Temple University
Beasley School of Law

Center for
Public Health Law
Research

MEASURING STATUTORY LAW AND REGULATIONS FOR EMPIRICAL RESEARCH

Evan D. Anderson, JD, PhD

Advanced Senior Lecturer, School of Nursing, University of Pennsylvania

Sue Thomas, PhD

Senior Research Scientist, Pacific Institute for Research and Evaluation (PIRE); Director of PIRE-Santa Cruz

Ryan D. Treffers, JD

Research Scientist

Alexander C. Wagenaar, PhD

Research Professor, Emory University Rollins School of Public Health; Professor Emeritus, University of Florida College of Medicine

A Methods Monograph for the Center for Public Health Law Research Temple University Beasley School of Law

OCTOBER 2023

MEASURING STATUTORY LAW AND REGULATIONS FOR EMPIRICAL RESEARCH

Evan D. Anderson

Sue Thomas

Ryan D. Treffers

Alexander C. Wagenaar

Summary

Effectively studying the relationship between law and population health requires variation in both the law and health outcomes over space as well as time, and reliable and valid methods for capturing variation and representing it in forms that allow comparison and analyses. 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 it relies on careful and consistent observation of the apparent features of legal texts. This approach produces data that is replicable through a process that is transparent. Transparency and replicability are essential attributes of scientifically defensible data.

There are many challenges in measuring law. Relevant legal texts can be hard to find and rife with ambiguous and conflicting meanings. Formulating reliable and valid ways of reducing complex bodies of law into numeric data can be difficult. There are also cultural and logistical hurdles to forming teams that combine legal scholars and social scientists, blending a full range of legal and scientific expertise. These challenges can be overcome by a methodical process of design, data collection, and analysis that adheres to scientific standards. Steps include the careful delineation of the scientific and legal questions to be addressed and the scope of the research; the iterative development and refinement of coding schemes; an intense focus on quality control; and the production of a transparent research protocol and codebook to accompany the resulting legal dataset.

Learning Objectives

- Describe the concept and value of systematic measurement of law.
- List the steps in a systematic process for measuring law and creating a numeric legal dataset.
- Identify important logisticl challenges in legal measurement projects.

Measuring law, as the term is used here, means determining dimensions or components of an area of law relevant for particular research studies 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 in both quantitative and qualitative social science research. Although few of these techniques are conceptually challenging, their application to the law can be difficult. The bulk of this chapter is devoted to providing a step-by-step guide for reflecting variation in laws across time, space, or both. To provide context and make clear the importance of these steps, the chapter 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 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. This policy experimentation is frequently a necessary step to identify 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 clear that restrictive laws saved lives. Now most states have adopted similarly restrictive laws, though important variation in the law across states remains. Teen crash rates have declined continuously since, marking GDL laws as one of the great modern examples of how policy innovation, rigorous evaluation, and evidence-based dissemination 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, hospital records, and other administrative data. But how does one measure and quantify something textual like the law? The answer is not too difficult using one or more

experienced legal researchers, careful consideration of a handful of persistent sources of error, and attention to the usual basic principles of science. As any qualitative researcher can attest, characteristics of texts are observable and these observations converted into numeric indicators, or *coded* as it is often called. At its core, the task of coding law is not altogether different from coding an open-ended interview transcript. In each instance, researchers strive to measure the features of the texts in ways that are consistent with scientific standards of reliability and validity.

There are two primary sources of difficulty in the process of 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. The method for measuring law offered in this chapter is more like observation than traditional legal interpretation, because the method emphasizes observing the text (what the law says) and minimizing interpretation (what the law means). 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 in public health law evaluations. Researchers' decisions that shape the laws that are collected and how they are understood increase 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 correct legal texts to collect and examine. Unlike the qualitative researcher who creates a file of transcripts for coding by, for example, interviewing a 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 relevant legal provisions, researchers also must consider how provisions interact within a broader legal framework. In most instances, lawyers are needed to help determine which 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 usually necessary.

Although 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 described here provides steps for operationalizing those principles and practices in legal research. Some standard scientific practices – such as the creation of a detailed 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 idiosyncratic to each practicing lawyer, and closely guarded as the lawyer's stock in trade. But our experience is that legal researchers can and do quickly internalize these rules and procedures drawn from scientific research; 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. An 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, 1983b). 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 will typically exclude legal information in these laws pertinent to the bus-driver ban. 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 major characteristics that vary in ways related to health. When conducted with appropriate rigor and transparency, legal mapping studies yield datasets that can be used in empirical evaluations examining many different hypotheses. These datasets are also the basis of ongoing policy surveillance (Burris, Hitchcock, Ibrahim, Penn, & Ramanathan, 2016). This chapter focuses on measuring law for hypothesis testing, but the same principles and practices are, with exceptions, required for legal mapping studies.

The Process for Measuring Law

As is the case in all scientific research, questions of interest and 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 used as independent or dependent variables, and the research design chosen for the study. Notwithstanding these differences, the process for measuring law depicted in Figure 11.1 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. The following sections describe each step in the figure except the first, which is addressed in other chapters.

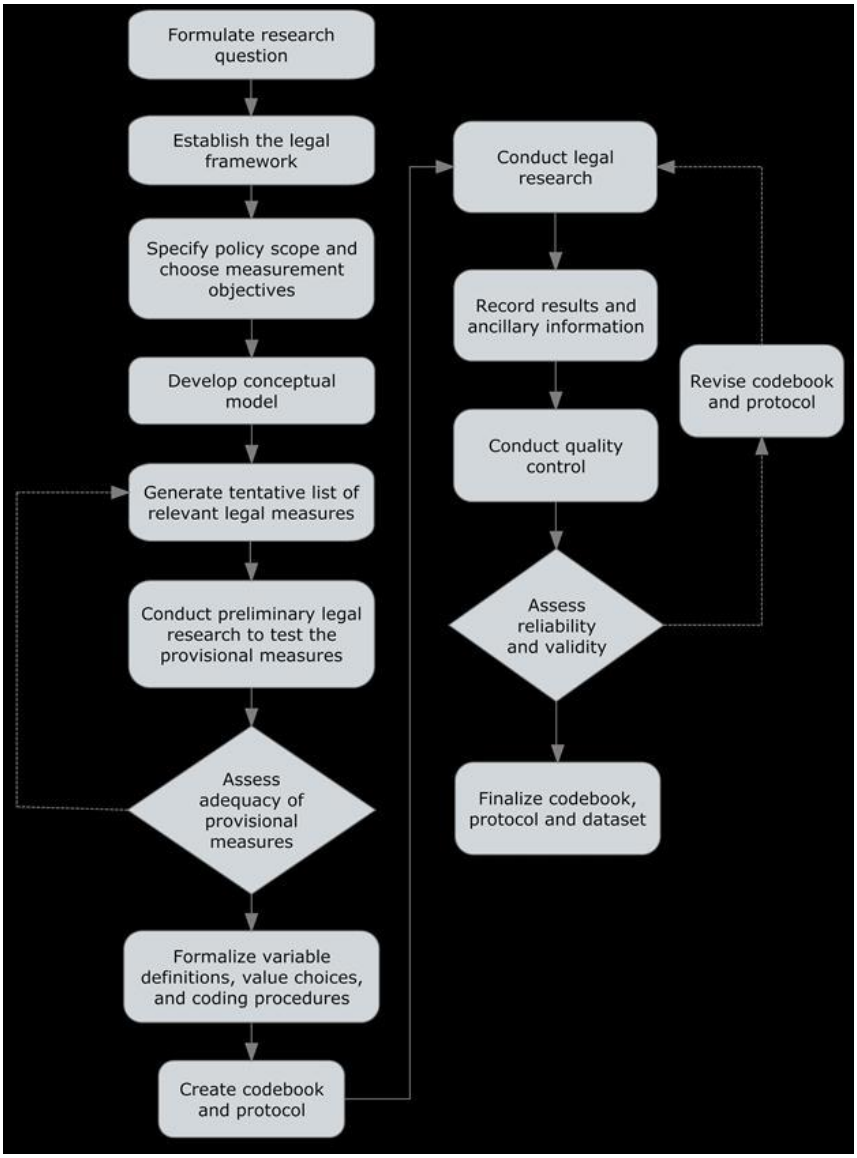

Figure 11.1. Process for Measuring Law.

Adapted from: Tremper, C., Thomas, S. & Wagenaar, A. C., (2010) Measuring Law for Evaluation Research, Evaluation Review, 34(3), 242-266.

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, if those drivers expect that injuring someone while using a device is likely to result in a successful lawsuit against them. Or it could be that law regulating insurance or employer liability is a plausible factor affecting 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 levels too (for example, 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 11.1 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) is equally as important to measure as statutes and ordinances.

	Federal	*State*	*Local*
Legislative	Statutes	Statutes	Ordinances
Executive	Regulations Executive orders Administrative judgments	Regulations Executive orders Administrative Judgments	Regulations Executive 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

Table 11.1. Types of Law by Level and Source.

Note: Dash indicates the absence of a law.

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.

SPECIFYING THE SCOPE AND CHOOSING MEASUREMENT OBJECTIVES

To conserve resources and ensure that 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 laws prohibiting driver activity with mobile communication devices have evolved over a 10-year period in regard to covered activity, devices, and classes of drivers. The preceding sentence is deliberately vague on one point: What is the right unit of analysis? Is it state statutes, federal motor vehicle regulations, local ordinances? The answer depends on the purpose of the measurement.

In most evaluations, the unit of analysis is a rule that applies to a certain population of organizations or individuals. That rule is defined in one more legal provisions (perhaps across bills and statutes). Each record – or row in a two-dimensional table – represents observed and recorded characteristics of that rule during a designated time period or at a specified point in time. Measuring these characteristics sometimes involves multiple provisions set out in numerous statutes. For example, one state might classify texting while driving as a Class A offense while another might classify it as a simple misdemeanor. Understanding the fines associated with those offenses – which might be important to understanding their effect – would then require reading the statute that defines those categories of offenses and associated penalties. A dataset reflecting variation in law for all US states and the District of Columbia would therefore have 51 rows excluding headings (and leaving aside the issue of data encompassing changes in the policy over time). The columns in the table would represent variables and the characteristics of state law those variables describe. Although that is generally 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 a single state.

Periodic discussion by the research team (including scientists and lawyers) clarifying the unit of analysis ensures that different ways of organizing the search for legal data 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 shorthand 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.

The primary goal is objective measurement of observable features in the law. The design of the measurement protocols and creation of a legal dataset should seek to eliminate both legal and conceptual interpretation during the data collection and coding process, with the necessary exception of a precise definition of the laws of interest and what legal terms used to represent it. It is essential to avoid conflating primary observations with subsequent, secondary analytic steps. For example, a research plan might be heading toward a classification of distracted driving laws as

weak and strong, or broad or narrow, or some other underlying theory-based conceptual differentiation. Assuming that such a second phase classification system can be devised and defended conceptually and empirically, such classifications are based on the previously observed and consistently coded observable characteristics of the legal text. Thus, in the distracted driving example, the first phase coding of observable provisions should start with a broad definition of what might be construed as "distracted driving," so that nuanced differences in definitions across states might be captured; such definitional differences might be important later when observed provisions are combined in various ways to represent underlying theory-based constructs. First one observes that the fine for violation in state A is $100 and in state B is $1000. It is a second analytic and conceptual step, independent from the coding of observed provisions, to classify A as a weak state and B as a strong state based on size of the fine (Burris, 2017). A third step is conceptually grappling with whether the larger fine is 10 times "worse" or 10 times more important than the smaller fine. Separating the first phase of coding of the observable characteristics of the text from later classification, scale building, or other forms of measurement development makes the development of measures more reliable and the analyses more transparent and reproducible – a *sine qua non* of good science.

DEVELOPING A CAUSAL DIAGRAM

Having defined the scope of laws to be collected, the next question is which provisions of the laws are to be measured. Causal diagrams – as described in more detail in Chapter 10 – are valuable tools for this purpose. By forcing researchers to identify and clarify plausible links between law and health outcomes, causal diagrams help flush out the legal inputs relevant to the research question. 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 relevance. Also important in that scenario are provisions specifying fines or other penalties such as suspensions for drivers on learner's permits. A causal diagram based in part or entirely on theory positing different mechanisms would suggest measurement of different legal provisions.

GENERATING A TENTATIVE LIST OF RELEVANT LEGAL MEASURES

In hypothesis-driven research, causal diagrams limit the laws to be studied on the basis of theories about how those legal inputs relate to some other outcome. The legal inputs themselves are often until this point understood as concepts such as manner of enforcement, scope, and severity. To capture how these conceptual legal inputs vary, the important components of each must be identified – specific legal provisions that must be coded to later be combined to best measure a theory-derived concept. Operationally, this is the point in the process to start creating variables, one for each specific provision coded. 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, cell phones, personal digital assistants, laptop computers, other) subject to the law.

CONDUCTING PRELIMINARY LEGAL RESEARCH

Whether the tentative measures align with relevant variation in extant 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, examining a random sample of jurisdictions increases the value of a preliminary review. If the structure or operation of law on the topic of interest is known to vary systematically across jurisdictions in terms of a small number of major characteristics, purposely sampling jurisdictions across such strata is best (eg, urban versus rural states). 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 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 legal provisions being studied and the resources needed to systematically collect and analyze.

ASSESSING 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 important dimensions of variation in the law and how to elicit 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 (that is, at ages 16, 17, 18, 19, and above) and laws that apply to all new drivers regardless of age. It might also become apparent that a handful of exceptions reduce the scope or enforceability of prohibitions, such as exceptions for hands-free device use or exceptions in statutes banning texting that permit typing keys to start or end a call.

FORMALIZING 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 for empirical research employs precisely defined and documented procedures used to represent each legal provision. From an operational perspective, this means first identifying all relevant ways that laws vary and then finding numeric schemes (or preliminary textual ones such as “yes” and “no”) to capture that variation. Although it is counterintuitive and may strike legally trained staff as odd at first, coding questions, and the

variables they define, must be precise and strive to eliminate human judgment. Precision means each item coded should be an elemental record of observation of a specific provision: is the provision present (yes/no)? Coding questions requiring interpretation undermine reliability because of ambiguities in legal interpretation and innate differences between coders. Consider a coding question asking whether a law prohibiting communication on wireless telephones by drivers applies to talking over an internet connection through a headset attached to a laptop. Different coders could reach entirely defensible but different conclusions depending on how they interpret the operative terms and ultimate meaning of the rule (that is, "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 an important safety valve, especially in early stages of coding. The "not sure" category gives coders an option for handling ambiguity that may or may not be resolved by additional research or subsequent developments such as a court ruling during the course of the study. Regular review of "not sure" cases by the entire research team often leads to revision of the coding protocol, increasing reliability, precision, and comprehensiveness. Sometimes a "not sure" or "unsettled" code will remain in the final dataset so that those records can be excluded from analyses or addressed separately in a sub-study.

The use of blanket dichotomous "law or no law" variables to code the presence or absence of a particular type of law is useful only in rare situations of limited research questions and study designs (Burris, 2017). This is distinct from the recommended practice of using simple dichotomous measures for each specific relevant provision or feature within a law. Without careful specification of provisions, dichotomous variables representing overall presence/absence of a law can easily obscure a great deal of legal variability and hence limit the value of a dataset for evaluating legal effects. Consider the example of underage alcohol purchase dataset from the National Institute of Alcohol Abuse and Alcoholism's Alcohol Policy Information System (2021). A dichotomous law or 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 most of the laws on the topic are clear and simple. Several jurisdictions, though, permit purchases by youths in some situations, such as acting in conjunction with law enforcement, or drinking in some situations, such as in the presence of parents. New York and Delaware do not prohibit underage purchase, but they do prohibit obtaining alcohol in connection with making a false statement. Rather than trying to shoehorn these laws into a dichotomous law or 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 should be separated into four dichotomous variables, each measuring one dimension, an approach adhering to the principle that a coded variable should always concern a single distinct, unitary attribute.

One of the preliminary measures identified 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 "Does the law prohibit texting?" and "Does the law prohibit talking?" Unidimensional variables can later easily be combined to reflect instances when both activities are prohibited or combined with other variables defining the existence of exceptions (for example, exception for hands-free use yes/no) to represent 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 (for example, low, medium, high). Interval variables are a special type of ordinal variable in which the difference between attributes is meaningful and assumed to be equal across the distribution (for example, 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 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 law.

Continuous variables, which can take any numeric value (for example, temperature in Celsius), provide even more analytic benefits than ordinal variables. Attributes such as length of jail terms, maximum fines imposed, appropriation amounts, and legal thresholds (for example, alcohol-impaired driving in terms of blood alcohol content), are often measurable with continuous variables. Careful definitions and coding protocols can also produce continuous variables based on a law's non-numerical features. For example, researchers might use primary observations of the length of jail sentence, the magnitude of fine, and the availability of defenses to create a composite measure of "stringency." Such composite measures are designed to represent concepts that come from a theory that underlies the research. It is critical for reliable and valid measurement to maintain a careful distinction between the "phase 1" coding – finding all instances of relevant laws and accurately coding the elemental observable provisions of each law – and "phase 2" measurement development, which combines the coded variables created in phase 1 to create measures of concepts derived from a theory underlying a particular study.

CREATING A 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 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 for coding laws. Elements include a description of the study; 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 laws); and details about the data (columns, text, numeric, Boolean).

A codebook alone is inadequate for any but the simplest study because critical decisions about coding conventions and procedures for legal data are rarely apparent from examination of a codebook alone. A comprehensive protocol includes information about how the laws were collected (for example, exact legal text databases searched, exact search terms and syntax used); 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 social scientists and legal researchers, up until this point, most activity should have been a joint effort with a lot 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 all of the coders. 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 such as reading provisions in context by carefully locating and considering legal definitions of all operative terms. Relevant provisions are often located by keyword searches; without explicit instruction some legal researchers will not inspect other provisions in the same part of the statutory code. The use of statutory tables of contents is not always emphasized when training law students but is indispensable for making sense of statutory schemes, especially across jurisdictions. Being observant for court decisions that influence status of the law is also essential. Key points of emphasis in discussions with attorney or law student coders 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 a high degree of clarity to guide the location, organization, and eventual coding of legal texts.

CONDUCTING 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 Fastcase and publicly available online sources such as the Library of

Congress website (www.congress.gov) for federal material. The commercial service StateNet (www.statenet.com) offers access to recently enacted statutes and pending bills. An excellent resource for studying legislative activity is LexisNexis Advanced Legislative Service, which catalogues bills eventually adopted. The returns 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 most instances. For large projects, and given the continued development and expansion of features and products offered by the many legal services available, 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 many jurisdictions publish their own materials directly online or through local law publishers and repositories including 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 can become unwieldy and require extensive investment of researcher time to search multiple databases and, on some occasions, query local governments directly (Sanner, Grant, Walter-McCabe, & Silverman, 2021). As Natural Language Processing, a branch of artificial intelligence, improves and becomes more widely available, searching many local governmental websites may become routine.

RECORDING RESULTS AND ANCILLARY INFORMATION

Well-designed data-collection software for gathering legal source material is crucial for a smooth process. An adequate data-collection system for coding the law stores much more than the final resulting codes for each variable. The system must retain ancillary information supporting the coding, typically in the form of extended blocks of legal text. The software should allow for changes to variables and codes as the project progresses, knowledge is gained, and previous decisions are updated, as well as keep a record of such changes. At a minimum, all relevant statutory, regulatory, and case law citations must be collected and recorded. Because coding of a single variable may depend 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 notes.

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, and software specifically designed for legal coding and measurement are also available, for example MonQcle (Center for Public Health

Law Research, 2022). 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, not only to the current research team but also to future users of the dataset. Maintaining records of each coding decision is essential. Although coding decisions should always have a solid basis in the observable features of the legal text, there may remain instances when the text of a law is explicit but important extrinsic information exists that changes the effect of the legal text. Law enforcement agencies may choose not to enforce a law through an organizational policy that is not included in the data such as, for example, widespread refusal to prosecute cannabis possession offenses in many cities. 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 reduce inaccuracies and confusion during later analyses.

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 objects provides coders with many opportunities to make mistakes. This sort of error can be reduced through the use of coding platforms that integrate codebooks, datasheets, and the legal text. Data entry forms designed in Microsoft Access or MonQcle, for example, enable researchers to create templates in which the legal text to be coded is visible next to the 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 not only must be thoroughly versed in conventions adopted at the design stage of the research project, they also must 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.

QUALITY CONTROL

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 errors. It is important to have two researchers redundantly and independently compile and code the laws.

Comparing the two resulting datasets for inconsistencies reduces the likelihood of undetected random or careless error and points to possible areas of underspecified coding conventions. Such double-coding also provides a direct assessment of inter-coder 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 the coder collects or codes law; likewise, coders often share characteristics that predispose them to similar patterns of observation or analysis, which could bias resulting data. A final important step to address these concerns is to have a third legal researcher who is totally naïve 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 a rate that adjusts for the probability of randomly selecting the correct answer. The statistical metric for assessing inter-rater reliability, 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. Unlike with survey research, there are no clear thresholds for deciding when divergence rates are too high. Generally speaking, anything more than the occasional discordance (that is, divergence rates of greater than 1% or 2%) is cause for concern.

After assessing inter-coder reliability, the codebook and protocol specifications typically are revised to improve coding of some of the measures. The research team then cycles back to conducting legal research with the newly revised protocols, repeating the quality control procedures and assessing reliability. The research team advances only when the highest practically achievable levels of reliability are attained, which often requires multiple rounds of revisions and testing.

A complementary approach to quality control avoids routine double-coding of all records by random sampling codes and double coding those sampled. This is done early in the process to identify ambiguous coding instructions or lack of clear specification of the elemental observable legal provisions the coding is intended to include. Rounds of revisions to the coding instructions are made and evaluated via double coding, until error rates are acceptably low. In this approach, a more-experienced coder or senior attorney is typically used for the second coding, to better identify errors by regular coding staff, and better point to effective ways for improving the coding instructions. At that point, final production coding is completed; achieved error rates are always documented and published with the final dataset.

When concluding the legal data collection and coding phase, the codebook and protocol documents should be carefully reviewed to ensure that they reflect all changes in definitions, coding conventions, or other matters that occurred during the legal data collection and coding process. The final codebook and protocol should be sufficiently specific to enable exact replication of the dataset if those procedures are implemented by a separate team at a later time. The codebook and protocol facilitate future updates and ensure comparability of the data collected at different times and by different teams. Norms of scientific publishing require final codebooks and

protocol documents be publicly available to other researchers for verification and replication studies.

Challenges and Next Steps

Challenges unique to measuring law for empirical research arise in the process we have described and deserve additional discussion.

COMPARING LAW ACROSS JURISDICTIONS

Jurisdictions have considerable authority to create law. This independence extends not only to the substantive features of law but also to the way in which policies are drafted as provisions and organized as statutory code. As a result, statutory regimes vary considerably across states and even more drastically across countries (Kavanagh, Meier, Pillinger, Huffstetler, & Burris, 2020). 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, McCrary, Swartz, Elbogen, & Van Dorn, 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 varies 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 where 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 syringes requires collecting not only 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 apply to residue in used syringes that are possessed prior to exchange (Fernández-Viña, Prood, Herpolsheimer, Waimberg & Burris, 2020). Challenges like this highlight the need for project teams to incorporate legal expertise early at the stages of conceptualization and design of the study, as well as during implementation.

TRACKING CHANGE OVER TIME

The weakness of purely cross-sectional comparisons for inferring causal effects is well known. Evaluations of changes in law, best with longitudinal data over many years, are much stronger. This chapter 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 date it legally takes effect. 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 or assumes any anticipatory effects of a law before it legally takes effect are small and conceptually distinct. For some studies, such as those examining correlates of policy choice in legislatures or those evaluating the relationship between legislation, public attention and social norms, the date of legislative passage may be more 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 abstruse and confusing. Identifying changes in the law over long periods can be time and labor intensive. Lexis, Westlaw, and a few specialized services such as 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 interjurisdictional 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 often do not appear in the text of legislation or statutes, or the legislation might refer to an extra-legal event, for example "60 days after the end of the legislative session". StateScape's free online 50-state chart is invaluable (https://www.statescape.com/resources/legislative/bill-effective-dates/), although only for current, not historical, practices. Retrospective research in some cases 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 important source of possible change. The judiciary, too, can invalidate a statute either in whole or in part. Such legal changes must be examined carefully to determine relevance to the research topic. Very high accuracy in coded effective dates is essential for legal evaluation research because errors in effective dates can invalidate studies.

RELIANCE ON SECONDARY SOURCES

At the start of a public health law 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 fifty-state lists and similar compendia of "the law." These secondary sources can be 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 rigorously defined protocols 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 often is not specified; lack documentation of the research process and coding conventions used to produce them (preventing replication); 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) found that original legal research produced more accurate results than key informant interviews or surveys of agency directors and staff. Error rates for some data collected by surveying agency personnel exceeded 50%.

There are few high-quality sources of legal data available. One of these, the Alcohol Policy Information System (APIS, alcoholpolicy.niaaa.nih.gov), developed by the National Institute on Alcohol Abuse and Alcoholism, relies on research attorneys to classify legal data on alcohol and recreational cannabis policy topics for all 50 states, the federal government, and the District of Columbia. The Prescription Drug Abuse Policy System provides data on state laws related to prescription drug abuse (pdaps.org). The LawAtlas portal (lawatlas.org) provides data on wide range of public health laws and allows users to interact with the data by creating custom charts and maps. Notwithstanding these examples, obtaining accurate legal data for an evaluation research project almost always entails conducting original legal research, which requires specialized legal and conceptual expertise beyond simple familiarity with a particular policy. Especially for multistate studies, anyone conducting research without considerable legal training and experience will seldom produce sufficiently reliable and accurate datasets.

CREATING COMPOSITE MEASURES

After reliable coding of observable legal provisions, individual provision indicators are often then combined in the construction of indices and scales thereby creating measures of theory-based constructs. Novice researchers can damage the reliability of the original legal coding by having coders make judgements to code levels of higher-level constructs such as breadth, stringency, or strength of laws. The objective is to always have lawyers and law students as coders tasked only with reliably and accurately coding observable provisions or characteristics of the law. Then, the scientists, based on theory and proposed study hypotheses, create measures (ranks, scales, combinations, continuous variables) that match the concepts desired in a way that is suitable for

the planned statistical analyses. Thus, the scientists take the elemental codes produced in phase 1 and build measures in phase 2. Imagine a law is reliably coded on a dozen provisions each as present or absent. Based on theory, the scientific team then designs a composite measure: if at minimum the law has provisions A and C, then it gets score 1=low, if it has three or more other provisions from the list as well along with A and C then score 2=medium, and if in addition to A and C the law has five or more of the other provisions then score 3=high. The legally trained coders do not make judgements on whether a law is low or high on such a dimension; they code the elemental provisions, and the scientist team specify combinations (summing, ranking, weighting) to create measures of concepts needed to test a particular theory or hypothesis.

Adding such composite variables to datasets often increases the value of empirical legal datasets to other users. The dataset for distracted driving laws (https://lawatlas.org/datasets/distracted-driving-1470663668) includes well over a hundred variables. Granular coding of legal features into separate variables provides the basis for classifying a jurisdiction's laws in many ways, matching divergent theories on the law's mechanisms of effect. Starting with granular coding not only enhances reliability and accuracy of legal coding. It also allows much more diverse sets of analyses later.

Developing good scales for legal measures remains particularly difficult, and few flawless examples exist (Moxham-Hall & Ritter, 2017). Among the best are rating systems such as Naimi and colleagues' (2014) alcohol policy environment scale. The development of the tobacco policy scale by Chriqui and colleagues (2002) 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 collaborated in constructing the scale. Third, a Delphi panel or other structured process was 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 are constructed are essential to a study's integrity. All coding conventions and scaling procedures as well as the scale's limitations must be documented and reported. If data for the scale are to be analyzed using statistical techniques that require interval-level data, it is important to in addition specify how increments between scale values are equalized.

Creating simple scales based on the number of statutes in a jurisdiction or otherwise treating all laws as equally important 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 laws 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 very 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. How a

scale is constructed is necessarily shaped by the underlying theoretical model, and a scale that is appropriate and useful for one purpose may be quite inappropriate for a different purpose.

RELIABILITY AND VALIDITY

The design and creation of reliable and valid empirical legal datasets requires following the same principles of measurement developed in the social sciences over the past half century (Nunnally, 1978). We have already emphasized the need to clearly specify simple observable provisions in legal texts, avoiding the need for coders to make conceptual or interpretive judgements, to achieve primary legal coding with high levels of inter-rater reliability. Following the same coding procedures on the same law at different times should also produce the same result – stability reliability. Subtly changing norms of legal interpretation over time might threaten stability reliability if coding procedures are not sufficiently specific. After construction of the primary legal dataset, the elemental variables are typically combined to create measures of higher-level constructs, and those composite measures should have high levels of internal consistency reliability – items purportedly measuring the same underlying construct should be highly correlated with each other and with the overall scale score.

Reliability is about getting the same result each time a procedure is followed. It is a prerequisite for a valid measure, but not sufficient. Validity is about making sure the procedures produce a measure that actually measures the concept intended. *Face validity* is simply whether the measurement procedures "on the face of it" appear to reflect the intended construct. More usefully, *content validity* addresses whether the procedures produce a measure that reflects the full domain of intended dimensions. Consider an intended measure of law's overall "strength" that only includes fines and jail time, missing dimensions of "strength" that reflect celerity (speed of penalty implementation), another key dimension from deterrence theory. At this point in measurement development, one also assesses *construct validity* – does the measure correlate well with other measures the theory states it should be correlated with (convergent validity), and is it not correlated with other measures theory states it should not be correlated with (discriminant validity). Finally, *criterion validity* is whether the constructed measure is highly correlated with a "gold standard" – an available measure widely accepted as the most accurate and valid. Unfortunately, such gold standard measures are rarely available in legal epidemiology. But the notion is still helpful – comparing new measures of law with existing measures, to help understand validity and accuracy of each.

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, Pomeranz, & Moran, 2008). Moreover, some municipalities have their own ordinances, and many school districts have adopted rules 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 law is included or ignored.

Conflicts among laws at different levels of government generally are 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 state law may conflict. For example, as of January 2022, 18 states have legalized adult recreational cannabis use, which remains illegal under federal law. Although federal officials could enforce federal law in these states, they have chosen not do so as a matter of discretion. 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 collaboration with legal experts is essential.

RESEARCH DESIGN, ASSUMPTIONS, AND INFERENCES

Although this chapter is devoted to measurement, a few comments about inference and analysis bear mentioning. For legal epidemiology studies there is an analog to the "If a tree falls in the forest" question: What if a law exists but no one follows it? Low 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 can create misleading evidence about whether the law "on the books" could produce different health outcomes "on the street." Although 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 assessments of a law'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 effects of laws as written.

Even when a mandate is clearly stated, implementation may not necessarily follow. Particularly in studies of laws that require resources for implementation – such as those 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 implementing and enforcing laws, may divert or delay funds with little recourse for policy makers. This chapter demonstrates that law is measurable for scientific study like other phenomena. Collecting and coding legal texts, however, is only the start. One also needs to measure a range of implementation factors and other features of jurisdictions to fully understand how law and health relate.

Conclusion

This chapter describes methods for measuring law and creating datasets for use in legal epidemiology research. The aim is to generate scientifically consistent and defensible measures 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. Although these legal coding methods are still developing and uniform standards and best practices are evolving, measurement and coding issues are an essential if underappreciated element of public health law research. Implementing the procedures offered in this chapter advances the field of evaluation research by increasing the utility and accuracy of research using legal data, ultimately improving public policy and its effectiveness in achieving important goals advancing population health and well-being. By blending the knowledge and skills of social, statistical and health scientists with those 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.

Further Reading

Burris, S., Hitchcock, L., Ibrahim, J. K., Penn, M., & Ramanathan, T. (2016). Policy surveillance: A vital public health practice comes of age. *Journal of Health Politics, Policy & Law, 41*(6), 1151-1167.

Kavanagh, M. M., Meier, B. M., Pillinger, M., Huffstetler, H., & Burris, S. (2020). Global policy surveillance: Creating and using comparative national data on health law and policy. *American Journal of Public Health, 110*(12), 1805-1810.

Policy Surveillance Program. (2022). Learning library. Retrieved May 16, 2022, from http://www.lawatlas.org/page/lawatlas-learning-library

References

Burris, S. (2017). Theory and methods in comparative drug and alcohol policy research: Response to a review of the literature. *International Journal of Drug Policy, 41*, 126-131.

Fernández-Viña, M. H., Prood, N. E., Herpolsheimer, A., Waimberg, J., & Burris, S. (2020). State laws governing syringe services programs and participant syringe possession, 2014-2019. *Public Health Reports, 135*(1_suppl), 128S-137S.

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. (2021). Underage drinking: Underage purchase of alcohol. *Alcohol Policy Information System (APIS)*. Retrieved February 15, 2022, from https://alcoholpolicy.niaaa.nih.gov/apis-policy-topics/underage-purchase-of-alcohol/43.

New State Ice Co. v. Liebmann, 285 U.S. 262 (1932).

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

Sanner, L., Grant, S., Walter-McCabe, H., & Silverman, R. D. (2021). the challenges of conducting intrastate policy surveillance: a methods note on county and city laws in Indiana. *American Journal of Public Health, 111*(6), 1095-1098.

Tremper, C., Thomas, S., & Wagenaar, A. C. (2010). Measuring law for evaluation research. *Evaluation Review, 34*(3), 242–266.

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.