



State Health RegData 1.0: A Quantification of State Healthcare Regulations

Kofi Ampaabeng and Elise Amez-Droz

October 2020

In this policy brief, we introduce State Health RegData, a dataset that contains the volume and characteristics of healthcare regulatory restrictions in US states. State Health RegData is part of a suite of datasets, called Health RegData, that quantifies healthcare regulations across the United States. The purpose of these datasets and their accompanying policy briefs is to shed light on the healthcare regulatory system and to provide researchers with tools to examine the impact of regulations on the practice, delivery, and outcomes of medical care.

Providing healthcare in the United States is complex, the result of various of levels of regulations—from federal to state to local—and the many professional associations and public-private partnerships involved in governing the industry. Almost every aspect of healthcare is heavily regulated, such as the rendering of services by medical professionals, the education of medical professionals, the provision of care by institutions, the manufacturing and dispensing of drugs, the financing of healthcare services provided to patients, and many others.¹ The often-stated goals for regulations of healthcare are to increase the quality of healthcare, improve access to healthcare, and to ensure that healthcare is affordable.² Pursuing these goals has, over time, led to a complex web of regulations with unintended consequences, not the least of which is an expensive healthcare delivery system that has struggled to innovate. In addition, the intricacy of these regulations favors industry incumbents that have the expertise and resources to maintain their position. It also constitutes a barrier to new entrants, shielding industry members from competition and innovation, which patients depend upon to ensure better care at a lower cost, year after year.³

The healthcare industry needs to be dynamic and needs to benefit from the openness and competition that have shaped other sectors of the economy, such as information technology and telecom-

munications. The first step toward a more competitive and dynamic healthcare system is shedding light on the regulatory system that governs its functioning. The Quantitative Health Lab (QHL), an initiative of the Open Health project at the Mercatus Center at George Mason University, developed Health RegData, a suite of datasets that contain the number and complexity of healthcare regulations across the country, both at the federal and state levels. The QHL builds upon almost a decade of research by the Mercatus Center's Policy Analytics project, which uses data analytics to quantify regulations, analyze the effects of regulations on the economy, and derive recommendations for reforms.⁴

We summarize the major features of state healthcare regulations from State Health RegData, which QHL researchers developed by using QuantGov, an open-source machine learning and text analysis platform for analyzing regulatory text. In addition to assessing the volume of healthcare regulations, we determine the industries that are likely to be affected by these regulations. Healthcare data abound, covering such domains as claims, health outcomes, clinical outcomes, expenditure, and others. However, to our knowledge, no data currently exist that would map the entire regulatory landscape. By engaging in this effort, we provide researchers, policymakers, entrepreneurs, and other stakeholders with a novel type of data to be used in driving better policy making and innovative initiatives.

The rest of the brief proceeds as follows: first, we describe how State Health RegData could be used by researchers; next we provide a brief overview of the methodology and a summary of the results; we conclude by outlining the next steps for State Health RegData.

POTENTIAL USES OF STATE HEALTH REGDATA

This policy brief is merely an introduction to State Health RegData and the first part of a long-term endeavor to understand the nature of healthcare regulations and their effect on the provision of healthcare in the United States. It is our hope that researchers will explore these data in detail and combine them with other types of healthcare data (outcomes, cost, coverage, etc.) to gain insights into the effects of regulation on the quality, access, and cost of healthcare and the effects of regulation on the provision of medical goods and services.

Identify and Track Healthcare Regulations

Using State Health RegData, researchers can track the evolution of state and federal healthcare regulations. State Health RegData is especially relevant in 2020 as a regulation-tracking device, as regulations rapidly change in the wake of the COVID-19 pandemic.

With the incorporation of topic modeling in future iterations, QHL researchers will be able to update State Health RegData to help identify areas of healthcare regulation that are currently

understudied, with the objective of uncovering opportunities for reform. State Health RegData allows researchers to compare the volume of regulation across states and thereby measure differences in the regulation of certain healthcare industries, professions, goods, and services.

Study the Role of Regulations, If Any, in Health Outcomes and Inputs across States

In addition to comparing healthcare regulatory burden across states, researchers can also examine whether cross-state differences in the delivery of healthcare can be explained by the volume and quality of healthcare regulations. State Health RegData allows for the incorporation of regulatory data into such analyses, which was previously not feasible owing to the textual format of such data.

METHODOLOGY

State Health RegData is a unique dataset that was developed using QuantGov's text analysis and machine learning algorithms to quantify and describe healthcare regulations in US states. In this section, we provide a brief overview of the methodology used in creating State Health RegData. We begin with our working definition of healthcare regulations, then describe the data sources, the data extraction process, and the algorithm. Finally, we describe the metrics or data elements that are available in State Health RegData.

Definitions

The key to State Health RegData is the definition of healthcare regulations. While many areas of public policy aim to improve human health (e.g., regulations on food, alcohol, air quality, and occupational safety), we limit our definition of healthcare to the provision of and payment for healthcare goods and services, such as pharmaceutical drugs and medical devices, hospitals, physicians, nursing care, telehealth, and health insurance. We do not include environmental health and safety regulations, even though they may be related to health, because these regulations are aimed at improving health but not necessarily aimed at the provision of healthcare services.

In addition to defining healthcare, we also define a unit of regulation. A unit of regulation is typically the body of text that is coherently related to a topic. States organize their regulatory codes into a hierarchy, dividing the content as appropriate. For example, Maryland organizes its administrative code into titles, subtitles, and chapters. Each combination of title, subtitle, and chapter deals with a topic or subject area. The titles correspond to agencies, while the subtitles reflect the function the agency is regulating. For example, the regulations for the Maryland Department of Health are filed under title 10, which covers most regulations pertaining to healthcare, including professional licensing. However, regulations pertaining to health insurance, which we include in our definition of healthcare, are filed under title 31, regulations for the Maryland Insurance Administration; the subtitles cover different types of insurance. Therefore, in the case of Mary-

land, title, subtitle, and chapter serve as a good unit of analysis. For the rest of this brief, we refer to the regulatory unit of analysis as a document.

It follows from this approach that the documents we identify as related to healthcare may contain text that does not pertain to healthcare. For example, it is common for states to include occupational safety and other environmental regulations within the hierarchy of regulations that we consider related to healthcare. Thus, only a portion of the document we identify as being healthcare related pertains to the regulation of healthcare. Unfortunately, this is a necessary compromise. As we describe in the next steps, subsequent versions of Health RegData will be built using models that parse units of regulation by various topics.

Data Sources

The primary data source for State Health RegData is the state regulatory codes of 44 states and the District of Columbia (DC). Data are not available for six states for various reasons.⁵ For this first version of State Health RegData, the data were collected from each state between May 2020 and July 2020. Therefore, these are the regulations that were in effect as of this period. This means that regulatory changes made since then are not included in this version of State Health RegData.

Algorithm Design

QuantGov determines healthcare regulatory restrictions in two steps. The first step is to use machine learning algorithms to identify regulatory texts that pertain to healthcare. (This first step is commonly known as training the algorithm.) In the second step, it applies text analysis tools to create the data elements—specifically, restrictions, complexity, and industry relevance.

We create an algorithm that determines the probability that a document pertains to healthcare.⁶ While we do not describe the details of the algorithm here, the key function of the classification methodology is to identify the words or combinations of words that are most common in healthcare-related documents.⁷ We train the algorithm using sample documents containing state and federal regulations that QHL researchers previously identified as having to do with healthcare; the algorithm demonstrated exceptional accuracy (an F1 score of 0.97 out of 1).⁸ We then apply the algorithm to all documents across the 44 states and DC. The output of the algorithm is the probability that a document pertains to healthcare (as we have defined it). A higher probability means that a document is more likely to be related to healthcare.

We determine a probability threshold for the inclusion of documents into the final State Health RegData dataset. All documents whose healthcare probability exceeds this threshold are included.

State Health RegData Metrics

After identifying the healthcare regulations, we proceed to create the data elements, or metrics, available in State Health RegData. The main elements are the total number of regulatory restrictions, the number of words in each document, the complexity of the document (average sentence length, number of conditional statements, and amount of information contained in each document), and the probability that the document pertains to the healthcare industry, as defined by the North American Industry Classification System (NAICS) codes.

We use QuantGov’s text analysis and machine learning algorithms to determine the number of regulatory restrictions in a document, determine the complexity the document, and further classify the document into the industries that could be affected by the regulations in that document.⁹ QuantGov identifies regulatory restrictions in a unit of regulation by counting the restrictive terms. These *regulatory restrictions* are instances of the terms “shall,” “must,” “may not,” “prohibited,” and “required,” and they can signify legal constraints and obligations. The total number of occurrences of these terms is counted in all Health RegData datasets.

In addition to counting healthcare regulatory restrictions, Health RegData also determines which industries, classified by NAICS code, are likely to be affected by a unit of regulation.¹⁰ Using the probabilities and the total number of restrictions identified in the unit of regulation, we derive the number of industry-relevant restrictions, which is the total number of restrictions multiplied by the probability that the unit of regulation applies to an industry.¹¹ The NAICS two-digit code “62” covers the healthcare and social assistance sector. We examine how the various industries in this broad category are affected by healthcare regulations. In addition, there are other industries, such as insurance carriers, that are affected by healthcare regulations but not in the healthcare and social assistance sector.

Finally, we use Health RegData to examine the complexity of regulations. Complexity refers to the readability of regulatory text and the ease of comprehension. Both are important because, all else being equal, text that is easy to read and understand will have lower compliance costs than text that is more difficult to read and understand.¹² As stated in the Federal Plain Language Guidelines of the US government, complex ideas are easier to grasp when they are presented in a manner that adheres to plain-language principles, such as using short sentences and few conditional clauses.¹³

We use three measures of complexity: sentence length (i.e., the average number of words per sentence), conditional statements (i.e., the average number of conditional clauses), and Shannon Entropy (a measure of the average information content in a document).¹⁴ In general, longer sentences are more difficult to read and understand, as are texts that include several conditional clauses and cross-references. And high Shannon Entropy scores mean that the content of the document spans a wide range of topics and concepts. Complex documents contain more information and therefore require more resources to understand and comply with. We discuss the complexity of regulations in the results section of this policy brief.¹⁵

RESULTS

We now present the preliminary data from State Health RegData 1.0. These data will be updated annually as the algorithms are updated. For each state, we report the total number of healthcare-related restrictions, the number of regulations that are relevant to the healthcare sector, and the share of total regulatory restrictions that pertain to healthcare.

HEALTHCARE REGULATIONS

The QuantGov algorithm identified 60,274 units of regulation that pertain to healthcare out of a total of 314,410 units of regulations across the 44 states and DC, representing a fifth of all regulations. Because the algorithm outputs the probability that a unit of regulation pertains to healthcare, the final list of documents we classify as healthcare related are those whose healthcare probability is greater than the decision threshold. Table 1 shows a random sample of the documents the algorithm identified as pertaining to healthcare, ordered by probability.

The second column, “Healthcare Probability,” is the predicted probability that the document pertains to healthcare. The third column, “Document Title,” is the title of the document as it appears in the administrative code of the state. Here we can surmise the content of the regulation. For example, in the first row, the title of the regulatory document is “Hospice Services,” and it is issued by the Oregon Health Authority. The algorithm determined with certainty that it pertains to healthcare. This document has 7,209 words (see sixth column), 72 restrictive terms (see fifth column), and a Shannon Entropy score of 8.56. These data are then gathered to create State Health RegData, which is described below.

STATE	HEALTHCARE PROBABILITY	DOCUMENT TITLE	AGENCY	NUMBER OF RESTRICTIONS	NUMBER OF WORDS	SHANNON ENTROPY
Oregon	1.00	Hospice Services	Oregon Health Authority	72	7,049	8.56
Maryland	0.95	Responsibilities Accreditations and Audits	Maryland Department of Health	21	1,587	7.81
Kentucky	0.91	Relative Responsibility Requirements for Medicaid	Cabinet for Health and Family Services	34	1,533	7.57
New Mexico	0.10	Calculation of Excise Tax	Taxation	8	1,923	7.49
Kentucky	0.04	Kentucky Building Code	Public Protection Cabinet	40	1,660	7.98
Missouri	0.04	Division of Finance – Association Branches and Other Facilities	Department of Commerce and Insurance	71	3,900	8.22

Table 1 (continued)						
STATE	HEALTHCARE PROBABILITY	DOCUMENT TITLE	AGENCY	NUMBER OF RESTRICTIONS	NUMBER OF WORDS	SHANNON ENTROPY
Wisconsin	0.04	Personnel Qualifications for Conducting Environmental Response Actions	Department of Natural Resources	43	3,406	8.29
Louisiana	0.03	Public Notification of Contamination	Environmental Quality	14	1,221	7.87
Maryland	0.03	Board of Boiler Rules	Maryland Department of Labor	422	21,378	9.39
Louisiana	0.02	Design of Pipeline Components 49 CFR Part 192 Subpart D	Natural Resources	160	9,335	8.58
Louisiana	0.02	School Bus Routes	Education	9	654	7.13
New Mexico	0.02	Penalties	Natural Resources and Wildlife	87	5,009	8.40
Iowa	0.01	Examination Reviews and Investigations	Commerce Department	19	1,636	7.45
Louisiana	0.01	Revocation or Suspension of Licenses of Registration	Professional and Occupational Standards	45	2,805	8.15
New York	0.01	Security Officer Training Tax Credit	Executive Department	26	1,701	7.55
Illinois	0.00	Interpretations of Statutory Language	Illinois Commerce Commission	1	853	7.42
Illinois	0.00	Special Permanent Program Performance Standards Coal Preparation Plants Not Located Within the Permit Area of a Mine	Department of Natural Resources	16	576	6.88
Nevada	0.00	Taxes on Agricultural Real Property and Open Space	N/A	62	4,731	8.37
New York	0.00	Prepayment of Sales Tax on Cigarettes	Department of Taxation and Finance	19	875	6.96
Utah	0.00	Driver Education Endorsement.	Education	8	1,069	7.91

Source: State Health RegData 1.0 (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, accessed August 2020.

Total Number of Healthcare Regulatory Restrictions

Table 2 shows the total number of healthcare restrictions and words in healthcare regulations and the total number of regulatory restrictions in each of the 44 states and the DC. Across the 44 states and DC, there are 805,817 healthcare restrictions as of July 2020 in more than 60,000 units of regulations deemed to pertain to healthcare. Healthcare regulatory restrictions make up 13.5 percent of all restrictions in these states. By contrast, federal healthcare regulatory restrictions make up only 5 percent of total federal regulatory restrictions. The number of regulatory restrictions varies widely across states. The state with the smallest number of healthcare regulatory restrictions is Montana (2,486 restrictions) and the state with the greatest number of healthcare regulatory restrictions is New York (46,184).

Table 3 shows the number of regulatory restrictions for each of the 44 states and DC and the proportion of each state’s regulatory code that applies to healthcare. In addition, we present the number of words contained in the regulatory texts. New York has the most healthcare regulatory restrictions, and these restrictions represent 15 percent of all regulatory restrictions in the state. The share of regulatory restrictions in a state that pertains to healthcare ranges from 4.2 percent

Table 2. Total Number of Healthcare Restrictions and Words in Regulations

METRIC	HEALTHCARE REGULATIONS	ALL REGULATIONS	PERCENT HEALTHCARE RELATED
Total restrictions	805,817	5,974,310	13.49
Total words	55,480,958	407,421,667	13.62

Source: State Health RegData 1.0 (dataset).

Table 3. Healthcare Regulatory Restrictions

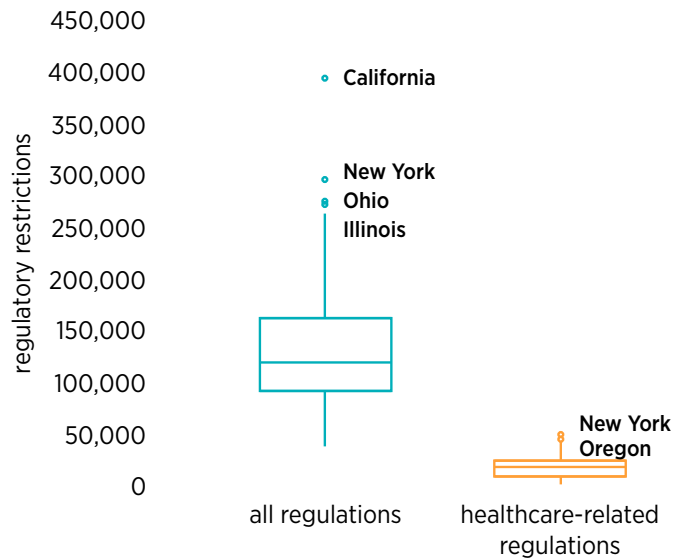
STATE	RESTRICTIONS			WORDS		
	HEALTHCARE REGULATIONS	ALL REGULATIONS	PERCENT HEALTHCARE RELATED	HEALTHCARE REGULATIONS	ALL REGULATIONS	PERCENT HEALTHCARE RELATED
Alabama	19,515	107,880	18.09	1,443,280	7,533,874	19.16
Arizona	12,860	64,319	19.99	1,762,790	6,004,954	29.36
California	28,548	395,608	7.22	1,357,919	21,284,860	6.38
Colorado	28,914	154,964	18.66	2,061,867	12,177,626	16.93
District of Columbia	25,087	137,185	18.29	1,424,467	8,128,682	17.52
Delaware	12,753	105,475	12.09	959,213	6,787,820	14.13
Florida	26,195	168,795	15.52	1,588,312	10,986,328	14.46
Georgia	15,113	109,848	13.76	703,536	6,067,499	11.60
Idaho	5,374	38,961	13.79	630,750	4,005,332	15.75
Illinois	21,828	273,989	7.97	1,408,633	18,213,395	7.73

Table 3 (continued)

STATE	RESTRICTIONS			WORDS		
	HEALTHCARE REGULATIONS	ALL REGULATIONS	PERCENT HEALTHCARE RELATED	HEALTHCARE REGULATIONS	ALL REGULATIONS	PERCENT HEALTHCARE RELATED
Iowa	21,092	160,603	13.13	1,230,041	9,816,474	12.53
Kansas	8,276	69,925	11.84	362,397	3,211,823	11.28
Kentucky	19,611	116,274	16.87	1,041,978	6,111,092	17.05
Louisiana	30,325	164,387	18.45	2,150,981	11,222,553	19.17
Maine	23,047	119,591	19.27	1,757,904	8,685,587	20.24
Maryland	28,033	137,695	20.36	2,228,113	9,989,257	22.31
Massachusetts	22,281	164,636	13.53	1,711,490	10,892,967	15.71
Michigan	4,767	76,236	6.25	261,873	4,367,835	6.00
Minnesota	8,608	98,067	8.78	488,854	5,696,249	8.58
Mississippi	24,798	116,153	21.35	1,680,591	8,880,233	18.93
Missouri	19,051	93,915	20.29	1,198,165	6,720,122	17.83
Montana	2,486	59,788	4.16	293,246	4,742,174	6.18
Nebraska	12,230	95,955	12.75	883,007	7,237,361	12.20
Nevada	3,153	64,265	4.91	238,612	4,981,924	4.79
New Hampshire	28,925	133,592	21.65	1,696,757	8,060,053	21.05
New Mexico	18,991	128,946	14.73	1,416,472	9,485,592	14.93
New York	46,184	296,296	15.59	2,650,671	17,622,247	15.04
North Carolina	11,017	107,092	10.29	565,169	5,970,940	9.47
North Dakota	10,047	52,385	19.18	716,553	3,709,157	19.32
Ohio	19,843	274,470	7.23	1,914,238	22,646,803	8.45
Oklahoma	14,196	142,604	9.95	945,613	9,211,617	10.27
Oregon	43,752	200,477	21.82	3,522,835	18,179,103	19.38
Pennsylvania	15,349	162,937	9.42	1,334,811	13,093,618	10.19
Rhode Island	13,830	94,051	14.70	877,471	5,789,509	15.16
South Carolina	10,082	78,727	12.81	462,394	4,765,126	9.70
South Dakota	3,980	43,521	9.15	350,443	3,380,173	10.37
Tennessee	22,247	119,272	18.65	1,508,555	8,276,640	18.23
Texas	36,365	263,369	13.81	2,400,793	17,117,088	14.03
Utah	8,904	91,517	9.73	464,022	5,672,997	8.18
Virginia	22,433	140,021	16.02	1,237,645	8,688,471	14.24
Washington	22,324	197,466	11.31	2,243,328	17,507,192	12.81
West Virginia	13,262	120,210	11.03	1,017,998	8,407,880	12.11
Wisconsin	9,574	161,549	5.93	708,526	12,250,243	5.78
Wyoming	10,567	71,294	14.82	578,645	3,841,197	15.06

Source: State Health RegData 1.0 (dataset).

Figure 1. Distribution of Regulatory Restrictions



Source: Authors' calculations based on State Health RegData 1.0 (dataset).

in Montana to 21.7 percent in New Hampshire. Three other states (Mississippi, Missouri, and Arizona) have roughly a fifth of their regulatory restrictions devoted to healthcare (21.4 percent, 20.3 percent, and 20.0 percent, respectively).

Figure 1 shows the distribution of the numbers of healthcare regulatory restrictions in table 3. Although California has the most regulatory restrictions (395,605) among the 44 states and DC, healthcare regulatory restrictions (28,548) comprise only 7 percent of California's total restrictions. California, Illinois, New York, and Ohio are outliers, with significantly more restrictions in general than other states, but only New York remains an outlier in terms of healthcare regulatory restrictions specifically.

REGULATIONS BY INDUSTRY

Table 4 shows the top 10 industries most affected by healthcare regulations. They are ordered by the number of industry-relevant restrictions, which are calculated by multiplying the total number of regulatory restrictions in each unit of regulation by the predicted probability that the regulations in that document affect an industry. As one would expect, industries that deal directly with healthcare are the ones most affected by healthcare-related regulatory restrictions. Across all 44 states and DC, the ambulatory healthcare services industry has the most industry-relevant restrictions (25 percent), followed by the administrative and support services industry (22 percent). Out of 84 industries (classified at the three-digit NAICS level), 10 industries account for 81 percent of all healthcare regulatory restrictions.

NAICS CODE	INDUSTRY	RESTRICTIONS	PERCENTAGE OF TOTAL	CUMULATIVE PERCENTAGE
621	ambulatory healthcare services	132,033	24.62	24.62
561	administrative and support services	119,180	22.22	46.84
624	social assistance	36,412	6.79	53.63
524	insurance carriers and related activities	33,628	6.27	59.90
541	professional, scientific, and technical services	33,073	6.17	66.07
322	paper manufacturing	24,505	4.57	70.64
445	food and beverage stores	18,720	3.49	74.13
623	nursing homes	17,983	3.35	77.49
622	hospitals	10,907	2.03	79.52
611	educational services	8,386	1.56	81.08

Source: State Health RegData 1.0 (dataset).

Table 5 shows the proportion of state healthcare regulatory restrictions that are relevant to some key healthcare industries from table 4 and to some nonhealthcare ones. More than half of the healthcare regulatory restrictions in Montana are relevant to the ambulatory healthcare services industry. By contrast, less than 5 percent of healthcare restrictions in North Carolina and New Hampshire apply to the ambulatory healthcare services industry.

STATE	AMBULATORY HEALTHCARE SERVICES	HOSPITALS	NURSING AND RESIDENTIAL CARE FACILITIES	SOCIAL ASSISTANCE	PROFESSIONAL, SCIENTIFIC, AND TECHNICAL SERVICES	ADMINISTRATIVE SUPPORT SERVICES
Alabama	20.93	6.56	2.09	3.67	7.07	17.50
Arizona	35.54	0.73	0.72	14.09	5.18	4.91
California	27.74	1.84	0.38	4.42	5.78	31.92
Colorado	43.39	0.55	2.17	3.04	2.58	18.72
District of Columbia	15.91	1.78	1.48	10.69	4.14	32.67
Delaware	23.71	1.31	0.91	7.51	2.90	17.60
Florida	20.19	7.05	8.26	5.30	7.20	19.11
Georgia	16.09	0.93	1.07	7.64	4.23	55.49
Idaho	47.02	0.69	7.88	5.02	2.37	12.82
Illinois	33.85	3.84	2.11	3.91	3.53	15.02
Iowa	24.16	1.67	1.11	10.94	1.43	24.36
Kansas	8.29	1.88	1.41	14.92	5.08	11.67
Kentucky	18.58	0.97	2.21	4.18	1.68	16.90
Louisiana	18.25	1.15	1.99	5.95	2.09	25.09

Table 5 (continued)

STATE	AMBULATORY HEALTHCARE SERVICES	HOSPITALS	NURSING AND RESIDENTIAL CARE FACILITIES	SOCIAL ASSISTANCE	PROFESSIONAL, SCIENTIFIC, AND TECHNICAL SERVICES	ADMINISTRATIVE SUPPORT SERVICES
Maine	32.45	1.31	0.88	5.75	2.85	41.05
Maryland	21.98	1.40	9.74	10.50	4.38	16.64
Massachusetts	29.33	1.46	2.95	11.58	4.51	14.75
Michigan	32.83	1.63	1.63	7.16	1.05	6.96
Minnesota	25.31	4.08	18.91	7.14	3.01	6.41
Mississippi	28.48	0.19	1.85	9.01	8.09	21.51
Missouri	17.55	1.14	0.75	6.05	2.14	27.70
Montana	52.17	1.35	4.56	11.48	0.43	11.64
Nebraska	35.82	2.79	2.79	2.61	6.79	16.70
Nevada	21.97	0.53	0.12	2.28	3.62	5.25
New Hampshire	4.53	0.60	1.02	4.98	32.82	37.63
New Mexico	27.24	1.65	2.37	4.56	8.22	13.56
New York	20.18	1.07	4.24	9.75	7.24	21.34
North Carolina	4.15	1.13	0.58	1.27	0.33	22.90
North Dakota	21.76	0.59	1.24	23.20	0.76	23.85
Ohio	19.32	0.50	8.95	4.50	2.11	4.96
Oklahoma	38.26	0.89	1.79	8.18	6.57	20.35
Oregon	24.13	3.61	2.22	10.92	7.75	16.83
Pennsylvania	29.34	2.46	5.52	9.45	9.14	11.56
Rhode Island	25.37	2.64	4.24	9.46	4.85	22.81
South Carolina	7.91	0.39	2.03	0.57	1.66	16.28
South Dakota	37.15	0.48	0.58	0.63	1.04	3.84
Tennessee	27.05	6.53	6.70	2.85	5.91	15.50
Texas	17.44	2.20	0.50	3.88	3.87	24.10
Utah	24.32	2.17	7.03	4.08	3.19	18.70
Virginia	18.97	3.25	0.82	9.86	6.21	42.70
Washington	31.65	3.00	11.33	5.69	5.33	7.15
West Virginia	27.02	1.71	8.38	4.88	2.82	14.19
Wisconsin	39.96	5.02	9.30	4.14	2.47	12.26
Wyoming	26.64	1.18	7.12	4.94	7.58	12.36

Source: State Health RegData 1.0 (dataset).

Complexity of Healthcare Regulations

Table 6 shows the averages for the three measures of regulatory complexity across the 44 states and DC.

METRIC	HEALTHCARE REGULATIONS	ALL REGULATIONS
Conditionals	74.06	52.36
Sentence length	24.10	24.14
Shannon Entropy	8.37	7.94

Source: State Health RegData 1.0 (dataset).

On the average, there are 74 conditional clauses in the identified healthcare regulatory texts of the 44 states and DC. By comparison, there are on average 52 such clauses in all the regulatory texts of the states. In terms of the other complexity metrics, the healthcare regulatory texts do not differ significantly from the other regulatory texts. On average, there are 24 words per sentence across all regulatory texts in the states. The Shannon Entropy score for healthcare regulatory text is 8.4, slightly higher than that for all regulatory text (7.9). Generally, documents with higher Shannon Entropy scores are more difficult to read because they contain a lot of varied information. For reference, Shakespeare plays typically have a Shannon Entropy score of 9.0 to 9.5.¹⁶ Table 7 shows the complexity metrics for all 44 states and DC. South Carolina's administrative code has the highest Shannon Entropy score (9.13), followed by Mississippi's (8.94) and Colorado's (8.76). South Caro-

STATE	AVERAGE SENTENCE LENGTH		AVERAGE NUMBER OF CONDITIONALS		AVERAGE SHANNON ENTROPY	
	HEALTHCARE REGULATIONS	ALL REGULATIONS	HEALTHCARE REGULATIONS	ALL REGULATIONS	HEALTHCARE REGULATIONS	ALL REGULATIONS
Alabama	20.43	21.60	36.78	24.32	8.40	7.64
Arizona	11.40	12.18	176.51	150.51	8.52	8.64
California	30.13	30.99	137.53	117.10	8.77	7.71
Colorado	24.49	28.21	131.48	103.11	8.76	8.56
District of Columbia	39.92	38.26	40.18	31.47	8.38	8.10
Delaware	26.05	26.90	47.98	40.53	8.43	7.85
Florida	22.64	22.92	31.19	27.84	8.07	7.74
Georgia	32.02	31.59	48.82	43.45	8.00	7.45
Idaho	13.49	13.13	76.79	46.48	8.63	8.42
Illinois	35.50	39.06	80.70	47.86	8.76	8.25
Iowa	22.08	21.57	59.29	28.55	8.15	7.68

Table 7 (continued)

STATE	AVERAGE SENTENCE LENGTH		AVERAGE NUMBER OF CONDITIONALS		AVERAGE SHANNON ENTROPY	
	HEALTHCARE REGULATIONS	ALL REGULATIONS	HEALTHCARE REGULATIONS	ALL REGULATIONS	HEALTHCARE REGULATIONS	ALL REGULATIONS
Kansas	23.29	19.39	37.10	19.22	7.95	7.25
Kentucky	19.63	18.42	22.11	9.12	7.98	7.44
Louisiana	20.60	19.18	27.92	21.20	8.10	7.66
Maine	20.73	21.72	78.66	36.74	8.35	8.19
Maryland	26.82	24.01	20.89	16.46	8.03	7.74
Massachusetts	27.89	28.63	42.24	42.11	8.39	8.07
Michigan	19.67	18.57	44.89	36.99	8.36	7.85
Minnesota	17.48	19.07	96.56	70.02	8.37	8.25
Mississippi	19.55	25.09	148.91	117.23	8.94	8.81
Missouri	27.09	29.77	63.76	26.96	8.57	7.57
Montana	25.55	24.19	145.24	84.29	8.66	8.42
Nebraska	22.98	25.56	60.51	26.06	8.50	7.48
Nevada	22.43	23.05	273.14	83.32	8.34	8.00
New Hampshire	32.89	30.61	87.01	55.13	8.20	8.10
New Mexico	30.76	31.54	23.55	19.72	7.83	7.62
New York	42.78	41.89	63.12	28.02	8.14	7.33
North Carolina	20.40	20.02	140.09	123.48	8.81	8.45
North Dakota	16.87	17.08	77.00	48.84	8.25	7.66
Ohio	38.48	34.81	105.41	89.05	8.39	8.08
Oklahoma	27.81	27.65	82.26	68.52	8.41	8.30
Oregon	24.38	24.23	52.78	40.56	8.26	7.99
Pennsylvania	16.00	16.57	59.13	43.39	8.48	8.10
Rhode Island	22.04	22.22	54.97	34.94	8.64	8.14
South Carolina	16.84	19.58	81.82	175.85	9.13	8.86
South Dakota	22.64	20.80	16.78	8.12	7.67	6.89
Tennessee	23.83	21.96	57.15	34.91	8.46	7.96
Texas	35.55	35.51	139.51	86.05	8.62	8.22
Utah	24.33	24.56	21.32	17.95	8.07	7.69
Virginia	16.63	16.33	88.78	49.69	8.69	8.00
Washington	24.33	22.26	61.04	39.39	8.20	7.74
West Virginia	15.98	15.20	40.01	32.31	8.17	7.95
Wisconsin	12.62	13.58	58.12	40.91	8.04	7.96
Wyoming	23.47	22.66	19.54	15.95	8.23	7.76

Source: State Health RegData 1.0 (dataset).

lina's is thus comparable to a typical Shakespeare play. By contrast, South Dakota's code has the lowest Shannon Entropy score (7.67), so South Dakota's code is easier to read and understand than all other analyzed states. In addition, the average unit of South Dakota's regulatory text contains less information than other states.

CONCLUSION AND NEXT STEPS

In this policy brief, we introduce the first edition of State Health RegData, a quantification of the volume of healthcare regulatory restrictions in 44 US states and DC. We provide a high-level view of the number of healthcare regulatory restrictions, the industries affected by these regulations, and the complexity of the healthcare regulatory text. These data can be used to examine the role of healthcare regulations in healthcare outcomes by comparing regulatory burdens across states. Over the next few months, QHL researchers will continue to improve State Health RegData, further classifying the healthcare regulations identified into various topics of interest. These include certificate of need, scope of practice, telemedicine, and insurance benefits, among others. Using this topic-based approach will allow researchers to examine the impact of specific regulations on outcomes of interest.

ABOUT THE AUTHORS

Kofi Ampaabeng is a research fellow and data scientist at the Mercatus Center at George Mason University. He specializes in curating data and generating policy-relevant insights from data. Before joining the Mercatus Center, he worked for IMPAQ International, LLC, where he evaluated the efficacy of government programs.

Elise Amesz-Droz is a program manager at the Mercatus Center at George Mason University. Previously, she has worked as a market research analyst for a start-up company in the pharmaceutical industry. She is a graduate of the Fuqua School of Business at Duke University.

NOTES

1. Robert I. Field, *Health Care Regulation in America: Complexity, Confrontation, and Compromise* (Oxford University Press, 2006).
2. Field, *Health Care Regulation in America*.
3. George Stigler, "The Theory of Economic Regulation," *Bell Journal of Economics and Management Science* 2, no. 1 (1971): 3-21; Omar Al-Ubaydli and Patrick A. McLaughlin, "RegData: A Numerical Database on Industry-Specific Regulations for All US Industries and Federal Regulations, 1997-2012," *Regulation and Governance* 11, no. 1 (2015): 109-23.
4. Al-Ubaydli and McLaughlin, "RegData: A Numerical Database."
5. Alaska's code website was unavailable, Arkansas has not published its administrative code online, Connecticut's website was inaccessible at the time of the web scraping, New Jersey's code is unavailable because it is licensed to LexisNexis, and Hawaii's code was incomplete.

6. We describe the details of the algorithm design in a separate methodology document
7. Patrick A. McLaughlin and Oliver Sherouse, "RegData 2.2: A Panel Dataset on US Federal Regulations," *Public Choice* 180 (2019): 43–55.
8. The F1 score is a means of measuring the accuracy of a classification algorithm that combines two concepts: precision and recall. Precision is the number of positive cases that were correctly identified, and recall is the proportion of actual positive cases that the algorithm can identify. The harmonic mean of these two measures is the F1 score. Higher F1 scores mean higher accuracy in classifying the input text.
9. In practice, we execute steps 1 and 2 in a single process; documents with low healthcare probability are not included in the final dataset.
10. McLaughlin and Sherouse, "RegData 2.2."
11. Al-Ubaydli and McLaughlin, "RegData: A Numerical Database."
12. Rudolph Flesch, "A New Readability Yardstick," *Journal of Applied Psychology* 32, no. 3 (1948): 221–33.
13. Plain Language Action and Information Network, *Federal Plain Language Guidelines*, rev. 1, May 2011.
14. Patrick A. McLaughlin, Scott Atherley, and Stephen Strosko, "RegData Canada: An Overview" (Mercatus Policy Brief, Mercatus Center at George Mason University, Arlington, VA, February 2019).
15. Patrick A. McLaughlin et al., "Is Dodd-Frank the Biggest Law Ever?" (Mercatus Working Paper, Mercatus Center at George Mason University, Arlington, VA, June 2020).
16. Marcin Lawnik, "Shannon's Entropy in Literary Works and Their Translations" (unpublished manuscript, n.d.) PDF file, <http://computer.scientific-journal.com/articles/1/23.pdf>.