The New Legal Empiricism & Its Application to Access-to-Justice Inquiries

D. James Greiner

Abstract: The United States legal profession routinely deals with evidence in and out of courtrooms, but the profession is not evidence-based in a scientific sense. Lawyers, judges, and court administrators make decisions determining the lives of individuals and families by relying on gut intuition and instinct, not on rigorous evidence. Achieving access to justice requires employing a new legal empiricism. It starts with sharply defined research questions that are truly empirical. Disinterested investigators deploy established techniques chosen to fit the nature of those research questions, following established rules of research ethics and research integrity. New legal empiricists will follow the evidence where it leads, even when that is to unpopular conclusions challenging conventional legal thinking and practice.

The U.S. legal profession routinely deals with evidence in and out of courtrooms, but the profession is not evidence-based in the scientific sense. Lawyers, judges, and court administrators, as they work in the U.S. justice system, make decisions that determine the lives of individuals and families. But they overwhelmingly rely on gut intuition and instinct, not on rigorous evidence. The choice to eschew evidence matters. The profession is asking the wrong questions about topics fundamental to our system of government, and “if they can get you asking the wrong questions, they don’t have to worry about answers.”

Here are two examples:

Judges decide which arrestees to incarcerate versus who to release pending disposition of criminal cases in so-called bail hearings. They frequently focus on information available only from an interview of the arrestee, such as how long they have lived in the community, their employment situation, and the presence of any family in the area. But arrestee interviews raise concerns of self-incrimination,
and are often unavailable because arrestees cannot be or do not choose to be interviewed. Scientific evidence suggests that if complete or nearly complete information on an arrestee’s past criminal history (if any) is available – information from existing administrative records – an interview with the arrestee is unnecessary and provides little additional relevant information.²

In some jurisdictions, most low-income survivors of domestic violence petitioning for civil protection orders will not receive full lawyer representation. Over-subscribed legal-services providers often decide to provide a few survivors/petitioners full representation and to provide something less (such as self-help materials and/or an explanatory telephone call) to most. But in making these triage decisions, lawyers often ask themselves which cases they can win (meaning obtain a civil protection order), not in which cases their representation is likely to make a difference.³ Winning a case and making a difference in a case are two different things. There is little evidence about how to identify cases in which full representation makes a difference.

These examples illustrate that the questions the legal profession chooses to ask about services it provides to poor and low-income people have substantial consequences about who spends time incarcerated and who obtains legal protection from abusers. In other words, these choices have immediate consequences for real people.

The new legal empiricism, which exists in pockets in the academy but only rarely outside of it, could transform the U.S. legal profession into an evidence-based field. As the examples above indicate, rationalizing those areas of law might reduce crime and incarceration and permit more effective triage decisions by legal-services providers. In other words, fewer people could go to jail, and those who do go might spend less time there, with no increase in crime rates or threat to the administration of justice. More survivors of domestic violence could have judicial protection orders. None of this would require additional resources.

As of the 1930s, U.S. medicine was not yet a science either. Physicians relied almost exclusively on gut instinct and intuition, informed by dubious claims from drug companies, to decide which drugs to provide patients. Over the next forty years, medicine transformed itself into a field in which physicians made decisions about treatment based on evidence from randomized controlled trials (RCTs).⁴ The U.S. population lives longer and healthier lives in part because of this transformation. The emergence of the new legal empiricism, one hopes, indicates that eighty-plus years later, law may begin to follow medicine’s example.

What is the new legal empiricism? The “new” here modifies “legal,” not “empiricism.” New legal empiricism is simply strong empiricism, as developed and implemented over the past decades in fields outside of law, now finally applied to law. It is “new” in the sense that law has followed neither medicine nor other social sciences (sociology, economics, political science, psychology) in demanding that strong empiricism become the standard for investigations that researchers conduct and the basis for decisions that those in the field make.

There are many kinds of empiricism, and empirical projects can further a variety of goals. Historians engage in empiricism, frequently based on archival records and oral interviews with the elderly, and have their own standards of inference. Literary scholars sometimes investigate the lives of the authors whose work they interpret. The new legal em-
Empiricism involves investigations into how the current legal system works, and how to change the world for the better, however “better” is defined.

The first step in strong empiricism is sharply defining the questions to address. The legal profession frequently fails here, and when it does, nothing that follows matters. Relevant questions emerge from conversation among empiricists, practitioners in a field, and their customers or clients. In the “bail” release example above, a judge or a legislator might consult an empiricist for help in easing jail overcrowding or ameliorating racial disparities in incarceration. The empiricist might examine jail rosters to discover that a substantial portion of jail residents are arrestees awaiting trial on the charges leveled against them, examine whether some racial groups are over- or under-represented, and ask how decisions are made about whom to incarcerate before trial.

Those questions might lead to the following information: In practice, after a law enforcement officer arrests someone, the arrestee is taken to a jail and held there while a prosecutor files charges against him. In most U.S. jurisdictions, the arrestee is brought before a judge (or the equivalent) who decides whether to release the arrestee as he awaits trial on those charges, and if so, what conditions the arrestee will have to meet to secure release. As a matter of broad policy, judges could release all arrestees, but doing so carries risks that some arrestees will fail to return for hearings or will commit new crimes while on release, either of which undermines the administration of the case and the public’s confidence in the judicial system. Judges could release no arrestees, meaning incarcerate them all, but that carries fiscal and human costs. Judges in all U.S. jurisdictions, following dictates of statutes and court decisions that vary in the degree of discretion delegated to judges and in the factors judges can or must consider, release some arrestees and not others, imposing conditions on all arrestees released. Bail is one well-known condition: the arrestee must arrange for the deposit of a certain amount of money, set by a judge, into a court account, with the money returned at the end of the case if the arrestee meets the conditions of release, but subject to forfeit to the court if the arrestee misbehaves.

My research suggests that judges making predisposition release decisions are attempting to minimize three rates: 1) the rate at which arrestees fail to appear at court hearings in their criminal cases; 2) the rate at which arrestees commit crimes (especially violent crimes) between arrest and case disposition; and 3) the rate at which arrestees are incarcerated. Judges and legislators are also often concerned about potentially unlawful racial disparities in the population of arrestees incarcerated at the bail stage and about whether criminal records or other risk factors justify racial disparities observed.

With that in mind, the empiricist and the judge or legislator might agree that the broad question is how to rank groups of arrestees according to the objective, unbiased risk that they will commit crimes or fail to appear if released. High-risk arrestees could be incarcerated, the remainder released. The empiricist might break this broader question into two: What observable factors classify arrestees according to their risk of committing crimes or failing to appear, without introducing racial bias into the bail decision? And does providing some kind of compilation or summary of those factors, for example, in the form of a score that classifies risk, result in reductions of some or all of the rates listed above?

To an empiricist, these last questions are of different types. The first, about
factors that classify (or predict) arrestee risk of misbehavior, gets at the way the world is. It is descriptive. The second, about whether providing information on risk to judges improves criminal justice outcomes, is causal. The causal question gets at whether doing something (providing risk information to judges) versus doing something else (not providing risk information to judges) alters a set of outcomes (the three rates above) in a desirable way.

Once the empirical questions to be explored have been identified, the empiricist must determine how to answer them. There will always be several options: qualitative techniques, such as structured interviews of potentially knowledgeable people; focus groups, in which an empiricist assembles groups of potentially knowledgeable people, provides them with open-ended questions, and elicits information from the resulting discussion; structured observation of relevant events; or reviewing relevant documents to look for patterns. Still other options include: quantitative techniques, such as surveys of randomly selected individuals or cases or judges; predictive models, which explore whether and how well precursor variables predict the value of ultimate variables; and randomized controlled trials, in which the empiricist randomly assigns cases, people, judges, or units of some kind to one condition or another and then measures which condition produces a more desirable set of outcomes. A practitioner of the new legal empiricism, like a practitioner of rigorous empiricism outside the legal context, chooses the method appropriate for the questions to be addressed.

To illustrate this second step, return to the two smaller empirical questions identified above: What factors accurately predict arrestees’ risk of misbehavior if released on bail? And does providing information to judges about those factors result in better reduction of the rates of failure to appear at hearings, new crimes committed between arrest and case disposition, and incarceration?

The empiricist will choose qualitative or quantitative techniques to address these questions. Which ones? In many situations, qualitative techniques are either superior to other options or an integral part of an overall research plan. For these two questions, however, I would look primarily to quantitative techniques. One reason is the objectivity of the information likely to be obtained by interviewing participants in the bail hearings. The goal of the empirical project is to improve the judges’ decision-making in these hearings, so there is reason to be cautious about relying exclusively on talking to people and observing the settings that are the target of improvement efforts.

Instead, researchers exploring this question have compiled data potentially available to judges making arrestee release decisions, such as information about charges, arrestee criminal history, their history of appearing or not at past hearings, their ties to the community, their race or ethnicity, and other factors. Researchers have connected this information to information about key outcomes, such as judges’ release decisions, arrestees’ failure to appear, and arrestees’ commission of new crimes. Applying statistical techniques to these data, researchers have created scoring systems or algorithms relating observable information about arrestees to release decisions, failures to appear, and new crimes. The scoring systems or algorithms are known as “risk assessment” instruments or scores. They aim to use variables that are observable at release hearings and are racially unbiased, along with a set of weights derived from the data analysis exercise just described, to classify arrestees according
to the risk that they will fail to appear or commit new crimes if released. This is the new legal empiricism at work.

The new legal empiricism is about more than selecting and implementing the right research techniques to obtain and analyze the right data; it is also about creating research norms that protect the credibility of the research and researcher.

New legal empiricism practitioners must follow general norms of social science research that have emerged to safeguard research integrity and the appearance of research integrity. For example, empiricists should not engage in investigations in which they have a financial or other interest that might affect their impartiality. Independent researchers ordinarily produce more credible results. Researchers must also follow the evidence. If the strong empirical techniques suggest or lead to truths that are unpopular among certain constituencies, or even normatively creepy, the empiricist must report those results.

New legal empiricism practitioners should stay cognizant of the limits of research techniques, and remember to include a discussion of those limits in the publications they produce. Where possible, empiricists specify their methods and research goals before they start doing any research, for example, by “registering” a proposed study on one of several websites that exist for this purpose. Where possible, consistent with confidentiality- and use-agreement limits frequently grounded in concerns of ethics and privacy, empiricists make their data and statistical code publicly available to allow replication of results.

New legal empiricism practitioners should also separate, to the extent possible, the facts they investigate and generate from value judgments required in any policy decision. Returning to the example of risk assessment instruments, recall that the aim was to minimize simultaneously the rates at which arrestees were incarcerated, failed to appear at court hearings, and committed new (and especially violent) crimes. Minimizing these rates might involve tradeoffs: more incarceration might mean few failures to appear and less crime. How to weigh these rates against one another in setting policy in this area is a value judgment, not an empirical question. Though 100 percent separation between empiricism and values is neither possible nor desirable, a fair amount of distance between them is achievable and essential to the empiricist’s credibility.

Returning to the running example of risk assessment instruments for bail hearings, assume that the initial step of constructing a risk assessment instrument is complete, so that an empiricist has identified a set of variables observable after arrest and not too closely associated with race that, weighted in a specified way, appear to classify arrestees according to risk of misbehavior. Assume that the researcher followed best practices with respect to research integrity.

The next question is whether sharing the risk assessment scores with judges will change their behavior, by facilitating release decisions that minimize racial imbalances along with rates of incarceration, failure to appear, and/or new criminal activity. This question is one of program effectiveness, suggesting that the backbone of research should be one or more RCTs, because answering the research question requires being able to compare judicial decisions (and results of those decisions) when the risk assessment instrument is available, and when it is not, to see which produces better outcomes. Only RCTs, with their random allocation of judicial decisions to a risk-assessment condition versus a
no-risk-assessment condition, can assure (up to statistical uncertainty) that differences observed in the outcomes are due to the difference in conditions and not some alternative factor.

In this example, without an RCT, a judge might choose to use a risk assessment only in cases the judge considers close or difficult. If so, and if defendants misbehave at higher rates in the cases the judge considers close or difficult, a comparison of misbehavior rates in cases with risk assessment scores to misbehavior rates in cases without such scores will show more misbehavior when risk assessment scores are considered, making it look like the risk assessment leads to worse outcomes. An RCT would prevent this kind of misrepresentation.

To answer a research question such as this, RCTs, though necessary, are not always sufficient. RCTs will disclose whether the availability of the risk assessment improves criminal justice outcomes but not why it does or does not. To find out why, a practitioner of the new legal empiricism should attempt to supplement the RCTs with other quantitative techniques, such as comparisons of rates of release, failures to appear, new criminal activity, and racial statistics before and after the risk assessment was adopted. A new legal empiricism practitioner should also deploy court observations, interviews, and other qualitative techniques. These techniques will generate information about how risk assessment scores work with judges’ decision-making, and thus why scores do or do not help. That information, in turn, will allow a researcher to theorize about when scores might work in other court systems.

Now imagine that several researchers were at work on this question and they all found the same thing: When they compared release patterns, failures to appear, new criminal activity, and racial balances before and after the implementation of the risk assessment program, they discovered improvements in some or all of these measures roughly coinciding with the implementation of a risk assessment score program. At the same time, however, RCTs showed no effect of the use of the scores on outcomes. Interviews with judges, prosecutors, defense attorneys, local government officials, and court administrators showed that communities ordinarily implement risk assessment scores as part of an overall package of criminal justice reforms. Such reform packages include reduction of time spent in jail before bail hearings, faster processing of information from law enforcement to prosecutors, and/or creation of programs providing differing levels of monitoring (ankle bracelets, drug testing, automated call-in services) that facilitate predisposition release. In the face of 1) a favorable before-after comparison, but 2) nothing statistically significant on the RCT comparison that evaluated the risk assessment instrument exclusively, a practitioner of the new legal empiricism might infer that the elements of the reform packages adopted contemporaneously with the risk assessment are probably responsible for the favorable changes, not the risk assessment itself.

The new legal empiricism means beginning with a specific set of questions. The questions to be investigated are not value judgments masquerading as factual inquiries; they are empirical. The investigation proceeds with an impartial investigator’s deployment of established techniques chosen to fit the nature of the research questions. The investigator implements these techniques in a manner that protects the integrity of the investigation and her own neutrality. Helpful practices include prespecification, transparency, making data and coding available for
The New Legal Empiricism & Its Application to Access-to-Justice Inquiries

replication, and defining variables clearly. The investigator follows the evidence where it leads, including to unpopular conclusions, and she is careful to explain the limits of the techniques she deploys.

All of this is new only to law.

The new legal empiricism has a great deal to offer to the field of access to justice. The field misses out when researchers and their partners in the U.S. legal profession choose not to render the results of their research credible.

Consider a question at the heart of access to justice: how much difference, if any, do different levels of legal assistance make? This broad question can be addressed by identifying a particular level of service – for example, offering self-help materials to individuals faced with a certain kind of dispute – and identifying an alternative level of service – for example, offering attorney-client relationships to individuals faced with that same kind of dispute. Having identified differing service levels, a researcher attempts to ascertain how much each level costs as well as the outcome variables the services are designed to affect. Examples of typical outcome variables include obtaining a favorable adjudicatory result, addressing the underlying socioeconomic issue that led to the adjudication, assuring that each litigant feels that she is treated with dignity and respect during the adjudication, and reminding government officials (the judge, the court staff) that human beings are involved in each of the cases they adjudicate.

In the past, many researchers would have proceeded by identifying a set of cases in which a litigant experienced one service level (say, self-help materials) and a set of cases in which a litigant experienced a different service level (say, full attorney representation), comparing the outcomes of litigants in each set, and then attributing any observed disparities in those outcomes to the difference in service levels. Sometimes, such researchers could measure observable background variables, such as race or gender or some measure of case complexity, and attempt to use statistical models to “control for” those background variables.

By contrast, a researcher working in the new legal empiricism proceeds by deploying an RCT, supplemented (ideally) by qualitative techniques, to understand how the adjudicatory system at issue functions. Only an RCT can assure, up to statistical uncertainty, that any differences observed on the outcome variables experienced by litigants with one service level (self-help materials) versus those experienced in another service level (full representation) are due really to the difference in service level offered as opposed to, say, differences in the individuals’ unobservable characteristics, such as motivation level, articulateness, or case characteristics.

This difference in methodology goes to the heart of what the new legal empiricism is and what it can offer. Practicing lawyers say that litigants who self-select into receiving self-help materials and litigants who obtain full representation do not have the same distribution of motivation, or articulateness, or case characteristics. Specifically, people who successfully search for a legal-services provider, find their way through its intake system, and persuade it to provide full representation – all sufficiently early in a matter for the attorney to provide real services – are likely more motivated, more articulate, and have cases that have different characteristics than those who do not. A new legal empiricism researcher’s use of an RCT-backboned study provides credible measurements of how much difference the disparate service levels make, unconfounded by the effects of the disparate background variables. The new legal
Empiricism offers credible evidence of the effectiveness, and cost effectiveness, of disparate legal-service levels. Other research offers little or nothing credible.

The difference in choice of method is real. A few years ago, statistician Cassandra Wolos Pattanayk and I compiled over one hundred studies comparing the effectiveness, and cost effectiveness, of legal-services providers’ disparate levels of service. This quantity of studies addressing the same research question demonstrates both the question’s importance and the vastness of the effort dedicated to answering it. But of the more than one hundred studies, only about seven (depending on how one counts) deployed RCT-backbone methodology. These seven studies reach seemingly contradictory conclusions, with some suggesting the politically unpopular conclusion that for some legal settings, and some sets of clients, higher (and more expensive) levels of legal services make little difference vis-à-vis lower, less expensive levels. The other studies are, as detailed above, not credible. The potential of, and need for, the new legal empiricism is evident.

Nonetheless, legal-services providers continue to produce (and, apparently, rely on) studies that do not deserve credence. Many of these are implemented by or commissioned by the programs themselves. They almost invariably reach laudatory conclusions. By way of example, a recent study of telephone advice programs in one state concluded that the evaluated programs “are achieving the primary goal of telephone-based legal assistance, which is to make legal assistance accessible to every eligible person . . . without sacrificing service quality and effectiveness in the process.” The researchers arrived at this conclusion by speaking to former clients they could reach and requiring the evaluated programs to conduct a “self-assessment.” The evaluators spoke only to individuals who received telephone services (and who they could reach), so there was no comparison group of individuals who did not receive telephone services. There could be no comparison necessary to conduct any kind of evaluation, much less the randomized evaluation needed for credibility. Here is the essence of self-evaluation: I am the greatest law professor on Earth. Just ask me. This is not the new legal empiricism.

To take another example from access to justice, consider the following narrative: Low-income individuals frequently encounter civil legal problems. Most would like to obtain attorneys to help them resolve those problems. The primary reason they do not consult or retain attorneys is that they cannot afford lawyers’ fees. To further access to justice, then, governments, philanthropists, and others should pump money into existing legal-services programs to fill the “justice gap.”

The narrative above has two basic points: low-income individuals frequently encounter civil legal problems, and most would like to obtain attorneys to help resolve those problems but do not do so because the cost is too high. Is the first point true? Is the second?

For many in the U.S. legal profession, the answers to these two questions are too obvious to require research. As suggested by the phrase “justice gap,” it simply must be the case that low-income individuals and families desire but cannot get legal services, specifically traditional attorney-client relationships. But a hypothetical new legal empiricism researcher would not accept the idea that truths this important are too obvious to investigate.

An editor of this volume, Rebecca Sandefur, did not accept that idea. She sought to find out whether low-income individuals encountered legal problems and, if so, whether they wanted attorney assistance, by asking low-income individuals.
Deploying a well-executed set of focus groups, she found that the answers to the two questions identified above were complex and nuanced. The evidence suggests that low-income individuals frequently encounter legal problems, but even when they recognize those problems as legal (not always), they generally prefer to involve neither lawyers nor courts. Cost is often not the primary reason for their reluctance to turn to formal law.8

Neither Sandefur’s research nor the RCTs on the effectiveness of different levels of legal services support the idea that legal services are worthless or that funding for legal services should be cut. This rigorous research does suggest, however, that standard narratives that exist in the U.S. legal profession are distorted in ways that matter.

The previous discussion suggests that the new legal empiricism has already provided much to challenge assumptions common in the U.S. legal profession. The approach could also offer much in the way of guidance for the profession’s policy-makers, regulators, funders, reformers, and revolutionaries. Challenges to accepted truths are helpful, but the approach should also deliver constructive ideas. Early indicators are that it will be able to do so. Here are some questions about which, with adequate funding and with the political will among the members of the legal profession, a new legal empiricism could provide useful guidance:

- What would be the effects of partial deregulation of the U.S. legal profession?
- Can limited licensed legal technicians, or other kinds of nonlawyer legal professionals, provide effective services at a cost accessible to low-income individuals?9
- Can online legal-service providers like LegalZoom, or free online legal-service providers, or nonprofits that provide a hybrid of online and traditional lawyer services provide an effective way for low-income individuals to benefit from the justice system?10
- Would changing legal ownership rules allowing lawyers, or unsupervised nonlawyer legal professionals, to work as salaried employees of corporations with convenient locations (think paralegals in offices at Walmart) improve access for low-income individuals and families?11
- What would be the effects of moving certain disputes to online or app-based adjudication?12
- Can algorithms and scoring systems, administered by human beings or computers implementing artificial intelligence programs, improve the functioning of courts, legal-services offices, court administrators, and other key actors within the justice system, as is already occurring in the medical profession?13
- Are there effective ways to divert individuals accused of crimes away from the traditional criminal law system?14

It is impossible to overstate the importance of these questions to the modern justice system. The new legal empiricism has much to offer.

If the new legal empiricism has already exploded some of the myths that previously masqueraded as truths, and if it has much to offer for the future, why has the U.S. legal profession yet to embrace it, and what can be done about the situation?

Because medicine and other disciplines have incorporated rigorous empiricism into their understandings of what counts as true, it cannot be that the new legal empiricism is inimical to the judgment-based reasoning that the legal profession offers in solving legal problems. Nor can
it be that rigorous empiricism is inconsistent with professional ethics, legal or otherwise. Again, the medical example shows as much. Rigorous empiricism can coexist with or alongside professional judgment.\(^{15}\)

In the 1930s, when medicine began to turn to rigorous empiricism (particularly the RCT), medicine was less a science than an individualized craft, especially as it was practiced outside major cities and teaching/research facilities.\(^{16}\) And in this period, published papers in both medicine and law began to make use of randomization to conduct empirically rigorous studies, suggesting that the intellectual foundation for transformation of the legal system into an evidence-based field was present then, just as it was in medicine.\(^{17}\)

In prior work, Andrea Matthews and I speculated that perhaps lawyers (and thus judges, who in the United States are ordinarily former lawyers) resist rigorous empiricism because 1) they are trained to pursue goals that clients provide and their thought processes are therefore fundamentally instrumental (in the service of advocacy) as opposed to analytical; and 2) there may be social value to having lawyers appear certain when they argue and to having judges appear certain when they make decisions, even when there is little basis for that certainty, and to appear certain, lawyers and judges must convince themselves that they are.\(^{18}\) The first observation, if true, might make it hard for lawyers and judges to embrace the new legal empiricism because they are trained more to argue than to analyze, and thus instinctively seek persuasion rather than truth. The second observation, if true, might make it hard for lawyers and judges to embrace this approach because certainty inhibits a desire for rigorous investigation.

Neither observation/argument is close to bulletproof. The duties of corporate general counsels are less about advocacy and more about strategic decision-making and policy-making than those of, say, courtroom litigators. Yet we see little if any evidence of rigorous empiricism in corporate counsel offices, despite the supposed existence of market forces that might put a premium on using a new legal empiricism to discover money-saving truths. When Matthews and I provided our speculation, we stated that these two observations were likely insufficient to explain fully the U.S. legal profession’s resistance to evidence-based thinking.

Medicine’s partial turn to evidence-based thinking reflected leadership shown by particular members of the medical profession; these leaders worked primarily in urban teaching centers and had strong connections to federal agencies.\(^{19}\) The legal profession needs leaders who can transform its thinking about what counts as useful knowledge. Law schools should create environments that introduce students to the issues discussed here. Legal academics aim to prepare the next generation of leaders in the U.S. legal profession to lead. A crucial element of that training is to teach about the need for, and about how to work with researchers to expand, the new legal empiricism.
ENDNOTES


3. Figures to support the statements about oversubscription are on file with author, from a survey of 2014 cases in an eight-county area around Akron, Ohio. The statements regarding lawyers’ focus on whether they can win are the result of conversations with legal aid attorneys around the country, including in two Ohio locations.


18. Greiner and Matthews, “Randomized Control Trials in the United States Legal Profession.”

19. Marks, *The Progress of Experiment*.