

Health RegData: Quantifying the Volume of Healthcare Regulations

Kofi Ampaabeng, Jonathan Nelson, and Elise Amez-Droz

April 2020

Regulations have played a significant role in shaping the healthcare industry in the United States. Through financing healthcare, licensing healthcare practitioners, educating healthcare professionals, establishing and operating healthcare facilities, and controlling the nature of the care to be provided, among other functions, federal and state healthcare regulations have over time created the complex US healthcare system of today.

While researchers over the years have examined the impact of specific regulations on various outcomes, they have paid very little attention to the volume or other quantifiable characteristics of healthcare regulations. This is because the sheer volume of regulations makes them difficult to analyze, and their format—text—calls for special tools to examine them. We aim to address this deficiency in data availability by employing the QuantGov and RegData platforms, developed by the Mercatus Center at George Mason University, to quantify the volume of healthcare regulations in the United States, including both federal and state regulations. In addition, we examine the industries that are most affected by healthcare regulations, and we measure the complexity of the regulations.

Since 2012, the Mercatus Center’s RegData project has made it possible for policymakers and stakeholders to discuss the effects of regulations by providing a quantifiable and replicable way of measuring regulations in a growing number of national and subnational jurisdictions, including the United States, Canada, and Australia. RegData measures the volume of regulations by counting the number of restrictions found in a unit of regulation.

By design, regulations impose restrictions on an agent of society by either preventing or mandating specific activities. Therefore, RegData quantifies the number of regulatory restrictions by counting the number of restrictive terms in a unit of regulatory text. In RegData, a regulatory restriction is defined as the occurrence any of the following terms, which are commonly associated with regulatory restrictions, in a unit of regulation: “shall,” “must,” “may not,” “required,” and “prohibited.” In addition to counting the number of restrictions, RegData also uses QuantGov, a machine learning platform created by the Mercatus Center, to estimate the probability that the text of a regulation applies to a specific industry. This unique feature of RegData allows for the examination of the impact of regulations on specific industries.

Health RegData extends the RegData approach to analyze healthcare regulations in the United States. The publication of this brief and the accompanying data marks the beginning of a long-term project to quantify the volume of healthcare regulations and the evolution of healthcare regulation since 1996. This first iteration of Health RegData examines the stock of healthcare restrictions as of December 2018. The main source of data is the *US Code of Federal Regulations* (CFR), published by the Government Publishing Office.

METHODOLOGY AND DATA

The CFR, which contains all US federal regulations, is organized into 50 titles, each of which covers a broad subject area. Each title is subdivided into parts, and parts are divided into sections. (Some large parts have subparts before sections.) In Health RegData, we evaluate regulations at the section level. In order to accurately identify regulatory restrictions, it is critical to determine the unit of analysis and the definition of a healthcare regulation. A unit of regulation is determined to pertain to healthcare if it concerns the provision of healthcare goods and services. Using this narrow definition, we exclude regulations that concern food, alcohol, air quality, labor conditions, and so forth, which may have the goal of improving the health status of Americans but do not directly relate to the provision of healthcare goods and services. Under this narrow definition, regulations related to pharmaceutical drugs are considered to be healthcare regulations. For Health RegData, we chose CFR sections to be the unit of analysis. This means a section is considered a unit of regulation.

A team of researchers reviewed all the CFR titles to identify parts and sections that regulate healthcare. They then extracted these regulations from the CFR and analyzed them using the QuantGov platform. This approach of manually identifying regulations is naturally prone to human error. However, it provides a reasonable approximation of the total number of federal healthcare regulations. In subsequent versions of Health RegData, we will use the QuantGov platform to train an algorithm to identify healthcare regulations at both the federal and state levels.

In addition to counting regulatory restrictions, RegData also determines the industries that are likely to be affected by a unit of regulation.¹ We use the North American Industry Classification

System (NAICS) codes to define the industries. We then use RegData’s machine learning algorithms to classify regulations into industries by assigning a probability that a unit of regulation applies to an industry identified by a NAICS code. Using the probabilities and the total number of restrictions identified in the unit of regulation, we derive the industry-relevant restrictions, which is the total number of restrictions multiplied by the probability that the unit of regulation applies to an industry.²

Finally, RegData allows us to examine the complexity of regulations. We use the term *complexity* to capture a few concepts related to compliance with regulations. These include the readability of the regulatory text and the ease of comprehension. The ease of reading and comprehending the regulatory text is important because, all else being equal, text that is easy to read and comprehend will have lower compliance costs than text that is more difficult to read and understand.³ 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.⁴

Using the CFR data, RegData provides a number of metrics to quantify the complexity of healthcare regulations. These are the unit of regulation’s readability (using the Flesch Reading Ease Score), the average sentence length, and the number of conditional statements in the unit of regulation. In addition, we borrow a concept from information theory known as Shannon entropy. Shannon entropy simply measures the rate at which new information is introduced in a unit of text. If a lot of new information is introduced, the text is more difficult to read and understand.

THE VOLUME OF FEDERAL HEALTHCARE REGULATIONS IN THE UNITED STATES

From the 2018 CFR, we identified 49,312 healthcare regulatory restrictions as of December 2018, comprising 4.5 percent of all federal regulatory restrictions (1,085,063). As one would expect, the Department of Health and Human Services (HHS) has issued 85.4 percent of all healthcare regulatory restrictions (42,088). Table 1, ordered from the agency that has issued the most healthcare restrictions to the one that has issued the least, shows that five agencies—mostly executive departments—have issued 98.5 percent of all federal healthcare regulatory restrictions. For comparison, table 1 also shows how many restrictions of any kind (including healthcare restrictions) these same agencies have issued.

Table 2 shows the total restrictions issued by offices within HHS. As one would expect, the Centers for Medicare and Medicaid Services and the Food and Drug Administration have issued most of the healthcare restrictions.

Table 3 shows the industries affected by the healthcare regulations identified, ordered by the number of industry-relevant restrictions. Industry-relevant restrictions, as mentioned previously,

Table 1. Share of Healthcare Regulatory Restrictions by US Agency, December 2018

AGENCY	HEALTHCARE REGULATIONS		ALL REGULATIONS	
	NO. OF RESTRICTIONS	% OF TOTAL	NO. OF RESTRICTIONS*	% OF TOTAL
Department of Health and Human Services	42,088	85.4	60,552	5.7
Department of Justice	2,519	5.1	17,996	1.7
Department of Veterans Affairs	1,601	3.3	10,395	1.0
Office of Personnel Management	1,160	2.4	8,628	0.8
Department of Labor	1,136	2.3	68,184	6.4
Department of Education	325	0.7	11,242	1.1
Federal Communications Commission	282	0.6	28,529	2.7
Social Security Administration	89	0.2	3,466	0.3
Department of the Interior	56	0.1	51,167	4.8
Department of the Treasury	30	0.1	99,372	9.3
Executive Office of the President	22	0.0	4,038	0.4
Environmental Protection Agency	4	0.0	169,895	15.9
TOTAL	49,312		533,464	

* The numbers in this column do not add up to the total because this table includes only the federal agencies that regulate healthcare. (The 533,464 restrictions from these agencies account for 49.2 percent of total federal regulatory restrictions.)
Source: Authors' calculations using Patrick A. McLaughlin and Oliver Sherouse, RegData US 3.1 Annual (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2018, <https://quantgov.org/regdata-us/>.

Table 2. Healthcare Regulatory Restrictions by Offices within the US Department of Health and Human Services (HHS), December 2018

HHS OFFICE	NO. OF RESTRICTIONS	% OF HHS RESTRICTIONS	% OF HEALTHCARE RESTRICTIONS
Centers for Medicare and Medicaid Services	16,484	39.1	33.4%
Food and Drug Administration	13,853	32.9	28.1%
Public Health Service	4,750	11.3	9.6%
Office of Inspector General—Health Care	402	1.0	0.8%
Joint regulations: Bureau of Indian Services and HHS	28	0.1	0.1%
Other HHS offices	6,599	15.7	13.4%
TOTAL HHS restrictions	42,088		85.4%

Source: Authors' calculations using Patrick A. McLaughlin and Oliver Sherouse, RegData US 3.1 Annual (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2018, <https://quantgov.org/regdata-us/>.

are the predicted probability that a unit of regulation affects an industry multiplied by the total number of regulatory restrictions in that unit. As one would expect, the industry most affected by healthcare regulations is ambulatory healthcare services. Industries that are directly associated with healthcare, such as hospitals and insurance providers, are in the top 10.

Table 3. The 10 US Industries Most Affected by Federal Healthcare Regulatory Restrictions

#	INDUSTRY	HEALTHCARE RESTRICTIONS	ALL RESTRICTIONS
1	621 - Ambulatory Health Care Services	3,265.88	20,031.20
2	524 - Insurance Carriers and Related Activities	1,415.70	18,509.99
3	325 - Chemical Manufacturing	958.36	102,732.55
4	445 - Food and Beverage Stores	942.58	15,178.82
5	622 - Hospitals*	818.97	N/A
6	481 - Air Transportation	719.59	53,549.13
7	452 - General Merchandise Stores*	665.71	N/A
8	522 - Credit Intermediation and Related Activities	622.04	64,257.37
9	624 - Social Assistance	555.43	22,134.57
10	623 - Nursing and Residential Care Facilities*	529.63	N/A

Note: The industry-relevant restrictions are excluded for these industries because the predicted probabilities were subject to noise. In general, all industries with low accuracy scores are excluded. NAICS = North American Industry Classification System.
Source: Authors' calculations using Patrick A. McLaughlin and Oliver Sherouse, RegData US 3.1 Annual (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2018, <https://quantgov.org/regdata-us/>.

COMPLEXITY OF REGULATIONS

RegData also provides tools to analyze the language used in writing regulations. As explained in the methodology and data section, we examine the complexity of healthcare regulations using three measures of complexity from the QuantGov platform. These are the average reading ease, the average sentence length, and the Shannon entropy score. In table 4, we show the averages of reading ease, sentence length (number of words), and Shannon entropy for US federal healthcare regulations. The reading ease measure uses the Flesch Reading Ease formula, which scores a unit of text on a scale up to 100. Generally, the higher the Flesch score, the more readable the text. There is no lower limit for the score, so it is possible for a unit of text to receive a negative score,⁵ and indeed many healthcare regulations issued by HHS do receive negative scores (see table 4).

Table 4. Complexity Metrics of US Federal Healthcare Regulations

AGENCY	FLESCH READING EASE SCORE		SENTENCE LENGTH (NO. OF WORDS)		SHANNON ENTROPY SCORE	
	HEALTHCARE REGULATIONS	ALL REGULATIONS	HEALTHCARE REGULATIONS	ALL REGULATIONS	HEALTHCARE REGULATIONS	ALL REGULATIONS
Department of Education	19.7	-61.3	29.0	24.0	6.2	7.4
Department of Health and Human Services	-2.4	-14.9	25.1	26.4	5.9	8.2
Department of Justice	-8.6	-3.3	28.1	24.9	6.0	7.9
Department of Labor	13.7	-1.6	29.9	26.0	6.5	8.1

Table 4 (continued)

AGENCY	FLESCH READING EASE SCORE		SENTENCE LENGTH (NO. OF WORDS)		SHANNON ENTROPY SCORE	
	HEALTHCARE REGULATIONS	ALL REGULATIONS	HEALTHCARE REGULATIONS	ALL REGULATIONS	HEALTHCARE REGULATIONS	ALL REGULATIONS
Department of the Interior	14.7	-0.8	26.7	24.6	5.2	7.9
Department of the Treasury	35.5	2.6	24.3	26.8	6.0	7.9
Department of Veterans Affairs	10.8	-2.1	24.3	22.2	6.3	7.6
Environmental Protection Agency	3.7	-10.1	29.4	26.9	5.1	8.4
Executive Office of the President	25.1	7.2	21.0	21.2	5.5	7.6
Federal Communications Commission	-7.9	-1.8	23.1	27.3	5.9	9.0
Office of Personnel Management	7.7	-3.6	23.8	22.7	5.7	7.45
Social Security Administration	24.6	26.4	28.6	20.8	5.2	7.04
Average	-1.5	-2.4	25.4	24.7	5.9	7.9

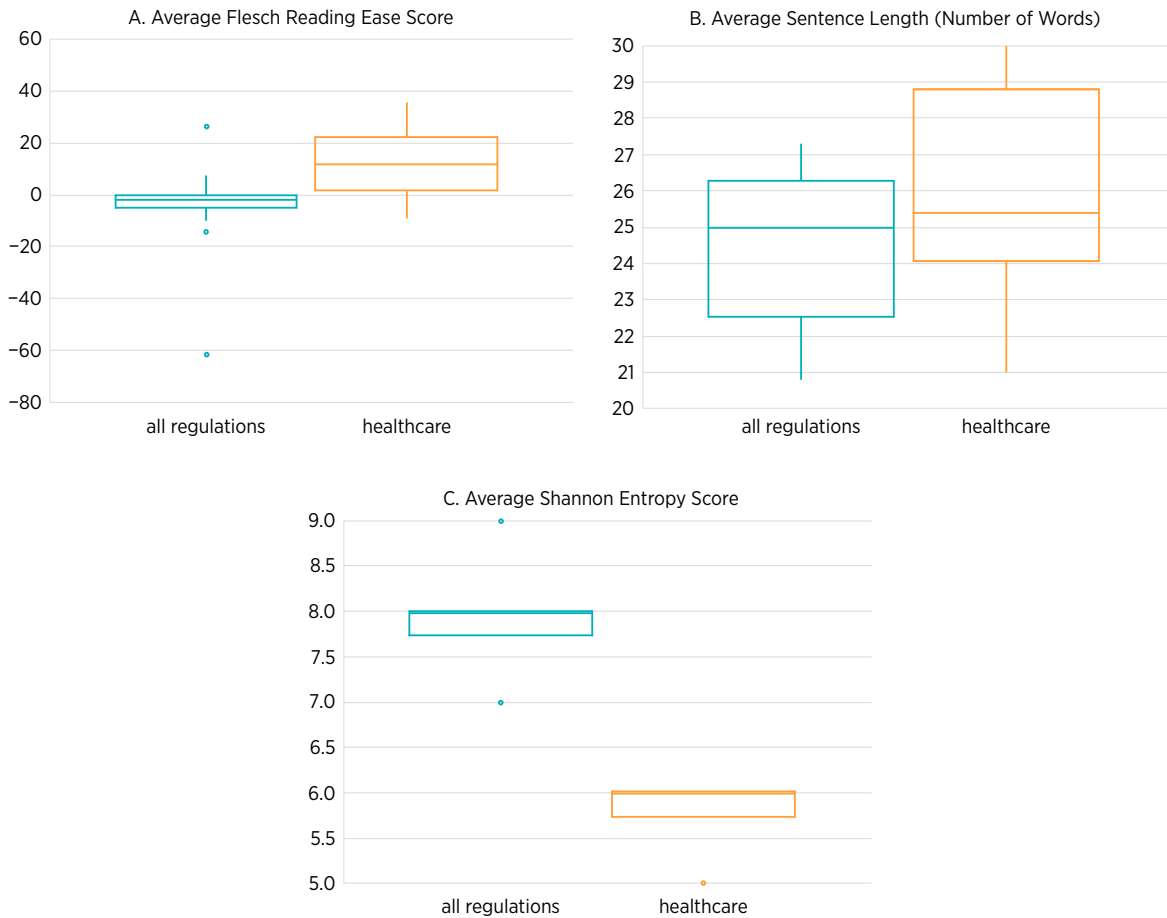
Source: Authors' calculations using Patrick A. McLaughlin and Oliver Sherouse, RegData US 3.1 Annual (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2018, <https://quantgov.org/regdata-us/>.

As a rule of thumb, documents with scores between 80 and 100 are considered easy to read. Documents with scores under 60 are considered difficult to read.

Table 4 shows that the average federal regulation is quite difficult to read and readers require high levels of education and subject-matter expertise to understand it. The average, however, masks significant differences across agencies. For example, HHS, which issues more than 85 percent of all healthcare regulations, has a Flesch score of -2.4, compared to scores of 35.5 and 25.1 for the Department of the Treasury and the Executive Office of the President, respectively. In general, the average sentence length correlates with the Flesch score, because the formula for the Flesch score considers average sentence length.

As figure 1 shows, the readability and complexity of healthcare regulations track with those of all other regulations in terms of reading ease. As shown in panels A and B, healthcare regulations have higher readability scores (meaning they are easier to read compared to other regulations) but slightly higher average sentence length. However, healthcare regulations are less complex in terms of the number of new ideas and concepts they contain compared to other regulations. In panel C of figure 1, we observe that healthcare regulations have lower Shannon entropy scores. To put

Figure 1. Readability of US Federal Healthcare Regulations



Source: Authors' calculations using Patrick A. McLaughlin and Oliver Sherouse, RegData US 3.1 Annual (dataset), QuantGov, Mercatus Center at George Mason University, Arlington, VA, 2018, <https://quantgov.org/regdata-us/>.

the Shannon entropy score into context, a typical Shakespeare play gets a Shannon entropy score of 8.0. At 5.9, US healthcare regulations are relatively less complex than the typical Shakespeare play and than other regulations.

CONCLUSION AND NEXT STEPS

Healthcare regulations have played an important role in shaping the healthcare sector over the years. Health RegData is an attempt to quantify the volume of healthcare regulations and their impact on the healthcare industry and health outcomes. Unsurprisingly, the Centers for Medicare and Medicaid Services within HHS has issued the most healthcare regulations—33 percent, followed by the FDA and the Public Health Service with 28 percent and 10 percent respectively. The industries most impacted by these regulations are ambulatory healthcare services, insurance carriers, and chemical manufacturing.

In this first edition of Health RegData, we have provided a snapshot of the volume of healthcare regulatory restrictions. This is merely the first step and the baseline for subsequent analyses. Over the next few months, we will examine the growth of healthcare regulations from 1996 to the present. In addition, we will examine the volume of healthcare regulations in the states and classify these regulations into various topics of interest. Some of these topics include occupational licensing, certificate-of-need laws, and public health. These analyses will allow researchers to examine the role of healthcare regulations in shaping the evolution of healthcare in the United States. In addition, the data on the complexity of regulations should be useful to legislators and regulators as they create or manage regulations.

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.

Jonathan Nelson is a software developer at the Mercatus Center at George Mason University. He writes code for the Policy Analytics team and has worked on RegData, Quantgov, and other related projects. He is a graduate of Grove City College, where he studied economics, mathematics, and philosophy.

Elise Amez-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. See Patrick A. McLaughlin and Oliver Sherouse, "RegData 2.2: A Panel Dataset on US Federal Regulations," *Public Choice* 180, no. 1-2 (2019): 43-55.
2. See Omar Al-Ubaydli and Patrick A. McLaughlin, "RegData: A Numerical Database on Industry-Specific Regulations for All United States Industries and Federal Regulations, 1997-2012," *Regulation and Governance* 11, no. 1 (2017): 109-23.
3. See Rudolph Flesch, "A New Readability Yardstick," *Journal of Applied Psychology* 32, no. 3 (1948): 221-33.
4. See *Federal Plain Language Guidelines*, rev. 1, May 2011, <http://www.plainlanguage.gov/howto/guidelines/FederalPLGuidelines/FederalPLGuidelines.pdf>.
5. Negative values for Flesch readability scores are possible because of the formula: $206.835 - (1.015 \times \text{ASL}) - (84.6 \times \text{ASW})$, where ASL = average sentence length and ASW = average number of syllables per word. Negative scores simply mean the unit of text is very difficult to read.