

THE YALE LAW JOURNAL

D. JAMES GREINER & CASSANDRA WOLOS PATTANAYAK

Randomized Evaluation in Legal Assistance: What Difference Does Representation (Offer and Actual Use) Make?

ABSTRACT. We report the results of the first of a series of randomized evaluations of legal assistance programs. This series of evaluations is designed to measure the effect of both an offer of and the actual use of representation, although it was not possible in the first study we report here to measure constructively all effects of actual use. The results of this first evaluation are unexpected, and we caution against both overgeneralization and undergeneralization.

Specifically, the offers of representation came from a law school clinic, which provided high-quality and well-respected assistance in administrative “appeals” to state administrative law judges (ALJs) of initial rulings regarding eligibility for unemployment benefits. These “appeals” were actually *de novo* mini-trials. Our randomized evaluation found that the offers of representation from the clinic had no statistically significant effect on the probability that unemployment claimants would prevail in their “appeals,” but that the offers did delay proceedings by, on average, about two weeks. Actual use of representation (from any source) also delayed the proceeding; we could come to no firm conclusions regarding the effect of actual use of representation (from any source) on the probability that claimants would prevail. Keeping in mind the high-quality and well-respected nature of the representation the law school clinic offered and provided, we explore three possible explanations for our results, each of which has implications for delivery of legal services.

We also conduct a review of previous quantitative research attempting to measure representation effects. We find that, excepting the results of two randomized studies separated by more than thirty years, this literature provides virtually no credible quantitative information on the effect of an offer of or actual use of legal representation. Finally, we discuss disadvantages, advantages, and future prospects of randomized studies in the provision of legal assistance.

AUTHOR. D. James Greiner is Assistant Professor of Law, Harvard Law School. Cassandra Wolos Pattanayak is College Fellow, Department of Statistics, Harvard University. Acknowledgements follow the Table of Contents.*



ARTICLE CONTENTS

INTRODUCTION	2121
A. The Research Program	2121
B. Offers Versus Actual Use	2127
 I. OUR STUDY	2132
A. The Unemployment System and First-Level Appeals in Massachusetts	2132
B. The Status of Research on Representation in First-Level Appeals Prior to Our Study	2139
C. The Harvard Legal Aid Bureau	2140
D. Study Methodology	2143
E. Results	2144
 II. THE CURRENT STATE OF OUR KNOWLEDGE	2175
A. Three Methodological Problems	2184
1. Failure to Specify the Intervention: When Is Representation Assigned?	2184
2. Selection Effects: How Is Representation Assigned?	2188
a. Judge-Induced Selection Effects: Outcome-Driven Counsel Appointments?	2190
b. Client-Induced Selection Effects, Part I: Only the Strong Want To (and Do) Survive?	2191
c. Client-Induced Selection Effects, Part II: Different Circumstances?	2193
d. Lawyer-Induced Selection Effects: Different Ways of Culling Clients?	2194
e. Selection Effects: Are Plausible Guesses Possible?	2195
3. Accounting for Uncertainty	2195
B. How Well-Run Randomized Experiments Solve These Problems	2196
 III. WHERE DO WE GO FROM HERE?	2198
A. The Limits of Randomized Studies	2198
1. Systemic Change	2199
2. Fielding Studies	2200
3. Provider Objections	2201
	2119

B. Maximizing Information	2204
1. Nonpecuniary Interests	2205
2. Outreach, Intake, and Client Choice	2206
3. System Accessibility and Accuracy	2207
CONCLUSION	2208
APPENDIX: "POWER CALCULATIONS"	2210

* The authors thank Carmel Arikat and Tim Taylor for first-rate research assistance, as well as Dr. Travis Coan for help in obtaining data from the Department of Labor. They also thank Laura Abel, Rachel Brewster, John Coates, Esme Caramello, Glenn Cohen, Scott Cummings, Lisa Dealy, Ted Eisenberg, Nora Engstrom, Steve Eppler-Epstein, Brian Flynn, Bob Ganong, John Goldberg, David Grossman, Michael Heise, Dan Ho, Anne Joseph O'Connell, Louis Kaplow, Adriaan Lanni, Andrew Martin, Dan Meltzer, Margaret Monsell, Alison Morantz, Rafael Pardo, Kevin Quinn, Ben Roin, Bob Sable, Ben Sachs, Rebecca Sandefur, Dave Schwartz, Jeff Selbin, Carroll Seron, Jed Shugerman, Donna Southwell, Jessica Steinberg, Matthew Stephenson, David Udell, Adrian Vermeule, David Wilkins, and Richard Zorza for criticisms and comments on drafts of this piece. This research was supported by a grant from the ABA Litigation Research Fund. The authors also give tremendous thanks to the Massachusetts Division of Unemployment Assistance, particularly Bob Ganong, for its work in providing the outcome data for this study. The biggest thanks of all goes to the students of the Harvard Legal Aid Bureau, whose desire to do right by their clients led them to engage in this study.

This paper benefited from feedback received at presentations and discussions at the Institute for Quantitative Social Science as well as at the law schools of the University of Connecticut, Emory, and Harvard. Also extremely helpful were the comments and criticisms received at a symposium on this paper hosted by concurringopinions.com.

INTRODUCTION

A. The Research Program

Particularly with respect to low-income clients in civil cases, how much of a difference does legal representation make?¹ The question is fundamental to the legal profession and to legal pedagogy, and it implicates overlapping areas of inquiry: from access to justice and the civil *Gideon* movement,² to the design of adjudicatory systems; from legal ethics and the unauthorized practice of law, to efficiency in the delivery of legal services (a subject addressed in the organic statute of the Legal Services Corporation).³ In 2007, we initiated discussions with various legal services providers to generate interest in a series of randomized trials⁴ that we hoped would provide gold-standard answers to the

1. We discuss in this Article the causal effect of both an offer of representation (from a particular service provider) and the actual use of representation (from any source). As we explain at multiple points throughout this Article, the two are different. Not all potential clients offered representation from a particular provider take it. Some potential clients not offered representation from a particular provider find representation elsewhere. We return to this point and its implications repeatedly.
2. The term “civil *Gideon*” comes from the Supreme Court’s decision in *Gideon v. Wainwright*, 372 U.S. 335 (1963), which held that indigent defendants have a right to counsel in criminal cases involving a substantial risk of incarceration for more than a year. No blanket right to counsel exists in the civil context; instead, generally speaking, a case-by-case assessment is required, in part owing to lack of information on how much a difference representation makes in various contexts. *See Turner v. Rogers*, 131 S. Ct. 2507 (2011) (holding that no per se right to counsel exists even when a civil litigant faces incarceration); *Lassiter v. Dep’t of Soc. Servs.*, 452 U.S. 18, 29-30 n.5 (1981) (stating that statistical studies regarding the effect of counsel on the fairness of parental termination hearings have thus far proven to be “unilluminating”); *see also* Tamara Audi, ‘Civil *Gideon*’ Trumpets Legal Discord, WALL ST. J., Oct. 27, 2009, <http://online.wsj.com/article/SB125659997034609181.html> (discussing the status of the civil *Gideon* movement, which seeks to mandate legal representation for indigent parties in civil cases); Carol J. Williams, *California Gives the Poor a New Legal Right*, L.A. TIMES, Oct. 17, 2009, <http://articles.latimes.com/2009/oct/17/local/me-civil-gideon17> (reporting on California’s new civil *Gideon* pilot program).
3. Legal Services Corporation Act of 1974 § 1007(a)(3), 42 U.S.C. § 2996 f(a)(3) (2006); *see also* LEGAL SERVS. CORP., PERFORMANCE CRITERIA 4, 22-28 (2007), available at <http://www.lsc.gov/pdfs/LSCPerformanceCriteriaReferencingABAStandards.pdf> (“The Criteria focus particularly on results and outcomes.” (emphasis omitted)).
4. On the need for and benefits from randomization in law, see generally Michael Abramowicz, Ian Ayres & Yair Listokin, *Randomizing Law*, 159 U. PA. L. REV. 929 (2011), in which the authors argue for the randomization of legal rules, particularly those promulgated by agencies. For an example of the power of randomization to reveal facts in the area of criminal counsel, see David S. Abrams & Albert H. Yoon, *The Luck of the Draw: Using*

question of how much of a difference both an offer of and actual use of legal representation make. We sought practice areas in which demand for legal services outstripped a provider's capacity, and in those areas, we suggested that the provider allow us to randomize which of several eligible potential clients would receive an offer of representation.⁵ The randomization would create "treated" (offered representation from this service provider) and "control" (no such offer) groups identical (up to random variation) in all ways except for the offer of representation. Examination of official records for the treated and control groups would allow rigorous measurement of the difference the offer of representation made with respect to the set of outcomes recorded in those records (an important but not exhaustive set of possible outcomes of interest). Fancier statistical techniques might allow inferences about the effect of actual use of representation for the set of persons who requested help from the particular provider.

As we explain in Part III, this provider-centered framework is not the only kind of randomized trial one can use to measure representation effects. Thus, one of our hopes was to use these early efforts as a springboard for additional randomized studies that were court-centered as well as spread across various other dimensions of interest, such as legal area (e.g., housing, family law, immigration, bankruptcy), level of representation (e.g., full attorney-client relationships, lawyer-for-the-day, advice), type of adjudicator (e.g., judge, administrative law judge (ALJ)), and style of adjudication (e.g., inquisitorial, adversarial).

Our primary hope was, and still is, that gold-standard research would produce information useful to the legal services community. We foresee several possible uses.⁶ First, our understanding is that both nationally and in every state in which the issue has been analyzed, demand (both nascent and

Random Case Assignment To Investigate Attorney Ability, 74 U. CHI. L. REV. 1145, 1152-53 (2007).

5. Randomization could occur with or without some kind of assessment of each case's merit or of whether the case in some sense "deserved" representation. What was essential was that any screen of this nature be applied prior to randomization.
6. Many of these themes are articulated in Russell Engler, *Toward a Context-Based Civil Right to Counsel Through "Access to Justice" Initiatives*, 40 CLEARINGHOUSE REV. 196, 200 (2006). On the need for the kind of research we advocate here, see, among others, Laura K. Abel, *Evidence-Based Access to Justice*, 13 U. PA. J.L. & SOC. CHANGE 295 (2010); Helaine M. Barnett, *2004 Bellwood Lecture: Justice for All: Are We Fulfilling the Pledge?*, 41 IDAHO L. REV. 403, 424 (2005); and Richard Zorza, *An Overview of Self-Represented Litigation Innovation, Its Impact, and an Approach for the Future: An Invitation To Dialogue*, 43 FAM. L.Q. 519, 541-42 (2009).

expressed)⁷ for legal services outstrips supply,⁸ posing an acute problem of resource allocation⁹ that has become the stuff of newspaper editorials.¹⁰ Rigorous assessments of how much of a difference an offer or actual use of representation made in different service areas would allow funders and providers themselves to allocate scarce resources more effectively. Second, rigorous evaluation of particular programs would help determine whether programs could be altered so as to make them more effective.¹¹ Of particular interest to us were provider outreach and intake systems, because depending on how these were designed, it struck us that some providers might not be reaching the populations of potential clients that most needed, and would most benefit from, help. Third, if we were to discover that an offer or actual use of representation made less of a difference in certain tribunals, perhaps those tribunals could be examined to see whether the relative lack of a representation effect might be due to tribunal characteristics or practices that made these systems accessible to pro se litigants. That, in turn, might provide direction for efforts from the legal services community (to the extent allowed)¹² to advocate for reform of other tribunals so as to increase their accessibility. Fourth, we believe that a gold-standard evaluation demonstrating that a particular legal

7. Nascent demand refers to persons who have legal problems but who, for whatever reason, do not seek legal assistance. Expressed demand refers to persons who do seek legal assistance for legal problems.
8. For a recent major study covering the nation as a whole, see Helaine M. Barnett, *Preface to* *LEGAL SERVS. CORP., DOCUMENTING THE JUSTICE GAP IN AMERICA: THE CURRENT UNMET CIVIL LEGAL NEEDS OF LOW-INCOME AMERICANS* (2009), available at http://www.lafla.org/pdf/justice_Gap09.pdf (“[F]or every person helped by LSC-funded legal aid programs, another is turned away. That was the primary finding in 2005 and LSC’s collection of data from LSC-funded programs across the country in 2009 reaffirms that finding.”). We have on file a collection of over two dozen national and state-specific studies reaching the same essential conclusion.
9. The allocation problem has become so difficult as to engender a fundamental reexamination of the premises of legal aid generally. *See* Richard Moorhead & Pascoe Pleasence, *Access to Justice After Universalism: Introduction*, 30 J.L. & SOC’Y 1 (2003).
10. Editorial, *Amid Epidemic of Foreclosures, A Crying Need for Legal Aid*, BOS. GLOBE, Nov. 3, 2009, http://www.boston.com/bostonglobe/editorial_opinion/editorials/articles/2009/11/03/amid_epidemic_of_foreclosures_a_crying_need_for_legal_aid; Editorial, *Need a Lawyer? Good Luck*, N.Y. TIMES, Oct. 15, 2010, <http://www.nytimes.com/2010/10/15/opinion/15fr12.html> [hereinafter Editorial, *Need a Lawyer?*].
11. Gregg G. Van Ryzin & Marianne Engelman Lado, *Evaluating Systems for Delivering Legal Services to the Poor: Conceptual and Methodological Considerations*, 67 FORDHAM L. REV. 2553, 2555-56 (1999).
12. *See* Omnibus Consolidated Rescissions and Appropriations Act of 1996, Pub. L. No. 104-134, § 504(a), 110 Stat. 1321, 1321-53 (limiting range of persons or entities to which the Legal Services Corporation (LSC) may provide assistance); 45 C.F.R. § 1610.8 (1997) (same).

services program was effective would be useful in cost-benefit analyses¹³ that might lead to arguments for increased funding,¹⁴ which we fervently endorse.

We pause to note that an evaluation of whether an offer or actual use of representation affects client outcomes is not the same thing as measuring the quality of a services provider's lawyering. There could be a variety of reasons why representation (offer or actual use) might not affect measured outcomes despite world-class advocacy, including, for example, (i) that the provider's outreach and intake system is producing only an unusually competent and motivated client base that does not actually need assistance, or needs it less, or (ii) that the adjudicatory system in which clients are operating is friendly toward pro se litigants, leaving less room for representation to make a difference, or (iii) that the issues to be litigated in a particular area are relatively straightforward (at least for the set of potential clients who seek representation), again leaving less room for representation to make a difference.

Returning to our research program, not all legal services providers embraced our suggestions to engage in gold-standard evaluation,¹⁵ but some did. In doing so, these providers demonstrated the courage necessary to subject their programs to evaluation that did not have a preordained result. Among the most courageous were the students of the Harvard Legal Aid Bureau (HLAB), a student-run, faculty-overseen legal services office that is part of the clinical educational program at Harvard Law School. The first study in our research program consisted of a randomized evaluation of HLAB's representation of claimants seeking unemployment benefits. As we explain, the quality of HLAB representation in the unemployment context is both high and well-respected.

This first study led to unexpected results. With respect to the claimants reached by the outreach and intake systems of HLAB's unemployment practice, and with respect to outcomes measurable from official records (which concerned a claimant's pecuniary interests), the randomized evaluation determined that an offer of HLAB representation had no statistically significant effect on the probability that a claimant would prevail, but that the offer did delay the adjudicatory process. This finding does not mean that we know that the HLAB offer had no positive effect on a claimant's probability of success. We can say, however, that any such effect is unlikely to have been large¹⁶ (or

13. See, e.g., J.J. Prescott, *The Challenges of Calculating the Benefits of Providing Access to Legal Services*, 37 FORDHAM URB. L.J. 303, 316 n.37 (2010).

14. Jeanne Charn & Jeff Selbin, *Legal Aid, Law School Clinics and the Opportunity for Joint Gain*, 11 MGMT. INFO. EXCHANGE J. 28, 29 (2007); Van Ryzin & Lado, *supra* note 11, at 2554-58.

15. We discuss reasons why this reluctance to our research program might have surfaced in Part III.

16. See *infra* Appendix for what we mean by "large."

else the data probably would have shown it), and that we do have a high degree of confidence in the delay finding.

Meanwhile, because roughly one-third of HLAB's client base consisted of claimants who were erroneously denied benefits as an initial matter and who would eventually have that erroneous denial reversed, the statistically significant delay probably¹⁷ meant that many of these claimants who were offered HLAB assistance suffered the harm of having to wait longer for their benefits to begin. We note that courts, legal services providers, commentators, and the U.S. Department of Labor have repeatedly and emphatically stated that the unemployment benefits system depends on speed, in a way other benefits programs do not, for its effectiveness.¹⁸ As we explain, there were also potential consequences for the financing of the unemployment system associated with cases in which there was an erroneous grant of benefits that would eventually be reversed. With respect to the actual use of, as opposed to an offer of, representation, we were able to determine that a delay effect was again present; we could come to no firm conclusion regarding a use-of-representation effect on the win rate. In sum then, in a purely pecuniary sense, the set of claimants HLAB's intake system allowed it to reach might have been better off not receiving the HLAB offer of assistance.

The unexpected results of the HLAB study prompted us to reexamine the literature purporting to assess quantitatively how much of a difference legal representation makes in civil cases. We found that despite the fact that the question has been studied dozens of times,¹⁹ very few of these studies (only two, really, both of which were randomized)²⁰ are worthy of credence. Almost all of these studies follow the same basic design, which is a comparison of the outcomes of cases with representation versus cases without representation, sometimes with regressions using predictors available in official case files. Partly as a result of this design, almost all such studies suffer from

17. As one commenter reminded us, we did not actually measure the date upon which unemployment claimants received checks, and it is possible in some cases that there were further issues to be decided between an eligibility determination and money in a claimant's hand. Our assertion regarding a potentially harmful delay thus depends on the assumption that a delay in the eligibility decision (which we did measure) on average results in a delay in when checks are delivered (which we did not measure).

18. See *infra* notes 61-68 and accompanying text.

19. See *infra* note 154.

20. W. VAUGHN STAPLETON & LEE E. TEITELBAUM, IN DEFENSE OF YOUTH: A STUDY OF THE ROLE OF COUNSEL IN AMERICAN JUVENILE COURTS (1972); Carroll Seron et al., *The Impact of Legal Counsel on Outcomes for Poor Tenants in New York City's Housing Court: Results of a Randomized Experiment*, 35 LAW & SOC'Y REV. 419 (2001).

methodological problems so severe as to render their conclusions untrustworthy,²¹ which (we hasten to emphasize) is different from wrong.²²

We find this dearth of credible, quantitative information troubling. Again, as we understand it, legal services budgets have been hit hard in recent years by reductions in charitable giving, state funding, and proceeds from interest on lawyers' trust fund accounts.²³ The resource allocation concerns identified above appear to be real. In the absence of credible, quantitative information about the effect legal services have on clients across different service areas, those who advocate for greater legal services funding, funders, and legal services providers themselves are left to rely on intuition and guesswork to make critical decisions. As our study shows, intuition and guesswork may be wrong in a way that matters. To our knowledge, prior to our study, no one's instincts resulted in a belief that an offer of legal representation could be adverse to the client's interests in any setting.

With this in mind, this Article has three purposes, which we tackle in turn. In Part I, we report the results of the randomized evaluation we conducted with HLAB of its unemployment practice. In Part II, we explain that the dozens upon dozens of observational (meaning non-randomized) studies purporting to measure the effect of representation in civil proceedings suffer from fundamental design problems, including (i) the failure to define the nature of the legal representation being studied; (ii) multiple layers of selection effects; and (iii) improper assessments of the uncertainty in estimation. These problems render the quantitative conclusions of these observational studies unworthy of credence. Concomitantly, we explain why the two randomized studies conducted thus far in the United States are trustworthy.

Finally, in Part III, beginning with the proposition that three randomized studies over more than fifty years across our nation's vast, multilayered legal system are not enough, we start the process of outlining a research agenda for the future. If the legal and quantitative communities could do no better than

21. See *infra* Section II.A. Occasionally, authors have acknowledged the methodological problems in case-file-based observational studies in this area; see, for example, Karl Monsma & Richard Lempert, *The Value of Counsel: 20 Years of Representation Before a Public Housing Eviction Board*, 26 LAW & SOC'Y REV. 627, 629-31 (1992). Some have recognized how to address these problems. See, e.g., Jean R. Sternlight, *Lawyerless Dispute Resolution: Rethinking a Paradigm*, 37 FORDHAM URB. LJ. 381, 389 n.31 (2001) ("Theoretically the solution to this quandary is to assign or not assign attorneys to disputes on a random basis, but accomplishing this end in the real world is difficult.").

22. The sole observational study we might believe in this area is Monsma & Lempert, *supra* note 21, which went to great lengths to address some of the challenges an observational design poses.

23. See, e.g., Barnett, *supra* note 8, at 412-14.

the observational studies we discuss in Part II, then perhaps those who make decisions in this area would be forced to rely on their conclusions, on the theory that unreliable quantitative information is better than instinct and conjecture. We are not certain that we subscribe to this theory, but we need not decide because we in the legal quantitative community can do better. This Article reports results from the first of five randomized trials in which we are participating. Besides unemployment, these studies concern housing/eviction,²⁴ Social Security disability, and divorce. They concern multiple models of service delivery, from full-time staff attorneys to pro bono referral services to lawyer for the day programs to student attorneys. They concern administrative adjudications (state and federal), specialized courts, and courts of more general jurisdiction. As we discuss below, randomized trials cannot answer all questions concerning how much of a difference legal representatives make, but if used creatively, they can provide credible answers on a far wider range of questions than is currently appreciated.

B. Offers Versus Actual Use

Before proceeding to the main body of this Article, we introduce a distinction between an analysis focusing on the causal effect of an offer of representation from a particular service provider and one focusing on the causal effect of actual use of representation from any source.²⁵ As we explained above and will explain in greater detail below, the fundamental research design we employed in the study we report here was a randomization of an offer of representation from HLAB. We could not ethically randomize actual use of

²⁴. D. James Greiner, Cassandra Wolos Pattanayak & Jonathan Philip Hennessy, *How Effective Are Limited Legal Assistance Programs? A Randomized Experiment in a Massachusetts Housing Court* (Oct. 24, 2011) (unpublished manuscript), <http://ssrn.com/abstract=1880078>; D. James Greiner, Cassandra Wolos Pattanayak & Jonathan Philip Hennessy, *The Limits of Unbundled Legal Assistance: A Randomized Study in a Massachusetts District Court and Prospects for the Future* (Oct. 23, 2011) (unpublished manuscript), <http://ssrn.com/abstract=1948286>.

²⁵. Readers familiar with an “intention-to-treat effect” and the statistical invalidity of an “as-treated” analysis will recognize that our discussion addresses issues raised in a variety of fields. See STEVEN PIANTADOSI, *CLINICAL TRIALS: A METHODOLOGIC PERSPECTIVE* 402 (2d ed. 2005) (medical research); Alan S. Gerber, Donald P. Green & Ron Shachar, *Voting May Be Habit-Forming: Evidence from a Randomized Field Experiment*, 47 AM. J. POL. SCI. 540 (2003) (political science). For the statistical perspective upon which we draw, see Joshua D. Angrist, Guido W. Imbens & Donald B. Rubin, *Identification of Causal Effects Using Instrumental Variables*, 91 J. AM. STAT. ASS'N 444 (1996); and Guido W. Imbens & Donald B. Rubin, *Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance*, 25 ANNALS STAT. 305 (1997).

representation from any source. The two—offer from HLAB and actual use from any source—are different. For example, in our study, HLAB (despite its best efforts) was not able to follow up with a small percentage of clients who had initially contacted HLAB and who had been randomly assigned to the ‘offer’ group, and thus not all claimants who were randomized to an offer of HLAB representation in fact used it. Meanwhile, some (probably around 39%)²⁶ of the potential clients not offered HLAB representation found legal assistance elsewhere. The two previous randomized studies of legal representation experienced similar crossover between treated and control groups.²⁷

In such a setting, there are at least two kinds of causal effects that might be of interest: the causal effect of an offer of HLAB representation, which measures the overall effectiveness of HLAB’s unemployment practice as well as the need for this practice among the community of potential clients reached; and the causal effect of actual use of representation from any source, which might provide some indication of the accessibility of the unemployment adjudicatory system in which we conducted our study. We believe that both effects (offer and actual use) are interesting, and our purpose in this Section is to explain why the effect of an offer is worth studying independently from actual use.

We recognize that we are discussing this concept at an unusually preliminary stage of this Article, before we have detailed what we did in our study or discussed the reasons why randomized experiments are at present the only way to provide credible quantitative information about the effect of representation (offer or actual use). Nevertheless, we take this unusual step because a great number of readers who reviewed drafts of this Article responded (with startling force) that we were silly to focus on offers at all and that only actual use of representation mattered. This group also argued that our data analysis techniques were incorrect because when we focused on the causal effect of HLAB offers, we included persons who received no

26. See *infra* note 135 and accompanying text for an explanation.

27. In the Stapleton and Teitelbaum “Zenith” study, 17.6% of the youths in delinquency proceedings randomized to an offer of representation by study lawyers ended up unrepresented, while 38.7% of those randomized to no such offer found representation elsewhere. STAPLETON & TEITELBAUM, *supra* note 20, at 51-52. In these authors’ “Gotham” study, the corresponding figures were 17.3% (treated group unrepresented) and 11.4% (control group represented). *Id.* In Seron et al., the relevant figures were 44.0% and 3.7%, with the high rate of the treated group unrepresented due to an initial attempt at triage, later abandoned. Seron et al., *supra* note 20, at 425. Thus, the crossover figures in our study are probably less than those experienced in the Zenith and Seron et al. studies, but more than those experienced in the Gotham study.

representation in our “treated” group, and we included persons who did receive representation in our “control” group. These readers requested that we produce a comparison of outcomes for potential clients who received representation (from any source) to those who did not receive representation; as we will explain, the quantitative invalidity of that analysis is so clear that we have never run it and have no intention of doing so.²⁸ We suspect that some of the insistent reactions we received on this score may have stemmed from the mistaken impression that studies such as ours necessarily measure the quality of the lawyering provided to potential clients. To the contrary, there are a variety of reasons why representation (offers or actual use) from a particular legal services program may not change outcomes despite high-quality service.

Perhaps the simplest way to think about why offer effects are interesting is to focus on the fact that issuing an offer of representation is what a service provider actually does to attempt to improve a potential client’s situation. Of course, the provider issues the offer in an effort to induce the potential client to use representation, but the provider typically cannot control whether the potential client does so, and some potential clients inevitably fail to respond. Nor, in the case of a provider decision not to issue an offer of representation, can the provider ethically prevent the potential client from obtaining representation elsewhere. Thus, focusing on the causal effect of the offer of representation allows an understanding of the effect of what the provider actually does, the effect of the event that the provider can control.

With this in mind, there are two primary reasons why we believe independently studying the effect of offers from a particular provider is worthwhile. The first is that studying the offer effect allows evaluation to capture aspects of a service provider’s operation that a focus on actual use does not. Among these aspects are the percentage of offers that a service provider makes that are ultimately accepted, as well as the availability of representation from alternative sources. Two extreme examples to make this point: first, suppose it were the case that only 3% of the clients to whom a service provider offered representation ended up ultimately receiving representation. Even if this 3% realized excellent results, would we call the provider’s operation successful?²⁹ Should an evaluation focus only on the 3%? Or would we say

28. We are not the first to make this choice. See STAPLETON & TEITELBAUM, *supra* note 20, at 51-52 (limiting analysis to offers); Pascoe Pleasence, *Trials and Tribulations: Conducting Randomized Experiments in a Socio-Legal Setting*, 35 J.L. & SOC’Y 8, 17-18 (2008) (same).

29. We should clarify here that what matters is the set of potential clients to whom a service provider has decided to make an offer, not those to whom it actually makes the offer. Thus, for this discussion, the following two circumstances are identical: (i) a service provider decides to make an offer of representation to a potential client who has completed its intake

instead that the provider's outreach and intake systems appear to be focusing on a potential client population that for whatever reason is unlikely to respond, and thus that the provider is wasting resources necessarily dedicated to outreach and intake?

A second extreme hypothetical, the flipside of the first: suppose that 98% of the potential clients to whom a particular service provider does *not* offer representation find legal assistance elsewhere. Would this fact suggest that the particular provider begin to explore the possibility that its resources are not needed in this practice area because there appears to be an adequate supply of representation from other sources, at least with respect to the client base its outreach and intake systems are allowing it to reach? In both of these extreme hypotheticals, if an evaluation were to focus solely on the effect of actual use of representation, it would miss critical facts about the provider's overall delivery system. In the first hypothetical, the critical missed fact might be that the provider is reaching out to a population that does not benefit from its services (because it does not make use of them), and in the second hypothetical the crucial missed fact might be the adequacy of other sources of representation.

The second reason to study the effect of offers of representation is that doing so allows us to avoid a choice between producing a study that is unworthy of credence and committing unethical conduct. As we explain in greater detail in Part II, randomization is critical to making credible inferences about causal effects, particularly in this area.³⁰ We can randomize whether a particular service provider makes an offer of representation to a potential client. But we cannot randomize who actually uses representation. We cannot ethically force a potential client randomized to receive an offer to follow up on the offer and become represented,³¹ and the kind of person who does not follow up is likely to be different from the kind of person who does. Similarly, we cannot ethically prevent a client randomized not to receive an offer from a particular provider from finding representation elsewhere, and the kind of person who finds representation from another source after being turned down once is likely to be different from the kind of person who does not. The upshot

system, but its repeated attempts to contact the potential client to communicate that offer are unavailing; and (ii) a service provider decides to make an offer of representation to a potential client who has completed its intake system, succeeds in reaching the potential client and makes the offer, receives an acceptance of the offer from the potential client, but from then on is unable to contact the client to effectuate representation.

30. See the answer to Question 6 at text accompanying notes *infra* 133-142, as well as all of Part II, *infra*.
31. See Drew A. Swank, *In Defense of Rules and Roles: The Need To Curb Extreme Forms of Pro Se Assistance and Accommodation in Litigation*, 54 AM. U. L. REV. 1537, 1546-47 nn.53-65 (2005) (collecting cases and other authority on the right to proceed pro se).

is that randomization is critical in this area, and we can randomize offers, but not actual use.

Speaking more generally, our argument here is that when one makes repeated attempts to do some action hoping to produce some effect, some percentage of the attempts will ordinarily not go as planned. Should one measure the effect per attempt, or should one measure only the effect of the attempts that go as planned? This question arises in a wide variety of settings. Three hypothetical examples follow.

Voter mobilization: A group of academics sends out canvassers to potential voters' homes to urge the potential voters to vote, but in fewer than 30% of the attempts are the canvassers actually able to talk to potential voters because most of the time, no one is at home when the canvassers visit. In a control group, no canvasser visits. Should the academics measure the effect of the attempt to contact each voter (i.e., of a canvasser visit), or should analysis be limited only to the effect of actual contact with would-be voters? Note that the canvassers are paid by the hour, and their hours unquestionably include time spent knocking on the doors of empty houses.³²

Football playcalling: the offensive coordinator of a football team is trying to decide whether he should call more passing plays or running plays when the team is on offense. The coordinator decides to do some kind of running play versus passing play comparison. Should the coordinator compare yards per play called as a pass versus yards per play called as a run, or should the coordinator compare yards per *completed* pass versus yards per *actual running* play? Remember that not all attempted passes will be completed; in fact, not all plays called as passes will end up in pass attempts because sometimes the quarterback will run or be sacked.³³

Flu shots: an insurance company is considering a program in which it would hire personnel to call doctors who have checkup appointments with the company's elderly insureds to ask the doctors to recommend flu shots to their patients. The thought is that more flu shots will lead to fewer hospitalizations, saving the company money. In a randomized study, the doctors of a treated

32. See Gerber et al., *supra* note 25, at 545-46. Here, the canvassers' attempt to talk to a potential voter is analogous to an HLAB offer of representation, while the voter's having a face-to-face conversation in which someone (perhaps one of the canvassers but also perhaps someone else, such as a campaign employee) urges the potential voter to vote is analogous to actual use of representation.

33. Cf. Robert Folleisbee, *The Long Pass, Is It Football's Best Offensive Weapon?*, RENAISSANCE POST (Jan. 15, 2011, 10:32 AM), <http://renaissancepost.com/sports/pro-football-all-time-passing-stats> (discussing the distinction between yards per attempted pass and yards per completed pass). Here, an analysis focusing on yards per play called as a pass/run is analogous to an analysis focusing on HLAB offers of representation.

group of insureds receive telephone calls, while no calls are made for a control group of insureds. Should the company measure the causal effect per attempt to contact a doctor (i.e., per telephone call), or should it measure only the effect per actual conversation with a doctor?³⁴

To repeat, we believe that one should measure both the effect of the attempt to do the action and the effect of the action when it works out as intended. In our case, that means we believe that the effects of both offers of representation (from HLAB) and actual use of representation (from any source) are worth study, and we try to measure both in this Article. As suggested previously, the lack of randomization makes studying the effect of actual use of representation more difficult. At this introductory stage, it is enough to say that the offer effect is worth examining independently.

I. OUR STUDY

In this Part, we present the results of our study. Section I.A provides basic background on the unemployment system and on how so-called “first-level appeals,” defined below, are conducted in Massachusetts. Section I.B discusses the status of research on the need for legal representation in first-level appeals prior to our study. Section I.C discusses the Harvard Legal Aid Bureau and its unemployment practice. Section I.D details the methodology used in our study. Section I.E provides quantitative results.

A. The Unemployment System and First-Level Appeals in Massachusetts³⁵

The Social Security Act of 1935³⁶ established the unemployment insurance system that covers most³⁷ of the nation’s workforce. According to the U.S.

34. Cf. Keisuke Hirano et al., *Assessing the Effect of an Influenza Vaccine in an Encouragement Design*, 1 BIOSTATISTICS 69 (2000).

35. Our description of the unemployment system, and the legal issues litigated in it, is designed to be sufficient only to understand HLAB’s unemployment practice, which focused primarily on individual cases as opposed to structural change. Thus, we do not discuss a variety of often complex issues that arise in the Massachusetts unemployment system, such as immigrant eligibility, overpayments, and employer requests for redetermination. For a complete analysis of these and other issues, see MONICA HALAS & MARGARET MONSELL, UNEMPLOYMENT ADVOCACY GUIDE: AN ADVOCATES’ GUIDE TO UNEMPLOYMENT IN MASSACHUSETTS (2010), available at <http://www.masslegalservices.org/unemploymentadvocacyguide>.

36. Pub. L. No. 74-217, 49 Stat. 620 (codified as amended in scattered sections of 42 U.S.C. (2006)).

Department of Labor (DOL), in 2009, when the bulk of the data for this study was gathered, there were approximately 17.2 million new claims for unemployment nationally, 79.6% of which resulted in payments.³⁸ Essentially, states administer unemployment programs funded through federal and state payroll taxes,³⁹ with a partial federal credit given to payments made for state taxes dedicated to the program.⁴⁰ The DOL oversees each state's program and releases funding to State Workforce Agencies (SWAs). The DOL also funds SWA administrative costs.⁴¹ To release funding, the DOL must certify each year that the state program meets federal statutory requirements.⁴² These requirements are fairly broad, however, so that in large part, a state's own law determines such critical items as eligibility for, amount of, and duration (within federal limits and subject to federal extensions) of benefits.⁴³

The Massachusetts SWA is the Division of Unemployment Assistance (Massachusetts DUA), a division of the Department of Workforce Development. As is true in most states, Massachusetts law provides certain eligibility requirements, the most important of which is that the employee must become unemployed through no fault of her own.⁴⁴ Massachusetts distinguishes between "quit" cases, in which the claimant left her job but argues that she did so for good cause (such as intolerable working conditions), and "discharge" cases, in which the employer admits that it fired the claimant but argues that it did so for good cause (such as absenteeism).⁴⁵ This distinction is important because in a quit case, the claimant bears the burden of persuasion on the issue of good cause for the separation, while in a discharge

- 37. At various times, railroad workers, servicemen, and other classes of employees have been covered via other programs. *See, e.g.*, Servicemen's Readjustment Act of 1944 (G.I. Bill), Pub. L. No. 78-346, 58 Stat. 284 (codified as amended in scattered sections of 38 U.S.C.); Railroad Unemployment Insurance Act of 1938, Pub. L. No. 75-722, 52 Stat. 1094 (codified as amended at 42 U.S.C. §§ 503, 1104, 45 U.S.C. §§ 351-368).
- 38. These figures were provided by an employee within the U.S. Department of Labor, Employment and Training Administration, Office of Unemployment Insurance. *See* E-mail from Travis Coan, Research Statistician, to D. James Greiner (July 12, 2010, 12:32 PM EST) (on file with authors). We requested corresponding figures for Massachusetts but were unable to verify the figures we received.
- 39. *See* I.R.C. §§ 3301-3311 (2006) (establishing the Federal Unemployment Tax).
- 40. *Id.* § 3302.
- 41. 42 U.S.C. §§ 501, 502(a), 1101(c)(1)(A).
- 42. *Id.* § 503(a).
- 43. Kevin Liebkemann & Raymond Cebula, *Interplay Among Unemployment Insurance, Welfare, Social Security Disability, and SSI Benefits*, 41 CLEARINGHOUSE REV. 305, 305 (2007).
- 44. MASS. GEN. LAWS ANN. ch. 151A, § 25(e) (2011).
- 45. *Compare id.* § 25(e)(1) (quit cases) *with id.* § 25(e)(2) (discharge cases).

case, the employer bears that burden.⁴⁶ As we will see, this difference is relevant to our data. In addition, Massachusetts (like most states) uses an “experience rating” system to determine how much an employer must pay in unemployment taxes. The term “experience rating” is a term of art for the idea that the more an employer “costs” the system (by having former employees file successfully for benefits, meaning the employer fired these employees or gave them good cause to leave), the more it must pay into the unemployment system.⁴⁷ This system gives employers an incentive to contest an unemployment claim by, for example, arguing that a claimant quit voluntarily.

To apply for benefits in Massachusetts, a former employee begins by filing a claim (in person or by phone) with the Massachusetts DUA at a “One-Stop Career Center.”⁴⁸ The previous employer has ten days to respond⁴⁹ and may include information as to why in its view the claimant should not receive benefits.⁵⁰ If either the employer or the employee raises an issue regarding eligibility, a claims adjuster⁵¹ first contacts the claimant and the employer(s) to

46. See *Speak to a Representative*, EXECUTIVE OFF. OF LABOR & WORKFORCE DEV., <http://www.mass.gov/lwd/unemployment-insur/filing-a-claim/issues-with-claim/speak-to-a-representative.html> (last visited Apr. 3, 2012).

47. MASS. GEN. LAWS ANN. ch. 151A, §§ 13-21 (provisions for employer contributions); Interview with Robert Ganong, Chief Counsel & Special Assistant Attorney Gen., Dep’t of Workforce Dev. (Oct. 18, 2010) [hereinafter Ganong October Interview]; *General Questions*, EXECUTIVE OFF. OF LABOR & WORKFORCE DEV., <http://www.mass.gov/lwd/unemployment-insur/ui-online-dua-quest/quest-project-info/system-faq/general-questions.html> (last visited Apr. 3, 2012). Note that the description above corresponds to what are called “contributing employers.” A different class of “optional” employers, principally nonprofits and state agencies, can opt to pay nothing into the system until one of their former employees successfully files for benefits. Ganong October Interview, *supra*; see MASS. GEN. LAWS ANN. ch. 151A, § 14A.

48. See MASS. GEN. LAWS ANN. ch. 151A, § 38(a); *Massachusetts One-Stop Career Centers*, EXECUTIVE OFF. OF LABOR & WORKFORCE DEV., <http://www.mass.gov/lwd/employment-services/career-services/career-center-services> (last visited Apr. 3, 2012).

49. MASS. GEN. LAWS ANN. ch. 151A, § 38(b). If the employer fails to respond in a timely manner without good cause, the employer suffers certain consequences. First, if the claims adjuster awards benefits, the employer cannot appeal that decision. Second, if the claims adjuster denies benefits and the claimant appeals, the employer is not a party to the appeal, and thus cannot cross-examine witnesses or offer documents in evidence (although the employer is invited to send witnesses to the hearing held in the appeal). Ganong October Interview, *supra* note 47. The employer may be able to request a reconsideration of a decision granting benefits and perhaps thereby reobtain party status. See HALAS & MONSELL, *supra* note 35, at 30.

50. MASS. GEN. LAWS ANN. ch. 151A, § 38(b).

51. The proper title for these officials is “Job Service Representative,” and they are colloquially referred to as “adjudicators.” E-mail from Robert Ganong, Chief Counsel & Special

request certain information (such as length of employment) and access to documents, as well as to ascertain the cause of the separation. The claims adjuster then issues an initial ruling, called a “determination,” granting or denying the claim.⁵² Thus, a key aspect of the decisionmaking process used here is that an employee of the decisionmaking body (in fact, the initial decisionmaker herself) shoulders the burden of gathering facts and documents, creating a record the parties can use as a foundation in later proceedings. The letter informing the claimant and the employer of the claims adjuster’s determination includes a form that the “losing” party can mail back to the Massachusetts DUA to initiate a first-level appeal. If the losing party requests an appeal, the claimant is sent a form that includes names and telephone numbers of organizations in the area, such as HLAB (for the Boston region), which represent claimants without charge.⁵³

Few if any legal services providers offer representation to claimants at the initial claim stage, and as we will discuss, HLAB does not do so. Instead, service providers wait until the claims adjuster has made an initial determination, then offer to represent in what is called the “first-level appeal,” as follows.

As a condition for receipt of federal grant funds, the Social Security Act requires that states provide an opportunity for a fair hearing to a claimant who believes her benefits application was erroneously denied.⁵⁴ As is true in most states, Massachusetts also provides a similar opportunity for an employer who believes that benefits were erroneously granted.⁵⁵ This is the so-called “first-level appeal,” but the term “appeal” is a misnomer in that what actually occurs is a *de novo* mini-trial before a “Review Examiner,” called for the purposes of this Article an “ALJ.” According to data provided by the Department of Labor, SWAs in 2009—when the bulk of our dataset was generated—decided approximately 1.5 million first-level appeals in the national unemployment system; Massachusetts decided 21,682 of these. The 14,291 first-level appeals initiated by claimants accounted for just short of two-thirds of the Massachusetts total, and when claimants initiated the appeals, their success rate was approximately 46.6%. When the employer was the party who lost before the claims adjuster and thus was appealing, the claimant success rate at

Assistant Attorney Gen., Mass. Dep’t of Workforce Dev., to D. James Greiner (Nov. 26, 2010, 4:50 PM EST) (on file with authors).

⁵². *Id.*

⁵³. Ganong October Interview, *supra* note 47.

⁵⁴. Social Security Act of 1935 § 303(a)(3), 42 U.S.C. § 503(a)(3) (2006).

⁵⁵. MASS. GEN. LAWS ANN. ch. 151A, § 39(b).

the first-level appeal was 74.6%.⁵⁶ We return to these figures below because they may provide a rough basis by which to assess whether the claimants who sought HLAB assistance are a distinct subset of the overall set of Massachusetts claimants participating in the first-level appeals process.

The adjudicative process in first-level appeals is a “tempestuous . . . marriage”⁵⁷ of inquisitorial and adversarial styles of judging.⁵⁸ The key event in a first-level appeal is the hearing before the ALJ. We personally observed three hearings during this study; each ran in essentially the same way, as follows. The ALJ began by explaining the purpose of the proceeding, clarifying who was appealing and whether the case was a quit or a discharge. The ALJ reviewed the documents the claims adjuster had previously gathered and asked whether there was any objection to including these documents in the hearing record (we never saw an objection). The ALJ turned to the party bearing the burden of proof in the proceeding (the claimant in a quit case, the employer in a discharge case). The ALJ asked questions, allowed the party to narrate additional facts, and provided the opposing party an opportunity to cross-examine. The ALJ asked the party with the burden of proof whether she had any witnesses to call, and if so, the ALJ questioned these witnesses first, then allowed the party who called the witnesses to do so, then the opposing party. The process repeated for the party not bearing the burden of proof. Notice that the ALJ always began the questioning. Prohibitions against hearsay did not apply, and documents were generally admitted without objection if reasonably authenticated. At the close of the evidence, the ALJ allowed the parties to make a brief summary statement. In two of the three hearings we witnessed, the proceedings lasted less than an hour; in a third, the proceedings had to be carried over to another day, because the hearing did not finish in the one-hour allotted time. A cassette tape recorder provided the official record of the proceedings.

Generally, claimants and employers, or more likely their representatives, could submit written briefs at the end of a hearing. These would have to have been written before the hearing itself, and thus have been based on what the representatives anticipated the evidence would show, which might or might not match the testimony actually heard. Because of this potential disconnect,

56. E-mail from Travis Coan, *supra* note 38.

57. Jerry L. Mashaw, *Unemployment Compensation: Continuity, Change, and the Prospects for Reform*, 29 U. MICH. J.L. REFORM 1, 16 (1996).

58. Cf. MASS. GEN. LAWS ANN. ch. 151A, § 39(b) (describing the adjudicative process). Section 39(b) references and adheres to the Massachusetts Administrative Procedure Act. See MASS. GEN. LAWS ANN. ch. 30A, §§ 9-10; 801 MASS. CODE REGS. 1.01-1.03 (LexisNexis 2011) (implementing the statute).

some ALJs deliberately did not read these written submissions until after they had composed a draft of their decisions (if they read them at all), fearing that they would confuse what was in the written submission with the evidence actually adduced at the hearing.⁵⁹ We note that this fact may implicate whether legal representation can make a difference in ALJ decisionmaking.⁶⁰

“[T]imeliness is next to godliness in the operation of unemployment compensation adjudicatory processes.”⁶¹ Courts,⁶² commentators,⁶³ legal services advocates,⁶⁴ and the DOL,⁶⁵ among others, all agree that speed is of extraordinary concern. In *California Department of Human Resources Development v. Java*,⁶⁶ the Supreme Court identified the purposes of the unemployment program as (i) maintaining the newly unemployed worker at subsistence levels without the necessity of her “turning to welfare or private charity,” (ii) allowing the worker to spend her time searching for a new, permanent job, instead of engaging in temporary work to make ends meet, and (iii) preventing a decline in the purchasing power of the worker, considered a way to stabilize the economy.⁶⁷ The *Java* opinion emphasizes over and over again that none of these purposes could be accomplished unless payments were as rapid as possible: “Probably no program could be devised to make insurance payments available precisely on the nearest payday following the termination,

59. Interview with Anne Will, Review Examiner, Mass. Div. of Unemp’t Assistance (Apr. 14, 2010).

60. Although federal law does not require it to do so, Massachusetts (like many states) provides an opportunity for a second-level appeal in the form of a petition for discretionary review to a Board of Review, as well as for judicial review of the Board’s decision. See MASS. GEN. LAWS ANN. ch. 151A, §§ 40-42. See generally ROBERT I. OWEN & EDWARD A. WOOD, *Timeliness in Deciding Second-Level Appeals*, in 3 NAT’L COMM. ON UNEMP’T COMP., UNEMPLOYMENT COMPENSATION: STUDIES AND RESEARCH 643, 643 (1980) (emphasizing the importance of timing at later stages of unemployment adjudication).

61. Mashaw, *supra* note 57, at 19.

62. See, e.g., *Wilkinson v. Abrams*, 627 F.2d 650, 661 n.14 (3d Cir. 1980); see also *infra* text accompanying notes 67 and 73.

63. See, e.g., HERBERT M. KRITZER, *LEGAL ADVOCACY: LAWYERS AND NONLAWYERS AT WORK* 26-27 (1998).

64. See, e.g., Sharon M. Dietrich & Cynthia L. Rice, *Timeliness in the Unemployment Compensation Appeals Process: The Need for Increased Federal Oversight*, 29 U. MICH. J.L. REFORM 235, 237 (1996).

65. See *infra* text accompanying notes 69 and 70.

66. 402 U.S. 121 (1971).

67. *Id.* at 131-32.

but to the extent that this was administratively feasible this must be regarded as what Congress was trying to accomplish.”⁶⁸

In the first-level appeals context, SWAs measure time in days and weeks, not months. In response to *Java*, the DOL promulgated regulations governing first-level appeals; these essentially require SWAs to complete 60% of first-level appeals within thirty days of filing and 80% within forty-five days.⁶⁹ State compliance with this mandate has varied. Massachusetts complied with these standards in 2007, but no doubt due to the recession beginning in 2008, the Massachusetts DUA in 2009 cleared only 17% of first-level appeals within the thirty-day limit and 45% within the forty-five-day limit.⁷⁰ Meanwhile, although its enforcement of these regulations has been criticized,⁷¹ the DOL viewed speed as sufficiently important to warn states against applying statewide hiring freezes to SWAs during the recent recession, under the rationale that inadequate staffing would lead to slower adjudication.⁷² Finally, there have been numerous lawsuits either directly related to SWA compliance with DOL timeliness regulations or motivated more generally by speed concerns.⁷³

68. *Id.* at 130; *see also id.* at 132 (“[e]arly payment”); *id.* at 133 (“promptly”); *id.* at 135 (“earliest point that is administratively feasible”).

69. 20 C.F.R. § 650.4(b) (2011). Massachusetts state law contains a similar emphasis on speed. *See MASS. GEN. LAWS ANN. ch. 151A, § 39(b)(5)* (2011) (requiring the Massachusetts DUA to “make every reasonable effort” to complete first-level appeals within forty-five days); MASS. GEN. LAWS ANN. ch. 151A, § 41(b) (same for Board proceedings).

70. In addition, the DOL has also issued regulations governing timeliness for initial benefit payments in response to claims the claims adjuster finds valid. *See* 20 C.F.R. § 640 (2011).

71. Emp’t & Training Admin., *Benefits: Timeliness and Quality (BTQ) Reports of State Workforce Agencies*, U.S. DEPARTMENT OF LABOR, <http://www.workforcesecurity.dolceta.gov/unemploy/btq.asp> (last visited Apr. 3, 2010).

72. Dietrich & Rice, *supra* note 64, at 237; Gregory Fortescue, *Running Late with Nowhere To Go: Searching for Enforcement of Federal Unemployment Insurance Timeliness Regulations After Gonzaga University v. Doe*, 21 LAB. LAW. 307, 308 (2006).

73. U.S. Dep’t of Labor, Application of State-Wide Personnel Actions to Unemployment Insurance Program, 74 Fed. Reg. 41165, 41165 (Aug. 14, 2009).

74. *See, e.g.*, Fusari v. Steinberg, 419 U.S. 379, 384 n.8 (1975) (noting that Connecticut’s unemployment system at the time was “the slowest in the Nation”); Cal. Dep’t of Human Res. Dev. v. Java, 402 U.S. 121, 121 (1971); Barcia v. Sitkin, 367 F.3d 87 (2d Cir. 2004); Jenkins v. Bowling, 691 F.2d 1225 (7th Cir. 1982); Wilkinson v. Abrams, 627 F.2d 650 (3d Cir. 1980); Burtton v. Johnson, 538 F.2d 765 (7th Cir. 1976); Mindo v. N.J. Dep’t of Labor & Indus., 443 F.2d 824 (3d Cir. 1971); Ross v. Giles, No. C-2-81-902, 1988 WL 252427 (S.D. Ohio Mar. 3, 1988); UAW v. Mich. Emp’t Sec. Comm’n, 517 F. Supp. 12 (E.D. Mich. 1980); Ortega v. Usery, 441 F. Supp. 100 (D. Conn. 1977); Graves v. Meystrik, 425 F. Supp. 40 (E.D. Mo. 1977); Phillips v. Dawson, 393 F. Supp. 360 (W.D. Ky. 1975); Nevills v. Illinois, 388 F. Supp. 622 (E.D. Ill. 1974); AFL-CIO v. Cal. Emp’t Dev. Dep’t, 152 Cal. Rptr.

Delay is considered particularly troubling in the case of a claimant whose initial claim for benefits was erroneously denied, meaning the denial will be reversed at the first-level appeal, because during the pendency of the first-level appeal proceeding, the claimant will not receive payments.⁷⁴ When the claimant prevails in the first-level appeal, the benefits will begin shortly thereafter, and the claimant will be paid the back amounts due from the date of the initial claim. Thus, the harm here is not a loss of money in absolute terms (the time periods involved are short enough to make interest essentially irrelevant) but the loss of the income stream during the time period in which it is potentially most needed.

B. The Status of Research on Representation in First-Level Appeals Prior to Our Study

Recognizing the political and economic importance of unemployment benefits, researchers have analyzed the effect of legal representation in first-level appeals by comparing success rates of claimants with representation to success rates of claimants without representation.⁷⁵ As discussed in Part II, most studies of the effect of legal representation use this same “design.” Two themes purportedly emerge from this research. First, representation ostensibly makes a large (positive) difference in a claimant’s probability of winning. Second, the legal representative purportedly need not be an attorney; rather, (lay) union representatives, paralegals, and law students, among others, can be effective in this setting. For example, with respect to law students, two authors wrote, “[t]ypically, the [law student] clinics handle a small volume of appeals

193 (Ct. App. 1979); *Miranda v. Cal. Unemp’t Ins. Appeals Bd.*, 111 Cal. Rptr. 419 (Ct. App. 1973); *Krupp v. Adm’r, Unemp’t Comp. Act*, 397 A.2d 1360 (Conn. 1978); *Myers v. Unemp’t Comp. Bd. of Review*, 300 A.2d 917 (Pa. Commw. Ct. 1973); *see also* *Torres v. N.Y. State Dep’t of Labor*, 410 U.S. 971 (1973) (Marshall, J., dissenting from denial of rehearing).

74. KRITZER, *supra* note 63, at 26; Marla D. Clark & Jesse S. Reyes, *Procedural Reform in the Unemployment Insurance System*, 29 U. MICH. J.L. REFORM 217, 231 (1996); Dietrich & Rice, *supra* note 64, at 236, 270; James M. Klein & Thomas E. Willging, *Beyond Java: Redistribution of Risks in the Administration of Unemployment Insurance*, 24 CASE W. RES. L. REV. 490, 514-18 (1973).

75. See, e.g., KRITZER, *supra* note 63, at 34-37; Maurice Emsellem & Monica Halas, *Representation of Claimants at Unemployment Compensation Proceedings: Identifying Models and Proposed Solutions*, 29 U. MICH. J.L. REFORM 289, 291 (1995-1996); Murray Rubin, *The Appeals System*, in 3 NAT’L COMM’N ON UNEMP’T COMP., *supra* note 60, at 625; Kenneth Owens, *Representation and Success: Understanding How the Two May Be Connected in Massachusetts Lower Authority Appeals*, NAUIAB NAVIGATOR, Spring 2007, at 6 (on file with authors).

with significant supervision and training in the preparation of the case . . . resulting in an extremely high quality of representation.”⁷⁶

These first two themes have led to calls from commentators⁷⁷ and government officials,⁷⁸ and to a lesser extent from the ABA,⁷⁹ for some form of representational guarantee in unemployment first-level appeals, at least for claimants whose applications are initially denied. A representational guarantee in first-level appeals is a goal of proponents of the civil *Gideon* movement.⁸⁰

C. The Harvard Legal Aid Bureau

The Harvard Legal Aid Bureau, or “HLAB,” is the oldest student legal services organization in the nation.⁸¹ Rising second-year law students apply⁸² to join the program; if admitted, they spend the next two years running a legal services office under the supervision of a Clinical Professor⁸³ and several

⁷⁶. Emsellem & Halas, *supra* note 75, at 327-28. *But see* KRITZER, *supra* note 63, at 63-66 (discussing difficulties students from a Wisconsin law school encountered in representing clients in first-level appeals).

⁷⁷. Ken Morris, *On Appeal: Claimant Advocacy and Full and Fair Hearings*, in 3 NAT'L COMM'N ON UNEMP'T COMP., *supra* note 60, at 665, 665; Susan Nofi-Bendici, *Representing Claimants in Unemployment Compensation Proceedings: Lessons Learned from Hearing and Deciding These Cases*, 43 CLEARINGHOUSE REV. 500 (2010).

⁷⁸. NAT'L COMM'N ON UNEMP'T COMP., BASIC STRUCTURE OF A FEDERAL-STATE UNEMPLOYMENT INSURANCE PROGRAM AND RELATED SUPPORTING PROVISIONS: RESPONSES TO THE NATIONAL COMMISSION ON UNEMPLOYMENT COMPENSATION 13 (1979). *But see* American Bar Association Endorses Proposal for Reform of Unemployment Compensation System, 154 Daily Lab. Rep. (BNA), at A-5 (Aug. 11, 1989) [hereinafter American Bar Association] (reciting Department of Labor concerns regarding whether providing representation in first-level appeals would slow case processing without altering outcomes).

⁷⁹. American Bar Association, *supra* note 78 (describing an ABA request to the Secretary of Labor to conduct a pilot program with an eye toward assuring representation for claimants in unemployment hearings).

⁸⁰. See, e.g., Russell Engler, *Connecting Self-Representation to Civil Gideon: What Existing Data Reveal About When Counsel Is Most Needed*, 37 FORDHAM URB. L.J. 37, 81 (2010); *see also supra* note 2.

⁸¹. HARV. LEGAL AID BUREAU, <http://www.harvardlegalaid.org> (last visited Apr. 3, 2012).

⁸². The application process in recent years has become increasingly competitive, with the number of applications reaching approximately quadruple the number of positions during the years in which we conducted our study.

⁸³. See David Grossman, HARV. LAW SCH. FACULTY PAGE, <http://www.law.harvard.edu/faculty/directory/index.html?id=373> (last visited Apr. 3, 2012).

Clinical Instructors and Fellows.⁸⁴ A student Board of Directors elected to various offices makes most decisions.⁸⁵ Our personal observations, the competitive nature of the HLAB application process, and HLAB's reputation within the community suggest that the HLAB students and the clinical staff provide high quality representation. HLAB represents clients only in the Boston area.

Incoming 2Ls specialize in either housing or domestic cases, but HLAB also handles unemployment first-level appeals. First-level cases are considered good pedagogical tools⁸⁶ because, as the cases cannot settle, they usually result in ALJ evidentiary hearings, providing students with an opportunity to do a direct examination and a cross-examination in addition to legal research and factual development. Moreover, relative to other case types, unemployment cases move through the system rapidly, so a student will usually take a case from intake to judgment in a semester.

During the study, HLAB intake, including unemployment intake, was done via telephone after a claimant who had received a claims adjuster determination called the general HLAB number. An HLAB student-attorney gathered information on financial eligibility, which was based on an income (but not an asset) screen,⁸⁷ as well as certain additional facts regarding the caller and the case.

HLAB students received training on unemployment cases at the beginning of their 2L year. Upon receiving an unemployment case, the course the HLAB student-attorney took depended on the facts and circumstances, so what follows is a rough schematic. The student-attorney ordinarily began by copying the Massachusetts DUA case file (the set of documents compiled by the claims adjuster), which often had information on, among other things, the employer's contentions regarding eligibility.⁸⁸ The HLAB student-attorney

⁸⁴. See *About Us*, HARV. LEGAL AID BUREAU, <http://www.harvardlegalaid.org/about-us> (last visited Apr. 3, 2012).

⁸⁵. See *News*, HARV. LEGAL AID BUREAU, <http://www.harvardlegalaid.org/news> (last visited Apr. 3, 2012).

⁸⁶. Cf. William D. Popkin, *Effect of Representation on a Claimant's Success Rate—Three Study Designs*, 31 ADMIN. L. REV. 449, 452 (1979) (making a similar point in the context of Social Security Administration appeals).

⁸⁷. Income was required to be less than 187% of the federal poverty level. Because the claimant was unemployed at the time of the intake interview, this requirement screened out few callers. This fact precluded us from exploring a regression discontinuity design, although we would have trusted the results of such a nonrandomized design less than the randomized results we present here.

⁸⁸. E-mail from Esme Caramello, Deputy Dir. and Clinical Instructor, Harvard Legal Aid Bureau, to D. James Greiner (Nov. 3, 2010, 11:41 AM EST) (on file with authors).

interviewed the client to ascertain basic facts. Often, the student-attorney attempted to obtain a copy of the case file from the employer. Ordinarily, this was done simply by sending a letter to the employer requesting the file,⁸⁹ but sometimes the student-attorney asked the claimant to request a copy of the file, if the student-attorney felt that the client was up to this task. The latter methods might be used so as to avoid unnecessarily reminding the employer of the fact of a first-level appeal, which might induce the employer to attend the ALJ hearing with a lawyer.⁹⁰ The student-attorney researched the law via traditional search methods,⁹¹ attempted to find and interview additional witnesses and, together with their supervising attorneys, made a judgment as to whether to attempt to convince witnesses to attend the ALJ hearing or to submit testimony via affidavit.⁹² To prepare for the hearing, the student-attorney prepared the client for a direct examination and a cross-examination. The student-attorney also prepared an opening statement in the event that one was allowed, a closing statement, and frequently, a memorandum of law (which may or may not have been submitted, depending on the student-attorney and supervising attorney's collective judgment about whether doing so would be constructive).⁹³ At the hearing, the student-attorney asked questions the ALJ did not already ask that were necessary to elicit relevant facts (on either direct or cross), introduced relevant documents, objected to testimony on evidentiary grounds, made an opening (if allowed) and closing statement, and sometimes (again, depending on a judgment) provided the ALJ with a memorandum of law.⁹⁴

89. See MASS. GEN. LAWS ch. 149, § 52C (2010); E-mail from David Grossman, Dir., Harvard Legal Aid Bureau, to D. James Greiner (Oct. 31, 2010, 1:49 PM EST) (on file with authors).

90. Interview with Alana St. Aude, Student-Attorney, Harvard Legal Aid Bureau, in Cambridge, Mass. (Apr. 12, 2010) [hereinafter St. Aude Interview]; *see also* Nofi-Bendici, *supra* note 77, at 503 (encouraging practitioners to have their clients obtain the file). One implication of this “stealth” effort is that whether an employer attends the hearing with an attorney may be an outcome, not a background variable, in the sense that it may be affected by whether the claimant has an attorney.

91. A primary source is HALAS & MONSELL, *supra* note 35. This volume is updated regularly.

92. Caramello E-mail, *supra* note 88.

93. St. Aude Interview, *supra* note 90.

94. Caramello E-mail, *supra* note 88.

D. Study Methodology

Requests for HLAB representation in unemployment first-level appeals have long outstripped HLAB capacity to provide assistance. This remained true during the study period.

The study process worked as follows: a claimant called the HLAB front desk, and HLAB arranged to conduct its usual screening interview. During the interview, the student-attorney informed the claimant that HLAB was conducting an evaluation and read the claimant a script describing the study⁹⁵ before requesting consent to participate.⁹⁶ HLAB transmitted to us a form that verified oral consent and contained basic information about the claimant. We randomized the case (see below) and informed HLAB whether or not to offer representation. If the randomization required not offering assistance, the student-attorney so informed the claimant by telephone and provided her with names and telephone numbers of other legal services providers in the area who might take her case. Regardless of the randomization result, we sent a letter to the claimant verifying her consent to participate in the study. After a few months, and pursuant to a prior arrangement, we sent the Massachusetts DUA a copy of the oral consent form and the follow-up letter along with an outcome form requesting information on the ALJ ruling, on relevant dates, and on whether the claimant was represented in the first-level appeal. Study intake began in the summer of 2008 and closed in the spring of 2010, by which time the 207 cases that form the basis of our analysis had cleared intake.⁹⁷

In any experiment, the randomization scheme is of considerable importance. In our study, the randomization scheme was entirely dictated by the needs of the HLAB program and the nature of unemployment first-level appeals. The speed at which first-level appeals move meant that HLAB needed to know within twenty-four hours (preferably sooner) whether it would offer to represent a potential client. Pedagogically, HLAB needed more cases in September through October than it did during other times of the year because

95. The script, approved by the Harvard Committee for Research on Human Subjects (as were all aspects of our study), is available on *The Yale Law Journal* website. *Telephone Consent Form*, YALE L.J., http://www.yalelawjournal.org/images/documents/greinerpattanayak_consent_form.pdf (last visited Apr. 3, 2012).

96. By report from the HLAB students, only two or three eligible claimants (out of over two hundred) refused to go forward once they found out about the study. These claimants were provided with names and telephone numbers of other service providers.

97. One claimant withdrew from the study after providing consent. A few others were deemed erroneously included when it became clear that their cases did not concern eligibility for benefits but some other issue, such as an overpayment, or when it became clear that their first-level appeals had already taken place.

in these months it had student-attorneys available who wanted to cut their teeth on unemployment cases. These factors led us to randomize each case separately using probabilities adjusted approximately every eight weeks to reflect the time of year.⁹⁸ This design was not our first choice. There are numerous ways to randomize, and a researcher should attempt to choose a design that will minimize variance, or the uncertainty in the estimation, because lower variance/uncertainty means greater knowledge. The design we used is associated with a higher degree of uncertainty than, for example, a design in which all 207 cases were available for randomization at the same time, with exactly 103 randomized to an offer of representation and exactly 104 to no offer.

E. Results

We organize our findings into answers to a series of eight questions. Our general strategy here is to present the results of several different statistical methodologies on each question, with results of the methodologies requiring the mildest assumptions first, followed by those requiring stronger assumptions. To be clear, in statistics-speak, “strong” assumptions are *less* desirable. All statistical techniques depend on some assumptions, so quantitative analysts attempt to minimize the number and strength of the assumptions needed to render results credible. Generally, stronger assumptions are required for modeling as opposed to nonmodeling techniques. The payoff is that modeling techniques tend to produce more precise answers, meaning less uncertainty. Thus, we present the results of the techniques with the weakest assumptions first. That said, “strength” of assumptions is a relative term. In this study, we are analyzing data generated from a randomized experiment. The hard work that went into generating randomized data means that all our primary results depend on assumptions that are quantum steps weaker (again, weaker is better) than those needed to render credible the dozens upon dozens of non-randomized studies we discuss in Part II. In short, we do not credit the findings in the observational studies discussed in Part II,

⁹⁸. At the beginning of the semesters, probabilities of acceptance were generally set high at .75 or .5. These were reduced over the semester to .5, or to .25. In a very small number of cases (during the summer, when HLAB was staffed in smaller part by incoming HLAB students and in larger part by students from other schools), probabilities were set as low as .15. Unfortunately, the number of cases in our study (207) did not all allow statistical analysis focusing on seasonal effects, although we did attempt (without real success) to explore whether callers in the November and December months (during the period of the Thanksgiving holiday, final exams, and between-semester break) had outcomes different from those who called at other times.

but we do believe all the results we present here, including those that require some modeling.⁹⁹

We have deliberately tailored our discussion so as to make it accessible to persons without quantitative training. In doing so, we have necessarily sacrificed the precision technical terms allow, as well as the ability to present all results our analysis produced.

Question 1: Did the randomization produce statistically equivalent “treatment” (an offer of HLAB representation) and “control” (no offer) groups?

In other words, so far as we can tell with observed variables, did the randomization do what it was supposed to do, i.e., create treated and control groups that are the same up to random variation except for the treatment? The answer appears to be yes, within expected limits. Given the number of variables we measured, random variation would be expected to produce dissimilarity between treated and control in an occasional variable; in our case, a misbehaving variable was gender in that the treatment group had significantly more men than did the control group.

Table 1 compares the treated group to the control group on the background variables listed and shows a measure of difference (which reflects the size of the population and the probability of being assigned to each group) called a standard deviation. As a very rough heuristic, a difference (positive or negative) of two standard deviations is suggestive of an imbalance. Only with respect to the gender variable is the difference in standard deviations close to two.

^{99.} For an approach even more skeptical than our own, one that would risk no modeling at all, see David A. Freedman, *Randomization Does Not Justify Logistic Regression*, 23 STAT. SCI. 237 (2008). Professor Freedman was perhaps the most articulate and gifted skeptic that the field of statistics has ever produced.

Table 1.

ALL CASES: COMPARISON OF BACKGROUND CHARACTERISTICS IN TREATED (HLAB OFFER) AND CONTROL (NO HLAB OFFER) GROUPS – WEIGHTED AVERAGES OR PROPORTIONS¹⁰⁰

CHARACTERISTIC	OVERALL	TREATMENT	CONTROL	STANDARD
	AVERAGE OR PROPORTION	(HLAB OFFER): AVERAGE OR PROPORTION	(NO HLAB OFFER): AVERAGE OR PROPORTION	DEVIATIONS DIFFERENCE
Number of cases ¹⁰¹	207	.78	.129	—
Log of days: Claims adjuster notice to HLAB intake interview	3.07	3.13	3.02	.97
Claims adjuster notice dated November or December	.14	.11	.16	-1.08
ALJ hearing scheduled at time of intake	.29	.31	.28	.52
Education: High school diploma or less	.56	.58	.55	.52
Number of dependants	.61	.66	.56	.72
Male	.53	.60	.46	2.28
Hispanic	.10	.11	.09	.49

¹⁰⁰. The figures in the third and fourth columns are weighted by inverse probability of selection, so the figures in column two are adjusted accordingly. Further details appear on *The Yale Law Journal* website. *Sampling Ratio Estimator*, YALE L.J., http://www.yalelawjournal.org/images/documents/sampling_ratio_estimator.pdf (last visited Apr. 3, 2012). The unweighted figures led us to the same substantive conclusion, i.e., that the treated and control groups were similar. The only variable with weighted values noticeably different from its unweighted figures was the variable “claims adjuster notice dated November or December.” Because, as explained at the end of Section I.D, the probability of receiving an HLAB offer depended on the time of year that the claims adjuster issues her decision, we place greater importance on the version of this variable that takes this fact into account, i.e., the weighted version.

RANDOMIZED EVALUATION IN LEGAL ASSISTANCE

CHARACTERISTIC	OVERALL	TREATMENT	CONTROL	STANDARD
	AVERAGE OR PROPORTION	(HLAB OFFER): AVERAGE OR PROPORTION	(NO HLAB OFFER): AVERAGE OR PROPORTION	DEVIATIONS DIFFERENCE
Discharge case (employer bears burden of persuasion) ¹⁰²	.48	.47	.49	-.33
English spoken at home	.87	.87	.86	.09
Needs interpreter at hearing	.08	.10	.05	1.39
Married	.24	.25	.22	.66
Non-Hispanic White	.42	.39	.43	-.67
Non-Hispanic Black	.45	.48	.43	.68
Non-Hispanic Asian	.03	.02	.04	-.78
Claims adjuster denied benefits (claimant appealing)	.62	.61	.63	-.35

¹⁰¹. Given the nature of the randomization scheme used (varying probabilities, many of which were not .5), we would not expect roughly equal numbers of cases in the treated and control groups, nor is balance on this number relevant to the inquiry.

¹⁰². See *supra* note 46 and accompanying text. Note that this variable (as were all others not determined by Massachusetts DUA records) was by claimant report; we had no way to verify whether the case actually stemmed from a claimant's having left her employment or having been discharged.

In addition to recording the variables listed above, we also asked the HLAB student-attorneys to rate each case after the intake interview on three criteria: the amount of time needed to prepare the case well, the potential client's ability to assist in the litigation, and the strength of the case. These were among the criteria that were used by HLAB in its case selection decisions prior to our study. We asked for ratings on a scale of one to five, where a one meant little time, low ability to assist, and a weak case, respectively. Although in the interests of space we do not show the results, balance on these three variables was good (the largest absolute difference in standard deviations was for the third of these variables, with a value of -.13). Including these variables as predictors in the "Separate Regressions" technique, described in the following subsections, did not change the results in any way. Perhaps more importantly, including them did not narrow the intervals we observed. One

With respect to missing data: the efforts of the HLAB student-attorneys, good communication and follow-up with HLAB staff, and a keep-it-simple approach to the intake process resulted in very, very little missingness on background variables. Only 6 of 207 (under 3%) of observations had any background, outcome, or other variable used in this Article missing, and only 11 of 3933 cell values used were missing. We used subject matter knowledge to state a random distribution that we used to fill each missing cell value five different times, creating five different “completed” datasets¹⁰³ to account for the uncertainty due to missing background variables. The fact is, however, that the amount of missingness was so small that the use of any reasonable method of accounting for missing data would result in the same substantive conclusions.

What can be done about the imbalance on the gender variable?¹⁰⁴ We adopted several approaches including using gender as a predictor in modeling, analyzing men and women separately (although doing so raised its own statistical challenges), and using methods designed for “broken” randomized experiments.¹⁰⁵ Although we do not show all these results, as it turned out, all these analyses made clear that the imbalance on gender did not affect the substantive conclusions we reached.

might criticize the five-point scale we used (and we would accept such criticism), but these results mildly suggest the possibility that it may be difficult to obtain useful information on whether representation is likely to make a difference in a particular case based solely on a telephone interview. *See Part III.*

103. Our inspiration here was Donald B. Rubin, *Multiple Imputations in Sample Surveys—A Phenomenological Bayesian Approach to Nonresponse*, in *PROCEEDINGS OF THE SURVEY METHODS SECTION* 20 (American Statistical Association ed., 1978). By “subject matter knowledge,” we mean that we spoke with HLAB students and with lawyers knowledgeable about the DUA process, then used the results of these conversations to guess what would constitute good predictions for missing values.
104. Some limited previous research has suggested that men are more likely to win first-level appeals than are women. *See* Catherine K. Ruckelshaus, *Unemployment Compensation for Victims of Domestic Violence: An Important Link to Economic and Employment Security*, 30 CLEARINGHOUSE REV. 209 (1996). Our own data suggested slightly the opposite.
105. Essentially, this involved using propensity score methods to compel balance on misbehaving covariates (which, as is often the case, involved degrading balance on other covariates). *See* Cassandra Wolos Pattanayak, *The Critical Role of Covariate Balance in Causal Inference with Randomized Experiments and Observational Studies* (May 2011) (unpublished Ph.D. dissertation, Harvard University) (on file with authors). To reiterate, the substantive results of these methods were the same as those shown above.

Question 2: Did an offer of HLAB representation increase the probability that the claimant would prevail?

There was no evidence that an offer of HLAB representation increased the probability that a claimant would prevail. In other words, if there was an increase (which no study, including ours, could rule out), it is unlikely to have been large. On (weighted) average, .76 of claimants who received an offer of HLAB representation prevailed in their first-level appeals, while a (weighted) average of .72 of those who did not receive an offer prevailed. This .05¹⁰⁶ difference is not statistically significant, and even this insignificant difference narrows with some relatively innocuous statistical modeling, where “innocuous” means far easier to credit than the assumptions required to make credible the studies we critique in Part II of this Article.¹⁰⁷ If we limit ourselves to inferences based almost entirely¹⁰⁸ on the randomization alone, a 95% confidence interval for the difference is (-.06, .16). If we risk some relatively innocuous statistical modeling,¹⁰⁹ the 95% interval shrinks to (-.06, .09). We credit the latter result.

A question immediately arises: when we say that it is unlikely that an HLAB offer caused a large increase in the probability that a claimant would prevail, how unlikely is “unlikely” and how large is “large”? We provide further details in a short appendix, but the basic message is that, without using the relatively innocuous statistical modeling referred to above, the most likely increase in the probability of a victory due to an HLAB offer is in the .00 to .05 range; the odds are about four to one against an increase of .10 in the probability of a victory; and an increase of .15 or larger is unlikely (by our estimate, the odds of such an increase are approximately nineteen to one against). If we do use the relatively innocuous statistical modeling, an increase of .09 or greater in the probability of a victory is unlikely (again, meaning roughly 19:1 against). See the Appendix for further details.

To assess the effect of a HLAB offer, we first present the results of a “Fisher,” or permutation, test. The idea here is that we begin with a “null

¹⁰⁶. The (weighted) treated group win rate for the HLAB offer group was .764. The (weighted) control group win rate was .717. These figures thus round to .76 and .72, but the difference of .047 rounds to .05.

¹⁰⁷. For further discussion of the “relatively innocuous” assumptions imposed by our modeling technique, please see *infra* Appendix.

¹⁰⁸. These figures come from the “Sampling, Ratio Estimator” technique appearing in Figure 2; further explanation appears *infra* note 116.

¹⁰⁹. These figures come from the “Separate Regressions” technique appearing in Figure 2; further explanation appears *infra* note 115.

hypothesis" that an offer of HLAB representation made no difference in whether each claimant won or lost. We then ask this question: if it were true that each claimant's outcome would have been exactly the same with or without an HLAB offer, how likely is it that the randomization alone would have produced an absolute difference in win rates as large as, or larger than, the 5% we actually did see? If we find that the randomization alone would have been *unlikely* to produce the difference we actually saw (or a larger one), then we would reject the "null hypothesis" and conclude (essentially) that the offer of HLAB representation made a difference. How unlikely is "unlikely"? For this particular analysis it makes little difference, because we were nowhere close, but generally, quantitative analysts like to see what they refer to as a "p-value" of something near .05 or less to begin considering the result sufficiently "unlikely." Note also that the Fisher Test requires only the mildest of assumptions for credibility. It depends only on the scheme randomly assigning claimants to "treatment" (HLAB offer) or "control" (no HLAB offer). We know of no test depending on fewer or milder assumptions.

Figure 1 presents a histogram for the Fisher test, with the observed result indicated by a vertical spike.¹¹⁰ The histogram represents 500,000 simulations of what we might have seen if the null hypothesis were true. If the observed results had been sufficiently "unlikely" in the sense described in the previous paragraph, this vertical spike would have been to the far right (or left) of the graph. Again, that would have indicated that the observed difference in outcomes was unlikely to be due to chance, assuming the null hypothesis that the HLAB offer had no effect. In fact, the spike is near the center of the histogram, indicating that the 5% difference we saw between treated and control groups, as indicated by the black line, is easily attributable to chance (under the null hypothesis). The p-value for this test is .51,¹¹¹ nowhere near .05. Thus, with apologies for the triple negative that the nature of statistical analysis imposes: the data give us no reason to disbelieve the "null" hypothesis that the HLAB offer made no difference in the probability of a claimant win. In less technical terms: the offer appears to have had no large¹¹² effect on win/loss.

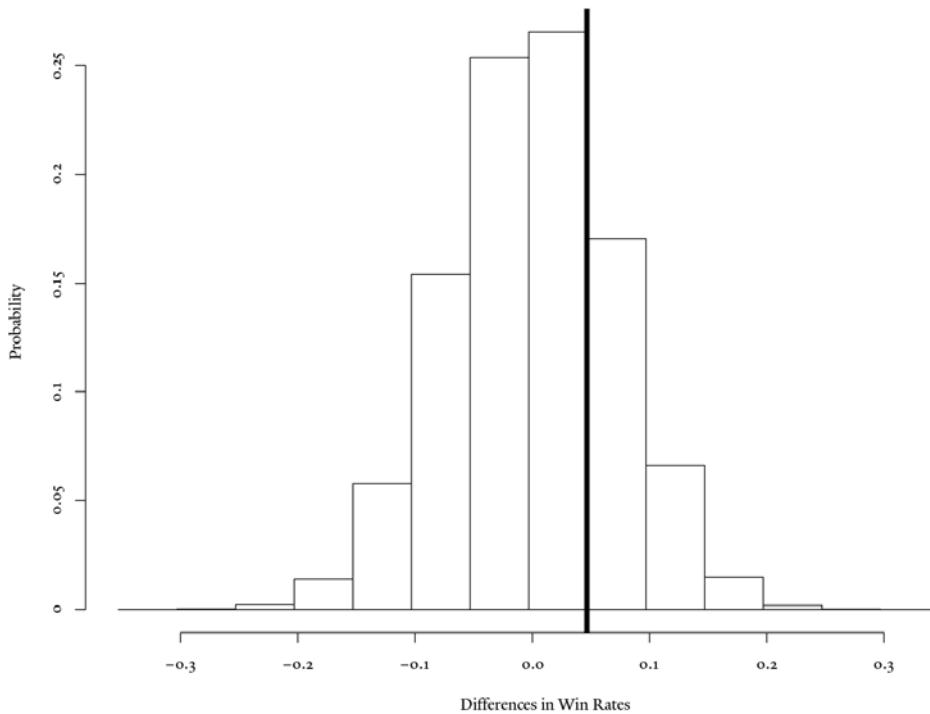
¹¹⁰. As there were 2^{207} (more than 2×10^{62}) possible randomizations, we did not calculate the probabilities and treatment effects for each one. Instead, we used computer simulation techniques; the specifics are available on the *Yale Law Journal* website. *HLAB Study Code*, YALE L.J., <http://www.yalelawjournal.org/images/documents/hlabstudycode2011009.r> (last visited Apr. 3, 2012).

¹¹¹. We had previously reported Fisher test results in terms of one-sided tests, which, obviously, had p-values half as large as two-sided tests (.255 in this case). At the urging of several reviewers, we have switched to a two-sided test.

¹¹². See *infra* Appendix.

Figure 1.

ALL CASES: FISHER TEST FOR DIFFERENCE IN WIN RATES CAUSED BY HLAB OFFER,
CASES WEIGHTED



As noted above, a strength of the Fisher test is that it depends on only the mildest of assumptions, but two of its weaknesses are that it cannot, for a 0-1 outcome such as claimant success,¹¹³ produce an interval into which the true effect of an HLAB offer is likely in some sense to fall,¹¹⁴ and it does not use any of the background information that we collected on each case. Accordingly, Figure 2 presents the results of two other statistical techniques, both of which produce intervals, and the latter of which uses the background variables

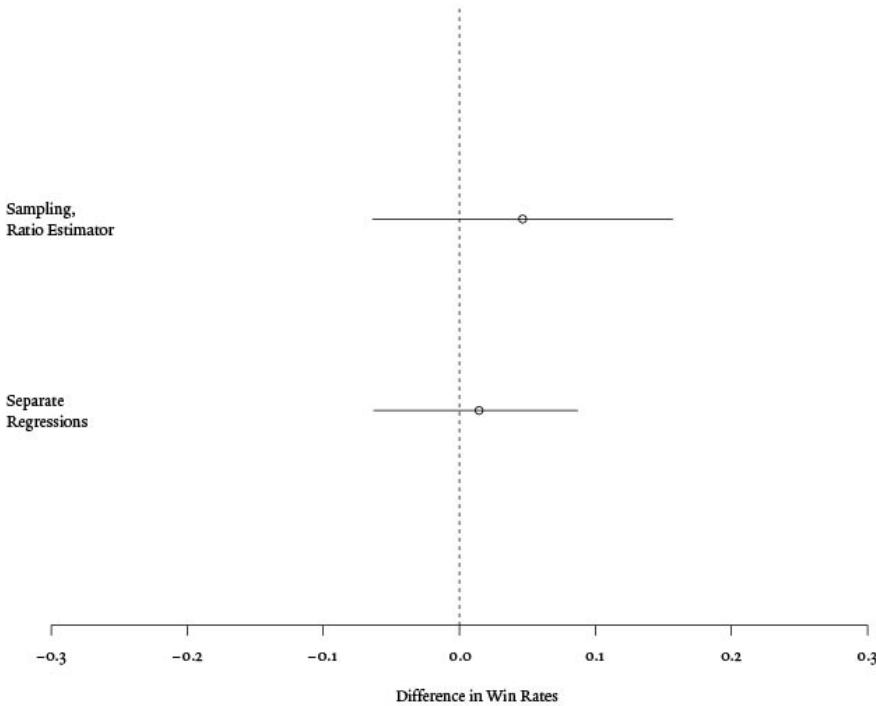
¹¹³. For a 0-1 outcome, we have discovered no intuitive way to fill in the missing potential outcomes so as to produce an interval for a treatment effect.

¹¹⁴. As we present both Bayesian and frequentist results here, we suppress philosophical debates about the meaning of intervals.

collected (see Table 1). The bottom result uses modeling¹¹⁵ and thus depends on stronger assumptions, but again, the assumptions required here are relatively innocuous.¹¹⁶

Figure 2.

ALL CASES: 95% INTERVALS FOR DIFFERENCE IN WIN RATES CAUSED BY HLAB OFFER



¹¹⁵ “Modeling” here refers to the use of mathematical equations that relate variables listed in Table 1 to outcome variables, such as win/loss and time to adjudication. Modeling requires us to assume that the equations used are similar enough to the process that generated the data.

¹¹⁶ For the “Sampling, Ratio Estimator,” the point estimate comes from a difference in ratio estimators for means of the treated and control potential outcome vectors. Details of this estimator and corresponding variance calculations appear on the *Yale Law Journal* website. *Sampling Ratio Estimator*, *supra* note 100. The estimator relies on a first-order Taylor approximation that may underestimate the variance. The interval is plus/minus 1.975 (the relevant t-distribution quantile) standard deviations. For the “Separate Regressions,” we used a Bayesian simulation analog to the Peters-Belson method. See William A. Belson, *A Technique for Studying the Effects of a Television Broadcast*, 5 J. ROYAL STAT. SOC’Y: SERIES C APPLIED STAT. 195 (1956); Charles C. Peters, *A Method of Matching Groups for Experiment with No Loss of Population*, 34 J. EDUC. RES. 606 (1941). Again, details are available on the *Yale Law Journal* website. *HLAB Study Code*, *supra* note 110.

The intervals here are 95% intervals, meaning that the true effect of an HLAB offer is in some sense 95% likely to fall within the interval, but that values nearer to either end of an interval are less likely to correspond to the truth than are values nearer the middle. The “Sampling, Ratio Estimator” method produces an interval of (-.06, .16), while the “Separate Regressions” method interval is (-.06, .09).

The pattern here is straightforward: both intervals in Figure 2 easily include zero, indicating that there is no evidence that an offer of HLAB representation increases (or changes) the probability that the claimant will win her first-level appeal. Stated another way, any effect due to the HLAB offer is unlikely to have been large.¹¹⁷ The “Separate Regressions” technique uses modeling, which involves stronger assumptions, so the interval is narrower than the interval using the Ratio Estimator, but the basic conclusion is the same. Our preference is for this “Separate Regressions” method, which we believe represents a reasonable balance between the risks associated with some modeling and the desire to obtain more precise information. But all the techniques presented here produce the same inference, and none suggests that the HLAB offer has much of an effect on the probability of a claimant victory.

One final note: all the information here on claim success is with respect to the ALJ adjudication; the Massachusetts unemployment system provides a second-level appeal to a Board of Review as well as a right to judicial review of the Board’s decision. We requested from the Massachusetts DUA information about the ultimate disposition of each case, i.e., the outcome of the claim if the case proceeded beyond the ALJ level to the Board of Review or to the judiciary. The Massachusetts DUA agreed to give us the relevant information. It turned out that approximately thirty-five of the cases in our dataset proceeded beyond the ALJ level. As of the time of this writing, approximately thirty-two cases have reached final decisions, all but one¹¹⁸ of which resulted in affirmation of the ALJ decision. With this in mind, there is essentially no chance that the substantive conclusions of this Article will change as the three pending cases reach final adjudication.

Question 3: Did an offer of HLAB representation delay adjudication?

The short answer is yes. On (weighted) average using a logarithmic mathematical transformation explained immediately below, cases receiving an

¹¹⁷. See *infra* Appendix.

¹¹⁸. See *infra* note 257 for an explanation of this case. We are not certain whether to code this as a full reversal of the ALJ decision, but because this is only one case, our results are substantively the same even if we do code this case as a full reversal.

HLAB offer took 53.1 days from HLAB interview to ALJ decision,¹¹⁹ while those not receiving the offer took 37.3 days. The sixteen-day difference¹²⁰ is statistically significant. With some relatively innocuous modeling the difference shrinks slightly: the median effect is a delay of about 12.1 days, with a median meaning that the HLAB offer would cause a delay longer than 12.1 in about half of the cases. On a different scale, a scale upon which we will place more reliance, the offer of HLAB representation effectively multiplied the time from HLAB interview to ALJ decision by about 1.4, corresponding to an increase of about 40%. All these results are statistically significant.

In presenting most of our results on time and delay, we will work with the “logarithm” of the number of days¹²¹ rather than the “raw” number of days. Why the logarithm? For most of the techniques we use, it is better to work with a set of numbers that approximates a symmetric, bell-shaped curve when graphed. For our raw measurements of time from HLAB interview to ALJ decision, the values did not look like a bell curve because several cases took much longer than was typical, giving the graph a long tail on its right side. When this happens, as it commonly does for measurements of time, quantitative analysts often apply a mathematical transformation called a logarithm to the raw numbers, hoping to transform the set of numbers into something that more closely resembles the classic bell curve. Here, that hope was realized, so we used the logarithm results.

A byproduct of applying this mathematical transformation is that most of our results will not be in terms of raw differences in number of days but rather in terms of multipliers. Recall that multiplying a number by one changes nothing. If on average an offer of HLAB representation effectively multiplied the length of time from HLAB interview to ALJ decision by a number greater than one (and the result was statistically significant), then we have evidence that the HLAB offer caused a delay. If the offer effectively multiplied the time from interview to decision by a number less than one (and the result was statistically significant), then we have evidence that the HLAB offer sped things up. One can think of this roughly in terms of percentage increases. A multiplier of, say, 1.25 means that the HLAB offer increased the length of time from HLAB interview to ALJ decision by about 25%.

¹¹⁹. Technically, we are measuring from HLAB interview to either ALJ decision or a withdrawal of the appeal by the party who brought it.

¹²⁰. The difference without the logarithmic transformation is around eleven days. We prefer the logarithmic and modeling results described in the text.

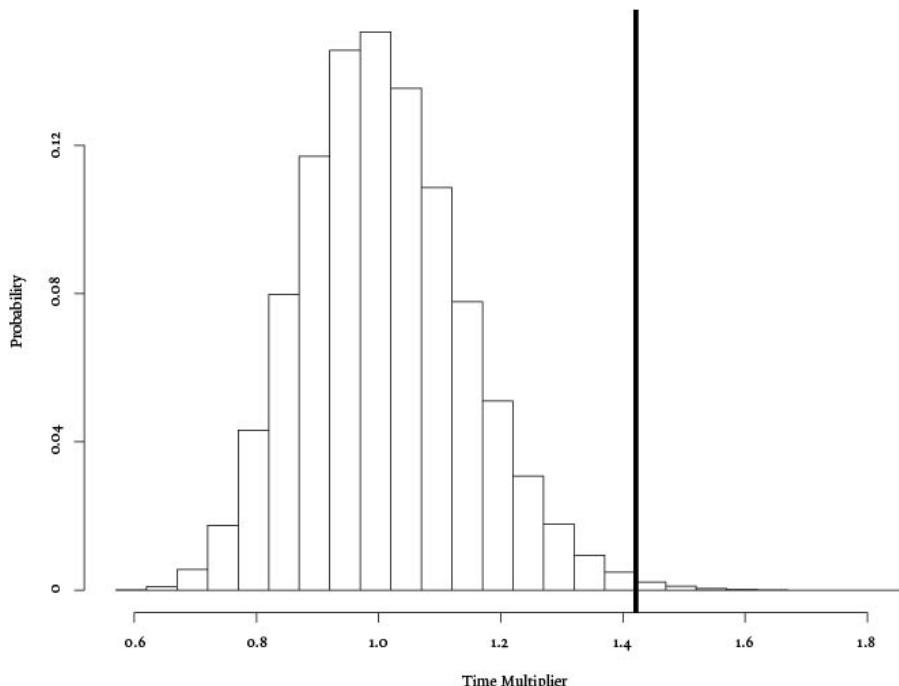
¹²¹. We actually use $\log(1 + \text{number of days})$ for mathematical convenience.

We proceed as before, showing results in order of increasingly strong assumptions, but with (again) the reminder that assumptions involved here are of a lesser order of magnitude than those involved in a non-randomized study. Figure 3 is the length-of-time counterpart to Figure 1, above. It shows the results of a Fisher test for average of the logarithm of number of days from the date of the HLAB interview (which was ordinarily within twenty-four hours of the date the case was randomized) to the date the ALJ rendered her decision. Analogously to Figure 1, the null hypothesis here is that the HLAB offer of representation had no effect on the average of the logarithm of the number of days from HLAB interview to ALJ decision, meaning that the multiplier due to an HLAB offer was one. Figure 3 is the histogram showing 500,000 simulations¹²² of what we might have seen if the null hypothesis had been true, and the observed value is indicated by a vertical spike. Here, in contrast to Figure 1, that observed value is on the far right of the histogram, suggesting that randomization is unlikely to have produced the observed result were the null hypothesis true. The P-value is approximately .008, far less than the .05 that often serves as a rough benchmark for statistical significance. Thus, we have reason to reject the null hypothesis that the HLAB offer caused no change in the log number of days. Figure 3 is on the multiplier scale, so values greater than one indicate a delay due to the HLAB offer. Again, we are aware of few statistical techniques for drawing causal inferences that depend on assumptions weaker than those of the Fisher test. Thus, the fact that this test shows statistically significant results is strong evidence of delay.

^{122.} See *supra* notes 110-111 and accompanying text for an explanation of how we implemented the Fisher test.

Figure 3.

ALL CASES: FISHER TEST FOR TIME (FROM HLAB INTERVIEW TO ALJ DECISION)
MULTIPLIER CAUSED BY HLAB OFFER (CASES WEIGHTED)



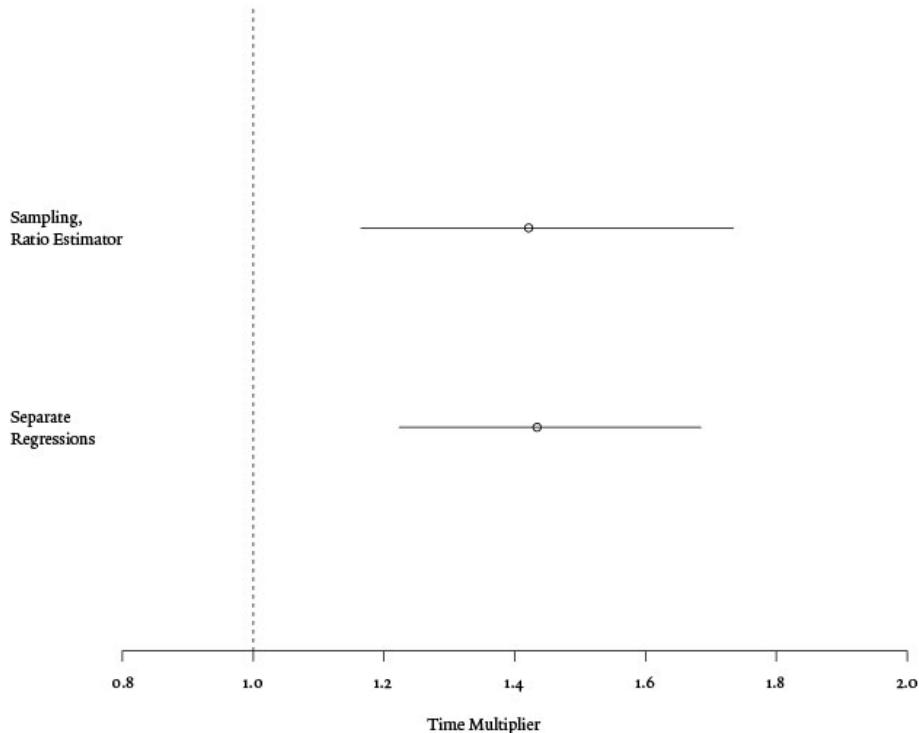
What about producing intervals and using the background variables listed in Table 1? Figure 4 is the analog of Figure 2, above, and it has the same structure. Thus, it presents the results of two other statistical techniques, both of which produce intervals, and the latter of which uses the background variables collected. The bottom result depends on modeling and thus depends on stronger assumptions, but again, the assumptions required here are comparatively innocuous.¹²³ Note that Figure 4 (like Figure 3) is also on the multiplier scale, so values greater than one indicate a delay due to the HLAB offer. The intervals here are 95% intervals, meaning that the true effect of an HLAB offer is in some sense 95% likely to fall within the interval, but that values nearer to either end of an interval are less likely to correspond to the truth than are values nearer the middle. The 95% interval for the Sampling, Ratio Estimator technique is (1.2, 1.7), and the corresponding interval for the

¹²³. See *supra* note 116 for an explanation of the estimation techniques used here. Further details are available on *The Yale Law Journal* website, *HLAB Study Code*, *supra* note 110.

Separate Regression technique is also (1.2, 1.7) (the latter interval is slightly narrower in a way that rounding masks).

Figure 4.

ALL CASES: 95% INTERVALS FOR TIME (FROM HLAB INTERVIEW TO ALJ DECISION)
MULTIPLIER CAUSED BY HLAB OFFER



The pattern here is straightforward. Both intervals in Figure 4 are above one, indicating that an offer of HLAB representation causes a statistically significant increase in the length of time from HLAB interview to ALJ adjudication. The bottom interval is the result of some modeling; because stronger assumptions are required, a slightly narrower interval results. Our preference is for the “Separate Regressions” results shown above, which we believe represent a reasonable balance between the risks associated with some modeling and the desire to obtain more precise information. But the results of all the techniques presented here are essentially the same, and all suggest that the HLAB offer causes a delay.

How much of a delay is caused? According to the “Separate Regressions” technique, the median number of delay days caused by the HLAB offer is

approximately 12.1, with the median meaning that half of cases will have a larger delay. Again, to put this result in context, DOL regulations suggest that 60% of all first-level appeals should be completed within thirty days of filing, and that 80% be completed within forty-five days.¹²⁴ In another context, a median two-week delay might not matter; in this system, the delay is nontrivial.

Question 4: Did the subset of claimants who were initially denied benefits also experience a delay in their proceedings with no apparent, or no apparently large, effect on the win rate?

Acknowledging that answers here will be less certain because we are working with a smaller dataset, the answer appears to be a tentative yes. Focusing on only the cases in which the claims adjuster initially denied the claim, thus making the claimant the “appellant” requesting the mini-trial de novo before the ALJ, we see the following results. First, the randomization produced reasonably comparable treated (HLAB offer) and control (no HLAB offer) groups, with one chance exception that, unfortunately, might be important: the (log of the) amount of time the claimant allowed to elapse between receiving notice of the initial claim and the HLAB interview. Second, the offer of HLAB representation had no statistically significant effect on the probability that the claimant would win, although the effects would have had to have been large for us to have seen them. Of those offered HLAB representation, .70 prevailed on weighted average, as compared to .65 of those not offered HLAB representation. The difference is not statistically significant, and it narrows when we do a small amount of modeling (which requires modeling assumptions but is probably worthwhile given the smaller number of observations and the imbalance in the time variable mentioned above). Third, on (weighted) average, cases receiving an HLAB offer took 45.9 days from HLAB interview to ALJ decision, while those receiving no HLAB offer took 37.6 days. Analyzed on the log scale, this eight-day difference is statistically significant, but the eight-day figure could be an understatement, perhaps stemming from the fact that (as evidenced by the imbalance in log time from initial claim to HLAB interview) the treated and control groups differed on when they contacted HLAB, which might indicate differing levels of procrastination or assertiveness among claimants. With some relatively innocuous modeling, the median effect is a delay of about 10.2 days, with the median meaning that the HLAB offer caused a delay longer than 10.2 days in half of the cases.

¹²⁴. See *supra* note 69 and accompanying text.

As we have stated several times before, all the results herein are limited by statistical uncertainty. Ordinarily, uncertainty goes up as the number of observations goes down. Thus we would expect to see, and do see, a greater degree of uncertainty when we examine subsets of the overall data, including the subset of initially denied claimants examined in this subsection. We expect, and in fact observe, less definitive results in this subset.

First, with respect to the comparability of the treated and control groups, Table 2 is the analogue of Table 1, above, in that it compares the composition of the treated (HLAB offer) and control (no HLAB offer) groups on background variables, but for initially denied claimants only. Comparing the fifth column in Table 2 to the fifth column in Table 1 shows that the standard deviations of the differences tend to be a bit larger in Table 2, suggesting that the treated and control groups in the initially denied claimant subset are not quite as well balanced as in the overall set of claimants. That said, the only variable that looks unbalanced is the logarithm of the number of days from the notice of the claims adjuster determination to the HLAB intake interview.

Table 2.

INITIALLY DENIED CASES ONLY: COMPARISON OF BACKGROUND CHARACTERISTICS IN TREATED (HLAB OFFER) AND CONTROL (NO HLAB OFFER) GROUPS – WEIGHTED AVERAGES OR PROPORTIONS¹²⁵

CHARACTERISTIC	OVERALL AVERAGE OR PROPORTION	TREATMENT (HLAB OFFER): AVERAGE OR PROPORTION	CONTROL (NO HLAB OFFER): AVERAGE OR PROPORTION	STANDARD DEVIATIONS DIFFERENCE
Number of cases	126	44	82	—
Log of days: Claims adjuster notice to HLAB intake interview	2.76	2.92	2.61	2.32
Claims adjuster notice dated November or December	.14	.15	.14	.15

¹²⁵. The discussion *supra* notes 100-102 applies equally to Table 2. We have deleted the Non-Hispanic Asian variable from this table due to the small number of such claimants in this subset of the data.

CHARACTERISTIC	OVERALL AVERAGE OR PROPORTION	TREATMENT (HLAB OFFER): AVERAGE OR PROPORTION	CONTROL (NO HLAB OFFER): AVERAGE OR PROPORTION	STANDARD DEVIATIONS DIFFERENCE
ALJ hearing scheduled at time of intake	.25	.31	.19	1.53
Education: High school diploma or less	.61	.65	.58	.77
Number of dependants	.71	.88	.57	1.48
Male	.58	.62	.54	1.01
Hispanic	.12	.15	.09	1.03
Discharge case (employer bears burden of persuasion)	.55	.52	.58	-.65
English spoken at home	.86	.83	.88	-.92
Needs interpreter at hearing	.11	.16	.07	1.63
Married	.24	.27	.21	.82
Non-Hispanic White	.43	.38	.47	-.100
Non-Hispanic Black	.43	.47	.40	.72

The imbalance on log time from notice of claims adjuster determination to HLAB interview may be problematic because the length of the time lapse from claims adjuster determination to HLAB interview might be a signal for characteristics, such as a tendency to procrastinate, that could affect our primary outcomes. Accordingly, we use some modeling to address this imbalance.

In terms of win rates, Figures 5 and 6 are the analogues to Figures 1 and 2 above, except the dataset is limited to the set of cases for which the claims adjuster initially denied benefits. The Fisher test,¹²⁶ with a p-value of .57 (nowhere near .05), shows no reason to disbelieve the null hypothesis that the HLAB offer caused no difference in the win rates. The intervals for the difference in win rates produced via both the ratio estimator and separate regression methods shown include zero, although the interval for the ratio estimator is wide, reflecting a fair amount of uncertainty. Thus, our

¹²⁶. See *supra* notes 110-111 and accompanying text for an explanation of the Fisher test and specifics of our implementation.

conclusions on this point depend to a greater degree on the modeling assumptions inherent in the separate regressions technique.

Figure 5.

INITIALLY DENIED CASES ONLY: FISHER TEST FOR DIFFERENCE IN WIN RATES CAUSED BY HLAB OFFER, CASES WEIGHTED

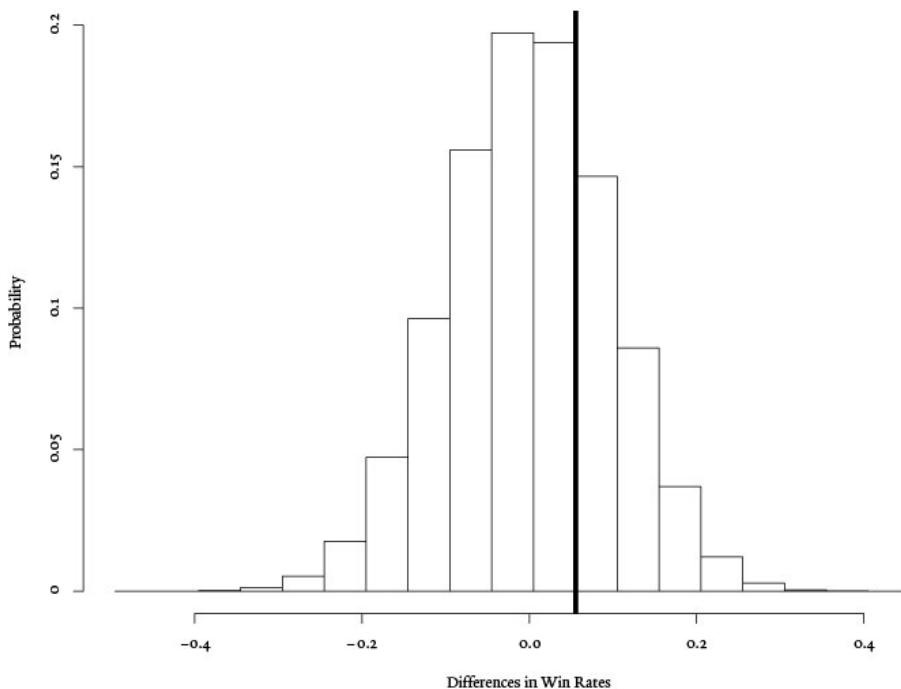
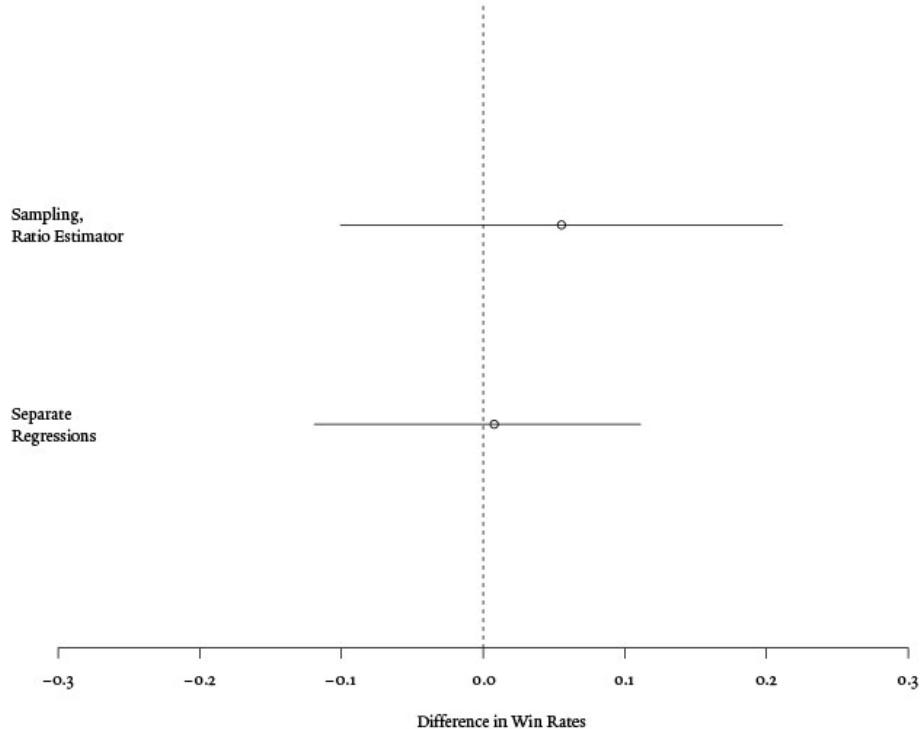


Figure 6.

INITIALLY DENIED CASES ONLY: 95% INTERVALS FOR DIFFERENCE IN WIN RATES CAUSED BY HLAB OFFER¹²⁷



With respect to the effect of the HLAB offer on delay in adjudication, the results are mostly the same as for the entire set of cases but less definite, with the lower degree of certainty attributable to the smaller size of the dataset being analyzed. Figures 7 and 8 correspond to Figures 3 and 4, above. As before, both figures are on the multiplier scale. Figure 7 shows the Fisher test on the log number of days from HLAB interview to ALJ decision. The observed value is on the right side of the histogram, although not on the far right. The p-value¹²⁸ is .16; if this had been the only test we had run, we would conclude that we lacked sufficient evidence to disbelieve the null hypothesis of no delay. As suggested above, however, the Fisher test makes very few assumptions. This fact, together with the smaller size of the dataset, means that we might expect to see less certainty than we would from the whole dataset or from a technique that risks more assumptions. Keeping in mind that the Fisher test does not

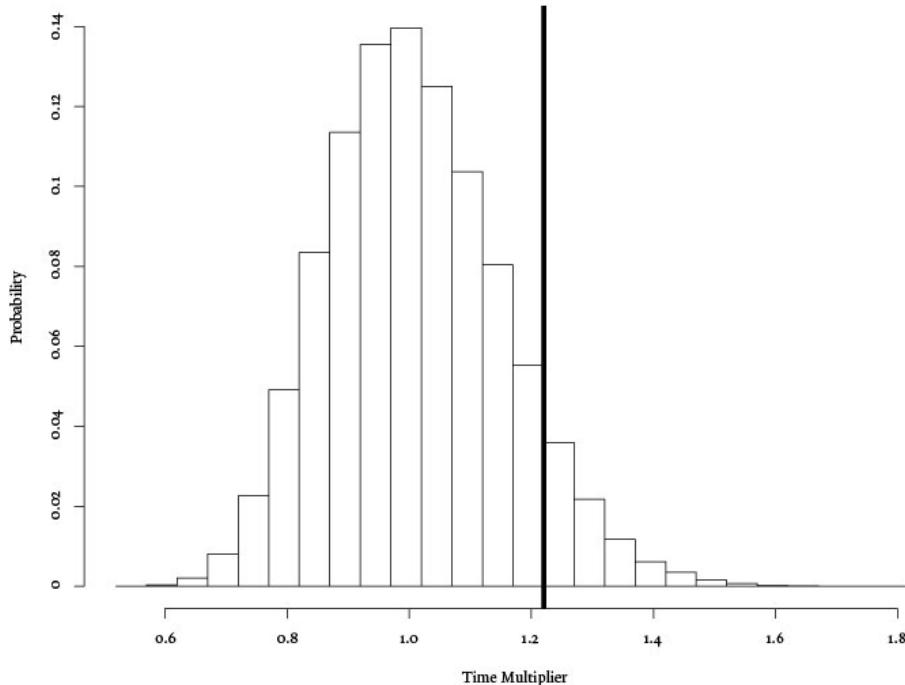
¹²⁷. See *supra* note 116 for an explanation of the techniques used.

¹²⁸. See *supra* note 110-111 and accompanying text.

adjust for the imbalance noted earlier on (log) time from claims adjuster determination to HLAB interview, we view the .16 P-value as inconclusive.

Figure 7.

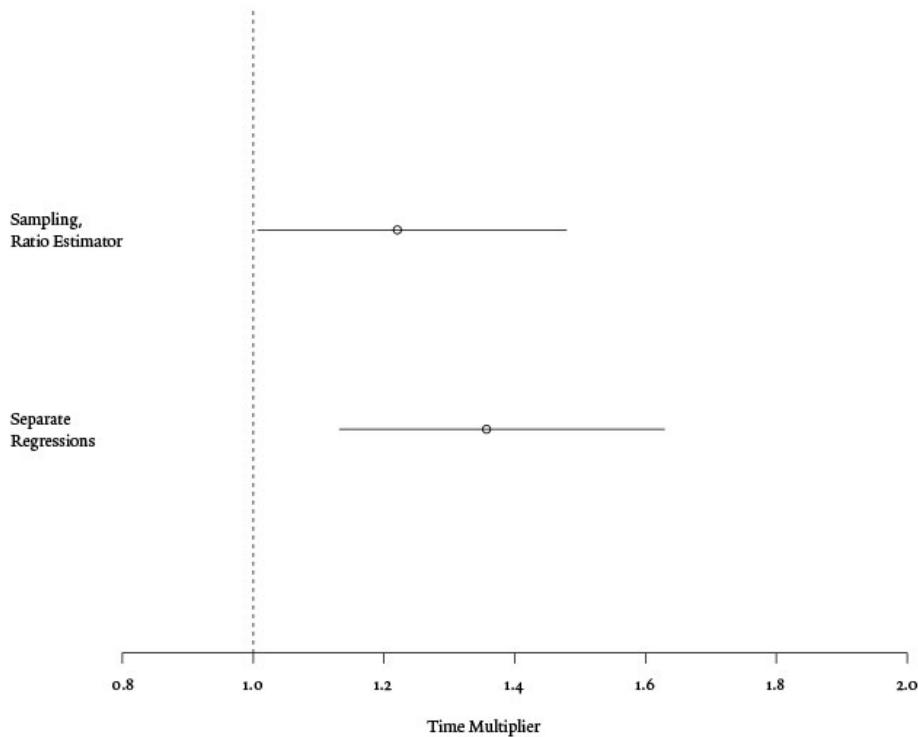
INITIALLY DENIED CASES ONLY: FISHER TEST FOR TIME (FROM HLAB INTERVIEW TO ALJ DECISION) MULTIPLIER CAUSED BY HLAB OFFER, CASES WEIGHTED



The other statistical techniques, the results of which appear in Figure 8, reach a different conclusion, which we credit here. Because of the multiplier scale, we look to see whether the intervals are greater than one and find that both are. In particular, keeping in mind the fact that the “Separate Regression” technique attempts to adjust for the imbalance on (log) time from claims adjuster determination to HLAB interview (at the risk of assumptions regarding the approximate correspondence of the data and the mathematical equations the models used), we credit the “Separate Regressions” result appearing in Figure 8. We conclude that the HLAB offer caused a delay among initially denied claimants.

Figure 8.

INITIALLY DENIED CASES ONLY: 95% INTERVALS FOR TIME (FROM HLAB INTERVIEW TO ALJ DECISION) MULTIPLIER CAUSED BY HLAB OFFER¹²⁹



Question 5: Did an offer of HLAB representation impose costs on the unemployment system?

Due to the delay noted in response to Question 3, the answer is a tentative yes. The Supreme Court's decision in *Java*¹³⁰ required that, if the claims adjuster granted benefits, such benefits should begin to flow immediately, even if an employer appealed. If, however, the employer appeal was successful, then payments paid during the pendency of the appeal were erroneous, and the longer the pendency of the appeal, the more money paid in error. The Massachusetts DUA attempted to recover such overpayments, primarily through recovery of other unemployment benefits or withholding of tax

¹²⁹. See *supra* note 116 for an explanation of the techniques used.

¹³⁰. Cal. Dep't of Human Res. Dev. v. Java, 402 U.S. 121 (1971).

refunds, but the agency could not succeed 100% of the time.¹³¹ Thus, for the set of claimants initially but erroneously granted benefits, delay placed a financial burden on the unemployment system, perhaps not in any individual case but certainly over the run of cases.

Putting aside recuperation, how much of a possible loss is involved here? The HLAB offer caused a longer delay for claimants initially granted benefits than for claimants initially denied. Using the “Separate Regressions” technique described above,¹³² an HLAB offer multiplied the length of the time period from HLAB interview to ALJ adjudication for initially granted claimants by about 1.4. In our dataset, there were thirteen claimants who were initially granted benefits but for whom the ALJ reversed in the first-level appeal; note that the small number of claimants with this profile suggests that this issue, to the extent that it exists, may not be a large one. Applying the delay multiplier of 1.4 together with the benefits rate for these claimants yielded dollar figures (ranked in increasing order and rounded to the nearest dollar) of cost due to delay of 163, 224, 244, 296, 318, 1106, 1546, 1797, 1888, 2200, 3013, 3935, and 4243. This is fairly back-of-the-envelope analysis, and the figures are sensitive to our choice to use a multiplier model; further, as one can see, these numbers are noisy. If one is willing to risk a rough inference on these thirteen observations, however, then an HLAB offer may cost the system something on the order of \$1546 (the median of the above figures) per case in which the claimant was initially granted benefits but for which the ALJ reversed, before the Massachusetts DUA’s recuperation efforts, which we were unable to measure.

¹³¹. Telephone Interview with Robert Ganong, Chief Counsel and Special Assistant Attorney Gen., Mass. Dep’t of Workforce Dev. (Nov. 29, 2010) [hereinafter Ganong November Interview]. We were unable to obtain figures regarding what percentage of overpayments are ultimately recovered.

¹³². See *supra* note 116 for brief explanation of this technique, as well as *The Yale Law Journal* website for our replication code. *HLAB Study Code*, *supra* note 110. Essentially, we applied the Separate Regressions technique to only the cases in which the claims adjuster initially awarded benefits. The result was a point estimate of a 1.4 times multiplier. We then applied this 1.4 time multiplier to the cases in which the claims adjuster had initially awarded benefits but the ALJ had reversed, in order to calculate a number of weeks of delay due to the HLAB offer for these cases (keep in mind here that times for cases in which the claims adjuster had initially granted benefits were generally longer than those for which the claims adjuster had initially denied). We then multiplied this number of weeks of delay to the weekly benefits amount being paid in the case prior to the ALJ reversal to obtain the dollar figures quoted above.

Question 6: When we shift focus from an offer of HLAB representation to the actual use of representation (from HLAB or some other source), do we still find a delay and no statistically significant effect on the win rate?

The fundamental results of this paper are that an *offer* of *HLAB representation* had no statistically significant effect on the win rate but did delay adjudication. Do these two fundamental results hold when we shift from an offer to *actual use* of representation and shift from HLAB to *any source or kind* of representative? In other words, what was the effect of representation of any kind? (Note that, as we have said repeatedly and will say again, we are still limited here to an inference for the set of claimants who sought representation from HLAB.) Again, offers of representation are different from actual use; not all persons offered representation took advantage, and some claimants not offered HLAB representation found assistance from other sources.

It turns out that the question is challenging to answer.¹³³ The reason is that, as noted above,¹³⁴ only the offer of, not actual use of, representation was randomized, and the randomization is what produces treated and control groups identical up to random variation. Some claimants (under 10%) randomized to receive an HLAB offer of representation did not in fact receive representation, ordinarily because HLAB was unable to contact them before their ALJ hearings. Some claimants¹³⁵ randomized not to receive an HLAB offer obtained representation from other service providers, such as private attorneys, Greater Boston Legal Services,¹³⁶ the Volunteer Lawyers Project,¹³⁷ or the

¹³³. Those familiar with the potential outcomes framework of causation will recognize that our discussion here is a wholesale adaptation of that in Hirano et al., *supra* note 34; our modeling strategy also comes from this piece. As this reference demonstrates, the techniques we use here are a form of instrumental variables modeling in which some of the less plausible assumptions typically involved in this technique are relaxed.

¹³⁴. See *supra* Introduction, Section B.

¹³⁵. The range of possible representation rates stems from the fact that some claimants and some employers withdrew from the appeals process before a hearing occurred. In such cases, the Massachusetts DUA was unable to tell us whether the claimants were represented. It is theoretically possible that as many as 60% of the no-offer group ended up represented, but the more likely value is around 39%. The uncertainty stems from a combination of data missing from the Massachusetts DUA as well as uncertainty as to whether claimants who withdrew their unemployment claims (or who did not appear at ALJ hearings) were represented. In the analyses reported in this Section, we alternatively assumed the 60% and the 39% figures. As it turned out, it did not matter what we assumed here because we could not produce useful inferences regarding the probability of obtaining benefits under any assumption. See *infra* note 142 and accompanying text.

¹³⁶. See GREATER BOS. LEGAL SERVS., <http://www.gbls.org> (last visited Apr. 3, 2012).

clinical program at Northeastern Law School.¹³⁸ We could not randomize the process by which claimants designated to receive an HLAB offer ended up unrepresented, nor could we randomize the process by which claimants designated to no HLAB offer ended up represented. There is no reason to guess that nature randomized these processes for us. In fact, there is reason to guess to the contrary. We find it plausible, even likely, that claimants the randomizer selected to receive no HLAB offer who nevertheless found attorneys represented a “Go-Getter” subset of the control group. This is essentially one of the “selection effects” we discuss in greater detail in Part II; we contend that this selection effect is one of the reasons why the observational studies done in this area previously are unworthy of credence, and this selection effect poses similar challenges for us here.

To make this point clearer, our intuition is that the process of finding an attorney once turned down by HLAB required effort, energy, articulateness, and persistence from a claimant. Our intuition also suggests that effort, energy, articulateness, and persistence are useful in winning a first-level appeal. The fact that the group of claimants randomized to no HLAB offer included some persons who obtained representation and some persons who did not, together with our guess about effort and persistence, lead us to guess that the group of claimants randomized to no HLAB offer was a mixture of (i) “Go-Getters” who did the extra work and had the extra skills to obtain attorneys despite disappointment at HLAB and (ii) “Regulars” who did not do this work and/or lacked these skills and thus who did not obtain attorneys. Thus, the control group (no HLAB offer) was a mixture of Go-Getters and Regulars, and we guess that the skills the Go-Getters possessed would have made them more likely to win their first-level appeals than the Regulars. But if the control (no HLAB offer) group was a mixture of Go-Getters and Regulars, so was the treated (HLAB offer) group. Why? Because random assignment of claimants to treated and control groups assured that the two groups were statistically the same on background characteristics. Thus, both treated and control groups were mixtures of Go-Getters and Regulars.

With this in mind, consider what would happen if we compared the success rates of (i) the set of all treated (HLAB offer) claimants as a whole versus (ii) a *subset* of control (no HLAB offer) claimants, specifically those who did not find legal representation. We would be comparing a set that contains a mixture of Go-Getters and Regulars (the whole of the treated group) to a set

^{137.} See VOLUNTEER LAWYERS PROJECT OF THE BOS. BAR ASS’N, <http://www.vlpnet.org> (last visited Apr. 3, 2012).

^{138.} See *Law School Clinics*, NE. UNIV. SCH. OF LAW, <http://www.northeastern.edu/law/academics/clinics/index.html> (last visited Apr. 3, 2012).

consisting of only Regulars (the claimants in the control group who did not obtain attorneys), even though we have reason to believe that Go-Getters might be inherently more likely to win their cases because they might be more able to find documents, convince witnesses to testify, and so forth. These two sets, (i) all of the treated versus (ii) only those controls who did not find attorneys, are *not* essentially identical in all ways except for the presence of representation. Thus, any differences we observed might be due as much to the difference in the kind of claimants we were comparing as to representation.

So, if we are interested in figuring out the effect of actual use of representation, we need to find two sets identical in all ways except that one set was actually represented and the other was not. In that case, why not simply remove the Go-Getters from the treated (HLAB offer) group and compare the Regulars¹³⁹ in the treated group, who got representation, to the Regulars in the control (no HLAB offer) group, who did not? This is exactly what needs to be done, and we attempt to do so.¹⁴⁰ However, it is not easy to do. The reason for

¹³⁹. Comparing the Go-Getters in the treated group to the Go-Getters in the control group will not provide any information about the effect of representation because, by definition, all Go-Getters (regardless of whether they were in the treated or control group) were represented.

Notice that to address fully the problem of non-randomness in the actual use of representation we should also recognize that a third type of claimant existed, namely, the claimant who would not have obtained representation even if assigned an HLAB offer, a “Never-Lawyer” claimant. Recall that just under 10% of the treated (HLAB offer) group ended up unrepresented; we know that these claimants were Never-Lawyer claimants. We also know that the treated (HLAB offer) and control (no HLAB offer) groups were the same up to random variation. From this, we infer that the control group also had some Never-Lawyer claimants in it. Unfortunately, the small size of the dataset and the small number of claimants known to be Never-Lawyers made the task of picking out Never-Lawyers in the control group hopeless, so we ignored this issue in the analysis above.

¹⁴⁰. This discussion clarifies that because we are attempting to remove the Go-Getters from the group randomized to no HLAB offer, we are not simply comparing the results of the represented group to the unrepresented group, as some in the legal services community had requested that we do.

Note that some have suggested that the Go-Getter/Regular story articulated above may not accurately describe the way that some claimants in our control group obtained counsel. According to a different account, in an unknown number of cases, HLAB students called another service provider, Greater Boston Legal Services (GBLS), to provide the names and contact information of persons randomized to no HLAB offer to see if GBLS (or another organization called the Volunteer Lawyers Project) might be able to provide representation. GBLS called each such claimant to offer representation. The implication here is that in those cases in which the HLAB student called GBLS, and GBLS called the claimant, “Go-Getter” effort was not required to obtain representation. The apparent upshot of this implication is that a comparison of claimants who ended up represented to claimants who did not end up represented would in fact be worthwhile. *See* Bob Sable, *What Difference Representation—A*

the difficulty is that we do not know who was a Go-Getter and who was a Regular in the treated (HLAB offer) group. We *do* know who is who in the control (no HLAB offer) group: recall that we distinguished the Regulars from the Go-Getters in the control (no HLAB offer) group by observing which of the control group claimants obtained representation despite being randomized to no HLAB offer. If, in the control group, a claimant obtained an attorney despite *not* receiving an offer from HLAB, we *know* that claimant was a Go-Getter. If, in the control group, a claimant did not obtain an attorney (after not receiving an HLAB offer), we *know* the claimant was *not* a Go-Getter and was thus a Regular. But by definition, all claimants in the treated group received an offer of (HLAB) representation. Thus, our problem is that we do not know what each treated (HLAB offer) claimant *would* have done had she *not* been offered HLAB representation, which is what we need to know to distinguish Go-Getters from Regulars. We do not observe the counterfactual state of the world, i.e., what claimants actually offered representation would have done had they not been so offered. We have to use statistical methods to guess whether each treated (HLAB offer) group claimant is a Go-Getter or a Regular. To do this well, we need observed background variables that do a good job of predicting whether a person was a Go-Getter or a Regular. And it turns out that in our study, the variables listed in Table 1 do not do the job very well, probably because they are not very good proxies for characteristics such as energy, articulateness, and persistence. In fact, as we explain in further detail in Part II, at present it appears that no one in the quantitative legal community (including us) knows what variables to measure as good proxies for characteristics such as energy, articulateness, and persistence.¹⁴¹

We do our best here to fit statistical methods that predict whether treated group claimants were Go-Getters or Regulars, even though we are challenged by the fact that we have only 207 cases. Our best is good enough to allow us to

Response, CONCURRING OPINIONS (Mar. 28, 2011, 12:10 AM), <http://www.concurringopinions.com/archives/category/representation-symposium>.

The difficulty here is that the HLAB students apparently made calls to GBLS in some cases but not others, and we have no information on the kind of cases for which calls were made because there was no HLAB policy for when such calls would be made; individual students chose the cases on their own. On what basis were such choices made? It strikes us as at least plausible that HLAB students might have made these calls for claimants having the most sympathetic facts, and claimants with sympathetic facts might be more likely to win regardless of representation. Thus, even if the Go-Getter story we describe above does not apply in all cases, the HLAB-to-GBLS process described might have simply substituted a different kind of selection effect. For a more general discussion of such themes, see Part II.

¹⁴¹. See Rebecca L. Sandefur, Elements of Expertise: Lawyers' Impact on Civil Trial and Hearing Outcomes 10-11 (Dec. 3, 2009) (unpublished manuscript) (on file with authors).

conclude with some degree of confidence that actual use of legal representation caused a delay in the ALJ adjudication. The modeling required to attempt to separate the Go-Getters from Regulars means that we cannot use the easily visualized techniques (such as the Fisher test) that we presented earlier, so we simply state that our best efforts produced a 95% interval for the multiplier due to the actual use of representation of (1.5, 2.4), well above 1.¹⁴² In fact, these delay figures are larger than those for the delay due to the HLAB offer alone. Thus, the delay that actual use of representation caused is sufficiently clear that even our decidedly imperfect models find it.

In contrast, our best is not good enough with respect to win/loss. A 95% interval for the difference in win rates due to actual use of representation is (-.14, .23); this interval is so wide as to be useless. Our suspicion that actual use of representation produces no large increase in the win rate for the set of claimants who contact HLAB is thus based on non-quantitative sources of information, sources we discuss in Question 8, below.

Question 7: Were there subsets—identifiable at intake—of cases where an offer of HLAB assistance did increase the probability of a win?

In other words, the data provide no evidence that an offer of HLAB representation increased the probability of a win for the set of claimants who sought HLAB assistance. But if there were easily identifiable subsets of claimants for whom the offer made a favorable result more likely, HLAB might as an experimental matter shift its practice to these subsets. Seeking such subsets was a primary reason for HLAB's decision to engage in this study in the first place.¹⁴³ Do such subsets exist?

Answering this question requires a departure from good statistical practice. It requires that we go fishing for subsets in which one sees good results. Even if the offer of HLAB representation had no effect on any subset of cases, if we examined enough subsets we might find some spurious statistically significant

¹⁴². See *supra* note 135. In the analysis presented above, we assumed that all the claimants who withdrew were unrepresented. If we assume that all claimants who withdrew were represented (a proposition we find unlikely), the 95% posterior interval for the time multiplier is (1.1, 1.8). Thus, we feel comfortable in concluding that there is a delay due to actual use of representation. The corresponding interval for the difference in win rate under the assumption that all claimants who withdrew were represented was (-.4, .1); again, hardly useful.

¹⁴³. See E-mail from Rachel Lauter, President, Harvard Legal Aid Bureau, to Harvard Law Sch. Community, reproduced in Dave Hoffman, *Harvard Clinic Responds to Greiner Study*, CONCURRING OPINIONS (Jan. 24, 2011, 2:06 PM), <http://www.concurringopinions.com/archives/2011/01/harvard-clinic-responds-to-greiner-study.html>.

increases in the probability of claimant success due to nothing more than random variation. Thus, the conclusions here are suggested avenues for evaluation of a future HLAB program, *not* a reason to believe that HLAB offers made a positive difference in any identified subsets. At issue here are potential ideas, not evidence.

With this in mind, we divided the data in over a dozen ways and analyzed each subset using the “Sampling, Ratio Estimator” and “Separate Regression” techniques identified in Figures 2, 4, and 6. At first, we examined the results hoping to find subsets of the data showing a statistically significant positive difference under both techniques in the probability that a claimant prevailed. We found none. In fact, for the Separate Regression technique, the method in which we would tend to place slightly more trust (given the small sizes of some of the datasets and the concomitant imbalances in background variables), none of the subsets showed a statistically significant advantage due to the HLAB offer. With that in mind, we resorted to looking at only the results from the “Sampling, Ratio Estimator” technique. Here, we found something somewhat close to significance in the following subsets of claimants: women;¹⁴⁴ claimants in “discharge” cases;¹⁴⁵ and non-Hispanic blacks.

Note that one variable we thought (hoped) might be related to ability to proceed effectively without legal representation, education, turned out to have little if any predictive value. As we discuss in Part II, the failure of this variable is a bad omen for those hoping to produce credible results regarding the effect of representation from observational studies.¹⁴⁶

To reiterate, nothing in these results supports the conclusion that an HLAB offer actually did benefit these subsets of claimants. Nevertheless, we reported these results to HLAB for it to consider in deciding whether to continue with its unemployment practice, and if so, what form that practice should take.

Question 8: What could explain our results?

One of the results in our study, the delay due to both an offer of and actual use of representation, is not difficult to explain. Attorneys, including student-attorneys, need time to investigate facts, to do research, and to otherwise prepare. The time frame for the first-level appeals process is tight, with

¹⁴⁴. *But see Stropnick v. Nathanson*, 19 M.D.L.R. 39 (Mass. Comm'n Against Discrimination Feb. 25, 1997) (holding that an attorney's practice of representing only female clients in divorce proceedings constituted unlawful sex discrimination).

¹⁴⁵. *See supra* note 45 and accompanying text (describing when claimants in “discharge” cases may be awarded unemployment benefits).

¹⁴⁶. *See infra* notes 236-237 and accompanying text.

hearings typically scheduled within a few weeks of the filing for an appeal by either the claimant or the employee. During that time, the claimant must find counsel, sign a retainer agreement, and provide relevant information before the attorney's work can seriously begin. Hearing postponements at the behest of legal representatives are thus to be expected.

The greater challenge is explaining the absence of evidence that an HLAB offer affected the probability that a claimant would prevail. At the outset, we acknowledge, but dismiss as implausible, two possible explanations. The first is the hypothesis that the absence of an effect is due to low quality HLAB lawyering. Our refusal to credit this hypothesis is due to our personal observations of the extraordinarily talented set of students HLAB recruits,¹⁴⁷ the students' dedication, the tremendous oversight and guidance the clinical staff at HLAB provides, and what we understand to be the strong reputation HLAB enjoys for its work in this area.¹⁴⁸

The second hypothesis we do not credit is the possibility that the control group had too many claimants who ended up finding representation, so that there was an insufficient contrast between the treated (HLAB offer) and control (no HLAB offer) group. The difference between the 90%+

¹⁴⁷. See *supra* note 82 and accompanying text.

¹⁴⁸. "The Harvard kids know their cases." Interview with Anne Will, Review Examiner, Mass. Div. of Unemp't Assistance, in Cambridge, Mass. (Apr. 14, 2010).

After observing several first-level appeals in another state, Kritzer concluded that "[b]ecause of the low fees involved, few attorneys handle [unemployment] cases, and . . . that many, perhaps most, of the lawyers are not familiar with the administrative law judges, or the rules governing eligibility for [unemployment] compensation." KRITZER, *supra* note 63, at 55. With respect to law students, Kritzer concluded that those he observed, all first-years participating in a volunteer organization with no selectivity, sometimes struggled with witness examination technique. *Id.* at 63-66.

As discussed in Section I.C, HLAB student-attorneys are 2Ls and 3Ls chosen from a highly selective process. They gained experience advocating on behalf of their clients in a variety of settings. During our study, HLAB student-attorneys were trained in the procedures applicable in Massachusetts, and they investigated the law and facts in their cases fully. We concede that HLAB student-advocates may have lacked the craftsmanship and knowledge of an attorney who has practiced for many years before ALJs in first-level unemployment appeals. But our sense based on studying this system for the past several years is that there were very, very few such attorneys (for the same reason that Kritzer identified). Thus, by saying that HLAB representation was strong and well respected, and therefore that it is implausible to attribute our results to low quality HLAB lawyering, we do not mean to suggest that an überlawyer specializing in first-level appeals might not have produced better outcomes. We do suggest, however, that a comparison involving a hypothetical überlawyer specializing in first-level appeals is of limited value, and that HLAB student representation was likely among the best that claimants in Massachusetts first-level appeals could expect to receive.

representation rate in the treated (HLAB offer) group and the around 39% representation rate in the control (no HLAB offer) group was sufficient to allow the techniques we used to show a statistically significant delay effect. Thus, the contrast was sufficient to make evident large effects, and as noted above, the crossover between treated and control groups in our study was similar to that experienced in other randomized studies of legal representation. No large effect on win/loss was, apparently, present. Thus, even if one believes, in some ill-defined way, that the high rate of representation in the control group “masked” the contrast between the treated and control groups, we suggest that any “true” increase in the success rate due to the HLAB program was unlikely to be large, because if it had been large, we probably would have seen it.

Among the many possible explanations for the result that an HLAB offer made no statistically significant difference in the success rate, we find three most plausible (although there certainly are others). The first begins with the observation that the win rate for the control group in our study (those not offered HLAB representation) was higher than the average success rate in Massachusetts first-level appeals as a whole. In our dataset’s control group, when the claimant appealed, she prevailed 65% of the time; in the overall Massachusetts system, the win rate for appealing claimants was 47%. In our dataset’s control group, when the claimant defended against an employer appeal, she prevailed 83% of the time, compared to a 75% rate in Massachusetts overall.¹⁴⁹ The comparison is not perfect for two reasons: first, HLAB accepts cases heard only in the Massachusetts DUA’s Boston office, whereas the comparison numbers quoted above are statewide; second, HLAB accepted only cases where eligibility for benefits was at issue, whereas the first-level appeals system as a whole includes cases presenting other issues (such as overpayments). Nevertheless, the disparity in the win rates, particularly for claimant-initiated appeals, is striking. This suggests to us either that the claimants who initiated contact with HLAB possessed personal characteristics making them more likely to win their cases or that the underlying facts in their cases were unusually strong. Either is possible; the former strikes us as particularly plausible given that the HLAB unemployment intake system depended on a telephone contact initiated by the claimant. In other words, perhaps HLAB and the other attorneys who assisted clients who had contacted HLAB were helping only those who did not need the help.

¹⁴⁹. See *supra* note 56 and accompanying text.

A second plausible explanation for our results lies in the “tempestuous[] . . . marriage”¹⁵⁰ of adversarial and inquisitorial styles of judging used to adjudicate first-level appeals. As discussed above, in this system, the claims adjuster shouldered the initial burden of collecting relevant documents. The ALJ initiated questioning of all witnesses. Perhaps the ALJ made a special effort (bent over backwards) for pro se litigants. Thus, our second proposed explanation is that the Massachusetts DUA had used and trained its staff effectively to create a system that was accessible to pro se litigants.¹⁵¹

A third and related plausible explanation might lie in the nature of the issues disputed in first-level appeals. Commentators are currently split on whether the issues ordinarily involved in first-level appeals are relatively simple or complex,¹⁵² but it appears to us that at least with respect to the cases HLAB handled, expert testimony was rarely needed, and ordinarily the cases could be adequately worked up over the course of a few (perhaps intense) weeks. If so, then again, perhaps the system was accessible to pro se litigants because the issues were accessible.

We return to these themes, particularly the first two, in Part II. For now, however, we note that without further study with different service providers and, perhaps, different study designs, we do not know which of these mechanisms (or something else we have not mentioned) is rendering an offer of representation unlikely to have resulted in a large effect on win/loss. And the three alternative explanations might apply to a lesser or greater extent in other legal services settings. A great many providers depend on client-initiated telephone systems for intake; some of these telephone lines are clogged, requiring client persistence. Even in the United States, other adjudicatory systems have in the past shared some salient characteristics of the ALJ process described above, such as a quasi-inquisitorial style of witness examination and evidence gathering.¹⁵³

^{150.} Mashaw, *supra* note 57, at 16.

^{151.} Cf. Popkin, *supra* note 86, at 451 (“[H]earing officers whose responsibility in nonadvers[arial] proceedings is to protect the unrepresented claimant are skeptical about a representative’s ability to help claimants win.”).

^{152.} Compare, e.g., Morris, *supra* note 77, at 665-66 (noting “intricate State laws” and “complexity”), and Emsellem & Halas, *supra* note 75, at 297-98 (“legally complex,” “trial-like setting”), with Nofi-Bendici, *supra* note 77, at 500 (“relatively straightforward”).

^{153.} See, e.g., *Richardson v. Perales*, 402 U.S. 389, 410 (1971) (noting that a Social Security ALJ acts as an examiner charged with developing the facts); Suzanne E. Elwell with Christopher D. Carlson, *The Iowa Small Claims Court: An Empirical Analysis*, 75 IOWA L. REV. 433, 463 (1996) (noting that in Iowa small claims court, the court “swears in the parties and their witnesses and examines them . . .”). See generally Jerry L. Mashaw, *The Management Side of Due Process: Some Theoretical and Litigation Notes on the Assurance of Accuracy Fairness and*

Our speculation is that the first of our explanations, that HLAB's intake system is allowing it to reach mostly those whose outcomes it cannot greatly affect, is playing a substantial role. If we are right about that (and we cannot know without further study), then a major focus of future research should be provider outreach and intake systems. Again, it appears to us that some service providers use intake systems that could screen out potential clients whose outcomes are most subject to change via representation.

It would be a mistake to overgeneralize the results of our study to conclude that offering free legal assistance is not worth the cost or time, or even that offers of representation make no difference in Massachusetts first-level appeals. If the first of our three possible explanations is playing an important role, then an altered intake system might result in a different set of potential clients. Meanwhile, a major access to justice problem could still exist in Massachusetts first-level appeals. On this point, we caution that studies of the kind we conduct here are dependent on context. One of our primary purposes in this paper is to persuade readers that we need numerous additional studies in a variety of different contexts. Hopefully, with numerous studies, we could begin to make more informed guesses as to which aspects of context matter most, and to base policy and funding decisions on these more informed guesses. This agenda recognizes, indeed it depends on, the idea that context matters.

It would also be a mistake to undergeneralize the results of our study by concluding that it has no implications outside of HLAB's first-level appeals practice. If the second or third (or some combination thereof) of our explanations is playing an important role, then our results might implicate a variety of legal assistance practices in a variety of different subject areas.

Perhaps most importantly, our study provides a sharp reason to reconsider other efforts to measure the effect of legal representation, as we do in Part II. We further suggest that the HLAB study provides reason to implement a large program of randomized control trials in the legal setting; we begin the discussion of what such a program should look like in Part III.

II. THE CURRENT STATE OF OUR KNOWLEDGE

The literature purporting to measure quantitatively the effect of legal representation in civil disputes is substantial.¹⁵⁴ In terms of subject area, studies

Timeliness in the Adjudication of Social Welfare Claims, 59 CORNELL L. REV. 772 (1974) (examining procedures used to adjudicate Social Security, Aid to Families with Dependent Children, and other claims).

¹⁵⁴. SUBCOMM. ON SOC. SEC. OF THE H. COMM. ON WAYS AND MEANS, 94TH CONG., RECENT STUDIES RELEVANT TO THE DISABILITY HEARINGS AND APPEALS CRISIS 1-124 (Comm. Print

1975) [hereinafter SUBCOMM. ON SOC. SEC.]; U.S. GEN. ACCOUNTING OFFICE, GAO/HRD-90-22BR, SPECIAL EDUCATION: THE ATTORNEY FEES PROVISION OF PUBLIC LAW 99-372, at 5 (1989) [hereinafter GAO, ATTORNEY FEES]; BD. OF IMMIGRATION APPEALS, THE BIA PRO BONO PROJECT IS SUCCESSFUL (Oct. 2004), *available at* <http://www.justice.gov/eoir/reports/BIAProBonoProjectEvaluation.pdf>; LISA PARSONS CHADHA, LAWYERS' COMM. FOR BETTER HOUSING, INC., TIME TO MOVE: THE DENIAL OF TENANTS' RIGHTS IN CHICAGO EVICTION COURT 35-36 (1996) (on file with authors); COURT STUDY GRP. OF THE JUNIOR LEAGUE OF BROOKLYN, REPORT ON A STUDY OF BROOKLYN LANDLORD AND TENANT COURT (1973) (on file with authors) [hereinafter COURT STUDY GRP.]; JAMES K. HAMMITT, RAND INST. FOR CIVIL JUSTICE, 2 AUTOMOBILE ACCIDENT COMPENSATION: PAYMENTS BY AUTO INSURERS (1985), *available at* <http://www.rand.org/pubs/reports/2006/R3051.pdf>; DEBORAH R. HENSLER ET AL., COMPENSATION FOR ACCIDENTAL INJURIES (1991); KRITZER, *supra* note 63; EDWIN M. LEMERT, SOCIAL ACTION AND LEGAL CHANGE 192-94 (1970) [hereinafter LEMERT, SOCIAL ACTION]; RAPHAEL L. PODOLSKY & STEPHEN O'BRIEN, LEGAL ASSISTANCE RES. CTR. OF CONN., INC., A STUDY OF EVICTION CASES IN HARTFORD 28-32 (1995); JOHN E. ROLPH ET AL., RAND INST. FOR CIVIL JUSTICE, 1 AUTOMOBILE ACCIDENT COMPENSATION: WHO PAYS HOW MUCH HOW SOON? (1985), *available at* <http://www.rand.org/pubs/reports/2007/R3050.pdf>; H. LAURENCE ROSS, SETTLED OUT OF COURT: THE SOCIAL PROCESS OF INSURANCE CLAIMS ADJUSTMENTS (1970); STAPLETON & TEITELBAUM, *supra* note 20; WOMEN'S LAW CTR. OF MD., INC., FAMILIES IN TRANSITION: A FOLLOW-UP STUDY EXPLORING FAMILY LAW ISSUES IN MARYLAND (2006) [hereinafter WOMEN'S LAW CTR.], *available at* <http://www.wlcmd.org/pdf/publications/FamiliesInTransition.pdf>; David P. Aday, Jr., *Court Structure, Defense Attorney Use, and Juvenile Court Decisions*, 27 SOC. Q. 107 (1986); Richard N. Block & Jack Stieber, *The Impact of Attorneys and Arbitrators on Arbitration Awards*, 40 INDUS. & LAB. REL. REV. 543, 546 n.5 (1987); George W. Burruss, Jr. & Kimberly Kempf-Leonard, *The Questionable Advantage of Defense Counsel in Juvenile Court*, 19 JUST. Q. 37 (2002); Ralph C. Cavanagh & Deborah L. Rhode, Project, *The Unauthorized Practice of Law and Pro Se Divorce: An Empirical Analysis*, 86 YALE L.J. 104 (1976); Stevens H. Clarke & Gary G. Koch, *Juvenile Court: Therapy or Crime Control, and Do Lawyers Make a Difference?*, 14 LAW & SOC'Y REV. 263 (1980); Alexander J.S. Colvin, *Empirical Research on Employment Arbitration: Clarity Amidst the Sound and Fury?*, 11 EMP. RTS. & EMP. POL'Y J. 405 (2007); Laura Cooper, *Goldberg's Forgotten Footnote: Is There a Due Process Right to a Hearing Prior to the Termination of Welfare Benefits When the Only Issue Raised Is a Question of Law?*, 64 MINN. L. REV. 1107 (1980); Stephen M. Crow & James W. Logan, *Arbitrators' Characteristics and Decision-Making Records, Gender of Arbitrators and Grievants, and the Presence of Legal Counsel as Predictors of Arbitral Outcomes*, 7 EMP. RESPONSIBILITIES & RTS. J. 169, 178 (1994); Clarence R. Deitsch & David A. Dilts, *Case Characteristics Affecting the Method of Grievance Dispute Settlement*, 1 EMP. RESPONSIBILITIES & RTS. J. 113 (1988); David Duffee & Larry Siegel, *The Organization Man: Legal Counsel in the Juvenile Court*, 7 CRIM. L. BULL. 544 (1971); David L. Eldridge, *The Construction of a Courtroom: The Judicial System and Autopoiesis*, 38 J. APPLIED BEHAV. SCI. 298 (2002); Jane W. Ellis, *Plans, Protections, and Professional Intervention: Innovations in Divorce Custody Reform and the Role of Legal Professionals*, 24 U. MICH. J.L. REFORM 65 (1990); Elwell with Carlson, *supra* note 153; Barry C. Feld, *The Right to Counsel in Juvenile Court: An Empirical Study of When Lawyers Appear and the Difference They Make*, 79 J. CRIM. L. & CRIMINOLOGY 1185 (1989); Barry C. Feld & Shelly Schaefer, *The Right to Counsel in Juvenile Court: The Conundrum of Attorneys as an Aggravating Factor at Disposition*, 27 JUST. Q. 713 (2010); Elyce Zenoff Ferster & Thomas F. Courtless, *Pre-Dispositional Data, Role of Counsel, and Decisions in a Juvenile Court*, 7 LAW & SOC'Y REV. 195 (1972); Elyce Zenoff Ferster,

Thomas F. Courtless & Edith Nash Snethen, *The Juvenile Justice System: In Search of the Role of Counsel*, 39 FORDHAM L. REV. 375 (1971); Anthony J. Fusco, Jr., Nancy B. Collins & Julian R. Birnbaum, *Chicago's Eviction Court: A Tenant's Court of No Resort*, 17 URB. L. ANN. 93 (1979); Lori Guevara, Cassia Spohn & Denise Herz, *Race, Legal Representation, and Juvenile Justice: Issues and Concerns*, 50 CRIME & DELINQ. 344 (2004); Steven Gunn, *Eviction Defense for Poor Tenants: Costly Compassion or Justice Served?*, 13 YALE L. & POL'Y REV. 385 (1995); Jan L. Hagen, *Justice for the Welfare Recipient: Another Look at Welfare Fair Hearings*, 73 SOC. SERV. REV. 177 (1983); Joel F. Handler, *Justice for the Welfare Recipient: Fair Hearings in AFDC – The Wisconsin Experience*, 43 SOC. SERV. REV. 12 (1969); Paula Hannaford-Agor & Nicole Mott, *Research on Self-Represented Litigation: Preliminary Results and Methodological Considerations*, 24 JUST. SYS. J. 163 (2003); David W. Hayeslip, Jr., *The Impact of Defense Attorney Presence on Juvenile Court Dispositions*, 30 JUV. & FAM. CT. J. 9 (1979); Elizabeth Hill, *Due Process at Low Cost: An Empirical Study of Employment Arbitration Under the Auspices of the American Arbitration Association*, 18 OHIO ST. J. ON DISP. RESOL. 777 (2003); Robert Kelly & Sarah Ramsey, *Do Attorneys for Children in Protection Proceedings Make a Difference? A Study of the Impact of Representation Under Conditions of High Judicial Intervention*, 21 J. FAM. L. 405 (1982) [hereinafter Kelly & Ramsey, *Do Attorneys for Children Make a Difference*]; Robert F. Kelly & Sarah H. Ramsey, *Legal Representation for Parents and Children in Child Protection Proceedings: Two Empirical Models of Acquisitorial Processes and a Proposal for Reform*, 89 DICK. L. REV. 605 (1985) [hereinafter Kelly & Ramsey, *Legal Representation in Child Protection Proceedings*]; Donald Kerwin, *Charitable Legal Programs for Immigrants: What They Do, Why They Matter and How They Can Be Expanded*, 4 IMMIGR. BRIEFINGS 1 (2004); David Kirp, William Buss & Peter Kuriloff, *Legal Reform of Special Education: Empirical Studies and Procedural Proposals*, 62 CALIF. L. REV. 40 (1974); Charles H. Kuck, *Legal Assistance for Asylum Seekers in Expedited Removal: A Survey of Alternative Practices*, in 2 U.S. COMM'N ON INT'L RELIGIOUS FREEDOM, REPORT ON ASYLUM SEEKERS IN EXPEDITED REMOVAL: EXPERT REPORTS 232 (2004), http://www.uscirf.gov/images/stories/pdf/asylum_seekers/ERS_RptVolIII.pdf; Gary LaFree & Christine Rack, *The Effects of Participants' Ethnicity and Gender on Monetary Outcomes in Mediated and Adjudicated Civil Cases*, 30 LAW & SOC'Y REV. 767 (1996); Leandra Lederman & Warren B. Hrung, *Do Attorneys Do Their Clients Justice? An Empirical Study of Lawyers' Effects on Tax Court Litigation Outcomes*, 41 WAKE FOREST L. REV. 1235 (2006); Edwin M. Lemert, *Legislating Change in the Juvenile Court*, 1967 WIS. L. REV. 421 [hereinafter Lemert, *Legislating Change*]; Angela Littwin, *The Affordability Paradox: How Consumer Bankruptcy's Greatest Weakness May Account for Its Success*, 52 WM. & MARY L. REV. 1933 tbls.3b-6b (2011); E. Patrick McDermott & Ruth Obar, *"What's Going On" in Mediation: An Empirical Analysis of the Influence of a Mediator's Style on Party Satisfaction and Monetary Benefit*, 9 HARV. NEGOT. L. REV. 75 (2004); Judith G. McMullen & Debra Oswald, *Why Do We Need a Lawyer?: An Empirical Study of Divorce Cases*, 12 J.L. & FAM. STUD. 57 (2010); Monsma & Lempert, *supra* note 21; Marilyn Miller Mosier & Richard A. Soble, *Modern Legislation, Metropolitan Court, Min[u]scule Results: A Study of Detroit's Landlord-Tenant Court*, 7 U. MICH. J.L. REFORM 8 (1973); Jane C. Murphy, *Engaging with the State: The Growing Reliance on Lawyers and Judges To Protect Battered Women*, 11 AM. U. J. GENDER SOC. POL'Y & L. 499 (2003); Rafael I. Pardo, *An Empirical Examination of Access to Chapter 7 Relief by Pro Se Debtors*, 26 EMORY BANKR. DEV. J. 5 (2009); William D. Popkin, *The Effect of Representation in Nonadversary Proceedings – A Study of Three Disability Programs*, 62 CORNELL L. REV. 989 (1977); Jaya Ramji-Nogales, Andrew I. Schoenholtz & Philip G. Schrag, *Refugee Roulette: Disparities in Asylum Adjudication*, 60 STAN. L. REV. 295 (2007) (condensed version with regression results available at <http://www.law.georgetown.edu/humanrightsinstitute/refugeeroulette.htm>); Charles E. Reasons, *Gault: Procedural Change and Substantive Effect*, 16

CRIME & DELINQ. 163 (1970); Jonathan I. Rose & Martin A. Scott, "Street Talk" Summons in Detroit's Landlord-Tenant Court: A Small Step Forward for Urban Tenants, 52 J. URB. L. 967, 999-1002 (1975); Bruce D. Sales, Connie J. Beck & Richard K. Haan, *Is Self-Representation a Reasonable Alternative to Attorney Representation in Divorce Cases?*, 37 ST. LOUIS U. L.J. 553 (1992); Austin Sarat, *Alternatives in Dispute Processing: Litigation in a Small Claims Court*, 10 LAW & SOC'Y REV. 339 (1976); Andrew I. Schoenholtz & Jonathan Jacobs, *The State of Asylum Representation: Ideas for Change*, 16 GEO. IMMIGR. L.J. 739 (2001); Seron et al., *supra* note 20; John Montague Steadman & Richard S. Rosenstein, "Small Claims" Consumer Plaintiffs in the Philadelphia Municipal Court: An Empirical Study, 121 U. PA. L. REV. 1309 (1973); Arup Varma & Lamont E. Stallworth, *Participants' Satisfaction with EEO Mediation and the Issue of Legal Representation: An Empirical Inquiry*, 6 EMP. RTS. & EMP. POL'Y J. 387 (2002); William C. Whitford, *The Small-Case Procedure of the United States Tax Court: A Small Claims Court that Works*, 1984 AM. B. FUND. RES. J. 797; Roselle L. Wissler, *Representation in Mediation: What We Know from Empirical Research*, 37 FORDHAM URB. L.J. 419 (2010); Barbara Yngvesson & Patricia Hennessey, *Small Claims, Complex Disputes: A Review of the Small Claims Literature*, 9 LAW & SOC'Y REV. 219 (1975); Perry A. Zirkel, *A Profile of Grievance Arbitration Cases*, 38 ARB. J. 35 (1983); Perry A. Zirkel & Philip H. Breslin, *Correlates of Grievance Arbitration Awards*, 24 J. COLLECTIVE NEGOTIATIONS 45 (1995); Ronald P. Hammer & Joseph M. Hartley, Note, *Procedural Due Process and the Welfare Recipient: A Statistical Study of AFDC Fair Hearings in Wisconsin*, 1978 WIS. L. REV. 145; Note, *Legal Services and Landlord-Tenant Litigation: A Critical Analysis*, 82 YALE L.J. 1495 (1973) [hereinafter Note, *Legal Services and Landlord-Tenant Litigation*]; Note, *Representation in Child-Neglect Cases: Are Parents Neglected?*, 4 COLUM. J.L. & SOC. PROBS. 230 (1968) [hereinafter Note, *Representation in Child-Neglect Cases*]; Owens, *supra* note 75; Melanie Archer, Access and Equity in the Due Process System: Attorney Representation and Hearing Outcomes in Illinois, 1997-2002, at 7-9 (Dec. 2002) (unpublished manuscript), <http://www.dueprocessillinois.org/Access.pdf> (last visited Apr. 3, 2012); Christopher R. Drahoszal & Samantha Zyontz, An Empirical Study of AAA Consumer Arbitrations, 51-55 (Mar. 13, 2009) (unpublished manuscript), <http://ssrn.com/abstract=1365435>; Russell Engler & Craig S. Bloomgarden, Summary Process Actions in the Boston Housing Court: An Empirical Study and Recommendations for Reform (May 20, 1983) (unpublished manuscript), <http://ssrn.com/abstract=969782>; Brook Gotberg, The Inner Workings of Unemployment Hearings in Massachusetts: A Statistical Approach (2008) (unpublished manuscript) (on file with authors); Phillip Granberry & Randy Albelda, Assessing the Benefits of Provision of Legal Services Through the Disability Benefits Project 3 (Aug. 2006) (unpublished manuscript), <http://www.nlada.org/DMS/Documents/1236007115.86/Disability%20benefit%20report%207-07.doc>; Rebecca Hall, Eviction Prevention as Homelessness Prevention: The Need for Access to Legal Representation for Low Income Tenants (May 1, 1991) (unpublished manuscript) (on file with authors); Richard J. Maiman, Private Attorneys in Juvenile Court: The Impact of *In re Gault* 9 (Aug. 29, 1974) (unpublished manuscript) (on file with authors); Neil Steiner, An Analysis of the Effectiveness of a Limited Assistance Outreach Project to Low-Income Tenants Facing Eviction 24-25 (Oct. 14, 1997) (unpublished manuscript) (on file with authors); see also Liz Elwart et al., *Increasing Access to Restraining Orders for Low-Income Victims of Domestic Violence: A Cost-Benefit Analysis of the Proposed Domestic Abuse Grant Program*, STATE BAR ASS'N OF WIS. 2 (Dec. 2006), available at <http://www.nlada.org/DMS/Documents/1176146724.92/WI%20appendix8.pdf> (asserting a fourteen percentage point increase due to counsel in request for restraining orders). For a discussion of the literature on mediation, see Lisa B. Bingham, Kiwhan Kim & Susan Summers Raines, *Exploring the Role of Representation in Employment Mediation at the*

have focused on automobile insurance claims,¹⁵⁵ bankruptcy,¹⁵⁶ disability (SSI/SSDI,¹⁵⁷ FECA,¹⁵⁸ and veterans claims¹⁵⁹), educational programs for disabled children,¹⁶⁰ employment (generally¹⁶¹ as well as focusing specifically on discharge/discipline¹⁶² and discrimination¹⁶³), family law (child neglect,¹⁶⁴ custody,¹⁶⁵ divorce,¹⁶⁶ and restraining orders¹⁶⁷), housing/eviction,¹⁶⁸ immigration disputes of all types,¹⁶⁹ juvenile delinquency,¹⁷⁰ small claims,¹⁷¹ special

USPS, 17 OHIO ST. J. ON DISP. RESOL. 341 (2001). This article also has data on settlement rates broken down by representation status.

Note that the focus of LaFree & Rack, *supra*, was not specifically on the effect of representation but rather on the effect of litigant immutable characteristics; as such, this study may be immune from some of the criticisms we articulate here but may be vulnerable to others. See D. James Greiner & Donald B. Rubin, *Causal Effects of Perceived Immutable Characteristics*, 93 REV. ECON. & STAT. 775 (2011).

¹⁵⁵ See, e.g., HAMMITT, *supra* note 154; HENSLER ET AL., *supra* note 154; ROLPH ET AL., *supra* note 154; ROSS, *supra* note 154.

¹⁵⁶ See, e.g., Littwinn, *supra* note 154; Pardo, *supra* note 154.

¹⁵⁷ See, e.g., SUBCOMM. ON SOC. SEC., *supra* note 154; KRITZER, *supra* note 63, at 111-50; Popkin, *supra* note 154; Granberry & Albelda, *supra* note 154.

¹⁵⁸ See, e.g., Popkin, *supra* note 154.

¹⁵⁹ See, e.g., *id.*

¹⁶⁰ See, e.g., GAO, ATTORNEY FEES, *supra* note 154; Archer, *supra* note 154.

¹⁶¹ See, e.g., Crow & Logan, *supra* note 154; Deitsch & Dilts, *supra* note 154; Hill, *supra* note 154; McDermott & Obar, *supra* note 154; Zirkel, *supra* note 154; Zirkel & Breslin, *supra* note 154.

¹⁶² See, e.g., KRITZER, *supra* note 63, ch. 2; Block & Stieber, *supra* note 154; Colvin, *supra* note 154.

¹⁶³ See, e.g., Varma & Stallworth, *supra* note 154.

¹⁶⁴ See, e.g., Kelly & Ramsey, *Do Attorneys for Children Make a Difference*, *supra* note 154; Kelly & Ramsey, *Legal Representation in Child Protection Proceedings*, *supra* note 154; Note, *Representation in Child-Neglect Cases*, *supra* note 154.

¹⁶⁵ See, e.g., WOMEN'S LAW CTR., *supra* note 154.

¹⁶⁶ See, e.g., Cavanagh & Rhode, *supra* note 154; Ellis, *supra* note 154; McMullen & Oswald, *supra* note 154; Sales et al., *supra* note 154; Wissler, *supra* note 154.

¹⁶⁷ See, e.g., Murphy, *supra* note 154; Elwart et al., *supra* note 154.

¹⁶⁸ See, e.g., COURT STUDY GRP., *supra* note 154; Eldridge, *supra* note 154; Fusco et al., *supra* note 154; Gunn, *supra* note 154; Hannaford-Agor & Mott, *supra* note 154; Monsma & Lempert, *supra* note 21; Mosier & Soble, *supra* note 154; Rose & Scott, *supra* note 154; Seron et al., *supra* note 20; Note, *Legal Services and Landlord-Tenant Litigation*, *supra* note 154; Engler & Bloomgarden, *supra* note 154; Hall, *supra* note 154; Steiner, *supra* note 154.

¹⁶⁹ See, e.g., BD. OF IMMIGRATION APPEALS, *supra* note 154; Kerwin, *supra* note 154; Kuck, *supra* note 154; Ramji-Nogales et al., *supra* note 154; Schoenholtz & Jacobs, *supra* note 154.

¹⁷⁰ See, e.g., LEMERT, SOCIAL ACTION, *supra* note 154; STAPLETON & TEITELBAUM, *supra* note 20; Aday, *supra* note 154; Burruss & Kempf-Leonard, *supra* note 154; Clarke & Koch, *supra* note 154; Duffee & Siegel, *supra* note 154; Feld, *supra* note 154; Feld & Schaefer, *supra* note 154;

education,¹⁷² federal tax (both small claims¹⁷³ and general¹⁷⁴), state tax,¹⁷⁵ unemployment,¹⁷⁶ and welfare.¹⁷⁷ The type of proceeding involved has varied from uncontested,¹⁷⁸ to claims adjustment,¹⁷⁹ to mediation,¹⁸⁰ to arbitration,¹⁸¹ to various types of administrative adjudications,¹⁸² to court proceedings (including specialized courts of limited jurisdiction).¹⁸³ When an external decisionmaking structure is in place, the adjudicatory style varies from less¹⁸⁴ to fully¹⁸⁵ adversarial, with several hybrids, including what once purported (and

Ferster & Courtless, *supra* note 154; Ferster et al., *supra* note 154; Guevara et al., *supra* note 154; Hayeslip, *supra* note 154; Lemert, *Legislating Change*, *supra* note 154; Reasons, *supra* note 154; Maiman, *supra* note 154.

^{171.} See, e.g., Elwell with Carlson, *supra* note 153; Hannaford-Agor & Mott, *supra* note 154; LaFree & Rack, *supra* note 154; Sarat, *supra* note 154; Steadman & Rosenstein, *supra* note 154; Yngvesson & Hennessey, *supra* note 154.

^{172.} See, e.g., Kirp et al., *supra* note 154.

^{173.} See, e.g., Whitford, *supra* note 154.

^{174.} See, e.g., Lederman & Hrung, *supra* note 154.

^{175.} See, e.g., KRITZER, *supra* note 63.

^{176.} See, e.g., *id.*; Owens, *supra* note 75.

^{177.} See, e.g., Cooper, *supra* note 154; Hagen, *supra* note 154; Handler, *supra* note 154; Hammer & Hartley, *supra* note 154.

^{178.} See, e.g., Cavanagh & Rhode, *supra* note 154.

^{179.} See, e.g., ROSS, *supra* note 154.

^{180.} See, e.g., McDermott & Obar, *supra* note 154; Varma & Stallworth, *supra* note 154; Wissler, *supra* note 154.

^{181.} See, e.g., KRITZER, *supra* note 63, at 151-92; Block & Steiber, *supra* note 154; Colvin, *supra* note 154; Crow & Logan, *supra* note 154; Deitsch & Dilts, *supra* note 154, at 53; Hill, *supra* note 154; Zirkel, *supra* note 154; Zirkel & Breslin, *supra* note 154; Drahozal & Zyontz, *supra* note 154.

^{182.} See, e.g., KRITZER, *supra* note 63, at 23-78, 111-50; Cooper, *supra* note 154; Hagen, *supra* note 154; Handler, *supra* note 154; Kirp et al., *supra* note 154; Kuck, *supra* note 154; Schoenholtz & Jacobs, *supra* note 154; Hammer & Hartley, *supra* note 154; Owens, *supra* note 75; Granberry & Albelda, *supra* note 154.

^{183.} See, e.g., COURT STUDY GRP., *supra* note 154; Aday, *supra* note 154; Clarke & Koch, *supra* note 154; Duffee & Siegel, *supra* note 154; Eldridge, *supra* note 154; Elwell with Carlson, *supra* note 153; Feld, *supra* note 154; Ferster & Courtless, *supra* note 154; Ferster et al., *supra* note 154; Fusco et al., *supra* note 154; Gunn, *supra* note 154; Hayeslip, *supra* note 154; Lederman & Hrung, *supra* note 154; Lemert, *Legislating Change*, *supra* note 154; Mosier & Soble, *supra* note 154; Pardo, *supra* note 154; Reasons, *supra* note 154; Sarat, *supra* note 154; Steadman & Rosenstein, *supra* note 154; Whitford, *supra* note 154; Yngvesson & Hennessey, *supra* note 154; Note, *Legal Services and Landlord-Tenant Litigation*, *supra* note 154; Note, *Representation in Child-Neglect Cases*, *supra* note 154; Hall, *supra* note 154.

^{184.} See, e.g., Popkin, *supra* note 154.

^{185.} See, e.g., Note, *Representation in Child-Neglect Cases*, *supra* note 154.

may still purport) to be a *parens patriae* approach in juvenile delinquency proceedings¹⁸⁶ as well as the particular focus of this paper, the “tempestuous marriage” in first-level appeals.¹⁸⁷ In terms of type of representative, most studies have focused on attorneys, but others have examined paralegals, law students, and union representatives.¹⁸⁸ With the exception of child custody,¹⁸⁹ juvenile delinquency,¹⁹⁰ welfare,¹⁹¹ employment arbitration,¹⁹² and divorce,¹⁹³ where results conflict, this research is often cited (despite occasional precautions from some authors)¹⁹⁴ to justify oft-repeated general assertions that legal representation makes all the difference.¹⁹⁵

One might think that after decades of research in such a wide variety of subject areas, types of proceedings, adjudicatory styles, and types of representatives, a substantial body of objective, quantitatively-based knowledge would be available regarding when, where, and how to deliver legal assistance, and how to structure adjudicatory systems to make such systems

^{186.} See, e.g., STAPLETON & TEITELBAUM, *supra* note 20, at 7-23.

^{187.} See *supra* note 57 and accompanying text.

^{188.} See, e.g., SUBCOMM. ON SOC. SEC., *supra* note 154; KRITZER, *supra* note 63, at 23-78, 111-50; Cooper, *supra* note 154; Monsma & Lempert, *supra* note 21.

^{189.} See, e.g., Kelly & Ramsey, *Do Attorneys for Children Make a Difference*, *supra* note 154; Kelly & Ramsey, *Legal Representation in Child Protection Proceedings*, *supra* note 154.

^{190.} See, e.g., STAPLETON & TEITELBAUM, *supra* note 20; Aday, *supra* note 154; Burruss & Kempf-Leonard, *supra* note 154; Clarke & Koch, *supra* note 154; Duffee & Siegel *supra* note 154; Feld & Schaefer, *supra* note 154; Ferster & Courtless, *supra* note 154; Ferster et al., *supra* note 154; Hayeslip, *supra* note 154; Lemert, *Legislating Change*, *supra* note 154; Reasons, *supra* note 154. Other studies purport to show representation producing favorable results in juvenile proceedings. See, e.g., Feld, *supra* note 154; Maiman, *supra* note 154, at 9.

^{191.} See, e.g., Cooper, *supra* note 154; Handler, *supra* note 154; Hammer & Hartley, *supra* note 154. Other studies purport to show representation producing favorable results in welfare. See, e.g., Hagen, *supra* note 154.

^{192.} See, e.g., KRITZER, *supra* note 63, at 151-92; Deitsch & Dilts, *supra* note 154; Hill, *supra* note 154; Zirkel, *supra* note 154. Other studies purport to show representation producing favorable results in employment proceedings. See, e.g., Block & Stieber, *supra* note 154; Colvin, *supra* note 154; McDermott & Obar, *supra* note 154; Zirkel & Breslin, *supra* note 154.

^{193.} See, e.g., Cavanagh & Rhode, *supra* note 154; Sales et al., *supra* note 154; Wissler, *supra* note 154. *But see* Ellis, *supra* note 154 (concluding that lawyers are associated with higher settlement rates and greater use of joint custody). The findings in McMullen & Oswald, *supra* note 154, at 81, are ambiguous.

^{194.} See, e.g., Drahoszal & Zyontz, *supra* note 154, at 53.

^{195.} See, e.g., AM. BAR ASS'N ET AL., REPORT NO. 112A, REPORT TO THE HOUSE OF DELEGATES 9 (2006), available at [http://www.legalaidnc.org/public/participate/legal_services_community/ABA_Resolution_onehundredtwelvea\[1\].pdf](http://www.legalaidnc.org/public/participate/legal_services_community/ABA_Resolution_onehundredtwelvea[1].pdf); Debra Gardner, *Justice Delayed Is, Once Again, Justice Denied: The Overdue Right to Counsel in Civil Cases*, 37 U. BALT. L. REV. 59, 70 (2008) (claiming “substantial difference”).

accessible to pro se litigants.¹⁹⁶ We suggest the opposite is true: we know almost nothing as a result of these studies,¹⁹⁷ and all but two provide no information on representation effects that would not already have been available from instinct and conjecture.¹⁹⁸ And as Part I of this Article demonstrated, instinct and conjecture can be wrong in a way that matters. In this Part, we discuss previous studies at modest length because we make the following claim: almost all this literature is unworthy of credence, and at least at present (and until we know more from better research), the only way to produce credible quantitative results on the effect of legal representation is with randomized trials.

We motivate our discussion in part with two apparent puzzles. First, a 2006 observational study, based on a review of case files, of first-level unemployment appeals in the Boston metropolitan region reported that representation resulted in a 19% to 24% increase in the probability that a claimant who appeared at a hearing would prevail.¹⁹⁹ As discussed above,

¹⁹⁶. This is the premise of the extraordinarily exhaustive review and synthesis of the literature in Engler, *supra* note 80.

¹⁹⁷. Sandefur has attempted to obtain information from case-file-based observational studies in this area by conducting a bounds-based meta-analysis of a well-chosen subset of these studies. See Sandefur, *supra* note 141. Although we admire the attempt to glean information from these studies, we are skeptical that it can be done. Sandefur limited her meta-analysis to studies that considered legal representatives who appeared at some kind of hearing or adjudication. As such, as we discuss in Subsection II.A.1, none of these studies identifies a well-defined intervention, and thus none appears to be measuring a causal effect. This alone is enough to suggest that at least without further, strong assumptions about whether legal representation early in a case's development affects the probability that the case reaches a hearing (and is thus included in the datasets Sandefur's meta-analysis covers), the underlying studies are themselves of little value. Limiting oneself to bounds cannot solve this problem. Moreover, even assuming this problem away, the bounds Sandefur calculates are necessarily wide (as she recognizes), and all we really know is that the "truth" lies somewhere in the bounds. Given the selection effects present, which we discuss in this Part and some of which Sandefur discusses herself, there is no reason to believe that across studies, the truth generally lies near the middle, or near the bottom, or in any particular place within these bounds.

¹⁹⁸. To clarify, we do not claim that the observational studies we list in note 154, *supra*, have no useful information. These articles often provide rich descriptive statistics and non-quantitative information on the litigants, the advocates, the adjudicative setting, the nature of the issues raised, and the sociological and economic contexts in which representation takes place. Our criticism is limited to the attempt to draw inferences of causation regarding the effects of offers or actual use of representation from the quantitative data these studies contain.

¹⁹⁹. Owens, *supra* note 75. A law student statistical compilation came to similar conclusions, at least with respect to discharge cases. See Gotberg, *supra* note 154, at 32 ("[B]ringing an attorney roughly doubled the claimant's chances of receiving benefits."). This latter study

however, our own study, which examined the effect of an offer of HLAB representation in first-level unemployment appeals (the same system) in the Boston metropolitan region (the same region) from 2008-10 (two to four years later), found no evidence that the offer had a large effect on the probability of a claimant win. The first puzzle: What explains this discrepancy, and which result is reliable?

The second puzzle: Can it be that assigning a lawyer to represent a juvenile in a delinquency proceeding increases the likelihood that the juvenile will be incarcerated or removed from his home? Such is the purported conclusion in seven²⁰⁰ of over a dozen²⁰¹ observational studies based on case files regarding the effect of counsel in juvenile delinquency proceedings, studies conducted after *In re Gault*²⁰² established that a juvenile could not be incarcerated for a significant length of time unless he had been represented by (or waived) counsel. Note that this result persists even after variables such as offense seriousness/type, race, sex, prior involvement in the delinquency system, type of adjudicator, the presence of a prosecutor, and plea entered are included in regression equations.²⁰³

We believe that the answer to these two puzzles is methodological. The 2006 Boston unemployment study, the seven juvenile delinquency studies, and the other non-randomized studies referenced above all relied on the same “design,” namely, a review of case files to determine the success rate for cases with representation versus the success rate for cases without representation. A few of these studies included regressions in an attempt to “control for” certain variables.

Our claim here is that this design cannot at present succeed in this area. More specifically, case-file-based observational studies referenced above generally suffer from three sets of methodological problems: the failure to

was careful to note some of the issues we discuss in this Part, include selection effects. *Id.* at 36-37 (clients may choose attorneys in more difficult cases); *id.* at 37-38 (noting that the bar may simply be skillful at “rejecting less desirable cases”).

²⁰⁰. See Burruss & Kempf-Leonard, *supra* note 154, at 60; Clarke & Koch, *supra* note 154, at 300-01; Duffee & Siegel, *supra* note 154, at 551-52; Feld & Schaefer, *supra* note 154, at 717; Hayeslip, *supra* note 154, at 12; Lemert, *Legislating Change*, *supra* note 154, at 441-43; Reasons, *supra* note 154, at 170. The numbers in Aday, *supra* note 154, at 112-16, would have supported the same conclusion if misinterpreted, but this author understood some of the methodological concerns we discuss in this Article and thus declined to draw a causal link between representation and a less favorable result.

²⁰¹. See sources cited *supra* note 170.

²⁰². 387 U.S. 1 (1967).

²⁰³. See Aday, *supra* note 154, at 112-16; Burruss & Kempf-Leonard, *supra* note 154, at 60; Clarke & Koch, *supra* note 154, at 300-01; Hayeslip, *supra* note 154, at 12.

define an intervention being studied, the failure to account for selection effects (which come in multiple layers), and the failure to follow basic statistical principles to account for uncertainty. Not all observational studies referenced above suffer from all these problems, but all suffer from at least one. None of these problems is technical in the sense of concerning whether to use logistic or probit regression or propensity score balancing, or whether to use interaction or squared terms in regression equations. Instead, each of these problems is fundamental to the nature of the question being asked and to whether a study is viable.²⁰⁴ We contend that the causal claims in these studies are unworthy of credence. We also believe that the failure to address these methodological concerns may cause, and probably has caused in many instances, the following, easy-to-understand consequence: the wrong answer.

In Section II.A, we discuss these three methodological problems endemic to observational studies in this area. We conclude that these problems are so severe as to render intractable any effort to measure the effect of representation via an observational study, at least at present. We suggest further that one cannot at present tell whether the methodological problems have resulted in overstatements or understatements of representation effects. In Section II.B, we demonstrate how a well-run randomized control trial solves each of these concerns. In Section II.C, we briefly review the two previous randomized control trials that measure the effect of representation in civil cases and find that most of the quantitative conclusions of these studies are reliable.

A. Three Methodological Problems

1. Failure To Specify the Intervention: When Is Representation Assigned?

When a study purports to measure the effect of representation, what is it measuring? We do not refer here to potential differences in effectiveness among types of representative, *i.e.*, attorney (legal aid versus privately paid versus pro bono, or specialist in the relevant area of law versus non-specialist) versus law student versus paralegal versus a bankruptcy “petition preparer,”²⁰⁵ versus something else. This question is obviously critical, and at least some

²⁰⁴. On the critical nature of clarifying the question under consideration in any observational study, see generally D. James Greiner, *Causal Inference in Civil Rights Litigation*, 122 HARV. L. REV. 533 (2008).

²⁰⁵. See A. Jay Cristol, *The Nonlawyer Provider of Bankruptcy Legal Services: Angel or Vulture?*, 2 AM. BANKR. INST. L. REV. 353 (1994).

researchers have considered it carefully.²⁰⁶ Rather, we refer instead to the question of *when* in the life of a case the representative intervenes and the implications of this question for what data should be examined.²⁰⁷ For example, the 2006 non-randomized study of Boston area first-level unemployment appeals referenced above examined data only on cases in which an ALJ hearing was held.²⁰⁸ In doing so, this study followed the practice common in this area of examining only cases that reach some kind of hearing, thus excluding those cases that settle or those that are dismissed for lack of evidence, those dismissed for a failure by one of the parties to follow proper procedures, and those in which one litigant withdraws or fails to appear.²⁰⁹ This practice makes sense only if legal representation has no effect on a case's probability of reaching the hearing stage, meaning that representation does not change the probability that a litigant defaults or formally withdraws,²¹⁰ that a case settles,²¹¹ that it is dismissed, and so forth. If this is not the case, if representation could affect whether a case reaches a hearing, then a researcher who examines only cases that reach hearings will ordinarily have the wrong dataset for the question she wants to address. No amount of statistical modeling can adjust for the wrong dataset.²¹² We find implausible the

^{206.} See, e.g., SUBCOMM. ON SOC. SEC., *supra* note 154; KRITZER, *supra* note 63, at 23-77, 111-49; Clarke & Koch, *supra* note 154.

^{207.} On the need to identify sharply the nature of the counterfactual being studied, see Van Ryzin & Lado, *supra* note 11, at 2564-65.

^{208.} Owens, *supra* note 75, at 6. We verified this orally. Interview with Kenneth Owens, Dir. of Hearings, Mass. Div. of Unemp't Assistance, in Cambridge, Mass. (June 28, 2010).

^{209.} See, e.g., COURT STUDY GRP., *supra* note 154; KRITZER, *supra* note 63, at 23-77, 151-92; Block & Stieber, *supra* note 154; Clarke & Koch, *supra* note 154; Cooper, *supra* note 154, 1160-72; Duffee & Siegel, *supra* note 154; Fusco et al., *supra* note 154; Gunn, *supra* note 154; Hagen, *supra* note 154; Hayeslip, *supra* note 154; Hill, *supra* note 154; Kirp et al., *supra* note 154, 130-33; Lederman & Hrung, *supra* note 154; Mosier & Soble, *supra* note 154; Sarat, *supra* note 154; Schoenholtz & Jacobs, *supra* note 154; Zirkel, *supra* note 154; Zirkel & Breslin, *supra* note 154; Granberry & Albelda, *supra* note 154; see also Monsma & Lempert, *supra* note 21, at 630 ("No author modeling outcomes at the final disposition stage among cases that survive dismissal examines whether estimates are biased by nonrandom selection at earlier stages."); Steadman & Rosenstein, *supra* note 154, at 1333 (acknowledging the danger in their choice to analyze only cases that did not settle).

^{210.} But see Monsma & Lempert, *supra* note 21, at 657 ("Another way in which lawyers may help tenants avoid eviction is by impressing on them the importance of attending hearings.").

^{211.} But see Suzanne Reynolds, Catherine T. Harris & Ralph A. Peeples, *Back to the Future: An Empirical Study of Child Custody Outcomes*, 85 N.C. L. REV. 1629, 1670-71 (2007) (finding that unrepresented parties are more likely than represented parties to enter litigation in child custody matters).

^{212.} To be clear, it is possible to think of a type of representational intervention that occurs only for cases that reach hearings: lawyer-for-the-day programs are now common in many

assumption that legal representation has no effect on a case's probability of reaching a hearing,²¹³ but for most legal settings, the rate at which cases reach hearings could be studied, rendering our beliefs irrelevant.

Consider another set of examples from two disparate legal fields, a 2009 study of the effect of counsel in bankruptcy proceedings²¹⁴ and a set of studies in affirmative asylum proceedings.²¹⁵ These disparate milieux share an important feature in common, namely, that in both cases the individual (i.e., the potential client) is the one who must initiate and sustain²¹⁶ the formal proceedings that constitute the basis for study. In other contexts, including housing eviction cases, the potential client is in a defensive posture, and thus she has "only" to respond to a proceeding someone else initiated. In bankruptcy and affirmative asylum proceedings, researchers attempt to measure the effect of counsel by examining official files of concluded cases, meaning again they look at cases in which the person who has to initiate the relevant proceedings (the potential bankrupt or the immigrant) actually manages to do so. Thus, a file-review-based observational study of a representation effect in these settings makes sense only if there are no substantial barriers to initiating and sustaining a bankruptcy proceeding or filing an asylum application, a proposition we find implausible.

The closest intervention we can imagine for such studies is a legal services provider that offers representation to potential clients who have already

settings, particularly in housing (at least in Massachusetts). In these programs, a stable of lawyers comes to court on the relevant day to offer representation (such as negotiation during a settlement or mediation conference or engaging in a colloquy with the court) on that day only. *See REBECCA L. SANDEFUR & AARON C. SMYTH, ACCESS ACROSS AMERICA: FIRST REPORT OF THE CIVIL JUSTICE INFRASTRUCTURE MAPPING PROJECT* 11, 31-132 (2011) (defining such programs and documenting their existence or non-existence in each state), available at http://www.americanbarfoundation.org/uploads/cms/documents/access_across_america_first_report_of_the_civil_justice_infrastructure_mapping_project.pdf. But in the studies cited at *supra* note 209, representation was not provided in this manner.

²¹³. For a contrary view, see Leandra Lederman, *Which Cases Go to Trial? An Empirical Study of Predictors of Failure To Settle*, 49 CASE W. RES. L. REV. 315, 338-40 (1999).

²¹⁴. Pardo, *supra* note 154.

²¹⁵. *See, e.g.*, Kerwin, *supra* note 154, at 6.

²¹⁶. Under § 301(a) of the Bankruptcy Code, 11 U.S.C. § 301(a) (2006), at least in the context of Chapter 7 filings, a would-be bankruptcy petitioner may initiate a proceeding by filing a short petition, a list of creditors, and other accompanying documents; the petitioner must also have attended a credit counseling course. Within fourteen days, however, the petitioner must file a schedule of assets and liabilities and a statement of financial affairs. *See FED. R. BANKR. P.* 1007(c). Our thanks to Rafael Pardo for educating us as to these requirements.

initiated bankruptcy or asylum proceedings. Even here, the datasets still do not match this hypothetical intervention.²¹⁷

The point here is that when conducting an observational study on the effect of counsel, a researcher must identify the nature of the representation that would hypothetically be offered. This principle, now basic in the business of conducting observational studies, does not appear to have been internalized in the effect-of-counsel literature.²¹⁸

Note that one cannot tell how a failure to specify the nature of studied representation affects results. For the bankruptcy and immigration examples, one might speculate that barriers to filing would suggest that study conclusions underestimate the effect of making representation available. The theory here is that only those litigants who are articulate, as well as persistent enough to negotiate their way through soul-crushing paperwork, ever initiate the relevant legal proceedings, and that this special group of litigants might need representation less. But legal services programs may themselves have barriers that persons desiring representation must overcome; as discussed in Part I, telephone intake systems might be one such barrier. Meanwhile, with respect to limiting datasets to cases that reach a hearing stage, representation may make it less (not more) likely that clients will attend hearings; legal services providers have argued that they increase judicial efficiency by encouraging litigants with weak cases to settle or give up.²¹⁹ Thus, one cannot tell in advance whether a failure to specify the kind of representation considered induces an understatement or overstatement of a representation effect. Indeed, without identifying the intervention with at least some degree of specificity, it is not clear what is being studied.

²¹⁷. Again, the difficulty is that the data probably include represented cases that were filed only because an attorney aided in the filing process.

²¹⁸. See, e.g., PODOLSKY & O'BRIEN, *supra* note 154, tbls.46-47 (cross-tabulating represented versus pro se outcomes in eviction proceedings, but only for tenants who appeared at the hearing (table 46) or for cases that were contested (table 47), both of which are intermediate outcomes that the presence of counsel likely affected); Hayeslip, *supra* note 154, at 13 (including the kind of plea entered in juvenile proceedings in a regression); Lederman & Hrung, *supra* note 154, at 1284 (dropping cases decided by summary judgment from an analysis of time to trial, despite the fact that lawyers probably affect whether the case gets decided on summary judgment); Monsma & Lempert, *supra* note 21, at 646, 657 (including in regressions the percentage of rent paid at the time of a hearing as well as a variable measuring case delay).

²¹⁹. See, e.g., Clark & Reyes, *supra* note 74, at 224 & n.36; Emsellem & Halas, *supra* note 75, at 312.

2. Selection Effects: How Is Representation Assigned?

In the business of measuring how much of a difference some intervention makes, selection effects are the oldest of old news. An example of a well-known selection effect is the fact that areas with greater police presence might have higher probabilities that crimes will be committed.²²⁰ We do not conclude that greater police presence causes more crime. Instead, we recognize that relevant decisionmakers may direct more police officers to neighborhoods deemed likely to suffer crimes.

As applied to the subject matter of this Article, the concern is that in a case-file-based observational study, the process by which potential clients sort themselves into represented versus unrepresented groups causes the two groups to differ in ways other than representation alone. A greater success rate (whatever “success” means) for the represented thus cannot be attributed to representation (as opposed to these other differences). The term “selection effects” comes from a focus on how litigants are selected (or select themselves) into the represented or unrepresented group.

Perhaps the fact that selection effects are such an old saw is to blame for the fact that few studies in this area take them seriously.²²¹ The response to the problem of selection effects in the literature regarding legal representation has essentially been three-fold: (i) ignore them, or what is essentially the same thing, recognize that they are a danger, but assert without elaboration that observed disparities in outcomes are too large to be explained by selection;²²² (ii) attempt to adjust for them via regression based on variables available in case files;²²³ and/or (iii) argue that nothing else can be done about the problem,

²²⁰. See, e.g., NAT’L RESEARCH COUNCIL, UNDERSTANDING CRIME TRENDS 107–08 (2008).

²²¹. One notable exception is Monsma & Lempert, *supra* note 21; these authors are acutely aware of selection effects and attempt to use modeling to remedy them. Thus, our criticism of this and a handful of other studies is that we do not believe modeling, particularly modeling based on variables available from case files, is an effective remedy.

²²². See, e.g., BD. OF IMMIGRATION APPEALS, *supra* note 154, at 13. The assertion by the Board of Immigration Appeals that its pro bono representation (as opposed to its case selection) is responsible for more favorable outcomes is breathtaking in light of its explicit recognition that “[g]iven its limited resources and current case screening design, the project selects [for representation] the most meritorious cases on appeal before the Board.” *Id.* at i; see also KRITZER, *supra* note 63, at 36 (asserting without elaboration that the observed disparities in outcomes are too large to be explained by selection); Kirp et al., *supra* note 154, at 67 (same); Popkin, *supra* note 154, at 1032–33 (same); Schoenholtz & Jacobs, *supra* note 154, at 744–46 (same); Steadman & Rosenstein, *supra* note 154, at 1333 (same).

²²³. See, e.g., HAMMITT, *supra* note 154, at 14–19; KRITZER, *supra* note 63, at 151–92; WOMEN’S LAW CENTER, *supra* note 154, at 28–29; Aday, *supra* note 154; Block & Stieber, *supra* note 154,

so the choice is between potentially-but-not-certainly wrong quantitative information and no quantitative information.

One of our aims in this Article is to persuade that these responses are inadequate. With respect to the first response, the idea that some differences are too large to be due to selection effects, we offer the contrast between the result of our study, which finds no evidence of a difference in win rates between those offered and those not offered HLAB representation, and the result of the 2006 observational study of first-level unemployment appeals in Boston, which reported differences of nineteen to twenty-four percentage points.²²⁴ As explained in Section II.B, below, randomization renders our results immune from selection effects, while the 2006 study (which included no background variables) is highly subject to them. The third response—premised on the assertion that only bad quantitative information is available—is simply wrong, as our own study and two others using randomization demonstrate.²²⁵

With respect to the second response, the use of regression (or other quantitative techniques) to “control for” differences in background variables, we discuss in this subsection some of the processes by which a party²²⁶ to a dispute could end up with representation. We suggest that at least three actors may be involved in these processes: the judge, who may appoint counsel; the litigant or potential client, who may seek representation; and the representative, who decides whether to provide representation. Our point here is that decisions by each of these actors induce selection effects that may be, and probably are, so strong as to undermine any attempt to adjust for them via regression or some more advanced method. Case files typically do not include enough of the right variables, such as (per Part I) proxies for effort, energy,

at 545; Clarke & Koch, *supra* note 154, at 271-74; Deitsch & Dilts, *supra* note 154, at 116-17; Eldridge, *supra* note 154, at 300; Feld, *supra* note 154, at 1294-1317; Hayeslip, *supra* note 154, at 11-12; Lederman & Hrung, *supra* note 154, at 1283-95; McDermott & Obar, *supra* note 154, at 104; Monsma & Lempert, *supra* note 21, at 646; Hammer & Hartley, *supra* note 154, at 191-94.

²²⁴ True, these two research efforts are studying different things, because the no-effect result we quoted above corresponded to offers of representation, while the 2006 study considered only claimants who attended hearings. But that merely restates our point in Section II.A: we do not know what intervention the 2006 study contemplated and thus do not know what question it purported to answer.

²²⁵ See, e.g., STAPLETON & TEITELBAUM, *supra* note 20, at 53, 185-92 (explaining and checking randomized design); Seron et al., *supra* note 20, at 423-24 (explaining randomized design).

²²⁶ Most of the parties in the studies cited above are low-income, so we focus much of our attention on low-income potential clients.

articulateness, and persistence, to allow any presently available adjustment technique to provide credible results.

a. Judge-Induced Selection Effects: Outcome-Driven Counsel Appointments?

To what kind of cases do judges appoint counsel? To understand the importance of this question, recall the second of the “puzzles” with which we began Section II.A: about half of the studies examining the effect of counsel in juvenile delinquency proceedings find an association between counsel and greater levels of state restriction (e.g., incarceration, removal from the home) at disposition, meaning less favorable case outcomes. This association persists in regression equations including variables such as offense seriousness/type, race, sex, prior involvement in the delinquency system, type of adjudicator, the presence of a prosecutor, and plea.²²⁷ Despite this persistence, the idea that lawyers cause their juvenile clients substantial legal harm, such that these clients would be better off facing serious criminal charges without counsel, strikes us as implausible. We would be among the last to say that a harmful counsel effect is completely impossible, but at a minimum we are driven to look for an alternative possible explanation.

The alternative explanation probably lies in the way in which cases are selected to receive counsel, which is another way of saying a selection effect: studies showing higher incarceration rates for represented juveniles “strongly suggest[] that the court [is] culling the docket for serious cases likely to receive severe dispositions and then requiring counsel in those cases but not others.”²²⁸ To clarify this statement, recall that *In re Gault*²²⁹ did not provide a right to counsel for juveniles charged, or even convicted, of serious crimes. Rather, *Gault* held only that a juvenile must be provided with (or must intelligently waive) counsel if the proceeding might “result in commitment to an institution in which the juvenile’s freedom is curtailed.”²³⁰ Given the aversion to lawyers in juvenile proceedings in some jurisdictions, it seems likely that at least some juvenile judges make an early decision in each case (probably on the basis of the initial reports by police and social workers and other information available early in the proceeding) regarding whether this child likely “needs”

^{227.} See *supra* note 203 and accompanying text.

^{228.} DONALD L. HOROWITZ, THE COURTS AND SOCIAL POLICY 192 (1977); *see also* Clarke & Koch, *supra* note 154, at 267 (suggesting the same explanation).

^{229.} 387 U.S. 1 (1967).

^{230.} *Id.* at 41.

incarceration, and if so, counsel is appointed. Thus, a researcher retrospectively examining case files in an observational study to measure the effect of counsel will find a “treated” group (those with counsel) fundamentally different from the “control” group (those without) because the ultimate decisionmaker has already assigned counsel to a higher percentage of those likely to be incarcerated.

A researcher might respond that she has “controlled for” a variety of factors, including offense level and prior involvement in the delinquency system, by including these variables in a regression equation, and thus that she has addressed the selection issue. Unfortunately, she has probably not addressed the selection issue (though no one ever knows for sure). The researcher is probably not controlling for the right variables, or enough of the right variables. Juvenile dispositions may turn on a judge’s conclusions regarding a child’s fundamental character, and this conclusion may be based in large part on reports of activities or actions other than the offense alleged or on the child’s official delinquency record. Possible sources of information include personal observation, school officials, arresting officers, or probation officers.²³¹ In short, regression (or some fancier adjustment technique) can only adjust for differences that the researcher sees and measures, and it appears that researchers are not seeing the right variables. Putting aside the juvenile delinquency context, important variables that come to mind are effort, energy, articulateness, and persistence. Without adequate proxies for these variables, no adjustment technique will work well.²³²

b. Client-Induced Selection Effects, Part I: Only the Strong Want To (and Do) Survive?

What kind of person forms the desire to obtain representation and is also capable of effectuating this desire?²³³ If this kind of person is different from the kind of person who never forms the desire to be represented (or from the kind of person who would like to obtain representation but cannot), and the

^{231.} See, e.g., Maiman, *supra* note 154, at 6-7.

^{232.} Some readers may be familiar with a technique called “instrumental variables.” For the reasons discussed in Angrist et al., *supra* note 25, and Imbens & Rubin, *supra* note 25, we have little faith in this technique’s capacity to solve the problems we identify above without the adequate variables we think are required. The adjustment techniques we discuss and attempt in “Question 6,” *see supra* Section I.E, are statistical generalizations of instrumental variables that relax some of the less plausible assumptions of the latter technique.

^{233.} See Jerome E. Carlin & Jan Howard, *Legal Representation and Class Justice*, 12 UCLA L. REV. 381, 423 (1965).

difference involved implicates how these persons would fare without counsel, then selection effects will cause observational studies based on review of case files to produce the wrong answer.²³⁴

Here, we distinguish two situations: those in which most represented persons in a dataset obtain representation by paying for it, and those in which most represented persons obtain counsel via a legal aid organization or via pro bono counsel. When measuring a counsel effect in a situation in which most or all persons with lawyers pay for representation, the danger of selection effects is straightforward: those who can afford lawyers are different from the less well off in ways that matter. We are not comparing two groups of people who are alike in all ways except for representation. We have no way of knowing how much of an observed improvement (if any) in outcomes between represented and unrepresented groups is due to representation as opposed to differences in income and related variables.

When measuring a counsel effect in a situation in which many persons who obtain representation do so via legal aid, pro bono, student assistance, or some other low- or no-cost²³⁵ model, we suspect that those who form the desire to obtain representation will constitute a disproportionately worldly, future-looking, and risk-averse subset of the general population. We made essentially this same point in our discussion above of “Go-Getters” and “Regulars,” so we will not repeat it here. The concept is clear: selection effects (strong ones) may be, and probably are, at work.²³⁶ Thus, a simple comparison of outcomes among the represented versus the unrepresented, the “design” used in most effect-of-representation studies, almost certainly is distorted by effects due to differences in the type of person who finds representation, in that those who manage to secure representation would be more likely to win in the absence of representation than those who do not secure representation.

To adjust for (and thereby eliminate) such type-of-person effects, a researcher needs measurements of the characteristics that induce litigants to obtain lawyers and to succeed in adjudicatory settings. That is difficult for traits such as tenacity and ability to communicate. One can try to substitute

²³⁴. See Laura K. Abel & Susan Vignola, *Economic and Other Benefits Associated with the Provision of Civil Legal Aid*, 9 SEATTLE J. SOC. JUST. 139, 156 (2010).

²³⁵. We speak here of “low-cost” or “no-cost” in terms of cost to the potential client.

²³⁶. See Michael Millemann, Nathalie Gilfrich & Richard Granat, *Limited-Service Representation and Access to Justice: An Experiment*, 11 AM. J. FAM. L. 1, 5 (1997) (“What distinguished the capable from the incapable *pro se* litigant . . . was not the difference between a high school or college education. Rather, it was more basic factors: the ability to speak and read English; a basic intelligence level; the absence of emotional and mental disabilities; and some degree of self-motivation, among other qualities.”).

observable and measurable characteristics that might possibly serve as proxies, such as education level, but using such proxies raises two issues: first, how good the proxies are,²³⁷ and second, whether even poor proxies are available to researchers conducting observational studies who primarily depend on case files. On the latter point, only a handful of the studies referred to above collect and use any variables about would-be clients in their datasets, with this deficiency no doubt in part due to the absence of relevant data in case files.²³⁸ In short, we are pessimistic that an observational study can arrive at the right measure of the effect of legal representation without including accurate measurements of strong proxies for the characteristics identified above.²³⁹

c. Client-Induced Selection Effects, Part II: Different Circumstances?

The previous Subsection discussed the potential client-induced selection effects stemming from the possibility that individuals who seek and obtain representation may be different from individuals who do not along dimensions that implicate the likelihood of success in a legal system. But selection effects can arise even if individuals who do and do not seek and obtain representation are the same in all relevant ways, or even when some third person or entity (perhaps a repeat player such as a union) decides whether the individual receives representation, because the circumstances of the cases in which representation is obtained may be different from those in which it is not. Perhaps individuals seek and obtain, or organizations provide, representation more often when the likelihood of a loss appears higher, or the consequences of a loss appear larger.²⁴⁰ If so, then again, an observational study comparing represented cases to unrepresented cases is comparing two sets of cases that are fundamentally different in ways other than the representation, and as a result, the comparison provides little useful information about the effect of representation.

^{237.} See *id.* As suggested in our own study, *see supra* text accompanying note 146, education was not an especially good predictor for whether representation would make a difference.

^{238.} Of the dozens of studies cited *supra* note 154, only a handful collected and used any information at all on personal characteristics of the person who was represented or unrepresented. Some of those include WOMEN'S LAW CTR., *supra* note 154; Aday, *supra* note 154; Block & Stieber, *supra* note 154; Clarke & Koch, *supra* note 154; Feld, *supra* note 154; Hayeslip, *supra* note 154; Littwin, *supra* note 154; and finally Monsma & Lempert, *supra* note 21, who recognize, discuss, and address this issue with great sophistication.

^{239.} Again, Monsma & Lempert, *supra* note 21, is an exception on this point.

^{240.} *Id.* at 639.

d. Lawyer-Induced Selection Effects: Different Ways of Culling Clients?

What sort of clients do lawyers (or law students or paralegals) choose to represent? The process by which lawyers accept cases almost certainly produces selection effects.

As was true in the previous subsections concerning client-induced selection effects, it is useful here to distinguish two situations, those in which most represented persons in a dataset obtain representation by paying for it, and those in which most represented persons obtain counsel via a legal aid organization or via pro bono counsel. First, the paid attorney context: here, we focus on situations in which the attorney's compensation includes a contingent fee arrangement, or a transaction fee arrangement, or a cap on the total amount payable. With such payment mechanisms in place, particularly in contingent fee arrangements, lawyers have an incentive to search for and accept cases carrying a high probability of a favorable outcome but requiring a minimum amount of work.²⁴¹ Thus, one might suspect paid lawyers to choose the strongest cases, using information from, e.g., client interviews and available documents to attempt to find the strong cases. Such information is typically not available to a researcher, and thus cannot be "adjusted for" via regression or some other method. If this is correct, then represented cases would have stronger underlying facts and/or law than unrepresented ones. Accordingly, the process by which paid attorneys select cases probably renders represented cases unlike unrepresented cases in a way relevant to the outcome.²⁴²

Now consider the situation in which most represented persons in a dataset are receiving legal assistance from low- or no-cost service providers, such as a legal aid entity or a student-based group. As noted above, most such providers are oversubscribed,²⁴³ so they must have some case selection mechanism. There are several possible mechanisms. In one, discussed by Monsma and Lempert,²⁴⁴ the provider searches for cases in which (it believes) it will make a difference. As Monsma and Lempert note, this mechanism, if successful (and there are questions as to whether providers can in fact choose the cases in which they will make a difference)²⁴⁵ can induce what they label a "curvilinear"

²⁴¹ If counsel's payment is by the hour, this incentive structure may reverse.

²⁴² See, e.g., Mashaw, *supra* note 57, at 18; Sternlight, *supra* note 21, at 389.

²⁴³ See *supra* note 8 and accompanying text.

²⁴⁴ Monsma & Lempert, *supra* note 21, at 639.

²⁴⁵ Monsma and Lempert report, for example, that a legal aid organization focused more of its resources in a housing/eviction practice on cases in which the tenant was being evicted for alleged misconduct, as opposed to for non-payment of rent, but that the organization's representation appeared to have no effect (or a negative effect) in misconduct proceedings.

relationship: the provider attempts to turn down cases that will fail or succeed regardless of its intervention and accept only cases in which representation will turn a less favorable outcome into a more favorable one.²⁴⁶ Other service providers may intentionally seek the most meritorious cases, at least in some settings. In one example, a court-based program cited “limited resources” to justify its attempt to “select[] the most meritorious cases”²⁴⁷ for representation, a statement we find breathtakingly ironic. Once again, all these efforts to base the choice of which cases will receive offers of representation on anticipated outcomes will undermine an observational study based on case files because the represented cases, by design, differ from the unrepresented in ways other than the representation. This can be avoided only if the researcher has access to variables encapsulating the information service providers use to make their representation decisions, and such variables are typically not available from case files.

e. Selection Effects: Are Plausible Guesses Possible?

Suppose, as we believe, some or all the mechanisms identified above are present in an effect-of-counsel observational study. Suppose further that at least in a study based on case files, the variables needed to address the resulting selection effects (particularly measurements of characteristics of the potential client) are rarely available. Can we at least say with confidence whether the selection effects are likely to cause an over- or understatement of a representation effect? In our view, broad statements in this area (e.g., “the benefits lawyers provide are typically overstated” or “estimates of the differences lawyers make are usually understatements”) are at present impossible. Some of the selection mechanisms described in the previous paragraphs might be expected to induce an overstatement of how much of a difference representation makes; some might be expected to induce an understatement.

3. Accounting for Uncertainty

There is little point in quantitative comparisons of results of cases with and without representation (even in a randomized study) unless the comparison

Id. at 639-48; *see also* Millemann, Gilfrich & Granat, *supra* note 236, at 4 (explaining the necessity of an exhaustive intake interview).

²⁴⁶ Monsma & Lempert, *supra* note 21, at 639.

²⁴⁷ BD. OF IMMIGRATION APPEALS, *supra* note 154, at i.

includes an honest assessment of uncertainty or, equivalently, a measurement of statistical significance. This point is so obvious that we will not belabor it, except to note that roughly half of the studies we found purporting to provide a quantitative assessment of the effect of representation reported one or more results (usually all of them) without calculating a measure of statistical significance.²⁴⁸

One key unappreciated point here is missing data. Almost every study ever attempted has some missing data, and unless the amount of missing data is truly de minimis (as in our study), it should be handled via what are by now standard statistical techniques²⁴⁹ to reflect the additional uncertainty missing data induces (techniques we used despite the small amount of missing data in our study).

We emphasize that more is at stake than statistical fastidiousness. Honest estimates of uncertainty matter; they are all that allows quantitative analysts to distinguish true patterns from random chance. Honest accounting for uncertainty due to missing data is part of this effort.

B. How Well-Run Randomized Experiments Solve These Problems

Section II.A discussed several methodological problems, at least one of which (and usually more than one of which) affected the credibility of each of the observational studies previously identified, particularly those based solely on case records. Briefly, these problems fell into three categories: (i) a failure to specify the nature of the representation/intervention contemplated in the study; (ii) a failure to account for selection effects of various types; and (iii) a failure to account honestly for statistical uncertainty. In this Section, we briefly discuss how a randomized control trial addresses the first two of these issues; the third is a matter of applying proper statistical technique.

²⁴⁸. E.g., Bd. of Immigration Appeals, *supra* note 154; Court Study Grp., *supra* note 154; Kritzer, *supra* note 63, at 79-110, 151-92; Lemert, Social Action, *supra* note 154; Rolph et al., *supra* note 154; Clarke & Koch, *supra* note 154; Cooper, *supra* note 154; Ferster et al., *supra* note 154; Fusco et al., *supra* note 154; Gunn, *supra* note 154; Handler, *supra* note 154; Hannaford-Agor & Mott, *supra* note 154; Hill, *supra* note 154; Kerwin, *supra* note 154; Kuck, *supra* note 154; Lemert, *Legislating Change*, *supra* note 154; McMullen & Oswald, *supra* note 154; Mosier & Soble, *supra* note 154; Murphy, *supra* note 154; Reasons, *supra* note 154; Steadman & Rosenstein, *supra* note 154; Yngvesson & Hennessey, *supra* note 154; Hammer & Hartley, *supra* note 154; Note, *Legal Services and Landlord-Tenant Litigation*, *supra* note 154; Note, *Representation in Child-Neglect Cases*, *supra* note 154; Granberry & Albelda, *supra* note 154; Hall, *supra* note 154; Maiman, *supra* note 154; Steiner, *supra* note 154.

²⁴⁹. See, e.g., Roderick J. A. Little & Donald B. Rubin, *Statistical Analysis with Missing Data* (2d ed. 2002); J. L. Schafer, *Analysis of Incomplete Multivariate Data* (1997).

First, with respect to the need to specify the nature of the contemplated intervention, because randomized control trials are forward-looking, a researcher knows exactly what intervention will be made for both the treated and control groups. In the HLAB study, the interventions were an offer of HLAB representation (treated) versus provision of names and telephone numbers of alternative legal services providers (control). And the researcher will know exactly when in the course of a case the intervention will be made. Here, the relevant moment was between the initial ruling by a claims adjuster and a hearing before an ALJ.

Second, as to the population being studied, the randomization destroys selection effects. Up to random variation, the set of persons offered representation is identical to the set of persons not offered representation. Because selection to an offer of representation is random, selection cannot be related to case outcomes. Note that as clarified above, even in a randomized study, the set of represented claimants is not identical (even within random variation) to the set of unrepresented claimants, so statements about the effect of actual use of representation itself require greater caution.

We close this Part by stating that the considerations discussed here lead us to credit the two previous randomized studies of which we are aware that measure the effect of legal representation in civil proceedings in the United States.²⁵⁰ The randomized control trial format compelled both sets of researchers to specify the nature and timing of the intervention being studied. In one study, the offer of representation was made to juveniles who had had delinquency petitions filed against them but who had not yet had their first scheduled court date;²⁵¹ in the other, offers were made to defendant/tenants responding to nonpayment of rent eviction proceedings as they waited in line before court appearances outside the court clerk's office.²⁵² The specification of the timing of the offer clarifies the critical issue of whether the study involved a potentially distinct subpopulation, and thus how generalizable it is. For example, the Seron et al. housing study concerned only tenants with the organizational skills, motivation, and perhaps confidence in the system and in the facts of their cases to attend the first calls at housing court in response to eviction proceedings. And of course, the randomization in both studies destroyed possible selection effects.

^{250.} STAPLETON & TEITELBAUM, *supra* note 20; Seron et al., *supra* note 20. To be completely clear, we are not certain that we credit the instrumental variables results presented in Seron et al., *supra* note 20, at 428-29, for the reasons discussed in Angrist et al., *supra* note 25. Nevertheless, the other results in this study are powerful enough.

^{251.} STAPLETON & TEITELBAUM, *supra* note 20, at 53-56.

^{252.} Seron et al., *supra* note 20, at 423.

In terms of findings, however, the substantive results of these studies are mixed. The juvenile delinquency study was actually two separate randomized trials that took place in cities the authors code-named “Gotham” and “Zenith.” In Gotham, offering legal counsel had no discernible effect on case outcomes; in Zenith, juveniles offered study lawyers were found delinquent in 40% of cases, while 56% of those not offered study lawyers were found delinquent, a difference the authors found statistically significant.²⁵³ The housing study concluded that an offer of counsel caused a case to take about a month longer to adjudicate, but that tenants offered lawyers defaulted less often, suffered fewer adverse judgments and warrants for eviction, and obtained more orders requiring repairs and rent abatement.²⁵⁴ The authors reported that these results were statistically significant.

Thus, these studies have mixed results in terms of whether counsel makes a difference. If put together with the results of our own study, which found no evidence that an offer of representation had a large effect on the outcome but did find that an offer of representation caused a potentially harmful delay, these studies suggest that we cannot as yet make reliable generalizations about the circumstances under which representation makes a positive difference.

III. WHERE DO WE GO FROM HERE?

If some nontrivial portion of our discussion in Part II is correct, then we currently have astonishingly little credible, quantitative information about the effect of representation (offers or actual use) in civil proceedings, and at present, such information can only be obtained via randomized trials. We therefore dedicate this final Part to beginning a conversation on the following two questions: what are the limits of randomized trials, and how can we structure them so as to maximize the information they produce?

A. *The Limits of Randomized Studies*

As is true of most methods designed to accumulate knowledge, randomized experiments have limits and drawbacks. In our view, there are three that are particularly important: an inability to evaluate the effects of systemic change,

²⁵³. STAPLETON & TEITELBAUM, *supra* note 20, at 51-52.

²⁵⁴. The authors found no statistically significant differences in number of court appearances, number of motions filed, and number of post-judgment motions filed. Seron et al., *supra* note 20, at 427.

operational and ethical challenges in the field operation, and the objections of some legal services providers to engaging in randomized studies.

1. *Systemic Change*

By far the most important limit of any sort of case-by-case quantitative evaluation, including a randomized one,²⁵⁵ is that it cannot measure the effect of lawyer efforts aimed at systemic change. Case-by-case quantitative evaluations operate and evaluate within systems, not across system structures, and thus they can provide little information about efforts to change the overall system as a whole. Case-by-case quantitative studies can provide information on whether a system as presently constituted is pro se accessible; they can provide less information on how the system came to be pro se accessible (if in fact it is), or what would be required to improve an inaccessible system. In the context of the Massachusetts DUA's first-level appeals system, for example, if it is the case that if the system is pro se accessible (the second of our three explanations articulated in response to Question 8, *supra*), it may be that this is so because legal services providers and lawyers more generally worked to make it so.²⁵⁶

To illustrate this point, one of the seventy-eight cases randomized to an offer of HLAB representation (an offer the potential client accepted) resulted in a systemic change applicable to certain employees who obtain, and then lose, part-time work.²⁵⁷ Our study, which was necessarily limited to outcomes such

^{255.} By "case-by-case evaluations," we mean analyses in which the study unit is an individual legal matter. Thus, all of the studies identified *supra* note 154 have this limit.

^{256.} Structural change need not always come from class action lawsuits or test cases. See Gary Bellow, *Steady Work: A Practitioner's Reflections on Political Lawyering*, 31 HARV. C.R.-C.L. L. REV. 297, 300 (1996) ("In other situations, we pursued aggregate results by filing large numbers of individual cases.").

^{257.} The particular fact situation was as follows: the employee was receiving full unemployment benefits; the employee obtained a part-time job, causing the DUA to reduce unemployment benefits to reflect a portion of the new income from the part-time job; the part-time job was seasonal, meaning it had a definite and firm end date, so in no event would the reduction due to the part-time job have applied beyond this definite and firm end date; the employee lost the part-time job through her own fault; and the DUA continued to apply a reduction in benefits due to the now-defunct part-time job, the idea being to disincentivize applicants from losing part-time employment through fault. Prior to the HLAB case, the DUA's regulations applied the continued reduction in benefits due to the now-defunct part-time job indefinitely. But after HLAB appealed a client's case through the administrative proceeding and filed a district court action, the agency agreed to revise its regulations so as to apply the reduction due to the now-defunct part-time employment only through that job's definite and firm end date. E-mail from Aaron Dulles, Student-Attorney, Harvard

as win/loss and time could not detect whether any of the cases randomized to the no-HLAB-offer group resulted in structural change. Even if we had been able to detect such outcomes, their impact would have been difficult to measure quantitatively without detailed information about the number of cases subject to the systemic changes involved.

Despite this acknowledged limit, a series of randomized evaluations should allow sharper focus on the question of whether, if structural change is the primary goal, that goal can be more economically achieved via targeted representation of test cases without an overarching program of services in a particular legal area, particularly if the overarching program provides little benefit to (or inflicts a harm on) the individual potential client.²⁵⁸

2. Fielding Studies

A second set of drawbacks of randomized trials consists of operational issues, including the time needed to effectuate them and ethical limits that arise in certain settings. As those involved in drug development are aware, randomized trials take time and require complex field operations to maintain. And ethically speaking, randomized experiments are not feasible in all settings. Our active but ultimately unsuccessful attempt to pursue a study in the area of affirmative asylum applications²⁵⁹ provides an object lesson. Despite a service provider's desire to engage in a randomized evaluation of its affirmative asylum practice, we discovered that intake for each potential case took approximately one day of attorney time. Thus, increasing (say, doubling) the number of cases as to which intake was accomplished, so that half could be randomized to an offer of representation and half randomized to no offer, was prohibitively expensive without a large grant. Perhaps even more importantly, we faced difficulties in constructing a system to collect outcome information for study subjects randomized not to receive an offer of representation. Consider that an outcome of considerable interest in this kind of study would be whether study subjects ever file asylum applications. For obvious reasons, some potential

Legal Aid Bureau, to Tom Feriss, Dir. of Comm'cns, Harvard Legal Aid Bureau (Mar. 24, 2011, 12:15 PM EST) (on file with authors).

²⁵⁸. By asking this question, we do not mean to imply that a handful of test cases are sufficient to induce structural change that sticks. Our sense is that structural change may require test cases, followed by changes in law and/or system design, followed by careful watchdogging (perhaps in the form of case-by-case representation) to make changes last.

²⁵⁹. The provider's "affirmative asylum application" practice focused on assessing whether persons already in the United States should file applications for asylum from immigration authorities, and if they should, on assisting these persons through the process.

study subjects were undocumented and thus did not have valid social security numbers, so our only option to find out whether study subjects filed asylum applications was to provide immigration authorities with names, birthdays, and (potentially) addresses. But providing this information to immigration authorities might potentially have exposed study subjects to deportation proceedings.

3. Provider Objections

The third drawback of randomized studies is the reluctance of some legal services providers to engage in them. We begin with this question: other than self-protectiveness,²⁶⁰ and fear of overgeneralization of unfavorable results by political enemies of legal services,²⁶¹ what other concerns have legal services providers articulated in reaction (or resistance) to randomized evaluation? In pursuing our research, we have heard several reactions. One is that randomizing potential clients is “heartless.” When we have probed this reaction, we have found that the perception of heartlessness comes from the fact that the randomized study forces the provider to perceive that it is turning potential clients away due to limited resources, a fact of life that has little to do with the randomized study. Previously, providers could “know” the fact of oversubscription but limit their perception of it by, for example, turning away potential clients upon receiving an initial telephone call.²⁶² The randomized

^{260.} See Gary Bellow, *Management of Legal Services: Legal Aid in the United States*, 14 CLEARINGHOUSE REV. 337, 343 (1980) (using this term). “No professional group easily supports research which may cast a shadow over its fundamental beliefs and practices.” STAPLETON & TEITELBAUM, *supra* note 20, at xii; *see also* Richard D. Schwartz, *Foreword* to W. VAUGHAN, *COUNSEL IN AMERICAN JUVENILE COURTS*, at xii (1972) (referring to judges who resisted a randomized study of the effect of representation in juvenile delinquency proceedings); Van Ryzin & Lado, *supra* note 11, at 2553 (referring to the “historical distrust of evaluation among many legal services providers”).

^{261.} *But see* Richard Zorza, *Avoiding the “Shut Down Effect” from Uncertain Research Results*, CONCURRING OPINIONS (Mar. 28, 2011, 5:25 PM), <http://www.concurringopinions.com/archives/2011/03/avoiding-the-shut-down-effect-from-uncertain-research-results.html> (“Lurking behind much of the debate in this symposium is anxiety that negative findings about access to justice services will strengthen and facilitate attempts to reduce resources for access to justice services. While it would be impossible to rebut the claim that this might happen, that cannot be an argument against conducting or reporting research. On the contrary, it has to be an argument for more and better research.”).

^{262.} “Some, perhaps many programs, manage the disconnect between the huge area of demand and the relatively scarce resources by establishing highly limited times for intake in priority areas.” Richard Zorza, *Access to Justice: The Emerging Consensus and Some Implications*, 94 JUDICATURE 156, 165 (2011).

evaluation requires more interaction with a caller before full representation is not offered. As one of the providers with whom we work put it, “We can see the blood now.”²⁶³

A second, less emotional reaction is as follows: the premise of our HLAB unemployment study was that once a set of eligible potential clients had been identified, HLAB ceded to a randomizer control over which among the eligible set would receive an offer of representation. One might object, however, that this design cannot capture the effect of representation because one of the tasks attorneys, particularly legal aid attorneys, perform is to pick out which cases will have their outcomes changed as a result of representation, and the randomizer prevents them from exercising their judgment in this manner. We confess to a degree of skepticism regarding this claim unless a provider invests substantial resources in case identification and evaluation prior to making a representation decision. But our skepticism is irrelevant because the premise that lawyers can discern who will benefit from representation is empirically testable via a double randomization scheme, as follows. Once determined to be eligible for a provider’s representation, the potential client is randomized to either a “lawyer-chooses” group or a “randomizer-chooses” group. If the randomization is to the “lawyer-chooses” group, then, obviously, the service provider exercises its judgment regarding whether the potential client “needs” representation in the sense that it believes the outcome in her case would be different (better) if representation is provided. If the randomization is to the “randomizer-chooses” group, then a second round of randomization decides whether the potential client is represented. Comparison of the success rates between the “lawyer-chooses” and the “randomizer-chooses” groups allows measurement of the lawyer’s ability to isolate cases that need representation. We are actively pursuing a study with this design.

A third objection to gold-standard evaluation, related but distinct to that just discussed, is that the randomization prevents the provider from exercising judgment, not on whether representation is likely to alter the outcome, but rather on the consequences of a win/loss.²⁶⁴ A housing example may clarify: a

²⁶³. Statement from Julie McCormack, Clinical Instructor, Wilmer Hale Legal Servs. Ctr. (late 2008). This quotation was subsequently verified by the speaker. E-mail from Julie McCormack to D. James Greiner (July 19, 2011, 8:35 AM EST) (on file with authors).

²⁶⁴. A related objection we have heard is that randomization requires a legal services provider to represent any and all persons requesting representation, regardless of the merits of case. This objection is simply wrong. Randomization is consistent with screening cases on any grounds, including merit, so long as the screening is done prior to randomization, and the screens applied leave a sufficient number of cases for study (that is, a number larger than what the service provider intends to represent).

randomizer might not allow a provider to select a tenant with three children over a tenant with none. The consequences of a favorable versus an unfavorable result may factor into a rational system of case selection once the effect (on the probability of a favorable result) of representation is well understood. But, per Parts I and II, that effect is not presently well understood. And the premise of an argument regarding consequences is that the provider's program as presently constituted increases the probability of a favorable result by a nontrivial amount. The results of the HLAB study, and the literature review we conducted in Part II, suggest that this premise deserves evaluation, particularly with respect to providers with intake systems that depend on client initiation and telephone communication.

A fourth objection to any quantitative case-by-case study, including a randomized one, is that it may be difficult in the litigation setting to define what constitutes a favorable outcome in each case (even putting aside the question of value identification discussed above). We note that this objection applies equally to nonrandomized attempts to quantify the effect of representation.²⁶⁵ All the issues discussed in this Part depend on context, but perhaps this one more so than the rest. There is probably a continuum of legal contexts in terms of difficulty of defining "success." We begin with the easy settings: in the context of government benefits, ordinarily the client/claimant wants the benefits, and wants them to begin sooner rather than later. The same is true of certain types of legal status (e.g., obtaining asylum, achieving a discharge in bankruptcy, maintaining parenthood). Tax cases are the same in reverse, that is, the client usually wants to minimize tax exposure. Thus, with respect to an enormous range of representation activity, a favorable outcome is easy to define.

Perhaps in the medium range of difficulty is housing/eviction. Here, one challenge may lie in determining exactly what the client wants, and the client's goals may change as a result of representation. Our conversations with housing litigators suggest, however, that in many cases a tenant or an eviction defendant has at least two measurable goals: (i) avoiding eviction, particularly a court-ordered eviction in the form of a writ of execution (the moniker in Massachusetts for an order empowering an official to remove the defendant and her possessions from a housing unit forcibly), and (ii) minimizing the amount of money the defendant pays to a landlord or unit owner. These may not be a housing client's only goals, and they may not be primary in all cases, and measurement of attainment of these goals may be complicated by the fact that clients may trade them against one another in different ways. But our

²⁶⁵. See *supra* note 154.

sense is that these two goals, both measurable, rank high enough in a sufficient number of clients' internal priority lists to allow meaningful quantitative evaluation.

Finally, perhaps the highest degree of difficulty in defining desirable outcomes is represented by the area of divorce and child custody. Even here, however, one study, after a thorough examination of the difficulties involved in defining a favorable outcome in the context of domestic litigation, measured success in terms of maintenance awards and the time needed to achieve a divorce.²⁶⁶ In addition, even if one cannot define "success" in all contexts, it is worth knowing whether litigants with representation end up with different outcomes than do the unrepresented, because of (not in spite of) potentially different value judgments about which outcomes would be better or worse.

A final objection to randomized evaluation among legal services providers is unique to law school clinics: the randomizer prevents a judgment as to the pedagogical value of a case. If pedagogy is the rationale, however, our view is that one must think carefully about whether facts such as no evidence of a substantial increase in the win probability and some harm (from delay, for example), if true, should be disclosed to the potential client before he or she agrees to be represented.²⁶⁷ And such facts cannot be disclosed if they are not known.

B. Maximizing Information

How much can randomized trials tell us regarding the effect of offers and actual use of representation? If we structure them well, randomized studies can tell us a great deal, and a great deal more than we have demonstrated in the present study. We discuss three groups of subjects that can be investigated with such designs: nonpecuniary client interests; service provider outreach, intake, and client-selection systems; and the accessibility and correctness of adjudicatory systems.

²⁶⁶. McMullen & Oswald, *supra* note 154, at 67-69.

²⁶⁷. One other objection deserves mention, one that we did not hear from legal service providers: perhaps providers need to choose their clients in order to obtain the mental state of dedication and belief in the rightness of their clients' cases necessary to litigate well. If so, then this is something the double randomization scheme described below will show, although it will be difficult to distinguish whether the greater win probability in the "lawyer-chooses" group is due to this "psyching up" effect or due to a lawyer's ability to choose cases in which win/loss will be affected by representation.

1. Nonpecuniary Interests

The present study primarily concerned representation effects on legal outcomes affecting the potential client's pecuniary interests (e.g., the probability of a win and the delay costs for claimants erroneously denied benefits) as well as costs borne by the system itself (e.g., the delay costs for claimants erroneously granted benefits). But representation in adjudicatory settings²⁶⁸ may serve other values; to what extent are effects on these other values measurable with a randomized experiment? The answer is that to a great extent, such effects are measurable.

We begin by identifying some of these other values. One value of representation might be in making adjudicatory systems run more smoothly (e.g., efficient case management). Indeed, some adjudicators have expressed a preference for represented as opposed to pro se litigants, contending that cases and hearings with the former group move more quickly and with less hassle.²⁶⁹ If so, then representation effects should be measurable via randomized studies that focus on the length, frequency, and characteristics of cases and hearings.

Representation may also promote nonpecuniary values that focus less on the adjudicatory system and more on the client. In this vein, representation may assure that each person subject to official decisionmaking and/or state coercive power is treated with dignity;²⁷⁰ it may promote a feeling on the part of the litigant that the process was fair and that her story was told, thereby increasing the litigant's willingness to accept the result of the adjudication (favorable or unfavorable).²⁷¹ It may educate the client as to her best interests or as to what is possible given legal and factual constraints, thereby adjusting the client's goals; and it may better the client's socioeconomic situation, even if

²⁶⁸ Our focus here is on the effect of representation in civil adjudications. Nevertheless, we believe that a randomized design can also inform nonadjudicative settings in similar ways. One might, for example, study the effect of credit and/or debt service advice on credit ratings, *Pleasence*, *supra* note 28, at 18-19, or the effect of end-of-life counseling on whether, after a certain period of time, a potential client has a valid will.

²⁶⁹ See, e.g., Editorial, *Need a Lawyer?*, *supra* note 10.

²⁷⁰ See, e.g., DEBORAH L. RHODE, *ACCESS TO JUSTICE* 9 (2004).

²⁷¹ Psychologist Tom Tyler has written extensively on the factors that affect, and the importance of, perceived fairness in official (including adjudicatory) systems. See, e.g., Tom R. Tyler, *The Role of Perceived Injustice in Defendants' Evaluations of Their Courtroom Experience*, 18 LAW & SOC'Y REV. 51 (1984); see also Nourit Zimmerman & Tom R. Tyler, *Between Access to Counsel and Access to Justice: A Psychological Perspective*, 37 FORDHAM URB. L.J. 473, 493-502 (2010) (collecting research and reporting the results of an independent analysis based on an observational study).

the legal outcome is the same, perhaps because the legal representative also acts as a coordinator of official and community resources (e.g., as a social worker).

Apart from the dignity value, which we cannot immediately see how to measure, we believe randomized trials can provide useful information on all these subjects. We propose the following: with respect to values that depend on a potential client's perception, ask the potential client.²⁷² Satisfaction and perception surveys are a mainstay of research outside the law, including in the medical arena. At present, we are actively pursuing a study that would survey potential client's perceptions of an adjudicatory process after hearings (if held) have occurred but before results have been announced. Our hope is that by asking clients to, say, score their perceptions of the fairness of the adjudicatory process on some scale, we can compare the perceptions of a treated group to a control group to determine the effect of an offer or actual use of representation on participant perceptions of the process. And pecuniary interests not directly connected to legal outcomes can also be measured with potential client interviews; such measurements may require tricky and long-term field operations, but they are possible.

2. Outreach, Intake, and Client Choice

Randomized designs can help to test the adequacy and efficiency of legal services programs' outreach, intake, and client-selection systems. As discussed in the response to Question 8, an important potential explanation for our HLAB findings is that the client base that HLAB's outreach and intake systems produced might not have needed the help. This result demonstrates the potential for randomized designs. Similarly, the double randomization scheme discussed in the previous Subsection would allow a rigorous evaluation of whether the outreach and intake schemes are producing a client base that actually needs representation, and whether a service provider is choosing those clients who will benefit from it.

²⁷². See, e.g., McMullen & Oswald, *supra* note 154, at 82. There may be reason to think that asking the potential clients could reveal surprising results. John M. Greacen, *Self Represented Litigants and Court and Legal Services Responses to Their Needs: What We Know*, CAL. ADMIN. OFF. OF THE CTS. 20 (2002), <http://www.courts.ca.gov/xbcr/partners/SRLwhatweknow.pdf> (last visited Apr. 3, 2012) (“The Van Nuys Legal Self Help Center evaluation concluded that litigants who had received Center services, who then lost their unlawful detainer cases, were more likely to perceive that they had not been prepared than litigants who had not visited the Center. In other words, visiting the Center appears to have increased a litigant’s expectations of his or her own ability to perform in court.” (citation omitted)).

3. System Accessibility and Accuracy

Third, randomized trials measuring the effect of representation can provide a valuable device to assess both the accessibility of an adjudicatory system and the system's accuracy. For both accessibility and accuracy, the randomized design can provide a baseline against which to measure system characteristics. Consider accessibility: one might define a system as pro-se-accessible if the outcomes of each adjudication in the system are the same regardless of whether a would-be litigant engages in a full attorney-client relationship with a competent²⁷³ attorney or appears pro se.²⁷⁴ If so, then one way to measure how well a system performs against this goal is via a randomized trial in which some litigants are offered full representation by competent attorneys and others are instead encouraged to use whatever resources the adjudicatory system provides to pro se litigants. Similarly, consider the "rightness" of results, *i.e.*, the adjudicatory system's accuracy. One problem is in discerning how to measure system accuracy without duplicating the entire system, an expensive proposition. One might decide that a reasonable proxy for the "right" outcome is the outcome reached in a proceeding in which the potential client is (or perhaps both sides are) represented by competent counsel. The randomized design just discussed would thus provide information on system accuracy as well.

If interest is in measuring either pro se accessibility or system accuracy, then a randomization scheme different from the single-provider-centered framework we employed in the present study might be preferable. As we have stated,²⁷⁵ an important limit of the HLAB study was that it concerned only those potential clients who contacted HLAB and made it through its intake system; although this was the correct set of persons to study to evaluate HLAB's program, there is reason to believe that this group of claimants may have been a distinct subset of those who participate in the Massachusetts DUA's first-level appeals system. To the extent that primary interest lies in the accessibility or accuracy of an adjudicatory system, however, a different design might be preferable, as follows. If one or more service providers are willing, and a sufficiently large set of litigants in an adjudicatory system are pro se, and

²⁷³. If the criminal context is any guide, attorney competence (and resourcing) will be a concern. See Mark C. Brown, Comment, *Establishing Rights Without Remedies? Achieving an Effective Civil Gideon by Avoiding a Civil Strickland*, 159 U. PA. L. REV. 893 (2011).

²⁷⁴. See Engler, *supra* note 6, at 197. The definition is not perfect; under it, for example, a system might be considered pro-se-accessible if all claimants lose. We suggest that in this case, however, the system has issues larger than pro se accessibility.

²⁷⁵. See text accompanying *supra* notes 149-150.

if contact information for members of this set is available shortly after they enter the system, then the group of pro se litigants can be randomized to either (i) an attempt to contact them to offer representation versus (ii) no such attempt. The benefit here is that the set of study participants subject to randomization is now broader than those who contacted a particular provider. The randomization determines whether a service provider “chases” the potential client, rather than whether representation is offered to a potential client who has already “chased” the service provider. This in turn allows inference to focus less on an evaluation of one provider’s practice and more on the system as a whole. Seron and her coauthors did something of this nature, taking advantage of a system-induced gathering of potential clients in one location to obtain partial information about system accessibility.²⁷⁶ One of the two housing studies in which we are participating has aspects of this design, and we have begun brainstorming about a study of this nature in the area of bankruptcy, where pro se petitioners are required to provide (hopefully valid) telephone numbers upon filing. We acknowledge that this design will not assess all aspects of the accessibility of the system, because only those who make it into the system will be subject to randomization.²⁷⁷ There could be other challenges, such as the rate at which those randomized to treatment accept the offer of representation. Nevertheless, limited but credible information is useful.

CONCLUSION

Until 2002, doctors routinely prescribed hormone replacement therapy (HRT) to healthy post-menopausal women as a protective measure, particularly to combat the potential for development of osteoporosis.²⁷⁸ In 2002, however, a large-scale randomized trial revealed that a certain kind of commonly used HRT increased the incidence of coronary heart disease, breast cancer, and stroke, while simultaneously decreasing the incidence of colorectal cancer and hip fractures.²⁷⁹ In terms of large-scale use, the hazards of HRT for healthy women with uteruses appeared to outweigh the benefits.²⁸⁰ More

²⁷⁶. Seron et al., *supra* note 20.

²⁷⁷. See *supra* notes 214-216 and accompanying text.

²⁷⁸. See Mayo Clinic Staff, *Hormone Therapy: Is It Right for You?*, MAYO CLINIC, <http://www.mayoclinic.com/health/hormone-therapy/WO00046> (last visited Apr. 3, 2012).

²⁷⁹. Writing Grp. for the Women’s Health Initiative Investigators, *Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women*, 288 JAMA 321 (2002).

²⁸⁰. Mayo Clinic Staff, *supra* note 278.

generally, the medical community has largely embraced the idea that to produce information and knowledge required to address the needs of its client base, it must make extensive use of the randomized trial.²⁸¹ Implicit in this extensive use is a judgment in the medical community that quantitative observational studies and qualitative day-to-day observations, while perhaps valuable in some situations, do not provide information sufficient to structure service delivery to clients. The dearth of randomized studies assessing the impact of outreach, intake, client selection, extent of service delivery (limited intervention versus full representation), nature of service delivery (non-attorney versus law student versus pro bono lawyer versus professional legal services staff attorney), and payment mechanisms, i.e., essentially all aspects of delivering legal assistance, suggests that we in law have thus far made a different judgment. The present study suggests that our judgment may be wrong.

We in the legal research community have particularly underappreciated and neglected to study the importance of service provider outreach, intake, and client selection systems. In almost any adjudicatory system it is probably the case that there will be some potential clients who will “lose,” or will realize outcomes that are unfavorable, even if they are represented. In many adjudicatory systems, it also seems likely that there will be some potential clients who will “win,” or who will realize outcomes that are favorable, even if they self-represent. And then there are cases in which some form of representation will make the difference between an unfavorable and a favorable outcome. The HLAB study provides reason to believe that outreach and intake systems may have a role in determining what mix of these hopeless, sure-win, or representation-makes-a-difference cases²⁸² a service provider will see, and the mix of cases a service provider sees will affect how much of a difference the provider’s services make.

Despite the best and continuing efforts of the civil *Gideon* and access to justice movements, and the need for greater funding for legal services provision, it may be time to face the fact that there will never be enough funding to provide a full attorney-client relationship with a competent lawyer to all low-income persons interacting with, or contemplating interaction with, the legal system. This is probably true even in areas of so-called “basic human

^{281.} See Abramowicz et al., *supra* note 4, at 931 n.3 (collecting references on the development and use of randomized trials in medicine).

^{282.} A fourth class of cases is possible: those in which a potential client will win if she proceeds pro se but will lose if she is represented. Such cases might be possible, for example, in adjudicatory systems in which a decisionmaker will proceed in an inquisitorial style if a litigant is pro se but will rely on a lawyer (and thus become passive) if one is present.

needs.”²⁸³ If these assertions turn out to approximate the truth, then triaging of legal services is unavoidable. We must learn to triage well. In an era of declining resources, and with respect to practices that focus at least in part on achieving outcomes for individuals (as opposed to those that focus exclusively on structural improvements), we should consider what we do not know, and that randomization provides a way to fill in canyons in our knowledge base.

APPENDIX: “POWER CALCULATIONS”

The primary result we report here is that there was no evidence that an offer of HLAB representation produced a statistically significant increase in the probability that an unemployment benefits claimant would prevail; any increase in the true probability of a victory due to the HLAB offer was unlikely to have been large, or our study would probably have detected it. The questions become: how “unlikely,” and how “large,” and what exactly is “probably”? Quantitative analysts tend to group questions of this sort under the label “power calculations.”²⁸⁴ Note that all of these questions are related. The larger the true increase in win probability due to an HLAB offer, the less likely it is that our study would have missed it.

Although we conducted a variety of analyses on these questions, we report two results here that capture the range of what we saw. First, the “Separate Regressions” technique shown in Figure 2 and briefly explained in the accompanying text and notes²⁸⁵ produced a 95% interval for the HLAB offer effect of (-.06, .09). The way to interpret this interval is that according to this

²⁸³. AM. BAR ASS’N ET AL., *supra* note 195, at 1.

²⁸⁴. Actually, the use of the term “power calculation” is wrong here, and it would also be wrong to use traditional power calculation methods to address the questions articulated above. See PIANTADOSI, *supra* note 25, at 175 (“Post hoc Power Calculations Are not Helpful.”). The reason is that our study is over, meaning we have data. Power calculations concern the ex ante probability that a statistical technique will conclude that an intervention (here, an offer of representation) produces no statistically significant effect when the true effect is of size X. Notice that in this statement, there was no reference to data; that is because, again, power calculations are ex ante, meaning that they are made before data has been collected. Because our study is over, we have data now. Accordingly, the right question to ask is as follows: how likely is it that we would have observed the data that we did had the true effect of an offer of HLAB representation been an increase of size X in the probability of a claimant victory? In technical terms, what is the probability of a Type II (false negative) error?

Because the abuse of the term “power calculation” is so common among quantitative analysts, we also use the wrong terminology here, with the clarifications that when we use the term, we intend it to mean the probability of erroneously reporting a null result had the true increase in win probability due to the HLAB offer been of size X.

²⁸⁵. See *supra* note 116.

technique it is 95% likely that the true “increase” (a negative value represents a decrease) in win probability due to the HLAB offer is between $-.06$ and $.09$, with the most likely values being near the middle of this interval (in the $.01$ or $.02$ range) and values near the ends of the interval becoming less likely. It is only 2.5% likely that the increase is larger than $.09$. Thus, we conclude that under this technique, it is unlikely that the true increase in win probability due to the HLAB offer is greater than $.09$.

Second, recall that we observed a (weighted) win rate in the HLAB offer group of $.764$ versus a win rate of the control group win rate of $.717$, for a difference of $.047$ (rounded). We conducted simulation studies as follows: we held the control group win rate constant at $.717$ and assumed that the “true” HLAB offer group win rate was $.05$ higher, i.e., $.767$. We simulated 10000 datasets, and for each simulated dataset, calculated (weighted) win rates in the HLAB offer group and the control group. We observed the proportion of this set of 10000 that had an absolute difference in win rates of $.047$ or less. The idea here is to see, under the assumption that the true increase in win rate due to the HLAB offer is $.05$, how likely it would be that we would observe treated and control group win rates of within $.047$ of each other. Unsurprisingly, under the assumption that the true increase in win rate due to the HLAB offer was $.05$, a fair number (40% or so) of the simulated datasets showed treated and control group win rates within $.047$ of each other. Under this technique, then, we conclude that a true increase in the win rate of $.05$ is consistent with the data we actually observed in our study.²⁸⁶

Next, still holding the control group win rate constant at $.717$, but we assumed the treated group win rate was $.792$, i.e., $.075$ higher. We simulated 10000 more datasets, and for each simulated dataset, calculated the (weighted) win rates in the HLAB offer group and the control group. We observed the proportion of this set of 10000 that had an absolute difference in win rates of $.047$ or less. Again, the idea here is to see, under the assumption that the true increase in win rates due to the HLAB offer is $.075$, how likely it would be that we would observe treated and control group win rates of within $.047$ of each other. A smaller number (30% or so), of the simulated datasets showed treated and control group win rates within $.047$ of each other. Under this technique, then we conclude that the true increase in win rate of $.075$ is also consistent, but less so than a $.05$ increase, with the data we actually observed in our study.

286. By the same token, a true “increase” of zero or some amount less than zero would also be consistent with the data we observed. That is what we mean when we say, at the end of this appendix, that the technique we are describing here is “symmetric.”

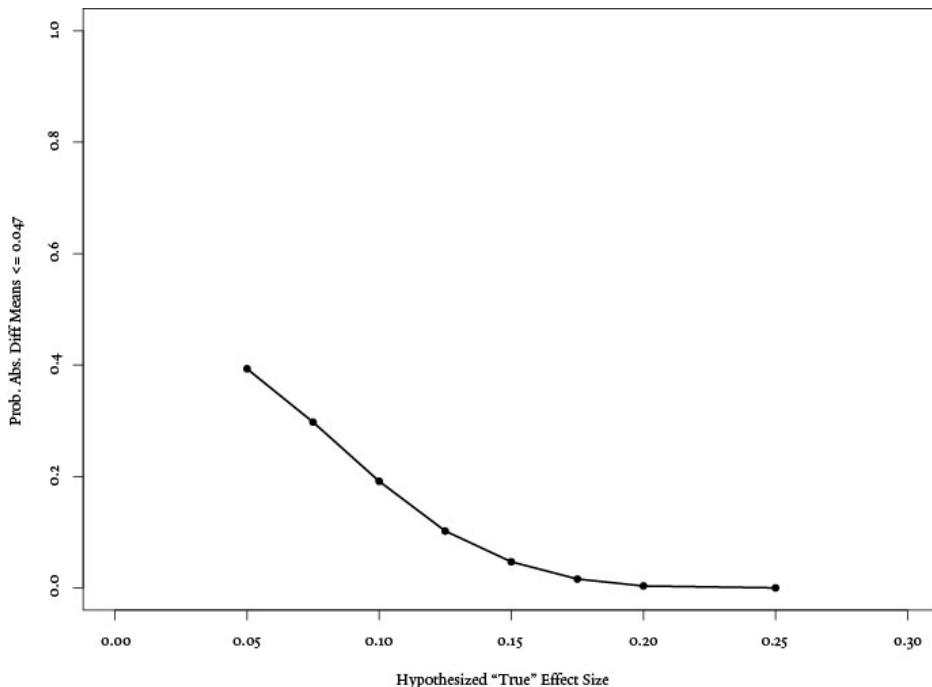
Next, still holding the control group win rate constant at .717, we assume that the HLAB offer group win rate was .817, or .10 higher. We repeated the simulation procedure, and found that the treated and control group win rates are within .047 of each of each other in about 20% of the simulated datasets. Thus, under this technique, the odds are about four to one against an increase of .10 or larger, given the data we observed.

We repeat the simulation technique assuming an HLAB offer group win rate of .842, or .125 higher than the control group win rate of .717. Now, only about 10% of the simulated datasets produced HLAB offer and control group win rates within .047 of each other, so the odds are nine to one against an increase of .125, given the data we observed. We repeat the simulation technique for HLAB offer group win rate of .867, or .15 higher than the control win rate; of .892, or .175 higher than the control group win rate, etc., all the way up to .967, or .25 higher than the control group win rate.

One can see the results of all of these simulations in Figure 9. On the x-axis is the assumed increase in win rates of the various simulations, starting at .05, then .075, then .10, then .125, etc. On the y-axis is the fraction of simulated datasets in which the HLAB offer and control group win rates were within .047 of each other. The 40% explained in the paragraph above becomes a .40 on this graph, the 30% becomes a .30 in the graph, etc. From this graph, one can see that the odds are about four to one against an increase of .10 or larger, and about 19 to one against an increase of .15 or larger.

Figure 9.

PROBABILITY THAT THE “TRUE” INCREASE WIN RATE DUE TO HLAB OFFER WAS X FOR VARIOUS VALUES OF X



Two final notes: First, there is a difference in the two techniques explained immediately above. For the “Separate Regressions” technique, an increase in the win probability due to the HLAB offer of .09 or larger was only 2.5% likely. In the simulation technique, an increase in the win probability due to the HLAB offer of .10 or larger was (roughly speaking) about 20% likely. Why this difference? The reason is that the “Separate Regression” technique uses some modeling (readers may be familiar with the term “logistic regression,” and its Bayesian analog), and modeling involves making certain assumptions about the nature of the data. If those assumptions are not reasonable approximations of the nature of the data, the results are untrustworthy. We have labeled these assumptions “relatively innocuous” in the text because they are far, far more believable than are the assumptions made in the studies we critique in Part II of this Article. But these assumptions are still there.²⁸⁷ The simulation technique

²⁸⁷. The most important assumption is that, for observations in the treated group, the relationship between observed background variables and the outcome can be reasonably

we described does not use modeling and thus does not depend on the assumptions required for the “Separate Regression” technique. The simulation technique still depends on assumptions; all quantitative techniques do. But these assumptions are even easier to credit than those of the Separate Regression technique. Thus, what we see here is a tradeoff, ubiquitous in quantitative analysis, between the assumptions one is willing to credit and the precision of the results.

Our second final note: we suspect that some practitioners and other readers might feel the urge to credit figures on the upper range of the Separate Regression techniques interval of $(-.06, .09)$, meaning to believe that the HLAB offer actually causes an increase of $.08$ or $.09$, or perhaps even higher. Alternatively, upon hearing that the Separate Regression technique involves some modeling assumptions, some readers might abandon it in favor of the simulation studies we presented in this appendix, and argue that increases larger than $.10$ are within some threshold of plausibility. We simply remind such readers that the Separate Regressions interval—and indeed all of the results we discuss in this Appendix—are largely symmetric. In other words, as far as the data and the Separate Regression technique show, it is (essentially) just as likely that the HLAB offer caused an “increase” in the win rate of $-.06$ as it is that it caused a $.09$ increase, with the most likely results being in the $.01$ to $.02$ range. Similarly, readers inclined to dislike the results of the Separate Regressions technique because of the assumptions it requires should positively disdain the results of the non-randomized studies we discuss in Part II.

approximated by a particular kind of mathematical equation relating an increase (or decrease) in a background variable to an increase (or decrease) in the outcome. The same is true for the observations in the control group, although the two approximations (treated and control) can and do differ from one another.