Measuring Statutory Law and Regulations for Empirical Research

A Methods Monograph

PHLR

Making the Case for Laws that Improve Health

February 28, 2012

Robert Wood Johnson Foundation

Measuring Statutory Law and Regulations for Empirical Research

A Methods Monograph
for the
Public Health Law Research Program (PHLR)
Temple University Beasley School of Law

By:

Evan Anderson, J.D.
Senior Legal Fellow
Public Health Law Research Program
Temple University

Charles Tremper, J.D., Ph.D.
Director
Perutilis Research & Consulting

Sue Thomas, Ph.D., M.Ed
Senior Policy Researcher
Pacific Institute of Research and Evaluation

Alexander C. Wagenaar, Ph.D.
Associate Director
Public Health Law Research Program
Professor of Epidemiology
University of Florida, College of Medicine

Summary

Effectively studying the relationship between law and population health requires (1) variation in both the law and health outcomes being studied, preferably over space as well as time, and (2) valid and reliable methods for capturing variation and representing it in forms that allow comparison. A rigorous method for measuring law generates numeric data representing variation in law. The key feature of the method – and that which distinguishes it most from traditional legal research – is that relies on observation of the apparent features of legal texts. By eschewing interpretation, this restrained observational approach produces data that is replicable through a process that is transparent. Transparency and replicability are essential attributes of scientifically defensible data.

There are challenges in measuring law. Relevant legal texts can be hard to find and when found can be rife with ambiguous and conflicting meanings. Formulating reliable and valid ways of reducing complex bodies of law into numeric data can be difficult. There may also be cultural and logistical challenges to blending legal and empirical expertise. These challenges can be overcome by a methodical process of design, data collection and analysis that adheres to general scientific standards. Steps include the careful delineation of the legal questions to be addressed and the scope of the research; the iterative development and refinement of a coding scheme; quality control; and the production of a transparent research protocol and codebook.

Introduction

Law affects health in many ways. Policy-makers use law to intervene between a health risks and harms. The effects of law on health, however, extend far beyond deliberate uses of law by health-oriented policy-makers. Few of our environments or behaviors remain untouched in some significant way by law. This monograph focuses on how to measure law to enable rigorous evaluation of its effects.

"Measuring law" is determining relevant dimensions or components of an area of law, categorizing the legal elements of a policy, and using the resultant categorization schema to produce accurate representations of the law in terms of counts and numeric indicators. The process for measuring law relies on techniques that are common practice in both quantitative and qualitative research. While few of these techniques are conceptually challenging, their application to the law can be difficult. The bulk of this monograph is devoted to providing a step-by-step guide for reflecting variation in laws across time and or space for purposes of evaluation research. To provide some context and make clear the importance of these steps, the monograph begins with a short section explaining some of the basic principles underlying the process. A final section describes common challenges and offers suggestions for addressing them.

The Impetus behind Measuring—or Mapping—Law

Supreme Court Justice Louis Brandeis famously remarked that "it is one of the happy incidents of our federal system that a single courageous state may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country" (New State Ice Co. v Liebman, 1932, p. 311). Innovation by states and other political units is vital to having an effective regulatory system. Policy experimentation by localities and states is almost always a necessary step on the way to identifying effective public health laws. Consider the problem of motor

vehicle crashes for teenagers and graduated driver licensing (GDL) laws. In the mid-1990s, a few states began experimenting with laws that restrict when and under what circumstances teenage drivers could operate a motor vehicle. As these laws proliferated in number and type, researchers evaluated their effects, first in studies comparing crash rates in single states before and after the adoption of a GDL restriction, and then in studies comparing changes in crash rates in states adopting GDL laws of varying restrictiveness. As these studies accumulated, it became very clear that restrictive laws saved lives. Less than fifteen years later, almost every state has adopted a similarly restrictive law. Teen crash rates have declined continuously since, marking GDL laws as one of the great modern examples of how policy innovation, rigorous evaluation, and evidenced-based dissemination can save lives (Preusser & Tison, 2007).

Two factors made GDL research possible. First, there was variation in both state laws and state motor vehicle crash rates. Second, there were methods for measuring that variation that enabled statistical comparisons. Measuring crash-related harms is accomplished by generating counts through police reports and other administrative data. But how does one measure and quantify something textual like the law? The answer turns out to be that is not too difficult with one or more legal researchers, careful consideration of a handful of persistent areas of error, and attention to the usual basic principles of science. As any qualitative researcher can attest, characteristics of texts can be observed and these observations can be easily converted into numeric indicators or coded as it is often called. At its core, the task of a coding law is not altogether different from coding an interview transcript. In each instance, researchers strive to measure the apparent features of the texts in ways that are consistent with scientific standards of validity and reliability.

There are two primary sources of difficulty in the process for measuring law. The first is typical of almost all research that involves analyzing the meaning or contents of texts. Law is, by its

nature, an abstraction with at best an uncertain underlying empirical foundation. For this reason, measurements of law themselves cannot be directly validated against observable phenomena in the natural world (see Singer, 1984). The method for measuring law offered in this monograph is more similar in many ways to observation than traditional legal interpretation. But few observations can be made of laws that do not require some predicate legal decision-making based on assumptions about the nature of the legal text being examined. The types of observations of laws that are defensible regardless of context or purpose – like the number of words in a statute or whether it includes a specific term – tend not to provide much value by themselves in public health law evaluations. The decisions that shape what laws are collected and how they are understood increases the importance of reporting how and, in some instances, why specific legal measures were adopted to represent a particular construct.

The second primary source of difficulty in measuring law is identifying the right legal texts to collect and examine. Unlike the qualitative researcher who creates a file of transcripts for coding by, for example, interviewing some defined group of people, the legal rules that regulate life in the United States are distributed over space, time, levels of government, and types of law. Determining the prevailing law governing the sale of sugar-sweetened beverages in a selection of cities, for example, might necessitate gathering laws from different sources (legislative, executive, judicial, electoral, and constitutional) and at different levels of government (federal, state, and local). In addition to difficulties finding the right provisions, researchers also have to consider how provisions interact within a broader legal framework. In most instances, lawyers are needed to help determine what laws are relevant, how to find them, and, in some instances, how they are to be interpreted – which means that for empirical researchers embarking on an evaluation of law, collaboration with legal colleagues is almost always necessary to produce optimal results.

While legal expertise is essential to measuring law it is not sufficient to generating valid data. The bulwark against error in legal measurement is a deliberate commitment to standard principles and practices of good science. The method below provides steps for operationalizing those principles and practices in legal research. Some standard scientific practices – like the creation of a research protocol that enables replication and updating of data – might seem odd at first to legal colleagues who are accustomed to the more normative and interpretive world of traditional legal research in which the process for doing research is closely guarded as the lawyer's stock-in-trade. But our experience is that legal researchers quickly internalize these rules; indeed, they often find them valuable in other areas of their work.

Projects that blend legal and empirical research fall into two primary categories. In the first, legal research measuring the features of law is driven by the goal of testing a specific hypothesis. A hypothesis-driven project might be, for example, investigating whether increasing the age at which individuals can drink will reduce fatalities from motor vehicle crashes (Wagenaar, 1983). In such instances, decisions about measurement are guided by, and typically limited to, the question of interest. Suppose that state law bans use of cell phones by bus drivers and novice drivers. Researchers evaluating the effect of the law on novices normally will not include the bus-driver ban in their data. In the second category, what we refer to as legal mapping studies, the purpose is to survey the legal terrain in a policy domain capturing all the characteristics that vary in ways plausibly related health. When conducted with appropriate rigor and transparency, legal mapping studies yield datasets that can be used in evaluations examining many different hypotheses. This monograph focuses on measuring law for hypothesis testing, but the same principles and practices are, with obvious exceptions, required for legal mapping studies.

A few final words about terminology bear mentioning. Throughout this book, law is understood in its broadest terms to encompass both legal texts and the whole range of institutions,

practices, and beliefs through which laws influence health and the determinants of health. This monograph is devoted exclusively to law as it is set out in Constitutions, statutes, regulations, ordinances and other legal texts. "Provisions," are clauses or sections of laws. We use the term "policy" to represent the operation of two or more laws. So, for example, a state policy regulating the use of mobile communication devices by drivers might include a statute specifying a prohibition (e.g., a ban on drivers under age 19 from "texting" while driving), a statute specifying the fine for a violation, and a statute specifying the extent to which police officers can enforce the prohibition as a primary or secondary offense (Ibrahim et al., 2011). This contrasts with definitions of policy that include a broader range of political outputs expressed as goals, preferences, long-range objectives and the means by which ends are sought (Plano et al., 1982).

The Process for Measuring Law

As is the case in all scientific research, the questions of interest and the availability of data define measurement objectives. The types of legal data pertinent to evaluation research vary widely depending on, among other things, whether legal measures are to be used as independent or dependent variables, and the particular research design chosen for the study. Notwithstanding these differences, the process for measuring law depicted in Figure 1 is composed of steps that are essential in all legal measurement projects. The process is generally iterative, with one or more steps being repeated as discoveries at one stage expose inadequacies of constructs developed at a previous stage. Each step in the figure is described below except the first which is addressed in other monographs.
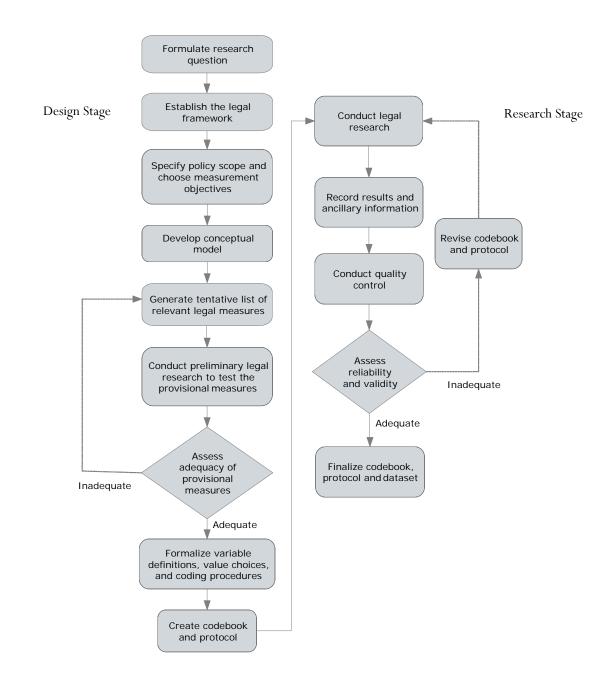
Formulate research question

Establish the legal framework

Specify policy scope and choose measurement objectives

Develop conceptual model

Generate tentative list of relevant legal measures

Conduct preliminary legal research to test the provisional measures

Assess adequacy of provisional measures

Inadequate

Adequate

Formalize variable definitions, value choices, and coding procedures

Create codebook and protocol

Conduct legal research

Record results and ancillary information

Conduct quality control

Revise codebook and protocol

Assess reliability and validity

Inadequate

Adequate

Finalize codebook, protocol and dataset

Figure 1. Process for Measuring Law

Establishing the Legal Framework

The first step in any legal measurement project is to identify the legal framework of interest. It is seldom practical or feasible to study every possible law related to a health issue. Consider the problem of distracted driving. It is clear that using a mobile communication device while operating a motor vehicle is a dangerous behavior. Many states and localities have responded by prohibiting activity with such devices for different groups of drivers. For empirical researchers interested in understanding the relationship between law and this high risk behavior, these interventional laws have obvious importance. But they are not the only or even necessarily the best place to start. The tort system could exert an equal if not greater influence on drivers' tendencies to answer a call or send a text message. Or it could be that law regulating insurance or employer liability is a plausible factor in distracted driving. Decisions about the type of law to study depend on the purpose of the study and theorized relationships of interest.

Assume that a researcher decides to focus on state laws that specifically prohibit activity with mobile communication devices. These laws exist at the local and federal level too (e.g., those that apply to long-haul truckers). Choice of legal framework is in this way also a choice of at which level or levels of government law will be studied. Table 1 displays a familiar categorization of law by level of government and by source, with the most common type of law in each cell. Statutes enacted by Congress or state legislatures are generally the easiest source of law to measure because they are readily accessible and—compared to common law created by courts—relatively straightforward. The ordinances of cities, counties, and other units of government below the state level offer similar advantages, although they may require more effort to locate. Bills under consideration by a legislature may also be of interest. For example, Wagenaar and colleagues (2006) used measures of bill introductions as an intermediate outcome in a national evaluation of statewide coalitions whose objective was to reduce the availability of alcohol to youth. In this study, bill introductions

functioned as one indicator of policy attention to an issue. For some policy domains, law emanating

from the executive branch, (for example, regulations, executive orders) can be as important to

measure as statutes and ordinances.

Table 1

Types of Law at Different Levels of Government

	Federal	State	Local
Legislative	Statutes	Statutes	Ordinances
Executive	Regulations Executives Orders Administrative Judgments	Regulations Executives Orders Administrative Judgments	Regulations Executives Orders Administrative Judgments
Judicial	Case Law (Common Law)	Case Law (Common Law)	Case Law (Common Law)
Electoral	--	Initiatives Referenda	Initiatives Referenda
Basic Law	Constitutions	Constitutions	Home Rule Charter

Note. Dash indicates the absence of a law by level and source.

All, none or some other combination of these legal frameworks might yield levers for effectively

reducing mobile communication device use by drivers. In choosing a legal framework, there is no

right or wrong answer a priori. But the decision should be a mindful one supported by plausible

theories – both legally and behaviorally or biologically, depending on the nature of the exposure –

about how laws within the framework relate to a health outcome of interest.

Specify the Policy Scope and Choose Measurement Objectives

To conserve resources and ensure that the ultimate legal measures generate data capable of illuminating the hypothesis of interest, it is important to define measurement objectives at the start even if they change during legal research and preliminary coding. In a study of distracted driving, for example, one might begin with the objective to capture how state laws prohibiting driver activity with mobile communication devices have evolved over a five year period (chosen perhaps because of the availability of crash data) in regard to covered activity, devices and classes of drivers. The sentence above is deliberately vague one on point: what is the right unit of analysis? Is it individual legal texts or state policy? The answer depends on the purpose of the measurement. An example of a study that focuses on specific laws as the unit of analysis is the Wagenaar and colleagues (2006) study, mentioned above, which used measures of bill introductions as an intermediate outcome in a national evaluation of statewide coalitions whose objective was to reduce the availability of alcohol to youth.

In most evaluations, the ultimate unit of analysis is state policy and not specific legal provisions. That is, each record – or row in a flat table – represents all the characteristics of state policy – for example, the legal drinking age and its enforcement – whether determined by one legal provision or multiple provisions set out in numerous statutes. A dataset reflecting variation in law for all U.S. states and the District of Columbia would therefore have 51 rows excluding headings (and leaving aside for the moment the issue of data encompassing changes in the policy over time). The columns in the table represent variables and the characteristics of state law those variables describe. While that is almost always the intended structure of the final dataset, seldom is it the easiest way for organizing and making sense of laws in the early stages of research. This is especially true in instances in which there are many related or similar provisions in one state.

Periodic discussion and clarity regarding the unit of analysis ensures that different ways of organizing the search returns in early phases of the research do not muddle the purpose of the legal measurement and therefore its ability to inform the hypothesis of interest. Casual and somewhat inevitable short-hand descriptions of the legal research increase the chance of confusion; it is easy to describe the example in this section simply as a study of "distracted driving laws." A truer articulation of the project and the underlying measurement objective is to: determine which activities with mobile communication devices are prohibited for specified drivers under state statutory law, how those prohibitions are enforced and what the fines are for associated violations.

Develop Causal Diagram

Having defined the scope of laws to be collected, the next question is simply which characteristics of the laws are to be measured. Causal diagrams – as described in more detail in the monograph by Swanson and Ibrahim – are valuable tools for this purpose. By forcing researchers to identify plausible links between the law and health outcomes, causal diagrams help flush out the legal inputs relevant to the question of interest. In the distracted driving context, for example, researchers might suspect that the primary legal mechanism mediating reductions in device use is deterrence. In other words, laws that are easier to enforce and carry higher penalties result in the greater reductions in high risk behavior. In that instance, provisions specifying whether police officers can enforce prohibitions as a primary offense – that is, without needing another pretext to make stops – have obvious importance. Also important in that scenario are provisions specifying fines or other penalties like suspensions for drivers on learner's permits. A causal diagram based in part or entirely on a different mechanism might suggest a different set of legal inputs.

Generate Tentative List of Relevant Legal Measures

In hypothesis-driven research causal diagrams limit the laws that will be studied based on theories about how those legal inputs relate to some other outcome. The legal inputs themselves are often, until this point, broadly understood as concepts like manner of enforcement, scope, and severity of fine. To capture how these conceptual legal inputs vary, the important components of each must identified. Then the task is to formulate simple schemes for describing that variation. Operationally speaking, this is the point in the process to start creating variables. They need not be dichotomous which is, as noted below, sometimes preferable for variables in the final dataset. For the case of distracted driving laws, a sufficient set of preliminary measures to describe the scope of the laws might be categorical variables reflecting activities prohibited (for example., only texting, only talking, talking and texting, other), classes of drivers covered (for example, all drivers, bus drivers, inexperienced drivers, other) and devices (for example, cellphones, personal digital assistants, laptop computers, other) subject to the law.

Conduct Preliminary Research

Whether the tentative measures align with relevant variation in the laws is an empirical question that should be tested on a sample of the jurisdictions to be studied. If little is known about the variation in the law across jurisdictions, a random sample of jurisdictions increases the value of preliminary review. If the structure or operation of law in the area of interest is known to vary systematically across jurisdictions, purposely sampling jurisdictions across strata is best. The size of the sample depends on the law being studied and the a priori knowledge that the research team possesses about legal variation in the area. In addition to examining a handful of jurisdictions, surveying literature that describes the policy environment whether in legal journals or other social science research sets the research on much firmer ground moving forward. Preliminary research illuminates relevant

dimensions of the law, and provides the research team with a basis for estimating the breadth and complexity of the law being studied and the resources needed to systematically collect and analyze it moving forward.

Assess the Adequacy of Provisional Measures

Pausing to test the adequacy of provisional measures reduces the likelihood of wholesale recoding that would be necessary were the coding scheme to be found later in the research process to be incapable of reliably and validly capturing relevant variation. At this stage in the process, insights about the important dimensions of variation in the law and how to elicit that variation through coding questions should be crystallizing in the minds of the research team. For the distracted driving example, researchers might decide that the provisional measure describing classes of covered drivers is insufficiently specific. Rather than lumping all inexperienced drivers into one group, the variable could be refined to distinguish between laws that cover drivers by age (for example, at ages 16, 17, 18, 19 and above) and laws that apply to all new drivers regardless of age. It might also become apparent, at this point, that a handful of exceptions reduce the scope or enforceability of prohibitions, such as exceptions for hands-free use or exceptions in statutes banning texting for the activity of typing keys to start or end a call.

Formalize Variable Definitions and Coding Procedures

It is in coding that measurement of law differs most significantly from traditional legal research. Traditional legal research typically produces narrative descriptions of how one or more laws differ in both their text and their meaning; measurement of law determines how laws vary on specific characteristics in specified units. From an operational perspective, this means first identifying all the relevant ways that laws vary and then finding numeric schemes (or preliminary textual ones like

"yes" and "no") to capture that variation. Although it is a bit counter-intuitive and may strike coders as odd at first, coding questions, and the variables they define, should strive to minimize human judgment. Coding questions requiring interpretation undermine reliability because of innate differences between coders. Consider a coding question asking whether a law prohibiting communication on wireless telephones by drivers applies to a talking over an internet connection through a headset attached through a laptop. Different coders could reach entirely defensible but different conclusions based on how they interpret the operative terms and ultimate meaning of the rule (i.e., a computer is not a wireless telephone, versus, anything that allows electronic transmission of oral communication is a telephone even if embedded in something that has other purposes).

Adding a "Not Sure" category to the coding choices for a variable provides a useful safety valve, especially in the early stages of coding. The "Not Sure" category gives coders an option for handling ambiguity and conflicts that may or may not be resolved by additional research or subsequent developments such as a court ruling during the course of the study. Sometimes, a "Not Sure" or "Unsettled" code will remain in the final data set so that those records can be excluded from analyses.

The use of dichotomous "Law/No Law" variables to code the presence or absence of a particular type of law is useful only for some limited types of research questions and study designs. Dichotomous variables can easily obscure a great deal of legal variability and, hence, limit the value of a dataset for analyzing, adopting, and evaluating policies. Consider an example from the Alcohol Policy Information System dataset (2011). A dichotomous Law/No Law variable initially might seem adequate for indicating whether a state prohibits people under age 21 from purchasing alcoholic beverages. Most states have such prohibitions and almost all of the laws on the topic are clear and simple. Several states, though, permit purchases by youth in some situations, such as acting in conjunction with law enforcement. The laws in New York and Delaware are less

straightforward. Although these two states do not prohibit underage purchase, they do prohibit obtaining alcohol in connection with making a false statement. Rather than trying to shoehorn these laws into a dichotomous Law/No Law variable, a polychotomous variable could be created, in this case with four codes to include an absolute prohibition, a prohibition with exceptions, a prohibition against purchase in connection with making a false statement, and no restrictions. Better still, the categorical variable could be separated into four dichotomous variables each measuring one dimension, an approach adhering to the principle that a measure should generally concern a single distinct, unitary attribute.

One of the preliminary measures identified above for distracted driving defined the possibilities for regulated activity as (1) only texting, (2) only talking, (3) talking and texting, and (4) other. A simpler approach is to create two dichotomous variables that respectively answer the questions (1) does the law prohibit texting, and (2) does the law prohibit talking. These uni-dimensional variables can be easily combined with each other to reflect instances when both activities are prohibited or combined with other variables defining the existence of exceptions (e.g., is there an exception for hands-free use) to represent all the regulatory permutations.

For the coding of legal texts, classifying features of laws into categories – that is, using categorical variables – is the most common way to reflect variance. The categories – or attributes – of a categorical variable have no natural numeric ordering or quantitative relationship. In contrast, the attributes of ordinal variables have a natural ordering (e.g., mild, hot, hotter). Interval variables are a special type of ordinal variable in which the difference between attributes is meaningful (e.g., separation into first, second, third and fourth quartiles). Some features of the law can be measured at ordinal or interval levels, thereby enabling dose-response analyses, possibly enhancing statistical power to detect the law's effects, or more closely matching the analytic model to the theory. Penalty

type, for example, can be coded as an ordinal variable, with civil infractions, misdemeanors, felonies, and capital crimes as values representing ascending severity as defined by the law.

Continuous variables, which can take any numeric value (for example, temperature in Celsius) provide the same types of benefits of ordinal and interval variables. Attributes such as length of jail terms, appropriation amounts, and legal thresholds (for example, alcohol-impaired driving defined as blood alcohol content over 0.08), are often measurable with continuous variables. Careful definitions and coding protocols can also produce continuous variables based on a policy's non-numerical features. This is distinguishable from combination of multiple variables to create a composite measure of a larger characteristic of the law. For example, researchers might use length of jail sentence, magnitude of fine and the availability of defenses to create a composite measure of "stringency." Composite measures are potentially powerful as analytic tools but difficult to validate. This synthetic approach – or second level coding – is discussed as a special case in the final section.

Create Codebook and Protocol

Creating a codebook and a well-defined, precise coding protocol to capture the decisions made during the design phase facilitates both the initial legal research process and any subsequent attempts to replicate or update the data. The codebook should reflect the standards of any good data collection documentation as well as the special considerations described above for coding laws. Elements may include a description of the study; the scope of data collection; variable definitions; values (codes and their definitions); algorithms for constructing scales; technical information about files—tables, records, relationships, number of records for each case (some jurisdictions have multiple related policies); and details about the data (columns, text, numeric, Boolean).

A codebook alone is inadequate for any but the simplest study because potentially critical decisions about coding conventions and procedures for legal data are rarely apparent from

examination of a codebook. A comprehensive protocol includes information about how the laws were collected (for example, exact legal text databases searched, exact search terms and syntax); inclusion and exclusion criteria for defining the body of legal texts to be coded; precise rules for coding the variables (which legal terms support classification into a specific category); conventions used regarding effective dates or other determinations about a law's operation; and standards for collecting legal citations.

For teams composed of empirical and legal researchers, up until this point, most of the activity should have been joint effort with lots of dialogue. In the next phase, however, the distribution of labor shifts considerably. Collecting the relevant laws requires legal training. Especially if legal researchers have not been integrated into the earlier steps, but even if they have, this is an ideal time for training. It is imperative that law students and other legal researchers understand not only what they are looking for but also why so that they can report on unanticipated nuances in the law. Legal researchers must follow good legal research practices like reading provisions in context by carefully locating and considering the definition of all operative terms. Relevant provisions are often located by key word searches; without explicit instruction some legal researchers will not inspect the other provisions in the same part of the statutory code. The use of statutory tables of contents is not always emphasized in training of law students but is indispensable for making sense of statutory schemes. Being observant for court decisions that substantially influence the legal meaning of status of the law is also essential. Key points of emphasis in discussions with law students include keeping records of any needed modifications in the research protocol, erring on the side of over-inclusion in the collection of law, and raising questions for discussion with supervisors. The training of legal researchers naturally coincides with collective review of the protocol and codebook. Whether the legal researchers are students or experienced

lawyers, from this point forward, the protocol and codebook should provide clarity to guide the location, organization and eventual coding of legal texts.

Conduct Legal Research

Having tested and refined the coding scheme, a project team can then proceed to the task of systematically collecting and coding the relevant law. Available resources to perform original legal research include Westlaw and Lexis, both comprehensive proprietary online legal research services. HeinOnline is another proprietary source that provides access to state session laws going back in some instances to the 1840s. Alternatives to these tools that can suffice for some studies include the commercial service Loislaw and publicly available on-line sources such as the Library of Congress web site for federal material. The commercial service State Net, offers access to recently enacted statutes and pending bills. An excellent resource for studying legislative activity is LexisNexis Advanced Legislative Service (ALS), which catalogues only bills that were eventually adopted. The returns from ALS provide a valuable picture of how law has evolved in a particular area and can be used for updating or checking established legal datasets. Conducting searches in these databases and making sense of returns requires legal training in almost all instances. For large projects, consulting with legal librarians and database vendors – which typically offer assistance free of charge for law students and many other legal researchers – can increase efficiency of research even for experienced lawyers.

Finding legal materials at the city and county levels is often difficult. There is currently no comprehensive or authoritative collection of local laws. Lexis, Westlaw, and FindLaw maintain partial collections of municipal ordinances and some jurisdictions publish their own materials online. Growing repositories are also available at www.municode.com, www.amlegal.com, and www.statelocalgov.net. If the research goal is limited to one or a few local jurisdictions, finding the

relevant ordinances is often feasible. As the number of jurisdictions expands, the task quickly

becomes unwieldy and will require extensive investment of researcher time to search multiple data

bases and, in some occasions, query local governments directly.

Record Results and Ancillary Information

Well-designed data collection systems for gathering legal source material are crucial for a smooth

measurement process. An adequate data collection system for coding the law stores much more

than the final resulting data for each variable. The system needs to retain ancillary information that

supports the coding, typically in the form of extended blocks of legal text. The system should allow

for changes to variables and codes as the project progresses, knowledge is gained, and previous

errors are corrected, as well as keep a record of such changes. At a minimum, all relevant statutory,

regulatory, and case law citations should be collected and recorded. Because coding of a single

variable may depend on a legal analysis based on several sources of law, (for example, three

regulations, two statutes, and a court case), a data field with no character limit is generally best for

citations. Ideally, citations should be stored in records that have a many-to-one relationship with the

coding record so that each citation can have its own field and be stored together with additional

information.

Along with citations, collecting relevant text facilitates subsequent review of coding

decisions. Microsoft Access, which permits up to 60,000 characters of text in each data field, can be

used for cataloguing small- or medium-sized laws. Custom-designed databases can also be

purchased from commercial vendors for large legal measurement projects. Rigorous and consistent

coding decisions are enhanced by recording rationales for coding in the data collection system,

especially in instances when there is latent ambiguity. In addition to storing legal information,

retention of the coders' notes, comments, and questions offers a big advantage.   Maintaining

records of each coding decision helps immensely. Although coding decisions should, as noted above, have a solid basis in the observable features of the legal text, there may also be instances when the text of a law is explicit but some important extrinsic information exists that changes the effect of the legal text – for example, a law enforcement agency choosing not to enforce a law through an organizational policy that is not included in the data. There may also be instances when the answer to a coding question is clear but some feature of the legal text is noteworthy. For example, consider a study measuring whether states include syringes within the definition of prohibited drug paraphernalia. Every state that defines a class of such objects refers to it as drug paraphernalia except Georgia which uses a different term, drug-related objects. Noting this sort of nuance in terminology is another appropriate use of comment boxes and can greatly reduce inaccuracies and confusion in later analysis.

As with all tasks that require repeated actions and fine-tuned manual manipulations, random human error is an inevitable threat to data integrity. The traditional model of coding texts involved three objects: a codebook describing the coding question, a datasheet in which coding decisions were placed (typically an Excel file or other sheet with lots of rows and columns), and the text to be coded. Moving between the three materials provides coders with many opportunities to make mistakes. This sort of error can be greatly reduced through the use of coding platforms that integrate codebooks, datasheets, and the legal text being coded. Data entry forms designed in Microsoft Access, for example, enable researchers to create templates in which the legal text to be coded visible next to coding questions that drive data to an underlying data table.

Clear coding roles for each person and adherence to rigorous implementation procedures reduce both outright errors and subtle distinctions that might otherwise go unnoticed. Regardless of how many coders work on the project, each will need to scrupulously follow clearly defined protocols and adhere to all coding guidelines. This is true for all types of research, and no less so

when measuring law. Coders must not only be thoroughly versed in conventions adopted at the design stage of the research project, they must also be empowered to alert the principal investigator to oddities that arise in the course of coding that may require modification of those conventions, or the use of caveats and elaborations in subsequent descriptions of the research.

Conduct Quality Control

Systematic quality control is essential to measuring law. Having at least two legal researchers concur on coding decisions is a minimal standard for achieving accurate codes. The two coders should work completely independently in coding each variable and recording results, without reference to each other's work. In a significantly less rigorous protocol, one researcher codes the data initially and the other then checks the coding. This more streamlined option catches many errors, but does little to find mistakes arising from laws that the first researcher missed.

A second important step in assessing accuracy of data gathering and coding is comparison to secondary sources or checking with government officials. In many instances, at least one legal survey exists which provides a good cross-reference. The value of these secondary sources is strengthened by the fact that many are created through different methods, like surveys of informants in state agencies or professional organizations. Measures adopted to minimize error should be included in the research documentation. Not only might such protocols facilitate future corrections of errors in specific legal compilations, their use will more broadly advance the quality of the science and the state of knowledge. Future researchers can adapt, modify and enhance the application of these protocols to improve understanding of the specific topic initially under study as well as a wide range of related policy issues.

Assess Reliability and Validity

Quality control measures are intended to test how well legal researchers have applied the protocol and coding scheme. Even perfectly executed protocols and codebooks can generate misleading data. Redundant coding reduces the likelihood of random or careless error. It also provides a reasonable assessment of reliability. But sometimes even independent coders share similarities that bias their treatment of the law. Being systematically integrated into a research project can subtly influence how a researcher collects or codes law; likewise the coders available to participate might share characteristics that predispose them to similar patterns of observation or analysis. A final important step to address these concerns is to have a third legal researcher who is totally naive to the project recode a randomly selected portion of the records. If desired, the rate of divergence can be reported as a crude rate or as rate that adjusts for the probability of randomly selecting the "right" answer. Cohen's Kappa provides a more conservative estimate of the reliability of a test than the crude rate of divergence by accounting for the fact that, for example, on a dichotomous variable independent coders picking randomly will get the same result half of the time simply by chance. The use of statistics like Cohen's Kappa depends on the dataset's context and the scale. In general, however, simple divergence rates may be sufficient. And unlike with survey research, there are no clear thresholds for deciding when divergence rates are too high. Generally speaking, however, anything more than the occasional discordance (i.e., divergence rates of greater than 1% or 2 %) is cause for concern.

Finalize Codebook, Protocol and Dataset

In the final phase of original legal research, the codebook and protocol should be reviewed to assure that they reflect any changes in definitions, coding conventions, or other matters that occurred during the research process. The final codebook and protocol should be sufficiently specific to

enable the exact replication of the dataset if those procedures are implemented by a separate team. The code and protocol also greatly facilitates future updates, and ensures comparability of the data collected at different times and by different teams.

Challenges and Next Steps

In the course of setting out the method above, some common areas of difficulty were highlighted in the steps in which they are typically most acute. Some challenges unique to legal measurement pervade a number of steps and therefore deserve a little more focused attention.

Comparing Law across Jurisdiction

Jurisdictions have considerable authority to create policy. This independence extends not only to the substantive features of policy but also to the way in which policies are drafted as provisions and organized as statutory code. As a result, statutory regimes often look different across states. States can accomplish identical policy positions through a variety of legal strategies and mechanisms. In some states, for example, a single comprehensive statute specifies different options for mental health care directives (Swanson et al., 2006); other states have a legal arrangement that creates a functionally equivalent policy through provisions that are scattered across probate codes, health and safety codes, and civil practice and remedies codes. In the distracted driving example, some states define the regulated activity, the fine for violations, and the manner of enforcement in one statute; others specify these details in multiple statutes. Some laws are detailed; in others a broad mandate is filled in by executive agency regulations.

It is not just that the text of provisions vary across states. Even if texts are identical, laws operate within regulatory structures and those structures differ, sometimes markedly, between jurisdictions. Failure to account for the broader legal context in which a law exists can produce

misleading comparisons. Consider, for example, a researcher interested in determining whether states in which syringe exchange programs are legal have lower incidence of HIV/AIDS. For that researcher, a reasonable way to start might be to collect all the laws that explicitly authorize syringe exchange programs. If the collection of law stopped there, however, the resulting findings would present an incomplete and inaccurate picture of the relevant state law. In some states, syringe exchange is legal under state law simply because no laws forbid it; categorizing such states as prohibiting syringe exchange because they have not explicitly authorized syringe exchange would be legally invalid. Accurately measuring how states vary with respect to the legality of syringe requires not only collecting public health laws that explicitly authorize syringe exchange but also criminal paraphernalia laws regulating possession and distribution of syringes, pharmacy statutes and regulations defining restrictions on the delivery of syringes, and criminal laws banning drug possession that could potentially apply to residue in used syringes that are possessed prior to exchange (Burris et al., 2010). These sorts of challenges highlight the need for legal expertise and at the stages of conceptualization and implementation of a study.

Tracking Change over Time

The weakness of purely cross-sectional comparisons for inferring causal effects is well known. Studies of changes in law, best with longitudinal data over many years, are much stronger. This Monograph takes as given the necessity of determining when a law was enacted or became effective and whether subsequent legislative or judicial action nullified it or modified it in ways relevant to the research – the question is which dates and how. In most states and at the federal level, except for emergency legislation, there is a lag between the date of enactment of legislation and the effective date. For evaluation research, the effective date of a law is usually the most appropriate measure for the law's onset, because many studies assume that a law cannot affect health outcomes until it legally

takes effect and is therefore enforceable. For some studies, such as those examining correlates of policy choice in legislatures or those evaluating the relationship between legislation and social norms, the date of enactment may be appropriate.

Effective dates are usually determined either by a specific clause in legislation or by the jurisdiction's legislative rules, which to the uninitiated can be quite obtuse and confusing. Identifying changes in the law over long periods can be time and labor intensive. Lexis, Westlaw, and a few specialized services like HeinOnline compile archived statutes and other legislative materials, which can make coding and validation easier. However, these historical materials tend to have more anomalies and inter-jurisdictional variations (for example, differences in years of coverage for archived statutes across states) than collections of current law. To perform historical legal research, coders typically need additional training. Effective dates tend not appear in the text of legislation or statutes and or refer to an extra-legal event, for example, "60 days after the end of the legislative session" or similar referent. StateScape's free online 50-state chart is invaluable, although only for current, not historical, practices. Retrospective research may be impossible at the local level or for state regulations because of the inaccessibility of historical records.

Amendments or other changes that occur after a law has been enacted and takes effect also require attention. Subsequent legislation that either directly amends a statute or repeals it entirely is the most obvious example of a modification. A sunset clause in a bill that nullifies it after a period of time is another source of possible change. The judiciary, too, can invalidate a statute either in whole or in part. Each such legal change must be carefully examined to determine its relevance to the research topic. Very high accuracy in coded effective dates is essential for public health law evaluation research because errors in effective dates can invalidate studies.

Reliance on Secondary Sources

At the start of an evaluation project, discovering that someone else has already produced a summary of applicable laws might seem like a windfall that obviates the need for painstaking legal research. Many advocacy and think-tank websites and publications offer authoritative-looking 50-state lists and similar compendia of "the law" on various topics. These secondary sources can be very useful for getting an overview of the law at a particular time and for use in the quality assurance process, but they are rarely sufficient sources of legal data for research projects. With few exceptions, these lists have one or more serious flaws including that they do not result from rigorous research and verification processes; lack effective dates or other indications of the period during which a law is in effect; provide data only for one point in time, which usually is not specified; lack documentation of the research process and coding conventions used to produce them (thereby preventing replication of the research); and often contain significant errors.

Another seemingly sensible shortcut is to use key informant interviews or surveys, perhaps targeting agency staff presumed to know the law they are charged with administering. Experience has repeatedly demonstrated, however, that agency staff members do not always have the right answers. Surveys or other instruments addressed to an appropriately knowledgeable official at an under-staffed agency may be completed by a subordinate and returned without review by the expert. A study by LaFond and colleagues (2000) comparing legal data compiled from key informant interviews, surveys of agency directors and staff, and rigorous, original legal research found that original legal research produced the most accurate results. Error rates for some data collected by surveying agency personnel exceeded 50 percent.

There are a few high quality sources of legal data are available. One of these, the [Alcohol Policy Information System](#) (APIS), developed by the National Institute on Alcohol Abuse and Alcoholism, relies on research attorneys to classify legal data on alcohol policy topics for all 50

states, the federal government, and the District of Columbia. Another example is the State Cancer Legislative Database Program, developed by the National Cancer Institute, which features a compendium of state statutes and resolutions on cancer-related policies. Notwithstanding APIS and a few others, obtaining legal data for an evaluation research project almost always entails conducting original legal research, which requires specialized expertise beyond familiarity with a particular policy. Especially for multi-state studies, anyone conducting research without considerable legal training and experience will be unlikely to produce reliable and accurate datasets.

Creating Composite Measures

The development of indices and scales can substantially increase the value of legal measurements. For some studies, collapsing multiple variables into summary measures is required for analytic purposes. The dataset for distracted driving laws includes well over 100 variables. No analysis would render meaningful findings with more than a few or these inserted as individual covariates. Granular coding of legal features provides the basis for classifying a jurisdiction's laws. Composite measures are then designed to represent constructs like stringency, severity, or comprehensiveness.

Developing good scales for legal measures remains particularly difficult and few flawless examples exist. Among the best are rating system such as the "Tobacco Policy Rating" based on data in the National Cancer Institute's State Cancer Legislative Database. Development of this scale bears several hallmarks that distinguish well-developed complex measures. First, the scale is firmly grounded in a causal diagram that links its components with tobacco use outcomes. Second, both legal and social science experts participate in constructing the scale. Third, a Delphi panel or other structured process is used for proposing, testing, and revising the scales. Because scales are by nature synthetic measures that encode assumptions along with observations, clarity and transparency regarding exactly how scales were constructed are essential to a study's integrity. All coding

conventions and scaling procedures should be reported. If data for the scale are to be analyzed

using statistical techniques that require interval-level data, it is important to specify how increments

between values are equalized as well as the scale's limitations (Chriqui et al., 2002, provides an

excellent example of a description)

Creating simple scales based on the number of statutes in a jurisdiction or otherwise treating

all laws as equally important produces a simple indicator that may be useful for certain purposes, but

simple counts may be misleading because of omnibus legislation and interjurisdictional variations in

codifying bills (for example, three statutes in one state may be equivalent to one statute in another or

equal to a combination of statutes and regulations in a third). In addition, some laws are likely to

have much more effect than others. For example, a researcher might identify a dozen different state

policies pertaining to child safety and use them in a study of childhood injuries. This design may

mask the reality that a single law or combination of a few laws accounts for all of the influence of

law on injuries; moreover, some of the specific laws in the scale may be inversely correlated with the

outcome measure.

Making Sense of Preemption and Federalism

The interplay among laws at the federal, state and local levels adds another dimension of complexity

to determining what the law "is" in any particular place. Sometimes the law being studied – say,

state law aimed at regulating sugar-sweetened beverages – is contingent on law at other levels of

government. At least two federal laws (the Child Nutrition Act of 1970 and the Child Nutrition and

WIC Reauthorization Act of 2004) address this issue, as do statutes or regulations in at least 34

states (Mello et al., 2008). Moreover, some municipalities have their own ordinances and many

school districts have adopted policies as well. Evaluations of state limitations on sales of sugar-

sweetened beverages in schools may be influenced by federal law and may show different results depending on whether local policies are included or ignored.

Conflicts among laws at different levels of government are generally resolved by their hierarchy, with federal law being supreme and state law trumping anything at the local level. This seemingly straightforward arrangement is more complex than it initially appears, however. In some situations, federal law and conflicting state law may both be applied in the same jurisdiction, and state authorities may decline to enforce federal law. For example, since California decriminalized medical marijuana in 1996, state authorities have not enforced federal drug laws against dispensers or users of medical marijuana. Federal officers in California continue to apply U.S. drug laws. The varieties of preemption – a term that encompasses different arrangements of authority between levels of government – present some of the most interesting and complex legal questions. Here again, the need for guidance from a legal expert is typically required.

Research Design, Assumptions and Inferences

Although this monograph is devoted to measurement, a few comments about inference and analysis bear mentioning. For empirical law evaluation studies there is an analog to the "If a tree falls in the woods" question: what if a law exists but no one follows it? Absence of compliance by relevant populations, varying enforcement by police or administrative personnel, unwillingness of prosecutors to bring charges, differing interpretations across jurisdictions, and inadequate funding for implementation or enforcement can create misleading evidence about the relationship between a law "on the books" and health outcomes "on the street". While the existence of a law supports an inference that it is being enforced, the possibility of non-enforcement or inconsistent enforcement can make a critical difference to compliance and measures of a policy's effect (see Wagenaar & Wolfson, 1994). Laws often have both deterrent and norms-shaping effects and the latter can occur

independently of enforcement. Research designs that include measures of enforcement and compliance better isolate the relative effects of different policies as written.

Even when a mandate is clearly stated, implementation may not necessarily follow. Particularly in studies of policies that require monetary appropriations, such as laws that create systems for providing treatment or give citizens a right to receive governmental services, another factor looms: whether adequate funds are available. Legislatures are more apt to pass authorizing legislation for a program than to pass appropriations to fund it. Executive agencies – which are charged primarily with enforcing laws – may divert or delay funds with little recourse for policy-makers.

Legal texts are measured to determine government-sanctioned policy. This information is seldom sufficient by itself to sustain inferences about the effectiveness or harm of specific policies. This monograph demonstrated that law is measurable like other phenomena. If you are reading this text, it is likely apparent that law is also very important to health. There often remains, of course, the need to measure a range of confounding factors including implementation of policy and other features of jurisdictions to understand how law and health relate.

Conclusion

This monograph describes a method for creating measurements of law and policy for use in public health law evaluations. The aim of the method is to generate scientifically defensible data that reflect in quantitative forms how laws vary over space and time. In addition to adding rigor to the study of legal texts, it provides a method for increasing the efficiency of studying legal change over time, a key requirement for evaluation research. While these methods are still developing and specific uniform standards and best practices are evolving, this approach is an essential if underappreciated element of public health law research. Implementing the suggestions and procedures offered in this

monograph can help advance the field of evaluation research by increasing the utility and accuracy of research that uses legal data, ultimately improving public policy and its effectiveness in achieving important goals in advancing population health and well-being. By blending the knowledge and skills of social scientists with that of legal experts, scholars can produce more accurate and more useful policy evaluations. As the field continues to advance, adopting minimum measurement standards along the lines suggested here will elevate the threshold of acceptable quality for designing, funding, and conducting evaluations of public policies embodied in law.

List of Figures and Tables

References

Burris, S., Anderson, E. D., Craigg, A., Davis, C. S. & Case, P. (2010). Racial disparities in injection-related HIV: A case study of toxic law. *Temple Law Review, 82*(5), 1263-1302.

Chriqui, J. F., Frosh, M., Brownson, R. C. et al. (2002). Application of a rating system to state clean indoor air laws. *Tobacco Control, 11*(1), 26-34.

Ibrahim, J. K., Anderson, E. D., Burris, S. C. & Wagenaar, A. C. (2011). State laws restricting driver use of mobile communications devices: "Distracted-driving" provisions, 1992-2010. *American Journal of Preventive Medicine, 40*(6), 659-665.

LaFond, C., Toomey, T. L., Rothstein, C., Wagenaar, A. C. & Manning, W. (2000). Policy evaluation research: Measuring the independent variables. *Evaluation Review, 24*(1), 92-101.

Mello, M. M., Pomeranz, J. & Moran, P. (2008). The interplay of public health law and industry self-regulation: The case of sugar-sweetened beverage sales in schools. *American Journal of Public Health, 98*(4), 595-604.

National Institute on Alcohol Abuse and Alcoholism. (2011). Underage drinking: Underage purchase of alcohol. *Alcohol Policy Information System (APIS).* Retrieved November 1, 2012, from http://alcoholpolicy.niaaa.nih.gov/Underage_Purchase_of_Alcohol.html

Plano, J. C., Riggs, R. E. & Robin, H. S. (1982). *The dictionary of political analysis.* Santa Barbara, CA: ABC-CLIO.

Preusser, D. & Tison, J. (2007). GDL then and now. *Journal of Safety Research, 38*(2), 159-163.

Singer, J. W. (1984). The player and the cards: Nihilism and legal theory. *Yale Law Journal, 94*(1), 1-70.

Swanson, J. W., McCrary, S. V., Swartz, M. S., Elbogen, E. B. & Van Dorn, R. A. (2006). Superseding psychiatric advance directives: Ethical and legal considerations. *Journal of the American Academy of Psychiatry and the Law, 34*(3), 385-394.

Wagenaar, A. C. (1983). Raising the legal drinking age in Maine: Impact on traffic accidents among young drivers. *International Journal of the Addictions, 18*(3), 365-377.

Wagenaar, A. C., Erickson, D. J., Harwood, E. M. & O'Malley, P. M. (2006). Effects of state coalitions to reduce underage drinking: A national evaluation. *American Journal of Preventive Medicine, 31*(4), 307-315.

Wagenaar, A. C. & Wolfson, M. (1994). Enforcement of the legal minimum drinking age in the United States. *Journal of Public Health Policy, 15*(1), 37-53.

Cases

Golden Gate Restaurant Ass'n v. City and County of San Francisco, 558 F.3d 1000 (2009) (en banc. 9th Cir., Smith, J., dissenting).

Medtronic, Inc. v. Lohr, 518 U.S. 470, 116 S.Ct. 2240 (1996).

New State Ice Co. v. Liebmann, 1932.

Please cite this document as:

Anderson, E., Tremper, C., Thomas, S. & Wagenaar, A. (2012). Measuring statutory law and regulations for empirical research. *PHLR Methods Monograph Series.*