Does Accountability Vary? Examining the Tenure of State Supreme Court Justices

Todd A. Curry
Department of Political Science
Western Michigan University
3438 Friedmann Hall
Kalamazoo, MI 49008-5346
todd.a.curry@wmich.edu

Mark S. Hurwitz
Department of Political Science
Western Michigan University
3434 Friedmann Hall
Kalamazoo, MI 49008-5346
mark.hurwitz@wmich.edu

Does Accountability Vary? Examining the Tenure of State Supreme Court Justices

Todd A. Curry
Mark S. Hurwitz
Western Michigan University

Abstract

At the core of the debate on how judges should be selected is the issue whether the judiciary should be accountable or independent. Among the various selection systems in the states, including partisan elections, non-partisan elections, gubernatorial and legislative appointment, and the so-called Missouri Plan, do any of these methods of selection provide for greater or lesser levels of accountability? It is important to answer this question with empirical evidence, so that policy makers can accurately choose the selection system that best addresses their goal of either accountability or independence. In this paper we seek to answer this question by analyzing the tenure length of state supreme court justices over a period of 25 years (1980-2005). We utilize Cox Proportional Hazard Models to test the duration of judicial tenures in the states and the likelihood justices are at risk of leaving the bench. By doing so we seek to examine which selection systems produce the most accountable justices. The results of our hazard models show that justices in partisan election states have the shortest tenures and are at the greatest risk of departing the bench, as this selection method is the most accountable of these intuitions. By contrast, the Missouri Plan (also called the merit system) is least accountable, even less so than appointive systems, as the retention elections employed by this selection system provide no legitimate threat to justices' tenure on the bench. Our findings thus demonstrate that accountability varies across the judicial selection systems in the states and does so in a predictable fashion.

Does Accountability Vary? Examining the Tenure of State Supreme Court Justices

Todd A. Curry
Mark S. Hurwitz
Western Michigan University

> "Judges have considerable discretion and should be held accountable for their choices, at least at the state level where we would expect a close connection between public preferences and public policy" (Bonneau and Hall 2009 p. 2).

> "An independent judiciary as provided by the Constitution has assured that the governed as well as the government are bound by the Rule of Law" (Sandra Day O'Connor 2003, 41).

Introduction

Political participation in government through voting is the "crucial attribute of democracy" (Riker 1982, 5). Indeed, the very essence of democracy is tied to free and popular elections. Yet, the judiciary in the United States has always been somewhat different. For instance, the federal system provides for presidential appointment of judges, who then serve essentially for life. Alexander Hamilton (*Federalist No. 78,* 398-399) argued that in order to avoid "arbitrary discretion in the courts" there must be "permanency of the judicial offices". Hamilton continued in *Federalist 78*: "temporary duration in office . . . would have a tendency to throw the administration of justice into hands less able, and less well qualified to conduct it with utility and dignity" (at 399).

Despite these strong testimonials in the *Federalist Papers*, with their ability to create their own judicial systems the several states have traditionally viewed judicial selection in a manner more varied than Hamilton envisioned for the federal courts. While some states in the Eighteenth and Nineteenth centuries followed the federal system of judicial appointment and retention, many others did not. In fact, in the early history of the United States appointment by state legislatures as well as various forms of judicial elections proved to be popular in the states.

Then, the early Twentieth century brought with it an increase in the use of judicial elections. Subsequently, judicial reform groups advocated for what they term merit selection as a way to remove politics from the courts.[1] Many states have since agreed, choosing to utilize some form of this selection system to for their judges. The result is great variety in the states today regarding judicial selection.

Thus, there is a long tradition in the United States of arguments on the subject of the various means of judicial selection, as well as those regarding judicial behavior more generally. Indeed, to assert there is a controversy regarding methods of judicial selection in the states would downplay the current policy landscape, in terms of both numbers of players and the level of rhetoric. A critical reason for the intensity of this debate is that judicial selection systems play directly into the normative issue of whether judges should be independent or accountable. Hamilton and the Founders of the Constitution favored judicial independence, a position more recently and strongly articulated by Justice Sandra Day O'Connor. Yet, accountability via political participation is essential to all democratic institutions, including the judiciary (Bonneau and Hall 2009; Hall 2001a). There is much at stake when it comes to the selection system each state chooses for its judiciary. Consequently, the terms of the debate tend to plunge into a level of hyperbole, often (but not always) with claims made based on a lack of empirical research by either side to support their assertions.

Accountability is one of the key claims made by those in favor of electoral systems. That is, because judges must stand for election, there is a greater likelihood that they will be held answerable for their actions and judgments on the bench. If a judge is making decisions that are

---

[1] While reformers and advocates refer to this as the merit system, it often is also called the Missouri Plan, as that state is credited with being the first to implement this system in 1940. At a minimum this selection mechanism incorporates nomination by commission but usually also includes retention elections. We use the terms merit selection and the Missouri Plan interchangeably.

out of tune with the prevailing political mainstream within the state, the electoral mechanism provides the public with a means of removal and replacement, thus enhancing judicial accountability (Bonneau and Hall 2009).

On the other hand, proponents of merit-based selection seek to maximize the goal of judicial independence, in order to shield judges from external influences deemed inappropriate when it comes to interpreting the law. These advocates, including Justice O'Connor and the American Judicature Society, among others, favor this selection system, believing that choosing and retaining judges in this manner enables them to make decisions regarding the law that are free from political or electoral constraints, thus augmenting judicial independence.

In addition to electoral and merit systems of selection, governors and legislatures have the power to appoint judges in a number of states. Judicial appointment by these political elites is often considered similar to merit selection in terms of the purposes behind these selection systems. That is, it is thought that appointive systems tend to promote judicial independence. Hamilton accordingly promoted this position over 200 years ago for the federal courts, and the analogy holds for gubernatorial and legislative appointment in the states.

As this debate has been constructed, then, conventional wisdom holds that elections promote accountability while appointive and merit systems instead advance independence. However, this notion that differing selection systems produce various levels of accountability has never been directly or systematically tested. The question thus becomes whether the various selection systems in fact have different effects on accountability. That is, do some selection systems produce more accountability than others? Our purpose in this paper is to examine empirically the various systems of judicial selection for differing levels of accountability by comparing the tenures of state supreme court justices across disparate selection systems.

Accordingly, we enter this debate not to take any particular side, but instead to test this particular issue within this controversy: whether different mechanisms of judicial selection produce varying levels of accountability.

Judicial Selection and Accountability

Methods of selection of supreme court justices vary by each state, such that nearly each state's selection system is unique. Nevertheless, the literature considers five broad, institutional selection mechanisms in the states: partisan elections, non-partisan elections, merit selection, gubernatorial appointment, and legislative appointment (Warrick 1993).

Partisan and non-partisan elections are similar vehicles for selecting judges, in that candidates usually must make it first through a primary election and then win a general election to attain a seat on the bench. The principal difference between these two types of electoral systems concerns the manner the individuals are listed on the ballot, either with or without a partisan label.

The common conception of Missouri Plan has been described as "a process in which a nonpartisan or bipartisan commission nominates a few individuals for a judicial position, for appointment (usually) by the executive based on the commission's recommended names, with subsequent tenure on the bench dependent upon a retention election at specific intervals" (Hurwitz and Lanier 2001, 86, fn. 11). While merit selection thus incorporates elements of both appointive and electoral systems, the vast majority of judges subject to retention elections are in fact retained (Hall 2001b), and thus the defining feature of merit systems is the initial nomination by commission (Hurwitz and Lanier 2001). Nevertheless, advocates of this selection mechanism

6

contend that retention elections are a critical feature as well, in that they allow citizens to determine whether judges continue in office.

In gubernatorial appointment states, judges are chosen by the governor, usually with confirmation by the state senate. This system is thus analogous to judicial selection in the federal courts (Canon 1972). Finally, legislative appointment occurs in but two states, Virginia and South Carolina, where their respective state legislatures have the authority to choose and retain justices on their supreme court.

These are the five broad categories of selection systems used in the states. In this paper we are interested in how these several systems of judicial selection promote accountability. As Hall (2001a, 1136) asserted in her study on how electoral pressures influence state judges to retire: "This study is only the first that seeks to unravel the fascinating and complicated nexus between democratic processes and career decisions in the states' highest courts. Countless questions remain [and] further inquiry will be fruitful, especially for examining and perhaps dispelling myths surrounding the politics of institutional design." In this paper we carry on this line of research by examining the interplay between accountability and judicial selection systems in state courts of last resort. In particular, we explore whether different judicial selection systems produce varying levels of accountability with respect to judicial tenures.

Theory

For the purpose of this study, we define accountability as the length of tenure for a justice on a state court of last resort. This definition recognizes that increased risk produces increased accountability (Hall 2001a). That is, greater risk of losing one's seat on the bench coincides with an increasing level of accountability. Each type of selection method has as one of its foundations

the goal of subjecting justices to varying levels of accountability, which is directly associated with the amount of risk of removal that the individual justices face.

With this definition of accountability in mind, our theoretical expectations are based in part on an institutional perspective (Aldrich 1995; Brace and Hall 1990; North 1990) and in part on conventional wisdom, both of which interestingly coincide here. We assume that both formal and informal institutions shape behavior in an endogenous setting. From an institutional perspective, the various judicial selection systems should produce variable levels of accountability in predictable ways, based on the unique institutional arrangements of each selection system.

As well, conventional wisdom plays a role in the rhetoric regarding judicial selection. Thus, there is much discussion, scholarly and otherwise, regarding the perceived benefits of different selection systems based notions of accountability and independence. For instance, both advocates and opponents of judicial elections generally agree that judges who stand for some form of election are more likely to be held accountable than judges in appointive systems. Stated otherwise, because of the institutional features of elections judges exposed to elections should be at greater risk of being forced off the bench than their colleagues in appointive systems.

While all the types of judicial elections potentially raise accountability to some degree, the occurrence of risk – and thus the greatest potential for accountability – should be highest in partisan elections. As Hall argued (2007, 1151), "voters vote when they have interest, readily available information, and choice." Such increased interest and information are likely to be highest in partisan elections, because 1) partisan elections produce the most expensive campaigns, which in turn increases the electorate's access to information (Bonneau and Hall 2009), and 2) partisan elections have the potential to field high quality challengers, which in turn

8

increases both information and choice (Hall and Bonneau 2006). Furthermore, even if voters do not absorb a sufficient amount of information about the judicial candidates from the campaign, candidates for these offices are listed with their partisan affiliation present on the ballot, a key source of information for many voters (Downs 1957).

Structurally, non-partisan elections are near mimics of their partisan counterparts, save for the fact that the candidates' partisan affiliation is not listed on the ballot in non-partisan elections. Because there is no partisan cue, and because non-partisan campaigns are less expensive than their partisan counterparts, voters consequently have comparatively less information and interest in non-partisan elections (Bonneau and Hall 2009).[2] Accordingly, due to the nature of competitive elections we hypothesize that non-partisan electoral systems should produce high levels of accountability. However, since partisan elections hold relatively more risk, accountability should be somewhat lower in non-partisan elections.

While retention elections are not associated with analogous levels of risk as competitive elections, advocates of the Missouri Plan make clear that retention elections provide voters with the ability to hold judges accountable. Notwithstanding, retention elections are generally low-salience affairs where candidates do not face any challengers and where there is strong evidence of ballot roll-off (Bonneau and Hall 2009). We thus would expect retention elections to produce lower levels of accountability than partisan or non-partisan elections.

However, retention elections remain, at their core, elections.[3] Consistent with advocates' claims, then, we would expect retention elections to produce higher levels of accountability than

---

[2] For these reasons, successful challengers in non-partisan elections usually spend quite a bit of money, certainly more than the norm, in order to overcome the information deficit that ordinarily accompanies non-partisan campaigns (Bonneau and Hall 2009).

[3] For instance, the American Judicature Society recently countered criticism that voters had no say in the merit system for the Kansas Supreme Court by asserting that the state's electorate is able to vote in retention elections every six years (Koranda 2010).

appointive systems, which are generally thought to produce judges who are independent, as occurs in the federal system (Hamilton, *Federalist No. 78*). That is, while gubernatorial and legislative appointment systems subject judges to some accountability when their terms come to an end (which distinguishes them somewhat from the federal system), the institutional mechanism of judicial appointment puts judges under less risk than electoral systems, as judges are only accountable to the executives and legislators responsible for their retention. If judges make decisions in line with their fellow political elites in these other branches, rarely should they be at risk of losing their positions. Of all these methods of selection, then, we would expect appointive systems to contain the comparatively lowest levels of accountability, since these judges are at the lowest level of risk.

In sum, the methods of selection and retention used in state supreme courts should affect the tenure of justices in predictable ways, as different institutional arrangements subject the justices to varying levels of risk that should be directly related to accountability. To reiterate our theoretical expectations, partisan elections should produce maximum levels of accountability, as the risk justices assume here is likely to be highest relative to the other selection mechanisms. Non-partisan judicial elections should similarly generate high levels of accountability, though perhaps not as much as that found in more salient partisan elections. While we expect retention elections, which by definition are non-competitive, to provide less risk and accordingly less accountability than the competitive electoral systems, we also presume that justices facing retention elections have greater levels of accountability than those confronting gubernatorial or legislative appointment. Stated otherwise, we anticipate that justices in appointive systems will have the most independence, the longest tenure and, thus, the lowest levels of accountability, since they are theoretically prone to less risk than any of the elective systems.

Data and Methods

We collected data on the length of tenure for every individual state supreme court justice

who served in one of sixteen states from 1980-2005. We categorized each state as employing

one of the following selection and retention systems: 1) partisan election, 2) non-partisan

election, 3) retention election, or 4) appointment. We selected these sixteen states because they

incorporate classic features of the respective selection systems without significant modification.

Further, we chose four states from each selection system so that all of the categories would be

equally represented. Table 1 provides the states we analyze herein and their respective methods

of selection.[4]

[Table 1 about here]

Table 1 also displays the number of justices and failures for each state in our study from

1980-2005. A failure occurs when a justice leaves the bench during the period of analysis for

any reason, which might include retirement, resignation, losing an election, or death. We are

keenly interested in the failures in our study, since we anticipate that the time a justice leaves the

bench is conditioned on her selection system.

Accordingly, our dependent variable is operationalized as the length of time a justice

served on the state supreme court, which takes into account when a justice's tenure failed. In

order to assess the effect of methods of selection upon the duration of the justices' tenure, we

incorporate variables to represent non-partisan elections, retention elections, and appointive

systems. In the full model this means that partisan elections serve as the statistical baseline,

---

[4]According to our coding rules, the Missouri Plan is thus represented by retention elections. We aggregated
gubernatorial and legislative appointment into a single "appointment" category, in part because only two states
utilize legislative appointment and only a few more employ gubernatorial appointment. Since both gubernatorial
and legislative selection methods are appointive in nature, and since we include all selection systems in our study, it
made theoretical sense to combine them.

allowing us to compare directly the accountability of each method of selection. We also include three dichotomous control variables in the full model, including each justice's political party, gender, and minority status (White or non-White). We have no theoretical expectations for these controls, but we include them for purposes of model specification. While some studies have examined dynamic effects of party or diversity in state courts, it remains to be seen whether these serve as potential influences on judges' tenure (Brace and Hall 1990; Hurwitz and Lanier 2003).

Intuitively, our study is akin to a medical experiment where a selection of patients has been given one of four different drugs, and our interest would be in learning how long patients generally survive on each of these drug and what their risk of death is. In our circumstance, the four different drugs are represented by the various methods of judicial selection we analyze, while the patients' lives and deaths are depicted by the justices' tenure length. Of course, we are not concerned with the patients' death; instead, our interest lies in how long the justices served at the time of their exit from the bench based on their particular selection system.

The appropriate statistical method to analyze data of this sort is an event history model, also known as hazard models (Collett 2003). There are a number of hazard models from which to choose, based on the assumptions of the model and the data utilized. The type we opted to employ is a Cox proportional hazard model. We utilized this semi-parametric technique because we make no assumptions about the form of the duration dependency. Parametric models assume specific distributions when modeling the hazard function. Our only assumption is that the hazard rate will be different in predictable ways across the various methods of selection, not that the hazard rate will have a specific distributional form. Thus, the Cox proportional hazard model is most appropriate for our research (Box-Steffensmeier and Jones 2004; Cox 1972).

Employing event history models to examine temporally-ordered data has a rich history in political science (*see* Box-Steffensmeier and Jones 2004). In particular, Cox proportional hazard models (hereafter, "hazard models" or "Cox models") have been applied in research on judicial politics. For instance, Shipan and Shannon (2003) used a hazard model to examine the duration of Supreme Court nominations and confirmation, while Langer et al (2003) applied this model to analyze how associate justices on state supreme courts select their chief justice. From a methodological perspective ours is analogous to these studies, as we examine accountability based on the duration of tenure of state supreme court justices.[5]

When dealing with duration models with data of this sort, issues of left truncation and right censoring become apparent. Left truncation occurs when an individual in the dataset joined the risk pool prior to the first observation. In our study, this means a justice was selected for a judicial position at some point before 1980, when we begin our analysis. These individuals do not enter at t=0, because we know from the data we collected when their tenure first began as a supreme court justice. Thus, a justice may have been unobserved by our study for 8 years, but when she enters our risk pool we code the data as if she began her tenure at t=9. On the other hand, right censoring occurs when a justice continues to serve after the end of observation period in 2005. In event history analysis employing Cox models, neither of these circumstances is problematic, because we are interested in the occurrence and non-occurrence of an event, in our case failures, during the period of analysis. That is, individuals who are coded as left truncated in the data contribute information to the model at the point they become observed, while right censored data contribute information to the model until they are no longer observed (Box-Steffensmeier and Jones 2004).

---

[5]Other examples of studies in judicial politics using event history models that are not Cox proportional hazard models include Patton (2007) and Savchak et al (2006).

In order to identify a Cox hazard model, the analyst must first assess the proportionality of hazard rates across different values of the independent variables. "The Cox Model assumes that the hazard function of any two individuals with different values on one or more covariates differ only by a factor or proportionality" (Box-Steffensmeier and Zorn 2001, 974). If this assumption does not hold true, the estimates of all the covariates in the model could be biased, not just the offending variables. Following the lead of Box-Steffensmeier and Zorn, we examined the assumption of proportionality by testing the scaled Schoenfeld residuals, and we found no evidence of non-proportionality.[6] We now turn to our Cox models in order to analyze accountability across selection systems.

Results

Figure 1 illustrates mean judicial tenures as a function of selection system. This figure shows that accountability across the methods of selection varies. In particular, justices in partisan electoral systems have the shortest tenures on average, while justices in merit-based systems (via retention elections) serve for the longest period of time. This lends initial support for our hypotheses, as partisan elections followed by non-partisan elections apparently provide for the most accountability among selection systems. However, this figure also suggests that retention elections, and not appointments, descriptively produce the least amount of accountability. In fact, a justice within a partisan election system can expect her tenure on the court to be about 25 percent shorter than a justice serving in a state which uses retention elections.

[Figure 1 about here]

---

[6] These results are available at the request of the authors.

14

The first step in testing our hypotheses is to examine whether the survivor functions of the four methods of selection are equal. Collett (2003) demonstrates that the log-rank test provides for such a test by pitting our hypotheses against the null hypothesis that there is no difference between the survivor functions of the four selection systems. The log-rank test in Table 2 shows that the survivor functions of the four methods of selections are clearly discrete. The Kaplan-Meier survivor function is a similar test that can be assessed graphically (Box-Steffensmeier and Jones 2004), as shown in Figure 2. While the survivor functions overlap briefly, they demonstrate that the four methods of selection are largely distinct. Moreover, once again we find that retention elections have the lowest level of accountability, as evidenced by the Kaplan-Meier Survivor function that decreases the slowest (signifying the longest tenure) for this selection system. Therefore, we can reject the null hypothesis of equal influence among the selection systems, as levels of accountability for each selection method are different. We accordingly can move on to the next step in our analysis.

[Table 2 and Figure 2 about here]

While the log-rank test and Kaplan-Meier function provides some information on the duration of the data, neither allows for direct comparisons of change in the hazard rate across the methods of selection. We accordingly ran our Cox model, which specifically provides for direct comparisons in rates of change across time within the model. In particular, our hazard model allows us to "determine if a variable increases duration by looking at its effect on the baseline hazard rate" (Shipan and Shannon 2003, 662). A variable with a negative coefficient signifies that it decreases the hazard rate, while a positive coefficient connotes that the variable increases the hazard rate. Stated somewhat differently, a negative coefficient indicates there is a decreasing likelihood of a justice leaving the bench when compared to the baseline hazard, while

15

a positive coefficient means that there is an increasing likelihood the justice will leave the bench compared to the baseline hazard rate.

Table 3 displays the results of our Cox model. All three of the substantive variables obtain statistical significance within the model and are negatively signed, including non-partisan elections, appointments, and retention elections, while two of the control variables, justice party and minority status, are statistically significant,[7] with the coefficient for justice party negative and that for minority status positive. Thus, all of these variables, except for minority status, serve to decrease the hazard rate relative to partisan elections, the baseline in this model.

[Table 3 about here]

Substantively, this signifies that when compared to the baseline hazard rate of partisan elections, all of these other methods of selection have a lower level of accountability; that is, each system elicits a decrease in the hazard rate, which corresponds directly to a decrease in the level of accountability. Moreover, since all of the substantive variables are negative and significant, the selection system with the highest level of accountability, as we hypothesized, is partisan elections.

The hazard models provide additional information, as we can use the statistics from the model to rank the order of the different selection mechanisms in terms of the accountability they produce. In particular, the last column of Table 3 depicts the relative change in the baseline hazard rate for each variable. The interpretation of this statistic is intuitive, in that a positive percentage signifies that, when compared to justices who stand for partisan elections, justices in states with that specific selection method are more likely to leave the bench, while a negative outcome indicates they are less likely to leave the bench relative to partisan elections.

---

[7] We also estimated our Cox model using PAJID scores (Brace, Langer, and Hall 2000) instead of the justice party variable, which did not change the results for the full model; but; the PAJID variable was positive and insignificant, the latter of which makes substantive sense.

16

Accordingly, we find that justices in non-partisan election systems are 56 percent less likely to leave the bench than justices in partisan systems, while justices in appointive systems are 76 percent less likely to depart than partisan election justices. Finally, justices subject to retention elections are 153 percent less likely to exit the bench than justices subject to partisan elections. Apparently, retention elections do not behave like their more competitive counterparts, as justices in retention election systems are one and one half times less likely to relinquish the bench for any reason than partisan justices.

Finally, Figure 3 plots the hazard rates for all of the methods of selection against each other, as this figure provides graphical comparison of the varying levels of accountability produced by these selection systems. To reiterate, the higher the hazard rate the more at risk justices are of leaving the bench, as that selection system thus produces greater levels of accountability; conversely, the lower the hazard rate the more insulated and less accountable are justices from that system. The selection system with the highest hazard rate, and thus the most accountability, is partisan elections, and no other selection system has a hazard rate that is relatively close. This provides additional evidence that judges subject to partisan elections are most at risk of exiting the bench, as this is the most accountable selection system. The second highest hazard rate belongs to non-partisan elections, followed relatively closely by that for appointive systems. Since the hazard rates for non-partisan and appointive systems are relatively close, these selection systems behave similarly, though non-partisan elections are somewhat more accountable institutions than appointive systems. Lastly, the selection system with the lowest hazard rate by far, and thus the system with the least amount of accountability, is retention elections. In fact, partisan election justices have a higher risk of leaving the bench in their tenth year in office than retention election justices have in their 25th year on the bench.

Clearly, the risk of departing the bench is far lower for justices in retention elections than for justices in any of the other selection systems.

[Figure 3 about here]

Conclusion

Our findings from the hazard models largely support our theoretical expectations on varying levels of accountability. Similarly, our results are chiefly, but not entirely, consistent with conventional wisdom in this regard. We expected partisan and non-partisan elections to be the most accountable judicial selection institutions in the states, with partisan elections at the top, and that is precisely what we found. Justices in partisan elections have the shortest tenures and the greatest failure rates – that is, they leave the bench more often and earlier than justices from states with different selection methods. Clearly, our research shows the selection method that provides the greatest amount of accountability is partisan elections. And, while non-partisan elections produce accountability at rates higher than either retention elections or appointment, they are not quite as effective in this regard as partisan elections. Nevertheless, our results demonstrate that competitive elections, whether partisan or non-partisan, are the most accountable institutions when it comes to judicial selection in state supreme courts.

Our theory, which relied in part on conventional wisdom, also claimed that retention elections should produce less accountability than their electoral counterparts, since retention elections are non-competitive by design. Here again the results comport with our expectations. However, we also expected retention elections to be more accountable institutions than appointment systems. After all, retention elections are still elections; as well, there is anecdotal reasoning that appointments are designed to produce independent but not necessarily accountable judges (*see, e.g., Federalist 78*). On this point our theoretical expectations, as well as

18

conventional wisdom, were incorrect. Appointments turned out to be a more accountable selection system than retention elections. Thus, our hypothesized rank order of judicial accountability was basically correct, except we found that retention elections provide less accountability than we anticipated.

In the states we examined that employ retention elections, only one justice was removed by the voters in 122 retention elections over 25 years of our analysis (1980-2005); and, the average tenure rate was significantly longer with retention elections than with other methods of selection. The American Judicature Society (2010) claims the institution of retention election "provides an opportunity to remove from office those who do not fulfill their judicial responsibilities." Opportunity, however, is not necessarily the equivalent of action. Our findings evince that retention elections produce very little accountability in the justices who are subject to run in them. Thus, if judicial accountability is the primary goal with respect to state courts of last resort, our findings suggest that merit systems with retention elections are not a wise choice. However, if judicial independence is the overriding objective, then retention elections provide the best mechanism for maximizing this goal. While our findings and conclusions are somewhat consistent with those of other scholars, particularly Bonneau and Hall (2009) and Hall (2001a), our research adds nuance to the literature on judicial selection in the states in that no other study has incorporated a research design as we did. Indeed, our hazard models proved well-suited to analyzing the varying levels of accountability among state courts.[8]

---

[8] In a recent paper, Reddick and Caufield (2010) of the American Judicature Society claim that merit selection systems produce higher quality judges than judicial elections. In support they showed how judges in merit systems are disciplined for ethical violations less often than elected judges, and that when judges are disciplined the sanctions – including removal from the bench – are more severe in merit systems. Our study does not specifically address the issues they raised. However, our findings do show that judges in states with partisan and non-partisan elections are much more likely to depart the bench *for any reason* than judges subject to the Missouri Plan.

Our purpose in this paper was not to take sides in the current debate on what is the best or most appropriate judicial selection system. Instead, our aim was to assess assumptions on both sides of the debate regarding accountability by subjecting those assumptions to rigorous empirical testing. We believe we have been successful in our ambition here. Perhaps our most interesting finding is that retention elections produce far less accountability than previously had been assumed. The belief that retention elections produce levels of accountability similar to other electoral systems is simply not borne out by our data and empirical tests. In fact, we do not think we are overstating the case when we say that the belief that retention elections provide even a modicum of accountability is simply a stylized fiction.

The title of our paper asks, "does accountability vary?" Our findings support an answer to this question in the affirmative. While it has been assumed that that methods of selection could be rank-ordered by level of accountability, our hazard models proved very capable in doing so empirically. Thus, in this paper we have followed Hall's (2001a) plea of further research by examining the nexus of democratic processes and institutional design regarding selection in state courts of last resort. There remains much to study with respect to this heated debate, and we hope that scholars continue rigorous empirical testing of the effects of the various judicial selection methods in the states.

References

American Judicature Society. 2010. Merit Selection: The Best Way to Choose the Best Judges. *American Judicature Society.* http://www.judicialselection.us/uploads/documents/ms_descrip_1185462202120.pdf (last viewed August 22, 2010)

Aldrich, John H. 1995. *Why Parties? The Origin and Transformation of Political Parties In America.* University of Chicago Press: Chicago.

Bonneau, Chris W. and Melinda Gann Hall. 2009. *In Defense of Judicial Elections.* New York: Rutledge.

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists.* New York: Cambridge University Press.

Box-Steffensmeier, Janet M. and Christopher J.W. Zorn. 2001. "Duration Models and Proportional Hazards in Political Science." *American Journal of Political Science*. 45: 972-88.

Brace, Paul and Melinda Gann Hall. 1990. "Neo-Institutionalism and Dissent in State Supreme Courts." *The Journal of Politics* 52: 54-70.

Brace, Paul, Laura Langer, and Melinda Gann Hall. 2000. "Measuring the Preferences of State Supreme Court Judges." *The Journal of Politics* 62: 387–413.

Canon, Bradley C. 1972. "The Impact of Formal Selection Processes on the Characteristics of Judges — Reconsidered." *Law and Society Review* 6: 579-93.

Collett, David. 2003. *Modelling Survival Data in Medical Research.* Boca Raton: CRC Press.

Cox, D. 1972. "Regression Models and Life Tables." *Journal of the Royal Statistical Society, Serial B.* 34:187-220.

Downs, Anthony. 1957. *An Economic Theory of Democracy.* New York: Harper and Row.

Hall, Melinda Gann. 2001a. "Voluntary Retirements from State Supreme Courts: Assessing Democratic Pressures to Relinquish the Bench." *The Journal of Politics* 63:1112-40.

Hall, Melinda Gann. 2001b. "State Supreme Courts in American Democracy: Probing the Myths of Judicial Reform." *American Political Science Review* 95: 315-30.

Hall, Melinda Gann and Chris W. Bonneau. 2006. "Does Quality Matter? Challengers in State Supreme Court Elections." *American Journal of Political Science* 50: 20-33.

Hamilton, Alexander, James Madison, and John Jay. 1982 [1787]. *The Federalist Papers*. New York: Bantam Books.

Hurwitz, Mark S. and Drew Noble Lanier. 2001. "Women and Minorities on State and Federal Appellate Benches, 1985 and 1999." *Judicature* 85: 84-92.

Hurwitz, Mark S. and Drew Noble Lanier. 2003. "Explaining Judicial Diversity: The Differential Ability of Women and Minorities to Attain Seats on State Supreme and Appellate Courts." *State Politics & Policy Quarterly* 3: 329-52.

Koranda, Jeannine. 2010. "How Kansas Picks its Supreme Court Justices Becomes Issue in Governor's Race." *The Wichita Eagle.* Aug. 10, 2010. http://www.kansascity.com/2010/08/10/2141178/how-kansas-picks-its-supreme-court.html (last viewed August 19, 2010).

Langer, Laura, Jody McMullen, Nicholas P. Ray, and Daniel D. Stratton. 2003. "Recruitment of Chief Justices on State Supreme Courts: A Choice between Institutional and Personal Goals." *The Journal of Politics.* 65: 656-75.

North, Douglass C. 1990. *Institutions, Institutional Change, and Economic Performance.* Cambridge: Cambridge University Press.

O'Connor, Sandra Day. 2003. *The Majesty of the Law: Reflections of a Supreme Court Justice.* New York: Random House.

Patton, Dana. 2007. "The Supreme Court and Morality Policy Adoption in the American States: The Impact of Constitutional Context." *Political Research Quarterly* 60: 468-88.

Reddick, Malia and Rachel Paine Caufield. 2010. "Judging the Quality of Judicial Selection Systems: An Empirical Analysis." Paper presented at the annual meeting of the Midwest Political Science Association, Chicago.

Riker, William H. 1982. *Liberalism Against Populism: A Confrontation Between the Theory of Democracy and the Theory of Social Choice.* Prospect Heights, IL: Waveland Press.

Savchak, Elisha Carol, Thomas G. Hansford, Donald R. Songer, Kenneth L. Manning, and Robert A. Carp. 2006. "Taking It to the Next Level: The Elevation of District Court Judges to the U.S. Courts of Appeals. *American Journal of Political Science* 50: 478-93.

Shipan, Charles R. and Megan L. Shannon. 2003. "Delaying Justice(s): A Duration Analysis of Supreme Court Nominations." *American Journal of Political Science.* 47: 654-68.

Warrick, Lyle. 1993. *Judicial Selection in the United States: A Compendium of Provisions* (2d. Ed). Chicago: American Judicature Society.

Table 1: Descriptive Statistics.

States	Method of Selection	Number of Justices	Number of Failures*
Alabama	Partisan Election	31	22
Louisiana	Partisan Election	18	11
Texas	Partisan Election	44	34
West Virginia	Partisan Election	19	14
Kentucky	Nonpartisan Election	21	13
Oregon	Nonpartisan Election	24	17
Washington	Nonpartisan Election	29	19
Wisconsin	Nonpartisan Election	18	11
Iowa	Retention election	20	13
Kansas	Retention Election	18	11
Oklahoma	Retention election	18	9
Nebraska	Retention Election	20	13
Maine	Gubernatorial Appointment	22	16
New Jersey	Gubernatorial Appointment	19	12
South Carolina	Legislative Election	14	8
Virginia	Legislative Election	19	12
	Totals	351	232

* Failure occurs when a justice leaves the bench for any reason during the period of analysis from 1980-2005, including retirement, resignation, losing an election, or death.

Table 2: Log-Rank Test for Equality in the Survivor Function.

Selection method	Failures Observed	Failures Expected
Partisan	81	54.31
Non-Partisan	60	58.29
Retention	45	68.20
Appointment	46	51.19

Null hypothesis: equal failures expected.

Chi2(3) = 23.77

Pr>Chi2 = 0.00

Table 3: Cox Proportional Hazard Model of the Duration of Judicial Tenure, by Selection System.

Variable	Estimate (s.e)	Change in Hazard Rate
Non-Partisan Election	- .449 (.179)*	-56%
Appointment	- .564 (.189)*	-76%
Retention Election	- .932 (.202)*	-153%
Justice Party	- .310 (.149)*	-36%
Gender	- .370 (.240)	-44%
Minority	.652 (.294)*	92%

Log Likelihood = -1106.136
LR Chi2 = 31.71
Prob>Chi1 = 0.00

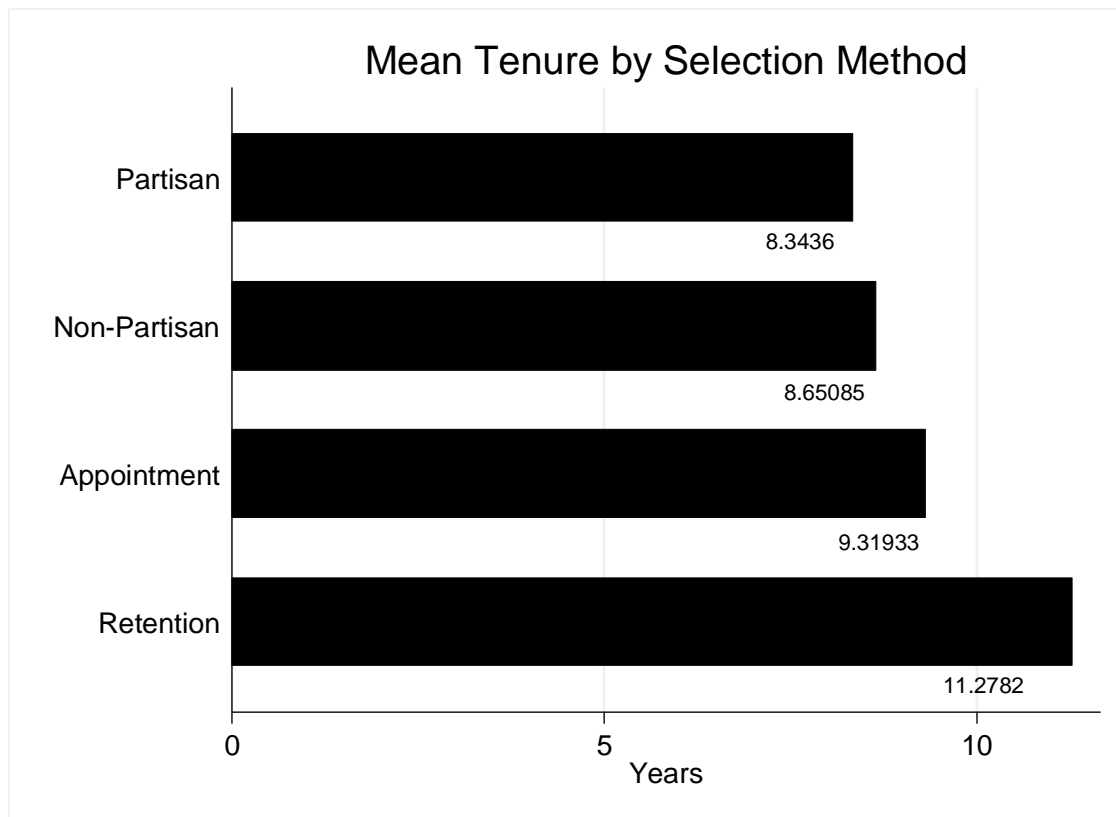**Figure 1: Mean Judicial Tenure as a Condition of Selection System.**

Mean Tenure by Selection Method

Partisan — 8.3436
Non-Partisan — 8.65085
Appointment — 9.31933
Retention — 11.2782

Years

Figure 2: Kaplan-Meier Test of Survivor Function Estimates.

Kaplan-Meier Survival Estimates

Legend:
— Partisan
········ Non-Partisan
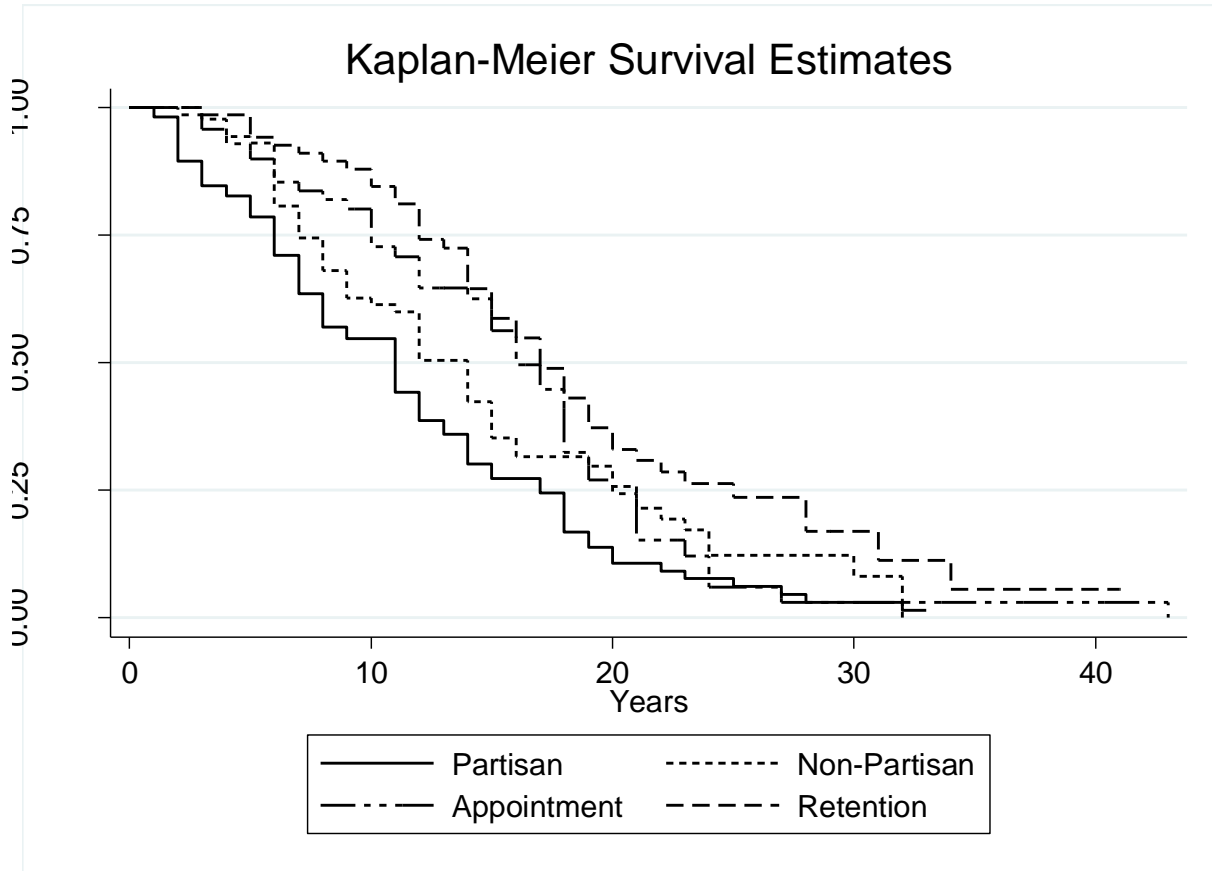—·—· Appointment
– – – Retention

**Figure 3. Cox Proportional Hazard Rates of the
Duration of Judicial Tenure, by Selection System.**

Baseline Hazard Rate for Judicial Selection Methods