



The Center on  
Health Insurance Reforms  
*Georgetown University Health Policy Institute*



Robert Wood Johnson  
Foundation

AUGUST 2015

# Big Data: A New Paradigm for Health Plan Oversight and Consumer Protection?

By Sabrina Corlette, Sandy Ahn, and JoAnn Volk

Support for this report was provided by a grant  
from the Robert Wood Johnson Foundation

## Authors



**Sabrina Corlette, J.D.**

*Research Professor and Project Director  
Center on Health Insurance Reforms  
Georgetown University Health Policy Institute*

**Sandy Ahn, J.D.**

*Research Fellow  
Center on Health Insurance Reforms  
Georgetown University Health Policy Institute*

**JoAnn Volk**

*Research Professor and Project Director  
Center on Health Insurance Reforms  
Georgetown University Health Policy Institute*

The Center on Health Insurance Reforms (CHIR), based at Georgetown University's McCourt School of Public Policy, is composed of a team of nationally recognized experts on private health insurance and health reform. We work regularly with a multidisciplinary group of faculty and staff dedicated to conducting research on issues related to health policy and health services.

CHIR faculty and staff study health insurance underwriting, marketing and products, as well as the complex and developing relationship between state and federal rules governing the health insurance marketplace. CHIR provides policy expertise and technical assistance to federal and state policy-makers, regulators and stakeholders seeking a reformed and sustainable insurance marketplace in which all consumers have access to affordable and adequate coverage.

## Introduction

Large data sets that can be analyzed to determine patterns of behavior – popularly called “big data” – are being used in ever-expanding ways. For example, big data is used to track consumer shopping patterns, understand environmental trends and prevent crime. In health care, physicians and scientists are using big data to help devise personalized treatments for diseases and track epidemics, and some major health systems are starting to use it to improve the quality of care.<sup>1</sup> More than a dozen states are using big data – through All-Payer Claims Databases (APCDs) – to better understand the pricing and use of health care services. Insurance companies were early adopters of big data, collecting and analyzing large volumes of information about the risks posed by prospective and current policyholders.

State insurance regulators have adopted the use of big data to conduct oversight of certain kinds of insurance, such as workers’ compensation and life insurance. And state and federal regulators regularly collect detailed data from health insurers to assess their financial solvency, as well as summary-level data to evaluate the reasonableness of premium rates and compliance with benefit standards. But those agencies providing oversight of health insurers have undertaken only modest efforts to collect, analyze and use large sets of claims, enrollment or sales data to understand market trends and how consumers are using their health insurance to access and pay for care.

### Big Data: *noun.*

**Extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.**

That could change, however, thanks to two yet-to-be-implemented provisions within the Patient Protection and Affordable Care Act (ACA) that contemplate a new regime of comprehensive data reporting by insurers and employer-based plans to both insurance regulators and

the public. The ACA ushered in sweeping reforms of the health insurance industry, prohibiting previously widespread practices such as the use of health status underwriting, gender rating, and the use of benefit design to discourage the enrollment of higher risk individuals. Policymakers recognized that with these practices prohibited, insurers might use other mechanisms to keep costs low that could undermine the ACA’s goal of ending health status discrimination. They also recognized that the reforms envisioned under the law are dependent on effective enforcement and greater transparency. Further, for the first time, federal taxpayers are subsidizing private health insurance companies through premium tax credits to defray the cost of coverage for millions of people. Such an outlay of tax dollars requires a high level of oversight. As a result, the ACA includes enhanced tools to monitor insurers’ compliance with the new standards.

Among these enhanced tools are expanded data collection authority and transparency requirements. Under the law, health insurers that market individual and group policies as well as employer-sponsored plans are required to report to the U.S. Department of Health and Human Services (HHS), states’ departments of insurance (DOIs) and the public a comprehensive range of information and data about their policies, practices and enrollee experiences.<sup>2</sup> Insurers that sell qualified health plans (QHPs) in the health insurance marketplaces must additionally report the information and data to the marketplaces.<sup>3</sup>

The U.S. Departments of Labor (responsible for the regulation of group health plans) and HHS (responsible for non-group plans and QHPs) have just begun a process to determine what data they will collect and how. But the agencies will require some very limited data reporting for some QHPs beginning in 2016 and sometime thereafter for off-marketplace individual and group plans.<sup>4</sup> While the long-term regulatory approach is uncertain, the Secretary of HHS has called more generally for the use of big data to “transform our healthcare system in unprecedented ways... our commitment... is to liberate data in every way we can.”<sup>5</sup>

A few states have adopted the data reporting provisions as part of their own implementation of the ACA, but most have not attempted to get ahead of federal regulators to operationalize the required data collection. There are likely three primary reasons for the delay. First, since enactment

of the ACA, federal and state officials have had to grapple with numerous pressing issues, such as the establishment and operation of health insurance marketplaces and the oversight of an insurance market undergoing dramatic changes in benefit design, marketing and pricing. Second, embarking on a comprehensive effort to collect, store, analyze and use large amounts of health plan data poses some policy, technical and resource challenges. Third, employer and insurance company interests are strongly opposed to implementation of these provisions, arguing they are too administratively burdensome.<sup>6</sup>

In this issue brief, we discuss how insurance regulators (primarily state DOIs and the federal Center for Consumer Information and Insurance Oversight, or CCHIO) and third parties are currently using data collection and how it could change under the ACA to improve health plan oversight and compliance. In

particular, we discuss how the new ACA requirements could prompt a sea change in regulatory oversight – and counterintuitively – reduce the regulatory burden on health plans. Not included in this brief, however, is a discussion of how data can be used by the general public, i.e., for purposes of comparing and shopping for health plans. We hope to revisit this topic in a future brief.

To prepare this paper, we analyzed state and federal requirements related to the collection and transparency of insurance company data, as well as guidance and reports from independent organizations that collect or receive health plan data, such as national accrediting bodies and state APCDs. We also conducted 15 interviews with consumer advocates, state and federal regulators, insurance company executives, and executives with organizations involved in the collection and analysis of health plan data.

## Current and Potential Uses of Health Plan Data

The ACA contemplates the reporting and collection of a broad range of health information, from enrollment to claims and underwriting practices to financial information. (See Figure 1).

### Figure 1. Data Categories Insurers and Plans Must Report under the ACA

- Claims payment policies and practices;
- Periodic financial disclosures;
- Data on enrollment;
- Data on disenrollment;
- Data on the number of claims that are denied;
- Data on rating practices;
- Information on cost-sharing and payments with respect to any out-of-network coverage;
- Information on enrollee rights; and
- Other information as determined appropriate by HHS.<sup>7</sup>

Some of this information, in summary form, is already reported to state or federal regulatory entities, and in some cases to the public. In addition, the health insurance marketplaces, both state- and federally run, have access to data about enrollments, disenrollments and premium payments for QHPs. California's marketplace has embarked on a first-of-its kind initiative to mine health

claims data beginning in the fall of 2015 in order to better assess insurers' benefit designs and whether consumers are receiving appropriate and timely health care services.<sup>8</sup>

With the exception of a small number of DOIs that have begun to use APCDs to support regulatory oversight, state DOIs and federal oversight officials are generally not collecting or using consumer-level data from health insurers, such as sales, enrollment, and claims data, as an oversight tool. As a result, much of what health plans do remains, as one consumer advocate put it, "a black box." What the ACA envisions via data collection is "fundamentally different [from what regulators do currently]." Depending on how ACA data reporting requirements are implemented, the new data disclosures hold the potential for regulators and the public to see in an in-depth way how insurance is really working for people.

In most states, the DOI is responsible for the oversight of insurers, ensuring their ability to pay claims and enforcing compliance with the ACA's market reforms. DOIs have broad authority to require insurers to submit data to help them perform their oversight and enforcement duties. However, a federal statute, the Employee Retirement Income Security Act (ERISA), regulates employer-based health plans. As a result, state DOIs do not collect data about self-funded employer plans (i.e., those where the employer shoulders the responsibility for paying claims). The interplay of state insurance regulation and ERISA is

complex and has been the subject of frequent litigation, including a case before the Supreme Court regarding Vermont's ability to compel employer plans to submit claims data to its APCD.<sup>9</sup>

Further, most health insurer data reported to DOIs are summary-level data reports about financials, plan design, rates, marketing and claims processing practices. Insurers of some products and services, such as life insurance, long-term care, and homeowner policies must file a market conduct annual statement (called the MCAS) with state insurance regulators. These annual statements are used to support more in-depth assessments of insurers' compliance with state laws. However, a similar statement is not currently required of those selling health insurance. The National Association of Insurance Commissioners (NAIC) is developing a MCAS for health plans, but progress has been slow. In the meantime, the NAIC has developed a survey and standardized data request form

that DOIs can use to better assess insurance company compliance with the ACA's market reforms. However, this effort relies on summary reports that insurers submit and is not designed to provide access to consumer-level sales, enrollment, claims or other health plan data.<sup>10</sup>

The federal agencies responsible for health plan oversight – the U.S. Department of Labor (DOL) and CCIIO – are also primarily receiving summary-level data about plans. However, big data is not new to federal health regulators. The Centers for Medicare and Medicaid Services (CMS) – CCIIO's parent agency – uses the sophisticated analysis of millions of health claims to detect and combat fraud in the Medicare program.<sup>11</sup> CMS has also recently released a massive data bank of provider charges to the Medicare program, allowing researchers and others to mine the data to better understand the pricing and use of health care services.<sup>12</sup>

**Table 1. Examples of “Big Data” vs. Summary-Level Data**

Big Data	Summary-level
Sales transactions	Total number of policies sold
Enrollment and disenrollment (i.e. 834 and 820 transaction forms)*	Total number of disenrollments
Medical, pharmacy, dental claims	Total number of claims paid Total number of claims paid to out-of-pocket network providers

\*An 834 form is a HIPAA-standardized transaction used by employers, government agencies and insurers to enroll and disenroll members in a health benefit plan. It includes information about the subscriber, the plan, and, if the member is disenrolling from the plan, the reason for disenrollment. Another standard transaction is the 820 transaction, which is used to provide premium payment information to insurers.

### *Third party data reporting*

Health insurers don't just report data to state and federal regulators. Those that sell QHPs on federal and state marketplaces share data about enrollment and disenrollment and premium payments and, as noted above, those selling QHPs in California will soon begin sharing claims data with the marketplace.

Insurers in many states are also reporting data to third parties such as APCDs, either voluntarily or as required by state law. APCDs provide a kind of “big data” – state-based databanks of paid medical, pharmaceutical and sometimes dental claims, submitted by both private and public payers. APCDs are currently operating or being implemented in 18 states.<sup>13</sup> APCDs have considerable experience working with health plans to improve the accuracy and usability of data reporting. While the information that APCDs collect does not perfectly overlap with the ACA's contemplated

data collection (for example, APCDs do not collect denied claims), states could leverage their experience and data to help implement the ACA's requirements.<sup>14</sup>

Some state DOIs are already beginning to use APCDs as an independent data source to buttress their regulatory oversight role. For example, DOIs in states such as Arkansas and Rhode Island intend to analyze APCD data to corroborate insurers' claims about price and utilization trends included in their proposed rate filings.

Many insurers also report important data elements to health plan accrediting organizations, such as the National Committee for Quality Assurance (NCQA) and URAC. The ACA's marketplaces require insurers to be accredited, as do many large employers. State and federal officials also often use accreditation as a proxy for an insurer's compliance with Medicare, Medicaid and commercial

plan standards. The accrediting bodies require plans to report clinical quality data, but also collect reports on plan policies and procedures. However, they do not generally collect claims, sales or enrollment data.

A few states also have government agencies or independent entities that collect health plan quality or complaint data and publish consumer-facing analyses or report cards. For example, Massachusetts' Health Policy Commission provides on its website annual reports noting the numbers of grievances and appeals filed against insurers in the state.<sup>15</sup> A new state law requires California's Office of the Patient Advocate (OPA) to collect, analyze and report on consumer complaint data drawn from state consumer assistance centers.<sup>16</sup> The state also publishes health plan report cards based on clinical quality and patient experience data.

### **The ACA's Data Categories**

The ACA doesn't prescribe what specific data should be collected within the outlined data categories, nor does it articulate the method of data collection. However, the comprehensive nature of the data categories listed in the law – financial, claims, enrollment, rating, benefit design and enrollee rights – gives state and federal regulators a powerful new ability to answer important questions about health insurers' behavior in the market and how consumers are accessing and paying for health care. To best answer these questions, regulators will need a data collection framework that captures a maximum amount of information in the most efficient and cost-effective way possible. This suggests a need to rely not just on traditional summary reports that must be individually read and analyzed over hundreds of staff hours but also to take advantage of the revolution in big data – those exceptionally large data sets that can be mined with a computer and sophisticated algorithms. Regulators need both in order to fully implement the ACA's vision for data collection. The following section discusses key data collection categories listed in the ACA and how a reporting scheme could be implemented to improve the efficiency and effectiveness of oversight.

#### ***Financial information***

One data category required by the law – periodic financial disclosures – appears to be fairly consistently collected from insurers and analyzed across state DOIs. A common refrain among insurance regulators is that the “number one consumer protection is insurer solvency,” and state DOIs take their solvency oversight duties seriously.

Insurers selling health policies to individuals and employer groups are required to report quarterly financial information to support their ability to cover the current and future claims costs of policyholders.

Within the federal government, CCIIO collects both rate filings and disclosures about insurers' expenses, premium revenue and claims in order to implement key provisions of the ACA. For example, CCIIO uses revenue and expense data to assess each plan's medical loss ratio (MLR), or the percentage of total premium revenue spent on paying for health care services. Under the ACA, insurers who don't meet a minimum MLR threshold must pay a rebate to policyholders. Insurers must also submit claims data to participate in the ACA's risk mitigation programs, which help compensate insurers who enroll people with high health care costs.<sup>17</sup> In both cases, CCIIO receives these data in summary reports. The agency has not engaged in any efforts to date to access large sets of claims, sales or enrollment data from insurers.

In addition to requiring financial reporting from insurers, the ACA also requires it from employer health plans, including those that are self-funded. Currently, state DOIs do not collect any data from employers that self-fund their employees' health benefits. Large employer-based plans (those with over 100 employees) must file a form with the U.S. Department of Labor (DOL) that contains financial disclosures. However, small employer-based plans (those with fewer than 100 employees) are largely exempt from the DOL requirement.<sup>18</sup> Yet small firms employ approximately 40 million people.<sup>19</sup> To fully implement the ACA requirement, the DOL will need to require some sort of financial filing from employers that are currently exempt.

Both DOL and HHS will also need to assess whether insurers need to submit new or different financial information in order to fulfill the ACA's data transparency objective. In general, both consumer advocate and insurance company stakeholders commented that, for most insurers, the current regime of financial reporting works reasonably well to protect consumers. It is less clear whether it is working for self-funded employer plans.

#### ***Enrollment information***

The ACA requires insurers to report data on enrollment and disenrollment. While enrollees cycle in and out of coverage on a regular basis, particularly in the non-group insurance market, regulators can use enrollment and disenrollment data to help identify outliers or potential

trends. A consumer advocate noted, “disenrollment is a good proxy for satisfaction with a plan...it could be an early warning signal that something is going on.”

To fully implement this requirement, regulators would benefit from both big data and summary-level data, which could help them gain a full picture of who is enrolling in or disenrolling from health coverage and why. To an extent, some insurers (i.e., those selling policies to individuals and small employers) already report limited summary-level enrollment data to state and federal regulators through rate and other filings, usually provided as the total number of members or policyholders. State DOIs can at any time ask for disenrollment data from insurers. Such a request might garner, for example, the number of policy terminations or cancellations initiated by the consumer and the number that occur because the consumer didn't pay the premium.<sup>20</sup>

However, this category is one in which summary-level reports from insurers have limited analytical potential by themselves. To effectively implement this provision and gain useful information about insurance company practices and consumer behavior, regulators should be allowed to access raw enrollment and disenrollment data – such as the 834 forms – at the transaction level. States that operate their own marketplaces already have this data for QHPs.

Regulators and the marketplaces could mine this enrollment data along with other data sources such as health claims, by characteristics such as zip code, subsidy eligibility, type of plan and diagnosis code, all of which they could use to flag whether a plan's marketing, utilization management, or other policies are worth a closer look. For example, if an unexpectedly high proportion of people with mental health diagnoses are disenrolling from a plan, regulators may decide the plan's mental health benefits, provider network or management of mental health claims require additional review. Or, if a plan is only enrolling individuals from zip codes in its service area known to have young, healthy residents, regulators may wish to investigate the company's marketing practices. “The nice thing about [big data mining],” noted one consumer advocate, “is that you don't have to know the answer ahead of time. You'll see trends you maybe hadn't even thought about.”

### *Rating practices*

Neither the statute nor federal rules define “rating practices,” another required reporting category. However,

former congressional staff interviewed for this paper suggest that Congress was interested in capturing information about the factors insurers use to set premium rates, such as age, industry, claims experience, and gender. DOIs can generally obtain information about a plan's rating practices from annual rate filings. The ACA bans many of the most egregious rating practices, such as health status and gender rating, but only for plans sold to individuals and small employers (currently defined 2-50 employees). Insurers selling policies to large employers, however, may still use these rating practices to set premiums. Yet many states do not require rate filings for plans sold to these larger employer groups, and states do not have authority to regulate the plans that employers self-fund.

Employer-based health wellness programs are permitted under federal rules to impose premium surcharges on employees for tobacco use, failure to participate in a wellness program, or failure to meet specified health goals, such as a target body mass index (BMI) or blood sugar level. The ACA's data collection provision could give state and federal regulators the authority to collect the information needed to better assess who is being charged these higher premium rates and on what basis.

### *Claims practices and denied claims*

The ACA requires insurers to report information about their claims policies and practices and the number of denied claims. Such disclosures could help federal and state officials discern whether discriminatory practices exist, particularly if stratified by diagnosis, zip code or type of service. Here again, state and federal regulators would benefit from capturing claims data – submitted, paid, and denied – at the transaction level. For example, access to claims data through an APCD in one state allowed officials to conduct a “targeted review” of how consumers were accessing substance use treatment. The APCD data enabled them to gain a comprehensive picture of how health plans were covering these services. In a similar vein, regulators may want to query denied claims to check whether insurers are denying coverage of certain types of services such as behavioral health or oncology more than other services.

The collection and use of information about insurers' claims and claims practices currently vary widely state-to-state. Most states do not collect data on how many claims are denied and for what reason. While all states record, categorize and store consumer complaints made to the DOI, and have the authority to ask insurers for data on the number of grievances and appeals filed by

policyholders, not all will ask for this information on a regular basis; some may only do so in preparation for a targeted audit. Yet regular access to data – even if provided in summary-level reports – regarding internal and external appeals and their disposition would give regulators the ability to compare how insurers handle enrollee grievances.

#### *Cost-sharing and provider network information*

The ACA requires insurers to report data on “cost-sharing.” This information could help policymakers and regulators better understand consumers’ experiences with deductibles, co-payments, coinsurance and out-of-pocket maximums. For example, if regulators were collecting claims data from health plans, they could query how many enrollees are hitting their out-of-pocket maximum in a given year. They could further refine such a query by diagnosis code or service category. Such data could also help answer questions such as: Are deductibles affecting the use of primary care services? Is a plan’s cost-sharing structure, such as the use of tiered formularies or provider networks, consistent with ACA rules prohibiting discrimination in benefit design? How is the use of cost-sharing affecting the use of brand-name vs. generic drugs? Are providers and plans appropriately handling claims for preventive services to ensure that enrollees don’t face cost-sharing, as required under the ACA?

The ACA also requires plans to report data on “payments with respect to any out-of-network coverage.” Data on enrollees’ use of out-of-network services could help regulators assess whether a plan’s network has a sufficient number and range of providers to deliver on promised benefits. For example, regulators could analyze claims data to determine how many enrollees receive services from an out-of-network provider, and whether there are meaningful differences by type of plan (i.e., open vs. closed network), zip code, type of service, or other enrollee or provider characteristic. Analysis could also provide information about the extent of balance billing associated with out-of-network claims. Such analyses could not only help regulators assess plans’ compliance with state and federal law, but also other broader policy

challenges such as provider workforce supply, provider market power and appropriate use of service settings.

Further, some states require plans to ensure that enrollees can access services within a maximum distance from their home or workplace or within a specified time frame in order to demonstrate an adequate network.<sup>21</sup> But it is difficult to know whether these “time/distance” standards ensure that a plan network is fully meeting enrollees’ needs. Understanding patterns of out-of-network use could be “the guts of a new and different network adequacy standard,” observed one consumer advocate. Instead of guessing at an appropriate network size, regulators could review data to give them an accurate picture of enrollees’ actual experiences using in- and out-of-network care.

Some states have unique data reporting requirements that are similar to the ACA that could be leveraged to answer some of these questions. For example, Massachusetts requires sellers of limited and tiered network plans to annually report summary-level data on use of services by provider tier and the use of out-of-network services.<sup>22</sup> However, state officials noted that reporting differences among insurers and problems with data integrity have led to delays in publishing reports. Five years after the requirement became effective, officials reported that they are still “refining” the data collection tool to make it more consistent across insurance companies. They noted that it has taken “lots of communication” between the state and insurers to get to a point “where the data is reliable.”

#### *“Other information as determined appropriate by the Secretary”*

The ACA includes a catch-all data category, providing the HHS with broad authority to determine what other information would be useful to collect from insurers. This could include, for example, information on marketing practices and broker commission structures, which can help assess compliance with the ACA’s prohibitions against discrimination based on health status. It could also include requests for information about market trends or problems that emerge over time but are not apparent today.<sup>23</sup>



## Regulation the Big Data Way: Implementation Issues to Consider

While health insurers were early adopters of big data in order to understand their current and prospective policyholders, for health insurance regulators it is new territory (state regulators do use big data for oversight of other lines of insurance, however). Instead of an oversight system that relies on insurers compiling summary data reports on finances, benefit design, rates and complaints, regulating via big data instead means using algorithms and sophisticated analytics to mine massive amounts of claims, sales and enrollment data to capture insurers' behavior in the market and policyholders' experiences with their plans. This approach would be largely new for state regulators, but advantages include improved data integrity, improved oversight, and greater efficiency.

### *Improved data integrity*

Currently regulators rely largely on summary reports from individual insurers. Insurers have different IT systems, methods of compiling the reports, and interpretations of key terms and data categories. Relying on insurers to compile these summary reports increases the risk of differences in interpreting data definitions among insurers as well as the submission of incorrect data. Such data integrity problems require regulators to spend considerable time and effort communicating with insurers to shore up the accuracy of the data. That time and effort would not be eliminated for regulators able to access standardized sales, claims and enrollment data, because all data collection efforts require a quality assurance program. But big data could mitigate some of the significant data integrity problems currently hindering effective regulation.

### *Improved oversight*

As noted above, big data allows regulators to conduct refined queries of large data sets and run analytics that allow for a more granular understanding of marketing trends and how policyholders with specific characteristics (i.e., diagnosis or geographic location) are faring under their plan. The data allows regulators to see details and trends that could be lost in summary-level reports. In addition, because regulators are not solely relying on summary-level reports that can vary from insurer to insurer, a big data approach can enable more apples-to-apples comparisons among insurance companies. "Insurers won't like [regulation through] big data because it means they're more accountable," predicted one consumer advocate.

### *Greater efficiency and reduced regulatory burden*

Instead of requiring insurers to spend hundreds of employee hours compiling summary-level reports, this approach can be automated. Regulators can learn from and build on the experience of APCDs. Although APCD officials noted that a big data reporting system is resource-intensive to establish, once it is in place and the scope and frequency of reporting decided, insurers can provide data to regulators (or, more likely, to a data consolidator acting on the regulator's behalf) via a pre-programmed feed. "Someone is really just pushing a button [to submit the data]," an APCD official told us.

However, the use of big data for regulatory oversight is not without challenges. These include the need to address privacy and security concerns and resource constraints, and to monitor and correct definitional problems.

### *Privacy and security concerns*

Any collection or transfer of data raises concerns about privacy and security. Regulators and the insurance marketplaces must balance these risks with the benefits of using big data as an oversight tool. No system can be completely invulnerable to those determined to break in, but regulators and insurers can take steps to protect sensitive information. Big data is, by its very nature, information about individual consumer transactions – it is only through this individual transaction data that data mining and predictive analytics are possible. Where data includes personally identifiable information, including highly sensitive health information, it should not be available to the public, and access must be highly controlled. There are federal (through the Health Insurance Portability and Accountability Act, or HIPAA) and state safeguards to protect personal information, but because they may not always apply to all data collection efforts, officials will likely want to ensure these and perhaps even stronger standards are enforced.

APCDs have implemented best practice safeguards to mitigate the possibility of breaches. These include encryption during data transmissions and data storage and the use of software programs to de-identify personal information either before or upon transmission. Regulators can also ensure that any publicly available reports derived from big data use non-specific information to reduce the risk of information being traced back to a particular

individual. Those who store and transfer data also must be prepared to adapt their protective measures as technological capabilities evolve.

### *Addressing resource constraints*

States and federal officials are unlikely to replace their current reliance on summary-level reports from insurers as an oversight tool. Such reports provide useful information and have long been the mainstay of insurance regulation. Rather, the use of large data sets can supplement that summary-level data, allowing regulators to stay on top of market trends in closer-to-real-time. Capturing, storing and analyzing millions – if not billions – of raw data requires resources, including sophisticated IT systems and experienced personnel. For DOIs already strapped for funding, these costs, especially if on top of the costs associated with reviews of summary-level data – could be perceived as a significant barrier. Regulators can gain significant economies of scale, however, by relying on a regional or national data consolidator, an entity that can store the data feeds from insurers and maintain a staff with the expertise to run the algorithms and analytics requested by state and federal regulators. States may also choose to make the data available to authorized third parties, such as researchers, who can use the data to identify market trends or emerging consumer protection concerns. CMS has successfully done this with its Medicare provider utilization and payment data set.

In addition, officials can conserve resources by avoiding duplication of data collection efforts already underway. For example, most individuals interviewed pointed to the regular financial disclosures that insurers are required to make to federal and state officials, and few could identify any additional value in requiring additional disclosures. As one insurer put it, “if federal regulators come out with something different [than what is already required], it won’t be fun.”

The new data collection requirements in the ACA include a number of data elements that are not being collected

anywhere, or are being collected and used only in limited circumstances or for specific purposes, such as for an audit or market conduct exam. For example, we could find no entity regularly collecting data on numbers or percentages of denied claims, and while some states require insurers to report data on the use of out-of-network services, most are not. These are data elements for which a big data approach would not only be less burdensome on plans and regulators than summary reports, but would also generate far more useful information about policyholders’ experiences. In addition, once built, the experience of APCDs suggests that the uses of the data will expand. Said one APCD official, “. . .there are uses for the [APCD] data now that we had never anticipated.”

### *Definitional challenges*

Those experienced with health plan data collection – whether via big data or through summary-level reports – universally remarked upon how challenging it can be to settle upon common definitions of data elements that all insurers can use. An APCD official told us that they had to do “lots of back and forth [with insurers] to define terms.” Similarly, a state official engaged in a data collection effort noted that a failure to clearly define terms early on in the project rendered the first couple of annual reports meaningless, because different insurers interpreted the information requested in different ways. “Definitions have to be tight and well understood,” he said.

Consumer advocates also support a “gradual” approach, with regular communications with insurers on data categories and definitions in order to set a solid foundation. Multiple parties noted that running a data collection and analysis enterprise is not a “once-and-done” proposition – it requires continual monitoring and assessment of data integrity. You have to “watch the shop very well,” one official observed. State officials further expressed concerns that federal implementation would incorporate different definitions than the ones used in the state, potentially complicating their own data collection and analysis efforts.

## Conclusion

The data reporting requirements included in the ACA are currently slated to begin for some marketplace plans in 2016 and sometime thereafter for other group and individual market plans. The broad sweep of the data categories laid out in the statute provides an opportunity for government officials to re-think how they use data for oversight purposes. As more and more industries seize on the advantages of big data to understand consumers, so too should those with the responsibility of protecting consumers consider a move in this direction.

Health insurance regulators don't often use big data, and doing so requires a real shift in the ways data is

collected and used. But the advantages are considerable, including a reduced regulatory burden on insurers and a richer understanding of insurer behavior and consumer experience. In particular, it can allow regulators to monitor and address market trends in real time and at a granular level that is unachievable via a summary-level report. The ACA requires the development of a new data collection infrastructure. As one expert on insurance regulation framed it, when implementing the ACA provisions "let's not institutionalize a 19th century view... when it should be a 21st century view."

## Acknowledgements

The authors thank the state and federal insurance regulators, insurance company executives, consumer advocates and health data experts that generously shared their time and expertise. We also gratefully acknowledge the editorial contributions of Kevin Lucia, Karen Pollitz, Timothy Jost, Christopher Koller, Peter Lee, Andrea Routh, Birny Birnbaum, Deven McGraw, Amy Costello and Josephine Porter.

## Endnotes

- 1 Toussaint JS, Mannon M. Hospitals are Finally Starting to Put Real-time Data to Use. *Harvard Business Review* [Boston, MA]. 2014 Nov. Available from: <https://hbr.org/2014/11/hospitals-are-finally-starting-to-put-real-time-data-to-use>.
- 2 42 U.S.C. §§ 300gg-15a, 18031(e)(3) (2010).
- 3 42 U.S.C. § 18031(e)(3).
- 4 80 Fed. Reg. 10750, 10829 (Feb. 27, 2015).
- 5 Burwell SM. On the Threshold of Tomorrow: How We Build a Better Health System and Empower People [speech to Health Datapalooza] Washington (DC); 2015 Jun 3 [cited 2015 Jun 15]. Available from: <http://www.hhs.gov/about/leadership/secretary/speeches/2015/threshold-tomorrow-how-we-build-better-health-system-and-empower-people.html>.
- 6 Karas RS. Phased Transparency Reporting Plan Delays Off-Exchange Collection. *Inside Health Policy* [subscription only]. 2015 Aug 13.
- 7 42 U.S.C. §§ 300gg-15a, 18031(e)(3) (2010).
- 8 Covered California [Internet]: Qualified Health Plan New Entrant Application for the 2015 plan year; 2014 Feb 20 [cited 2015 Jul 9]. Available from: <http://hbex.coveredca.com/solicitations/2015-QHP/documents/QHP%20New%20Entrant%20Application.pdf>. See also Terhune C. California's Obamacare exchange to collect insurance data on patients. *Los Angeles Times* [Internet]. 2015 Jun 21. Available from: <http://www.latimes.com/business/la-fi-obamacare-patient-privacy-20150622-story.html>.
- 9 *Gobeille v. Liberty Mutual Insurance Company*, 746 F.3d 497 (2nd Cir. 2014), cert. granted, 83 U.S.L.W. 3119 (U.S. June 29, 2015) (No. 14-181)
- 10 National Association of Insurance Commissioners [Internet]. Kansas City (MO): Market Analysis Procedures (D) Working Group Conference Call Minutes; 2015 May 6. On file with authors.
- 11 Agrawal S, M.D. The Use of Data to Stop Medicare Fraud [written testimony before the U.S. House of Representatives Committee on Ways & Means, Subcommittee on Oversight] Washington (DC); 2015 Mar 24 [cited 2015 Jun 15]. Available from: <http://www.hhs.gov/asl/testify/2015/03/20150324.html>.
- 12 Centers for Medicare & Medicaid Services [Internet]. Baltimore (MD): Medicare Provider Utilization and Payment Data; 2015 [cited 2015 Jul 2]. Available from: <http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/index.html>.
- 13 All-Payer Claims Database Council [Internet]. Durham (NH): Interactive State Report Map; 2015 [cited 2015 Jun 3]. Available from: <http://www.apcdouncil.org/state/map>.
- 14 Center for Consumer Information and Insurance Oversight [Internet]. Bethesda (MD): State-by-state summary of rate review grants; 2015 [cited 2015 Jul 2]. Available from: <https://www.cms.gov/CCIIO/Resources/Rate-Review-Grants/index.html>.
- 15 Massachusetts Health Policy Commission [Internet]. Boston (MA): Office of Patient Protection Annual Report 2013; 2014 Nov [cited 2015 Jun 7]. Available from: <http://www.mass.gov/anf/budget-taxes-and-procurement/oversight-agencies/health-policy-commission/patient-protection/2013-opp-annual-report.pdf>.
- 16 State of California Office of the Patient Advocate [Internet]. Sacramento (CA): Complaint Data Reports; [cited 2015 Jun 15]. Available from: <http://www.opa.ca.gov/Pages/ComplaintDataReports.aspx>.
- 17 The risk mitigation programs under the ACA are: risk corridors, reinsurance, and risk adjustment, otherwise known as the "3Rs." The risk corridor program is a temporary, three-year initiative in which the government collects revenues from insurers with high profits and redistributes it to insurers with significant losses. Reinsurance is a temporary three-year program that protects insurers with extremely high cost enrollees. Using fees collected from all insurers, it pays insurers a percentage of the costs for their sickest enrollees. Risk adjustment is a permanent ACA program that adjusts for differences in the health of insurers' enrollees by redistributing funds from companies with healthier-than-average customers to companies with sicker-than-average customers.
- 18 29 C.F.R. § 2520.104-20 (2013). Employers with 100 or fewer employees that pay for any portion of benefits from their general assets (rather than a segregated trust) are exempted from all filing requirements.
- 19 U. S. Census Bureau [Internet]. Washington (DC): Statistics of U.S. Businesses (2012) [cited 2015 Jun 18]. Available from: <http://www.census.gov/econ/subs/>.
- 20 See e.g., National Association of Insurance Commissioners [Internet]. Kansas City (MO): Data Call Spreadsheet template; 2014 Dec 11 [cited 2015 Jun 26]. Available from: [http://www.naic.org/documents/committees\\_d\\_mapwg\\_related\\_141112\\_data\\_call\\_no\\_redline.pdf](http://www.naic.org/documents/committees_d_mapwg_related_141112_data_call_no_redline.pdf).
- 21 Giovannelli J, Lucia K, Corlette S. Implementing the Affordable Care Act: State Regulation of Marketplace Plan Provider Networks. *The Commonwealth Fund* [New York, NY]. 2015 May. Available from: [http://www.commonwealthfund.org/-/media/files/publications/issue-brief/2015/may/1814\\_giovannelli\\_implementing\\_aca\\_state\\_reg\\_provider\\_networks\\_rb\\_v2.pdf](http://www.commonwealthfund.org/-/media/files/publications/issue-brief/2015/may/1814_giovannelli_implementing_aca_state_reg_provider_networks_rb_v2.pdf).
- 22 211 Code Mass. Reg. § 152.09 (2010).
- 23 Under the current reporting requirements for QHPs using the federal marketplace platform, CCHIO lists the following under its discretion to request "other information:" policies related to grace periods, retroactive denials, enrollee recoupment of overpayments, drug exceptions timeframes and enrollee responsibilities, information on the explanation of benefits, coordination of benefits, and insurer contact information. Center for Consumer Information and Insurance Oversight [Internet]. Bethesda (MD): Transparency Paperwork Reduction Act Supporting Statement; 2015 Aug. Available from: <https://www.cms.gov/Regulations-and-Guidance/Legislation/PaperworkReductionActof1995/PRA-Listing-Items/CMS-10572.html?DLPage=1&DLEntries=10&DLSort=1&DLSortDir=descending>.



Georgetown University Health Policy Institute  
3300 Whitehaven Street, N.W., Suite 5000  
Washington, DC 20007  
Telephone (202) 687-0880  
<http://chir.georgetown.edu/>