Health Regulators as Data Stewards

Kristin Madison
HEALTH REGULATORS AS DATA STEWARDS*

KRISTIN MADISON**

INTRODUCTION ..........................................................1605
I. MAKING DATA BIGGER......................................................1609
   A. Federal Government as Data Generator .......................1611
   B. Federal Government as Data Collector .......................1612
   C. Federal Government as Data Aggregator ....................1614
   D. Federal Government as Data Facilitator ....................1617
   E. Federal Government as Data Funder ........................1620
II. MAKING DATA SMALLER ..................................................1621
   A. Patient Decision Aids .............................................1622
   B. Patient-Centered Outcomes Research Institute .............1623
   C. Meaningful Use Regulations .................................1624
   D. Health-Related Quality Reporting ...........................1625
III. REMAKING DATA ..........................................................1628
CONCLUSION ........................................................................1636

INTRODUCTION

“Big data” has come to health care.¹ It is now widely agreed that rapidly improving abilities to pull together and analyze massive datasets will have the potential to transform health, health care, and the health care system. Corporate websites tout the benefits of big data-based technologies for improving patient care, expanding health care access, and managing health care costs.² Countless conferences

---

* © 2014 Kristin Madison.
** Professor of Law and Health Sciences at Northeastern University. I thank Melissa Jacoby for her suggestions, Frank Pasquale for helpful conversations, and Joan Krause and Richard Saver for their contributions to the symposium for which this Article was written. I also thank the staff of the North Carolina Law Review for their work on the symposium and for their improvements to this Article.


have discussed the future of big data in health care. Consultants have trumpeted the arrival of a big data “revolution” in the health industry, and the mainstream press, policymakers, and legal academics have all now turned their attention to the topic.

Big data’s transformative potential arises from the information it could generate for many different types of users, including health care providers, payers, patients, and regulators. Health care system stakeholders make countless decisions every day that influence the care that patients receive and ultimately patients’ health. Those decisions will nearly always turn on the information available to the decision maker. What types of information exist, who is generating that information, and how that information is gathered can have a profound effect on the choices that are made.

In the health care setting, private actors have often played a leading role in launching big data initiatives. Large, private insurers and large, sophisticated health care providers are well positioned to harness the power of big data. Private insurers have used their data to test the relationship between medical treatments and patient


outcomes\(^8\) and to engage in systematic study of cost trends.\(^9\) Private health care providers have looked to big data as a tool for improving their patient care\(^10\) and general operations.\(^11\) One of the biggest holders of big data, however, is the federal government, which processes over one billion health care claims each year for Medicare alone.\(^12\) If the data embodied within these claims could be transformed into useful information, the data would have the potential to affect countless health care decisions made all over the United States.

The federal government is more than just a big data repository, however; it has become a data steward.\(^13\) The term "data stewardship" can have many meanings, and it is sometimes associated with the responsibility for protecting the integrity and confidentiality of data.\(^14\) This is certainly one task that the federal government has taken on, both with respect to its own programs and with respect to data held by other regulated entities. Many health policy and law

---


11. See Terry, supra note 1, at 409 (describing big data initiatives).


13. To a lesser extent, state governments and other public and private entities have also become data stewards. Given the extent of the federal government’s ongoing data-related activities, however, this Article will focus on federal initiatives in this area.

14. See, e.g., SUSAN BAIRD KANAAN & JUSTINE M. CARR, U.S. DEP’T OF HEALTH & HUMAN SERVS., HEALTH DATA STEWARDSHIP: WHAT, WHY, WHO, HOW 1 (Sept. 2009), available at http://www.ncbi.nlm.nih.gov/090930lt.pdf (defining “health data stewardship” as “a responsibility . . . to ensure the knowledgeable and appropriate use of data derived from individuals’ personal health information” and highlighting as “essential practices and principles . . . transparency about use; . . . security safeguards and controls; de-identification (when relevant); data quality, including integrity, accuracy, timeliness, and completeness; limits on use, disclosure, and retention; oversight of data uses; accountability; and enforcement and remedies”).
scholars, including Nicolas Terry, Frank Pasquale, Barbara Evans, Deven McGraw, and Alice Leiter, have explored the implications of current federal privacy and confidentiality laws in a world of big data. The work of these authors and others makes clear that addressing concerns related to privacy and confidentiality is a critical step in efforts to take full advantage of the promise of big data. Many individuals value privacy in the health care sphere and may be reluctant to share information if it is at risk of being disseminated too widely. To address the concerns of such patients, policy makers and others may seek to limit the collection, aggregation, and use of patient data, or they may instead seek to develop robust privacy and confidentiality policies that offer the protections some patients prize.

While privacy and confidentiality issues are important and need to be addressed, this Article looks beyond them to consider the broader responsibilities associated with data stewardship. One of the meanings of “steward” is “one who actively directs affairs”—stewards manage things. In recent years, through its laws, policies, and programs, the federal government has taken on an increasingly important role in managing the flow of health-related data. By doing so, it has affected health care decision making and accelerated the process of health care reform.

Part I of this Article explains that one of the federal government’s most important functions as a data steward has been to make data bigger. Most obviously, the government adds to big data through the generation and sharing of claims data from Medicare. But it also expands data in myriad other ways; through numerous recent initiatives, the federal government has served as a data collector, aggregator, facilitator, and funder.

Part II suggests that at the same time, federal agencies have also sought to make data smaller. While amassing data can be an important first step in generating the information critical for health

15. See generally Barbara J. Evans, The Ethics of Postmarketing Observational Studies of Drug Safety Under Section 505(o)(3) of the Food, Drug, and Cosmetic Act, 38 Am. J. L. & Med. 577 (2012) (discussing HIPAA’s impact on pharmaceutical companies’ use of claims data and electronic health records to conduct postmarketing observational studies); Deven McGraw & Alice Leiter, A Policy and Technology Framework for Using Clinical Data to Improve Quality, 12 Hous. J. Health L. & Pol’y 137 (2012) (discussing implications of HIPAA, the federal Common Rule, and other laws for the use of electronic health records to improve health care quality); Pasquale, supra note 7 (emphasizing the importance of well-designed intellectual property and privacy laws and policies to ensure that the promise of health information technology is realized); Terry, supra note 1 (describing changes needed in law and policy to protect patient privacy in an era characterized by the growth of big data).

care reform, these data need to be analyzed and distilled before they can be used effectively by health system stakeholders. Such analysis is particularly challenging for patients, and the federal government has taken numerous steps to make information for patients both understandable and actionable. Part II examines several examples of programs that help patients tackle complex decision making, including the Patient-Centered Outcomes Research Institute ("PCORI") and the public posting of health care quality metrics.

Part III argues that while the efforts to make data bigger and smaller have done much to lay the foundation for improved decision making by all health care stakeholders—including payers, providers, and patients—more remains to be done to improve decision making by one other key health care system stakeholder: the government itself. As data stewards, health care regulators should ensure that they manage data in such a way as to better inform their own regulatory decisions. Some initiatives in this area are already underway, but health regulators, and, by extension, patients and taxpayers, would likely benefit from focusing more attention to efforts to build an evidence base for regulatory and programmatic interventions. Such efforts will often involve making data bigger and may sometimes involve making data smaller. Ultimately, however, they must include a systematic effort to remake data.

I. MAKING DATA BIGGER

While the term "big data" is of relatively recent vintage, it has arguably been integral to health care for many years. Data are a critical input into many aspects of the health care delivery system. The claims submitted to Medicare, Medicaid, and other public programs and private insurers are used not just to perform these entities' core payment functions but also to manage their broader operations. Data extracted from these claims have also long been

17. See infra Part III.
used by providers, researchers, and others seeking to learn more about health and health care delivery. But the sheer volume of these data, especially when combined with limited computing capacity, has often meant that users, as a practical matter, could perform analyses with only a subset of potentially useful data.

In this era of big data, however, the technical constraints on computing have loosened, allowing data to be more easily collected, stored, and analyzed. The lower cost associated with these tasks has allowed data to get even bigger and has made data-intensive analyses much more feasible in many settings. Entities in a position to collect data as part of their operations, such as payers and providers, are capable of collecting and storing more data than ever before, and data not systematically collected previously—such as data about purchasing patterns—can be gathered. These data can inform product development, marketing, community health needs assessments, health care quality evaluation, health regulation, and research in the areas of medicine, health care, and public health.

The federal government has worked to expand the availability of data that could be useful for many of these functions. In addition to

---


23. See id.

generating and sharing its own data, the federal government has acted as a data collector, data aggregator, data facilitator, and data funder.

A. Federal Government as Data Generator

The federal government continues to contribute significantly to the growth of big data through its role as a data generator, and equally importantly, through sharing of the data it generates. The Medicare program makes much of its data available in various forms to researchers and the public. Given patient privacy concerns, there are limits to how much federal claims data can be shared. Recent developments suggest an increased willingness to make claims data more widely available, however. One example is that for many years, an injunction blocked the release of Medicare physician claims to the public based on the implications of such a release for physician privacy. This injunction was lifted in 2013, and the Centers for Medicare and Medicaid Services (“CMS”) subsequently solicited public comment about the policies that should be adopted with respect to the release of physician claims data. In 2014, CMS announced a new policy under which it would determine on a case-by-case basis whether physician payment data could be released in response to Freedom of Information Act requests. It subsequently released claims data for over 880,000 health care providers.


26. See supra note 15 and accompanying text.


29. See id. at 1357.


B. Federal Government as Data Collector

Medicare claims data are generated as a byproduct of program operations, not as a result of a deliberate effort to expand data available for researchers or policy makers. The federal government has undertaken a number of initiatives, however, to collect other types of health-related information. Health care researchers often rely on data collected through the census and a number of other important surveys conducted through the Centers for Disease Control and Prevention ("CDC") and other agencies. Commonly used surveys include the National Health Interview Survey,\(^{33}\) the National Health and Nutrition Examination Survey,\(^{34}\) the Medicare Current Beneficiary Survey,\(^{35}\) and the Medical Expenditure Panel Survey.\(^{36}\)

One of the fastest-growing forms of federal data collection, however, involves information that is associated with public insurance programs but goes beyond the bare-bones fee-for-service claims data that are at the traditional core of program operations.\(^{37}\) One example is health care quality data. While insurers, accreditation organizations, state governments, and other entities may all seek quality-related information from providers,\(^{38}\) the federal government has often acted as a leader in this area. In 2003, Congress altered


\(^{37}\) See THORPE, PEREIRA & ROSENBAUM, supra note 19, at 2 ("Medicare claims data, like those maintained by private health insurers, are used primarily to pay claims.").

Medicare payment formulas to encourage hospitals to participate in a reporting system.\textsuperscript{39} Today, over 1,300 hospitals are participating in a value-based purchasing program in which reimbursement levels are tied to data the hospitals provide about quality, including infection rates and mortality rates.\textsuperscript{40} Also included in the program are data drawn from patient responses to surveys about their own experiences.\textsuperscript{41}

CMS collects quality-related information from other providers as well. In 2006, physicians were given payment incentives to voluntarily participate in a reporting system.\textsuperscript{42} By 2015, physicians covered by the program will receive a reduction in payment if they do not participate, and by 2017, all physicians who participate in Medicare will be subject to the value-based payment modifier.\textsuperscript{43} CMS has quality initiatives underway for home health agencies\textsuperscript{44} and nursing homes,\textsuperscript{45} and the Patient Protection and Affordable Care Act ("ACA") mandated the creation of quality reporting programs for long-term care hospitals and hospice programs.\textsuperscript{46} As the federal government moves away from a public insurance payment system based only on the quantity of services rendered toward systems that


\textsuperscript{41} Medicare Program, 78 Fed. Reg. at 50,678–80 (listing patient experience measures among quality measures used in program).

\textsuperscript{42} See Madison III, supra note 39, at 79.


\textsuperscript{46} See Patient Protection and Affordable Care Act § 3004, 42 U.S.C. § 1395ww(j), (m) (2012).
involve more careful examination of the nature of these services, the pool of data available for analysis will necessarily expand.

C. Federal Government as Data Aggregator

Data generated in connection with Medicare and Medicaid are likely to fit within most people's definitions of big data, given the size and scope of these programs. But for many potential uses of big data, these data are not big enough. The data are associated with beneficiaries of these public programs, not the majority of Americans who are privately insured, and claims data do not capture the full wealth of data available in electronic medical records. Private insurers and health care providers have access to vast stores of data on millions of Americans that could help guide health care system reform and support decision making among many system stakeholders. Access to data would provide researchers and analysts with a more comprehensive picture of health, health care, and health care financing in the United States. There are many barriers to sharing these data, including concerns (and laws) related to patient privacy, proprietary interests in the data, and the transaction costs involved in attempting to reach out to many different entities in our fragmented health care system, but the federal government has worked to overcome these barriers.

As an initial step, CMS has been able to promote aggregation by releasing its data so that it can be added to data held by others. The ACA required the release of Medicare claims data involving hospital services, physician services, prescription drugs, and other services and supplies to public and private entities seeking to pool these data with

47. CARMEN DENAVAS-WALT, BERNADETTE D. PROCTOR & JESSICA C. SMITH, U.S. CENSUS BUREAU, INCOME, POVERTY, & HEALTH INSURANCE COVERAGE IN THE UNITED STATES: 2012, at 26 tbl.8 (2013) (showing that in 2012, nearly sixty-four percent of people had private coverage and that fifty-two percent had private coverage only).

48. See Panzer et al., supra note 38, at 1972–73 (discussing limitations of claims data relative to "rich clinical data").


50. Cf. GROVES ET AL., supra note 4, at 12 (noting that big data innovations often make use of proprietary data).

other data for the purposes of evaluating provider performance.\textsuperscript{52} These entities would be required to release the quality ratings to the public.\textsuperscript{53} By facilitating the pooling of public and private data, this provision would increase the information available about the performance of individual providers, helping to address the data limits that so frequently plague efforts to develop reliable provider quality measures.\textsuperscript{54}

Another way in which the federal government can engage in data aggregation is by facilitating interactions among private entities. One example of this approach is the Sentinel Initiative,\textsuperscript{55} under which "the FDA seeks to create a scalable, efficient, extensible, and sustainable system . . . that leverages existing electronic health care data from multiple sources to actively monitor the safety of regulated medical products."\textsuperscript{56} In the Food and Drug Administrative Amendments Act of 2007,\textsuperscript{57} Congress required the Secretary of Health and Human Services ("HHS") to work with "public, academic, and private entities" to "develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources" that would include at least one hundred million patients by 2012.\textsuperscript{58} The statute further requires the creation of procedures that would allow for monitoring for adverse drug events using federal data such as Medicare and Veterans Affairs data as well as private sector data such as data from drug purchases


\textsuperscript{54} Limited data affect many efforts to measure provider quality, particularly at the level of an individual provider. See, e.g., David J. Nyweide et al., \textit{Relationship of Primary Care Physicians' Patient Caseload with Measurement of Quality and Cost Performance}, 302 \textit{JAMA} 2444, 2444 (2009) (concluding that "[r]elatively few primary care physician practices are large enough to reliably measure 10% relative differences in common measures of quality and cost performance among fee-for-service Medicare patients").

\textsuperscript{55} For a thorough discussion of the Sentinel Initiative and an evaluation of the possibility that the FDA might be able to require participation in it, rather than merely convene voluntary participants, see generally Barbara J. Evans, \textit{Authority of the Food and Drug Administration to Require Data Access and Control Use Rights in the Sentinel Data Network}, 65 \textit{FOOD \& DRUG L.J.} 67 (2010).


and health insurance claims. The procedures must also permit the identification of and reporting on trends and patterns of adverse drug events. Under a pilot program, the mini-Sentinel, the data would not be joined into a single database, but instead would be held by the institutions in which the data originated and then transmitted across “a distributed data network that is linked by a coordinating center.” Participants in this initiative include numerous entities affiliated with major insurers.

Another example of a federal aggregation effort is the eMERGE Network, an initiative funded by the National Institutes of Health (“NIH”) that aims to “develop, disseminate, and apply approaches to research that combine DNA biorepositories with electronic medical record . . . systems for large-scale, high-throughput genetic research” that “brings together researchers . . . from leading medical research institutions across the country.” The network is intended to foster collaboration in genetic research through shared expertise, access to shared tools, and the use of pooled data. Participants in the network agree to submit genetic data to a coordinating center that will then combine the data with the network dataset and submit them to the database of Genotypes and Phenotypes, which makes individual-level genetic data available to researchers.

Another example of an NIH-sponsored data aggregation program is the Health Care Systems Research Collaboratory. Like the eMERGE Network, this program features a coordinating center

60. See id. § 355(k)(3)(C)(IV)-(V).
that facilitates the dissemination of "data, tools, and resources." Members of the Collaboratory, which currently include organizations such as the Duke Clinical Research Institute, the Harvard Pilgrim Health Care Institute, and the Group Health Research Institute, "work with the NIH to produce, document, and disseminate standards, and to create durable infrastructure that facilitates multicenter studies and reuse of data." One Collaboratory trial involves a study of the impact of hemodialysis sessions of at least four hours; it will involve several hundred dialysis facilities where care is routinely provided, and it will use data routinely collected as part of care provision. One goal of the Collaboratory is to make better use of routinely collected data in real-world care settings.

D. Federal Government as Data Facilitator

When the federal government supports aggregation through programs such as the eMERGE Network or the Collaboratory, what it is really doing is facilitating interactions among private entities. By exercising leadership and designating an entity to serve in a coordinating role, it helps interested institutions surmount barriers that might otherwise prevent collaboration and the sharing of data. By contrast, other federal initiatives accelerate the growth of data by facilitating their development and dissemination. The federal government directly develops data when it collects data as part of its own operations or mandates reporting as part of its regulatory functions. At the same time, it may facilitate data growth among private entities. One of the best examples of this phenomenon is the federal effort to promote the growth of electronic health

---

67. Id.
70. See supra notes 63–65 and accompanying text.
71. See supra notes 66–69 and accompanying text.
72. For a discussion of how it is also funding these interactions, see infra notes 90–93 and accompanying text.
73. See supra Part I.B.
records (and the data they contain) through the HITECH Act and the associated "meaningful use" regulations. The HITECH Act sought to vastly expand the use of electronic health records through what might be thought of as three fundamental strategies: subsidies, standards, and supports. It made billions of dollars in payments—more than $40,000 per physician—available to providers who adopt and use electronic health records. To receive the rewards, providers are required to meet a series of objectives, such as recording a certain percentage of patients' demographic data as structured data or submitting a certain percentage of prescriptions electronically. These objectives are collectively known as "meaningful use" requirements. They are complemented by a set of standards by which electronic health record systems will be certified as supporting compliance with these requirements. Supports for the expansion of electronic health records include workforce training programs and regional extension centers that offer assistance to health care providers seeking to make the transition to electronic health records.

Changing the medium of medical record storage from paper to electronic media does not in itself make more data available to users. But the meaningful use regulations require that records be used for certain purposes, such as reporting clinical quality measures or maintaining active medication lists, which will likely mean that more

75. See 42 C.F.R. § 495.6 (2013) (establishing meaningful use objectives); 45 C.F.R. § 170.314 (establishing electronic health record certification criteria).
76. See HITECH Act, HEALTHIT.GOV, http://www.healthit.gov/policy-researchers-implementers/hitech-act (last visited May 7, 2014) (noting that the HITECH Act "seeks to improve American health care delivery and patient care through an unprecedented investment in Health IT" and that its "programs collaboratively collectively build the foundation for every American to benefit from an EHR").
79. See Blumenthal & Tavenner, supra note 77, at 501.
80. See EHR Incentives & Certification, supra note 78 (describing the role of certification); see also 45 C.F.R. § 170.314 (establishing electronic health record certification criteria).
81. See Melinda Beeuwkes Buntin, Sachin H. Jain & David Blumenthal, Health Information Technology: Laying the Infrastructure for National Health Reform, 29 HEALTH AFF. 1214, 1216 (2010).
data will be captured initially than might otherwise be the case.\textsuperscript{82} Moreover, the requirements associated with HITECH will ensure that, once captured, the data will flow more easily to other health-related entities. For example, the Stage 2 meaningful use criteria push data sharing forward by including as objectives the ability to submit electronic data to immunization registries, cancer registries, and public health agencies.\textsuperscript{83} HITECH also provided more than a half billion dollars to support the development of health information exchanges, through which electronic data could flow from one entity to another.\textsuperscript{84} Improved technologies that allow for less costly access to data can potentially support many different activities, including health care quality monitoring, medical research, and public health surveillance.

The federal government can also facilitate the dissemination of data through programs that make data easier to find and access. The most prominent example of this kind of project is the website HealthData.gov, which brings together over a thousand health-related datasets, from Medicare cost report data to CDC data to Food and Drug Administration ("FDA") recall information.\textsuperscript{85} The purpose of the website is to "mak[e] high value health data more accessible to entrepreneurs, researchers, and policy makers in the hopes of better health outcomes for all."\textsuperscript{86} The website provides access to data not previously available and seeks to offer data in a form that is "machine-readable, downloadable and accessible via application programming interfaces" so that it can be more easily used.\textsuperscript{87}

In addition, just this year, CMS announced these data would be available more rapidly and at a lower cost through a new initiative,
the Virtual Data Research Center. Given the traditionally heavy reliance of health services researchers on Medicare claims data, this initiative is likely to further promote the publication of studies examining the U.S. health care delivery system.

E. Federal Government as Data Funder

One of the most important ways that the federal government makes big data bigger is through direct funding of projects involving data generation, collection, or dissemination. In recent years, annual congressional appropriations for the NIH, which funds health research, have been in the range of thirty billion dollars. The Agency for Healthcare Research and Quality conducts and sponsors research on health care delivery. Many of the previously discussed initiatives that involved private entities were supported by grant funding, and the billions of dollars appropriated under the HITECH Act undoubtedly have accelerated the adoption of electronic health records. Federal funding has played an important role in expanding the availability of data.

89. See id. ("Studies that help us better understand the healthcare ecosystem are published almost weekly in major medical journals based on CMS data.").
93. See Blumenthal & Tavenner, supra note 77, at 501 (noting that incentive payments will total up to $27 billion).
II. MAKING DATA SMALLER

Federal agencies are uniquely positioned to foster the growth of data, given the nature of their services and the scope of their authority. The traditional governmental role as a provider of public goods is consistent with the federal role in supporting research activities including data collection, analysis, and dissemination. The federal role as a sponsor of public health insurance covering a large fraction of the American population will often mean that federal agencies will play a central role in efforts to secure access to large, comprehensive datasets. But these are not the only tasks that federal entities have taken on. While much federal effort is focused on making data bigger, federal agencies have also sought to make data smaller. In recent years, the government’s enthusiasm for promoting wider availability of data has been matched by a zest for distilling it into smaller, more usable forms.

The idea that data stewardship might involve efforts to make data smaller makes perfect sense in a policy and practice environment characterized by an emphasis on patient-centeredness. The concept of patient-centeredness, identified by the Institute of Medicine in a 2001 report as one of six goals for health care delivery, involves a focus on the “needs, values, and expressed preferences of the individual patient,” and patient education is often mentioned as an important dimension of patient-centered care. Initiatives that seek to communicate health and health care-related information in a way that patients can understand and use are therefore consistent with the goal of patient-centeredness. Patients frequently face daunting challenges in navigating the complexities of the health care system at a time of particular personal vulnerability. To be effective, the communication process will therefore often involve simplifying the presentation of information, tailoring information to individual

95. See Rexford E. Santerre & Stephen P. Neun, Health Economics: Theories, Insights, and Industry Studies 230 (3rd ed. 2004) (“One legitimate function of government is to provide public goods . . . . Government ensures that public goods are produced in either the private or public sector and collects the necessary funding through taxation.”).


patient needs and preferences, and targeting information directly to the patient, so that the information reaches and engages the patient. Simplifying, tailoring, and targeting could all be thought of as ways of making data smaller. A variety of federal initiatives have taken on these tasks.

A. Patient Decision Aids

The ACA demonstrates both a commitment to patient-centeredness and an associated emphasis on making data smaller.98 As one commentary explains, the ACA "repeatedly refers to patient-centeredness, patient satisfaction, patient experience of care, patient engagement, and shared decision-making in its provisions."99 One example of a provision fitting this description concerns patient decision aids, which the ACA defines in part as "an educational tool that helps patients . . . understand and communicate their beliefs and preferences related to their treatment options, and to decide . . . what treatments are best."100 The provision requires the creation of standards against which decision aids will be evaluated and the creation of a process for certifying those meeting the relevant standards.101 In addition, it calls for a program that would fund the development of patient decision aids that would, among other things, present up-to-date clinical evidence about the risks and benefits of treatment options in a form and manner that is age-appropriate and can be adapted for patients . . . from a variety of cultural and educational backgrounds to reflect the varying needs of consumers and diverse levels of health literacy.102

In other words, the ACA embraces a federal policy goal of ensuring that clinical information is made available to patients in a form tailored to their needs. Thus, while the federal government helps make data bigger by supporting clinical research,103 it also helps make

99. Id. at 1.
102. Id. § 299b-36(d).
103. See supra notes 90–94 and accompanying text.
data smaller by increasing the likelihood that research results are delivered to patients in an easily usable format.

B. Patient-Centered Outcomes Research Institute

Another ACA program that falls within the “making data smaller” category is the Patient-Centered Outcomes Research Institute (“PCORI”). PCORI has often been described as an institution that sponsors comparative effectiveness research; it funds studies that compare the effects of multiple approaches to treating a particular disease or condition. As its official name indicates, however, patients—not treatment methods—are at the very center of its mission. PCORI describes its mission as “help[ing] people make informed health care decisions, and improv[ing] healthcare delivery and outcomes, by producing and promoting high integrity, evidence-based information that comes from research guided by patients, caregivers and the broader healthcare community.” Its description of its vision reinforces the small data nature of PCORI’s work: “Patients and the public have information they can use to make decisions that reflect their desired outcomes.” The information must be ultimately usable by patients, not just researchers or medically trained clinicians, and it must take a form that allows patients to understand its implications for the outcomes they care about.

PCORI’s early projects suggest a close adherence to this vision. For example, one funded study will evaluate a toolkit that helps clinicians identify “the type of treatment most likely to be successful based on the different pain experiences reported by the patient.”

104. See 42 U.S.C. § 1320e(b).

105. See, e.g., Alan M. Garber, How the Patient-Centered Outcomes Research Institute Can Best Influence Real-World Health Care Decision Making, 30 HEALTH AFF. 2243, 2243 (2011) (describing PCORI as “created to set priorities and fund comparative effectiveness research while ensuring the production of high-quality studies that adhere to rigorous standards”).


108. Id.

while another will "provide culturally tailored information for Latina adolescents and their parents to help in making decisions on whether or not to receive the human papilloma-virus vaccination." The choices PCORI has made evidence a federal commitment to tailoring information to the needs of individual patients.

The patient decision aid and PCORI examples illustrate a common phenomenon: efforts to make data smaller often accompany efforts to make data bigger. PCORI will promote the growth of comparative effectiveness research in general but will also target research toward the populations that most benefit from it. Medical research continues to produce new results, but patient decision aids are critical to ensuring that individuals use results appropriately.

C. Meaningful Use Regulations

The meaningful use regulations might also be said to fit the small-data-as-counterpart-of-big-data mold. The HITECH statute sought to vastly expand the adoption of electronic health records through subsidies, standards, and supports, a step that will facilitate future research, including big data projects. But the meaningful use standards that accompanied the HITECH statute would simultaneously help to make data smaller by promoting the sharing of data tailored to an individual patient's needs.

Like PCORI and patient decision aids, meaningful use regulations are guided by the concept of patient-centeredness. These regulations encourage medical professionals to reach out to patients based on patients' likely needs. According to the objectives specified by the regulations, medical professionals should have the ability to use their electronic health records to generate lists of patients by condition, which would permit better monitoring and follow-up with patients. Professionals should be able to identify patients who would benefit from reminders; they should also be able

110. Id.
111. See supra notes 74–84 and accompanying text.
to identify educational resources tailored to the needs of specific patients.\textsuperscript{114}

Other objectives focus on facilitating patients' ability to obtain their own data. Electronic health records must support the provision of clinical care summaries to individual patients and allow patients to download their own data.\textsuperscript{115} The specific standards required for achieving meaningful use reinforce these general objectives; one of the measures of patient use, for example, is whether five percent of patients actually view, use, or transmit their own records.\textsuperscript{116}

\textbf{D. Health-Related Quality Reporting}

A final example of a federal initiative in which a focus on making data smaller is embedded within a broader goal of making data bigger is health care quality reporting. As discussed in Part I, federal health agencies have sought to collect health care quality-related data, to aggregate public and private data so as to support the development of better quality metrics, and to facilitate the sharing of data more generally, some of which could be used in analyzing quality-related issues.\textsuperscript{117} Entities that access these data can use them to analyze provider quality, which can be an important input into many different sorts of decisions by health system stakeholders. Health care providers may be interested in assessing their own quality relative to their peers, while payers and policymakers may use the information to get a better sense of the value of health care provided and how it has changed over time. Sophisticated system stakeholders will often have the tools to analyze these data on their own.

For individual patients who must select health care providers, however, information about quality must be conveyed in a simple and straightforward way in order to be usable. CMS has committed itself to providing health care quality information for a growing number of

\begin{itemize}
  \item \textsuperscript{114} Id.
  \item \textsuperscript{115} Id.
  \item \textsuperscript{117} See supra Parts I.B–I.D.
\end{itemize}
types of health care providers. Its systematic, web-based health care quality reporting began in 2005 with the publication of ten quality measures for hospitals across the country. In 2008, it added patient experience ratings based on data from the Hospital Consumer Assessment of Healthcare Providers survey, as well as information about mortality rates for heart attacks, heart failures, and pneumonia. Since then, it has added a number of other hospital quality measures, including hospital readmissions. CMS also provides quality ratings for nursing homes, home health agencies, and dialysis facilities. Individual patients can visit these websites, type in their zip codes, and view quality measures associated with the nearest providers.

Evidence suggests that the number of individuals actually using quality ratings is relatively limited. A 2012 survey found that the percentage of respondents who had consulted online rankings or reviews of doctors and hospitals was in the range of fifteen percent, with the most educated and the middle aged being disproportionately more likely to consult such ratings. Nevertheless, the growth of federally sponsored quality reporting continues. The ACA, for example, mandated public reporting of physician quality information, along with quality reporting in other areas. Authors of a recent report examining the ACA’s focus on patient-centeredness identify nine distinct ACA provisions that require the use of measures of patient-centered care, four of which involve public reporting.

118. See supra notes 42–46, 52–54, 88–89 and accompanying text.
120. Id.
121. Id.
126. MILLENSON & MACRI, supra note 98, at 7.
Consumers have stressed that the usability of quality reporting hinges on a simple presentation of information. The complexity of quality metrics—especially when combined with the complexities of the health care system and the medical conditions prompting patients to seek ratings in the first place—can quickly overwhelm patients. CMS has therefore sought to make quality data smaller not just by producing quality metrics from its vast troves of data, but also by streamlining its presentation of these metrics. One way it has done so is by producing star ratings rather than just presenting statistical data. For nursing home care, for example, CMS presents ratings such as "above average" or "much below average," accompanied by ratings on a five-star scale; more detail is available only with additional clicks of a mouse. CMS also provides star ratings for Medicare Advantage plans and Medicare drug plans in which some Medicare beneficiaries choose to enroll. In 2012, CMS decided to make data even smaller for Medicare beneficiaries by sending enrollees of health and drug plans rated "poor" or "below average" for at least three years a personal letter stating, "We encourage you to compare this plan to other options in your area and decide if it is still the right choice for you."

In all of these examples, federal agencies have fostered the growth of data as part of their broader efforts to reform the health care system. Expanding medical research, increasing knowledge about the effectiveness of health-related interventions, building an electronic health record infrastructure, and increasing information about provider quality are all important ways that federal agencies have contributed to the data that will undergird future health care system operations. By extracting from these data-building initiatives information that is potentially relevant to patient decision making—

---

129. See id.
by making data smaller—these agencies have helped to ensure that patients can exercise more influence over their own care and play a more significant role in shaping the future health care system.

III. REMAKING DATA

Making data bigger and making data smaller are both important steps toward facilitating better decision making. More data, and more usable data, are beneficial for many different health system stakeholders. But careful thought should also be given to the nature of data being generated and the purposes these data serve.

Many of the previously discussed examples of data stewardship were useful for decision making by private actors. Federal research funding and data aggregation initiatives could change treatment decisions made by private physicians and coverage decisions made by private insurers.\textsuperscript{132} Mandates to collect and report data about aspects of health care delivery may alter providers’ decision making about the care they provide and patients’ decision making about the care they receive.\textsuperscript{133} Many federal data initiatives, including the HITECH statute’s facilitation of electronic health records\textsuperscript{134} and all of the initiatives to make data smaller, reduce the costs of information acquisition, analysis, and use for all system stakeholders. HITECH could facilitate the work of health care providers, streamline health insurance claims processing, and greatly lower the costs facing health care researchers. PCORI’s projects will ultimately facilitate patients’ decision making.

The potential impact of the federal government’s data stewardship efforts, however, is not limited to their influence on private individuals and entities. Information can be as valuable for government decision makers as it is for private decision makers. Medicare claims data are of course critical to the daily “decisions”


\textsuperscript{134} See supra notes 74–84 and accompanying text.
made by Medicare about whom to pay and how much, as is the information about quality that now finds its way into federal payment formulas. But the sphere of public decision making extends far beyond the mundane operations of public insurance programs. Congress and federal agencies make countless legislative and regulatory decisions that have a profound effect on the health care system and ultimately on population health. These decisions, like the decisions made by private providers, would also benefit from careful data stewardship.

Federal agencies regularly make use of the data they collect to analyze the effects of their own programs. The Government Accountability Office, for example, recently used Medicare claims data to investigate the implications of physician ownership of imaging equipment for the frequency of imaging. This study's results added to the evidence policy makers might take into account as they consider the appropriate scope of the prohibitions and exceptions of the Stark law, which limits the financial relationships among referring physicians and providers of certain health care services. There are many other examples of data collected in connection with program operations being used in identifying program issues and analyzing potential program reforms. Among the institutions created by the ACA was the Center for Medicare and Medicaid Innovation ("Innovation Center"), which was charged with the task of "test[ing] innovative service and delivery models to reduce program expenditures... while preserving or enhancing the quality of care." The Innovation Center will be heavily involved in


140. 42 U.S.C. § 1315a(a)(1).
analyzing data generated in connection with new models of service delivery and payment within public programs.\textsuperscript{141}

While claims data can be useful for decision-making purposes, data initiatives specifically designed to elicit data relevant for regulatory decision making would be even more helpful. Such initiatives are rare. The information void is particularly apparent in settings outside of public payment programs, where regulators do not have pre-existing databases to draw from when making regulatory decisions. The ACA imposes many new health-related obligations and limitations: it requires calorie labeling on chain restaurant menus,\textsuperscript{142} it mandates that drug and device manufacturers disclose financial relationships with physicians,\textsuperscript{143} and it limits employers' use of financial incentives to encourage healthy behaviors.\textsuperscript{144} What it does not do is establish federal programs to systematically collect data that would allow federal regulators to predict or assess the impact of the requirements they impose. Federal health agencies have devoted considerable attention to expanding data for use by a variety of decision makers and to distilling data for use by patients and others,\textsuperscript{145} but historically they seem to have devoted less attention to generating data for their own use in regulatory decision making.

Federal agencies have long sought to assess the likely impact of regulations before enactment; executive orders require regulators proposing new, economically significant regulations to conduct analyses of the regulations’ likely costs and benefits.\textsuperscript{146} These analyses reflect agencies’ predictions of regulatory consequences, based on evidence drawn from a variety of sources.\textsuperscript{147} Recently, there has been

\begin{footnotesize}
\begin{enumerate}
\item[143.] See 42 U.S.C. § 1320a-7h(a).
\item[144.] See id. § 300gg-4(j).
\item[145.] See supra Parts I, II.
\item[147.] See OFFICE OF MGMT. & BUDGET, supra note 146, at 2 (agencies must provide “a careful and transparent analysis of the anticipated consequences of economically significant regulatory actions”); id. at 3 (the agency should use “the best available scientific, technical, and economic information” and “generally rely on peer-reviewed literature”).
\end{enumerate}
\end{footnotesize}
more emphasis on the need to systematically assess regulations after they have been put in place. An executive order from 2011 states as a general principle that “[o]ur regulatory system ... must measure, and seek to improve, the actual results of regulatory requirements.”

It requires agencies to “consider how best to promote retrospective analysis of rules that may be outmoded, ineffective, insufficient, or excessively burdensome” and to develop “a preliminary plan ... under which the agency will periodically review its existing significant regulations.”

Cass Sunstein has emphasized the importance of assessing regulations’ actual effects, both intended and unintended.

Efforts to engage in some form of retrospective evaluation have begun; hundreds of regulatory reviews have already been completed. Cary Coglianese criticizes these retrospective reviews, however, as being “ad hoc and largely unmanaged.” To foster more systematic review, he proposes, among other steps, the adoption of a requirement that agencies “include in each prospective regulatory impact analysis” conducted as part of the regulatory process “a plan for the subsequent evaluation of the proposed rule.” This plan should specify metrics that could be used to assess whether regulatory objectives were met, to identify existing data or propose data that could be developed for use in the assessment, and to discuss potential research designs (including “sources of cross-sectional or longitudinal variation, other potential explanatory factors that might need to be controlled, and possible statistical approaches to estimating counterfactuals”).

Coglianese’s proposal is ambitious. Researchers in many fields can testify to the practical and financial barriers involved in identifying or developing relevant data and settling on appropriate

---

150. Id. at 3822. An executive order from 2012 reinforces these requirements and calls for further steps to be taken. See Exec. Order No. 13610, 77 Fed. Reg. 28,469, 28,469 (May 14, 2012).
151. Sunstein, supra note 148, at 1389 (“[R]egulations should be revisited and reviewed retrospectively, to ensure that they are promoting their intended functions, and are not producing excessive costs or unintended adverse side effects.”).
153. Id. at 59.
154. Id. at 62.
155. Id. at 62–63.
research methodologies. Assessing an intervention that occurs outside of a controlled environment, and that is thus subject to many confounding factors, is particularly challenging. The advantage of developing such research plans, however, whether done in conjunction with issuing a proposed regulation or even earlier, when the possibility of future regulation becomes apparent, is that it would allow for a much more robust assessment of potential regulatory impacts.

Both Coglianese and Sunstein cite to the work of Michael Greenstone, who has called for a "move toward a culture of persistent regulatory experimentation." Making an analogy to the FDA's drug approval process, Greenstone stresses the need for greater testing of regulations. He calls for more funding for evaluations and the creation of an independent review board that would assess the effectiveness of regulations. Greenstone also calls for small-scale implementation of regulations, which would accommodate the variation necessary to allow rigorous testing of the effects of regulation. He refers briefly to the possibility of quasi-experiments or randomized controlled trials of regulations, an idea that has been advocated as potentially feasible by policy groups and explored by other legal scholars. Sunstein notes that randomized experiments have "particular advantages" and that "experimental or quasi-experimental studies are preferred to focus groups," although focus groups can sometimes be useful in assessing regulations.


157. Coglianese, supra note 152, at 60 (citing Greenstone for a proposal to create a regulatory institution focused on retrospective review); Sunstein, supra note 148, at 1391-92 (discussing Greenstone's insights on regulatory experimentation).


159. See id.

160. See id. at 119–21.

161. See id. at 118–19.

162. Id. at 119.


165. Sunstein, supra note 148, at 1371.
In an ideal world, regulators would have access to data that would allow them to test potential regulations before they are implemented, as well as to continuously collect data that would allow them to monitor post-implementation effects. The executive order calling for retrospective analysis of rules "that may be outmoded, ineffective, insufficient, or excessively burdensome" is a step forward, but as Coglianese suggests, it would be useful to develop study methodologies and data collection plans prospectively. In addition, the executive order's call for retrospective review sounds somewhat akin to the FDA's requirements for the evaluation of safety and effectiveness of a drug, where the goal is to see if the drug works and what its problematic side effects might be, rather than to determine what drug works best. In an ideal world, regulatory evaluation would push toward the comparative evaluation now stressed in the health care setting. We do not want to know only whether a regulation works; we want to know which regulation works best.

To develop the informational foundation necessary for both prospective and retrospective evaluation, agencies must structure regulations to facilitate appropriate data generation, and Congress must give them authority to do so. In another article, I explore some mechanisms that federal agencies could potentially use to generate useful data in the health context. One possibility is the conditioned waiver: a waiver of existing regulations that would allow regulated entities to undertake otherwise prohibited activities, provided that they supply data that permits assessment of the impact of the waiver. Another possibility is regulating with variation, such as by allowing federal regulators to test the impact of restaurant menu labeling regulations by imposing different requirements in different geographic areas and/or by changing regulations in a systematic way over time.

A third possible approach that could provide at least some helpful data to regulators would be to make greater use of regulations that condition activities on detailed reporting. In other work, I have discussed recent regulations of employer health plans that make use

166. Coglianese, supra note 152, at 62–63 (proposing development of evaluation plans).
169. See id. at 806–14.
170. See id. at 800–06.
of financial incentives contingent on health standards, such as premium discounts for individuals with a body mass index below a certain threshold. Rather than simply raising the ceiling on the magnitude of incentives, regulators could condition the use of high levels of incentives on the disclosure of information that would give regulators more information about these programs, even if it is not enough to conduct a full evaluation. This approach would be less costly to regulated entities than a blanket reporting requirement, but it might still generate data that is useful in assessing whether further regulation might be warranted.

The prospect of systematically collecting information to support regulatory evaluation seems daunting, but there are some efforts underway to do so. In 2012, the Office of Management and Budget ("OMB") issued a memo that directs agencies to "include a separate section" in their budget submissions "on agencies' most innovative uses of evidence and evaluation" and notes that "[t]he Budget also will allocate limited resources for initiatives to expand the use of evidence." The memo calls for the implementation of evaluations using administrative data and evaluations linked to waiver authorities, among other steps, to encourage more evidence-based policy making. While recognizing the existence of legal, financial, and practical constraints of randomized trials of regulations, the Departments of Treasury, Labor, and Interior have all stated that they will consider using experimental designs to determine the impact of regulations.

One example of a commitment to regulatory assessment is found in the recently established Tobacco Centers of Regulatory Science, which will not just "increase understanding of the risks associated with tobacco use," but also "aid in ... evaluation of tobacco product regulations" and "help ... assess the impact of FDA's prior, ongoing, and potential future tobacco regulatory activities." But perhaps the

---

172. Id. (manuscript at 1–2).
174. Id. at 1–2.
175. See Sunstein, supra note 148, at 1392.
clearest example of a statutory and regulatory framework that supports systematic assessment is that associated with the Consumer Financial Protection Bureau (the "Bureau"). In creating the Bureau, Congress established a regulatory framework structured to permit systematic collection of data, encourage prospective regulatory experimentation, and require retrospective regulatory evaluation. Congress required the Bureau to monitor the risks posed to consumers in the financial product and service markets and granted the Bureau the authority to "gather and compile information from a variety of sources," including surveys and database reviews, and to require certain entities "to file ... annual or special reports, or answers in writing to specific questions ... as necessary for the Bureau to fulfill the monitoring, assessment, and reporting responsibilities imposed by Congress." In other words, Congress mandated systematic collection of data in an area in which the Bureau had the authority to regulate—data that the Bureau could potentially use to lay the foundation for future regulation.

In addition, Congress gave the Bureau authority to permit experimentation. More specifically, it granted authority to allow trial disclosure programs in which covered entities would be permitted to attempt to "improve upon any model form" issued by the Bureau. In essence, this authority allows regulated entities to propose their own experiments. The Bureau has expressed a willingness to authorize such experiments if the information they produce will be helpful. In a policy notice issued in 2013, the Bureau notes that "in-market testing, involving companies and consumers in real world situations, may offer particularly valuable information with which to improve disclosure rules and model forms." To obtain a waiver of otherwise applicable federal disclosure requirements, companies must test whether disclosure has indeed improved areas such as consumer use or understanding, and share this data with the Bureau.

---

177. The author thanks Melissa Jacoby for suggesting this example.
179. Id. § 5532(e)(1). In fact, the statute pushes the Bureau to not just allow, but encourage trial disclosures; the statute states that "standards and procedures issued by the Bureau shall be designed to encourage covered persons to conduct trial disclosure programs." Id. § 5532(e)(2).
180. Among the criteria used for granting waiver approval is the extent to which the proposed program will help the Bureau improve its disclosure rules. See Policy to Encourage Trial Disclosure Programs, 78 Fed. Reg. 64,389, 64,393 (Oct. 29, 2013) (policies B.1 and B.2).
181. Id. at 64,389.
182. Id. at 64,393 (policies A.3 and A.9).
Finally, within five years of a significant rule’s effective date, Congress has required the Bureau to assess the rule’s effectiveness in meeting statutory objectives and the Bureau’s goals based on “available evidence and any data that the Bureau reasonably may collect.”\textsuperscript{183} Together, these three aspects of the statutory framework undergirding the Bureau—the authorization and requirement to collect data, the waiver authority that would allow at least some variation in regulations, and the requirement for retrospective evaluation—ensure a long-term role for the Bureau as a data steward. Similar authorities and requirements could be put in place for health agencies.

The ultimate scope and impact of the Bureau’s data stewardship remain to be seen. Data collection and waiver authorities can be used sparingly or expansively; retrospective analyses may vary in their comprehensiveness. Nevertheless, the very existence of these data stewardship functions reflects a federal interest in developing and using new data sources to better inform the regulatory process. This orientation toward purposeful data collection and analysis is consistent with recent calls for retrospective regulatory evaluation, and if it is more broadly embraced by health regulators, it will provide a firmer foundation for future health-related regulation. By reshaping the data environment surrounding health regulatory functions—by remaking data—health-related agencies would be better positioned to ensure that their regulatory efforts are well spent.

CONCLUSION

It is apparent that the federal government’s role as a health data steward is rapidly growing. Numerous federal initiatives have focused on expanding data in ways that could ultimately be used by public and private payers to transform the health care system. At the same time, federal entities have embraced a commitment to distilling vast quantities of data into much smaller amounts of information tailored for and targeted to the needs of individuals. However, much remains to be done to cultivate data that could be used not just by external health system stakeholders or patients, but internally, by agencies themselves, to evaluate actual or proposed regulations. As health data stewards, federal agencies have an obligation to identify ways to gather the data necessary to support this critical function.

\textsuperscript{183} 12 U.S.C. § 5512(d).