STATISTICS IS A PLURAL WORD

Steven L. Willborn and Ramona L. Paetzold***

Responding to D. James Greiner, *Causal Inference in Civil Rights Litigation*, 122 HARV. L. REV. 533 (2008).

Proving causation in civil rights cases is tough. We are not always certain what causation means, we do not know what goes on inside people's heads, sometimes we are dealing with institutional actors, the relevant time period can be very long, and on and on. Statistics can be very useful in addressing these problems. It can help us describe a situation carefully and precisely and it can help us decide if a legal standard is met. But statistics is a homely tool, not a magic wand. Statistics is imperfect as proof of causation in the same way that every other type of proof is imperfect — it is messy, indirect, uncertain, and subject to varying interpretations. In addition, and perhaps most importantly, "statistics" is a plural word offering up a variety of ways to view data and, in turn, a variety of ways to hone in on the legal issue of causation. Hence, it is more a *toolbox* than a tool.

In a recent article in the *Harvard Law Review*, Professor James Greiner argued that a particular statistical technique (regression) is ill-suited and overused as a tool to prove causation in civil rights cases.[1] He suggested that another statistical technique (potential outcomes) should be used instead. We agree with much of what he said — that regression has serious limitations, that potential outcomes can be a useful technique, that courts should pay great attention to timing and detail in accepting statistical analysis. We disagree, however, with his claim that potential outcomes necessarily provides a superior way of examining data in civil rights cases. We also disagree that potential outcomes should be viewed as the universally preferred analysis. The best approach is to allow experts to use the full toolbox of statistical techniques so that the cathedral can be viewed from multiple angles. Statistics will be more valuable in civil rights cases when properly treated as a plural word.

In Part I, we consider Professor Greiner's claim that the potential outcomes approach is superior to regression. We like the potential outcomes approach and think it promises great value in some cases.

* Dean and Schmoker Professor of Law, University of Nebraska College of Law.
** Professor & Mays Research Fellow, Mays Business School, Texas A&M University.

[1] D. James Greiner, *Causal Inference in Civil Rights Litigation*, 122 HARV. L. REV. 533 (2008).

On the other hand, most of the problems Professor Greiner attributes to regression also apply to the potential outcomes approach. This is almost inevitable since potential outcomes would *use* some form of regression as part of the analysis in most (maybe all) civil rights cases. It is also inferior to regression in handling some issues that commonly arise in civil rights cases. In Part II, we examine Professor Greiner's claim that regression lacks an adequate framework for making causal inferences. We disagree with that claim, too. Certainly, regression cannot *prove* causation; no statistical method can do that, including the potential outcomes approach. Instead, regression calls on the same basic causal framework used throughout discrimination law and, indeed, the same basic causal framework used by potential outcomes. The issue here, really, is whether we want to hold statistical proof of discrimination to much higher standards of proof than are used elsewhere. We think Professor Greiner's article moves in that direction, and we think that's unfortunate.

I. REGRESSION, POTENTIAL OUTCOMES, AND EXPERT DISCRETION

In *Causal Inference*, Professor Greiner proceeds on two fronts: he mounts an aggressive critique of the predominant statistical technique used in civil rights litigation (regression) and he offers an alternative method (potential outcomes).

In beginning to address these linked arguments, it is important to distinguish between research design and statistical methodology. Research design comes first. In the civil rights context, the question is how do the data from the case arise? Do they arise from some form of experiment or quasi-experiment, or do they take some other form? The second question is how best to analyze that data statistically. Should we use regression analysis (and, if so, what type of regression), analysis of variance, structural equation modeling, survival analysis, potential outcomes analysis, or some other technique? The answer to the second question depends in part on the answer to the first, but only in part.

Professor Greiner's article may be misleading to the uninitiated because it tends to conflate research design and statistical methodology. But the debate here is only about statistical methodology, not about research design. The data in civil rights cases are not experimental in nature; they are observational. This means that the *design* giving rise to the data did not control for systematic variation adequately. As a result, the techniques used to analyze the data must attempt to control for those sources of variation statistically. Although Professor Greiner's potential outcomes approach may sound as though it creates an experimental design for the data, it is in fact a statistical approach for reducing sources of variation. As such, it joins a long line of other statistical approaches that attempt to do the same thing, just in different ways.

But even though the data in civil rights litigation are observational, it is worth thinking about the advantages of controlled experiments. The principal advantage is that in a true experiment it is theoretically possible to make causal inferences about outcomes because only one explanation of cause and effect should exist: with everything else held constant, the one manipulated factor (the "treatment") must have caused the observed difference in outcomes. In true experiments, all other possible explanations for the difference have been held constant by using complete randomization and "treatment" and "control" groups. Complete randomization means that the participants are selected at random, they are assigned to two different groups at random, and then the groups are assigned to the "treatment" and "control" conditions at random. Every participant is measured on an outcome and the average outcome for the treatment group is compared with the average outcome for the control group. If those two average outcomes are different from each other,[2] then it is inferred that the treatment *caused* this difference. Note that this is an inference based on the *research design*. The final step — the hypothesis testing to evaluate the difference between the treatment and control groups — is statistical, and statistical inferences in and of themselves cannot prove causation.

Even in controlled experiments, however, the validity of causal inferences is based on a set of assumptions. One such assumption is that randomization will render irrelevant all variables not measured as part of the study. For example, if a study looks at the effect on recovery of getting a drug treatment versus not getting the drug treatment and complete randomization is used as described above, then any differences in recovery will be inferred as being due to the drug treatment. In reality, however, the study participants will have differed on many measures — age, sex, socioeconomic class, race, attitudes, beliefs, etc. It is assumed that randomization renders these measures irrelevant because they should be distributed in the same way across the two groups — there should be no *systematic* variation on these measures between the two groups. But for a variety of reasons, randomization does not always accomplish this.[3] Thus, even for Professor Greiner's "gold standard" of controlled experimentation,[4] causation is not known

---

[2] This typically means that a test for statistical significance has been conducted to see if they are "significantly different" from each other. *See, e.g.*, RICHARD C. SPRINTHALL, BASIC STATISTICAL ANALYSIS 237 (8th ed. 2007).

[3] For example, randomization could, by chance, put all females in the treatment group and all males in the control group. If the drug's effectiveness were based on sex, causal inferences about the effectiveness of the treatment could not be distinguished from inferences based on sex.

[4] Greiner, *supra* note 1, at 564.

for sure. That is an important reason why replication is essential and virtually universal in scientific research. Multiple experiments are done over time and across a variety of settings to determine if the same results are always obtained. However, randomization and replication are clearly not possible in a given civil rights case.

Professor Greiner is not proposing that civil rights cases be handled by complete randomization, however, because he cannot. Instead, he is proposing a statistical methodology — potential outcomes — that attempts to simulate an experimental design by matching subjects to control for relevant sources of variation.[5] In this approach, the protected class status, such as being (perceived) female, is used as the treatment and being (perceived) male is used as the control. Note that this approach does not create a randomized experiment. First, the participants in the analysis are not selected at random: they are the relevant employees of the organization. Second, the set of employees available for "study" cannot randomly be split into two groups for study. Third, perceptions about their sex cannot be randomly assigned to them.[6] Thus, the potential outcomes approach does not begin to achieve the *actual* randomization necessary to a true experiment.[7]

The potential outcomes approach attempts to simulate a true experiment. In Greiner's example it does that by "matching" male and female[8] employees on variables describing them (covariates) to try to isolate what is "caused" by their sex. Even in his simple example, this matching process requires the statistical expert to make many complicated (and potentially controversial) choices. First, what are the variables to be used for matching?[9] Second, when are a male and a female "reasonably close" on those variables so as to constitute a match?[10] Third, what does the analyst do if there are multiple choices that are "reasonably close" to constitute a match?[11] Fourth, just because the individuals are reasonably close to be considered a match, does it

---

[5] Potential outcomes does not involve a "matched design" in the experimental sense, but merely simulates it.

[6] Professor Greiner acknowledges this. Greiner, *supra* note 1, at 576 n. 94. He uses sex as the treatment variable, which we and others (see his comment) view as problematic, but we do not discuss this point further.

[7] We therefore contest Professor Greiner's claim that the potential outcomes approach creates "mini-randomized experiments," even within pairs. *Id.* at 573.

[8] We drop the "perceived" aspect here, as does Professor Greiner.

[9] Professor Greiner notes this subjective selection of variables by the researcher as a problem with regression. It is. But potential outcomes has the same problem.

[10] Greiner, *supra* note 1, at 571. Sometimes data are missing on one or more of these variables. As a result, a protocol on how to deal with missing data must be adopted, just as in regression analysis.

[11] This is critical, since the outcomes (say, salary) for two "reasonably close" males may be quite different.

make sense to assign the male's salary to the female, or could there be other factors that suggest that this is unreasonable?

This process becomes even more complicated when there are a large number of variables or covariates to use for matching, as Professor Greiner noted. When this occurs (which would be in virtually every civil rights case), a "propensity score"[12] must be estimated to attempt to balance the covariates. Different assumptions can lead to different propensity scores, which in turn can lead to different conclusions about whether a sufficient balance has been obtained. The final selection of propensity scores affects the inferences to be drawn about the outcome, however, so different inferences about discrimination could result.[13] And, once again, just as in the case of randomized experiments, issues of joint causation or sequential causation (in the scientific sense) add to the complexity, and to the choices that must be made.[14]

In *Causal Inference*, Professor Greiner emphasizes the role of expert judgment and discretion when regression is used in civil rights cases, but minimizes it when he discusses potential outcomes analysis. As we indicated above, if he had engaged in a fuller description of the potential outcomes methodology, the central role of expert judgment in potential outcomes analysis would have been more evident. But his claim was not, and could not be, that no expert judgment is required by potential outcomes analysis. Instead, his was a comparative claim — potential outcomes provides less room for expert discretion (and mischief) than regression. For several reasons, we question that claim, too.

First, potential outcomes analysis requires the expert to "match" as many variables as the "firm used (or claims to have used) in its decisionmaking."[15] Inferences about outcomes are only as good as the covariates that have been included. A similar point is also true for regression — variables that are not included in the regression model are not controlled and their effects on the outcome cannot be ascertained. Consequently, both types of analyses require complex decisions about which variables to include.

---

[12] A propensity score is a numerical value that attempts to control for differences in the covariates between the treatment and control group. *See, e.g.*, Kosuke Imai & David A. van Dyk, *Causal Inference With General Treatment Regimes: Generalizing the Propensity Score*, 99 J. AM. STAT. ASS'N 854 (2004).

[13] A range of potential outcomes could be given, based on different assumptions made by the analyst. Of course, regression analysis can also lead to different outcomes regarding discrimination based on the analyst's different assumptions about the role of the covariates.

[14] For example, the problems associated with "intermediate outcomes" need to be explored carefully, but Professor Greiner does not address precisely how to incorporate these issues into the potential outcomes approach. *See* Greiner, *supra* note 1, at 565.

[15] *Id.* at 574.

Second, in cases that involve multiple variables (which, again, will be every civil rights case), the potential outcomes approach requires propensity scores to be estimated.[16] This is done by experimenting with statistical models until an acceptable level of balancing of covariates has occurred. Some form of regression — involving assumptions and choices by the expert about functional form — is typically used to accomplish this.[17] Hypothesis testing is then used to determine when adequate balancing has been achieved. Thus, even in the potential outcomes approach, inferences are based on an expert's subjective choices about a regression model and the presence or absence of statistical significance via hypothesis testing. In the typical regression approach to analyzing civil rights litigation data, these choices are all evident at the surface-level. In the potential outcomes approach, these choices and their effect on inferences may not be as transparent.[18]

On this point, Professor Greiner might counter that the potential outcomes approach is preferable because it involves balancing covariates without knowing anything about the outcome values (e.g., salaries).[19] Estimation of the propensity scores within the potential outcomes approach occurs before the expert needs to see any of the outcome data, and does not require that data for the analysis.[20] The implication is that having outcome data available to the expert automatically results in bias on the part of the expert. We doubt the implication from both sides. On the regression side, access to the outcomes data is relevant to determining the fit and appropriateness of the

[16] Professor Greiner only briefly mentions the need to balance multiple variables and therefore includes no discussion of the choices and processes involved. *Id.* at 574–75.

[17] A final set of propensity scores is selected when it appears that there are no statistically significant differences in the distribution of the covariates for strata created by the scores. This is determined by repeated hypothesis testing. This is very similar to the process for selecting the correct model specification in traditional regression analysis.

[18] This conflicts with Professor Greiner's oft-repeated claim that the potential outcomes approach will be more easily understood by judges and juries. Greiner, *supra* note 1, at 538, 557, 580 & 597. The choices made about these types of underlying issues will be part of the litigation, and they will be complicated.

[19] Professor Greiner supports this claim with a legal process point. Since the outcome values do not need to be known at the outset, the court should prohibit their disclosure until later in the litigation, coordinate communication between experts, and monitor the analysis in other ways. *Id.* at 579–80. We ran this idea by a small sampling of federal district judges who reacted to it with horror. The reasons ranged from doubts about their legal authority to manage the trial in this way, to concerns about whether it was a proper role for them even if they had the authority, to concerns about their expertise to engage in such tasks. This suggestion is a non-starter. Professor Greiner also suggests that civil rights litigants might "wish to ask their testifying expert to proceed via potential outcomes as opposed to regression." *Id.* at 580. We would suggest instead that civil rights litigants ask their experts to consider the full range of statistical options for analyzing the data in the case. We doubt that many civil rights litigators will be able to make the subtle judgments necessary to determine the best statistical approach. That, after all, is why they get statistical experts in the first place.

[20] *Id.* at 579.

model. Any impulse the expert has towards bias will be constrained both by the standard criteria used for assessing regression models and by the certain knowledge that his choices will be closely examined later in the litigation. On the potential outcomes side, if the analyst is inclined to bias, there will be ample opportunity to exercise it there, too, so minimizing one avenue is unlikely to be effective. But even here, the opportunity is not cut off. Even in potential outcomes analysis, outcome values may need to be accessed to deal appropriately with those outcome variables that are properly considered to be intermediate outcomes.

Covariates, by definition, precede treatment and outcomes. Intermediate outcomes may be affected by treatment, but they precede the outcome. Professor Greiner under-addressed the issue of separating covariates from intermediate outcomes.[21] On one hand, he dismisses too easily the possibility that some variables may be intermediate outcomes. For example, in discussing years of education, he asserts that it is reasonable to conclude that such a variable predates perception of race and therefore would be a covariate.[22] However, educational level may be perceived or valued differently once the employer is aware of an employee's race or gender. For example, an employer may discount the value of an educational level for a female applicant while not doing so for a male applicant.[23] Distinguishing covariates from intermediate outcomes is a complex and uncertain task.

On the other hand, Professor Greiner states that the effects of intermediate outcomes should not be controlled away.[24] In the potential outcomes approach, that means they should not be balanced. He does not say what to do with intermediate outcomes, however. How can the expert incorporate the full range of intermediate outcomes into the analysis? In the traditional regression approach, intermediate outcomes are important variables in the regression models.[25] Such vari-

[21] Note that he does describe the difficulties of separating the two as a "tug-of-war" that requires the expert to make assumptions "supported by reasonable judgment" about which variables are affected by treatment. Greiner, *supra* note 1, at 583. We highlight this point to make clear again that there is plenty of room for bias to invade the potential outcomes approach.

[22] *See id.* at 577; *see also id.* at 581.

[23] A number of experiments indicate that this is, sadly, a common phenomenon. *See* Christine Jolls, *Is There a Glass Ceiling?*, 25 HARV. WOMEN'S L.J. 1 (2002). More generally, we are increasingly coming to understand that implicit bias is a common and powerful phenomenon. *See* Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CAL. L. REV. 945 (2006).

[24] *See* Greiner, *supra* note 1, at 565.

[25] We disagree that regression methods "mask[]" these issues. *Id.* at 589. Due to explicit ways of testing for mediating variables, most regression methods explicitly force the expert to consider these issues. *See, e.g.*, Reuben M. Baron & David A. Kenny, *The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations*, 51 J. PERSONALITY & SOC. PSYCHOL. 1173 (1986).

ables are typically referred to as intervening or mediating variables, and a variety of statistical methods exist for testing their effects. Thus, in situations involving intermediate outcomes (that is, the majority of civil rights cases), regression provides standard approaches for analysis, that both improve the analysis and impose restraint on potential expert bias. The treatment of intermediate outcomes is less developed in the potential outcomes approach.

Even if the set of covariates can be clearly identified, the issue of which data to include in the analysis arises. Although the potential outcomes approach attempts to remove some sources of variability and bias, it may create others. It does this by creating a *hypothetical* set of data that is not the *actual* set of data. It discards some actual data on employees. If there is not a common range of values on covariates to match, then those male and female employees without matching counterparts are eliminated from consideration. It is not at all obvious that this makes sense. For example, suppose the only covariate to match is years of work experience — it is the only variable the firm claims to use in salary determination — and one female in the data set has only ten years of experience, while everyone else has twelve or more years. Professor Greiner would eliminate this female from consideration, regardless of her salary level.[26] Suppose she was paid $10,000 per year while the closest male (twelve years of work experience) was paid $40,000 and females with twelve years of work experience were paid at least $25,000. Could we seriously claim that her situation does not provide important evidence of salary discrimination? Should we exclude her from the remedy because no male was a close enough match? In some instances, the potential outcomes method can require discarding a large percentage, if not most, of the actual data. Thus, the inferences being made for a few employees can come at the expense of failing to consider the plight of many (or most) of the others. Professor Greiner calls this the "price to be paid," while acknowledging that discarding data will tend to reveal fewer instances in which a statistical inference of discrimination can be made.[27] We consider this a high price, and an unnecessary one.

Having said all this, we want to make clear that we think the potential outcomes approach is a useful way to evaluate data in civil rights cases. We do not seek to disparage the approach — it has its statistical merits and, hence, it deserves a space in the toolbox of available methods. But we do question the extent to which Professor Greiner underplays the amount of discretion and judgment required

[26] This is assuming that he does not consider ten to be sufficiently close to twelve to consider it a match. One must either estimate or impute a salary value for this female based on comparable male salaries or eliminate the female from consideration.

[27] Greiner, *supra* note 1, at 567.

for potential outcomes analysis and overplays it for regression. The use of statistical techniques is part art, part science. *All* statistical methods involve a host of underlying assumptions. In an ideal world, whenever a particular method is used, all of its underlying assumptions would be perfectly and fully met, especially in situations involving issues as important as civil rights. But in practice, with real-world data, this simply does not happen.[28] For this reason, among others, experts have to make choices. There is always an element of subjectivity when an expert decides which statistical approach is appropriate and how well the available data fit the model(s) based on that approach. These choices are informed by the expert's experience and judgment. Both the regression and the potential outcomes methods can be misused, or used well. Neither is exempt from subjectivity; both are susceptible to bias. What is important in both social science and litigation is that the expert reveal the choices that were made and the extent to which assumptions are met.

II. STATISTICS AND CAUSAL INFERENCE

Stepping back from these fairly technical arguments, we have a more general disagreement with Professor Greiner about the nature of causation in civil rights cases and, more generally, about the role of statistical proof in addressing causation issues.[29]

Causation is one of the law's most difficult concepts.[30] We agree wholeheartedly with Professor Greiner that statistical evidence must be "honed by an insistence on a definition of a causal effect."[31] We also agree that the complications of statistical analysis can sometimes cause courts to "lose track of what it is that they need to decide, of the questions they need to answer."[32] But, unfortunately, we do not think Professor Greiner does a good job of addressing the causation issue. For example, he argues that a significant benefit of his proposed potential outcomes approach is that it would resemble but-for causation,[33] even though that is not the proper standard for most civil rights litigation.[34] But our problem is more basic. Central to Professor Greiner's

---

[28] As Professor Greiner notes, "the ideal data are never available." *Id.* at 589.

[29] Or at least we think we do. Professor Greiner does not speak very directly about these more foundational issues.

[30] The most famous and influential work on the topic is H.L.A. HART & TONY HONORÉ, CAUSATION IN THE LAW (2d ed. 1985). For a good exploration of causation in employment discrimination law, see Martin J. Katz, *The Fundamental Incoherence of Title VII: Making Sense of Causation in Disparate Treatment Law*, 94 GEO. L.J. 489 (2006).

[31] Greiner, *supra* note 1, at 543.

[32] *Id.* at 556.

[33] *Id.* at 537, 560 & 581.

[34] In most civil rights litigation, liability can be established if an actor makes a decision using an improper reason, even if it is not a but-for cause of the decision. *See, e.g.*, Desert Palace, Inc.

argument is the claim that regression does not present an adequate causal framework. We think that it does present an adequate framework, or at least one as adequate as his proposed alternative.

Let's begin with nonstatistical evidence of discrimination. First, compare two types of evidence of discrimination: (1) the supervisor says he didn't hire a woman because of her sex and (2) the woman presents evidence that she was much more qualified for the position than the man who was actually hired. The first type of evidence is direct and ideal. But these days it is seldom if ever available. The second type of evidence is different. The plaintiff, in essence, is saying, "assuming no discrimination, how likely is it that the supervisor would hire a person who is clearly less qualified?"[35] The plaintiff is claiming that the answer is "not very likely" and therefore that the court should find that discrimination was the real reason for the decision. In this simple case, the strength of the plaintiff's case will depend on how favorably she compares to the man. If she is clearly more qualified on every dimension and there is no other explanation for the decision, she has a very good case. On the other hand, if the man is more qualified in some ways (or even if there are non-qualification reasons for the decision other than gender), then her case is weaker.

This second example presents the causation question in the same form as in a statistics-based case. The twist is that statistics permits comparison between large groups of men and women. The problem in making this comparison is that the groups of men and women are not exactly the same. Note that this is the same type of problem as in the second example above; where the woman and the comparative man might not be exactly the same (or, equivalently, the woman might not be superior on every relevant dimension). The factfinder has to use her intuition and judgment to decide if the comparison is close enough and, if so, to make the inference that, in the absence of discrimination, one would not see the man preferred over the woman. In a statistics-based case, the statistical analysis is an attempt to equalize the two groups to make them more comparable. At the end of the day, the ultimate "cause" question is the same: after making the two groups as

---

v. Costa, 539 U.S. 90, 94–95 (2003) (employment discrimination); Mt. Healthy City Sch. Dist. Bd. of Educ. v. Doyle, 429 U.S. 274, 287 (1977) (First Amendment); *see also* Martin J. Katz, *Reclaiming* McDonnell Douglas, 83 NOTRE DAME L. REV. 109, 124–42 (2007).

[35] Alternatively, one could say that the plaintiff is saying, "since the man hired was less qualified, how likely is it that the decision was discriminatory?" For the purposes of our comparison of regression and potential outcomes, this is not an important distinction since both depend on the form of inferential reasoning in the text. There is another school of statistics, Bayesian analysis, that permits analysis of the question in this footnote. RAMONA L. PAETZOLD & STEVEN L. WILLBORN, THE STATISTICS OF DISCRIMINATION, ch. 12 (2008). Although not central to our comparison here, the availability of Bayesian analysis is another reason to be skeptical of claims of a universally favored form of statistical analysis in civil rights litigation.

comparable as possible, would we see this result in the absence of discrimination (such as, women on average being paid $X less each year)?

We should note that we have made a leap here, though it is not a leap that distinguishes the statistical and nonstatistical cases. The precise question the statistical case is answering is not "how often would you see this outcome in the absence of discrimination," but rather "how often would you see this outcome by chance."[36] Judicial acceptance of statistical cases means that the courts are willing to infer that "discrimination" is the reason for a disparity if a well-done statistical study demonstrates that it is highly unlikely to occur by chance. This is the same thing that is happening in the nonstatistical case. There, when the factfinder decides that the woman was better qualified and so it seems really unlikely the supervisor would have hired the man over her, the courts permit the factfinder to infer that "discrimination" was the real reason.[37] Again, then, the structure of the two types of proof of discrimination is the same.[38]

There are two main differences between the statistical and nonstatistical cases, but neither affects the basic form of the causation decision. First, the statistical case is much harder to grasp. There are a lot of employees involved, with a large number of relevant variables (education, years worked, salary, etc.), and a sophisticated statistical analysis to facilitate the comparison (but also to make it more complicated and abstract). In the nonstatistical case, we think we know more about the elements of the necessary comparison — for example, about how managers make employment decisions, how humans process information, and the like. Given the modern psychological literature, we probably know less about these things than we think we do.[39] Thus, the difference here may be more about our awareness of limitations than about the limitations themselves. Another difference is that only the statistical case permits us to quantify precisely our estimate of how likely it would be that we would see a particular result in the absence of discrimination. The same general type of estimate is being made in the nonstatistical case; it just does not have a probability number attached to it. Thus, in the nonstatistical case, a jury might find that the woman really was considerably more qualified than the man, so it was quite unlikely that the manager would have hired the

---

[36] Even this is not the precise question, but it is close enough for our purposes. For discussion, see PAETZOLD & WILLBORN, *supra* note 35, § 2.04, at 2-9 to 2-14 (2008).

[37] In *St. Mary's Honor Ctr. v. Hicks*, 509 U.S. 502 (1993), the Court held that the factfinder was not absolutely required to make this inference. *Id.* at 511. Nor is a court absolutely required to make the inference in a statistical case. In both cases, however, the courts regularly and appropriately make this leap.

[38] From now on in this article, we will generally make this leap without comment.

[39] *See* sources cited *supra* note 23.

man absent discrimination. By contrast, in the statistical case, an expert might testify that, after controlling for relevant variables, one would see this difference in wages by chance in only one out of 5,000 cases. But at their base, both judgments are the same: one would not expect to see these results absent discrimination.

An important characteristic of the "cause" determination in both statistical and nonstatistical cases is that the normal approach does not directly answer the question we would really like to have answered. The question we would ideally like to have answered is: did the employer intend to discriminate? But, as we indicate above, we rarely have direct evidence of that these days. Instead, the question we are stuck with is: would you see this outcome in the absence of discrimination? As a practical matter, that is the question that is being answered in both statistical and nonstatistical cases.[40]

Both regression and Professor Greiner's preferred potential outcomes approach are methods for answering this question. And, although neither can answer the direct discrimination question, they both answer the practical question. However, they in do so in different ways. Regression uses real data, controls for relevant non-sex differences between men and women as well as possible (with lots of judgment thrown in about what to control for and how), and then produces a coefficient for sex that is either significant or not. If done perfectly, which it never is, the analysis provides a direct and precise answer to the practical question of how likely is it that you would see this outcome absent discrimination? The potential outcomes approach attempts to match men and women on relevant variables (using lots of judgment about what to match and how), then produces hypothetical data for each matched man and woman, and then produces a difference between the matched groups of men and women on the variable of interest (such as wages) that is either significant or not. Again, if done perfectly, the analysis provides a direct and precise answer to the same question of how likely it is that you would see this difference between men and women absent discrimination. But the important point here is that the approach does *not* avoid statistical inference or the kinds of probabilities reported as statistical (non)significance. The

[40] For reasons that are outside the scope of this article, the practical question may be the better one to answer anyway. The "intend to discriminate" question puts a narrow psychological gloss on discrimination, and, because of that, it is underinclusive. Discrimination should not be limited to conscious discrimination. There may well be circumstances in which an employer could answer truthfully that it did not intend to discriminate, but nevertheless, because of structural factors or implicit association discrimination or something else, women are treated differently and less favorably than men. Our conception of discrimination should be sufficiently broad to encompass those forms of discrimination, too, and the second, practical question may well do a better job of that the first, "ideal" question. *See generally* Owen M. Fiss, *A Theory of Fair Employment Laws*, 38 U. CHI. L. REV. 235, 297–99 (1971).

basic form of the legal cause argument is the same, the potential outcomes approach uses regression modeling within it, and the ultimate result is either significant or not. Do not get us wrong. We think the potential outcomes approach is promising. But we do not think it avoids the problems endemic to discrimination cases — statistical and nonstatistical — and we do not think it avoids most of the problems with regression that Professor Greiner points out. Sometimes it might ease those problems and so be the preferred technique, but not always.

Thus, we disagree with Professor Greiner that regression does not have an adequate causal framework. In general terms, the causal framework is the same one used in nonstatistical cases. It is the same one used by his proposed potential outcomes approach. All three frameworks are answering the question of how likely it is that we would see this outcome in the absence of discrimination. They answer the question by using and manipulating data in different ways, but the general approach is the same. Certainly, regression can be used badly and inappropriately, and it is easy to find cases where that has occurred. If potential outcomes ever becomes commonly used in civil rights litigation (and we hope it does),[41] it will be used badly and inappropriately on occasion. But that will not be reason to abandon it. Instead, it will be reason to remind people of the theory, value, and limits of potential outcomes analysis and of statistical analysis in civil rights cases more generally.

III. CONCLUSION

Professor Greiner's article implies that statistical proof of discrimination is considerably different and more dangerous than other types of proof, and so it must be held to much higher standards. Special steps must be taken with statistical evidence to guard against expert bias and other forms of contamination. If sometimes those especially high standards cannot be met, then the proof cannot be used to try to prove discrimination. That is simply the "price to be paid."[42]

We think that is incorrect and worry that it will nudge courts in the wrong direction. First, the basic structure of a statistics-based case is the same as every other type of case. On this level, statistical and nonstatistical cases are the same, not different. Second, while of course we should be concerned about expert witnesses who might be biased and attempt to mislead the courts, supervisors, plaintiffs and others may also be biased and misleading when they testify in nonstatistical cases. Again, the two situations are similar. We should do our best to limit this and to trust in the judicial process to do its usual good job, instead

[41] We know of no cases to date in which it has been used.

[42] Greiner, *supra* note 1, at 567.

of throwing out the baby with the bathwater. Third, though statistical proof is complicated and hard to understand, as we find out more and more about garden-variety proof we find it is complicated and hard to understand, too. Eyewitness testimony is not as straightforward as it might seem,[43] many of us have implicit and unconscious biases,[44] and so on. Both statistical and nonstatistical cases call on judges and juries to make subtle and difficult judgments. That is no reason to shy away from the proof or impose special extra-tough standards for one category only. Finally, and most importantly, statistical proof of discrimination has great value. Indeed, sometimes it is the only way to detect systemic discrimination that would go unaddressed otherwise. Limiting the tools that might help us to parse our complicated world, especially on issues as important as civil rights, is simply ill-advised.

---

[43] *See generally* 2 HANDBOOK OF EYEWITNESS PSYCHOLOGY (Rod C. L. Lindsay et al. eds., 2007).

[44] *See* sources cited *supra* note 23.