals. Judicial career paths are one possible political process with the same stage structure, with "District Court" corresponding to Stage 1, "Appellate Court" to Stage 2, and "Supreme Court" to Stage 3. We use an elapsed-time formulation for our durations.[29]

We chose a true parameter value for (1) each transition's baseline hazard and (2) the effect of a covariate, x, on each transition (reported in Online Appendix I's Table 9). We then ran Monte Carlo simulations. Specifically, we generated a data set using Table 9's parameter values, estimated the multistate model, and recorded our quantities of interest. We repeated this procedure 1000 times and averaged our quantities of interest across all 1000 simulations.

[26] For data set organization details, see Cleves et al. (2010, 378–381), Jones and Metzger (forthcoming, Online Appendix A), and Wreede, Fiocco, and Putter (2010).
[27] `mstate` is capable of simulating transition probabilities for semi-Markovian setups as well. For details, see Online Appendix H.
[28] We use the `msm` package's simmulti function to generate our simulated data (Jackson 2011). `msm` uses the user-specified transition intensities for each transition to generate the observed data, by calculating transition probabilities as an exponentiated function of the intensities and time (Jackson 2011, 2). Additionally, we kept our x values fixed across all the simulated draws.
[29] Our model is Markovian as a result of (1) recording our durations in this way and (2) not including other covariates regarding a subject's transition history; for details, see Online Appendix C.

Table 1 Baseline hazard comparisons—simulated data

Stage 1 Exiting transitions		*Stage 3 Entering transitions*	
Quantity	p-value	Quantity	p-value
$\beta_{(1\rightarrow 3?)}$	0.000	$\beta_{(2\rightarrow 3?)}$	0.228
PH test	0.456	PH test	0.528

Notes. PH test uses Schoenfeld residuals, with t transformed by (1 – Kaplan–Meier). p-values are averaged across 1000 simulations. To collapse the two transitions, the column's p-values must both be greater than 0.05.

3.1.1 Stratified baseline hazards

We begin by checking if the various baseline hazards are different from one another. We look at two different situations for illustrative purposes, though we easily could have chosen others; theory and substance should guide which baseline hazards to check. In the first situation, we examine whether α_{12_0} and α_{13_0} are different from one another. It amounts to seeing if the two transitions *out* of Stage 1 have equivalent baseline hazards. In the second, we examine α_{13_0} and α_{23_0}, which checks whether the two transitions *into* Stage 3 have equivalent hazards. We know, in truth, that the first situation's baseline hazards are different, whereas the second situation's are the same, based on the parameter values we chose to generate the data.

The checking procedure has two parts. First: for each situation, we estimate a model in which (1) the two transitions in question share the same stratum, meaning that they share the same baseline hazard, but (2) we include a dichotomous variable for one of the two transitions in question (Wreede, Fiocco, and Putter 2010, 265).[30] Including the variable allows us to see whether the transition we dichotomize has a significantly different effect on the hazard than any other transition in its stratum by explicitly parameterizing the effect.[31] Since the transitions we are collapsing share the same stratum, this amounts to testing whether one of the collapsed transitions has a significantly different baseline hazard than the other collapsed transition (e.g., Fig. 2a versus 2b). A statistically significant coefficient means the two baseline hazards are, in fact, different, and we should stratify (Fig. 2b). We report the relevant coefficient's p-value, averaged across all the simulations, in Table 1's β row. As expected, the p-value for Stage 1's exiting transitions is smaller than 0.05, suggesting a difference exists between α_{12_0} and α_{13_0}. This evidence alone is sufficient to justify stratifying these two transitions. Also as expected, the p-value for Stage 3's entering transitions is larger than 0.05. α_{13_0} and α_{23_0} are statistically indistinguishable from one another. This is one piece of evidence against stratifying transitions $1\rightarrow3$ and $2\rightarrow3$, but it is not sufficient.

Second, we also need to check whether the dichotomous variable's coefficient has proportional hazards (e.g., Fig. 2b versus 2d). This tells us whether the *rate* at which the collapsed transitions' hazards change is the same. We can imagine situations in which the two hazards are statistically indistinguishable from one another, but their rates of change are vastly different (e.g., Fig. 2c). If this is the case, the two transitions' hazards clearly are not equivalent, and we should stratify them. We test for proportional hazards using Schoenfeld residuals, calculated for the dichotomous variable's coefficient (Box-Steffensmeier and Jones 2004, 131–37; Therneau and Grambsch 2000, chap. 6; Keele 2010; Licht 2011; Park and Hendry 2015). Statistical significance indicates a violation of the proportional hazards assumption, and the need to stratify. We report the p-values in Table 1's "PH Test" row. Neither PH p-value is statistically significant, implying no proportional hazards violation for either set of collapsed hazards.

Taking the two previous paragraphs together, we can conclude that the baseline hazards of $1\rightarrow2$ and $1\rightarrow3$ are different from one another (Table 1's left half). Because the underlying rates at which subjects exit from Stage 1 differ from one another, these exiting transitions should therefore be

[30]For example, if we are checking to see if the $1\rightarrow2$ transition has the same baseline hazard as $1\rightarrow3$, we generate a variable coded 1 if an observed transition is $1\rightarrow3$, and 0 otherwise. It does not matter which transition is dichotomized.
[31]This model should still include all transition-specific covariate effects. Otherwise, the PH test may come back statistically significant due to model misspecification, instead of a true PH violation (Keele 2010).

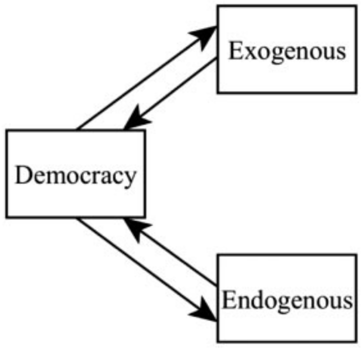

Fig. 4 Stage diagram—Maeda.

stratified. Additionally, we can conclude that the baseline hazards between $1 \rightarrow 3$ and $2 \rightarrow 3$ are statistically indistinguishable from and proportional to each other (Table 1's right half). The underlying rates at which subjects enter Stage 3 are not different. As a result, we could estimate a single baseline hazard for this pair of transitions.[32]

3.2 *Maeda: Modes of Democratic Reversal*

We use Maeda (2010)'s study on democratic reversals as our second application. Maeda's central theoretical contribution is that democratic regimes may end in multiple ways. A democratic regime could be exogenously terminated, from outside the government itself (e.g., military coups). It could also be endogenously terminated, from inside the government (e.g., self-coups). Maeda shows that his covariates of interest have different effects depending on the type of democratic reversal. Maeda uses a CR setup to assess his hypotheses empirically, with coups and self-coups as the competing failure events.

Maeda is specifically interested in democratic reversals alone. However, if we are ultimately interested in what factors contribute to or hinder the presence of democratic regimes in countries, it also makes sense to look at the entire democracy reversal-restoration process, instead of one piece of it. For instance, the role of current political institutions and the society in which they are embedded appear in narratives describing both democratic reversals and democratic restorations (e.g., Linz and Stepan 1978, 1996; O'Donnell and Schmitter 1986). Therefore, it may be reasonable to expect that the type of the democratic reversal may condition the political and economic factors associated with restoration. As a consequence, we reexamine Maeda's study, but in addition to democratic reversal we simultaneously consider "democratic restorations" (see Fig. 4).[33] We ask: If a democracy becomes a nondemocracy, does it revert back? If so, how long before it does?[34] We use a gap-time formulation for our durations.[35] Maeda uses the same formulation. We think gap time makes theoretical sense, because the duration a country spends in each stage is likely to be pertinent. The time that elapses since a coup is likely to have a strong bearing on the probability of a

[32]At this juncture, though, we would hesitate in entirely collapsing the specific transitions into Stage 3—that is, same baseline hazards *and* the same covariate effects. x's effect could still differ across transitions.

[33]To obtain as many observed transitions as possible, we use replication data from Svolik (2015), which contains information on democratic reversals and reversal type worldwide from 1789 to 2007. Maeda's replication data set is also worldwide, but only from 1950 to 2004. We include four covariates in our model: economic development (measured as GDPPC), economic growth (measured as GDP growth), a dummy variable for whether the democracy is a presidential system, and a dummy for "whether a military dictatorship preceded the current democratic spell" (Svolik 2015, 724–25). Maeda's full models use more covariates, but we chose a more parsimonious approach in order to place the focus first and foremost on our estimating procedure.

[34]Online Appendix D's Table 5 contains information about all 123 transitions in the expanded data set, based on what stage the state is transitioning *from* (the current stage) and what stage it is transitioning *to* (the next stage).

[35]This makes our multistate model semi-Markov; see Online Appendix C for further discussion.

Table 2 Multistate model of democratic reversals-restorations

(a) Unique covariate effects

	$D \to Ex$	$D \to En$	$Ex \to D$	$En \to D$
Economic development	−0.678**	−0.283	0.524*	1.023*
	(0.175)	(0.250)	(0.245)	(0.509)
Economic growth	0.101	0.002	−0.294	0.217
	(0.095)	(0.107)	(0.238)	(0.542)
Presidential system	0.020	0.713	−0.244	0.927
	(0.301)	(0.439)	(0.430)	(0.599)
Military—previous ND type	1.345**	−0.393	0.115	1.971
	(0.328)	(0.482)	(0.401)	(1.095)
Log-likelihood (partial)	−450.242			

(b) Collapsed covariate effects

	$D \to Ex$	$D \to En$	$Ex \to D$	$En \to D$
Economic development	−0.559**	−0.559**	0.460*	0.460*
	(0.141)	(0.141)	(0.204)	(0.204)
Economic growth	0.033	0.033	0.033	0.033
	(0.065)	(0.065)	(0.065)	(0.065)
Presidential system	0.188	0.188	0.188	0.188
	(0.200)	(0.200)	(0.200)	(0.200)
Military—previous ND type	1.262**	−0.222	0.158	0.158
	(0.317)	(0.465)	(0.358)	(0.358)
Log-likelihood (partial)	−456.631			

Notes D, democracy; Ex, exogenous; En, endogenous; ND, nondemocracy. Shaded cells in the same row are constrained to be equal during estimation. *$p \leq 0.05$, **$p \leq 0.01$, two-tailed tests.

democratic restoration, for example. Table 2's top half contains the results of our multistate model when we allow covariate effects to vary by transition.[36]

3.2.1 Transition-specific covariate effects

Whether we *need* to allow covariate effects to vary by transition is the matter we turn to first. We take the model we report in Table 2a and compare it to a second model in which we force every covariate to have the same effect across every transition. Our aim is to see whether *any* transition-specific covariate effects are necessary. We use a likelihood-ratio test to assess this proposition (Aalen, Borgan, and Gjessing 2008, 135–36). The second model is a restricted version of Table 2a's model, since we are constraining the parameter estimates to be equal across transitions. The null hypothesis is that Table 2a's coefficients and the second model's coefficients are equal. A significant likelihood-ratio test means that the coefficients are *not* equal, implying that our unconstrained model from Table 2a is the better bet. Our test comes back with a p-value smaller than 0.05 ($p = 0.000$, $\chi^2 = 40.84$ with 12 d.f.). Therefore, our use of transition-specific covariates appears to be justified, more broadly.

We can also assess whether a covariate exerts the same effect on specific subsets of transitions. We examine transition pairs to illustrate the procedure, similar to our test for baseline hazard equivalence. As before, there are several different scenarios we could check, and theory should guide our choices. Here, Maeda has theoretical predictions about economic development and

[36]We test for equivalent baseline hazards for the two transitions *out* of Democracy, and also for the two transitions back *into* Democracy. Our tests suggest the transitions should remain separate.

presidential systems—specifically, that these variables' effects will differ across the two outward transitions from Democracy. We can assess whether the two individual coefficients are significantly different from one another using a Wald test. The Wald test's null hypothesis is that the coefficients are equal ($\beta_{(\mathrm{D}\rightarrow\mathrm{Ex})\mathrm{DEV}} = \beta_{(\mathrm{D}\rightarrow\mathrm{En})\mathrm{DEV}}$). Surprisingly, this Wald test is statistically insignificant ($p = 0.185$, $\chi^2 = 1.76$ with 1 d.f.). Counter to Maeda's argument, economic development's effect does not significantly differ. We could estimate a single effect β_{DEV} for the D→Ex and D→En transitions. A similar truth exists for presidential systems. Assessing whether $\beta_{(\mathrm{D}\rightarrow\mathrm{Ex})\mathrm{PRES}} = \beta_{(\mathrm{D}\rightarrow\mathrm{En})\mathrm{PRES}}$ returns a p-value of 0.189 ($\chi^2 = 1.73$ with 1 d.f.). The effect of presidential systems on coups is no different than its effect on self-coups. We could also collapse these two effects into one.[37]

In this specific instance, the wide-ranging likelihood-ratio test evidence, when coupled with the evidence from the individual Wald tests, suggests checking a middle ground to finalize the model's specification—permitting some transition-specific covariates, but collapsing others. Collapsing could mean, for a specific covariate, estimating only one effect for every transition (e.g., 1 β, four transitions) or estimating more than one effect but still fewer than the number of transitions (e.g., 2 βs, four transitions). On top of this, each covariate may exhibit different "collapsing" patterns. We encourage scholars to methodically test the various permutations in their model. We can again use likelihood-ratio tests to adjudicate between all these various possibilities.[38] We perform all the permutations to arrive at our final specification, which we report in Table 2's bottom half.

We discuss only two scenarios here, for illustrative purposes. First, we check if we can estimate a single effect for economic development for all the model's transitions, while permitting all other variables to have transition-specific effects. We compare this restricted model to Table 2a's unconstrained model. The likelihood-ratio test comes back statistically significant ($p = 0.000$, $\chi^2 = 23.75$ with 3 d.f.). It is better to estimate four transition-specific effects for economic development, not one overall effect. Second, we check the same situation for presidential systems. The likelihood-ratio test is statistically insignificant ($p = 0.178$, $\chi^2 = 4.91$ with 3 d.f.). We could estimate only one effect for presidential systems across all four transitions in our model.

In sum, multistate models allow us to easily include transition-specific covariate effects and use likelihood-ratio and Wald tests to assess whether a covariate's effect differs across transitions. We point out the difficulty of performing these specification tests using classic CR setups, because each transition is estimated as a separate model. By contrast, multistate models estimate all the transitions as part of one model. We demonstrated this ability by checking the coefficients associated with economic development and presidential systems. Maeda argues that each covariate should exhibit a different effect on transitioning to nondemocracy via coup versus via self-coup. We find no evidence to support this claim. $\beta_{(\mathrm{D}\rightarrow\mathrm{Ex})\mathrm{DEV}}$ and $\beta_{(\mathrm{D}\rightarrow\mathrm{En})\mathrm{DEV}}$ are statistically indistinguishable from one another, as are $\beta_{(\mathrm{D}\rightarrow\mathrm{Ex})\mathrm{PRES}}$ and $\beta_{(\mathrm{D}\rightarrow\mathrm{En})\mathrm{PRES}}$.

3.2.2 Transition probabilities

An exclusive focus on the risk of each individual transition, in isolation, is quite restrictive for two reasons. First, it is poorly suited to assessing the larger democratic process, which contains recursive transitions—that is, both the risk of democratic reversal *and* the risk of democratic restoration following a reversal. Substantively, we may be interested in the probability that a country is a democracy in five years, but this simple query belies the fact that a country could be a democracy in five years either because it remained a democracy *or* because it experienced a democratic reversal

[37]We performed the same Wald tests on Maeda's original replication data set using his full model specification. We find similar results for a continuous-time formulation, which is the comparable scenario with our multistate model. Economic development's effect is not statistically different between the two transitions, and neither is presidential system's effect. For a discrete-time formulation, economic development remains statistically insignificant, but presidential system's effect becomes statistically different across the two transitions.

[38]Provided that the two alternative models are nested, as always. As a culminating evaluation criterion, scholars can use non-nested goodness-of-fit metrics like AIC and BIC to arrive at a final specification.

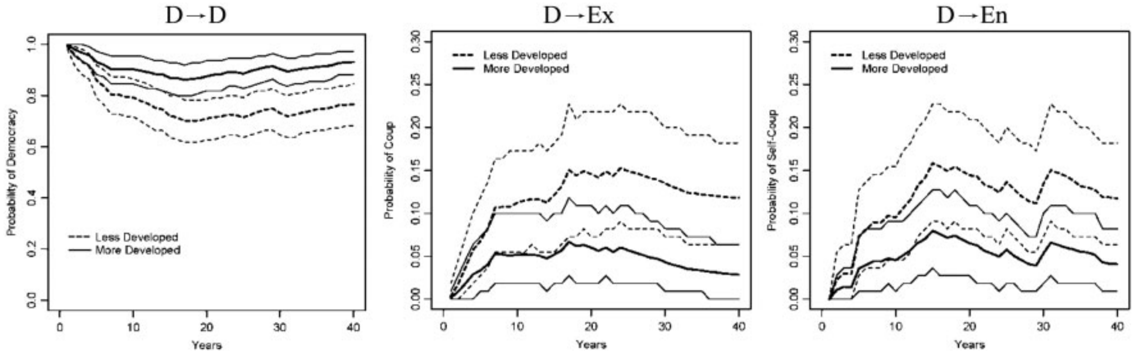

Fig. 5 Transition probabilities from democracy.
Notes. Lines represent a country's probability of occupying the corresponding stage by t. All estimates begin with the current stage as Democracy and $s = 0$. Quantities computed using simulation. Thin lines = 95% confidence intervals.

and subsequently recovered within that five year period. Second, focusing on one transition does not permit us to examine a covariate's net effect. If a particular covariate has a positive effect on some transitions and a negative effect on others, we may want to assess that covariate's effect on the overall trajectory of a subject through the process. Transition probabilities are well suited to address both these concerns.

We estimate a set of simulated transition probabilities using Table 2b's model.[39] We begin by setting the observation time to zero and placing the country in the Democracy stage. Figure 5 contains three plots, each showing the probability that the country will occupy a different stage of the process over time. In essence, this set of transition probabilities captures: given that country i is a new democracy ($s = 0$), what is the probability it will remain a democracy (Democracy, left), experience a coup and become nondemocratic (Exogenous, middle), or decide on its own to cease being a democracy (Endogenous, right)? We compute these probabilities for the 25th (less developed) and 75th (more developed) percentiles of economic development and hold all other variables at their median values.

Figure 5's probability estimates take into account all of the possible paths through which a country could begin in Democracy at $s = 0$ and arrive elsewhere by t. For example, the probability that a country is democratic at $t = 10$ reflects the probability that the country remained democratic over all ten years, but also the probability that it experienced a coup at time 2 and subsequently redemocratized by time 10, and the probability that it experienced a self-coup at time 4 and subsequently redemocratized by time 10, and so on. Therefore, by estimating transition probabilities, we obtain economic development's net effect on the larger democratic process, because we are simultaneously taking into account that economic development makes democratic reversals less likely and also makes recoveries more likely (Table 2(b)). Figure 5's estimates clearly bear this out. They show that democratic countries tend to remain democratic over time, but wealthier democracies are even more likely to do so than poorer democracies (left panel). For instance, after approximately 10 years, wealthier democracies have a 91% probability of being democratic, compared with only a 79% chance for poorer democracies.[40]

To demonstrate the broader implications of studying the process as a whole, we compare Fig. 5's reported transition probabilities from our democratic reversal-restoration multistate model (Fig. 6's dashed line) with comparable transition probabilities from a classic CR model of only democratic reversals (Fig. 6's solid line).[41] The classic CR model only contains the two exiting transitions from Democracy, mirroring Maeda (2010). The CR model's data set contains seventy-three total transitions, evident from Online Appendix D's Table 5. Notably, the CR data set still contains

[39]For why we must simulate, see fns. 21, 27, and 35.
[40]For more on substantively interpreting transition probabilities, see Online Appendix G.
[41]We set economic development to its 75th percentile value and all other covariates to their median values.

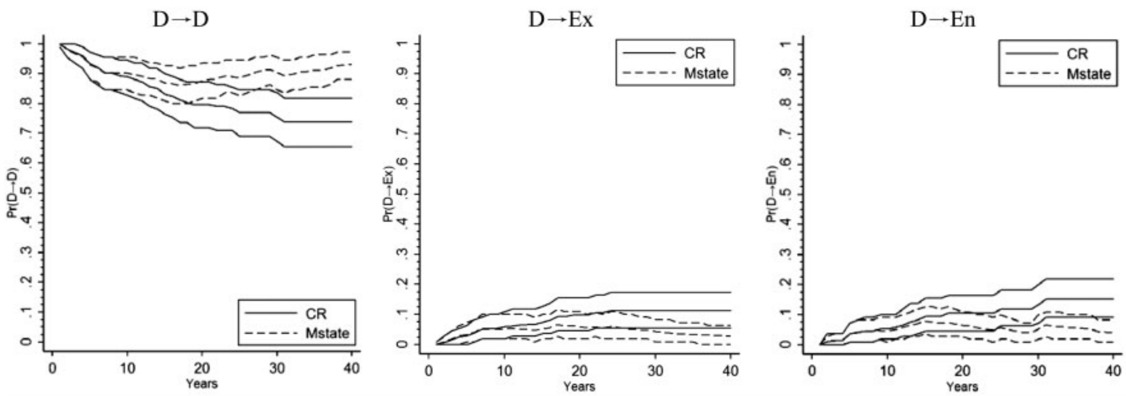

Fig. 6 Transition probabilities: CR versus Mstate.
Notes. Lines represent the probability of a country occupying the corresponding stage by t. All estimates begin with the current stage as Democracy, and $s = 0$. Quantities computed using simulation. Thin lines = 95% confidence intervals.

observations for any country that was a democracy, experienced a reversal, and then became democratic once more. The CR model simply does not *model* the restoration—a crucial distinction with serious ramifications.

On the whole, the models' transition probabilities are not the same. Across all three of Fig. 6's panels, the two transition probabilities are comparable for the first ten years, but then begin to diverge. The divergences make sense, and dramatically underscore many of the points we have made. Consider the probability that a country remains democratic (left panel). In our multistate model, countries can both leave and return to democracy. Accordingly, we see our model's transition probability both increase and decrease across time. By contrast, in the CR model, countries can only leave democracy. The CR model's transition probability reflects this fact by systematically decreasing across time. The substantive implication is that we would underestimate the probability of a country being democratic in the long term if we used a CR model, because the model itself does not acknowledge the possibility of democratic restorations. Importantly, this implication highlights that such underestimation will occur *even if* our argument is only concerned with democratic reversals.

A similar truth holds for the second and third panels. Even if we do not believe that nondemocracy (via coup or self-coup) is *substantively* a stage countries cannot leave, we are effectively modeling it as such *empirically* with a CR model. As a result, the CR model overestimates the probability of a country being nondemocratic because the model does not permit countries to become anything but in the long run.

To close, we acknowledge these dramatic differences between CR models' and multistate models' transition probabilities will not always exist. Specifically, the two models will always perform identically when the all stages except the first are truly absorbing. However, Fig. 6 makes strikingly evident that differences *can* exist, and it is this very possibility that should give scholars serious pause and encourage them to add multistate models to their methodological toolkit. Our discussion of Fig. 5 demonstrates a small sampling of the inferences we can make using transition probabilities.

4 Conclusion

How can we model processes characterized by causal complexity that occur across time? We suggest multistate survival models are one answer. Estimated as a stratified Cox model, a multistate model permits researchers to examine all of the transitions in a process. Multistate models are incredibly flexible and are able to capture many different transition structures. They can easily accommodate processes with competing, repeated, recursive, and sequential transitions. The end result is a more holistic take on the process of interest. Simpler survival models, like standard Cox models and competing risks, are less holistic because they are more restrictive. Standard Cox models only examine a single transition, and CR models only examine a single stage (and every transition out of that stage).

We highlighted three features that make multistate models particularly attractive: transition-specific baseline hazards, transition-specific covariate effects, and transition probabilities. The first two make accurate estimation of the last possible. The last epitomizes multistate models' holistic perspective. The models can estimate the probability of a subject occupying a particular stage at t by accumulating the probabilities associated with every possible transition sequence—both direct and indirect—ending in that stage. Importantly, transition probabilities permit researchers to estimate a covariate's net effect on the entire process, which may differ from its direct effect on a single transition.

We used two different applications to showcase these features. First, we used simulated data to demonstrate the well-established statistical tests for assessing baseline hazard equivalence. Second, we extended Maeda's (2010) study of democratic reversals by adding democratic restorations. We used the example to demonstrate how to check for different covariate effects and the versatility of transition probabilities. The transition probabilities suggest that wealth, in the form of economic development, is important. Wealthy democracies are more likely to stay democracies than poorer democracies. We also showed how a classic CR model of only democratic reversals would underestimate a country's long-term probability of being democratic and overestimate its long-term probability of being nondemocratic.

Multistate models' potential for political science research is clear. The models extend methodological tools scholars already use widely, and, in so doing, give us new leverage over substantively important questions about political processes. The leverage comes in the form of myriad possible inferences: the models permit researchers to make inferences about how subjects move through the process while also enabling researchers to make inferences about the nature of the process's causal complexity—Do covariate effects differ across transitions? Should certain stages be modeled separately, or can they be collapsed? What is a covariate's net effect across all the transitions in a process? Such inferences afford researchers new ways to assess the implications of their arguments. Further, modeling the complexity across time provides us with more accurate estimates and, thus, more accurate hypothesis tests. With the ubiquity of both time and causal complexity in political processes, multistate models are therefore a powerful tool for expanding our current understanding and breaking new ground.

Conflict of interest statement. None declared.

References

Aalen, Odd, Ornulf Borgan, and Hakon Gjessing. 2008. *Survival and event history analysis: A process point of view*. New York: Springer.

Allison, Paul D. 1984. *Event history analysis: Regression for longitudinal event data*, 1st ed. Thousand Oaks, CA: Sage.

Beck, Nathaniel, Jonathan 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.

Beyersmann, Jan, Arthur Allignol, and Martin Schumacher. 2011. *Competing risks and multistate models with R*. New York: Springer.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event history modeling: A guide for social scientists*. Cambridge, UK Cambridge University Press.

Box-Steffensmeier, Janet M., Suzanna Linn, and Corwin D. Smidt. 2014. Analyzing the robustness of semi-parametric duration models for the study of repeated events. *Political Analysis* 22(2):183–204.

Box-Steffensmeier, Janet M., and Christopher J. W. Zorn. 2002. Duration models for repeated events. *Journal of Politics* 64(4):1069–94.

Brambor, Thomas, William Roberts Clark, and Matt Golder. 2006. Understanding interaction models: Improving empirical analyses. *Political Analysis* 14(1):63–82.

Braumoeller, Bear F. 2003. Causal complexity and the study of politics. *Political Analysis* 11(3):209–33.

Carter, David B., and Curtis S. Signorino. 2010. Back to the future: Modeling time dependence in binary data. *Political Analysis* 18(3):271–92.

Chiba, Daina, Nils W. Metternich, and Michael D. Ward. 2015. Every story has a beginning, middle, and an end (but not always in that order): Predicting duration dynamics in a unified framework. *Political Science Research and Methods* 3(3):515–41.

Cleves, Mario, William Gould, Roberto Gutierrez, and Yulia Marchenko. 2010. *An introduction to survival analysis using Stata*, 3rd ed. College Station, TX: Stata Press.

Cox, David R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society B* 34(2):187–220.

Crowder, Martin J. 2012. *Multivariate survival analysis and competing risks*. Boca Raton, FL: Chapman and Hall/CRC.

Daniel, William T. 2015. *Career behaviour and the European parliament: All roads lead through Brussels?* Oxford: Oxford University Press.

Epstein, David L., Robert Bates, Jack Goldstone, Ida Kristensen, and Sharyn O'Halloran. 2006. Democratic transitions. *American Journal of Political Science* 50(3):551–69.

Fine, Jason P., and Robert J. Gray. 1999. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association* 94(446):496–509.

Fukumoto, Kentaro. 2009. Systematically dependent competing risks and strategic retirement. *American Journal of Political Science* 53(3):740–54.

Geskus, Ronald B. 2015. *Data analysis with competing risks and intermediate states*. Boca Raton, FL: Chapman and Hall/CRC.

Gordon, Sanford C. 2002. Stochastic dependence in competing risks. *American Journal of Political Science* 46(1):200–17.

Greene, William H. 2012. *Econometric analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall.

Hansen, Holley E., Sara McLaughlin Mitchell, and Stephen C. Nemeth. 2008. IO mediation of interstate conflicts: Moving beyond the global versus regional dichotomy. *Journal of Conflict Resolution* 52(2):295–325.

Hansford, Thomas G., Elisha Carol Savchak, and Donald R. Songer. 2010. Politics, careerism, and the voluntary departures of U.S. District Court judges. *American Politics Research* 38(6):986–1014.

Hensel, Paul R. 2001. Contentious issues and world politics: The management of territorial claims in the Americas, 1816–1992. *International Studies Quarterly* 45(1):81–109.

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

Huth, Paul K., and Todd L. Allee. 2002. *The democratic peace and territorial conflict in the twentieth century*. Cambridge, UK Cambridge University Press.

Jackson, Christopher H. 2011. Multi-state models for panel data: The msm package for R. *Journal of Statistical Software* 38(8):1–28.

Jones, Benjamin T. 2013. *The past is ever-present: Civil war as a dynamic process*. PhD dissertation, The Ohio State University.

Jones, Benjamin T., and Shawna K. Metzger. Forthcoming. Evaluating conflict dynamics: A novel empirical approach to stage conceptions. *Journal of Conflict Resolution*.

Keele, Luke. 2010. Proportionally difficult: Testing for nonproportional hazards in Cox models. *Political Analysis* 18(2):189–205.

Licht, Amanda A. 2011. Change comes with time: Substantive interpretation of nonproportional hazards in event history analysis. *Political Analysis* 19(2):227–43.

Juan J. Linz and Alfred Stepan, eds. 1978. *The breakdown of democratic regimes*. Baltimore, MD: Johns Hopkins University Press.

Juan J. Linz and Alfred Stepan, eds. 1996. *Problems of democratic transition and consolidation: Southern Europe, South America, and post-communist Europe*. Baltimore, MD; Johns Hopkins University Press.

Maeda, Ko. 2010. Two modes of democratic breakdown: A competing risks analysis of democratic durability. *Journal of Politics* 72(4):1129–43.

Mattiacci, Eleonora, and Benjamin T. Jones. 2016. (Nuclear) change of plans: What explains nuclear reversals? *International Interactions* 42(3):530–58.

Metzger, Shawna K., and Benjamin T. Jones. 2016. Replication data for 'surviving phases: Introducing multistate survival models.' Harvard Dataverse.

Mills, Melinda. 2004. Stability and change: The structuration of partnership histories in Canada, the Netherlands, and the Russian Federation. *European Journal of Population* 20(2):141–75.

O'Donnell, Guillermo, and Philippe C. Schmitter. 1986. *Transitions from authoritarian rule: Tentative conclusions about uncertain democracies*. Baltimore, MD: Johns Hopkins University Press.

Park, Sunhee, and David J. Hendry. 2015. Reassessing Schoenfeld residual tests of proportional hazards in political science event history analyses. *American Journal of Political Science* 59(4):1072–87.

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.

Putter, H., M. Fiocco, and R. B. Geskus. 2007. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine* 26(11):2389–430.

Reed, William. 2000. A unified statistical model of conflict onset and escalation. *American Journal of Political Science* 44(1):84–93.

Slantchev, Branislav L. 2004. How initiators end their wars: The duration of warfare and the terms of peace. *American Journal of Political Science* 48(4):813–29.

Svolik, Milan W. 2015. Which democracies will last? Coups, incumbent takeovers, and the dynamic of democratic consolidation. *British Journal of Political Science* 45(4):715–38.

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

Vollset, Stein Emil, Aage Tverdal, and Håkon K. Gjessing. 2006. Smoking and deaths between 40 and 70 years of age in women and men. *Annals of Internal Medicine* 144(6):381–89.

Wreede, Liesbeth C. de, Marta Fiocco, and Hein Putter. 2010. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Computer Methods and Programs in Biomedicine* 99(3):261–74.

———. 2011. mstate: An R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software* 38(7):1–30.