IMS INSTITUTE FOR HEALTHCARE INFORMATICS

December 2015

Closing the Healthcare Gap

The Critical Role of Non-Identified Information

Introduction

The dramatic increase in the amount of digitized healthcare information being generated brings new opportunity to close the "healthcare gap" – the difference between today's reality and what is possible from a clinical, patient and economic perspective. Appropriately accessing and utilizing information for research purposes is a critical element of closing this gap. In this context, there is growing importance in understanding the role of non-identified data (data that does not identify the patient, because the data has been de-identified) in advancing research and understanding our connected healthcare system.

The purpose of this report is to showcase examples of the use of non-identified longitudinal patient-level data to help address a range of important public health issues. The examples reinforce the important value derived from the use of patient-level data that has been de-identified. The report also highlights widely accepted patient privacy and security frameworks that advance appropriate use of big data across healthcare stakeholders. Finally, we call for action across four critical fronts to accelerate progress in the safe, secure, and effective use of non-identified information in the United States and around the world.

Finding new approaches to preventing and combating disease remains a top priority for health systems. At the same time, developing and implementing best practices for healthcare management and operation of health systems – and closing the healthcare gap – can bring enormous benefit to patients and societies. The critical role that non-identified data can play in this effort remains an exciting and invaluable part of achieving success.

This report was produced independently by the IMS Institute for Healthcare Informatics as a public service, without industry or government funding. The contributions from David M. Cutler, Otto Eckstein Professor of Applied Economics, Harvard University; Dr. Ann Cavoukian, Executive Director of the Privacy and Big Data Institute, Ryerson University, and former Information and Privacy Commissioner for Ontario, Canada; Dr. Khaled El Emam, CEO of Privacy Analytics; and Sharon Terry, MA, President and CEO of Genetic Alliance are gratefully acknowledged.

Murray Aitken
Executive Director
IMS Institute for Healthcare Informatics

IMS Institute for Healthcare Informatics
100 IMS Drive, Parsippany, NJ 07054, USA
info@theimsinstitute.org www.theimsinstitute.org

Find out more

If you wish to receive future reports from the IMS Institute or join our mailing list, please **click here**

Contents

Executive summary

The opportunity to close the healthcare gap – the difference between today's reality and what is possible from a clinical, patient and economic perspective – remains a promise for all health systems. The dramatic increase in the amount of digitized health data generated each day brings new possibilities to apply that information to advance research and understanding in our connected healthcare system.

Sources of big data in healthcare range from individual "wearable" technologies and mobile applications to hospital electronic medical records, transaction data and insurance claims. The greatest opportunity to address our healthcare gaps comes from accessing and utilizing real-world, non-identified, patient-level, longitudinal data. This "real-world data" goes beyond typical clinical trial data and can provide insights into the efficacy and safety of medical treatments in every-day practice and across large populations.

More broadly, non-identified data (data that does not identify the patient, because the data has been de-identified) is now used to support an evidence-based healthcare system in a range of ways to improve patient outcomes and also reduce waste and avoidable costs. Patient outcomes can be improved directly through the application of this non-identified data in areas such as comparative effectiveness and best practices development and dissemination; patient populations, response and risk management; health technology assessments; and benchmarking performance and quality of physicians and facilities. While some of these areas also contribute directly to cost savings, other applications of non-identified data can relate to hospital systems seeking to remain within the bounds of their budgets, payers identifying ways to reduce the total cost of care per member, and patients looking to reduce their out-of-pocket costs and improve their overall health. The range of potential uses of non-identified data in this context includes measuring and comparing "total cost of care"; developing better, less costly clinical trials; detecting waste and inappropriate use of resources; and adjusting payments based on performance measures.

Non-identified data can also be used to identify health disparities between sub-populations (age, socio-economic status, geographic location) and therefore facilitate provider and community responses to address them. The most critical benefit of sharing non-identified patient-level information is the ability to connect the various interactions, treatments and outcomes over time and undertaken by different participants within a health system. Connecting information across a connected healthcare system is an essential requirement for accelerating closure of our healthcare gap.

The benefits of utilizing non-identified patient data are documented in peer-reviewed research publications used by policy-makers, academics and healthcare industry participants.

Some of the best uses of this non-identified data to benefit patients and the healthcare system are seen when applied to:

- Guiding public health strategy, for example reducing antimicrobial resistance by supporting antibiotic appropriate use initiatives
- Identifying public health issues, such as protecting the elderly from preventable falls through identification of excessive use of tranquilizers
- Ensuring drug safety and preventing serious drug side effects, through the use of a distributed data network
- Targeting interventions to improve their efficacy, for example combating the prescription drug abuse epidemic by understanding patient behaviors
- Identifying health policy changes to improve care and lower costs, for example identifying issues like hospital readmissions that can respond to policy modifications

While the vast array of available healthcare data and increasingly sophisticated data analytic techniques can help improve healthcare and reduce costs, these benefits must be balanced with appropriate respect for individual privacy. Because useful patient data is derived from individuals and their health experience, safeguards must be taken to ensure that an individual's identifiable health information is not distributed or revealed outside of appropriate and permitted situations. The duty of healthcare stakeholders to protect patient privacy is paramount and taken very seriously, as demonstrated through renewed efforts to codify privacy frameworks and provide implementation guidance.

Techniques to render patient medical information appropriately non-identified (and therefore appropriately protect patient privacy) include a combination of removing, generalizing and disguising direct patient identifiers (such as social security numbers, names, email addresses, medical IDs, and genomic information), along with privacy and security safeguards and contractual limitations to ensure there are sufficient controls over information to keep it non-identified and ensure use in a responsible manner (this overall process is called de-identification). Since the value derived from patient data may be related to the presence or absence of specific data elements (such as age or gender), which are indirect patient identifiers, it is critical to balance carefully the value of research data with appropriate protection of patient privacy.

The removal of all individual identifiers to reduce the risk of patient re-identification to zero would correspondingly reduce the societal benefits of research. Instead, risk-based approaches that remove identifiers to make the risk of re-identifying any individual very small, while preserving the value of the overall patient data, allow data to be analyzed usefully on a longitudinal basis for individuals and on an aggregated basis for groups or populations. In conjunction with appropriate contractual protections and technical, physical and administrative safeguards, this approach provides a broader range of benefits to the healthcare system while still protecting patient privacy.

Obtaining the maximum value from non-identified data requires all stakeholders to join in the quest to further progress in four key areas:

- Increasing the availability and accessibility of high-quality, non-identified patient-level data sources
- Increasing the use of non-identified patient-level data to conduct research
- Universally applying best-practice privacy and security standards
- Strengthening the impact of evidence through increased collaboration

Taken together, these initiatives, combined with the rising volume of healthcare data, will enable healthcare systems globally to benefit from closing the healthcare gap.

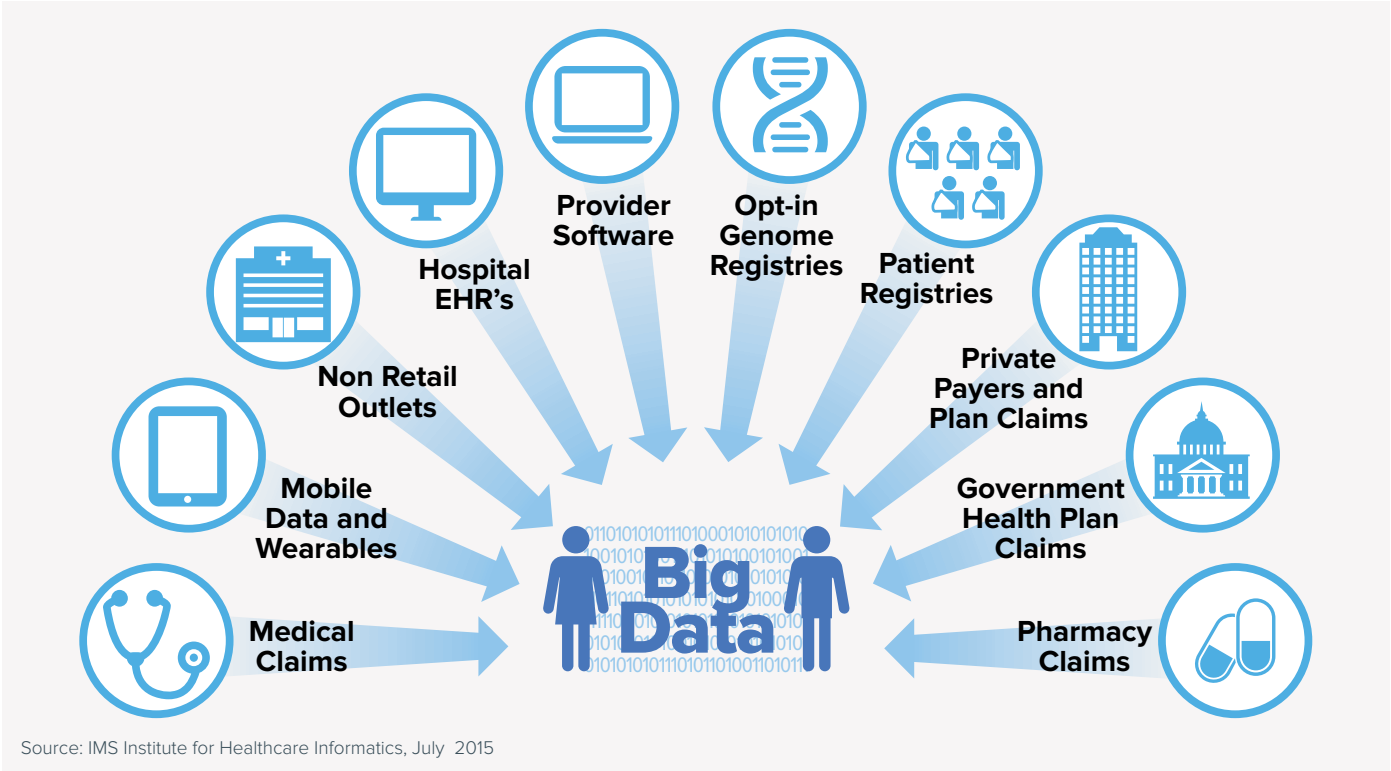
Health systems connected through data

The explosion of big data

The explosion of information in our society brings with it the promise of developing smarter and better systems. "Big data," now available in many fields, and across many areas of enterprise, promotes new understanding of how things currently work, and enables us to take actions to make systems, and the organizations within them, better meet our needs.

In healthcare, this explosion of information is linked to the sheer amount of transactional data now being collected digitally. As physicians move away from the paper-based systems of the past, more data is now being captured through Electronic Health Records (EHRs) than ever before. EHRs are now being used by 76% of non-federal hospitals and over 78% of office based physician practices, recording health information in real time on patient health conditions, vital signs, medicine use and lab test results.[1,2] Additionally, data on the 4.3 billion prescriptions now filled each year in the U.S. market, the 1.2 billion visits Americans make to the doctor's office, and the 629 million patient visits to a hospital (including 460 million outpatient visits) flow from other sources such as billing claims (see Exhibit 1).[3]

Exhibit 1: Sources of Big Data in Healthcare



Source: IMS Institute for Healthcare Informatics, July 2015

This information creates new opportunities to close the healthcare gap by identifying issues, guiding solutions and monitoring improvements to patient care and access, and provides a path towards affordable and optimized medicine with attention focused increasingly on the patient.

In our current healthcare system, not every dollar spent contributes toward improving health. Avoidable costs due to waste, errors, inefficiencies and misallocations of resources still exist. Fixing these issues is critical for the sustainability of healthcare, which now accounts for over 17% of our GDP, and approximately 10% of the median household income in terms of out-of-pocket costs.[4,5] The optimization of healthcare delivery would not only mean bringing down costs for consumers and taxpayers but also saving lives while doing so. The availability of information collected in the real-world at the patient level, and its ability to reveal how patients respond to treatments, brings us closer to accomplishing these goals.

Real-world data

This "real-world data" goes beyond typical clinical trial data that can demonstrate the impact of one or two treatment options on a few thousand individuals. By providing insights into the efficacy and safety of medical treatments in everyday practice and across much larger populations, it can be used to improve the health system's ability to save lives by getting the right care to the right patient at the right time and identify and address major healthcare gaps.[6]

> "The insights from big data have the potential to touch multiple aspects of health care: evidence of safety and effectiveness of different treatments, comparative outcomes achieved with different delivery models, and predictive models for diagnosing, treating, and delivering care. In addition, these data may enhance our understanding of the effects of consumer behavior, which in return may affect the way [healthcare] companies design their benefits packages." [7]
>
> *Nilay D. Shah, Ph.D., Associate Professor, Division of Health Care Policy and Research, Mayo Clinic and Jyotishman Pathak, Ph.D., Director of Clinical Informatics Services and Professor of Biomedical Informatics, Mayo Clinic*

Research performed on real-world data yield benefits to all participants in the healthcare system. The data provides researchers with a robust pool of information that can answer many questions, and be queried repeatedly. By revealing when optimal patient care has occurred and where suboptimal, wasteful or harmful interventions still occur, real-world data helps health systems and payers promote best practices and efficiencies. Armed with this knowledge, researchers and health care industry policymakers can reduce inappropriate care and costs within the medical system, and make policy adjustments to important government safety net programs.

Allowing providers and public health authorities to better visualize and understand treatment variability also helps combat health disparities. "Descriptive studies" showing what is actually occurring in the healthcare system can reveal surprising cases of variability in care, inequality, or prescription bias, and help create initiatives to address them. It also serves as an objective source of information on treatment variability and health disparities, facilitating community responses to address them.

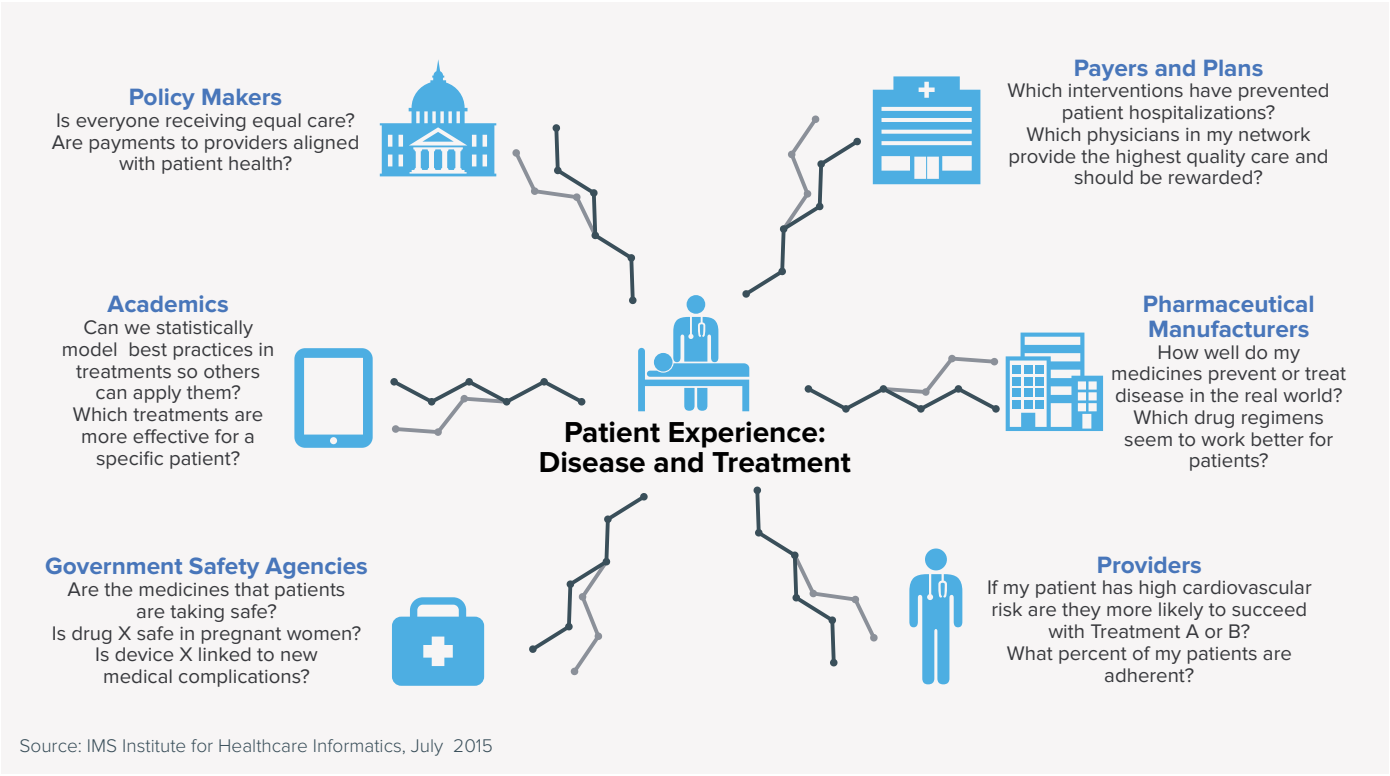
Unlike government statistics that provide a static view of the medical system or trends over time, real-world data is not static. Monitoring new data flowing in real time can help safety agencies minimize adverse events and preventable deaths from medical treatments. It can also be used to track disease patterns to preserve population health, for example to monitor outbreaks of flu or detect and control infectious disease.

For physicians, the movement toward individualized healthcare tailored to a given patient is made increasingly possible through research using real-world data, which provides enough information about subpopulations to understand what is relevant to a specific patient seeking treatment; what is right for one patient may not be the right treatment for another based on history, risk, age, and a host of other factors. While clinical studies are generally not large enough to provide information to tailor treatment to specific patients and deliver on the promise of personalized or precision medicine, larger real-world sources can. They can also provide a means to study narrow populations or rare diseases, because it is only when you can view the breadth of populations across multiple settings that you can obtain a depth of information on smaller subsets of patients.

> "Many factors have converged to make now the right time to begin this ambitious project [the President's Precision Medicine Initiative] . . . . [Individuals] are engaging in improving their health and participating in health research more than ever before, electronic health records have been widely adopted, genomic analysis costs have dropped significantly, data science has become increasingly sophisticated and health technologies have become mobile. We have to seize this moment to invest in these promising scientific opportunities to help [people] live healthier lives." [8]
>
> *Dr. Francis S. Collins, Director, National Institutes of Health*

Patients also gain an improved ability to interact with the health system using tools based on this data. Provided with information that gives them access to quality and cost measures about physicians and facilities, they are empowered to make decisions better about their own healthcare.

Exhibit 2: Elements of a Connected Healthcare System



Source: IMS Institute for Healthcare Informatics, July 2015

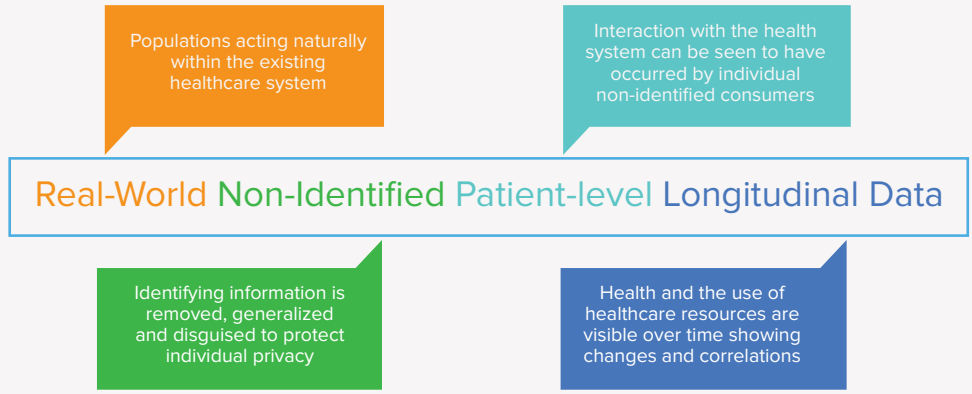
Connected healthcare

For a healthcare system notoriously slow at adopting innovation and best practices, the most critical benefit of shared information is the ability to enable connected healthcare through data. Healthcare is essentially a series of interactions, treatments and outcomes over time and undertaken by different players. These need to be analyzed and understood in a connected fashion to fully understand what works, what doesn't, at what cost, and with what benefits for individuals and populations.

When all actors in the healthcare system share visibility to the same information about the medical system, this helps them to jointly identify critical health problems, informs a common understanding and encourages coordinated action (see Exhibit 2). Whether stakeholders work together within research partnerships or work in parallel on the same problem, the data itself provides a common language and enables them to tackle the problem together to more rapidly close healthcare gaps.

Connected healthcare also fundamentally needs a connected understanding of the patient. To ensure that the entire healthcare system delivers integrated care for the benefit of an individual and larger populations, action based on shared information that tells a complete and accurate story, is needed. Only with a shared understanding of patient experience and what benefits patients across the medical system can healthcare stakeholders deliver connected healthcare.

Exhibit 3: Data For Healthcare Research



Source: IMS Institute for Healthcare Informatics, July 2015

Non-identified patient-level data

At the foundation of this opportunity to improve health are non-identified patient-level data sources. These are real-world datasets built from millions of healthcare transactions over long periods of time that describe the ways in which healthcare services are used (see Exhibit 3). "Non-identified" healthcare data has been stripped of patient identifiers and protects patient privacy through de-identification steps to allow the data to be shared more widely and safely for research purposes (following best practices). This permits the creation of robust real-world datasets that allow comparisons of treatments at multiple levels: from one physician, facility, health system, geography, or country to another.

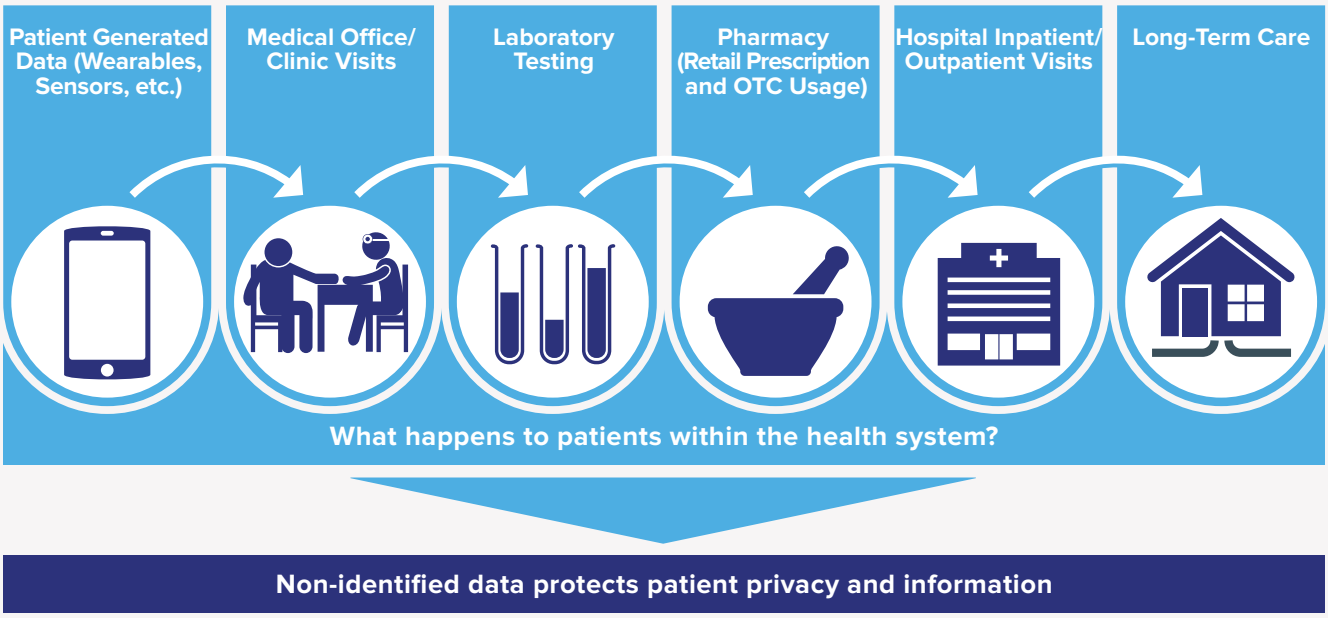
All data starts within the healthcare delivery system as patient identifiable information, used appropriately for treatment, payment and other permitted activities. Physicians, health systems, hospitals and payers obtain information that includes patient names, social security numbers and other direct identifiers for treatment and billing purposes. Facilities that use patient identifiable data for research typically can see only the limited data that they collect, making this data's value somewhat lacking for understanding the healthcare system or medical treatment as a whole. A specific health system, for instance, may have a clear picture only of a patient's use of healthcare for several years within their system, or for one specialty health condition. Or, they may have a biased or partial view due to the demographics of their particular treatment population. Since these diversely collected pockets of data may look very different from one another, each facility essentially holds an incomplete clinical picture of a patient or the population — each holds its own facts.

Moreover, while this identifiable information can be used for research in certain contexts and following certain rules, these efforts often also are cumbersome and require attention to specific regulatory and compliance requirements. In addition, this patient identifiable research typically is limited in volume and in the array of healthcare settings providing data. The promise of using non-identified information that has been de-identified through standards frameworks presents new and increased opportunities to learn and benefit from this data, by encouraging a broader collection, use and analysis of data that has been de-identified and therefore provides a more robust and complete picture of overall health care.

Additionally, the ability to understand patient health progression can only occur if the patient is seen within the data in a connected fashion over many years — or longitudinally. For instance, although some of the most robust clinical information we have in the U.S. medical system comes from Medicare, it is a limited dataset that only reflects the real-world experience of the elderly in the United States. It lacks the ability to provide a view into the experience of these patients before they age into the program. It lacks the ability to provide a view into the two-thirds of the insured population with private health insurance. It lacks the ability to assess correlations with patient disease over a lifetime.[9]

In order to understand healthcare holistically, a broader base of knowledge must be accessed and treatment information from more than one system must be integrated. This requires the creation of health datasets (using a consistent data schema) that can show healthcare delivery across locations and settings — i.e. across multiple pharmacies or hospitals — and longitudinally over time (see Exhibit 4). Such datasets can include identified or non-identified data (depending on the purpose), and can either be distributed datasets that can be queried jointly for research purposes or consolidated datasets.

Exhibit 4: Non-Identified Data Connecting Fragmented Healthcare



Source: IMS Institute for Healthcare Informatics, July 2015

"The volume and landscape of research involving human subjects have changed considerably. Research with human subjects has grown in scale and become more diverse. Examples of developments include . . . the growing use of electronic health data and other digital records to enable very large data sets to be analyzed and combined in novel ways." [10]

Federal Policy for the Protection of Human Subjects, Notice of Proposed Rulemaking

Data sources

Non-identified patient data are derived from identified patient data sources. These include the following.

- **Places where healthcare is received** – e.g. physician offices, clinics, hospitals and their clinical software systems including ambulatory and inpatient electronic medical records (EMR).
- **Ways that healthcare is paid for** – e.g. billing or medical claims made to both private-sector and government health insurers through an adjudication process, including those at pharmacies, physician offices, or collected by payers themselves.
- **Newer sources** – e.g. patient wearable sensors, unstructured data, and self-reported mobile health data.

Data sharing projects

To reduce the fragmented view of the healthcare system, efforts to aggregate or create networks of reliable data for research are underway, and most of these sources will be non-identified when shared. Growth in the use of EHR systems, now required for provider participation in the Medicare and Medicaid EHR Incentive Programs, and efforts at EHR standardization, are enabling the creation of multi-system interoperable data networks.

Examples of these include the Health Care Cost Institute (HCCI), a payer initiative with Aetna, Humana, Kaiser Permanente and United Healthcare contributing data, and efforts being made by the Patient-Centered Outcomes Research Institute (PCORI) to create a representative national network of distributed datasets for clinical outcomes research called PCORnet.[11,12] PCORnet pools data from healthcare systems, Clinical Data Research Networks (CDRNs), and disease-focused patient groups - called Patient-Powered Research Networks (PPRNs), to ensure high-quality, comprehensive data exists for specific disease populations.[13] Other regional and local initiatives are also growing through state-legislated formation of all-payer claims databases (APCDs) and health information exchanges (HIEs).[14] These can contribute significantly to medical knowledge, especially those projects aiming to gain deep knowledge of specific disease or sensitive patient populations such as PEDSnet, a collaborative effort between health systems to establish a national pediatric learning health system.

Several of these efforts are intended to be shared with only select participants in the healthcare system for research – i.e. just physicians or just government agencies. However, no single healthcare stakeholder can maximize the use of healthcare information. To speed our health system's ability to understand core issues and improve patient care, a broader ecosystem of all healthcare stakeholders, including academics, policymakers, and commercial and private sector participants must be able to share access to non-identified pools of patient-level data for research, each approaching a health problem with their unique perspective.

Progress through data – the benefit of non-identified health data research

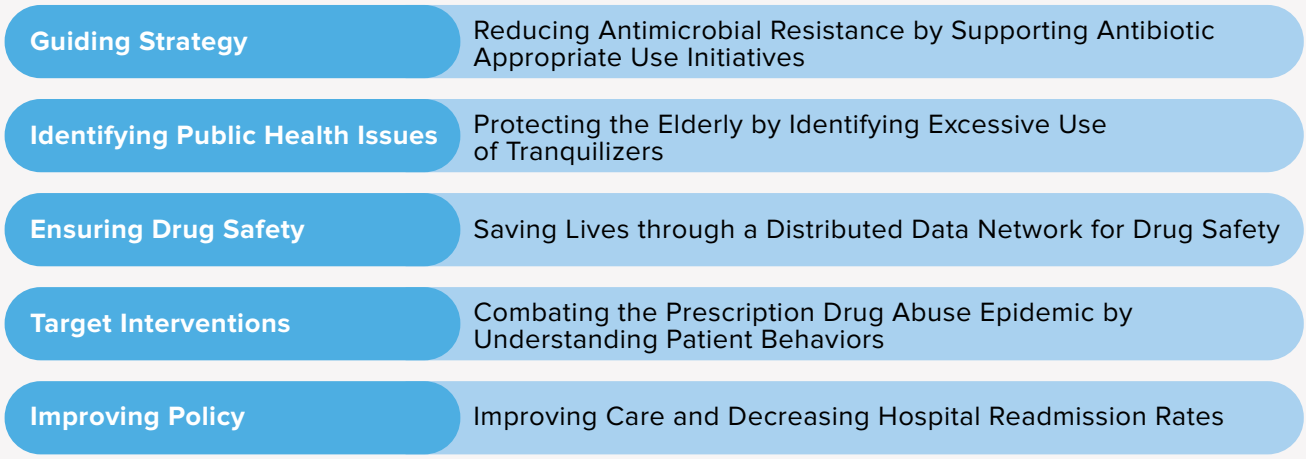
Research has shown the versatility of non-identified patient data to provide a vast range of tangible benefits to society. Without the data we would lack a complete understanding of our medical system and how it is performing. Just 10-15 years ago, when this data was rarely available, identifying inappropriate care was a difficult task. Now, however, healthcare stakeholders can routinely employ this data to examine healthcare delivery more closely.

> "No business ever got better without knowing what it was doing. And that is basically where healthcare has been – we didn't know what we were doing so of course we couldn't get better... That has shifted in the last decade as we have come to have more access to these sorts of data."
>
> *David M. Cutler, Otto Eckstein Professor of Applied Economics, Harvard University*

Top 5 uses of non-identified patient-level data

A review of peer-reviewed publications using this data finds that multiple actors — policy-makers, academics and industry participants — currently rely on non-identified patient-level data to conduct a broad range of studies. The following five examples demonstrate some of the best uses of this data to benefit patients and the healthcare system, and show the progress that is already being made through data (see Exhibit 5).

Exhibit 5: Top Uses of Non-Identified Patient-Level Information



Source: IMS Institute for Healthcare Informatics, July 2015

Reducing antimicrobial resistance by supporting antibiotic appropriate use initiatives

The inappropriate use of antibiotics is a known public health issue. The emergence of deadly antibiotic-resistant bacteria is a direct result of indiscriminate use of these valuable medicines, and has led to a global public health crisis. "The untreatable infection" is a reality, driving an urgent need to address the issue. Antibiotic resistant bacteria cause over two million illnesses and 23,000 deaths per year in the United States and over 700,000 worldwide.[15,16]

The U.S. Centers for Disease Control and Prevention (CDC) estimates that up to 50 percent of antibiotics prescribed are either not medically necessary, or are misused – for instance given at the wrong dose, or for the wrong pathogen, or in the wrong sequence.[17] Uncomplicated upper respiratory tract infections that are usually viral, for instance, are a common inappropriate use of antibiotics. Despite clinical awareness that treatment of such viral, inflammatory, or other specific diseases with antibiotics is inappropriate, prescribers continue to dispense antibiotics.

Descriptive or "observational" studies using non-identified data reveal what is actually occurring in the healthcare system, and can therefore clarify whether medical resources are being used appropriately, and aid public health officials to combat inappropriate prescribing. In the case of antibiotics, public health agencies use non-identified patient data to track antibiotic use and overuse, to deploy educational outreach programs to change behavior – such as reducing inappropriate physician prescribing or patient misuse – and to assess the success of these interventions. With data-guided public health strategies, the efficacy of these essential medicines can be maintained.

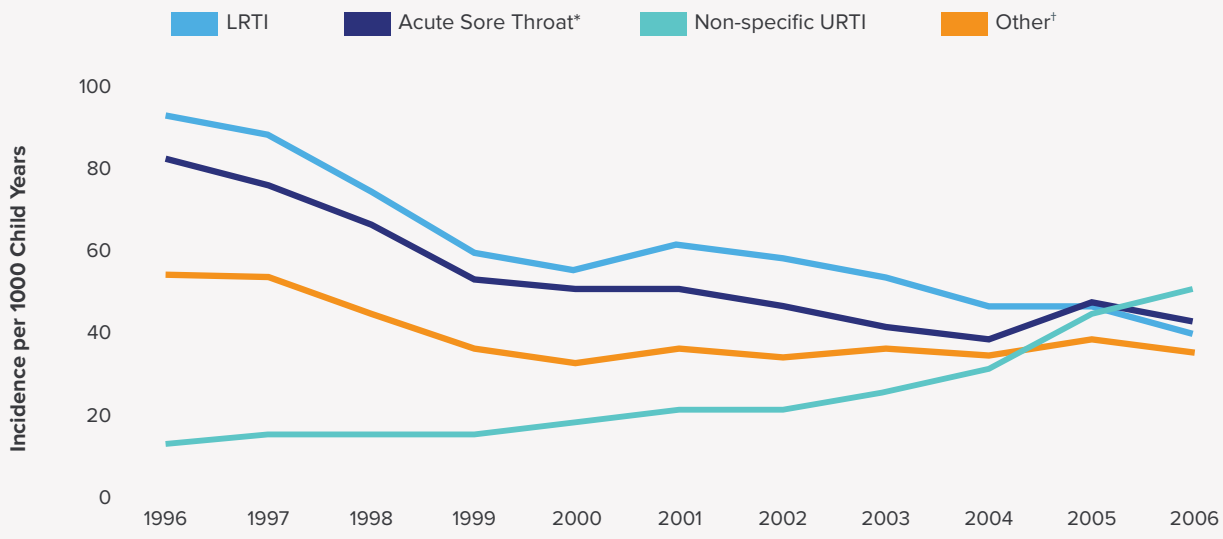
The CDC uses real-world data to understand antibiotic prescribing rates, examine patterns of geographic variability across the country and seasonal variation in use, and understand links between prescribing rates and socioeconomic and population health factors. For instance, research headed by Dr. Lauri Hicks, head of the Get Smart program at the CDC, determined there was an almost three times higher rate of antibiotic prescribing in the U.S. South across all age groups (931 prescriptions per 1000 persons)

> "Data is critical. And we have some valuable data already. We know that 5 out of 6 Americans are prescribed antibiotics each year. That adds up to 262 million antibiotic prescriptions annually. We also know that some doctors prescribe antibiotics far more frequently than others. And studies have consistently shown that a lot of America's antibiotic use is unnecessary. But we need to know more. We need to track antibiotic use and the spread of drug-resistant bacteria even more closely… If we can see where these drugs are being over-prescribed, we can target our interventions where they're needed most." [18]
>
> *President Barack Obama*

compared with other regions, with the highest prescribing among family practitioners.[19] Although much of the data used in the United States is physician longitudinal data, which lacks the ability to assess the appropriateness of prescribing for individual patients, other countries are making use of robust non-identified patient-level data that identifies patient diagnoses to gain greater control over antibiotic misuse.

In the UK, the IMS Health Mediplus dataset of non-identified patient-level data tracks hospital use of antibiotics and assesses how often they were prescribed inappropriately for non-bacterial conditions in pediatric patients (0–18-year-olds). The National Institute for Health and Care Excellence (NICE) guidelines recommend delayed or no prescribing for common respiratory tract infection diagnoses (acute otitis media, acute cough/bronchitis, acute sore throat, acute sinusitis and common cold), deeming them inappropriate. Researchers found that total antibiotic prescribing in the UK declined by 24% between 1996 and 2000, reflecting decreased prescribing for specific respiratory tract infections such as 'acute sore throat' (tonsillitis/pharyngitis) and otitis but increased again by 10% from 2003–2006, due to a fourfold increase in prescribing for non-specific upper respiratory tract infections (URTI) (see Exhibit 6).[20]

Exhibit 6: Respiratory Indications for Antibiotic Prescribing in 0–18-Year-Old Children in UK Primary Care



*Included the indications tonsillitis, pharyngitis and nasopharyngitis.
†Included unspecified respiratory disorder, allergic asthma and asthma unspecified.
Note: Lower respiratory tract infection (LRTI) Upper respiratory tract infections (URTI)

Source: P L Thompson et al. Arch Dis Child, 2009;94:337-340. Adapted by permission from BMJ Publishing Group Limited

Although top-line usage data seemed to show successful reductions in inappropriate prescribing for specific non-bacterial conditions and lower respiratory tract infections (LRTI) – this study alerted the UK government that general practitioners may have simply shifted their billing codes towards unspecific codes to avoid scrutiny, ignoring NICE's guidance on upper respiratory tract infections. This also helped quantify the size of the healthcare gap, allowing researchers to surmise that if NICE guidelines were fully followed, it would lead to a 17–34% reduction in prescribing overall in this age group.

Prior to this study, no published data on the specific clinical indications tied to antibiotic prescribing for children in primary care existed, making it impossible to determine the impact of guidelines and educational initiatives. Because the data used in this study included patient visits and diagnoses, researchers were able to directly link antibiotic prescriptions to patient clinical indication and determine the appropriateness of prescribing.

Having proved its value for broad based public health surveillance, the use of non-identified patient data to combat antibiotic resistance is also now expanding. In Europe, it will be used to build a broad surveillance program called English Surveillance Programme for Antimicrobial Utilisation and Resistance (ESPAUR), which will link antimicrobial resistance information to prescribing in the hospital setting.[21] In the United States, the research conducted by the CDC has allowed educational programs for healthcare providers and patients on appropriate antibiotic use and stewardship to be deployed more effectively including the CDC's "Get Smart: Know When Antibiotics Work" and "Get Smart for Healthcare" programs targeting the community and inpatient settings respectively. This research has suggested that targeting interventions to the South census region and family practice may be most needed and have the most impact.[19,22]

With greater awareness of this issue, critical players are strengthening public health initiatives. For instance, President Barack Obama released a "National Action Plan for Combating Antibiotic-Resistant Bacteria" in March 2015 that sets out lofty goals of reducing inappropriate antibiotic use by 50 percent in outpatient settings and by 20 percent in inpatient setting over the next five years.[15] It also calls for strengthening national surveillance efforts, which are likely to include the use of non-identified data.

Protecting the elderly by identifying excessive use of tranquilizers

The use of benzodiazepines such as alprazolam and diazepam in the elderly has become a major public health issue. Studies show that long-term use of these tranquilizers by elderly patients puts them at increased risk for falls and fractures, motor vehicle accidents and even a type of cognitive impairment that mimics dementia.[23,24] However, these drugs are still commonly used to treat anxiety, insomnia and a range of other behavioral issues despite care guidelines recommending use of psychotherapy and antidepressants for anxiety, and behavioral modification to treat sleep problems.[25]

The U.S. government recognized this safety issue and excluded benzodiazepines from the Medicare Part D program in 2006, ensuring that Medicare would not pay for these medicines in Americans age 65 and over. Although it was assumed that this action would nearly eliminate benzodiazepine use for seniors, research published in JAMA Psychiatry this year led by Dr. Mark Olfson of Columbia University and co-investigators showed continued use of dangerous tranquilizers.[26]

The study examined real-world use of benzodiazepines in the United States by non-identified patients in 2008. Using a dataset representative by sex, age, and insurance coverage of the total U.S. population — the IMS Health Lifelink longitudinal patient prescription database — the research showed that benzodiazepine use still increases steadily with age (see Exhibit 7). Those age 65-80 years old, who are often at greatest risk of injury, constituted the highest users.

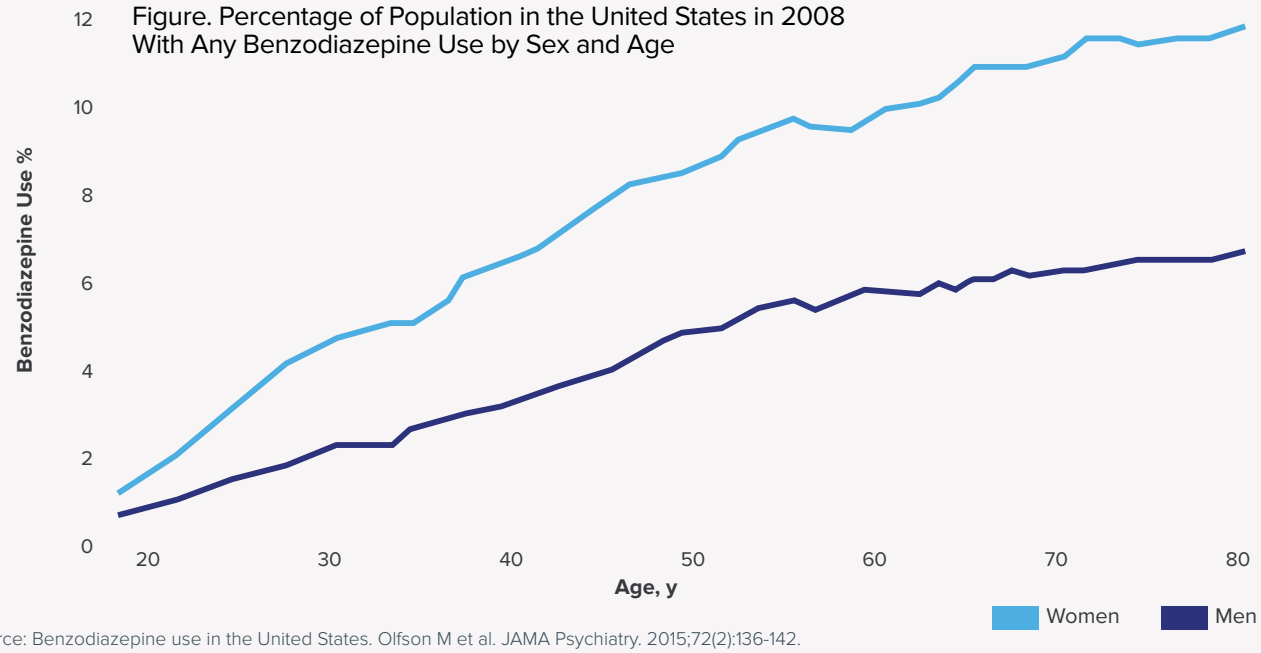
Exhibit 7: Benzodiazepine Use in the United States

Table 1: Prevalence of any Benzodiazepine use, long-term Benzodiazepine use, and use of long-acting Benzodiazepines by sex, and age group in the United States in 2008[a]

Variable	Mean age, y, %			
	18-35	36-50	51-64	65-80
US Population				
With any benzodiazepine use, y	2.6	5.4	7.4	8.7
Among men	1.7	3.7	5.3	6.1
Among women	3.6	7.1	9.2	10.8
Among persons with any benzodiazepine use				
With long-term benzodiazepine use[b]	14.7	22.4	28.0	31.4
Among men	15.6	22.8	28.4	28.8
Among women	14.2	22.2	27.8	32.6
With any long-acting benzodiazepine use, y	24.1	25.4	25.4	23.8
Among men	26.9	29.5	29.4	27.1
Among women	22.7	23.3	23.4	22.4

a. The data source was 2008 LifeLink Information Assets-LRx Longitudinal Prescription Database 2008 (IMS Health Inc).
b. Long-term use defined as 120 days or more supply of Benzodiazepine during 2008.

Figure. Percentage of Population in the United States in 2008 With Any Benzodiazepine Use by Sex and Age



Source: Benzodiazepine use in the United States. Olfson M et al. JAMA Psychiatry. 2015;72(2):136-142.

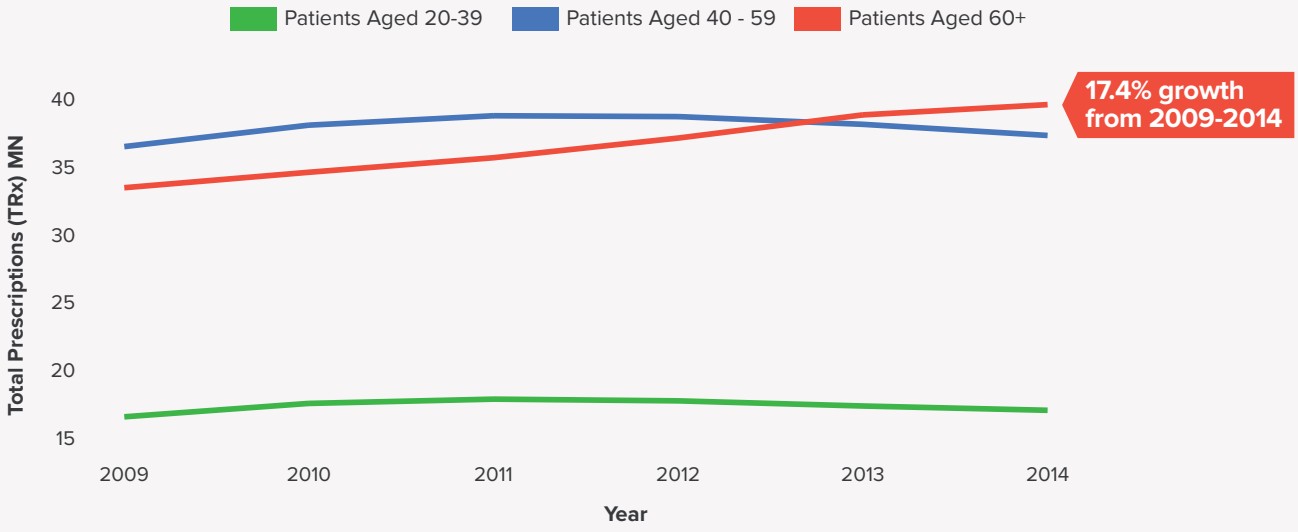
The data also showed disparate use by patient gender in addition to age: use among women was nearly twice the use rate for men. The data demonstrated that many used these medicines over the long term, increasing their risk of harm, and that long-term use (≥120 days) increased with age from 14.7% (18–35 years) to 31.4% (65–80 years).

Another issue uncovered by the research: a great majority of prescriptions for benzodiazepines are not written by psychiatrists. Prescriptions written for approximately 9 out of every 10 older adults using benzodiazepines were issued by primary care physicians or other non-psychiatrists, who sometimes view long-term benzodiazepine use as less of a clinical threat.[27]

The use of non-identified patient-level data provided new insights about a lingering patient safety issue that had not been well described by earlier studies — and one that is increasingly critical to address as the population ages. Prior studies using narrower databases, such as those focusing on Medicare-paid prescriptions alone with no benchmark use for younger populations, had generally supported the view that the 2006 Medicare disincentives had been effective. Visibility in this new study to use among all segments of the elderly population by age (including cash payers and those receiving care through Medicare, Medicare Advantage, Medicaid, or employer retiree plans) enabled comparison and benchmarking of use rates with those for younger populations (predominantly in commercial plans) showing higher use in older groups. With other similar evidence demonstrating the payment ban failure to curb benzodiazepine prescribing, Medicare ultimately removed the exclusion in 2013.

The issue continues, and continues to grow, with the total volume of prescriptions in the elderly rising in comparison with other age groups (see Exhibit 8) from 33 million in 2009 to 39 million in 2014, an increase of 17.4% among the 60+ age cohort. This compares with increases of 3% and 2% for those ages 20–39 and 40–59, respectively.

Exhibit 8: Use of Benzodiazepines in Older Adults



Source: IMS Health, National Prescription Audit, May 2015

Although formal efforts to educate physicians about the advisability of reducing benzodiazepine use in the elderly have been ongoing for more than a decade, this piece of research alerted stakeholders anew to widespread use of a drug class that bears significant risk for elderly patients.[28] It has reinforced that suboptimal care is currently being delivered, raised questions about why safer alternatives are not being used, and it can serve as a renewed warning to clinicians, drawing their attention to the importance of adhering to a set of longstanding best-practice guidelines.

The National Institutes of Health is in the process of re-transmitting this research for precisely these purposes.[29] Now aware that current levels of use by the elderly still exceed the level considered "appropriate and safe," the National Institute of Mental Health has issued statements describing these new data as "worrisome patterns in the prescribing of benzodiazepines for older adults, and women in particular."[30-32]

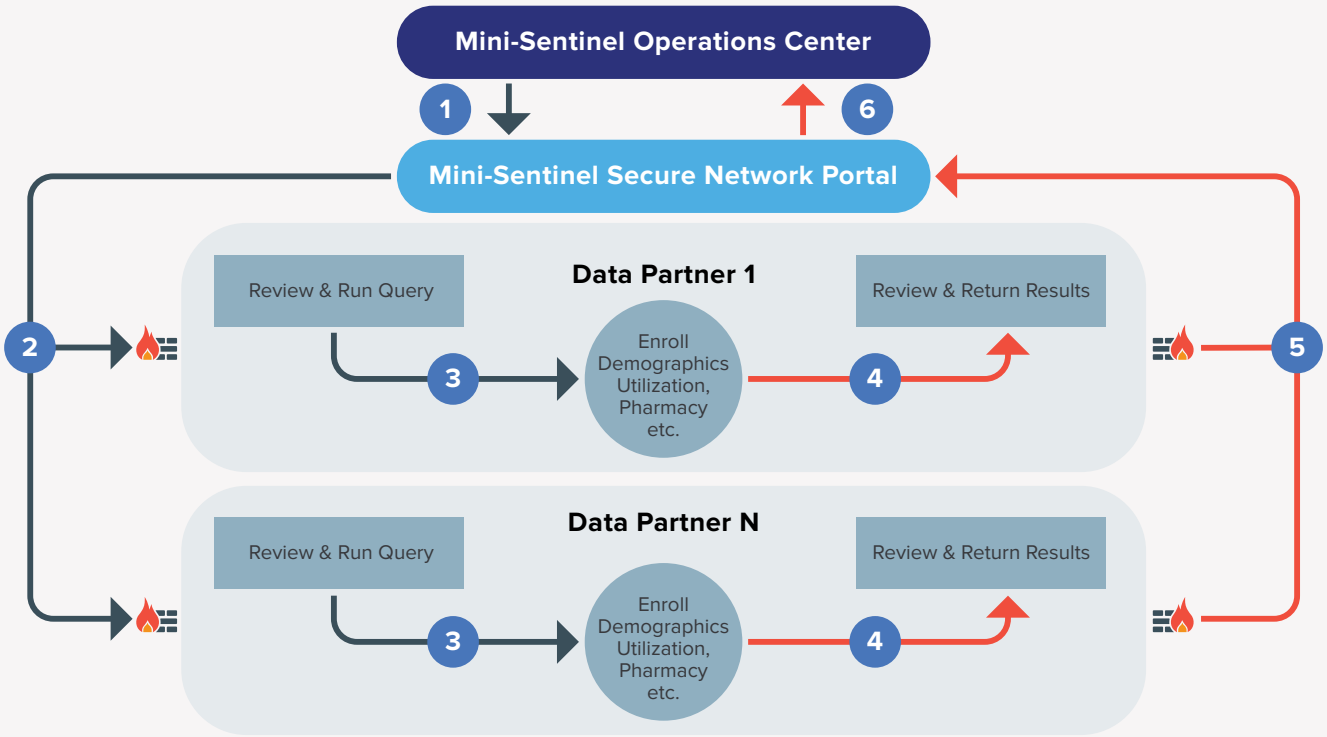
Saving lives through a distributed data network for drug safety

The development of innovative medicines saves lives and improves the quality of life for many worldwide, treating and eliminating chronic and acute diseases. Despite the benefit medicines provide, a proportion of individuals may still develop sensitivities to these therapies or experience adverse events. Some rare drug toxicities may only be revealed once a broader patient population of 100,000 is exposed to a drug, or after a drug has been marketed for some time.[33] It is estimated that there are over 2.2 million serious adverse drug reactions per year in the United States, resulting in approximately 1.5 million hospitalizations and 106,000 deaths, making the role of post-marketing surveillance of drug risk-benefit profiles critical to public health.[34,35]

In Europe, the European Medicines Agency frequently requests drug utilization studies and two-thirds of those studies are based on real-world data sources such as EMRs and claims data.[36] The Food and Drug Administration (FDA) too, grappling with the experience of the rofecoxib arthritis drug withdrawal and other safety concerns of the early 2000s, recognized the importance of tracking additional patient experience once medicines are approved and in the market (pharmacovigilance). In response, the FDA Amendments Act of 2007 mandated the creation of a new computer-based safety tracking system — the FDA Sentinel Initiative — a new system to monitor drug safety based on non-identified patient data.[37]

Sentinel uses non-identified patient-level data to detect potential safety-related signals earlier and more comprehensively than alternative monitoring approaches. The data available to the Sentinel Initiative (and MINI Sentinel, a prior five-year pilot program), comes from provider electronic health records, health plan claims (including Medicare and Medicaid databases), and other electronic sources, with the FDA receiving summary or aggregate information from these separate data partners in response to its pharmacovigilance queries (see Exhibit 9).

Exhibit 9: Mini-Sentinel Distributed Analysis



1 User creates and submits query (a computer program)
2 Data partners retrieve query
3 Data partners review and run query against their local data
4 Data partners review results
5 Data partners return results via secure network
6 Results are aggregated

Source: Seventh Annual Sentinel Initiative Public Workshop, February 5, 2015

While the Sentinel system does not replace other FDA tools such as the FDA's voluntary Adverse Event Reporting System (FAERS), where healthcare partners or foreign regulatory bodies notify the FDA of new safety concerns alongside concerns seen in clinical trials, meta-analyses, and case reports, it supplements these tools to determine whether an issue exists by enabling larger population studies to confirm these signals.

Sentinel's ad-hoc access to data on a population of 178 million people receiving prescriptions enables the FDA to run rapid, one-time, retrospective safety assessments to investigate if specific drug adverse events have occurred.[38,39] It also allows the prospective (sequential) monitoring of patient experience for the 48 million patients seen by their data partners, helping them to track concerning events over time as additional data accrues. It can also be used to clarify when drug risks do not exist and when restrictions on marketed drugs can be eased, thus preserving therapeutic options.

These queries are also particularly important to reduce or prevent unrecognized risk to vulnerable patient populations such as the elderly, pediatric populations, and childbearing women. Patients with numerous co-morbidities who may take potentially interacting drugs are also a concern for drug monitoring. These sensitive populations are rarely studied in company-sponsored clinical trials, however, physicians will frequently prescribe on an off-label basis in the hope of treating a patient successfully.

With the ability to interrogate non-identified patient data, the FDA is now able to watch the use of these drugs off-label and understand how many patients are being exposed, and how they fare. They can consider whether the occurrence rates of adverse events necessitate deeper investigation, especially when drug indications are correlated to diseases often impacting children or elderly. For this reason, one goal for the Sentinel system has been to develop a reusable tool to monitor drug use over time among pregnant women who deliver a live-born infant, and is the largest dataset currently tracking use by this population.[40]

A wide range of applications for Sentinel is possible because the data included in the project is at the patient level and can show drug dispensing and use, doctors' visits and outcomes (morbidity or mortality) on a large scale, and even put them in the context of patient risk using demographic information. Since the data is derived partially from EMRs and claims datasets, additional clinical data is also available on vital signs and laboratory test results, showing the detailed impact of medicines on patient health.[41]

So far, Sentinel studies have enabled the FDA to add a warning to the anti-hypertensive drug olmesartan's label after confirming risk of a severe intestinal problem (spruelike enteropathy), determine that there was no increased risk of febrile seizures with a type of flu vaccine, and no increased risk of bleeding with Dabigatran, an atrial fibrillation drug (in fact it proved to have lower rates of both thrombotic and hemorrhagic strokes than warfarin), thus maintaining use of a valuable drug.[40,42] The system also showed that a Human Papillomavirus Vaccine does not pose higher risk of venous thromboembolism after concerns were presented to the FDA Pediatric Advisory Committee.[43] Further, the FDA has been able to assess the impact of some regulatory activities or label changes on the medical community's prescribing or on patient outcomes. For instance, using Sentinel, the FDA confirmed that the use of prasugrel diminished versus the use of clopidogrel in patients with prior TIA/stroke after a change in prasugrel's labeling.[44]

While Sentinel has had a number of accomplishments, some criticisms exist of the program – that it has cost hundreds of millions of dollars but has led to few new significant findings, and that data from partner sites often provided conflicting assessments. Of 137 drug assessments, only 4 led to the issuing of drug safety communications through February 2015.[38] Nonetheless, the original pilot program called Mini-Sentinel pilot project has been considered a success, and the FDA, along with Harvard Pilgrim and its other data partners, is transitioning now to a full scale program. It already is one of the largest distributed data health networks in the United States and will be used to run hundreds of queries per year with the intention of making our system safer and ensuring that drugs are used safely.[45]

Combating the prescription drug abuse epidemic by understanding patient behaviors

Non identified patient-level data can reveal patterns of health behaviors among patients and physicians. Behaviors such as physician prescribing patterns, the doctor or hospital visits made by patients, or use of preventative care and medicines are all visible in such data for analysis. This ability to see patient behavior has been put to good use to combat the drug abuse epidemic by identifying and modeling patterns of prescription filling known as "shopping behavior," employed by abusers to gain access to prescription drugs inappropriately.

While drug abuse is often associated with illicit substances, the non-medical use of prescription drugs such as painkillers, tranquilizers and stimulants is widespread. Over 52 million people in the United States are estimated to have used prescription drugs non-medically in their lifetime resulting in as many as 1.4 million emergency department visits per year.[46,47] The number of prescription medicine abusers over age 12 in 2013 was 6.5 million, with as many as 4.5 million people abusing painkillers, such as oxycodone and hydrocodone, alone.[48]

Manufacturers have worked to combat this issue through the recent launch of several abuse-deterrent formulations of painkillers, but despite these efforts, this and other classes of drugs remain highly addictive and prone to abuse.[49] It remains critical to understand patterns of misuse of prescription drugs, gauge which agents pose a greater risk of abuse, and assess the impact of efforts by various stakeholders – including the Drug Enforcement Administration, pharmacy chains, wholesalers and manufacturers – to combat this major public health issue.[50,51]

Non-identified patient-level data has been critical in developing an understanding of the ways drug abusers interact with the medical system, particularly in understanding patient shopping behavior. For opioids and attention deficit hyperactivity disorder (ADHD) stimulants – both subject to abuse – patients are known to obtain prescriptions from multiple prescribers ("doctor shopping"). Good behavioral models of this phenomenon – intended to define shopping behavior while avoiding the inappropriate flagging of individuals with legitimate use – have been created using non-identified patient data by comparing patient use of drugs prone to abuse with patterns of drugs not typically abused. For instance, one study compared patient use of ADHD drugs to that of asthma drugs, and an earlier study for opioids similarly compared the use of opioids to diuretics (non-abused) using longitudinal patient data.[52,53]

In both studies, researchers found that being a patient with overlapping prescriptions written by two or more prescribers and filled at three or more pharmacies was the best predictor of abuse. In the ADHD study this pattern was 400% more frequent among those shopping for ADHD drugs than those prescribed asthma medications. Applying this definition, this study was also able to show that shopping was most common in younger subjects aged 10–39 years, and that a small number of abusers accounted for most shopping behavior. Among patients who shopped, 9.2 % of them shopped six or more times and accounted for 42.0 % of all shopping.

Other studies using these models have studied how soon shopping behavior was observed after a patient's first prescription, the typical number of events per shopper, their preferred drugs within a class, and their preferred methods of payment. Opioid studies on non-identified data have shown that abusers tend to fill prescriptions at multiple pharmacies across state lines, avoid combination products, and often pay cash.[54-56]

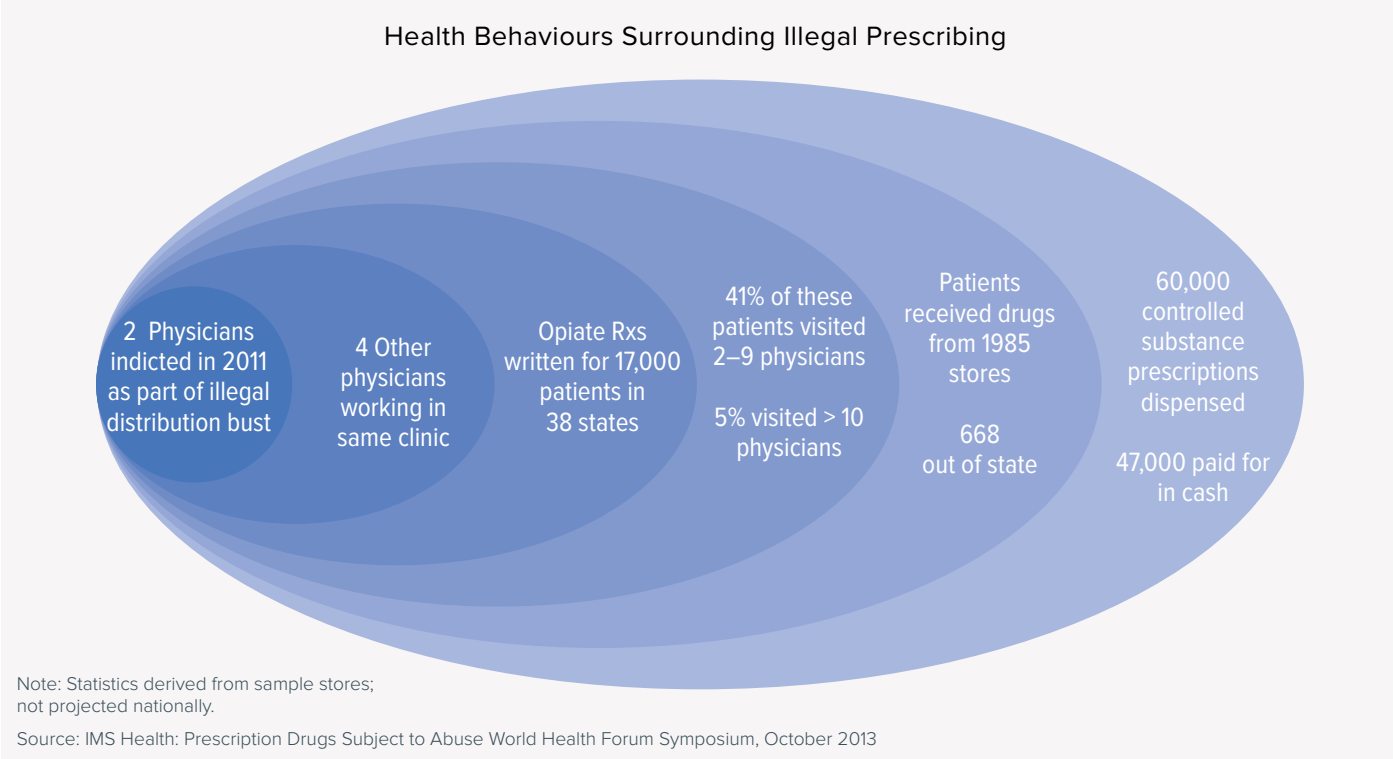
Unlike in previous studies that had looked at simple counts of the number of prescribers or pharmacies a subject had gone to within a year to define shopping behavior, modeling this data using non-identified patient longitudinal sources allowed researchers to distinguish successive prescribers from concomitant prescribers (or overlapping dispensing within the same day).[53] By developing more objective and accurate measures based on the number of distinct pharmacies and overlapping prescribers, this model has become a tool for stakeholders to identify where shopping behavior is most prevalent and formulate strategies to prevent it, and to apply this definition in a more objective and targeted way than subjective measures of abuse or dependence. This enables health care providers, insurers and pharmacies to implement monitoring that can decrease abuse or diversion.

The models generated using non-identified patient-level data also provide better means to guide physician education programs. One study, using the modeled shopping definition, examined which prescribers had significantly higher numbers of shoppers. Results showed that prescribers who were male and in their 70s, or who prescribed schedule II opioids, had an increased likelihood of having shoppers. Prescribers with greater numbers of opioid patients had a greater proportion of shoppers, with prescribers having 66 or more opioid patients (25 percent of prescribers) prescribing for 82 percent of all shoppers. This indicates that targeted educational programs to such doctors would be the most cost effective.[57]

One unpublished study examining the practice of two physicians charged with illegal activity using non-identified patient data can be seen below in Exhibit 10. This straightforward use of such data, extrapolated from the physicians, reveals details of both physician and patient behavior around these controlled substances.

The societal costs for prescription opioid abuse were estimated at $55.7 billion in 2007, including lost work productivity, treatment and criminal justice.[58] And according to the Drug Abuse Warning Network, in 2010 there were 15,585 emergency department visits related to nonmedical use of ADHD stimulants incurring costs to the medical system.[59] With tools such as patient behavioral models, healthcare stakeholders become better at countering drug abuse and saving the lives of young people, while also reducing costs to the medical system.

Exhibit 10: Application of Non-Identified Information



Note: Statistics derived from sample stores; not projected nationally.

Source: IMS Health: Prescription Drugs Subject to Abuse World Health Forum Symposium, October 2013

Improving care and decreasing hospital readmission rates

Policymakers can use non-identified patient data to understand flaws in the current medical system and implement policy changes. They can also measure the impact of these health policy changes to see if they are working. A good example of such use is research led by Dr. Stephen Jencks, published in the New England Journal of Medicine, that revealed unexpectedly high readmission rates at hospitals. Armed with non-identified data from Medicare claims, researchers discovered that patients with severe illnesses were frequently released from the hospital only to return to the hospital weeks later when their condition worsened again.[60]

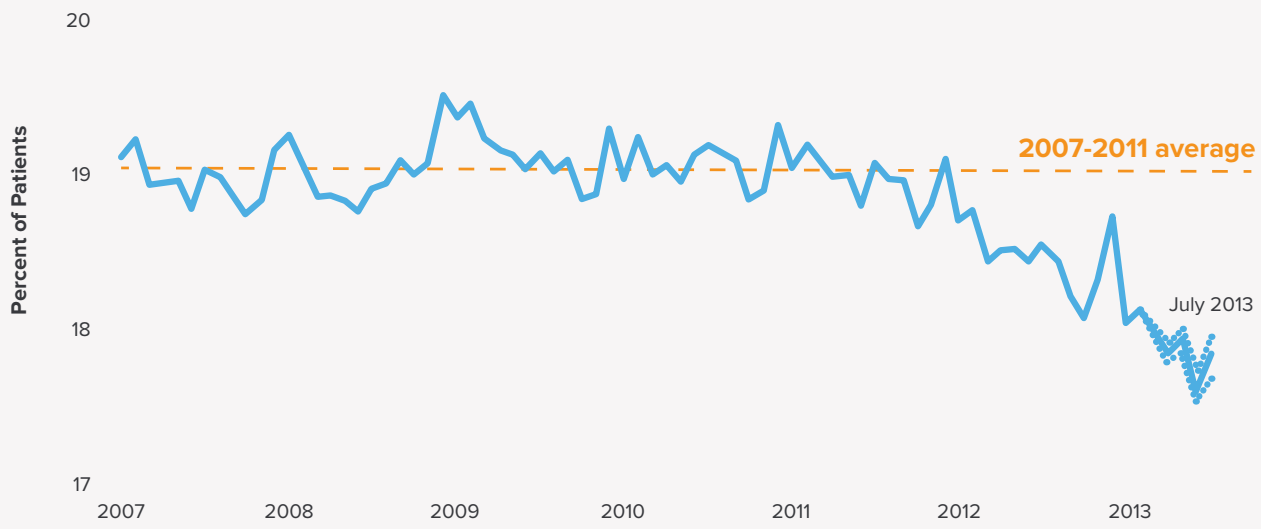
During the time period examined in this study (2003–2004), Medicare paid on a per hospitalization basis (diagnosis related group) reimbursing each time a patient was seen or treated. Using non-identified patient data, researchers discovered that approximately 20% of all Medicare patients discharged from a hospital subsequently returned to the hospital and were readmitted within 30 days. This strongly implied that little patient follow-up was being provided once the patient left the hospital.

The readmissions study found that half of the patients readmitted had no interactions with the medical system between discharge and readmission, and had seen no physician during their time outside the hospital. The discovery of this pattern, shocking to policymakers, drove policy changes in Medicare payment systems aimed at keeping patients healthy and out of the hospital by encouraging patient follow-up. Authorized by the Affordable Care Act (ACA), CMS created the Hospital Readmissions

Reduction Program (HRRP), which reduces payments to hospitals that readmit a larger number of patients with specific diseases soon after discharge. This had the effect of shifting the system from one that rewarded, and even paid more for, poor care (since hospitals could lose money by avoiding readmissions) to one where hospitals understand that money can be taken away.

Until Jencks' use of non-identified patient-level data to tackle the topic, granular information on the frequency and patterns of re-hospitalization in the United States had not been readily examined. Use of this data provided not only a critical policy improvement to a government program, but also a strong impetus for innovation in the medical system – the shift from a system that pays for volume to one that pays for value. The changes spurred by this study helped re-align financial incentives to provide better health outcomes for patients, and because changing incentives is one of the strongest ways to change behavior, the downstream effects have been noticeable. Hospital readmission rates have turned sharply lower since the ACA began penalizing hospitals, as tracked using non-identified patient data from Medicare, with a direct result of keeping patients out of hospitals, as well as reducing costs to the health system overall (see Exhibit 11).

Exhibit 11: Medicare 30-Day, All-Condition Hospital Readmission Rate, 2007-2013



Source: Centers for Medicare and Medicaid Services, The Annual Report of the Council of Economic Advisors, March 2014 [61]

Optimizing care for better outcomes

The concept of "evidence based medicine" developed by Dr. David Sackett and colleagues at McMaster University in the 1980s, with its goal of a healthcare system solidly founded upon data, has rapidly become the foundation of our medical system. This shift away from qualitative understanding and conventional wisdom to one that holds high-quality evidence in highest regard was rightfully acknowledged as one of the top 15 advances in medicine.[62]

Non-identified data is now used to support an evidence-based healthcare system in a range of ways to improve patient outcomes (see Exhibit 12).

Exhibit 12: Uses of Non-Identified Patient Data to Optimize Care

Source: IMS Institute for Healthcare Informatics, July 2015

"Sometimes people are receiving more intense treatments over time and it seems to be benefitting them… and in other cases it seems to be harmful… People are receiving all sorts of treatments that are not particularly medically recommended, don't seem to be standard in the literature and don't seem to be particular effective for that patient. …So, it helps to understand both the good features and bad features about the medical system."

David M. Cutler, Otto Eckstein Professor of Applied Economics, Harvard University

Comparative effectiveness and best practices

Traditional interventional/randomized controlled clinical trials and observational trials (where patient groups unintentionally receiving different care are compared) provide essential evidence on the value and safety of medicines and optimal treatment patterns, but there is a growing role for real-world non-identified patient data, queried retrospectively, to test similar clinical hypotheses and gain immediate answers without the high cost of clinical trials. Such analyses typically create statistically matched cohorts of non-identified patients to draw comparison, or compare outcomes of cohorts of differing patients.

Prospective clinical trials vs. retrospective real-world data studies

Clinical trials that the NIH, hospitals, and other organizations run require facilities to build an infrastructure and enroll patients – taking time and staff to answer a single question. Patient recruitment into these studies can also be a challenge, particularly in the case of rare diseases where few patients may be available in a specific area for enrollment. The larger national populations visible in some non-identified patient sources – their data readily available – is therefore enormously valuable for research.

The value of information gleaned from large population-based studies tracking patients over many years is widely recognized. Such longitudinal prospective cohort studies, such as the Framingham Heart Study and the Nurses' Health Study that assessed risk factors for cardiovascular disease and cancer, and looked at broader patient groups numbering 5,209 and 122,000 respectively, are prized in the realm of research.[63,64] However, these were monumental efforts and in such trials a determination of whether some participants develop the disease outcomes of interest can often only occur after a long time lapse.

Examining non-identified patient-level data retrospectively looking at large populations and their health changes over time is particularly useful to find such correlations rapidly and inexpensively, and to explore the impact of treatment or patient behavior on patient health. Still, the nuanced data (e.g. custom surveys and comprehensive lab work) collected in prospective clinical trials often go beyond what is regularly available in clinical EMR databases and the non-identified data they generate, and remain invaluable for such research.

Patient populations, response and risk

Cohort analyses also provide information that health providers use to personalize care to their patients based on their individual risk. As diagnostics, tumor typing and other technologies usher in an era of personalized medicine, non-identified research on subpopulations and their individual risk has helped establish best practice guidelines for specific patient types. For instance, population-based cohort analyses using health plan claims data, have identified that patients with inflammatory bowel disease (IBD) (n=108,000) are at risk for many other conditions including pneumonia, herpes zoster and melanoma and non-melanoma skin cancer, and separately, that women with polycystic ovary syndrome (PCOS) are at risk for thromboembolism if given contraceptives.[65-69] Such data informs physicians to monitor these patients more closely and see that they receive appropriate screening. As we move to a system that smartly relies on preventative care to reduce healthcare costs and avoid poor and costly patient outcomes, the ability to identify which patients are at risk using already-generated data will be a powerful tool.

Health technology assessment

Non-identified patient data also contributes to health technology assessments that help determine how medicines and other therapies should be used and whether they should be paid for. The explosion in the availability of such data has helped agencies that conduct health technology assessments begin including analyses of patient outcomes and comparative effectiveness of therapies in the real-world. While in the past, decisions on the role of medicines and formulary inclusion might have been based solely on the dossiers of clinical trial data submitted to Pharmacy and Therapeutics Committees of Payers and PBMs (often performed by pharmaceutical companies only on select patients), most of these agencies now take into account broader effects on patients and health system, including evidence on patient adherence, alternate drug regimens used in every day practice, and efficacy in specific patient populations as demonstrated in non-identified patient-level data.

For instance in 2008, Wellpoint (now Anthem Inc.), became the first company to begin evaluating pharmaceuticals considering the medical cost offsets they generate by curing patients, preventing illness, and improving patient quality of life.[70] Such assessments of the role of medicines are dependent on a wealth of outcomes research and comparative effectiveness studies produced by pharmaceutical companies, clinicians, government and private payers alike, often using non-identified patient-level data. These can address the relative efficacy of therapies, rates of adverse events, typical patterns of patient adherence, cost offsets, or the impact on specific patient outcomes over the long term.[71-78]

Benchmarking performance and quality of physicians and facilities

The process of optimizing care for better health requires tools to recognize when optimal practice is occurring and when it is not. Non-identified patient-data can be used to compare one facility's or one physician's performance (measured in terms of patient outcomes and costs) to another or to best practices, correctly risk-adjusting outcomes to account for differences in patient population, disease severity, and case-mix. This provides direct and tangible benefits to both patients seeking to select the best doctors and health administrators who would otherwise not know how their physicians measure up.

Payers and health systems are able to measure the performance of their physicians against a range of metrics rolled up from the patient level, and then compare their treatment patterns to best practice overall or to their peers to show whether they are delivering high-quality care. These also enable administrators to have real conversations with doctors about their performance. By showing them their own patterns of treatment, and how they compare within a distribution of other doctors, the conversation is no longer a theoretical one about what best practices look like. Rather, physicians can see clearly whether they are using health resources more or less than is considered best practice, or if other doctors have better outcomes. These tend to produce more complex and grounded conversations to influence physician behavior.[79]

This influence is particularly important since changing physician behavior to align with new best practices can be an exceedingly long process. It may take up to 17 years for advances to be incorporated into ordinary clinical practice.[80] The path to speeding best practices is through such performance measurements, the process of sharing these results, and tying incentives to these – such as implemented with the ACA.

> "We realize that in order to capture the 30 to 40 percent of value which is now lost because of stuff in healthcare that doesn't help human beings, actually hurts them... we have to fundamentally change behavior. ...The way to change behavior is not just changing the incentives... You have to change the incentives but you also have to enable people with timely feedback about what they're doing, and the consequences of that, and what they should be doing. And that's all about data." [81]
>
> *Dr. Glenn D. Steele Jr., President and CEO, Geisinger Health System*

Curing the health system

The examination of non-identified patient data, sometimes without a hypothesis, can also lead researchers to unexpected realizations about gaps in our knowledge about the health system. They may discover where further research is needed or where there may be a societal concern. Studies on non-identified data can show that pregnant women frequently take statins through their pregnancies although clinicians have poor understanding of downstream effect, or that use of mental health medications in the United States has great variability, with use in some areas by only 1% of residents and in others by 40% of residents, driving questions whether discrepancies are being driven by variability in access to healthcare and insurance coverage across regions.[82,83] Possibly the greatest value for policymakers comes from understanding such health disparities among subpopulations and across regions. Variations in access to healthcare, utilization of healthcare resources, and the differences in disease or health status can be clear indicators of health disparities that need to be addressed, often resulting in avoidable deaths, and costs to the healthcare system.

> "De-identified health information is particularly valuable for detecting and measuring variations in the availability and utilization of health services and understanding how these variations contribute to health disparities that impair quality of life, reduce productivity, and result in premature death in different communities and in different segments of our population… There is ample evidence of measurable disparities in access to health care, the delivery of health services and health status among the nation's growing ethnic and racial minority populations… With the widespread availability of de-identified health information, it is possible to assess how variations in access to, and utilization of, health services contribute to ethnic and racial health disparities, and develop strategies and techniques to address these disparities." [84]
>
> *Dr. Louis W. Sullivan, Former U.S. Secretary of Health and Human Services, CEO and Chairman of The Sullivan Alliance, and President Emeritus of the Morehouse School of Medicine*

Non-identified patient-level data can also be used to argue for revisions to existing policies. For instance, when the government cut Medicare payments for Dual-energy X-ray absorptiometry (DXA) bone density tests in 2007, studies using de-identified data provided evidence that Medicare policy was in conflict with cost-effective preventative healthcare. These studies showed that DXA testing in fact led to fracture prevention, benefiting older women and producing cost savings to the Medicare program.[85-88] At a higher level, using global non-identified patient-level data (where available) enables us to compare our national policy against that of other nations to understand whether a larger policy gap exists.

The discovery of these gaps, and the need to close them, helps to initiate the process of crafting new policy and contributes to the improved functioning of health systems. As described in prior examples on readmissions and antibiotic misuse, the discovery of when and where waste occurs helps guide our ability to make policy – both in public and private settings of care. Without this data our ability to craft policy would be 'substantially hindered'.[79]

Identifying savings opportunities – paying for value not waste

As the costs of healthcare have risen, concerted efforts are being made to protect the sustainability of the medical system by speeding the adoption of best practices, preventative care, and optimized approaches that benefit the patient while also eliminating money spent on wasted/ineffective treatment. Both hospital systems seeking to remain within the bounds of their budgets and payers seeking to reduce the total cost of care per member, aim to implement cost effective measures while maintaining or improving patient outcomes. Patients are similarly looking to save money as health plans have shifted to insurance models with greater member cost-sharing. They benefit from access to physician and hospital cost comparisons produced from summarized non-identified patient data, which enable them to make informed economic decisions about their health.

Nearly all of the examples provided on the use of non-identified patient-level data to optimize the healthcare system and patient care can also serve to reduce costs. For example this data serves to reduce costs across several broad areas:

- Reducing waste
- Reducing or eliminating payments for inefficient care while incentivizing cost-effective care
- Optimizing educational initiatives and clinical trials
- Enabling population health strategies

More specifically, it contributes to savings through the following uses (see Exhibit 13).

Exhibit 13: Uses of Non-Identified Patient Data to Identify Savings Opportunities

Source: IMS Institute for Healthcare Informatics, July 2015

Detect waste and inappropriate use of health resources

Making healthcare more effective and more affordable does not mean providing low quality care; in fact this aim often serves as a stimulus to improve the quality of care through the identification of cost drivers – which are often indicative of poor care. Just as reducing the high rates of hospital readmissions or infections with real-world data can cut costs while providing a straightforward benefit, so too does the reduction of healthcare services that are harmful and inefficient such as the overuse of antibiotics.

Studies using non-identified patient-level Medicare data have also shown the use of post-acute care services (skilled nursing facilities and inpatient rehabilitation facilities) after hospital stays to be associated with lower patient survival. Evidence such as this, combined with new payment and delivery models (Accountable Care Organizations and bundled payments), have propelled efforts to shift patient care to more effective settings and reduce costs.[89,90]

Pay for performance and population health

Patient-level data – both identified and non-identified – have directly contributed to cost management efforts by supporting this shift from volume to value-based payment models. These arrangements often include a component of 'at risk' pay tied to performance measurements from aggregated patient data metrics. Such payment models have in turn propelled the focus of institutions more strongly toward prevention and the rollout of population health management initiatives that also rely on non-identified patient data. By creating and applying behavioral models using this data, health systems and payers can better predict patient risk of costly adverse events, create new strategies to help patients who are heavy users of healthcare, or find areas to reduce preventable costs.

Examining healthcare costs

To understand where treatments or initiatives are saving the health system money, it is first necessary to understand the baseline prevalence of disease or baseline costs of treatment ('total cost of care'). Understanding the incidence or prevalence of diseases, including rare ones such as Castleman's Disease (a disease of the Lymph nodes), can be a challenge and can be assessed by non-identified patient-level data.[91] Such baseline estimates can then be used in predictive analytics, to understand how costs will change over time, or show how treatments impact these over time. For instance, one study assessed the rates of progression from Hepatitis C to advanced liver disease in the United States and projected the future numbers of patients affected, helping to clarify the value of cost offsets of newly emerging therapies.[92] The data also enables comparisons of costs, or the cost offsets, of one treatment versus another, such as a study examining how the total healthcare costs of treating sleep disorders vary with the use of different agents.[93]

Inform, target and measure interventions ensuring appropriate use

Such baseline measurements are also necessary to measure the impact of initiatives. Applications of non-identified patient-level information can include targeting educational or policy initiatives more directly to areas where they will be effective, providing an understanding of the current state of therapy usage, or even measuring the success of these initiatives or policy changes. One example of this is a study by the FDA to measure whether a new black box warning on atomoxetine, an attention-deficit/hyperactivity disorder medicine, diminished its use.[94]

Develop better clinical trials

Finally, this data has also been put to good purpose to develop better clinical trials. Non-identified patient-level data has been used to help pharmaceutical companies and providers locate areas of high patient prevalence for rare diseases (to facilitate recruitment), and to find physicians who frequently treat specific conditions and see many patients who might be able to enroll.[95]

De-identification frameworks to create non-identified patient-level data and maintain patient privacy

Protecting patient privacy through de-identification

While the vast array of available healthcare data coupled with increasingly sophisticated data analytic techniques can help improve healthcare and reduce medical costs, these benefits must be balanced with appropriate respect for individual privacy. Because useful patient data derives initially from individuals and their health experience, safeguards must be taken to ensure that an individual's identifiable health information is not distributed or revealed outside of appropriate and permitted situations. With the emergence of big data, healthcare stakeholders have taken their duty to protect patient data very seriously and have renewed efforts to codify privacy frameworks and guide implementation.

One of the most effective routes organizations can take to appropriately protect individual privacy for "Big Data research" is through the use of non-identified healthcare information.[96] By implementing effective steps to remove identifiers (de-identify) and protect personal data through contractual limitations and appropriate safeguards, the ability to evaluate data at a patient or aggregate level can be preserved while making the risk of any future identification of any individual very small. The removal of identifiable information serves to protect patient privacy as data is shared (including protecting identities in the event of a security breach), thus providing significant benefits for most research purposes and making this de-identification process an effective risk management tool.[97]

Techniques to render patient medical information appropriately non-identified include a combination of removing, generalizing and disguising some information, along with privacy and security safeguards (administrative, technical and physical) and contractual limitations to ensure there are sufficient controls over information to keep the information non-identified and ensure use in a responsible manner (see Exhibit 14).

> "Organizations endeavoring to share such data might consider employing a combination of several approaches to mitigate re-identification risk. These include technical controls, such as removing quasi-identifiers and other kinds of information that might be used to re-identify the data subjects; continuously surveying for data that could be linked to the de-identified information that they are sharing; controls on the de-identified data, such as data use agreements and click-through agreements that prohibit re-identification, linking to other data, or sharing with others; and technical controls that limit the activities of data recipients." [97]
>
> *Simson L. Garfinkel, Ph.D., National Institute of Standards and Technology*

Exhibit 14: Techniques to Render Information Non-identified

De-identification techniques that alter specific data elements	
Removal or masking/disguising of data	Typically applied to Direct Patient Identifiers such as social security numbers, names, email addresses, medical IDs, and genomic sequence information
Generalizing information	Applied to indirect patient identifiers such as dates (date of birth, treatment dates, other dates), age, sex, race, ethnicity, postal codes, socioeconomic data. – i.e. Age-ranges can be provided for grouped patients such as 'Patients Aged 30-35' and Postal-Codes can be truncated such as changing 36487 to 364XX Methods are tied to the risk of re-identification for a specific dataset

Continued security controls to protect privacy and maintain de-identification	
Administrative controls	Policies/procedures, contractual requirements (e.g., an obligation to not re-identify individuals, specific privacy and security training, mandatory data classification schemes with access restrictions)
Technical controls	Encryption, access management, etc.
Physical controls	Firewalls, badge access to facilities, etc.

Non-Identified Information

Win-win approaches to de-identification

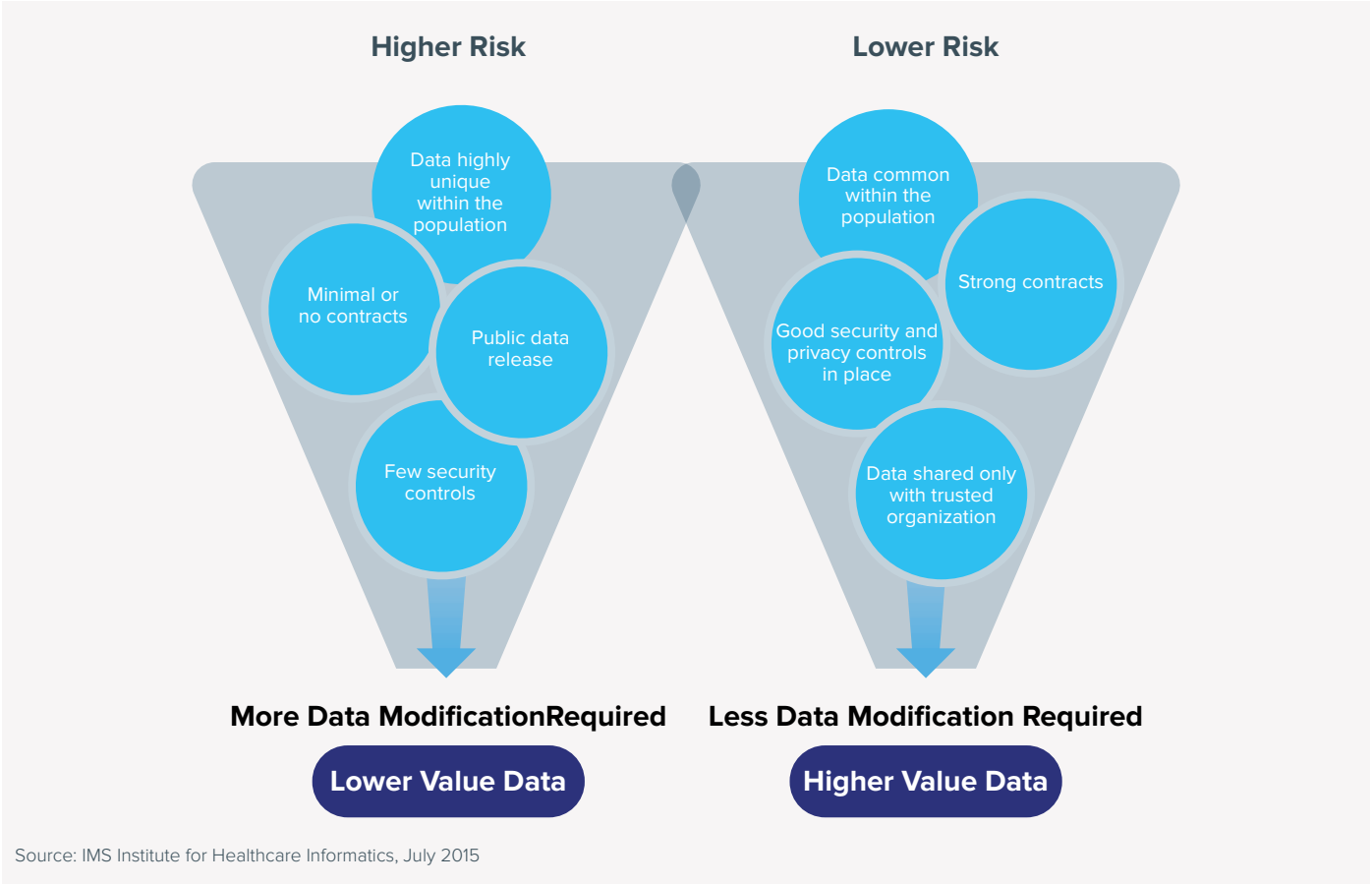
Since the value derived from patient data may be related to the presence or absence of specific data elements (such as age or gender) which are indirect patient identifiers, it is critical to balance carefully the value of research data with appropriate protection of patient privacy. "Poorly de-identified" data revealing too much patient detail may maintain all of its research value, but increase the risk of patient re-identification unacceptably. On the other hand, some approaches that de-identify data may go too far and make the data unusable or of limited value for research. For example, with one extreme, all patient identifiers are removed such that there would be no basis to re-identify an individual. This step – while effective at fully protecting patient privacy – creates data sets with extremely limited uses and eliminates most societal benefits derived from healthcare data.

Rather than mandate the removal of all individual identifiers to reduce the risk of patient re-identification to zero, reducing the societal benefits of research, other risk-based approaches to protect patient privacy seek a more reasonable balance. These approaches remove identifiers such that the risk of re-identifying any individual is very small while still preserving the value of the overall

patient data – allowing examination both on a longitudinal basis for individuals, and on an aggregated basis for groups or populations. Where taken with appropriate contractual protections and appropriate technical, physical and administrative safeguards – this approach to creating non-identified data provides a broader range of benefits to the healthcare system while still protecting patient privacy.

This approach also allows for variable application of de-identification methods and level of controls based on the risk of re-identification of individuals for a specific dataset and use. Privacy risk typically correlates to the identifiability of the data elements (either alone or in combination) and the nature of the data use (e.g., received by trusted organization, public posting to the internet, etc.). For instance, if the same dataset were duplicated for different purposes – one to be posted publicly online and the other to be used internally in a highly secure data warehouse – they would be treated as two different datasets for the purposes of a de-identification analysis. The public dataset would require more modification since there are fewer controls in place, whereas the quality of data kept in the data warehouse could be much higher (see Exhibit 15).[97]

Exhibit 15: Creation of Appropriately Non-Identified Data



Source: IMS Institute for Healthcare Informatics, July 2015

"De-identification is not a single technique, but a collection of approaches, algorithms, and tools that can be applied to different kinds of data with differing levels of effectiveness. In general, privacy protection improves as more aggressive de-identification techniques are employed, but less utility remains in the resulting dataset." [97]

Simson L. Garfinkel, Ph.D., National Institute of Standards and Technology

Legal guidelines

The importance of protecting patient privacy has made it critical to develop frameworks for the appropriate de-identification of health care data – outlining specific actions to be taken with the data and overall contractual and security safeguards needed based on a risk assessment appropriate to the specific use at hand. De-identification combined with a consistent, broadly applicable policy framework to govern its use by healthcare stakeholders can achieve the best of both worlds – important benefits from valuable data analytics combined with strong privacy protections.

De-identification frameworks have been developed by both governments and healthcare stakeholders alike, in the form of legal guidelines and codified approaches to protect patient privacy rights. The primary existing legal frameworks – both in the United States and around the world – adopt a contextual, risk-based approach to ensure an acceptable and insignificant re-identification risk level. In the United States, the HIPAA (Health Insurance Portability and Accountability Act) Privacy Rule and related HHS guidance documents provide the most detailed approach to the "de-identification" of individual data.

"Some people say you can never completely eliminate privacy risks, so you shouldn't make any data available for research. I consider that to be nonsense. There is no such thing as zero risk – it simply doesn't exist – it's a myth. It doesn't exist anywhere in the world, so why invoke it here [in Healthcare]? Do you go outside in a rainstorm even though it is possible, (but highly unlikely) you might be hit by lightning? Of course you do! The risk is so minimal that you it doesn't stop you from doing what you need to do."

Dr. Ann Cavoukian, former Information and Privacy Commissioner for Ontario, Canada

Pursuant to the HIPAA Privacy Rule, patient information can be "de-identified" using two different approaches– the "safe harbor" approach where 18 specific identifiers are removed from the data or the "expert determination" risk-based approach where an expert determines that the risk of re-identification of an individual is "very small," based on a variety of relevant factors. Where either of these approaches is taken, the resultant non-identified data is no longer considered individually identifiable and ceases to be subject to the HIPAA rules.[98]

Typically there is alignment within global statutes to define data as not being individually identifiable information (i.e. non-identified) if it neither identifies an individual nor can be used to identify an individual at a future time using other information sources. These principles also typically recognize that de-identification can prevent re-identification but may not eliminate risk altogether; some risk is tolerated because of the public benefits of data use, but risk of re-identification must be actively assessed (see Exhibit 16).

> "What we need are strong de-identification standards combined with a risk of re-identification framework. Yes, you can point to cases where data was de-identified and then re-identified, but when you look at those cases, they all involve cases where the data were poorly de-identified. Like anything – you can encrypt data very well or poorly, it all depends on how you do it."
>
> *Dr. Ann Cavoukian, former Information and Privacy Commissioner for Ontario, Canada*

The risk of re-identification

While the theoretical possibility of re-identification exists, virtually all public reports of re-identification related to health care data have involved data sets that were not de-identified according to any standardized frameworks.[103] Rather than discrediting de-identification in any way, these isolated examples instead focus attention on the critical need for appropriate de-identification frameworks that can be applied in a consistent, manageable manner.

When data is properly de-identified, there is proof to show that efforts to re-identify the data have failed. For instance, during the Heritage Health Prize competition, a vendor was hired to de-identify a dataset of 113,000 individuals to be used in the competition. Before data was shared with entrants and unbeknownst to them – the sponsor of the competition hired a well-known academic to attack the data. Despite attempts, he was unsuccessful in re-identifying a single person within the dataset.[99]

De-identification frameworks

Beyond the broad HIPAA approach in the United States, which remains the most explicit standard for appropriate de-identification, standards and frameworks are now being put forward by other healthcare stakeholders to clarify guidance on "de-identification" – laying out specific de-identification steps to be taken to ensure data cannot reasonably be re-identified, and to enhance adoption across the industry.

Exhibit 16: Alignment on Definitions of Non-Identified Data and Risk

Country	Referenced Documents	Non-Identified Data	Reducing Reidentification Risk
Australia	Australian Privacy Principles[100]	"De-identification involves removing or altering information that identifies an individual or is reasonably likely to do so."	"[d]e-identification can be effective in preventing re-identification of an individual, but may not remove that risk altogether." "[t]he risk of re-identification must be actively assessed and managed to mitigate this risk. This should occur both before an information asset is de-identified and after disclosure of a de-identified asset."
United States	HIPAA Privacy Rule[101]	"Health information is de-identified, or not individually identifiable, under the Privacy Rule, if it does not identify an individual and if the covered entity has no reasonable basis to believe that the information can be used to identify an individual."	"a covered entity may demonstrate that it has met the standard if a person with appropriate knowledge and experience …makes and documents a determination that there is a very small risk that the information could be used by others to identify a subject of the information."
United Kingdom	The Information Commissioner's code of practice[102]	"We use the term 'anonymised data' to refer to data that does not itself identify any individual and that is unlikely to allow any individual to be identified through its combination with other data."	The applicable privacy rules do "not require anonymisation to be completely risk free – you must be able to mitigate the risk of identification until it is remote." This Code describes specific approaches to identifying and analyzing "re-identification risk," and prescribes an overall approach that takes into account the overall context of the data, how it is being used and by whom
Canada	Privacy Commissioner Whitepaper: "De-Identification Protocols: Essential for Protecting Privacy"[103] referencing Personal Health Information Act, SNL 2008, c P-7.01, s 5(5).	"Information is de-identified if it does not identify an individual, and it is not reasonably foreseeable in the circumstances that the information could be used, either alone or with other information, to identify an individual."	"de-identification often involves the implementation of a robust re-identification risk management framework. An assessment must be conducted to identify the risks of re-identification in the particular circumstances involved, having regard to such factors as the motives and capacity of the organization or individual to re-identify the information."

All of these have consensus around the need for risk-based de-identification, including guidance from the Institute of Medicine (U.S.) which recently published a report on de-identification and the sharing of clinical trial information; the Council of Canadian Academies, which published on the sharing of health data that ties the degree of de-identification to the circumstances under which the data is made accessible for research, and by PhUSE (Pharmaceutical Users Software Exchange), a not-for-profit consortium in Europe that has put forward de-identification standards.[104-106]

But perhaps the strongest recent effort furthering a broadly applicable de-identification framework comes from the Health Information Trust Alliance (HITRUST), which focuses heavily on implementation, guidance, assessment and compliance management for multiple stakeholder types in connection with a consistent and repeatable de-identification approach. The HITRUST De-Identification Framework supports HIPAA's requirements but also offers organizations additional guidance beyond the regulatory provisions, putting forward twelve criteria to guide the establishment of a successful de-identification program.[107]

It provides details on administrative controls to govern de-identification and how the organization can actually arrive at a reasonable conclusion that data has been appropriately de-identified. By also offering training and certification programs, HITRUST holds new promise for gaining broad adoption of best-practice risk-based de-identification methods across the healthcare industry.

While these de-identification frameworks vary in certain details, the overall goal is consistent: to adopt standards that can reliably protect individual privacy through reducing risks of re-identification, while still preserving the important societal benefits that can be achieved through research and analysis of this data. Where these steps are followed, applying objective standards, appropriate security, and relevant contractual requirements, non-identified healthcare information can be used to benefit a wide variety of healthcare purposes, for public and private benefit, while preserving appropriate protections for individual privacy.

> "All of the attacks on data – general data and health data – were done on datasets that were not de-identified properly... they didn't follow any of the standards or best practices... so those datasets are easy targets. For those datasets that have been de-identified properly the ability to attack them and to re-identify individuals have been small or zero... As far as we know, no one has been successful in attacking this data, but you have to use best practices."
>
> *Khaled El Emam, CEO of Privacy Analytics*

Obtaining the most value from non-identified data

"Although the health-care industry has lagged behind sectors like retail and banking in the use of big data --partly because of concerns about patient confidentiality --it could soon catch up. First movers in the data sphere are already achieving positive results, which is prompting other stakeholders to take action, lest they be left behind." [108]

Basel Kayyali, David Knott, and Steve Van Kuiken, McKinsey & Company

For the healthcare system to improve patient care and become more efficient – to save lives and money – all stakeholders must join in the quest of maximizing big data's contributions to evidence based medicine. There has been growing recognition by all stakeholders that non-identified health data is fundamental to improving healthcare, as they witness the progress made to identify best treatments, protect patient safety and combat heath disparities and waste. However, to maximize the value and privacy benefits this data can provide, further progress will need to come through the following approaches (see Exhibit 17):

- Increasing the availability and accessibility of high-quality non-identified patient-level data sources
- Increasing use of non-identified patient-level data to conduct research
- Universally applying best-practice privacy and security standards
- Strengthening the impact of evidence through increased collaboration

Increased availability and accessibility of high-quality real-world data sources

Given their essential value to improving healthcare, it is important to continue to create non-identified health data sources and improve their completeness. Although many sources of non-identified patient-level data exist, the U.S. still is limited by data fragmentation, gaps in information, and the inability to see across the healthcare system as a whole. For research purposes, many stakeholders currently can still only obtain an incomplete view of the health system or patient experience.

"The big issue is that the data is scattered here and there. The challenge is going to be to combine the data to get a whole look at the medical system..."

David M. Cutler, Otto Eckstein Professor of Applied Economics, Harvard University

Exhibit 17: Approaches to maximize the value of big data in healthcare

Increasing the availability and accessibility of high-quality non-identified patient-level data sources
Standardization of data and interoperability to allow meaningful analyses
Data sharing and infrastructure projects to create larger or richer pools of data for research – i.e. view the health system in its entirety, or all aspects of a given disease
Continued processing of patient health data into non-identifiable forms for research by health information owners
Ensuring availability of unique and new types of health information to contribute to healthcare understanding
Increasing the use of non-identified patient-level data sources to conduct research
Broaden use of non-identified patient data for research where possible, including at facilities that use patient identifiable information for other purposes
Universally applying best-practice privacy and security standards
Acceptance of best-practice de-identification frameworks across the industry
Broad adoption of these frameworks by all stakeholders
Adherence to these best-practice frameworks to ensure patient privacy
Greater transparency and accountability by data processors about the frameworks they use and how they handle non-identified data
Strengthening the impact of evidence through increased collaboration
Connect healthcare stakeholders through data to speed consensus and remove barriers to action
Multi-stakeholder alignment (commercial/academic/government/provider) as partners contributing to medical, scientific and statistical healthcare research
Creative and tangible ways to ensure the output from this data has greater effects downstream on patient treatment

Data standardization

Obtaining an integrated view of a patient through data is the first step in this process of creating a connected health system. The movement of health records onto a digital platform, encouraged through the Health Information Technology for Economic and Clinical Health (HITECH) Act and the Medicare and Medicaid EHR Incentive Programs, has advanced the standardization of data and has helped to improve connected patient care.

> "We now have all 10 million people on the electronic health record... Our physicians have been able to study, through the data, the efficacy of care and to look at the practices that produce different outcomes. We've been able to aggregate data to better understand how to take care of, for example, different ethnic populations. We're more efficient and effective because now we're targeting to different diseases, to different populations, to different outcomes. That is allowing us to do things very differently." [109]
>
> *Bernard J. Tyson, Chairman and CEO of Kaiser Permanente*

However, the lack of standardized data formats among health systems and their EMRs has prevented full integration of patient health records such that they can contribute to an integrated view of the patient through data and enable connected healthcare. For research, the transformation and integration of disparate data sources into consistent, usable forms is a huge undertaking often riddled with issues of how data was originally collected at the point of service. The same terminology may still be used at different sites to mean different things, making future analyses challenging without data adjustments. Additionally, errors and gaps in information collected at the point of collection – sometimes unavoidable – may prevent a complete analysis from being conducted.

Improvements to standards and software are necessary to help this process and achieve interoperability. Already, data standardization efforts developed by the Nationwide Health Information Network (NwHIN) have been developed to enable the secure exchange of health information but it will be incumbent upon EMR providers and health systems to align to such standards, and make other improvements against this challenge, including finding ways for physicians – frequently strained for time – to enter data into EMRs more easily and completely.[110]

Data sharing and accessibility

Increased sharing of data is also critical to gaining further value from it for research. Our research suggests that the most significant gains in the future will be made from cross-site data infrastructure projects where data can be queried across multiple locations, with appropriate safeguards in place. The conversion of valuable repositories of personally identifiable patient healthcare data into non-identified forms by health information owners will be needed to make this data searchable across stakeholders for research.

Such community and cross-stakeholder data sharing projects require significant investment and effort to create these larger pools of research data, which have a goal to either obtain a more cohesive view of the health system, or a deeper understanding to help specific disease populations (like Patient-Powered Research Networks). Because this infrastructure is costly, adequate funding, awards and incentives from public and private sector sources may be necessary to help health information owners process

> "To accomplish the goal of optimizing the system and care, nationally merged datasets are strongly needed, since if you if you are only seeing only one part of a patient's experience (some part of time) then you cannot optimize."
>
> *David M. Cutler, Otto Eckstein Professor of Applied Economics, Harvard University*

health data into non-identifiable forms, build research systems or contribute their data to larger pools of research data – to make the data more widely available for beneficial use. Ideally in the future, such community or regional sharing will continue to be integrated into broader regional and even national repositories whose breadth will enable deeper research on rare diseases.

The non-identified data that already exists can also be put to better use, enabling researchers to gain a more comprehensive view of disease populations (i.e. for disease specific initiatives). For instance, government and commercial payers will be critical to making data more readily available for research outside of their institutions. The ability to see medical care across both public and private payers – to see the system as a whole – is critical and yet combining public (Medicare, Medicaid, VA, etc.) with private sector data has been difficult, and public sources including Medicaid and Veterans Administration (VA) data still remain unavailable for research at the national level.[79]

However, in the past few years, sharing data in order to improve the health system has been a high priority for the Centers for Medicare & Medicaid Services (CMS), which is in the process of making their data more publicly available. CMS has published all part D provider claims, which details information on the prescription drugs prescribed by each physician, and Part B medical payments within Medicare. Still, there remain gaps in the data picture. In addition, modifications to the Medicare Qualified Entity program in April 2015, through the passage of Medicare Access and CHIP Reauthorization Act of 2015 (MACRA), have cleared the way for a broader number of researchers to gain access to samples of Medicare data for research, and combine them with other sources including private sector claims.[111,112]

Patient-level Medicaid files are still unavailable, however, although MACRA set a path for the data to be made available, if deemed appropriate, beginning July 1, 2016. In line with progress made by CMS, more data sharing by other stakeholders is needed.

> "Governments are always underfunded in terms of internal resources to analyze data because that is the first thing you cut. So in countries with very strict privacy concerns that have led to not making data available for research they know much less about their healthcare system and they are in a much worse position to make positive change in healthcare."
>
> *David M. Cutler, Otto Eckstein Professor of Applied Economics, Harvard University*

New types of health information

As scientific knowledge advances, new types of health information may also become available that can contribute to healthcare understanding and improvement. While the value from imaging data, wearable mobile technologies and unstructured non-identified data have recently been become available, genomic sequence data (i.e. single nucleotide polymorphisms, exomes, whole sequences) has been left behind.

Due to the inherent identifiability of genomic sequence data, the value that can be derived from research on these sources is more limited, with any research currently requiring patient permission for use on an identified basis. Promising technologies such as secure computation that de-identify genomic information and other innovations will need to be promoted through competitions and prizes to make this and future advances available for research.[98]

> "Genomic data would be the next frontier. If we can figure out how to share genomic data, coming up with models and technologies to facilitate that would be the next big step for data sharing."
>
> *Khaled El Emam, CEO of Privacy Analytics*

Increasing the use of non-identified patient-level data sources for research purposes

To protect patient health information, the use of privacy-enhancing technologies like de-identification should continue to grow for research purposes, even at facilities that use identified data for other purposes. While certain types of research at these facilities may require patient identity or require that the original EMR records be directly assessed (such as research intended to impact the care of specific patients), this is not the case for many research purposes, and wherever appropriate, patient-level non-identified data rather than identifiable information should be used to protect patient privacy.

> "It would be our hope that covered entities, their business partners, and others would make greater use of de-identified health information than they do today, when it is sufficient for the research purpose. Such practice would reduce the confidentiality concerns that result from the use of individually identifiable health information for some of these purposes. The selective transfer of health information without identifiers into an analytic database would significantly reduce the potential for privacy violations while allowing broader access to information for analytic purposes, without the overhead of audit trails and IRB review." [84]
>
> *Dr. Louis W. Sullivan, Former U.S. Secretary of Health and Human Services, CEO and Chairman of The Sullivan Alliance, and President Emeritus of the Morehouse School of Medicine*

Universally applying best-practice privacy and security standards

There is growing acceptance that frameworks that take a risk-based approach to the process of de-identification and creation of non-identifiable data provide a win-win approach to maintain both the benefits of healthcare research and healthcare privacy. These paradigms do not ask society to sacrifice the benefits of non-identified health-evidence sources in pursuit of a myth of zero risk, and growing acceptance of these will yield the best balance for society, which should not be forced to choose between healthcare research and information privacy, when both can be achieved.

With consensus reached by healthcare stakeholders around data privacy frameworks that appropriately address the risk of re-identification, the healthcare industry will need to make efforts to ensure these standards – and the best practices embodied in these standards – become the norm and are broadly adopted and followed. To increase the use of non-identified data and assure the public that their privacy is ensured, organizations handling healthcare data for secondary research purposes should work to adopt a best-practice risk-based approach (if they have not already), gain certifications where available, and ensure the highest quality of data (with value retained) can be used for specific purposes and released for research while minimizing risk.

> "Science has continued to advance, as has the dialogue regarding the changing nature of research and the preferred balance of protections for research participants… Research that poses greater risk to subjects should receive more oversight and deliberation than less risky research…[with] the level of review more proportional to the seriousness of the harm or danger to be avoided." [10]
>
> *Federal Policy for the Protection of Human Subjects, Notice of Proposed Rulemaking*

In addition to adhering to one of the de-identification standards or frameworks that have come out, if data processors embrace these with greater transparency and accountability, being clear about their adherence to these standards/frameworks and how they handle non-identified data, this is likely to maintain and increase public confidence that their data is being used appropriately and their privacy is being protected. Such actions are likely to yield benefits to patient self-reporting of health information and increase willingness to share this and other sources of information. This, in line with public education awareness campaigns on the value of data use, could help the public to recognize the role of non-identified patient data to benefit themselves and their families, and the security measures protecting them.

"The public wants to know their data is being handled properly and the opportunities for inappropriate disclosure are being minimized. They want to trust that the organizations handling their data are doing this in a responsible manner…to trust that the right things are being done."

Khaled El Emam, CEO of Privacy Analytics

Strengthening the impact of evidence

Although non-identified data yield insightful findings that can benefit the health system, like with all medical advances, fully incorporating new information into clinical practice can take an exceedingly long time. Whether evidence comes from randomized controlled trials or real-world data, even clear research findings may be ignored or simply fail to be recognized and have downstream impact.

"A considerable lag frequently exists between advances in the health sciences and the incorporation of new knowledge, techniques, and treatments into physicians' practices. This lag contributes to increased costs in health care, as well as losses in productivity and quality of life and an increase in premature deaths. Research shows, however, that the collection and use of de-identified health information about physicians' practices can reduce this lag, benefit patients, and improve public health." [84]

Dr. Louis W. Sullivan, Former U.S. Secretary of Health and Human Services, CEO and Chairman of The Sullivan Alliance, and President Emeritus of the Morehouse School of Medicine

The use of non-identified data can help to speed this process by connecting healthcare stakeholders through data, thereby enabling research partnerships between commercial and academic actors so all may contribute to medical and scientific and statistical research. With a shared understanding of what is occurring in the health system, a first barrier to adoption is removed. However, additional downstream tools are needed to pull through medical findings.

Distribution of information through caregiver organizations or safety agencies is often not enough to change physician behavior. Policymakers, both public and private, must find creative and tangible ways to ensure the output from this data has greater effects downstream on patient treatment. All healthcare stakeholders should leverage the tools at their disposal to speed a change in practice in response to evidence derived from real-world non-identified cohort analyses or other similar high-quality sources. For policymakers and payers, this may mean altering incentives, including financial incentives, to align with new findings. For health systems, educational outreach programs to prescribers, prescribing audits, electronic prescribing alerts, and feedback on prescribing patterns can be helpful. Patient education can also play a role in ensuring downstream impact of new research, as can medication reviews to ensure appropriate use.[113] Ultimately, only if we measure the impact of these initiatives will we know if these initiatives are succeeding in closing the healthcare gap.

References

1. Charles D, Gabriel M, Searcy T. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2014 Quick Stat #47. 2015 Apr. Retrieved from: http://healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf
2. Hsiao CJ, Hing E. NCHS Data Brief No 143, Use and Characteristics of Electronic Health Record Systems Among Office-based Physician Practices: United States, 2001–2013. 2014 Jan. Retrieved from: http://www.cdc.gov/nchs/data/databriefs/db143.htm
3. Aitken M, Kleinrock M, Lyle J, Nass, D, Caskey L. Medicines Use and Spending Shifts: A Review of the Use of Medicines in the U.S. in 2014. IMS Institute for Healthcare Informatics. 2015 Apr.
4. Hartman M, Martin AB, Lassman D, Catlin A, the National Health Expenditure Accounts Team. National Health Spending In 2013: Growth Slows, Remains In Step With The Overall Economy. Health Affairs. 2015 Jan;34(1):150-160.
5. Schoen C, Radley D, Collins SR. State Trends in the Cost of Employer Health Insurance Coverage, 2003–2013. The Commonwealth Fund. 2015 Jan.
6. Aitken M, and Gorokhovich L, Advancing the Responsible Use of Medicines: Applying Levers for Change. IMS Institute for Healthcare Informatics. 2012 Sept.
7. Shah N, Pathak J. Why Health Care May Be Finally Ready for Big Data. Harvard Business Review. 20145 Dec 3. Available from: https://hbr.org/2014/12/why-health-care-may-finally-be-ready-for-big-data/
8. National Institutes of Health. NIH framework points the way forward for building national, large-scale research cohort, a key component of the President's Precision Medicine Initiative®. 2015 Sep 17. Available from: http://www.nih.gov/news/health/sep2015/od-17.htm
9. Health Care Cost Institute (HCCI). About. 2015. Retrieved from: http://www.healthcostinstitute.org/about
10. National Archives and Records Administration. Federal Register. 1985. Available from: http://www.gpo.gov/fdsys/pkg/FR-2015-09-08/pdf/2015-21756.pdf
11. Health Care Cost Institute (HCCI). Data Contributors. 2015. Retrieved from: http://www.healthcostinstitute.org/data-contributors
12. Patient-Centered Outcomes Research Institute. Clinical Data and Patient-Powered Research Networks: Awarded Projects. 2015 May 14. Retrieved from: http://www.pcori.org/research-results/pcornet-national-patient-centered-clinical-research-network/clinical-data-and-0
13. The National Patient-Centered Clinical Research Network. Clinical Data Research Networks. 2015. Retrieved from: http://www.pcornet.org/clinical-data-research-networks/
14. Love D, Custer W, Miller P. All-Payer Claims Databases: State Initiatives to Improve, Health Care Transparency. The Commonwealth Fund. 2010 Sep. Retrieved from: http://www.commonwealthfund.org/publications/issue-briefs/2010/sep/all-payer-claims-databases
15. The White House, Office of the Press Secretary. Executive Order -- Combating Antibiotic-Resistant Bacteria. 2014 Sep 18. Retrieved from: https://www.whitehouse.gov/the-press-office/2014/09/18/executive-order-combating-antibiotic-resistant-bacteria
16. O'Neill J. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. The Review on Antimicrobial Resistance. 2014 Dec. Retrieved from: http://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf
17. U.S. Department of Health and Human Services, CDC. Antibiotic Resistance Threats in the United States. 2013. Atlanta, GA; 2013. p. 7, 36-7
18. WebMD. Exclusive: Obama on Antibiotic Resistance. 2015 Mar 27. Retrieved from: http://blogs.webmd.com/breaking-news/2015/03/exclusive-obama-on-antibiotic-resistance.html
19. Hicks LA, Bartoces MG, Roberts RM, Suda KJ, Hunkler RJ, Taylor TH Jr, Schrag SJ. U.S. Outpatient Antibiotic Prescribing Variation According to Geography, Patient Population, and Provider Specialty in 2011. 2015 Mar 5. Retrieved from: http://cid.oxfordjournals.org/content/early/2015/03/05/cid.civ076.abstract
20. Thompson PL, Spyridis N, Sharland M, Gilbert RE, Saxena S, Long PF, Johnson AP, Wong ICK. Changes in clinical indications for community antibiotic prescribing for children in the UK from 1996 to 2006: will the new NICE prescribing guidance on upper respiratory tract infections just be ignored? Arch Dis Child. 2008 Dec 9. Retrieved from: http://adc.bmj.com/content/94/5/337.long
21. Public Health England. English Surveillance Programme For Antimicrobial Utilisation And Resistance (ESPAUR) Report 2014. 2014 Oct 10. Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/362374/ ESPAUR_Report_2014__3_.pdf
22. Hicks LA, Taylor TH Jr, Hunkler RJ. U.S. outpatient antibiotic prescribing, 2010. N Engl J Med. 2013 Apr 11;368(15):1461-2.
23. CBS News. Commonly Prescribed Drugs Could Raise Risk for Alzheimer's. 2014 Sept 10. Retrieved from: http://www.cbsnews.com/news/anti-anxiety-drugs-could-raise-risk-for-alzheimers/
24. Hampton LM, Daubresse M, Chang H, Alexander G, Budnitz DS. Emergency Department Visits by Adults for Psychiatric Medication Adverse Events. JAMA Psychiatry. 2014;71(9):1006-1014. doi:10.1001/jamapsychiatry.2014.436.
25. NIH News in Health. Many Older People Take Anti-Anxiety Meds Despite Risks. 2015 Feb. Retrieved from: http://newsinhealth.nih.gov/issue/feb2015/Capsule1
26. Olfson M, King M, Schoenbaum M. Benzodiazepine Use in the United States. JAMA Psychiatry. 2015 Feb 1;72(2):136-142.
27. Cook J, Biyanova T, Masci C, Coyne J. Older Patient Perspectives On Long-Term Anxiolytic Benzodiazepine Use And Discontinuation: A Qualitative Study. J Gen Intern Med. 2007. 22(8):1094-1100.
28. Pimlott NJG, Hux JE, Wilson LM, Kahan M, Li C, Rosser WW. Educating physicians to reduce benzodiazepine use by elderly patients: a randomized controlled trial. CMAJ: Canadian Medical Association Journal. 2003;168(7):835-839.
29. National Institutes of Health. Benzodiazepine Often Used in Older People Despite Risks. U.S. Dept of Health & Human Services. 2015 Jan 12. Retrieved from: http://www.nih.gov/researchmatters/january2015/01122015benzodiazepine.htm
30. National Institutes of Health. Despite risks, benzodiazepine use highest in older people. U.S. Dept of Health & Human Services. 2014 Dec 17. Retrieved from: http://www.nih.gov/news/health/dec2014/nimh-17.htm

31. Mad in America. Government Calls U.S. Benzodiazepine Prescription Levels "Worrisome". 2014 Dec 30. Retrieved from: http://www.madinamerica.com/2014/12/government-calls-benzodiazepine-prescription-levels-worrisome/
32. Olfson M. Benzodiazepine use most common among older adults. JAMA Psychiatry. 2014 Dec 18. Retrieved from: http://www.healio.com/psychiatry/anxiety/news/online/%7B737d68af-e91e-4077-b25f-43e586682592%7D/benzodiazepine-use-most-common-among-older-adults
33. U.S. Food and Drug Administration. Characterization of New Drug's Safety Profile Before Marketing. 2014 06. Retrieved from: http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm110632.htm#ADR Reporting
34. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. Journal of the American Medical Association. Apr 15, 1998; 279: 1200 - 1205.
35. U.S. Food and Drug Administration. ADRs: Prevalence and Incidence. 2014 06. Retrieved from: http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DevelopmentResources/DrugInteractionsLabeling/ucm110632.htm#ADR Reporting
36. Schroeder C, Keja J, Hughes B, Ehlken B, Toussi M. Understanding Patterns of Drug Utilization Studies (DUS) Requested by the European Medicines Agency (EMA). [Oral Presentation] 30th ICPE Conference, Taipei, Taiwan, Oct 24-27, 2014.
37. U.S. Food and Drug Administration. Center for Drug Evaluation and Research. The Sentinel Initiative. 2010 Jul. Retrieved from: http://www.fda.gov/downloads/Safety/FDAsSentinelInitiative/UCM233360.pdf
38. Health Policy Brief. The FDA's Sentinel Initiative. Health Affairs. 2015 Jun 4. Retrieved from: http://www.healthaffairs.org/healthpolicybriefs/brief.php?brief_id=139
39. HIPAA and Common Rule Compliance in the Mini-Sentinel Pilot Authored by the Mini-Sentinel Privacy Panel: Kristen Rosati, Barbara Evans and Deven McGraw 1
40. Engelberg Center for Health Care Reform at Brookings. Sentinel Initiative Public Workshop. 2015 Feb 5. Retrieved from: http://www.brookings.edu/~/media/events/2015/02/05-fda-sentinel-initiative-workshop/2015-sentinel-initiative-annual-meeting-slide-deck.pdf
41. Platt R, Carnahan RM, Brown JS, Chrischilles E, Curtis LH, Hennessy S, Nelson JC, Racoosin JA, Robb M, Schneeweiss S, Toh S, Weiner MG. The U.S. Food and Drug Administration's Mini-Sentinel program: status and direction. 2012 Jan 19. Retrieved from: http://onlinelibrary.wiley.com/enhanced/doi/10.1002/pds.2343
42. Southworth MR, Reichman ME, Unger EF. Dabigatran and Postmarketing Reports of Bleeding. New England Journal of Medicine. 368;14 April 4, 2013 pg 1272-4.
43. Mini-Sentinel Coordinating Center. Mini-Sentinel Website. All Reports. Gardasil Vaccination and Venous Thromboembolism (PRISM). 2015 Apr 23. Retrieved from: http://www.mini-sentinel.org/Reports/
44. Mini-Sentinel Coordinating Center. Mini-Sentinel Website. All Reports. Clopidogrel and prasugrel use. 2013 Jan 11. Retrieved from: http://www.mini-sentinel.org/Reports/
45. Harvard Pilgrim Health Care. FDA awards $150 million Sentinel contract to Harvard Pilgrim Health Care Institute [Press Release]. 2014 Oct 1.
46. NSDUH: Results from the 2011 National Survey on Drug Use and Health: Summary of National Findings. 2012 Jun 25. Retrieved from: http://www.samhsa.gov/data/NSDUH/2k11Results/NSDUHresults2011.htm
47. Substance Abuse and Mental Health Services Administration, Emergency Department Data / DAWN. The Dawn Report: Highlights of the 2011 Drug Abuse Warning Network (DAWN) Findings on Drug-Related Emergency Department Visits. 2013 Feb 22. Retrieved from: http://www.samhsa.gov/data/sites/default/files/DAWN127/DAWN127/sr127-DAWN-highlights.pdf
48. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. The NSDUH Report: Substance Use and Mental Health Estimates from the 2013 National Survey on Drug Use and Health: Overview of Findings. 2014 Sep 4. Retrieved from: http://www.samhsa.gov/data/sites/default/files/NSDUH-SR200-RecoveryMonth-2014/NSDUH-SR200-RecoveryMonth-2014.htm
49. Schwarz A. Painkillers Resist Abuse, but Experts Still Worry. The New York Times. 2015 Jun 6. Retrieved from: http://www.nytimes.com/2015/06/07/us/painkillers-resist-abuse-but-experts-still-worry.html?emc=eta1&_r=1
50. Cepeda MS, Fife D, Ma Q, Ryan PB. Comparison of the Risks of Opioid Abuse or Dependence Between Tapentadol and Oxycodone: Results From a Cohort Study. J Pain. 2013 Oct;14(10):1227-4.
51. Cepeda MS, Fife D, Kihm MA, Mastrogiovanni G, Yuan Y. Comparison of the risks of shopping behavior and opioid abuse between tapentadol and oxycodone and association of shopping behavior and opioid abuse. Clin J Pain. 2014 Dec;30(12):1051-6.
52. Cepeda MS, Fife D, Berwaerts J, Yuan Y, Mastrogiovanni G. Shopping Behavior for ADHD Drugs: Results of a Cohort Study in a Pharmacy Database. Drugs R D (2014) 14:205–211.
53. Cepeda MS, Fife D, Chow W, Mastrogiovanni G, Henderson SC. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the U.S. Drug Saf. 2012 Apr 1;35(4):325-34.
54. Cepeda, MS, Fife, D, Chow W, Mastrogiovanni G, Henderson, SC. Opioid Shopping Behavior: How Often, How Soon, Which Drugs, and What Payment Method. Journal of Clinical Pharma, 2013. 53: 112–117.
55. Cepeda MS, Fife D, Berwaerts J, Friedman A, Yuan Y, Mastrogiovanni G. Doctor shopping for medications used in the treatment of attention deficit hyperactivity disorder: shoppers often pay in cash and cross state lines. Am J Drug Alcohol Abuse. 2015 May;41(3):226-9.
56. Cepeda MS, Fife D, Yuan Y, Mastrogiovanni G. Distance traveled and frequency of interstate opioid dispensing in opioid shoppers and nonshoppers. J Pain. 2013 Oct;14(10):1158-61.
57. Cepeda, MS, Fife D, Berlin JA, Mastrogiovanni G, Yuan Y. Characteristics of prescribers whose patients shop for opioids: results from a cohort study. J Opioid Manag. 2012 Sep-Oct;8(5):285-91.
58. Birnbaum HG, White AG, Schiller M, Waldman T, Cleveland JM, Roland CL. Societal costs of prescription opioid abuse, dependence, and misuse in the United States. Pain Medicine. 2011; 12: 657-667.
59. Substance Abuse and Mental Health Services Administration, Emergency Department Data / DAWN. The Dawn Report: Emergency Department Visits Involving Attention Deficit/Hyperactivity Disorder Stimulant Medications. 2013 Jan 24. Retrieved from: http://www.samhsa.gov/data/sites/default/files/DAWN073/DAWN073/sr073-ADD-ADHD-medications.pdf
60. Jencks SF, Williams MV, Coleman E. Rehospitalizations Among Patients in the Medicare Fee-for-Service Program. N Engl J Med 2009;360:1418-1428.
61. The White House. The Annual Report of the Council of Economic Advisers. Transmitted to the Congress March 2014. Retrieved from: http://www.whitehouse.gov/administration/eop/cea/economic-report-of-the-President/2014

62. Godlee F. Milestones on the long road to knowledge. BMJ. 2007. 334 (Suppl 1):S2–S3. Retrieved from: http://ocw.tudelft.nl/fileadmin/ocw/courses/PublicHygieneandEpidemiology/res00025/!303420424d4a323030375f4d65646963616c4d-696c6573746f6e6573.pdf
63. Framingham Heart Study. History of the Framingham Heart Study. 2015. Retrieved from: https://www.framinghamheartstudy.org/about-fhs/history.php
64. The Nurses' Health Study. History. Retrieved from: http://www.channing.harvard.edu/nhs/?page_id=70
65. Long MD, Farraye FA, Okafor PN, Martin C, Sandler RS, Kappelman MD. Increased risk of pneumocystis jiroveci pneumonia among patients with inflammatory bowel disease. Inflamm Bowel Dis. 2013 Apr;19(5):1018–24.
66. Long MD, Martin C, Sandler RS, Kappelman MD. Increased risk of pneumonia among patients with inflammatory bowel disease. Am J Gastroenterol. 2013 Feb;108(2):240–8.
67. Long MD, Martin C, Sandler RS, Kappelman MD. Increased risk of herpes zoster among 108 604 patients with inflammatory bowel disease. Aliment Pharmacol Ther. 2013 Feb;37(4):420–9.
68. Long MD, Martin CF, Pipkin CA, Herfarth HH, Sandler RS, Kappelman MD. Risk of melanoma and nonmelanoma skin cancer among patients with inflammatory bowel disease. Gastroenterology. 2012 Aug;143(2):390–399.e1.
69. Bird ST, Hartzema AG, Brophy JM, Etminan M, Delaney JA. Risk of venous thromboembolism in women with polycystic ovary syndrome: a population-based matched cohort analysis. CMAJ. 2013 Feb 5;185(2):E115–20.
70. Sweet, B. Revised WellPoint HTA Guidelines for Formulary Submission: 2008 Standards and Recommendations. Presentation. Retrieved from: http://www.amcp.org/WorkArea/DownloadAsset.aspx?id=12235
71. Lin SJ, Hatoum HT, Buchner D, Cox D, Balu S. Impact of 5-HT3 receptor antagonists on chemotherapy-induced nausea and vomiting: a retrospective cohort study. BMC Health Serv Res. 2012 Jul 23;12:215.
72. Wise SJ, Nathoo NA, Etminan M, Mikelberg FS, Mancini GB. Statin use and risk for cataract: a nested case-control study of 2 populations in Canada and the United States. Can J Cardiol. 2014 Dec;30(12):1613–9.
73. Bird ST, Delaney JA, Brophy JM, Etminan M, Skeldon SC, Hartzema AG. Tamsulosin treatment for benign prostatic hyperplasia and risk of severe hypotension in men aged 40–85 years in the United States: risk window analyses using between and within patient methodology. BMJ. 2013 Nov 5;347:f6320.
74. Bird ST, Etminan M, Brophy JM, Hartzema AG, Delaney JA. Risk of acute kidney injury associated with the use of fluoroquinolones. CMAJ. 2013 Jul 9;185(10):E475–82. doi: 10.1503/cmaj.121730.
75. Yeaw J, Watson C, Fox KM, Schabert VF, Goodman S, Gandra SR. Treatment patterns following discontinuation of adalimumab, etanercept, and infliximab in a U.S. managed care sample. Adv Ther. 2014 Apr;31(4):410–25.
76. Slejko JF, Ho M, Anderson HD, Nair KV, Sullivan PW, Campbell JD. Adherence to statins in primary prevention: yearly adherence changes and outcomes. J Manag Care Pharm. 2014 Jan;20(1):51–7.
77. Jiang Y, Ni W, McGinnis JJ. Comparison of health services use associated with ziprasidone and olanzapine among schizophrenia and bipolar disorder patients in the USA. Clin Drug Investig. 2014 Jul;34(7):491–9.
78. Best JH, Hoogwerf BJ, Herman WH, Pelletier EM, Smith DB, Wenten M, Hussein MA. Risk of cardiovascular disease events in patients with type 2 diabetes prescribed the glucagon-like peptide 1 (GLP-1) receptor agonist exenatide twice daily or other glucose-lowering therapies: a retrospective analysis of the LifeLink database. Diabetes Care. 2011 Jan;34(1):90–5.
79. From an interview with David M. Cutler, Otto Eckstein Professor of Applied Economics, Harvard University. 2015 May 20.
80. Nieva VF, Murphy R, Ridley N, Donaldson N, Combes J, Mitchell P, Kovner C, Hoy E, Carpenter D. From Science to Service: A Framework for the Transfer of Patient Safety Research into Practice. 2005 Feb. Retrieved from: http://www.ncbi.nlm.nih.gov/books/NBK20521/
81. Healthcare IT News. Geisinger CEO gives tips for smarter BI. 2014 Mar 11. Retrieved from http://www.healthcareitnews.com/news/geisinger-ceo-gives-tips-smarter-bi
82. Knox CA, Delaney JA, Winterstein AG. Anti-diabetic drug utilization of pregnant diabetic women in us managed care. BMC Pregnancy Childbirth. 2014 Jan 17;14:28.
83. King M, Essick C. The geography of antidepressant, antipsychotic, and stimulant utilization in the United States. Health & Place. 2013; 20: 32–8.
84. Sullivan LW. On the Value of De-Identified Information in the Healthcare System. Paper presented at: The Future of Privacy Forum, Washington, DC. 2011. Retrieved from: http://www.thesullivanalliance.org/forms/speeches/4_Future-of-privacy-forum-speech_12-5-11.pdf.
85. King AB, Fiorentino DM. Medicare Payment Cuts for Osteoporosis Testing Reduced Use Despite Tests' Benefit In Reducing Fractures. Health Affairs. 2011 Dec. Retrieved from: http://content.healthaffairs.org/content/30/12/2362.full
86. Burge R, Dawson-Hughes B, Solomon DH, Wong JB, King A, Tosteson A. Incidence and Economic Burden of Osteoporosis-Related Fractures in, the United States, 2005–2025. JBMR. 2006 Dec 4.
87. King AB, Saag KG, Burge RT, Pisu M, Goel N. Fracture Reduction Affects Medicare Economics Impact of Increased Osteoporosis Diagnosis and Treatment. Osteoporos Int (2005) 16: 1545–1557.
88. King AB, Tosteson AN, Wong JB, Solomon DH, Burge RT, Dawson-Hughes B. Interstate Variation in the Burden of Fragility Fractures. J Bone Miner Res. 2009 Apr;24(4):681–92.
89. Post-Acute Care in the Medicare Program: Testimony Before House Committee on Ways and Means Subcommittee on Health. Friday June 14, 2013. Retrieved from: http://www.hhs.gov/asl/testify/2013/06/4481.html
90. Govern, P. Study links post-acute care costs with lower survival rates. Vanderbilt University. 2015 Apr 30. Retrieved from: http://news.vanderbilt.edu/2015/04/study-links-post-acute-care-costs-with-lower-survival-rates/
91. Munshi N, Mehra M, van de Velde H, Desai A, Potluri R, Vermeulen J. Use of a claims database to characterize and estimate the incidence rate for Castleman disease. Leuk Lymphoma. 2014 Sep 29:1–9.
92. Zalesak M, Francis K, Gedeon A, Gillis J, Hvidsten K, Kidder P, Li H, Martyn D, Orne L, Smith A, Kwong A. Current and future disease progression of the chronic HCV population in the United States. PLoS One. 2013 May 21;8(5):e63959.
93. Carlton R, Lunacsek O, Regan T, Carroll CA. Healthcare costs among patients with excessive sleepiness associated with obstructive sleep apnea, shift work disorder, or narcolepsy. Am Health Drug Benefits. 2014 Sep;7(6):334–40.
94. Du DT, Zhou EH, Goldsmith J, Nardinelli C, Hammad TA. Atomoxetine use during a period of FDA actions. Med Care. 2012 Nov;50(11):987–92.

95. IMS Health. Product Information. Clinical Trial Optimization Solutions. IMS Health SiteOptimizer. 2015. Retrieved from: http://www.imshealth.com/deployedfiles/imshealth/Global/Content/Information/Applications/CTOS%202014/IMSHealth_SiteOptimizer_f.pdf
96. Cavoukian A, Emam KE. De-identification Protocols: Essential for Protecting Privacy. Privacy by Design. 2014 Jun 25. Retrieved from: https://www.privacybydesign.ca/index.php/paper/de-identification-protocols-essential-protecting-privacy/
97. Garfinkel, "De-Identification of Personal Information," National Institute of Standards and Technology, NISTIR 8053 (October 2015), available at http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf.
98. From an interview with Khaled El Emam, CEO of Privacy Analytics. 2015 May 13.
99. US Department of Health and Human Services. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Retrieved from: http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html#_ednref1
100. Office of the Australian Information Commissioner. Privacy business resource 4: De-identification of data and information. 2014 Apr. Retrieved from: http://www.oaic.gov.au/images/documents/privacy/privacy-resources/privacy-business-resources/privacy_business_resource_4.pdf.
101. U.S. Department of Health and Human Services. Standards for Privacy of Individually Identifiable Health Information; Final Rule. 2002 Aug 14. Retrieved from: http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/privruletxt.txt
102. UK Information Commissioner's Office. Anonymisation: managing data protection risk code of practice. 2012 Nov. Retrieved from: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf
103. Sweeney L. Matching Known Patients to Health Records in Washington State Data. Harvard University, Data Privacy Lab. 2013 Jun. [White Paper] 1089-1.
104. Council of Canadian Academies. Accessing Health and Health-Related Data in Canada. 2014 Dec. Retrieved from: http://www.scienceadvice.ca/uploads/ENG/Assessments%20and%20Publications%20and%20News%20Releases/Health-data/HealthDataFullReportEn.pdf
105. Institute of Medicine (IOM). 2015. Sharing clinical trial data: Maximizing benefits, minimizing risk. Washington, DC: The National Academies Press. Appendix B: Concepts and Methods for De-identifying Clinical Trial Data. Retrieved from: http://www.nap.edu/catalog/18998/sharing-clinical-trial-data-maximizing-benefits-minimizing-risk
106. phuse. Data Transparency Material. 2015 May. Retrieved from: http://www.phuse.eu/Data_Transparency_download.aspx
107 Health Information Trust Alliance (HITRUST). HITRUST De-Identification Framework. 2015. Retrieved from: https://hitrustalliance.net/de-identification/.
108. Kayyali B, Knott D, Van Kuiken S. The big-data revolution in US health care: Accelerating value and innovation. McKinsey & Company. 2013 Apr. Available from: http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care
109. The Wall Street Journal. Noseworthy, Tyson on Balancing Care and Cost. 2015 Jun 22. Retrieved from: http://www.wsj.com/articles/noseworthy-tyson-on-balancing-care-and-cost-1434924122
110. The Standards and Interoperability (S&I) Framework. S&I Framework Explored. Retrieved from: http://www.siframework.org/framework.html
111. Library of Congress. H.R.4015 – SGR Repeal and Medicare Provider Payment Modernization Act of 2014. 113th Congress (2013-2014). Retrieved from: https://www.congress.gov/bill/113th-congress/house-bill/4015
112. Centers for Medicare and Medicaid Services. Qualified Entity Program. C2014 [updated 2015 Apr 13]. Retrieved from: http://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/QEMedicareData/index.html?redirect= /QEMedicareData
113. Gould RL, Coulson MC, Patel N, Highton-Williamson E, Howard RJ. Interventions for reducing benzodiazepine use in older people: meta-analysis of randomised controlled trials. Br J Psychiatry. 2014;204(2):98-107.

Authors

Murray Aitken
Executive Director, IMS Institute for Healthcare Informatics

Murray Aitken is Executive Director, IMS Institute for Healthcare Informatics, which provides policy setters and decision makers in the global health sector with objective insights into healthcare dynamics. He assumed this role in January 2011. Murray previously was Senior Vice President, Healthcare Insight, leading IMS Health's thought leadership initiatives worldwide. Before that, he served as Senior Vice President, Corporate Strategy, from 2004 to 2007. Murray joined

IMS Health in 2001 with responsibility for developing the company's consulting and services businesses. Prior to IMS Health, Murray had a 14-year career with McKinsey & Company, where he was a leader in the Pharmaceutical and Medical Products practice from 1997 to 2001. Murray writes and speaks regularly on the challenges facing the healthcare industry. He is editor of Health IQ, a publication focused on the value of information in advancing evidence-based healthcare, and also serves on the editorial advisory board of Pharmaceutical Executive. Murray holds a Master of Commerce degree from the University of Auckland in New Zealand, and received an M.B.A. degree with distinction from Harvard University.

Deanna Nass
Senior Researcher, IMS Institute for Healthcare Informatics

Deanna Nass is senior researcher and project manager at the IMS Institute for Healthcare Informatics leading the development of reports and performing analyses on biopharmaceutical and healthcare trends. Deanna joined the IMS Institute in 2013 with 16 years of experience in the Biopharma industry. She has worked at IMS Health since 2004, first as a Senior Consultant responsible for market assessments and subsequently as a Senior Account Manager responsible for business development. Prior to IMS Health, Deanna worked as a freelance market research consultant and writer of industry publications for Medical Data International, Clinical and Theta Reports. Deanna holds a B.A. in Biology from Yale University with a specialization in neurobiology and a Certificate in International Affairs from NYU.

About the Institute

The IMS Institute for Healthcare Informatics leverages collaborative relationships in the public and private sectors to strengthen the vital role of information in advancing healthcare globally. Its mission is to provide key policy setters and decision makers in the global health sector with unique and transformational insights into healthcare dynamics derived from granular analysis of information.

Fulfilling an essential need within healthcare, the Institute delivers objective, relevant insights and research that accelerate understanding and innovation critical to sound decision making and improved patient care. With access to IMS Health's extensive global data assets and analytics, the Institute works in tandem with a broad set of healthcare stakeholders, including government agencies, academic institutions, the life sciences industry and payers, to drive a research agenda dedicated to addressing today's healthcare challenges.

By collaborating on research of common interest, it builds on a long-standing and extensive tradition of using IMS Health information and expertise to support the advancement of evidence-based healthcare around the world.

Research agenda

The research agenda for the Institute centers on five areas considered vital to the advancement of healthcare globally:

The effective use of information by healthcare stakeholders globally to improve health outcomes, reduce costs and increase access to available treatments.

Optimizing the performance of medical care through better understanding of disease causes, treatment consequences and measures to improve quality and cost of healthcare delivered to patients.

Understanding the future global role for biopharmaceuticals, the dynamics that shape the market and implications for manufacturers, public and private payers, providers, patients, pharmacists and distributors.

Researching the role of innovation in health system products, processes and delivery systems, and the business and policy systems that drive innovation.

Informing and advancing the healthcare agendas in developing nations through information and analysis.

Guiding principles

The Institute operates from a set of Guiding Principles:

The advancement of healthcare globally is a vital, continuous process.

Timely, high-quality and relevant information is critical to sound healthcare decision making.

Insights gained from information and analysis should be made widely available to healthcare stakeholders.

Effective use of information is often complex, requiring unique knowledge and expertise.

The ongoing innovation and reform in all aspects of healthcare require a dynamic approach to understanding the entire healthcare system.

Personal health information is confidential and patient privacy must be protected.

The private sector has a valuable role to play in collaborating with the public sector related to the use of healthcare data.

IMS INSTITUTE FOR HEALTHCARE INFORMATICS

IMS Institute for Healthcare Informatics
100 IMS Drive, Parsippany, NJ 07054, USA
info@theimsinstitute.org
www.theimsinstitute.org

We invite you to download IMS Institute reports in iTunes