



CALIFORNIA HEALTHCARE FOUNDATION



In Plain Sight:
Is Open Data Improving
Our Health?

JANUARY 2015

Contents

About the Author

Cheryl Wold, MPH, is an independent consultant who provides strategic research and evaluation services to nonprofit and public-sector organizations. Previously, Cheryl was chief of the Health Assessment Unit in the Los Angeles County Public Health Department, where she directed the Los Angeles County Health Survey and developed and disseminated health indicators and data to guide health improvement activities.

About the Foundation

The California HealthCare Foundation (CHCF) is leading the way to better health care for all Californians, particularly those whose needs are not well served by the status quo. We work to ensure that people have access to the care they need, when they need it, at a price they can afford.

CHCF informs policymakers and industry leaders, invests in ideas and innovations, and connects with changemakers to create a more responsive, patient-centered health care system.

For more information, visit www.chcf.org.

©2015 California HealthCare Foundation

3 Introduction

Understanding the Impact of Open Data

5 Preventing Foodborne Illness in Chicago

Innovation

Early Results

6 Childhood Obesity Surveillance in New York

Innovation

Early Results

8 Reducing Pedestrian Injuries in San Francisco

Innovation

Early Results

9 Public Reporting of Hospital-Acquired Infections

Innovation and Early Results

California

New York

Going Forward

12 How Do We Evaluate Impact?

14 Acknowledgments

15 Endnotes

Introduction

Many in the health community believe that making data accessible to a broad swath of stakeholders — including health practitioners, technology entrepreneurs, researchers, journalists, and government staff and officials — will catalyze innovations to improve the US health care system.

This belief is affirmed by the enthusiasm at events showcasing innovative uses of open health data — the Health Data Consortium’s annual Health Datapalooza, California’s Open DataFest, and Illinois’s Putting Health Data to Work, among others.¹ These events attract thousands of participants who come to hear government officials discuss their work to make data more accessible, talk with experienced open data innovators about how to partner with entrepreneurs and civic-minded organizations, and observe vendors demonstrate tools for visualizing data or cloud services. Other types of events, such as code-a-thons and design challenges, take a more hands-on approach and encourage the public to contribute to the body of innovative uses of data. All of these events help create and support a network, or ecosystem, of data use and innovation.

In the spirit of being open with information, those putting open data to use are also eager to share what they’ve learned. The Open Data Handbooks from California (PDF) and New York State and the nonprofit Open Data Institute’s online guides on topics ranging from “How to

Open Health Data Initiatives

Initiatives to promote open data — the dissemination of data in machine-readable formats — aim to accelerate innovative uses of information. It worked with weather and global positioning system data, so why not health?

This series of case studies describes open health data initiatives from across the country and looks at the impact they are making on people’s health and well-being. They include a pedestrian safety effort in San Francisco; childhood obesity prevention in New York; and the public reporting of hospital-acquired infections.

These case studies were originally published online in installments from January through March 2015.

Make a Business Case for Open Data” to “The Open Data Consumer’s Checklist” are examples of a growing repository of resources to advance practices in the field.² Two of these groups have made their resources publicly available via GitHub, a web-based hosting service for software projects that is widely used by technologists but less well known among health data officials.

There has been a noticeable snowball effect. According to an article by Emily Shaw of the Sunlight Foundation, the more health data are released by governments, the more data-fueled applications benefiting health care’s “triple aim” (better quality, lower cost, and population health) are generated.³

For example, a ProPublica article by Charles Ornstein and Ryann Grochoswki Jones, “Top Billing: Meet the Docs Who Charge Medicare Top Dollar for Office Visits,” discusses how the Centers for Medicare & Medicaid Services’ release of hospital and physician payment data invited exploration into how to address the wide cost variation that was revealed.⁴

And an article by Erika G. Martin, Natalie Helbig, and Nirav R. Shah, “Liberating Data to Transform Health Care: New York’s Open Data Experience,” published in *JAMA* in 2014, referenced the opening of the clinical registry of cardiac surgery and subsequent efforts to reduce mortality and other poor outcomes in New York as compelling evidence in the case for opening access to data.⁵

Applications using government datasets also empower consumers with new ways to take a more active role in maintaining their own health. For example, people can now compare health care quality among physicians and nursing homes, find information about clinical trials and community services, manage health behaviors and medication, and learn about food safety and product recalls.⁶

Understanding the Impact of Open Data

But is opening the door to health data having an impact on health? Can we pinpoint where open health data initiatives are resulting in observable improvements in outcomes, policies, practices, or behaviors? How are such improvements being measured?

These questions were posed to local and state health officials engaged in open data efforts. The most concrete lessons they've learned so far focus on operational aspects of the work — for example, how best to motivate government agencies to release their data and how to involve others outside of government — and the most common outcome measures are narrowly focused and process related. Organizations are tracking the type and number of datasets released, number of downloads, and innovative applications using the data. In other words, while the field is doing a good job of documenting the uses of open data, it is still lacking an understanding of the long-term impact of open health data.

It will take time to measure impact from this relatively new field. Yet even at this stage, several common lessons emerged from the interviews with health officials, which together provide some early signals as to how open data might be assessed:

Most view their open data work as in its early stages.

Health officials would like to understand how data are being used and if those uses have resulted in specific health impacts, but few have studied those outcomes. Interviewees talked about the datasets that were being used for specific public health or health care improvement projects, but the work was either not yet far enough along to have generated outcomes or to have been evaluated systematically. Process outcomes, such as studies of the uptake and use of open data across various health datasets, weren't always collected or examined. All of those interviewed expressed great interest in hearing about what others were doing to track outcomes.

Agency buy-in takes time. All of those interviewed emphasized that liberating datasets at the state and local level required a lot of work that was sometimes impeded by bureaucratic processes, internal resistance, or simply a lack of resources. Several health officials were currently focused on securing support from different leaders within their own agencies for open data efforts. One official described the process as challenging, but critical: “[It takes] so much time just to make people comfortable with putting the data out.”

When interviewees were asked which datasets were easier to secure agency buy-in for release than others, a common theme emerged: Datasets not involving people, such as resource listings (e.g., farmers market or WIC program locations) and environmental health data, were

much easier to publish compared to datasets that involve personal information, even though these datasets would be aggregated, cleansed of any identifying data, or both.

External partnerships are important. Interviewees agreed that another key to success was the development of external partnerships, such as those formed with technical and nontechnical stakeholders like universities, private companies, and nonprofits. These partnerships sometimes occurred spontaneously at “hack nights” or civic coding events, and they also resulted from planned outreach, such as organized forums. Such exchanges often led to creative uses of the data and brought new expertise to the table.

Audiences' varying needs must be understood. Many of those interviewed emphasized the importance of learning how different audiences want to use the data and involving those audiences in the project development process. Since first releases are rarely perfect, the iterative process of preparing and publishing datasets offers multiple opportunities for input from a wide range of stakeholders.

For example, the general public may be interested in summary information, such as the name of the restaurant with a letter grade as an overall health score, along with a map or other visualization of the data. Researchers or data analysts, however, may want access to trend data and detailed information on the types of health violations. Open data initiatives have adapted over time to include steps that assure the appropriate use of the information through complete metadata and other documentation.

Potential internal efficiencies may also have costs.

Open data efforts were seen as positively impacting government services, making them more responsive to public needs, and creating valuable efficiencies — for example, by encouraging self-service through user-friendly access to influenza vaccination sites or restaurant inspection information. However, the effort required to make data available — converting, formatting, maintaining, and assuring its relevance to varied customers — should not be minimized. Improved information management, such as producing more comprehensive metadata, also requires an investment in resources for additional coordination and oversight, as well as translations of the provenance, meaning, and limitations of the information by experts.

In many cases, the self-service afforded by open data made it easier to respond to data requests. In a few cases, however, the effort to respond to related inquiries grew. For example, New York experienced an increase in Freedom of Information Law (FOIL) requests for radon exposure data after those data were published on the data portal. In other cases, the growth in demand necessitated refinements of the information, such as more or different groupings — this was seen as a positive development but also required additional staff time.

The following four case studies, representing diverse health issues and best practices from local governments across the nation, illustrate ways in which practitioners are assessing the potential impact of open health data on public health and health care delivery.

Preventing Foodborne Illness in Chicago

Restaurants rely on consumer support to stay in business.

But people generally do not want to eat at establishments with poor health ratings. That's why initiatives to release these ratings data have helped increase restaurant compliance with food safety regulations. This is one example of how improvements in food safety can be traced directly to open data initiatives.



The US Centers for Disease Control and Prevention (CDC) estimates that 48 million (one in six) Americans experience a food-related illness each year, resulting in 128,000 hospitalizations and 3,000 deaths.⁷ In response to calls for open access to Food Safety Inspection Service (FSIS) data, a National Academy of Sciences (NAS) review found big potential benefits and minimal risks to making establishment-specific FSIS data available, and concluded that “public releases could result in increased compliance with regulatory requirements.”⁸

At the local and state levels, food establishment inspection and grading, which are followed by investigation and control activities, are the primary tools for the prevention

of foodborne illness. Proactive restaurant inspection, the assignment of letter grades, and the public dissemination of those grades have led to greater consumer awareness, declines in the number of serious food safety violations, and decreases in foodborne illness.

A study by Paul A. Simon et al. on the “Impact of Restaurant Hygiene Grade Cards on Foodborne-Disease Hospitalizations in Los Angeles County,” published in the *Journal of Environmental Health* in 2005, found 13% fewer hospitalizations in the three years following Los Angeles County’s launch of a restaurant health inspection grading program.⁹

Another study by A. Blake Waters et al. on the “Impact of Internet Posting of Restaurant Inspection Scores on Critical Violations,” which was published in 2013 in the *Journal of Environmental Health*, found significant decreases in serious restaurant violations in Salt Lake City after posting health grades online.¹⁰

According to the US City Open Data Census, published by the Open Knowledge Foundation, as many as 30 cities publish food establishment inspection scores publicly, often in machine-readable formats.¹¹ The cities of New York and San Francisco have even partnered with Yelp to develop Local Inspector Value-Entry Specification (LIVES), an open data standard that enables restaurant inspection scores to be published in the consumer-focused application.¹²

Innovation

Chicago’s Food Inspections dataset has been publicly available on the city’s open data portal, and managed by the Chicago Department of Public Health (CDPH), since January 2010.¹³ This dataset was noted by the CDPH informatics manager as one of the most frequently updated health-related data sources on the portal. It is the centerpiece of the city’s initiative to prevent foodborne illnesses.

Chicago officials credit the success of this initiative to the pairing of strong technology with strong community partnerships.

Representatives from the Civic Consulting Alliance and Allstate Insurance volunteered their time to help the city’s public health department and Department of Innovation and Technology (DoIT) design a predictive model using

open datasets to prioritize food establishments for inspection. Using this model, city public health staff members can identify establishments that would be most likely to fail an inspection or to incur violations. Those establishments are targeted for visits, and ideally, code violations and related illnesses are prevented.

Chicago health officials and their partners are also using social media to increase the reporting of suspected food-related illnesses. For example, using the city's FoodBorne Chicago application, health officials are mining the tweets and online reviews of consumers who mention being sickened by food establishments and contact these consumers to file a report with the CDPH.¹⁴ Developed in partnership with the Smart Chicago Collaborative, this program's innovative use of social media has increased reports of food poisoning and identification of restaurants violating health codes, according to the report "Health Department Use of Social Media to Identify Foodborne Illness — Chicago, Illinois, 2013–2014," published in *Morbidity and Mortality Weekly Report*.¹⁵ Other cities, such as New York, are testing similar uses of consumer rating websites such as Yelp.¹⁶

Early Results

To assess their predictive modeling approach, officials in Chicago randomly assigned inspectors to two groups. One group used the standard inspections protocol. The other group followed the predictive modeling program. Compared to those using the traditional protocol, inspectors using predictive modeling identified 5% more critical violations and discovered them earlier in the inspection cycle, reducing the risk of potential public exposure to foodborne pathogens.

This model is spreading. Chicago's approach of combining open data with predictive analysis to develop more-effective inspection protocols is being used to address other health problems in the city, such as tobacco and lead poisoning prevention as well as rodent control. The model is also being considered for replication by other cities and by the CDC. Described by colleagues in other states as "light years ahead of other jurisdictions when it comes to leveraging open data," Chicago is setting an example for others to carry out similar work in this field.

Childhood Obesity Surveillance in New York

Childhood obesity is linked to a number of negative health outcomes. Obese children are more likely to have high cholesterol and high blood pressure, to develop bone and joint issues, and to suffer from social and psychological problems.



According to the Centers for Disease Control and Prevention (CDC), obese children are at risk of obesity as adults and of developing multiple chronic diseases such as heart disease, diabetes, and arthritis.¹⁷ These chronic conditions are responsible for 70% of deaths and for 75% of health care costs in the United States each year.¹⁸

Understanding the magnitude of the problem and who is affected by it is half the battle in disease prevention and control. The health sector has a long history of disease and risk factor surveillance work, which includes health behavior surveys as well as mandated reporting of certain diseases and conditions. The data collected by these programs are then made available to researchers, policy-makers, and other stakeholders for analysis and program development. Many of these public health surveillance programs have leveraged evolving information technology to make their data even more easily accessible.

For example, school districts are using student health data to design student wellness plans, and state health officials are using these data to monitor and design program and policy interventions to curb obesity.

Innovation

The state of New York is considered a health data pioneer. In 2013, it published the first state-level, open data site in the United States devoted solely to health.¹⁹ New York health officials made the data from its statewide Student Weight Status Category Reporting (SWSCR) system available on that portal as soon as they could.

The New York SWSCR system was launched in 2008 as part of the state's efforts to address childhood obesity.

Amendments to the New York State Education Law in 2007 required the inclusion of body mass index (BMI) and students' weight status as part of the student health certificate for selected grades — K, 2, 4, 7, and 10.²⁰ Since the implementation of this program, BMI data are reported for all of these students through a secure reporting system every two years for nearly all school districts in the state. This includes close to 600 school districts, excluding New York City, which has a separate surveillance and reporting program, and excluding all private schools. The New York State Department of Health (NYDOH) facilitates the analysis and dissemination of the data, a process that was established by amendments to the same law in 2007.

Before they used the online portal to release these data, New York health officials would produce a report on the data and share it as a PDF document. These static documents did not allow for easy access to the raw data or for creative analysis. The online portal did.

New York health officials identified several elements that were critical to the successful implementation of the SWSCR data portal:

- ▶ Using clean, well-documented data and metadata
- ▶ Producing data visualizations, such as maps and charts, to help explain the tabular data, which was also available online
- ▶ Helping people use the information

The New York open data portal includes a how-to video to guide people through the site; visitors browsing the SWSCR data are offered a link to that video to help them get the information they are seeking. The video is also frequently used by NYDOH staff members. The team found that users preferred the video over the live webinars, which were also offered as a source of guidance. As a result of the success of the general video guide, the program is planning to produce another video guide about use of the Student Weight Status Data.

Early Results

Making the SWSCR data available through the open data portal has resulted in more people using the data. As of June 2014, over 50,000 people have viewed the data since their release on the portal. In less than one year, the data have drawn more than five times the number

of views than the PDF data reports since the start of the surveillance program.

The open data platform enables users to download the data in standard formats for analysis. More than 5,000 users have used this download feature. Over 20% of the requests for SWSCR data are served automatically to developers and other applications via the application programming interface (API).

Using the CDC Framework for Evaluating Public Health Surveillance Systems, a widely accepted evaluation standard, NYDOH staff examined the impact of employing the open data portal on the program's surveillance activities.²¹ The program compared several outcomes, including the timeliness, extent of adoption, and usefulness of its SWSCR system, before and after the open data portal was used, noting several improvements. The time required to produce district-level reports had been cut in half, from two years to one. The time required to analyze and prepare the data was cut by one-third, from 15 months to 10 months, and the time required for approval of the data release was reduced by one-half, from 4 months to 2 months.

These results were shared with the Council of State and Territorial Epidemiologists in Tennessee on June 23, 2014, by Rachael A. Ruberto and Ian F. Brissette, in their presentation "Harnessing the Power of Open Data: Results from Dissemination of School District Student Obesity Data in New York State."²²

Compared to prior releases in PDF format, the electronic release of the data resulted in a greater number of media outlets using the data, allowed for the creation of customized data visualizations, and resulted in an increased number of stories published about the issue. Users took advantage of the data that were provided in a flexible rather than static format.

Several media outlets produced in-depth stories about an issue that the data uncovered — the glaring disparities in the prevalence of obesity by school district. Journalists were able to link these discrepancies to socioeconomic differences between school districts, using the data on rates of poverty and free- and reduced-price lunch eligibility.

Ian Brissette, director of the state's Bureau of Chronic Disease Evaluation and Research, talked about the

media attention: “The media grasped onto the data, and a lot of the stories that came out were pointing out the disparities in obesity. The media took our data and without prompting told the story. We, at the state health department, didn’t have to package it for them — they did it themselves and in compelling ways that spoke to local regions.”

Reducing Pedestrian Injuries in San Francisco

San Francisco is recognized as one of the most walkable cities in the country. Many of San Francisco’s 825,000 residents walk around the city for exercise and because parking is expensive. But pounding the pavement can also be dangerous.



According to the city’s Department of Public Health, San Francisco pedestrians are the victims of half of all motor vehicle-related fatal injuries, compared to the national rate of 13%. In 2013, Mayor Ed Lee outlined a San Francisco Pedestrian Strategy, which includes goals to reduce serious pedestrian injuries by 25% and fatal pedestrian injuries by 50% by 2016 and 2021, respectively.²³

Innovation

Given its history of using data to develop tools to inform policies and plans to create healthier environments, such as health impact assessments, the San Francisco Department of Public Health (SFDPH) emerged as a key partner in the mayor’s directive, and created a set of public data tools to help San Francisco transportation, health, and other officials implement changes to meet the city’s pedestrian safety goals.²⁴

SFDPH started with the motor vehicle traffic collision data that were available on the California Highway Patrol’s (CHP) website²⁵ and created an interactive map highlighting the city’s high-injury street corridors — the 6% of the city’s street miles on which 60% of severe and fatal pedestrian injuries occur.²⁶ The analysis for the interactive

map weighted the injury by severity and assigned the intersection totals to the nearest street segments, so that the street corridors contributing to the highest number of fatal and serious injuries could be identified.

San Francisco’s open data efforts didn’t stop there. Following the success of the interactive map, SFDPH leveraged the city’s available open datasets in an inter-agency data-sharing initiative. With funding from the CDC, the SFDPH integrated over 200 geospatially-referenced variables from multiple datasets, such as vehicle traffic and bus volumes from the transportation dataset, and age, sobriety, day of the week, and other details from the collisions dataset, to develop TransBASE, an integrated transportation geodatabase.²⁷

This tool gives users easy access to multiple data sources that are linked through standardized geographical references. For example, concerned neighbors could search for their residential block and find the number and types of pedestrian injuries that have occurred, the number of motor vehicle accidents, the location of nearby liquor outlets and health clinics, and other community data.

Early Results

These data products — the early mapping tool and TransBASE — have been central to ongoing policy development and planning in San Francisco. For example, pedestrian injury maps were used to raise awareness and citizen participation at community meetings held in collaboration with the San Francisco Police Department and other city departments. TransBASE was later used to analyze where injuries occur, what factors are involved, and who is most vulnerable. The data tools were also used to engage residents in “WalkFirst,” a multiagency, multi-stakeholder initiative coordinated by the San Francisco Planning Department to increase pedestrian safety in the city.

Megan Wier, MPH, senior epidemiologist and SFDPH lead on health, transportation, and equity, described the high-injury corridor data and TransBASE as “fundamental for the city’s understanding of where injuries are concentrated and how can we best target our resources — whether that’s engineering, education, or enforcement — to reduce injury.”

According to Wier, bringing this additional information to the other three lead government agencies

— transportation, planning, and the controller’s office — was a logical extension of their earlier data and mapping work: “We sat down with our transportation agency and said, ‘Look, we have this resource,’ and the response was, ‘That’s exactly what we’re looking for.’” The database was critical in the data analyses and creation of profiles of high-injury corridors that were necessary to develop engineering solutions and immediate and long-term recommendations for actions to reduce pedestrian injuries and fatalities.

San Francisco city agencies have committed \$17 million for improvements so far. Among these improvements, which are outlined on the WalkFirst pages of the San Francisco Planning Department’s website,²⁸ specific actions include:

- ▶ Upgrading 44 miles of streets where injuries are most concentrated, 5 miles per year through 2021
- ▶ Giving pedestrians extra crossing time at 800 intersections citywide, at least 160 annually
- ▶ Improving safety around schools and senior centers with high pedestrian injury
- ▶ Upgrading 13,000 curb ramps in the next 10 years
- ▶ Targeting police enforcement efforts on the city’s most hazardous corridors and intersections

In addition, the San Francisco mayor, County Board of Supervisors, and six city agencies recently adopted “Vision Zero,” an international framework for improving traffic safety.²⁹

Open datasets and collaboration on the part of government agencies have successfully informed policies and engineering improvements designed to create safer streets and reduce pedestrian injuries in San Francisco.

Public Reporting of Hospital-Acquired Infections

Mandating the public reporting of hospital-acquired infections was the first step. Learn how two states are starting to use these data to effect change.



Hospital-acquired infections (HAIs), as the name suggests, are infections acquired by a patient during a hospital visit, and are likely the most common complication of hospital care. These infections lead to extended hospital stays, increased costs for the patient and institution, and in some cases, death. According to the US Office of Disease Prevention and Health Promotion’s national action plan to prevent health care-associated infections, HAIs affect approximately 1 in every 20 inpatients, and the costs of HAIs are estimated to reach billions of dollars each year.³⁰

Because HAIs are, for the most part, preventable, high rates of HAIs are a marker of poor quality care. Preventing HAIs is a leading health priority for health care institutions and for the nation as organizations are searching for ways to reduce costs and to improve their quality ratings.

Specific interventions, such as hand washing, antiseptic procedures, and prompt removal and inspection of catheters, have proven to be effective at preventing HAIs, according to the Agency for Healthcare Research and Quality’s patient safety network.³¹ However, the implementation of these practices and rates of infection vary greatly by facility.

Public reporting of some of the most common hospital-acquired infections was mandated in some form in roughly two-thirds of all states between 2005 and 2009. New York, the first state to implement public reporting using the Centers for Disease Control and Prevention’s (CDC) National Healthcare Safety Network (NHSN), the most widely used HAI tracking system in the US, served as an early model for HAI prevention programs in other states, including California.³²

Although these two states differ significantly in the sizes of their population — for example, California is responsible for surveillance in nearly 420 hospitals versus New York’s approximately 180 hospitals — and in specific mandates and procedures, both programs have followed a similar trajectory: working to improve the completeness of the data reported by hospitals, such as the number and type of infections captured; reporting those infection rates publicly; and finally, using reported data to focus prevention efforts.

Innovation and Early Results

These public reporting programs are a form of positive peer pressure. The idea is that mandating hospital reporting and publicizing a given hospital’s HAI rates encourages actions in the public and private sector, as well as by consumers, to reduce those rates. It appears that public reporting, combined with ramped-up preventive practices and reimbursement policies, among other factors, is positively affecting rates of these preventable hospital infections. The NHSN’s progress report on health care-associated infections shows some notable improvements in the most dangerous and costly infections between 2008 and 2013 — a 46% reduction in central line–associated bloodstream infections, or CLABSI, occurred in aggregate for the nation and for many states reporting, including California and New York.³³

What is the potential for open data to accelerate further declines, especially in rates of other types of infections? For example, could providing hospital- and infection-specific rates in a standardized, machine-readable format, similar to restaurant inspection and other data, increase adoption of HAI prevention practices?

While no studies have been conducted to answer this question, a look at reporting practices by two large, state-based programs in California and New York suggests a strong case for open data to have a positive influence on this front.

California

The California Department of Public Health (CDPH) Healthcare-Associated Infections Program launched in 2010 following a legislative mandate in 2008. Program Chief Lynn Janssen said the program “works harder year by year to make the data as useful as possible to multiple audiences” to drive prevention. Currently, the CDPH

program provides data to the public in multiple ways to reach different audiences and to account for the complexity of the information from each of over 400 hospitals.

One way the program reports the hospital data to the public is through the creation of annual reports, which include over 85 tables and are in PDF format.³⁴ Another way that the data are shared is through an interactive map, which currently includes summary measures — symbols indicating whether the rate is higher, lower, or unchanged as compared to the state average.³⁵

According to CDPH, mandated reporting has facilitated important programmatic actions to prevent HAIs, including the establishment of the HAI program itself, improved reporting and surveillance by health care facilities, and the establishment of a liaison program to improve the adoption of recommended HAI prevention practices by health care facilities. “CDPH encourages hospitals to use the annual HAI public report to compare their infection incidence and assess areas for targeting local HAI prevention efforts,” said Corey Egel, a department spokesperson. They also invite consumers to use the report to make decisions about where to seek care.

This information is in demand: The program’s website received 17,000 visitors in 2010, growing to over 282,000 in 2013; 31% (88,000) of these visits in 2013 were directly to the HAI interactive map, designed for the lay user through a grant from the California HealthCare Foundation, while the remainder of visits were to the report designed for health professionals.

Janssen discussed the program’s impacts: “Over the past three years, reported data clearly show that overall, California hospitals are making progress in the prevention of HAIs. A recent state-by-state progress report (PDF) posted on the CDC’s website showed that when compared to 2008 national baselines, by 2013, California experienced a 48% decrease in CLABSIs, a 18% decrease in SSI [surgical site infection] following colon surgery, and a 28% decrease in SSI following abdominal hysterectomy surgery.”³⁶

While rates are improving for California hospitals in aggregate, mandated data provide insight into where to focus continued effort. For example, in response to rising *Clostridium difficile* (*C. difficile*) infections (CDI), which have increased in California, CDPH is launching community-based prevention initiatives in two counties

to reduce CDI among patients moving between long term care facilities and hospitals. (*C. difficile* is a bacterium that can cause serious and fatal intestinal disease, especially among the elderly and people with prolonged use of antibiotics. According to a study published in the *New England Journal of Medicine*, this bacterium was the cause of an estimated half a million infections in the US in 2011, mainly among people in hospitals and long term care facilities.)³⁷

The program intends to publish 2013 rates on California's new data portal, which launched in August 2014.³⁸ Publishing hospital-specific data on an open portal will enable multiple audiences to search, filter, and download HAI data in flexible formats.

New York

Similar to California's program, The New York State Department of Health (NY DOH) conducts surveillance to monitor HAI rates and to encourage the adoption of prevention and control practices. Also the authoritative expert for its state, NY DOH has been required by law to publish an annual report on HAI infections by hospital since 2007. In addition, NY DOH has conducted intensive audits of data quality since the onset of the program, as required by the state law.

NY DOH has also made it a priority to make these data available for public use by publishing an annual comprehensive summary report in PDF form on its website and on the state's open data portal every year.³⁹ The report Hospital Acquired Infections: New York State 2013 noted significant declines in the rates of various HAIs since 2007 and outlined the associated cost impacts.⁴⁰ For example, CLABSIs declined 52% overall in various intensive care unit (ICU) settings, with associated cost savings estimated to be between \$20 million and \$79 million since 2007. Surgical site infections (SSI) decreased by 23%, resulting in a cost savings of between \$10 million and \$28 million.

In addition to the annual summary report, New York has published hospital-specific data on HAIs on the state's open data portal, since 2012.⁴¹ Valerie Haley, PhD, director of the New York State HAI Reporting Program, pointed out the value of both types of dissemination: The data published on the portal provides both current and historical summary statistics whereas the full report provides in-depth interpretation of the current data, including

how they are collected, analyzed, and risk-adjusted, as well as guidance about how to use the data.

Haley reported that HAI rates by hospital had been downloaded over 2,800 times in the two-year period ending February 2015. It is difficult, however, to tell who is downloading the HAI data due to the nature of the open data portal — when a user downloads data, no identifying information is recorded.

However, there are other uses that can be tracked: A site-analytics function on the Health Data NY portal enables trending of published data on third-party websites. "Our ability to do this trending occurs when another website inserts a Health Data NY hyperlink within their article, application, or other content, which directs the reader back to the NYS DOH HAI data or a specific visualization, such as a map," explained Patricia Lynch, the project manager for Open Data New York. Using this feature, New York officials found that health care consumers, reporters, universities, policy analysts, and hospitals were among the various users of the HAI data.

Two notable examples of data use include:

- ▶ DocSpot is a consumer website that allows users to search for providers based on specific needs.⁴² The site integrates data from over 600 sources, including state medical boards, hospital and clinic physician directories, and reviews, into one unified interface.
- ▶ Crain's New York Business newsletter is a subscription service that focuses on various business sectors within the state, including health care.⁴³ A two-part series featured interactive graphics and extensive information about the rates and costs of hospital infections⁴⁴ and an in-depth look at the *C. difficile* infection rate at local hospitals.⁴⁵

These are just two examples of how the availability of open data on HAI supported digital media and applications by third parties that would not have been possible if the data were disseminated using traditional static means alone.

Going Forward

Asked whether there was resistance to publishing the rates, both Haley and Janssen agreed that there was not. While there may have been concerns when reporting programs first started, Janssen observed: "I think everyone

is over that now — the data are out there on CMS,” referring to data comparing over 4,000 hospitals published by the Centers for Medicare and Medicaid Services.⁴⁶

One potential source of concern for reporting hospitals is the appearance of increased rates of infection year to year. Such fluctuations can sometimes be attributed to more complete identification and reporting of infections within a given facility. Janssen said that most of California’s hospitals have worked hard to improve reporting, and noted that CDPH is completing a three-year validation effort that helps hospitals evaluate and improve the quality of their reported data.

Many challenges remain. While rates of the dangerous and costly infections that have been under surveillance are declining overall, the burden of the leading 13 causes of HAIs is large, impacting 4% of hospital inpatients in the US in 2011, according to a multistate survey published in the *New England Journal of Medicine*.⁴⁷ In addition, newer multidrug-resistant organisms are on the rise. State and federal agencies need to regularly evaluate the types of infections that should be monitored and adjust reporting requirements accordingly.

The lessons learned from these states and other sectors can be used to accelerate the use of HAI data to improve hospital conditions and ensure the health and safety of patients.

Publicly-reported data are being actively put to use for prevention purposes with few concerns about publishing format or disclosure of facility names. The investment of resources to prepare and document the information is considerable but well worth it, as the data have the potential to help generate huge health dividends.

How Do We Evaluate Impact?

Documenting health-related outcomes associated with open data is in the early stages of development. These approaches from outside the health field could be instructive to the open health data movement.

The majority of published articles focused on the potential economic impact of open government data. Efforts to quantify the economic value from government health data include, for example, a widely cited McKinsey

& Co. study, “Open Data: Unlocking Innovation and Performance with Liquid Information,” that estimated the economic value of machine-readable, accessible health data at upward of \$300 billion annually.⁴⁸

New York University’s GovLab is compiling a catalog of new business development that is based on open data, called the Open Data 500.⁴⁹ Findings from these studies may help inform future research on return on investment (ROI) related to open data initiatives.

A study from the Transparency Policy Project at the Kennedy School of Government, “Transit Transparency: Effective Disclosure Through Open Data,” examined the release of operational data by transit authorities in five US cities and the resulting impact on operations and customer service.⁵⁰ That paper outlined four major lessons, which may be applied to the health sector:

- 1. Identify the problem to be solved and whether the data available can help solve it.** The case studies featured were trying to solve specific health problems through data. For example, San Francisco focused on generating more specific, actionable data about pedestrian injuries to lower the occurrence of those injuries.
- 2. Prioritize data for which there is demand.** Transit, weather, and restaurant inspection data were cited as examples of demonstrated public demand. Those interviewed had different viewpoints about which datasets were most in demand. In some cases, the analytics for the demand were either not managed by the health department or simply not quantified.
- 3. Ask whether solving the problem requires a third party — an information intermediary such as a skilled software developer.** For most problems, health officials tended to rely on internal capacities for data management and analysis. In some cases, however, tapping external expertise and partners ultimately lead to more effective dissemination, as in the case of New York, and interventions in Chicago and San Francisco.
- 4. Make data available in a nonproprietary data standard.** The case studies feature data that could easily be made available in widely accepted and open-source formats, and published using data standards to ensure uniform meaning, such as those collected

through surveillance activities like hospital infections reporting, injury, and obesity.

The bottom line is demonstrating whether open data initiatives are helping health officials and practitioners do their jobs better. Is opening data netting positive results and health outcomes? Different evaluation approaches could be used to test the impact of open data initiatives. Some of those used in the case studies and elsewhere may be worth considering.

► **Experimental risk-based design.** Chicago used an experimental design to test whether new food establishment inspection protocols based on predictive modeling enabled by open data resulted in earlier intervention. The design also enabled researchers to quantify these efficiency gains. (Related descriptive research, which is described in a CDC *Morbidity and Mortality Weekly Report* article, “Health Department Use of Social Media to Identify Foodborne Illness — Chicago, Illinois, 2013–2014,” examines the potential contribution of social media such as Twitter.)⁵¹ The use by the Fire Department of the City of New York (FDNY) of a similar risk-based model to predict fires and assign building inspectors was described in a *Wall Street Journal* blog entry by Elizabeth Dwoskin, “How New York’s Fire Department Uses Data Mining.”⁵² And New York City is using eviction data to predict (and prevent) homelessness among families. Quantifying how open data may be leading to smarter health protection and enforcement systems could have many potential health impacts.

► **Studies examining events over time revealed changes in the frequency of serious food establishment violations before and after Internet publication of inspection scores in Salt Lake City, and in an analysis of food-related hospitalizations in Los Angeles County after posting of letter grades in restaurant windows.** (The Salt Lake City study, “Impact of Internet Posting of Restaurant Inspection Scores on Critical Violations,” was published in the *Journal of Environmental Health* in 2013; the LA study, “Impact of Restaurant Hygiene Grade Cards on Foodborne-Disease Hospitalizations in Los Angeles County,” was published in the same journal in 2005.)⁵³ Similar approaches could be taken to test the impact of public reporting on infection control and reporting practices by hospitals, for example. These studies need to account for important

temporal factors, such as changes in financial reimbursement policies, but can generate valuable hypotheses and contribute to practices.

► **CDC Framework for Evaluating Public Health Surveillance.** New York’s employment of a surveillance system evaluation approach using metrics (how timely, useful, and effective), as well as the analysis of media coverage, revealed impressive results and may have additional applications to other datasets.⁵⁴ As San Francisco demonstrated, open data not only improved the data about the location of serious and fatal pedestrian injuries but also facilitated the creation of a more data-driven, evidence-based approach to street improvements to reduce them. The targeting of improvements and resulting rates will be further evidence of impact.

Open data releases are a recent phenomenon — evidence takes time to accumulate, and evaluations are difficult and costly. Yet there are some promising examples that show the types of outcomes that may start to emerge from open health data efforts and support the conversations.

Acknowledgments

Author Cheryl Wold thanks the following people for their contributions to these case studies.

Preventing Foodborne Illness in Chicago

Jay Bhatt, DO, MPH, MPA, FACP, managing deputy commissioner and chief innovation officer, Chicago Department of Public Health

Bechara Choucair, MD, MS, former commissioner of health, Chicago Department of Public Health

Raed Mansour, MS, executive assistant to the commissioner, Chicago Department of Public Health

Matthew Roberts, MPH, informatics project manager, Chicago Department of Public Health

Tom Schenk Jr., MS, director of analytics and performance management, Department of Innovation and Technology, City of Chicago

Childhood Obesity Surveillance in New York

Ian Brissette, PhD, director, Division of Chronic Disease Research and Evaluation, New York State Department of Health

Patricia Lynch, PMP, project manager, Health Data New York, New York State Department of Health

Reducing Pedestrian Injuries in San Francisco

Megan Wier, MPH, senior epidemiologist and lead in Health, Transportation, and Equity for the Environmental Health Branch, San Francisco Department of Public Health

Devan Morris, data analyst and cartographer, San Francisco Department of Public Health

Public Reporting of Hospital-Acquired Infections

Corey Egel, spokesperson, Office of Public Affairs, California Department of Public Health

Valerie Haley, PhD, director, Hospital-Acquired Infection Reporting Program, New York State Department of Health

Lynn Janssen, MS, CIC, CPHQ, chief, Healthcare-Associated Infections Program, California Department of Public Health

Patricia Lynch, project manager, Health Data New York

General contributions

James Garrow, MPH, director of digital public health, City of Philadelphia

Michael Porter, PhD, director of data use and strategic analysis, New York City Department of Health and Mental Hygiene

Nirav Shah, MD, senior vice president and chief operating officer for clinical operations, Kaiser Permanente – Southern California

Endnotes

1. "Health Datapalooza," Health Data Consortium, healthdatapalooza.org; "California HHS Open DataFest II – March 2015," Stewards of Change, stewardsofchange.com; "Health Data Consortium Partners with State of Illinois to Convene Forum on 'Putting Health Data to Work'," Illinois Department of Public Health, November 8, 2013, www.idph.state.il.us.
2. Open Data Handbook, California Health and Human Service Agency (October 2014), www.chhs.ca.gov (PDF); "Open Data Handbook," New York State Open Data Initiative, ny.github.io/open-data-handbook.
3. Emily Shaw, "The Impact of Open: Keeping You Healthy," Sunlight Foundation, June 25, 2014, sunlightfoundation.com/blog.
4. Charles Ornstein and Ryann Grochowski Jones, "Top Billing: Meet the Docs Who Charge Medicare Top Dollar for Office Visits," ProPublica, May 15, 2014, www.propublica.org.
5. Erika G. Martin, Natalie Helbig, and Nirav R. Shah, "Liberating Data to Transform Health Care: New York's Open Data Experience," *JAMA* 311, no. 24 (June 25, 2014): 2481-82, jama.jamanetwork.com.
6. "Healthgrades," www.healthgrades.com; Charles Ornstein and Lena Groeger "Nursing Home Inspect: Find Nursing Home Problems in Your State," updated August 2014, projects.propublica.org; "TrialX.com for Patients," trialx.com; Erine Gray, "Aunt Bertha: About Us," www.auntbertha.com; "FLOR'NCE - The Digital Coach," mHealth Coach, accessed October 13, 2014, mhealthcoach.com; "iPharmacy," MedConnections, accessed October 13, 2014, medconnections.com; Taha Kass-Hout, "OpenFDA Provides Ready Access to Recall Data," US Food and Drug Administration, updated August 8, 2014, accessed October 13, 2014, open.fda.gov.
7. "Estimates of Foodborne Illness in the United States," US Centers for Disease Control and Prevention, www.cdc.gov.
8. The Potential Consequences of Public Release of Food Safety and Inspection Service Establishment-Specific Data (Washington, DC: National Academies Press, 2011), www.nap.edu.
9. Paul A. Simon et al., "Impact of Restaurant Hygiene Grade Cards on Foodborne-Disease Hospitalizations in Los Angeles County," *Journal of Environmental Health* 67, no. 7 (March 2005), kuafu.umd.edu (PDF).
10. A. Blake Waters et al., "Impact of Internet Posting of Restaurant Inspection Scores on Critical Violations," *Journal of Environmental Health* 75, no. 10 (June 2013), www.thefreelibrary.com.
11. "US City Open Data Census," Open Knowledge Foundation, us-city.census.okfn.org.
12. "Local Inspector Value-Entry Specification (LIVES)," Yelp, www.yelp.com/healthscores.
13. "Data Portal," City of Chicago, data.cityofchicago.org.
14. "Foodborne Chicago: Report Food Poisoning to the Chicago Department of Public Health," Smart Chicago Collaborative and Chicago Department of Public Health, www.foodbornechicago.org.
15. Jenine K. Harris et al., "Health Department Use of Social Media to Identify Foodborne Illness — Chicago, Illinois, 2013–2014," *Morbidity and Mortality Weekly Report* 63, no. 32 (August 15, 2014): 681-85, www.cdc.gov/mmwr.
16. Cassandra Harrison et al., "Using Online Reviews by Restaurant Patrons to Identify Unreported Cases of Foodborne Illness — New York City, 2012–2013," *Morbidity and Mortality Weekly Report* 63, no. 20 (May 23, 2014): 441-45, www.cdc.gov/mmwr.
17. "Childhood Obesity Facts," US Centers for Disease Control and Prevention, www.cdc.gov/healthyouth.
18. "Chronic Disease Publications," US Centers for Disease Control and Prevention, www.cdc.gov/chronicdisease.
19. "Open Data Health," New York State Department of Health, health.data.ny.gov.
20. "Student Weight Status Data," New York State Department of Health, www.health.ny.gov.
21. Robert R. German et al., "Updated Guidelines for Evaluating Public Health Surveillance Systems: Recommendations from the Guidelines Working Group," *Morbidity and Mortality Weekly Report: Recommendations and Reports* 50, no. RR-13 (July 27, 2001): 1-35, www.cdc.gov/mmwr.
22. Rachael A. Ruberto and Ian F. Brissette, "Harnessing the Power of Open Data: Results from Dissemination of School District Student Obesity Data in New York State," New York State Department of Health, cste.confex.com.
23. San Francisco Pedestrian Strategy, Mayor's Pedestrian Safety Task Force, www.sfmayor.org.
24. "Pedestrian Safety," San Francisco Department of Public Health, www.sfhealthequity.org.
25. "SWITRS — Internet Statewide Integrated Traffic Records System," California Highway Patrol, www.chp.ca.gov.
26. "Pedestrian-Vehicle Injuries and High Injury Corridors," San Francisco Department of Public Health, sfgov.maps.arcgis.com.
27. "TransBASE," San Francisco Department of Public Health, transbasesf.org.
28. "WalkFirst," San Francisco Planning Department, sf-planning.org.
29. "Vision Zero," Walk San Francisco, walksf.org.
30. "National Action Plan to Prevent Health Care-Associated Infections: Road Map to Elimination," US Office of Disease Prevention and Health Promotion, www.health.gov/hcq.

31. "Health Care-Associated Infections," Agency for Healthcare Research and Quality, www.psnet.ahrq.gov.
32. "National Healthcare Safety Network," US Centers for Disease Control and Prevention, www.cdc.gov/nhsn.
33. "Healthcare-Associated Infections (HAI) Progress Report," US Centers for Disease Control and Prevention, www.cdc.gov/hai.
34. "HAI Annual Reports : What Can We Do To Prevent HAI?," California Department of Public Health, www.cdph.ca.gov.
35. "My Hospital's Infections," California Department of Public Health, gis-apps.cdph.ca.gov.
36. "Healthcare Associated Infections: Progress," US Centers for Disease Control and Prevention, www.cdc.gov/hai (PDF).
37. Fernanda C. Lessa et al., "Burden of Clostridium difficile Infection in the United States," *New England Journal of Medicine* 372 (February 26, 2015): 825-34: www.nejm.org.
38. "CHHS Open Data Portal," California Health and Human Services, chhs.data.ca.gov.
39. "Open Data," New York State Office of Information Technology Services, data.ny.gov.
40. "Hospital-Acquired Infection (HAI) Rates in New York State Hospitals," New York State Department of Health, www.health.ny.gov.
41. "Open Data," NY State Office of IT Services.
42. "DocSpot," www.docspot.com.
43. "Crain's New York Business," www.crainsnewyork.com.
44. "Pulse Extra January 15, 2015," Crain's New York Business, www.crainsnewyork.com.
45. "Pulse Extra January 22, 2015," Crain's New York Business, www.crainsnewyork.com.
46. "Official Hospital Compare Data," Centers for Medicare & Medicaid Services, data.medicare.gov.
47. Shelley S. Magill et al., "Multistate Point-Prevalence Survey of Health Care-Associated Infections," *New England Journal of Medicine* 370 (March 27, 2014): 1198-1208, www.nejm.org.
48. James Manyika et al., "Open Data: Unlocking Innovation and Performance with Liquid Information," McKinsey & Co., October 2013, www.mckinsey.com.
49. "The OD500 Global Network," New York University, www.opendata500.com.
50. "Transit Transparency: Effective Disclosure Through Open Data," Kennedy School of Government, www.transparencypolicy.net.
51. Harris et al., "Health Department Use."
52. Elizabeth Dwoskin, "How New York's Fire Department Uses Data Mining," *Wall Street Journal* Digits blog, blogs.wsj.com.
53. Waters et al., "Impact of Internet Posting"; Simon et al., "Impact of Restaurant."
54. German et al., "Updated Guidelines."