

A Flawed Search for Bias in the American Bar Association's Ratings of Prospective Judicial Nominees: A Critique of the Lindgren Study

*Michael J. Saks & Neil Vidmar**

*A finding of fact is only as good as the methods used to find it.*¹

I. INTRODUCTION

Professor James Lindgren, using data gathered by the Federalist Society, has analyzed a data set pertaining to the nominees of the George H. W. Bush presidency (1988-1992) and the presidency of William Jefferson Clinton (1992-2000) for appointment to federal courts of appeals.² The purpose of the project was to test the hypothesis that the ABA's ratings of federal judicial nominees were biased. The study sought to discover the extent to which the ABA's ratings of

* Michael Saks, Ph.D. (Ohio State University, 1975), MSL (Yale Law School, 1983), is Professor of Law and Professor of Psychology, Arizona State University. He is co-editor of MODERN SCIENTIFIC EVIDENCE, past Editor-in-Chief of LAW & HUMAN BEHAVIOR, and a present or past member of the following editorial boards: JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY; LAW & SOCIETY REVIEW; LAW & POLICY; APPLIED SOCIAL PSYCHOLOGY ANNUAL; SOCIAL BEHAVIOUR: AN INTERNATIONAL JOURNAL OF APPLIED SOCIAL PSYCHOLOGY; ETHICS AND BEHAVIOR; EXPERT EVIDENCE; and LEGAL & CRIMINOLOGICAL PSYCHOLOGY.

Neil Vidmar, Ph.D. (University of Illinois, 1967), is Russell M. Robinson II Professor of Law at Duke Law School and Professor of Psychology. He is a present or past member of the editorial boards of LAW & HUMAN BEHAVIOR; LAW & SOCIETY REVIEW; PSYCHOLOGY, PUBLIC POLICY AND LAW; LAW & SOCIAL INQUIRY; the JOURNAL OF APPLIED SOCIAL PSYCHOLOGY; PSYCHOLOGY, CRIME AND LAW; LEGAL AND CRIMINOLOGICAL PSYCHOLOGY; and the CANADIAN JOURNAL OF LAW AND SOCIETY.

The two authors were asked by the ABA Section of Litigation to undertake a review of the Lindgren study that is the subject of this article. Neither author is actively involved with the ABA or its Section of Litigation, but we were compensated for the time we spent reviewing the Lindgren study and writing this article. Professor Stephan Landsman of DePaul Law School assisted us in obtaining information about ABA rating procedures and provided minor editorial commentary on drafts of this article. The analyses and conclusions in this critique are solely those of the authors. We are content to have the logic of our review be weighed against any suspicions of bias that compensation or assistance from Professor Landsman may raise. We wish to thank Professor Lindgren for providing a copy of the data set that he used for his study.

¹ A maxim from the world of research methodology and statistics.

² James Lindgren, *Examining the American Bar Association's Ratings of Nominees to the U.S. Courts of Appeals for Political Bias, 1989-2000*, 17 J.L. & POL. 1 (2001).

prospective judicial nominees can be explained by the politics of the president who made the nomination. While ratings of nominees are intended to be a reflection of the nominee's integrity, professional competence and judicial temperament, party politics may play a role in the ratings of these nominees.³ The Lindgren study looks at the 108 persons nominated and confirmed to the federal appellate bench during the four years of the Bush Presidency and the eight years of the Clinton Presidency.

The Lindgren study concludes that, among appeals court nominees *with* prior judicial experience, Bush appointees "appear to have been equally treated compared to Clinton appointees. Indeed, Bush nominees got an insignificant advantage."⁴ However, for nominees *without* prior judicial experience, Professor Lindgren reports that his "study shows – for the first time – strong evidence of differential treatment of nominees by the ABA's rating committee."⁵ According to the Lindgren study, virtually nothing mattered except which president made the nomination: "the patterns revealed in the data are consistent with a conclusion of strong political bias favoring Clinton nominees."⁶

³ The study is much like a Title VII job discrimination study in which the job being applied for is that of appellate judge and the discrimination alleged is that ratings of fitness for the job are improperly influenced by who proposed the person for the job.

⁴ Lindgren, *supra* note 2, at 6 (citation omitted).

⁵ *Id.* at 3.

⁶ *Id.* at 26. Such strong conclusions are asserted despite Professor Lindgren's own apparent realization that his study might not be able to support them. Professor Lindgren acknowledges that "cautious academics would not determine conclusively that such a bias exists" and that "one should always be cautious in drawing conclusions, especially where there are important unmeasured variables not in the models, such as integrity or judicial temperament." *Id.* Professor Lindgren disregards his own reasonable cautions when he moves on to giving policy advice: "One obvious policy response would be for the ABA to cease rating judges until it can eliminate the apparent bias or show that no bias exists." *Id.* at 27. Another policy option would be for "the Senate, the White House, and the press to de-emphasize or ignore entirely the ABA ratings because of probable political bias." *Id.*

Professor Lindgren's strong conclusions and recommendations are not confined to the original article. Professor Lindgren has expressed them, sans cautions, in a *Wall Street Journal* piece. James Lindgren, *Yes, the ABA Rankings on Judicial Nominees Are Biased*, WALL ST. J., August 6, 2001, at A13. In that article he states:

I've just completed a statistical study of the ABA's ratings of appointees to the U.S. Courts of Appeals during the Clinton and first Bush administrations and can report that the facts don't support the ABA's claim of objectivity. The ABA may once have been objective, but it's not anymore.

Id. Later in the column he asserts that:

The data suggest that when Bill Clinton took office, the ABA softened its standards, possibly emphasizing credentials such as temperament and

Of course, the hypothesis that is the focus of the Lindgren study – that the ABA’s judicial evaluations were not politically neutral – might or might not be true. However, a careful examination of the study indicates that it suffers from numerous serious and interconnected flaws that place its conclusions in considerable doubt. Informed by a proper understanding of the study’s methods, a knowledgeable and disinterested student of empirical social science research would not be able rationally to conclude that the ABA’s evaluations were politically biased.

The present Article provides a detailed review, a methodological critique, of the Lindgren study. We identify and explicate the study’s principal flaws and explain their consequences for the conclusions reached by Professor Lindgren.

II. THE PROBLEM OF DISAPPEARING JUDGES, OR, DATA DREDGING

To understand the methodological points that we are about to make it is important to recognize the stages of the decision making process that lead to confirmation of the judges who are the subject of the Lindgren study. In the administrations of the first President Bush and President Clinton the evaluation of potential judicial nominees generally began when the Justice Department and the White House reviewed names of potential judgeship candidates. Potential nominees who survived that initial vetting were referred to the ABA for evaluation by its Standing Committee on Federal Judiciary. With the ABA evaluations in hand, the White House and the Department of Justice reviewed the potential nominees once again, and actual nominations were made. The final stage of the process is Senate consideration of the nominees. Those confirmed by the Senate were ready to take their seats as federal judges.⁷

philosophy that are harder to measure than experience and educational success. Now the ABA is back to rating Republican nominees – and is apparently back to its old harsh ways.

Id. In his concluding paragraph he admonishes the ABA by asking:

Why hasn’t the ABA itself noticed the large political differences in its evaluative processes and worked harder to understand, explain or eliminate them? Now that there are hard data that support the claims of its critics, it would be good to see fewer denials and more introspection and reform.

Id.

⁷ Sheldon Goldman et al., *Clinton’s Judges: Summing Up the Legacy*, 84 JUDICATURE 228 (2001) at 229-231, provide details of the judicial nomination process during the Clinton presidency.

To guide the reader as we examine this matter we have constructed Table 1. The table reflects the several stages of the evaluation process. The first row of Table 1 represents the prospective nominees who are referred to the ABA for its evaluation. These are the people being evaluated by the ABA whether they are being considered for seats on the district courts (cell A) or the courts of appeals (cell B). The area of the cells in the table is a rough approximation of the number of people represented by the cell. Thus, for example, because there are on average approximately three times as many district court nominees as circuit court nominees, there is about three times as much area contained under the "District Court" heading as under the "Court of Appeals" heading.⁸ The President's nominees (cells C and D) are a

Evaluation of potential candidates began when names of potential candidates were received by the Justice Department. The potential candidate was instructed to fill out extensive questionnaires, including the ABA forms. Initial screening was conducted by Department of Justice lawyers with the goal of weeding out clear problem cases. The screening included a preliminary FBI investigation and a personal interview. If the person passed this initial hurdle, he or she was instructed to submit the ABA form to the ABA committee at the same time that the FBI began a full field investigation. In four to six weeks, the ABA Standing Committee on Federal Judiciary would send a letter to the Justice Department reporting the rating of the candidate. In the next stage of the process a joint Justice Department/White House committee made a decision regarding whether to move forward with the nomination. With respect to appeals court candidates, more than one prospective nominee was considered for each specific slot. Political considerations played an important role in determining which candidate would be nominated. The confirmation process involved additional scrutiny by the Senate, of course, sometimes involving highly partisan political decisions, before a candidate was confirmed. *Id.* at 231-241.

The judicial selection process employed by the first Bush administration is described in Sheldon Goldman, *Bush's Legacy: The Final Imprint*, 76 JUDICATURE 282 (1993). The description reveals procedures that invested decreasing time, staff, and other resources in the process of selecting potential nominees, certainly compared to those of the Reagan administration, *id.* at 285-86, and apparently also compared to those of the Clinton administration. *See generally*, Sheldon Goldman, PICKING FEDERAL JUDGES: LOWER COURT SELECTION FROM ROOSEVELT THROUGH REAGAN (1997); JOEL GROSSMAN, LAWYERS AND JUDGES (1965); William G. Ross, *Participation by the Public in the Federal Judicial Selection Process*, 43 VAND. L. REV. 1 (1990).

Professor Lindgren never considers this obvious possibility, namely, that the Bush administration's internal selection process yielded a group of prospective nominees who were collectively not as well qualified as the Clinton group simply because the Bush administration invested less effort in the process. If that is the case, then ABA bias would seem to cut the other way: despite receiving a less qualified pool of prospective nominees from the Bush administration, the ratings did not differ significantly. (That they did not differ significantly is shown *infra* at note 22.)

⁸ Based on available data, it is apparent that these numbers fluctuate over time (for example, the ratio of district court nominees to appellate court nominees ranges from about 2.2:1 to about 4.6:1 during the Bush I and Clinton eras) and with the details of the selection process. Table 1 therefore can present nothing more than an approximate and average sense of the relative magnitudes of the numbers and how they change in the courts as people move from being prospective nominees being evaluated by the ABA to confirmed judges. For the data, see CITIZENS FOR INDEPENDENT COURTS, UNCERTAIN JUSTICE: POLITICS AND AMERICA'S COURTS (2000). *See also* Goldman et al., *supra* note 7.

subset of the prospective nominees. In the next stage the Senate confirms most, though not all,⁹ of the nominees (cells E and F), who then are sworn in as judges. The bottom row of Table 1 represents the subset of the judges affirmed by the Senate who have no prior judicial experience (cells G and H). The data on which the Lindgren conclusion of bias depends come entirely from cell H. Cell H is further divided to indicate that the Lindgren study chose one of two different measures of the ABA's evaluation.¹⁰

Consideration of the total process leading to confirmation of a judge suggests that any bias by the ABA would be most evident in cells A and B. That is where the ABA makes its decisions. That is also where the sample size is largest, providing the most sensitive tests and most stable estimates.¹¹ By the time a candidate reaches cells E and F, in which the Lindgren data begin, the original pool of candidates has been winnowed by political and other considerations. The persons in cells A and B constitute the full set of people who would have been subjected to discriminatory review, according to the Lindgren hypothesis. The relationships present in this group would not have been distorted by numerous later removals made by decision-makers in the Justice Department, the White House and the Senate.¹² Had data on the prospective nominees in cells A and B been used, Professor Lindgren would have tested the hypothesis on the most appropriate sample.¹³ To the best of our knowledge, however, the Lindgren project made no

⁹ For example, in 1999-2000, the Senate confirmed only 57 of 83 (68.7%) district court nominees and only 13 of 32 (40.6%) court of appeals nominees. See Goldman et al., *supra* note 7.

¹⁰ The measure chosen was the more extreme and probably the less meaningful one. See *infra* notes 26-32 and accompanying text.

¹¹ As discussed *infra* notes 77-78 and accompanying text, the sample size on which the Lindgren study relied is far too small for the analyses employed, a situation that produces unreliable findings.

¹² As the contents of a sample change, the statistical relationships found in the sample change, sometimes dramatically. See further discussion of this point, *infra* notes 19-20 and accompanying text, and Section V.

¹³ There is no reason, in principle, that this could not have been done. It must be noted that the total pool of candidates represented in cells A and B, and the ratings given them by the ABA, are held in confidence by the ABA. In particular, names are never released of persons who progress no further in the nomination process. But private and governmental organizations often do make otherwise confidential data available to researchers, taking appropriate care to preserve the confidentiality of data. If sensitive medical, criminal, and corporate data, and jury deliberations can be made available for appropriate scrutiny by researchers, the ABA could be asked to do so as well. One way this sometimes is accomplished is to have the owner of the data do the coding and provide researchers with a data set from which identifying information has been removed. In that way the statistical data are made available while confidentiality is preserved.

**Table 1. Potential Nominee Samples Ignored and Analyzed
(Areas Approximately Proportionate to Sample Size)**

	District Courts	Court of Appeals
Prospective Nominees (evaluated by the ABA)	A	B
Nominees (nominated by the President)	C	D
Confirmed Nominees (confirmed by the Senate)	E	F
Confirmed nominees without prior judicial experience	G	H

effort to gain access to those data. Nor does Professor Lindgren discuss the potential impact on the study's conclusions of analyzing a subset of confirmed judges¹⁴ rather than the full sample of prospective nominees.¹⁵

Moreover, the Lindgren project begins by completely ignoring the district court nominees. This omission is never explained. If nothing else, the same analysis could have been conducted for confirmed district court nominees (cell E) that was conducted for the appellate judges (cell F). That would have been an informative replication, a check on the conclusions of this "first time" finding of bias,¹⁶ and either a confirmation or refutation of the findings that Professor Lindgren has asserted. But the district court pool was simply ignored.¹⁷ About 75% of the potential data available for testing the hypothesis of ABA bias were thus eliminated.¹⁸

Within the narrower domain of the appellate court, the ideal group to study would have been the prospective nominees who had been reviewed by the ABA (cell B), for all the reasons discussed above. Of course, the data would have required more effort to collect, but the findings would have been commensurately more reliable, and those who wished to make policy recommendations based on them would have been standing on firmer ground. Assuming those data were unobtainable, the Lindgren research might have turned its focus to

¹⁴ Namely, seated appellate judges with no previous judicial experience.

¹⁵ For further discussion of this impact, see *infra* notes 19-20 and accompanying text, and Section V.

¹⁶ "[T]his study shows – for the first time – strong evidence of differential treatment of nominees by the ABA's rating committee." Lindgren, *supra* note 2, at 3.

¹⁷ Professor Lindgren's rush to announce his "first time" demonstration of bias in the ABA's treatment of prospective judicial nominees is reminiscent of other premature announcements of "discoveries," such as the cold fusion fiasco. Researchers at the University of Utah announced their "discovery" of cold fusion in a rush to publication without adequate confirmation and replication, only to be shown soon after to have found nothing more than errors and illusions. See JOHN R. HUIZENGA, *COLD FUSION: THE SCIENTIFIC FIASCO OF THE CENTURY* (1992); GARY TAUBES, *BAD SCIENCE: THE SHORT LIFE AND WEIRD TIMES OF COLD FUSION* (1993).

¹⁸ As noted earlier, there are on average about three times as many nominees to the district courts as to the appellate court, and so we extrapolate accordingly to the entirety of the two columns of Table 1. See *supra* note 8.

One could imagine an argument that the Lindgren study might have made (but does not make) to support its decision to ignore prospective nominees for the district court, namely, that the ABA's bias is centered on nominees to the appellate courts. Of course, that argument would need to be fleshed out; merely asserting it does not make it so. But it is not necessary to explore that path because in every way – the wide embrace of its hypothesis, its language, its conclusions, and its policy recommendations – the Lindgren study makes clear that it sought to test broadly and it concluded that ABA political bias was and is ubiquitous.

those appellate court nominees put forward by the presidents (cell D), but it did not. Or to those nominees confirmed by the Senate (cell F), but it did not. The focus of the study was actually on a subset of confirmed nominees, namely those without prior judicial experience (cell H, containing 49 persons). Remarkably, then, the conclusions of the Lindgren study are based on a subset of a subset of a subset of a subset of the most obvious and appropriate pool of data.

This attrition from the most appropriate and largest sample has important consequences. As the pool is winnowed and transformed – from group B to group D to group F to group H – the statistical relationships to be found within it are almost certain to be transformed.

Consider a simple illustration, in a different domain, involving applicants to an elite law school. In the original applicant pool of students the LSAT scores will range quite widely. Since LSAT scores are thought to be related in an important way to ability to learn law, if all of the students who applied were admitted to the elite school the relationship between LSAT scores and first year grades would be quite high.¹⁹ But elite law schools do not, of course, admit all of the applicants. After the pool is narrowed and reshaped by the selection process, the class that enters is of far higher ability than the original applicant pool. Not only do the admitted students have higher LSAT scores, but also the range and variability of the scores is much smaller than it had been in the original pool. Because the LSAT scores now fall within a narrower span, when a correlation is computed between LSAT scores and first year law school grades the correlation will be much smaller.²⁰ Someone who looked only at the final group of selected law students and the small correlation it produced between LSAT scores and first year law school grades might be tempted to denounce the law school for paying little attention to measures of ability (measured by LSAT scores) in its admissions process and might suggest that the

¹⁹ The typical (median) correlation between LSAT scores and first year average law school grades is around .52 to .53. Lisa C. Anthony et al., *Predictive Validity of the LSAT: A National Summary of the 1995-1996 Correlation Studies* (Law School Admission Council LSAT, Technical Report 97-01) (August 1999).

²⁰ Although the median correlation for all 165 law schools participating in the study (with an average of more than 500 students per school) was around .52 to .53, at some schools the correlation was as low as .02, due largely to decreased variability in entering classes. "[T]he range among admitted students becomes restricted relative to the applicant pool." *Id.* at 4. "Because there is less variability in the scores of admitted students than in the scores of all applicants, correlations are smaller than they would have been had the class been admitted randomly from the total applicant pool." *See id.* The situation we have just described is known to statisticians as *attenuation*.

school relied on other (perhaps “political”) criteria. But such a charge would be the result of a poor understanding of the way relationships among variables are affected by high selectivity of the group being studied. The Lindgren study almost certainly suffers from just such a misunderstanding.

Let us return to our tour of the Lindgren study with the aid of Table 1. Why the study did not focus on cell D (the presidents’ appellate nominees) is never explained. But Professor Lindgren does explain why it did not focus on cell F (all confirmed appellate nominees). The study eliminates nominees with prior judicial experience from its core analyses because: “Bush appellate appointees who were lower court judges appear to have been equally treated compared to Clinton appointees.”²¹ That does not really explain why the sample was split. When the entire data set of 108 judges (cell F) is analyzed together, no significant difference is found between the ABA ratings of Bush and Clinton judges. The details of this finding, not reported by the Lindgren study, are provided in the margin.²² And what about those 55% of successful nominees who were analyzed by the Lindgren study and found to have been “equally treated”? Rather than saying: “here is

²¹ Lindgren, *supra* note 2, at 6.

²² What would have been learned if the Lindgren study had analyzed its entire data set of 108 judges confirmed by the Senate (all of the people in cell H), without controlling for any variables that might explain the assumed gap between ABA evaluations of Bush versus Clinton judges? Put differently, is there a gap to explain in the first place? We conducted this analysis counting as receiving positive ABA evaluations those rated “Well Qualified” (whether determined through unanimous or majority votes) versus all lesser ratings. The result is non-significant. Fisher’s Exact Probability Test, *p* (two-tailed) = .202. In other words, there is no difference between the ABA ratings given to Bush and Clinton nominees.

Although statisticians conventionally use the more conservative two-tailed test whether or not they have good grounds for believing the outcome will go in one direction rather than another, imagine that the hypothesis being explored in the Lindgren study were derived from a well-developed theory or from a dependable body of data, and therefore one could begin to justify using a one-tailed test. Using such a test, there still is no difference between the likelihood that a Bush or a Clinton judge will have earned a “Well Qualified” rating from the ABA. Fisher’s Exact Probability Test, *p* (one-tailed) = .127.

Suppose we use the more extreme measure of positive ABA ratings favored by the Lindgren study and count as a “Well Qualified” rating only those resulting from a unanimous vote of the committee. See the discussion of this choice between measures, *infra* notes 26-32 and accompanying text. Using this measure, Fisher’s Exact Probability Test, *p* (two-tailed) = .112 (still non-significant). Or, for the one-tailed test, Fisher’s Exact Probability Test, *p* (one-tailed) = .064. Finally, a marginally significant difference.

In sum, using “all” 108 judges, there is no way to find a statistically significant difference between ratings of Bush and Clinton nominees. If the limits of statistical reasonableness are pushed as far as possible, a marginally significant difference can be “found.” This suggests why the most attractive strategy was to remove 55% of the judges from the data set and move the analysis on from there.

some evidence that refutes the hypothesis of ABA bias,” the study instead sets aside those 55% of successful nominees and moves on to the next subset (cell H). The problem is not only that the total sample was not analyzed, nor that it was further subdivided. The problem is that the group in which no bias was found was discarded as being no longer relevant to testing the study’s hypothesis. Rather than evidence of non-bias, it becomes non-evidence.²³

What is to be made of the study’s finding that, for those nominees with prior judicial experience, there is no evidence of bias?²⁴ Doesn’t this undermine the hypothesis of political bias? Nominees who have been judges, more so than those without prior judicial experience, have records that might reveal their thinking more clearly and thus lead partisans of one party to prefer them and partisans of the opposite persuasion to disapprove of them. To that improved clarity on how the

²³ Subdividing and discarding data as one explores one’s data, and then basing one’s conclusions on the portion of the data most consistent with one’s hypothesis, is widely considered to be an unsound research practice. Relatedly, see *supra* note 22.

²⁴ What Professor Lindgren actually tells the reader is: “Bush appellate appointees who were lower court judges appear to have been equally treated compared to Clinton appointees. Indeed, Bush nominees got an insignificant advantage.” Lindgren, *supra* note 2, at 6. He repeats this basic point three more times in the paper, each time emphasizing in one way or another that it is not a real tilt in favor of President Bush’s nominees, because the “differences” or “advantage” are “not statistically significant” or “insignificant.” “Indeed, Bush nominees got an insignificant advantage.” *Id.* “The ABA showed no substantial differences one way or the other in evaluating candidates who were former judges; Bush lower court judges fared about as well as Clinton lower court judges in ABA ratings for open federal appellate posts – in some statistical models insignificantly better.” *Id.* “Bush nominees fare roughly as well and sometimes even better than Clinton nominees (though the differences are not large enough to be statistically significant.)” *Id.* at 28. This last quotation raises another statistical error. Magnitude of effect is a reason, not the *only* possible reason that some observed differences are not statistically significant.

There is a small problem with this and a large problem. The small problem is that Professor Lindgren’s analysis does not reflect an understanding of the purpose and meaning of significance testing. To put it simply, to say that there is a difference, but that it is not significant, is incoherent. A difference that is not significant is not a difference. It is noise, that is, random error. That’s why the test was done: to assess whether one is looking at a difference or at noise.

The large problem is that elsewhere in the article Professor Lindgren has told readers that the data he is dealing with comprise the complete population and therefore statistical significance tests are irrelevant. See Lindgren, *supra* note 2, at 10. Any difference is a difference. If so, a tilt favoring Bush nominees is a tilt favoring Bush nominees. Period. There is no minimizing the point.

It appears that when the article encounters a difference that is inconsistent with its hypothesis, that difference becomes a non-difference ‘because – the statistical test is non-significant.’ But when it encounters a difference that is consistent with its hypothesis, that is treated as real because – ‘it is a population we are dealing with’ and then significance tests are ‘meaningless.’

Although Professor Lindgren is continually inconsistent about whether he is dealing with a population that does not need significance testing or a sample that does, the most reasonable way to conceive the data in his study is as samples that do require significance testing. Later in the paper we explain this further. See *infra* notes 67-71 and accompanying text.

nominee thinks, add the fact that judgments such as those being made by the ABA can hide behind mountains of background information, conflicting outside opinions, subjectivity and ambiguity. If bias is operating, these are cases in which it might easily assert itself. If nothing else, if there is strong evidence *falsifying* the bias hypothesis (from the 59 nominees with prior judicial experience) alongside Professor Lindgren's assertion of strong evidence *supporting* the bias hypothesis (from the 49 nominees without prior judicial experience) a major puzzle has presented itself: the data are contradictory.

But the Lindgren study does not pause over this data-based puzzle. Professor Lindgren's faith in the hypothesis of political bias is unshaken even though his own data refute it. The hypothesis of bias is kept safe from the threat of counter-evidence by the speculation that prior judicial experience served as a silver bullet that magically and completely protected nominees from the bias that is presumed to exist. Professor Lindgren does not disown his findings that the ABA treated the majority of Bush and Clinton nominees "equally." But he explains this as resulting from biased evaluators being stopped in their tracks from doing political evil by the simple expedient of giving them nominees with prior judicial experience. The only evidence or reasoning Professor Lindgren offers in support of his interpretation is a conversation with a colleague.²⁵ For the Lindgren study, the hypothesis of political bias can withstand even the clearest falsifying data.

The reader might pause here and consider again how small the numbers are in cell H. The entire edifice of the study rests on this

²⁵ See Lindgren, *supra* note 2, at 28-29. "These results are consistent with reports from some participants in the process of judicial selection. Stephen Calabresi, a Northwestern law professor who was involved in judicial selection in the Reagan-Bush era, commented, 'After the Democrats regained control of the Senate in 1986, we made determined efforts to pick sitting federal or state judges for appellate court vacancies. We assumed that lower court judges would have an easier time being reviewed by the ABA.' Asked if this had an effect on who was picked, Calabresi said, 'Definitely, yes.'" *Id.*

That exchange is worth re-reading. It tells us of a change in White House selection strategy. It tells us nothing about the issue here, which is the decision-making of the ABA evaluators. The study in fact makes no effort to do any empirical testing on its new sub-hypothesis that (1) assuming political bias exists, having lower court judicial experience, regardless of anything else, stops the political bias dead in its tracks. An alternative hypothesis is that (2) prior judicial experience, regardless of anything else, wins uniformly high ratings from ABA evaluators. Yet another hypothesis might be that (3) in the face of an unfriendly Senate, screeners decided to switch to nominees with the best credentials they could find (and give fewer rewards to political friends with weaker credentials). Any one of these, or others, might profitably be explored. But Professor Lindgren accepts the first as true without any testing or even any argument beyond the non-sequitur offered in its support.

sample of 18 judges nominated by President Bush and 31 judges nominated by President Clinton.

Before moving to its focus on these 49 successful appellate court nominees without prior judicial experience, which is where the study spends virtually all of its analysis and discussion, Professor Lindgren makes a final sort of subdivision. This involves the choice of dependent measure. How shall the ABA ratings be rendered for the purposes of analysis? Since bias often is subtle and hard to discern, one would expect that the most sensitive measure would be chosen in order to maximize the chances of detection, such as a scale of favorableness to a candidate. Instead, the study chose to categorize as favorable decisions by the ABA only those nominees whose evaluations of "Well Qualified" had been the products of a unanimous vote, and to put every other evaluation of any kind into the negative category. Since ABA evaluations of "Well Qualified" – whether the votes that led to that recommendation were unanimous or majority votes – are normally treated by the Senate as functionally equivalent,²⁶ then all ABA recommendations of "Well Qualified" (rather than only a subset of them) arguably should have constituted the positive category. Instead, Professor Lindgren chose to count some ABA recommendations of "Well Qualified"²⁷ as if they were the functional equivalent of a vote of "Not Qualified."

What is the consequence of this choice between counting all ABA "Well Qualified" recommendations the same and counting as positive only the ones that were the products of unanimous votes? If the former choice had been made, the entire premise of the Lindgren study would have collapsed.²⁸ If one chooses the unanimous-or-nothing option, then one will conclude that Clinton nominees received the ABA's "Well Qualified" rating more often than Bush nominees did. But if one chooses to include as winning ABA approval those nominees whose ABA evaluations of "Well Qualified" were by majority as well as by unanimous votes, then the fundamental difference in ABA evaluations

²⁶ A former chair of the ABA Standing Committee on Federal Judiciary, Ronald Olsen, could recall no instance in which a divided "Well Qualified" vote had had any effect on the outcome of the Senate evaluation of a judicial nominee. Personal communication between Mr. Olsen and Stephan Landsman, shared with us by Mr. Landsman, October, 2001.

²⁷ Namely, the ones resulting from less than unanimous votes.

²⁸ That is, there would be no point in conducting a study to try to account for the apparent bias, because there would be no apparent bias.

between Bush and Clinton nominees evaporates into statistical non-significance.

To properly understand this point, one must understand that the purpose of statistical significance testing is to help answer the question of whether two measures which superficially appear to be different reflect a real difference or are merely illusory (i.e., noise, sampling error). If one chooses the unanimous-or-nothing option, then one can properly say that the percentage of Clinton nominees in the sample who received the ABA's "highest rating"²⁹ is greater than the percentage of Bush nominees who did (65% v. 17%) – not only in the sample but also in the actual population.³⁰ But if the measure includes all of those rated "Well Qualified," whether by majority or unanimous votes, then the percentage of Clinton nominees in the sample who received the ABA's "highest rating" can be said to be no different from the percentage of Bush nominees who did. (The apparent sample difference of 74% v. 56% must be said to be nothing more than sampling error.)³¹ Had the Lindgren study not chosen the extreme definition of an ABA "Well Qualified" rating there would be no basis at all for saying that Bush and Clinton nominees were treated differently.³²

III. THE PROBLEM OF CONSTRUCT UNDER-REPRESENTATION

Professor Lindgren correctly summarizes the ABA's criteria for evaluating judges. He notes that the evaluations are "directed to professional qualifications – integrity, professional competence and judicial temperament,"³³ and then continues with the following summary of ABA criteria:

²⁹ Lindgren, *supra* note 2, at 6.

³⁰ By Fisher's Exact Probability test, p (two-tailed) = .002 and p (one-tailed) = .001. Because the sample sizes on which these two values are based are so small, the standard error around these numbers will be quite large. But one can confidently reject the null hypothesis that there is in reality no difference between the rate at which the two sets of prospective nominees were rated "highly qualified."

³¹ The difference does not exceed what would be expected by chance. Fisher's Exact Probability Test, p (two-tailed) = .217 or p (one-tailed) = .153.

³² If it were found that the prospective nominees received different evaluations, the next question, of course, becomes whether the different ratings are justified by differences in the qualifications of the prospective nominees. In a rational process, better qualified nominees would receive higher ratings.

³³ Lindgren, *supra* note 2, at 3 (quoting *The American Bar Association Standing Committee on Federal Judiciary: What It Is and How It Works* 3 (March 1991)).

As for integrity, the ABA notes, “the prospective nominee’s character and general reputation in the legal community are investigated, as are his or her industry and diligence.” “Professional competence encompasses such qualities as intellectual capacity, judgment, writing and analytical ability, knowledge of the law and breadth of professional experience.” For appellate court nominees, the ABA states that “the Committee may place somewhat less emphasis on the importance of trial experience as a qualification,” yet appellate nominees “should possess an especially high degree of scholarship and academic talent and an unusual degree of overall excellence.” Regarding judicial temperament, the ABA states that the Committee “considers the prospective nominee’s compassion, decisiveness, openmindedness, sensitivity, courtesy, patience, freedom from bias, and commitment to equal justice.”³⁴

Professor Lindgren has rendered a fair summary, but it is useful to consider the actual process of applying these criteria. The nature of the ABA process (at least until the present administration) is described in a publication entitled *The ABA Standing Committee on Federal Judiciary – What It Is And How It Works*.³⁵ That publication states in relevant part:

After a judicial vacancy occurs and prior to any nomination to fill that vacancy, the Chair of the Committee receives the name of a prospective nominee for evaluation from the office of the Attorney General or the office of the White House Counsel. The investigation of the prospective nominee is usually assigned to the circuit member of the Committee in the judicial circuit in which the judicial vacancy exists. . . .

The Attorney General’s office sends to each prospective nominee a comprehensive ABA-designed questionnaire (called the “Personal Data Questionnaire”) that seeks wide-ranging information

³⁴ *Id.* at 4 (citations omitted).

³⁵ *The American Bar Association Standing Committee on Federal Judiciary – What It Is and How It Works* (1999).

related to fitness for judicial service. The responses are sent to the U.S. Department of Justice, the Committee Chair, and the circuit member. Receipt of the Personal Data Questionnaire is usually the starting point for the investigation. The circuit member makes extensive use of it in the investigation. . . .

The circuit member examines the legal writing of the prospective nominee and personally conducts extensive confidential interviews with those likely to have information regarding the integrity, professional competence, and judicial temperament of the prospective nominee including, where pertinent, federal and state judges, practicing lawyers in both private and government service, law school professors and deans, legal services and public interest lawyers, representatives of professional legal organizations, community leaders, and others who are in a position to evaluate the prospective nominee's integrity, professional competence and judicial temperament. In addition, comments from groups involved in the merit selection or evaluation of prospective nominees for the federal judiciary may be received and considered.

Interviews are conducted under an assurance of confidentiality. If information adverse to the prospective nominee is discovered, the circuit member will advise the prospective nominee of such information if he or she can do so without breaching the promise of confidentiality. If not, the Committee will not consider those facts in its evaluation unless the persons disclosing them authorize disclosure, or the information is otherwise known by, and discussed with, the prospective nominee. Sometimes a clear pattern emerges early in the interviews, and the investigation can be briskly concluded. In other cases, conflicting evaluations as to professional competence may be received, or questions may arise as to integrity or temperament. In those instances, the circuit member pursues the leads and problems as necessary to reach a fair and accurate assessment of the prospective nominee. This may

involve a large number of interviews as well as the examination of transcripts and other relevant records.

A meeting of the circuit member, and in appropriate cases one or more other members of the Committee, is held with the prospective nominee. The circuit member discusses with the prospective nominee his or her qualifications for a judgeship. During the interview the circuit member also raises any adverse information discovered during the investigation and discusses it with the prospective nominee. The prospective nominee is given a full opportunity to explain the matter and to provide any additional information bearing on it. The circuit member may need to conduct additional interviews in order to complete this stage of the investigation. . . .

The circuit member prepares a written informal report to the Chair containing a description of the prospective nominee's background, summaries of all interviews conducted, including the interview with the prospective nominee, an evaluation of the prospective nominee's qualifications, and a recommended rating. After receiving the informal report and discussing it with the circuit member, the Chair discusses the informal report with the Attorney General's office. While protecting the confidentiality of those interviewed, the Chair passes on the substance of the report and gives a tentative evaluation – indicating that, if a formal report is requested, the prospective nominee will probably be found Well Qualified, Qualified or Not Qualified.

If the office of the Attorney General so requests, the circuit member prepares a formal or final report. The circuit member then sends the written formal report to all members of the Committee together with the response to the Personal Data Questionnaire and copies of any other relevant materials. After studying the formal report and its enclosures, each member sends a vote to the Chair. If questions are raised, the

Committee may discuss the prospective nominee by telephone conference call or at a meeting.³⁶

Any study of the evaluation of judicial nominees soon confronts the problems inherent in capturing and measuring whatever it is that is meant by “judicial qualifications” or “evaluation criteria” so that the researcher can use measures of those as predictors of what the evaluators are doing. No study can rise above the efficacy of its measures. This brings us to the problem of whether the Lindgren measures capture the construct at issue, or whether they miss their target.

The problem of construct under-representation³⁷ is analogous to that of unrepresentative sampling. In the same way that a researcher could draw an unrepresentative sample of the population of people or things to be studied, a researcher also can draw an incomplete and unrepresentative sampling of the attributes that constitute a concept being studied. Conversely, just as good survey sampling captures a representative sample of the population of people or things one is trying to learn about, good measurement captures a representative sample of the domain of attributes of the thing one is trying to measure. Consider an elementary illustration of construct under-representation.

Suppose you wanted to buy a certain house, but suspected that it was in poor condition. Consequently, you hired a home inspector to find out. Suppose further that the home inspector did no more than to inspect the impressive paint job on the outside of the house, and based his overall assessment only on that, ignoring other systems of the house that were of importance to you, such as the condition of the furnace and the electrical wiring. You would immediately realize that he measured too few of the attributes that go into an assessment of the quality of a home, and that you could not rely on his conclusion. The inspector’s “study” failed to capture a reasonably complete or representative sample of the attributes of interest.

The Lindgren study suffers from this same shortcoming. Compare the concept domain reflected in the ABA judicial rating process that we

³⁶ *Id.* at 5-8.

³⁷ THOMAS D. COOK & DONALD T. CAMPBELL, *QUASI-EXPERIMENTATION: DESIGN AND ANALYSIS ISSUES FOR FIELD SETTINGS* 64-65 (1979). See generally, Norbert Schwartz, Robert Groves, and Howard Schuman, *Survey Methods*, in, *THE HANDBOOK OF SOCIAL PSYCHOLOGY*, (Daniel Fisk, Susan Gilbert & Gardner Lindzey, eds., 4th ed. 1998); and Charles M. Judd, *Measurement*, *in id.*

have quoted above at length³⁸ to the variables measured by the Federalist Society and analyzed by Professor Lindgren. While judges are supposed to be evaluated by the ABA with regard to their “integrity, professional competence and judicial temperament,” Professor Lindgren describes the failure of his study to capture all of the attributes relevant to the concept of judicial qualifications:

Most of the variables examined in this study are both important in themselves (such as having judicial or practice experience) and potentially important markers of intellectual or other sorts of ability (such as attending an elite law school or serving on law review). *Some of the ratings criteria, however, cannot be measured empirically, such as integrity and judicial temperament.*³⁹

Thus, Professor Lindgren acknowledges almost at the outset of the article that his study examines only one of three major criteria of interest to the ABA evaluation committee, namely professional competence.⁴⁰ At most, then, his findings speak to only a portion of what is sought to be evaluated in the ABA’s evaluation process, and – like the home inspection service that did not look at the whole house – likely results in conclusions that fall short of reflecting the reality of the situation of real interest.

Rather than respecting the limitations of its under-measurement of the domain of real interest and exercising appropriate caution, as the article unfolds Professor Lindgren’s incomplete measures come to be regarded more and more as if they were interchangeable with the full reality of judicial qualifications, or as if they *were* the reality of judicial qualifications.⁴¹ Although Professor Lindgren perfunctorily recognizes

³⁸ See *supra* notes 35-36 and accompanying text.

³⁹ Lindgren, *supra* note 2, at 4 (emphasis added).

⁴⁰ Professor Lindgren seeks to justify under-sampling of the domain of relevant measurement by suggesting that what he does not measure cannot be measured empirically. (“Some of the ratings criteria, however, cannot be measured empirically, such as integrity and judicial temperament.” *Id.*) Professor Lindgren’s argument merely confuses empirical measurement with easy measurement. There is no reason in principle why integrity and judicial temperament cannot be measured empirically, but to do so, and to do so well, requires a level of effort, sophistication, and resources that are not found in his limited study. However, the detailed blue print for such an analysis is spelled out by the ABA process, as discussed at *supra* notes 35-36 and accompanying text. Indeed, although the ABA does not call what it does “measurement” that is actually what its rating process is.

⁴¹ For examples: “relevant measurable credentials,” Lindgren, *supra* note 2, at 18, “extensive

the need for caution, given the limitations of his measures – “[O]ne should always be cautious in drawing conclusions, especially where there are important unmeasured variables not in the models, such as integrity or judicial temperament”⁴² – at no other point does the article conform to its own wise counsel.

In short, most of what needs to be measured is not measured. Professor Lindgren sees this problem momentarily, but does not pause long over it or its consequences.

IV. THE PROBLEM OF PROXY VARIABLES

But for the moment let us assume, along with the Lindgren study, that its more limited definition of judicial qualifications is the proper notion of what it means to be a promising jurist. The next question is how well the particular measures used in the study capture the limited conceptual definition chosen for the study. This brings us to one of the most serious, pervasive, and easily understood problems of the study: the choice of variables employed to represent the general concept of “professional competence.”

Consider the measures gathered by the Federalist Society. The flaws in those measures are readily apparent. Yet these measures are not questioned or qualified in Professor Lindgren’s article. The Federalist Society data contain six measures of “professional competence.” These involve whether the nominee has

served as a private practitioner;
served as a government lawyer;
already served as a judge;
attended a law school ranked as one of the ten best in the
current rankings from *U.S. News and World Report*;
served on a law review while in law school; and
served as a law clerk to a federal judge.

The primary observation about these measures is that they are what social scientists call “proxy” variables.⁴³ In the Lindgren study they are

relevant experience,” *id.*, “appointee with good credentials,” *id.*, “candidates with no measured credentials,” *id.* at 20, “Despite having no better measured credentials than Bush nominees, the Clinton nominees were rated as more qualified.” *Id.* at 28.

⁴² *Id.* at 26.

⁴³ Proxy variables are used in a wide array of settings for social science and economics research

stand-ins for whatever we might really want to measure in the name of professional competence. In other words, they are substitutes for more direct measures of competence, such as a close examination of a nominee's written and oral work, the collective judgments of peers, explicit tests of ability, or other such possibilities that are reflected in the ABA's statement.⁴⁴ Proxy variables are only as good as the degree to which they reflect the measures for which they are substitutes.⁴⁵

Let us look at each of these proxy measures, and think about how likely they are to capture what a reasonable lawyer or president or senator is likely to mean by "professional competence."

Proxy Variable 1: Prior Service as a Judge

Consider the criterion of having served as a judge. This would seem to be one of the best of the proxy variables. Professor Lindgren asserts: "Practically, the most important credential for being a judge is already being a judge."⁴⁶ Arguably, having been a judge of some kind might seem to be good preparation for being a good appellate judge. But an argument needs to be made – if hard evidence is not going to be offered – in support of the notion that a correlation is presumed to exist between prior judicial experience and performance as a federal appellate judge. It is not enough merely to assert this as a conclusion. That lower court experience is the "most important" preparation for appellate performance is not as obvious as it seems at first blush.

In and of itself, having served as a judge tells us nothing except that the person had successfully passed some other confirmation process and had then worn a black robe at work. It does not tell us how good a judge the nominee has been or will be. Such a determination might require an examination of the legal and judicial acumen reflected in the judge's written opinions, an assessment by peer judges and knowledgeable lawyers, or other measures. More important for

and appear in many forms when social and economic research is used in litigation. Some of the problems with proxy variables are discussed in *THE EVOLVING ROLE OF STATISTICAL ASSESSMENTS AS EVIDENCE IN THE COURTS* 24, 103, 116, 185-86, 246 (Stephen E. Fienberg, ed., 1989). Additional discussion can be found in Peter Stahlecker & Götz Trenkler, *Some Further Results on the Use of Proxy Variables in Prediction*, 75 *REV. ECON. & STAT.* 707 (1993).

⁴⁴ See *supra* notes 35-36 and accompanying text.

⁴⁵ See, e.g., Fienberg, *supra* note 43. This is also known as "operationalization," which is discussed in elementary texts on social and psychological research. See EARL BABBIE, *THE PRACTICE OF SOCIAL RESEARCH* (6th ed. 1992) at 113-129; Charles Judd, Eliot Smith & Louise Kidder, *RESEARCH METHODS IN SOCIAL RELATIONS* 42-49 (6th ed. 1991).

⁴⁶ Lindgren, *supra* note 2, at 6.

purposes of research and reality – as opposed to easy assumptions – it does not tell us whether, in general, persons who “have been judges” make better appellate judges than persons who have not been judges, but have all of the other qualifications.⁴⁷ Professor Lindgren appears to sense the shortcoming of the argument and tries to shore up his assumption by saying that it is unlikely that any president would knowingly nominate people who had been bad judges.⁴⁸ This certainly is a reassuring thought, even if it is a rather back-handed compliment to the nominees. But it does not add support to the empirical claim.⁴⁹

Federal district judges being considered for nomination to the appellate bench would have the advantage of having previously gone through a very similar process for their ascent to the trial bench. Indeed, federal district court judges have previously been screened for professional competence as well as for integrity and judicial temperament. This does not necessarily hold as strongly for nominees who are elected or appointed state judges, but nevertheless, a president would almost certainly engage in screening for these characteristics to eliminate someone whose performance as a judge has been noticeably poor. On balance, we would expect that nominees with prior judicial experience would fare well in the ABA ratings. And generally they do.⁵⁰ As already noted, however, none of this has anything to do with the assumed empirical relationship between prior judicial experience and superior performance as an appellate judge, which is presumably why the variable was chosen by the Federalist Society in the first place.⁵¹

⁴⁷ That, after all, is the test of whether prior judicial experience is a useful predictor of performance as an appellate judge: As between two groups of prospective nominees, equal in all respects except that one group has prior judicial experience and the other group does not, does the former group turn out to be better appellate judges than the latter group?

⁴⁸ *Id.* Interestingly, the Federalist data set contains a few examples from both the Bush and Clinton nominees of judicially-experienced persons who received split “Qualified/ Not Qualified” ratings from the ABA and some who received only a “Qualified” rating.

⁴⁹ To remind the reader: The unstated empirical assumption hiding behind this variable is that, in general, persons who “have been judges” make better appellate judges than persons who have not been judges (yet have all of the other qualifications). Of course, prior experience as a judge could include nominees who served on state courts of appeals. In this instance the correspondence between past and future judicial experience might be quite good. However, these fine distinctions also are absent from Professor Lindgren’s discussion and his data set.

⁵⁰ See *supra* note 24.

⁵¹ Because this point is so easily lost in the din of the refrain that being-a-judge-prepares-you-to-be-a-judge, we will say it one more time: The unstated empirical assumption hiding behind this variable is that, in general, persons who “have been judges” make better appellate judges than persons who have not been judges (yet have all of the other qualifications). And neither evidence nor argument is offered to support this claim.

The skills, habits of mind, and work style required are quite different for the job of a trial, as opposed to an appellate judge. There is no evidence or argument in the article to support the assumption that the former is good preparation for the latter. Indeed, given the differences in the nature of trial and appellate judicial work, one could as readily believe that being good at the former is inversely related to being good at the latter.

Interestingly, although President Bush had a far larger pool of active federal district judges from which to choose,⁵² there is no difference in the proportion of nominees with prior judicial experience nominated by the two different administrations.⁵³ This lack of difference would seem to be quite surprising, given what Professor Lindgren has suggested about the importance of prior judicial experience and the difference in the size of the respective pools of district court judges. Yet the study pays no attention to this fact or what it means for the hypothesis being studied.

Proxy Variables 2 and 3: Private Practitioner and/or Government Lawyer

Does serving as a private practitioner or a government lawyer tell us anything about competence? As with the predictor variable of prior judicial experience, the implicit claim is that persons who have one or both of these attributes will make better appellate judges than persons who lack these attributes. Also, as before, no evidence is offered in the article, nor even an argument made for why the asserted relationship should be expected to exist. And with these variables new problems present themselves.

The first problem is that the study offers no operational definition of "private practice" or "government lawyer." Stating operational definitions of variables is standard practice in empirical social science research.⁵⁴ Here, the reader is left to guess what these measures are and what they mean. At least with prior judicial experience, we can be

⁵² When President Clinton arrived in office, 76% of the active federal district judges had been appointed by Republican presidents. See Sheldon Goldman (1993), *supra* note 7, at Table 6 (showing that of 560 active federal district judges, 423 had been appointed by Republican presidents and 137 had been appointed by Democratic presidents).

⁵³ The numbers are as follows: 57% of successful Bush nominees had prior judicial experience versus 53% of Clinton nominees. Fisher's Exact Test, *p* (two-tailed) = .697 and *p* (one-tailed) = .413.

⁵⁴ See Marilynn Brewer, *Research Design and Issues of Validity*, in HANDBOOK OF RESEARCH METHODS IN SOCIAL AND PERSONALITY PSYCHOLOGY 6-10 (Harry T. Reis & Charles M. Judd, eds., 2000). See also Judd et al. *supra* note 37, at 43-49.

reasonably confident that we will know it when we see it and that others will concur with our intuitions.⁵⁵

But does the “private practice” of law include solo practice, small firm practice, criminal defense work, and corporate counsel – or only large white shoe law firms? If all of those constitute “private practice,” what is the rationale for claiming this measure is a marker of special ability? If only the latter is “private practice,” such employment may be in large part a reflection of an elite law school degree and law review membership and is redundant with those measures.

The same question can be asked about being a “government lawyer.” What is its definition? What government legal jobs are defined in or out of this measure? Is every lawyer whose paycheck comes from the government, at any level of government (from local to federal) a government lawyer? What is the rationale for that as a marker of special preparedness for the appellate bench? The broader the category, the harder it is to develop a rationale for it. The narrower the category, the more important it is to specify what the variable includes, what the rationale is for that definition, and what empirical evidence or arguments support the rationale.

Whatever the definition of these terms, at least one of them is rendered almost totally useless because it has so little variation.⁵⁶ A variable that is highly skewed can disclose no relationships. Take another elementary example. Suppose researchers sought to identify correlates of high grades in college, and one of the measures they used was whether students owned the assigned books in a course. If all or nearly all of the students did own the books, the research would find little if any correlation between having the books and the grades achieved in the courses. An unthinking researcher might conclude that how well students did in their courses was unrelated to whether or not they possessed the books assigned in the courses.

The Lindgren study has done much the same thing with the variable of private practice experience. According to the database, 93 of the 108 nominees had private practice experience. Consequently, this variable could have had little or no relationship with the ABA ratings.

⁵⁵ One could ask, however: Does the Lindgren/Federalist definition of judge include administrative law judges, justices of the peace, municipal judges, small claims court judges, judges of tribal courts, etc.?

⁵⁶ In other words, that the great majority of the judges have the same score on the variable.

Let us be clear on the implications of including the “private practice” variable in the study and retaining it in spite of its marked skewness. This variable was pre-ordained to show virtually no relationship between ABA ratings and whether one did or did not have the attribute. Thus, when the Lindgren study concludes that the ABA ratings cannot be accounted for by – or were made in almost complete disregard of – whether a nominee had private practice experience or not, that conclusion tells us a lot less about the ABA’s rating behavior than it tells us about the employment by the study of a variable that could hardly produce any “finding” other than the one it did.

Put most simply: why would anyone think that these two variables would correlate with being an excellent federal appellate judge? Note that we are not saying that a person lacking any legal background and legal work experience could be a good appellate judge. Some background in thinking long and hard about the law, and in appellate judge-craft is necessary, though not sufficient. But what sort of correlation could such broad and common measures produce? Undoubtedly very small ones. And that is what the Lindgren study finds. However, the “finding” does not reveal that the ABA pays no attention to legal occupational experience.⁵⁷ Rather, the measures used in the Lindgren study are by their nature largely incapable of finding such a relationship.

If these two legal occupation settings – private and government practice – are as broad as they appear, how many legal jobs do not fall into one or the other of them? One such job is that of law professor. By this study’s lights, then, it appears that one of the few law jobs that is not a “marker for intellectual ability” and depth of legal knowledge is that of law professor.⁵⁸

Proxy Variable 4: Elite Law School

The measure of “elite law school” has the virtue of being the only measure that actually is defined by the study. An elite law school is defined as one ranked in the top ten by *U.S. News and World Report* in its “current rankings.”⁵⁹ But the “elite law school” criterion deserves special scrutiny.

⁵⁷ Maybe the ABA does and maybe it doesn’t. The present authors can remain totally agnostic on the question.

⁵⁸ Only a few of the nominees in the Lindgren/Federalist data set had neither private practice nor government practice experience.

⁵⁹ Presumably that means 2001.

Though the rankings may be popular and entertaining, thoughtful educators and researchers have serious concerns that these rankings are poor indicators of the quality of a legal education.⁶⁰ Moreover, nominees who attended the elite law schools of their day in past decades would now be mis-categorized as graduates of non-elite schools if their schools slipped out of the top ten. And vice versa. Why were the *U.S. News* rankings used despite serious challenges to the validity and stability of those rankings as measures of school quality? Why were rankings from the wrong decade used? After all, most judicial nominees to appellate courts graduated decades before being appointed.⁶¹ The answer obviously is that the current rankings were handy, not that they were the best measure available. For the moment, though, let us accept them at face value and consider them as a meaningful predictor to see where they take us.

If graduating from one of the top ten law schools, as compared to all the rest, is an accurate predictor of appellate judging ability, as the Lindgren study asserts, then the alumni of Northwestern University Law School, Professor Lindgren's institution, would be among those not expected to perform well as federal appellate judges since Northwestern was ranked only 13.⁶² No doubt the same unhappy prediction would be made for seven of the 20 most recent appointees to the U.S. Supreme Court, who graduated from law schools other than Professor Lindgren's top ten.⁶³

The point is not that it is unreasonable to assume some correlation between the likelihood of being a first rate appellate judge and the law school pecking order – whether the correlation reflects the quality of education received at the school or the self-sorting and selection process that places highly qualified students into those schools. The point here is that the measure is so crude. The Lindgren study took

⁶⁰ See Brian Leiter, *Measuring the Academic Distinction of Law Faculties*, 29 J. OF LEGAL STUD. 451 (2000); Richard Schmalbeck, *The Durability of Law School Reputation*, 48 J. OF LEGAL EDUC. 568 (1998).

⁶¹ Indeed, Bush appointments began after he was elected president in 1988. Assume conservatively a minimum of two decades following graduation from law school was the norm for nomination, i.e. 1968. As a measure of law school standings the "current" rankings are more than 30 years out of date.

⁶² *Best Graduate Schools 2001*, U.S. NEWS & WORLD REP., April 9, 2001.

⁶³ There is yet another way to think about the shortcomings of these law school categorizations that may help put this measure into perspective. By the reasoning of the Lindgren study, the top graduate from a Northwestern Law School class would be coded as not as qualified to be an appellate judge as the lowest ranked member of the Harvard graduating class.

ranks and reduced them to two categories: top ten versus all the rest. Valuable information was thrown away, such as the ability to distinguish a graduate of school #11 from a graduate of school #175.

There is no good research or statistical or analytical reason to have thrown away so much information in this fashion, and the effect on the study's analysis and conclusions can be pernicious. The manner in which the Lindgren study measured school quality reduced the correlation between school rank and potential as a judge, thereby depressing the ability of this measure to explain differences between ABA ratings of Clinton and Bush nominees on proper grounds.⁶⁴ And so, when the study concludes that it cannot account for the difference in ratings between Clinton nominees and Bush nominees, it is telling us about the effects of its own imperfect measures, not about the behavior of the ABA Committee.⁶⁵

Proxy Variables 5 and 6: Law Review Service and Judicial Clerkship

The criteria of serving on law review and having a judicial clerkship are subject to problems similar to those we already have seen. The definition is never given. What counts as an editorship? Do secondary as well as the principal law journals count? The larger problem is the assumption that people who served on law reviews or as judicial clerks are, in general, going to be better appellate judges than people who did not engage in such activities. The assumption is an easy one to make, but no argument is made on its behalf and no data are presented in support of it. It may be true, but it may not be true.

⁶⁴ As an illustration, we conducted this simple analysis. We created a list of ranks 1 through 175. Each of these was paired with a hypothetical measure of judicial potential (on a 9-point scale) which assumed that no school that ranked lower in the pecking order would ever produce graduates with more potential than a school above it in the pecking order, though of course many closely ranked schools had the same scores on the 9-point scale. The correlation between these ranks and the hypothetical measure of judicial quality was .989. We then converted the rank information just as the Lindgren study did: the top ten all were coded as one group and the bottom 165 all were coded as a second group. The correlation then shrank to .386.

⁶⁵ Saying that the study's measures are defective is not to say that the ABA's evaluations are not politically biased. It says only that the failings of the Lindgren research should not be misinterpreted as failings of the ABA.

V. THE PROBLEM OF MAKING DISTINCTIONS WITHIN A HIGHLY SELECTED POPULATION

Suppose the president were sitting in the Oval Office, waiting for the arrival of the ten candidates from among whom he would choose a person for the position of White House science adviser. And suppose a helpful aide made the following suggestions to the president: The people likely to know the most about science have Ph.D.'s, have spent a good bit of their lives in laboratories, and have subscriptions to *Science*. This entirely plausible advice will turn out to be useless when the ten Nobel Laureates who passed the initial, intermediate, and final screenings arrive for their interviews.

The Lindgren study makes the same sorts of assumptions as the helpful White House aide – and no doubt with the same effect. The measures proposed to be relied on are plausible for the general population of lawyers. But those measures are incapable of making fine distinctions among the small elite corps of lawyers who have reached the threshold of a presidential nomination to the Courts of Appeals. Let us make this point somewhat more statistically.

Let us assume that the proxy variables used in the Lindgren study are highly correlated with the real measures of interest in the general population of lawyers within a year or two following their graduation (which is where half of the six proxy measures are focused: school rank, law review, clerkship). That does not mean those measures are correlated in the slightest for the far more highly selected population of seasoned lawyers given serious consideration for a nomination to the federal bench. It could be that all of the lawyers under consideration for circuit court seats have highly impressive records, so much so that the real measures of interest (and the ABA's ratings) do not correlate with whether the nominee attended an elite school, whether the nominee had been on law review, and whether the nominee had a judicial clerkship. That would be true if the screeners do nothing more than make sure that no one gets serious consideration unless he or she has high intelligence and is known to have a solid knowledge of law – regardless of what law school the prospective nominee attended, and regardless of whether the prospective nominee served on law review or had a clerkship.

Thus, even if there were not the major problems of validity of the proxy variables identified above in Section IV, the measures used in the Lindgren study fail to detect the finer distinctions that need to be made

by the White House and the ABA Committee because, for this elite group of lawyer nominees, everyone scores high on those measures (or, more correctly, the real traits that the Lindgren proxy measures purport to reflect). The Lindgren measures no longer correlate with the attributes of interest to the ABA or the President or the Senate. The measures are too crude for the fine-grained task at hand.⁶⁶

VI. FURTHER STATISTICAL AND TECHNICAL PROBLEMS

That we have subsumed certain problems under a single heading in this section and labeled them “technical” does not mean that they lack potentially serious consequences for the conclusions of the Lindgren study.

A. Are the Data of the Study a Sample or a Population?

Professor Lindgren asserts that he is dealing with a population, not a sample. This means that we do not need to infer from the study sample to the population of real interest, because the data of the study *are* the population of interest. The consequence of this is that we can disregard inferential statistics (typically, tests of significance) and take all descriptive findings and patterns as real.⁶⁷ He says:

Because this database contains all the nominated and confirmed judges, not a sample of them, statistical significance is literally meaningless. Significance testing is designed to assess the degree of confidence one can have that estimates obtained from a sample will approach the true parameters of the population from which they were drawn. Thus, little weight should be given to statistical significance here, since our means are

⁶⁶ Professor Lindgren could prove us wrong on this point simply by showing that a correlation does exist. That is, he could develop a measure (or measures) of the actual judicial performance of the 108 judges in this database. Then he might test whether the six predictors employed in his study correlate with observed judicial performance. We think the answer will be the following: no correlation reliably different from zero. This is simply because the six measures used in the study are too crude to detect the differences that make a difference for this highly selected group. Note that the issue we are discussing here is similar to our example of elite law schools having a restricted range of LSAT scores. *See supra* notes 19-20 and accompanying text.

⁶⁷ Assuming, of course, that everything else in the design of the study does not suffer from serious defects. Otherwise, another research methods and statistical analysis maxim comes into play: garbage in, garbage out.

exactly the population means. What are important here are the strength of the relationships and the explanatory power of the variables.⁶⁸

Notionally, Professor Lindgren is certainly correct in his statement about the role of inferential statistics. But he is incorrect when he says that he is dealing with a population. He also is inconsistent. At important moments in his study he acts as if he has been dealing with a sample. And those inconsistencies allow him to say things about his data, and the implications of the study, that could not properly be said if he were dealing with a population.

For example, as we have already seen, Professor Lindgren dismisses the fact that the ABA's ratings of President Bush's nominees with prior judicial experience were higher than those of President Clinton's nominees because, he tells us, the difference is statistically non-significant. However, if that is a population, then indeed "statistical significance is literally meaningless," the difference is noteworthy, and it must be confronted.⁶⁹

The reader should be clear that Professor Lindgren actually is dealing with a sample and not a population. The judges who are the focus of Professor Lindgren's analyses are the 49 people represented by cell H of Table 1. They are not the "population" of persons evaluated by the ABA. Because cell B is where the behavior of real interest to the study is to be found, Professor Lindgren has to be assuming that the people and data in cell H are representative of all the prospective nominees in cell B. That is a sample.⁷⁰

Moreover, if cell H truly were a population, and Professor Lindgren treated it as a population, his study would offer conclusions about this one particular set of 49 judges, confined to their time and circumstances, and would take care to caution readers that no more can be said beyond that. Yet, that is not what the study does at all. It seeks

⁶⁸ Lindgren, *supra* note 2, at 10-11.

⁶⁹ Of course, given the hypothesis of the study, the *lack* of a difference between the ABA ratings of Bush and Clinton nominees is extremely noteworthy, yet Professor Lindgren saw little of importance in it, beyond quickly explaining it away. See *supra* notes 22-23 and accompanying text.

⁷⁰ That they are in fact *representative* is extremely unlikely. They have gone through quite a lot of selection to get from cell B to cell H, so the ability to infer back is highly problematic. This is a major problem for the study. See *supra* notes 19-20 and accompanying text and Section V. Perhaps that is another reason Professor Lindgren benefits from referring to the 49 cell H judges as the "population" of interest: he is excused from having to deal with the problems of relying on an unrepresentative, biased, sample.

to test a general hypothesis about ABA political bias.⁷¹ The asserted findings are unhesitatingly generalized to ABA judicial evaluation behavior generally, to all of the appellate nominees, to district court nominees as well as to appellate nominees, and to future nominees as well as past nominees. All of that presupposes that the study is a sample of those other things, so that those other things can be inferred from the sample, and so that the findings from this one little sample can be generalized and applied to future ABA judicial evaluation activities. In other words, the Lindgren study tries to have it both ways.

B. The Problem of Drawing Inferences from an Incompletely Specified Model, Or, What We Might Have Learned from the Missing Variables

Textbook advice on the conduct of logistic regression is that if the model is not fully specified (that is, if all relevant predictor variables are not present) then the model may be defective and its implications misleading.⁷² As a practical matter, the requirement of full specification rarely is met. But the present study falls so far short of full specification that one ought to be quite cautious about the findings and consider them to be exploratory, preliminary, and tentative.

In other words, as in correlational analysis of virtually all kinds, the variables that aren't there create a major threat to the validity of the study's conclusions. This study measures six variables that are so intercorrelated (overlapping) as to be functionally fewer than six. And together these measured variables cover only a small portion of the domain of relevant attributes. The addition of variables not included in the study could change the picture dramatically.⁷³

Think again of the analogy to Title VII litigation.⁷⁴ A plaintiff offers a multivariate model suggesting that people have been discriminated against on the basis of an improper consideration. It is quite common for the defendant to respond by showing that if an additional variable

⁷¹ Not about bias toward these 49 judges at one point in history.

⁷² See generally Raymond E. Wright, *Logistic Regression*, in *READING AND UNDERSTANDING MULTIVARIATE STATISTICS* (Laurence G. Grimm & Paul R. Yarnold eds., 1995).

⁷³ Professor Lindgren is not unaware of this danger, but dismisses it with a wave of his hand: "If the ABA's highly subjective process is somehow consistent with ABA standards, it would have to be because differences in unmeasured criteria (e.g., integrity and judicial temperament) are so huge, important, and identifiable that they entirely swamp the measured criteria of judicial experience, law school background, and legal practice experience. Not only does that seem extraordinarily unlikely, but also whether such a sanguine state of affairs exists is a matter of faith, not evidence." Lindgren, *supra* note 2, at 26-27.

⁷⁴ See *supra* note 3.

or variables were added to the model, they could explain decisions as being based on reasonable and proper grounds, decisions that in the plaintiff's model had appeared to be the product of bias.⁷⁵

C. The Problem of Redundancy and Multicollinearity

The preceding problem concerned unmeasured variables. This problem concerns the fact that several of the measures that are used in the Lindgren study seem obviously to overlap in what they are measuring. Attending an elite law school and serving on law review are predictive of being appointed to a federal clerkship. If "private practice" were defined as being a member of a large Wall Street law firm, the first three variables are predictive of the fourth. The problem is not merely that there is redundancy causing the study to be measuring less than it seems to be measuring. At least as serious, predictor variables that are inter-correlated among themselves produce an analysis that suffers from unstable (untrustworthy) estimates. This is known as multicollinearity, and steps should be taken to avoid it.⁷⁶ The report of the Lindgren study discloses no attempt either to test for the problem or to try to cure it if it is present.

D. Insufficient Sample Size

Textbook advice on logistic regression is that sample sizes should be at least 50 for each predictor variable.⁷⁷ In this study that would mean a sample of 250-300 or more rather than the sample of 49 that was used. That defect risks producing results that are unstable and unreliable. This might explain why the highly processed findings of the Lindgren study are at war with simpler and more compelling findings based on the same data, as shown in the margin.⁷⁸

⁷⁵ If this were litigation instead of scholarship, Professor Lindgren might well say: "I've presented the factual case for the plaintiff. If variables other than the ones I've used could account for the bias I think I've demonstrated, it is up to the defendant to bring forward its own data and its own model."

⁷⁶ See Mark H. Licht, *Multiple Regression and Correlation*, in *READING AND UNDERSTANDING MULTIVARIATE STATISTICS* 45-48 (Laurence G. Grimm & Paul R. Yarnold eds., 1995).

⁷⁷ Wright, *supra* note 72.

⁷⁸ Let's look at some very simple counts of the data, using only the Lindgren/Federalist criteria. Consider the 49 successful nominees without prior judicial experience, the group regarding whom Professor Lindgren believes he found the greatest evidence of political bias. The weakest 19 of those 49 had only one or two of the Lindgren "measurable qualifications." Fifty-six percent of President Bush's nominees are found in this "least qualified" group compared to only 29% of President Clinton's nominees.

Let's turn next to the 18 "most qualified" nominees (those having 4 or 5 of the Lindgren

E. *The Problem of Coding Accuracy*

Another methodological problem arises in Professor Lindgren's acceptance of the Federalist Society's database. It is not unusual in social science research to use databases collected by others, but one must always consider sources of error or bias in the data. In this particular instance there is cause for concern. To the extent that the Federalist Society had already asserted its faith in the hypothesis of ABA bias,⁷⁹ its own data coding might inadvertently have resulted in systematic errors supporting its hypothesis.⁸⁰ Good research practice would have been to double-check all of the data, ideally by using research assistants who are blind to the hypothesis.⁸¹

VII. "REMEDIES"

Having concluded that there is bias in the ABA ratings, Professor Lindgren proposes remedies for the future, all of which amount to making less or no use of the ABA evaluations of prospective judicial nominees.⁸² His offer of "remedies" again reveals that Professor Lindgren is dealing with samples and not with populations.⁸³ With a population, there is nothing to generalize to; the study and the analyses are about the population and nothing more. But if one thinks that the data are predictive of other decisions at other times, such as the future,

criteria). Only 28% of President Bush's nominees are found in this group, compared to 42% of President Clinton's nominees.

These differences are not statistically significant. By Fisher's Exact Test, p (two-tailed) = .184 and p (one-tailed) = .114. This is fortunate for the Lindgren study, or Professor Lindgren would have been compelled to ask why the ABA was giving ratings that on average did not differ between Bush and Clinton nominees (*see supra* notes 22-23 and accompanying text) when it was obvious that Clinton's nominees were better qualified (according to the Lindgren "measurable criteria").

⁷⁹ Such expressions of belief can be found in back issues of the Federalist Society's *ABA Watch*, available at <www.fed-soc.org/abawatch.htm>.

⁸⁰ *See generally*, ROBERT ROSENTHAL, *EXPERIMENTER EFFECTS IN BEHAVIORAL RESEARCH* (rev. ed. 1976).

⁸¹ *Id.* Professor Lindgren reports that he double-checked about two-thirds of the data against various sources and corrected errors that he found in the data base. *See* Lindgren, *supra* note 2, at 1, n.1, 4. His having done so alleviated this problem to a greater or lesser extent.

⁸² Professor Lindgren suggests that academics would not "favor continuing to use the results of an evaluative process that seems to be so strongly biased," *id.* at 24, that "the ABA [should] cease rating judges until it can eliminate the apparent bias or show that no bias exists," *id.* at 25, advises "the Senate, the White House, and the press to de-emphasize or ignore entirely the ABA ratings because of probable political bias," *id.* at 25, and that perhaps "the ABA should consider withdrawing from the process of rating federal court nominees" altogether, *id.* at 25.

⁸³ This concept is dealt with in a confused and inconsistent way throughout Professor Lindgren's article, and we address it more fully *supra* at subsection VI.A.

then one is dealing with a sample. Accordingly, one should act as if one is dealing with a sample. But these are relatively minor points.

In his penultimate draft,⁸⁴ but deleted from the published article, Professor Lindgren put forward a remedy that borrowed from racial norming in school admissions and employment testing. We nonetheless address this remedy here because it shows so clearly why no remedy would be needed in the present context even if there were a confirmed problem. Professor Lindgren suggested that the asserted ABA bias could be “corrected” by certain statistical adjustments to the ABA’s ratings. Doing so would “level” the field, causing the nominees of the first Bush administration and the Clinton administration to be treated more equally.

But in this “discrimination” case, even if the charges of bias are true, the suggested remedy makes no sense because it misconceives the “hiring” process. Unlike the factory hiring or school admissions situations, regardless of the ABA ratings, each administration will make as many nominations as it can to fill open seats, and fill as many of those seats as the Senate confirms nominees. Assuming each president was considering equally qualified pools of prospective nominees, and assuming the ABA ratings were politically biased, the Lindgren study asserts no finding other than that one president’s nominees got inflated ratings (or that the other’s got deflated ratings). A linear transformation such as that changes nothing in the nomination process.

The problem, as perceived and asserted by the study, amounts to the following: One president is, in effect, advised by the ABA to send his A people to the Senate and leave his B and C people behind, while the other president is advised to send his B people to the Senate and leave his C and D people behind – while in reality one president’s A nominees were equal to the other president’s B nominees.⁸⁵ Assume the charge is true. Because this is a comparative (or rank ordering) process *within* each group of prospective nominees (not *between* groups of nominees), it still means that the best people from each president’s groups get the highest “grades” for their “class,” and to the degree that the White House accepts the advice, these are the people who move on to Senate consideration.⁸⁶

⁸⁴ On file with the editors.

⁸⁵ We are using A, B, C, and D “people” in the same sense that one would refer to A, B, C, or D “students.”

⁸⁶ That the hypothesized ABA bias, assuming it existed, did Bush nominees no harm and gave the Clinton nominees no advantage is perhaps reflected in the fact that in its four years the Bush

Our point is simply that a linear shift on the rating scale, when the rank orders remain the same, changes nothing in the end.

The kind of charge that would raise serious concerns, which the paper does not make, would be that the ABA ratings are so perverse that they downgraded Bush's better prospective nominees below his weaker nominees, thereby inverting the rank order, so that the President was being urged to nominate, and the Senate to confirm, the poorer nominees and to send the better nominees home. There is no reason why anyone, political friend or political foe, would want to urge nomination and confirmation to life tenure appellate judgeships of the worst nominees a president has to offer, while rejecting the best nominees. Nor does the study make any suggestion that this is happening. But we mention this because it is the only scenario (a non-linear shift, inverting the rank order) that would make a practical difference to the process.

VIII. CONCLUSIONS

The Lindgren study offers remarkably overheated conclusions⁸⁷ considering how thin, unstable, and contradictory are the foundations on which those conclusions stand.

Using all judges in the study's data set (100% of the sample), there is no statistically significant difference in the ratings of Clinton versus Bush judges. This strong evidence against the hypothesis of bias is not reported by the study at all.

For nominees with prior judicial experience (55% of the sample), the Lindgren study found no difference in ABA ratings of Bush versus Clinton nominees. This strong evidence against the hypothesis of bias was set aside and was all but ignored.

For nominees without prior judicial experience (45% of the sample), if ABA recommendations of "Well Qualified" are counted whether or not they are the products of unanimous or majority votes, again there is no difference in ABA ratings of Bush versus Clinton nominees.

administration had 47 (11.75 per year on average) nominees ascend to the appellate bench, while in its eight years, the Clinton administration had only 61 (7.65 per year on average) confirmed to the appeals courts. See Goldman et al., *supra* note 7.

⁸⁷ "[T]he patterns revealed in the data are consistent with a conclusion of strong political bias favoring Clinton nominees" and the "effect sizes are extremely large." Lindgren, *supra* note 2, at 26.

A difference is found only when the definition of "Well Qualified" is limited to those resulting from unanimous votes of the ABA committee and applied to nominees without prior judicial experience. There, at last, is the "difference" that the study purports to try to explain (and on failing to do so draws its harsh conclusions). But the failure to account for that "difference" is as likely to be found in the weaknesses of the study as in the supposed weaknesses of the ABA's evaluation process.

Though the study begins with a broad concern about the ABA evaluation process, it almost immediately turns its attention away from the pool of prospective nominees which constitute the subjects of the ABA's evaluations, away from prospective nominees for the district court, down past the smaller group the White House nominates, down even past the confirmed nominees, finally narrowing its focus to a subset of 49 seated judges.⁸⁸

The Lindgren study failed to capture most of the construct of judicial qualifications, and concedes that failure. Yet it proceeds to use its desiccated version as the one to test the real thing.

The more limited construct that it did try to measure, namely competence, was assessed by a handful of ill-defined and undefended (neither empirically nor theoretically) measures which plainly were chosen because they were easily acquired, not because they were the most appropriate measures for the task at hand.

The measures are defective in other ways. One is so skewed that it is incapable of correlating with ABA ratings or anything else. Because some of these predictors are correlated with each other, they not only are less informative than they appear to be, but the statistical models that are created out of them will also tend to produce unstable and misleading results. The measures are so crude that, while they quite plausibly would be useful for making distinctions among the general population of lawyers, they will yield little more than noise in any effort to use them to distinguish among an already highly selected, elite subgroup of lawyers.

The Lindgren study's most appropriate role would be as a preliminary, exploratory inquiry which, at most, reveals some tentative evidence supporting the bias hypothesis. That, in turn would have suggested the potential value of conducting more and better research directed at the hypothesis. But there is little if any justification for

⁸⁸ Those who did not have prior judicial experience.

reading the study to support its strongly expressed and extreme conclusions.

The most reasonable position one could hold, on present knowledge, is to remain agnostic on the issue of whether the ABA ratings are politically biased. Perhaps they are, perhaps they are not. But the Lindgren study casts little if any light on the question. No literate consumer of empirical research would rely on its conclusions for any serious purpose. It tells us nothing of value about the controversy.