Public Health Law Monitoring and Evaluation in a Big Data Future

SCOTT BURRIS[*]

I. INTRODUCTION

Law is important to public health. It provides government health agencies with their jurisdiction and regulatory authority.[1] Laws and regulations are routinely used in the name of health to regulate behavior and foster safer environments.[2] More fundamentally, law's influence in shaping everyday life and the socioeconomic and physical environments in which it unfolds has a powerful impact on both the level and distribution of health.[3] Despite law's importance, and despite the strong orientation toward scientific evaluation in public health, the study of the impact of laws and legal practices on health ("public health law research") has been uneven. While research of the highest quality has been sustained in a few areas like auto safety and

[1] LAWRENCE O. GOSTIN, PUBLIC HEALTH LAW: POWER, DUTY, RESTRAINT (University of California Press, 2d ed. 2008).

[2] Scott Burris & Evan Anderson, *Legal Regulation of Health-Related Behavior: A Half Century of Public Health Law Research*, 9 ANN. REV. OF L. & SOC. SCI. 95 (2013).

[3] Scott Burris, et al., *Integrating Law and Social Epidemiology*, 30 J. L. MED. & ETHICS 510 (2002); Scott Burris, *From Health Care Law to the Social Determinants of Health: A Public Health Law Research Perspective*, 159 U. PA. L. REV. 1649 (2011).

tobacco control, it has been infrequent or truncated in others, like gun control and HIV/AIDS. The research that has been supported is almost entirely aimed at evaluating deliberate legal interventions. Epidemiological research on unintended health effects of non-health laws has been almost entirely neglected.[4] Overall, the national investment in rigorously separating the laws that help from the laws that hurt has been insufficient in comparison to its importance to the nation's health.

The Public Health Law Research Program was established at Temple Law by the Robert Wood Johnson Foundation in 2009.[5] Its mission is to fund and support scientific research documenting the impact of law on public health. From the perspective of this mission, I will discuss two complementary ways that big data fits into our thoughts and work in public health law research.   The first is straightforward: our single biggest challenge in evaluating the health impact of laws and legal practices is the lack of data measuring intermediate and ultimate outcomes. The second is, in a big data discussion, perhaps paradoxical: old-fashioned scientific methods of human coding statutes and regulations may be the most feasible short- and intermediate-term means of getting law into the big data mix, and may have implications for legal practice and research that transcend the field of health.

II. BIG DATA FOR LEGAL EVALUATION

Traditional science has often done very well in evaluating the impact of interventional health laws.[6] The classic example is traffic safety, where over many years there was the necessary thought and funding invested in research to produce the human and data

---

[4] *See* S. Burris, et al., *Making the Case for Laws that Improve Health: a Framework for Public Health Law R*esearch, 88 MILBANK Q. 169 (2010) (defining public health law research); Burris and Anderson, *supra* note 2 (reviewing five decades of public health law research).

[5] Scott Burris & Evan Anderson, *Making the Case for Laws that Improve Health: The Work of the Public Health Law Research National Program Office*, 39 J.L. MED. & ETHICS 15 (Suppl. 1) (2011).

[6] In the public health law research framework, we have distinguished three categories of public health law: *interventional*, consisting of laws that are enacted explicitly to promote better health; *infrastructural*, consisting of the laws that define the jurisdiction, authority and duties of health agencies; and *incidental*, consisting of laws of all sorts that have unintentional or unanticipated effects on health. *See* Burris et al., *supra* note 4, at 175.

infrastructure necessary to do robust evaluations.[7] This included the creation of the Alcohol Policy Information System, a model dataset of legal interventions,[8] and, on the outcome end, the National Traffic Safety Administration's (NHTSA) Fatality Analysis Reporting System (FARS), an invaluable compendium of crash data.[9] It also included sustained support for career researchers and for the development of methods and measures of important intermediate factors like alcohol involvement and compliance with safety belt laws.[10]

This level of investment in evaluation of interventional health law has been more the exception than the rule. In many areas of legal intervention (say, laws regulating the sexual behavior of people with HIV) there is no infrastructure of intermediate measures, and even outcome data (HIV incidence) has serious limits. And when we think in terms of "legal etiology" – the study of the health effects of laws not specifically intended to produce a health outcome the difficulties multiply. It has been a long and difficult struggle, for example, to bring research funding and attention to the impact of mass incarceration on the health of those locked up and those left behind.[11] Adopting a social determinants framework (i.e., understanding health as flowing from basic social structure) adds to the challenge, since the logic of the enterprise is to look for structural characteristics of social organization that shape individual experiences, vulnerabilities and

---

[7] For an excellent collection of examples of outstanding studies, see this special virtual issue of the JOURNAL OF SAFETY RESEARCH, VIRTUAL SPECIAL ISSUE: 100TH ANNIVERSARY OF THE NATIONAL SAFETY COUNCIL, *Elsevier* (2014), at http://www.journals.elsevier.com/journal-of-safety-research/virtual-special-issues/100th-anniversary-of-the-national-safety-council/.

[8] *Alcohol Policy Information System (APIS),* NAT'L INST. ON ALCOHOL ABUSE AND ALCOHOLISM (2011), http://www.alcoholpolicy.niaaa.nih.gov.

[9] *Fatality Analysis Reporting System,* NAT'L HIGHWAY TRAFFIC SAFETY ADMIN. (2014), http://www.nhtsa.gov/FARS.

[10] Timothy Heeren, et al., *Surrogate Measures of Alcohol Involvement in Fatal Crashes: Are Conventional Indicators Adequate?*, 16 J. SAFETY RES. 127 (1985); Uniform Criteria for State Observational Surveys of Seat Belt Use, 23 C.F.R. pt. 1340 (2014).

[11] *See, e.g.,* K. M. Blankenship, et al., *Black-White Disparities in HIV/AIDS: the Role of Drug Policy and the Corrections System*, 16 J. HEALTH CARE POOR UNDERSERVED 140 (2005); N. Harawa & A. Adimora*, Incarceration, African Americans and HIV: Advancing a Research Agenda*, 100 J. NAT'L MED. ASSOC. 57 (2008).

immunities over a life course, and that can ultimately manifest themselves as a cause of death in a wide range of ways.[12]

Ultimately, these challenges persist less because of the scientific difficulties than from a lack of human and monetary resources for overcoming them. From a methods point of view, we have a wide variety of strategies for modeling and measuring the effects of laws.[13] In PHLR, we have stressed the value of causal models[14] and of using the many available, well-tested theories that reveal mechanisms of legal effect.[15] These approaches allow us to identify a large and various roster of theoretically grounded events that could serve as measures of legal effect. Generally speaking, these theories instantiate legal effect in attitudes, behaviors, and environmental changes that can be measured.

The potential for big data to help here is plain: we now produce huge streams of information, from traffic tickets to tweets to the purchase of condoms. If we can begin to draw on this data to quantify implementation and to measure changes in daily life and surroundings that our theories tell us are valid indicators of legal effect, we may be able to overcome the traditional limits imposed by the cost of observation and the gaps in NIH funding. Beyond that, new methods of analysis suited to big data may/should allow us in time to deal better with common variation in enforcement, the typically diffuse and partial impact of law, and the fact that few social objectives are pursued one law at a time.

III. LITTLE BIG DATA FROM LEGAL RESEARCH

For our project, supporting strong scientific research meant, in part, defining and spreading reliable research methods suited to legal questions. When law is being evaluated, the very first methods

---

[12] *See, e.g., Bruce G. Link & Jo Phelan, Social Conditions as Fundamental Causes of Disease,* Spec. No. J. HEALTH & SOC. BEHAV. 80 (1995); LISA BERKMAN & ICHIRO KAWACHI, SOCIAL EPIDEMIOLOGY (Oxford University Press, 2000).

[13] *See generally* ALEXANDER WAGENAAR & SCOTT BURRIS, PUBLIC HEALTH LAW RESEARCH: THEORY AND METHODS (Joseph Wiley and Sons, 2013).

[14] Jeffrey Swanson & Jennifer Ibrahim, P*icturing Public Health Law Research: The Value of Causal Diagrams, in* PUBLIC HEALTH LAW RESEARCH: THEORY AND METHODS, (Alexander C. Wagenaar & Scott Burris eds., 2013).

[15] Scott Burris & Alexander C. Wagenaar, *Integrating Diverse Theories for Public Health Law Education, in* PUBLIC HEALTH LAW RESEARCH: THEORY AND METHODS, (Alexander Wagenaar & Scott Burris eds., 2013).

question is how to capture the attributes of law – the independent variable – in a way that will be accepted as reliable by the scientific community. That means using transparent and consistent methods to transform the text of laws into numbers. For decades, a small cadre of scientists, including lawyers, has been creating scientifically reliable legal datasets,[16] but there was little in the literature in the way of articulated, shared standards cutting across topical silos.[17] Early in the work of PHLR, we commissioned a paper intended to fill this gap, which ended up taking us down an unexpected and still unfolding path of innovation (and which led us to this conference).

The methods paper was written by three researchers with considerable experience in creating legal datasets for public health law evaluations.[18] This article focused on articulating a standard, scientific approach to defining the scope of a data set, collecting the law, creating and implementing a robust coding scheme, controlling quality, and maintaining transparency and reproducibility. The process was characterized as iterative, and focused on measuring the apparent characteristics of legal texts, rather than interpreting their legal meaning. Whether the term "cell phone" in a traffic law (for example) would cover a wi-fi-enabled iPad being used for a Skype call could be quite important for a lawyer applying that law to a particular case, but for purposes of creating legal data, it would normally suffice in the initial coding to observe that the term "cell phone" is used to specify the device whose use the law regulates.

Armed with this explicit methods guide, we thought we should test it by using it to create new datasets. We selected as our pilot "distracted driving" laws – laws directed explicitly at the use of mobile communications devices by drivers – from the first instance of such a

---

[16] Examples include the Alcohol Policy Information System, *see supra* note 8, and the compendium of smoke-free laws created and maintained by Americans for Non-Smokers' Rights, *Smokefree Lists, Maps and Data*, AMERICANS FOR NON-SMOKERS' RIGHTS (2014), http://www.no-smoke.org/goingsmokefree.php?id=519. For a earlier methods pieces, see C. LaFond, *et al., Policy Evaluation Research. Measuring the Independent Variables*, 24 EVAL. REV. 92 (2000); Melanie A. Wakefield & Frank J. Chaloupka, *Improving the Measurement and Use of Tobacco Control "inputs",* 7 TOB. CONTROL 333 (1998).

[17] Jamie F. Chriqui, *et al., What gets Measured, gets Changed: Evaluating Law and Policy for Maximum Impact*, 39 Suppl. 1 J.L. MED. & ETHICS 21 (2011).

[18] Charles Tremper, *et al., Measuring Law for Evaluation Research*, 34 EVAL. REV. 242 (2010); *see also* Evan Anderson, *et al., Measuring Statutory Law and Regulations for Empirical Research, in* PUBLIC HEALTH LAW RESEARCH: THEORY AND METHODS (Alexander Wagenaar & Scott Burris eds., 2013) (elaborating on original article).

law in 1992 to the present.[19] Given 39 states that had passed more than 300 iterations of these laws (i.e., a distinct version of the law in effect from time x to time y), which we were coding for 20 variables, the project was not small, either in the collection of the law or its coding. At that point, we were faced with the question of what data collection and analysis tools to use. Law was early as a profession in the development of databases of basic professional texts,[20] so the law was available in electronic form, but how would we do the coding? The old standby (still used by some in the field) would be paper forms, which leave a clear research trail for quality control and can be easily constructed to avoid clerical error in coding (though there is then a second chance at error in data entry). Some researchers do the coding directly onto spreadsheets, but avoiding input errors in a large dataset requires a high degree of care and skill, and using a spreadsheet requires that redundant coding[21] be manually merged or compared. We tried MS Access, which allows the use of forms, but also requires that redundant coding output be merged. When that proved cumbersome, we tried Google Forms, which allowed us both to custom design our entry form and supported simultaneous redundant coding. There was a limit, though, to how much customizing we could do on Google's platform. If this account was tedious to read, imagine actually running all of these coding experiments, which led to a null outcome: none of these options were ideal.

We wanted something that would store the text of the laws we were coding, and let us create coding forms to make the coding faster and more accurate. So we built it ourselves. The LawAtlas Workbench is, as far as we know, the only software designed for coding statutes and regulations.[22] It allows the creation of custom coding forms, stores the legal texts, allows the coder to code and view the text on the same window, and supports simultaneous redundant coding managed via an easy-to-use dash board. The forms dump to a database that can be downloaded into any of the usual analytic software packages, or a spreadsheet.

---

[19] Jennifer K. Ibrahim, *et al., State Laws Restricting Driver Use of Mobile Communications Devices: "Distracted-Driving" Provisions, 1992-2010,* 40 AM. J. PREVENTATIVE MED. 659 (2011).

[20] Olufunmilayo Arewa, *Open Access in a Closed Universe: Lexis, Westlaw, Law Schools, and the Legal Information Market,* 10 LEWIS & CLARK L. REV. 797 (2006).

[21] Independent coding by two reviewers is an important quality control procedure.

[22] *The LawAtlas Workbench,* PUBLIC HEALTH LAW RESEARCH PROGRAM (2014), http://lawatlas.org/workbench#.U1aEjfldXbw.

Our original aim was to promote wider availability of scientific legal datasets for use by researchers, which the Workbench largely accomplished once we posted the data, codebooks and protocols on our website. It did not take long, though, to realize that we had a trove of digitized legal information that we could easily share with the public, policymakers, the media and anyone else interested in seeing legal trends and the distribution of laws across the country. So we created an interactive public portal where we could publish our data.[23] This was the last element in the construction of a feasible model for "health policy surveillance," the "'ongoing, systematic collection, analysis, interpretation and dissemination' of information about a given body of public health law and policy."[24] Policy surveillance satisfies two basic conditions for the effective use of law and law reform to improve health: the creation of data for evaluation, and the rapid dissemination of health policy activities to speed the diffusion of innovation.[25] LawAtlas is now a robust content management software system for coding and displaying statutory information. By developing reliable routines for research, and building custom software for coding and publishing, we have cut the cost of doing surveys of laws across states – and counties, and the world for that matter. This is big for public health, because of how important law is, and because of the need for evaluation data. But I consider in this paper whether and how what we have been doing might be relevant to law's "big data" future. We hope that what we have done is big for public health, but is it big data? The answer is neither obvious nor sure, and works in part via paradox.

---

[23] *LawAtlas: The Policy Surveillance Portal*, THE PUBLIC HEALTH LAW RESEARCH PROGRAM OF THE ROBERT WOOD JOHNSON FOUNDATION (2012), http://www.lawatlas.org.

[24] *See* Chriqui, *supra* note 17, at 21 (quoting R. R. German, L. M. Lee, J. M. Horan, R. L. Milstein, C. A. Pertowski, and M. N. Waller, *Updated Guidelines for Evaluating Public Health Surveillance Systems: Recommendations from the Guidelines Working Group*, 50 MMWR RECOMMENDATIONS AND REPORTS 1, 2 (No. RR-13) (2001)).

[25] This was the view of an Institute of Medicine Committee, which wrote:

> "Legal interventions merit study for their effectiveness and comparative
> effectiveness (both against other legal intervention and compared to
> other kinds of interventions). Furthermore, a system of surveillance
> could be developed and pilot-tested to track the progress of efforts to
> expand the geographic reach of effective policies and laws, and to
> identify unmet needs for policy development and advocacy strategies."

INSTITUTE OF MEDICINE, FOR THE PUBLIC'S HEALTH: REVITALIZING LAW AND POLICY TO MEET NEW CHALLENGES 11 (The National Academies Press, 2011).

Normally, when we think of Big Data in the realm of legal rules, interpretations or procedures, we think about huge piles of digital legal information, such as:

- All published and many unpublished federal and state court decisions since the beginning of the Republic.

- Current and historical statutes going back – well, HeinOnline says its historical statutes go as far back as 1717.

- Current and at least some historical regulations.

- Ordinances from thousands of cities and counties.

- Federal and state dockets and case records.

- More law reviews than even Fred Rodell[26] could shake a stick at. The big data connection lies in the fact that the piles are high and can be attacked using machine learning and other data mining/big data tools either to extract specific kinds of documents (e.g., in discovery) or to identify patterns in the otherwise impenetrable mass of instances. The emblematic big data scientists are using network theory or neural network programming to find the nuggets sampling never could. Or something like that – I don't know, because that's not what we do.   We legal coders are doing the apparent opposite of high-tech analysis. The strategy we adopted may be radical for law, but it is pretty conventional human-based research: manual coding of the apparent features of the text. The use of a scientific method and computers gives us efficiency advantages, but it is still people making decisions and inputting the data. I will stop and describe that briefly.

Some of the key elements of the process:

- We are not interpreting:[27] we are coding apparent features of the text – what the law says rather than what it means. This may

---

[26] Rodell, a Yale Law professor, was famous for his pithy dismissal of papers like this in journals like this: "There are two things wrong with almost all legal writing. One is its style. The other is its content. That, I think, about covers the ground." Fred Rodell, *Goodbye to Law Reviews*, 23 VA. L. REV. 38 (1936).

[27] Of course, a coding scheme is an interpretation, but the emphasis in scientific coding is to observe rather than to interpret. *See supra* text accompanying note 18. As lawyers

sometimes produce data that do not fully answer a user's legal question, but the process is not designed to eliminate the need for lawyers or legal analysis. Rather, in capturing functional elements of a law descriptively, coding from a lawyer's perspective simply gets the lawyer to the interpretive questions more efficiently. That said, coding the apparent features of a law does create data that, in theory, could support computerized interpretation based on algorithms defining the conditions for a given answer set.

- We use a standardized process to increase efficiency and accuracy with training and built-in quality control. This is practice making perfect, but it is also about reducing costs. We believe policy surveillance is a public good, but that just puts it into competition with all the other under-funded public goods in public health. The systematic approach of science is just going to produce marginally more accurate results for marginally less money than the cottage craft model prevailing in law.

- We use software designed for the purpose. This is also, first and foremost, an efficiency element. Reducing error and simplifying the research, coding and management process produce better results cheaper. Publication can also be rationalized, through the creation of standard web pages and a catalog of infographics that can be populated with little custom work.

- We aim to maximize the usability of the data. This may be gilding the lily, since the whole point of making law into data is to take advantage of the innumerable ways data can be used once it is created for a particular purpose. Small steps, like using FIPS or other standard codes to denominate jurisdictions, can facilitate re-use of the data for other descriptive and analytic purposes.

- We pursue efficiency in updating. A stable, user-friendly content management system and explicit, documented methods minimize the cost and effort required to monitor legislative activity and update datasets. Datasets for evaluation are often built in longitudinal form, which better supports quasi-experimental analysis. Even when the original dataset was cross-sectional,

---

learning to do this sort of research, and teaching other lawyers, we came to realize how well law students are taught to combine observation and interpretation as they read a statute or regulation. Typically, on finishing a first reading, they can tell you what the law means more convincingly than they can recount what it says.

ongoing surveillance gradually builds longitudinal data, because changes are added to – but don't replace – the prior law.

For all these virtues, we are still describing a process in which human beings collect and code information into data. This is a mode of data creation, and a level of output, that looks like the 97-pound weakling on Big Data Beach. And yet, … There are a number of respects in which coded statutes might count as big data. Let's start with the idea that the bigness of big data is not just the total volume of information, or even the size of any particular pile in the warehouse, but in the interrelationship between the amount of information, the range of information, the accessibility or manipulability of the information, and the potential for the information to answer important questions. If I get you that far, then coded statutory information might fit into a number of flavors of bigness.

- **Intrinsically big:** Laws are important.

- **Marginally big:** Right now, there is virtually no statutory info in the datasphere. As big data analyses draw upon new forms, sources, and combinations of information to form impressions of what matters in the world, it is not a stretch to think that law in its many roles (regulatory tool, structuring agency, cultural product) will be playing a role that can be captured, if law is available as data.

- **Relatively big**: While the statutory and regulatory data that we are capturing is only a small part of the corpus of available legal information, it is arguable that statutes and regulations selected for their perceived importance to various users could be disproportionately useful to analysts and other users.

- **Introductory big**: Machine coding of legal text is cool and could create huge big heaps of data – but is really hard. While we wait for brilliant people to solve that one, working with human-coded material has the potential to identify or stimulate needs for more legal data and more ways to manipulate it.

- **Conceptually big** (as law goes): There seems to be some kind of interesting advance in the idea of descriptive coding by humans. Lawyers are trained to seamlessly combine observation and interpretation. By contrast, the scientific

method we demonstrate aims simply to capture the apparent features of the text within a minimalist semantic framework (eg., prohibition, regulated persons, etc.). Given a descriptively coded dataset, it becomes at least theoretically possible for legal research algorithms to supplement human analysis in the interpretative process. This is just one way that law as data might change a lot of things in government, law practice and daily life in ways we don't see now.

IV. Conclusion

Though we often have our attention focused on the controversial measures, like gun control or soda cup limits, law is actually a rather popular tool for improving public health,[28] and in any case it is hard to imagine making a great deal of progress without it in areas like injury and non-communicable diseases.[29] It has been important to health for a long time, but it has not been well-integrated into the data dashboard. Big data seems to offer a lot of scope for more and better evaluation of the impact of laws and legal practices on health. Not only do we not have evidence for a lot of uses of law in health, but it is often hard for interested people – citizens, the media, health officials – to find or keep track of developments in law, let alone to innovate together. A combination of humble, tried and true techniques for making data, and the amazing ways we now have to slice, dice and combine data, creates the possibilities of policy surveillance and the integration of statutory and other policy data into the full range of platforms, informatics systems, mashups and uses that constitute our big data future. Ours is not, at the moment, very big data, but maybe there is a lesson in that. What makes data big is not necessarily size for its own sake, but its capacity to guide decisions and choices and to help us learn about our world.

---

[28] Stephanie Morain & Michelle M. Mello, *Survey Finds Public Support For Legal Interventions Directed At Health Behavior To Fight Noncommunicable Disease*, 32 HEALTH AFF. 486 (2013).

[29] *See generally* Burris & Anderson, *supra* note 2 (recounting rise and importance of law in modern public health).