

RiskScape: A Data Visualization and Aggregation Platform for Public Health Surveillance Using Routine Electronic Health Record Data

Noelle M. Cocoros, DSc, MPH, Chaim Kirby, MA, JD, Bob Zambarano, PhD, Aileen Ochoa, MPH, Karen Eberhardt, BS, Catherine Rocchio, SB, W. Sanouri Ursprung, PhD, Victoria M. Nielsen, MPH, Natalie Nguyen Durham, MPH, John T. Menchaca, BA, Mark Josephson, BS, Diana Erani, MBA, Ellen Hafer, MTS, MBA, Michelle Weiss, MPH, Brian Herrick, MD, Myfanwy Callahan, MD, MSPH, Thomas Isaac, MD, MBA, MPH, and Michael Klompas, MD, MPH

 See also Perlman, p. 180.

Automated analysis of electronic health record (EHR) data is a complementary tool for public health surveillance. Analyzing and presenting these data, however, demands new methods of data communication optimized to the detail, flexibility, and timeliness of EHR data.

RiskScape is an open-source, interactive, Web-based, user-friendly data aggregation and visualization platform for public health surveillance using EHR data. RiskScape displays near-real-time surveillance data and enables clinical practices and health departments to review, analyze, map, and trend aggregate data on chronic conditions and infectious diseases. Data presentations include heat maps of prevalence by zip code, time series with statistics for trends, and care cascades for conditions such as HIV and HCV. The platform's flexibility enables it to be modified to incorporate new conditions quickly—such as COVID-19.

The Massachusetts Department of Public Health (MDPH) uses RiskScape to monitor conditions of interest using data that are updated monthly from clinical practice groups that cover approximately 20% of the state population. RiskScape serves an essential role in demonstrating need and burden for MDPH's applications for funding, particularly through the identification of inequitably burdened populations. (*Am J Public Health*. 2021; 111:269–276. <https://doi.org/10.2105/AJPH.2020.305963>)

State and local health departments are responsible for monitoring the magnitude, trends, and patterns of infectious diseases, chronic conditions, and health behaviors over time and within various populations. The efficiency and timeliness of data available to public health agencies and processes for managing and interpreting these data, however, are variable. While notifiable diseases are often infectious and reported rapidly and electronically to health departments, data on nonnotifiable conditions such as asthma, obesity, and hypertension are more limited.

Public health agencies use systems such as the Behavioral Risk Factor Surveillance System (BRFSS), the National Health and Nutrition Examination Survey (NHANES), all-payer claims databases, hospital-based data sources, and electronic laboratory reporting for data on chronic disease and health behaviors. The BRFSS is a self-reported, telephone-based survey that provides important public health data but has relatively small sample sizes and delays of about 1 to 2 years between data collection and publication. The NHANES combines self-reported data with

physical examinations, including laboratory testing, but the sample size is also relatively small, so it does not provide state or local level results; it also involves a wait of 2 or more years before results are disseminated. Moreover, none of these major public health surveillance systems include user-friendly, interactive visualization tools as part of the system. By contrast, continual, automated analysis of electronic health record (EHR) data is emerging as a complementary tool for public health surveillance of infectious diseases, chronic conditions, and health behaviors. Novel and emerging

infections—such as COVID-19—require new, timely sources of data.

EHR-based surveillance has the promise of providing health departments with rich, timely, and clinically detailed data from large populations. Examples include New York City's Macroscopic System¹ and the Colorado Health Observation Regional Data Service network.² EHR-based surveillance can serve as the source for data visualization systems that allow public health practitioners to monitor and explore health indicators at the aggregate level. We describe in this article the Massachusetts Department of Public Health's (MDPH's) RiskScape platform, a Web-based interactive data portal for displaying and analyzing near-real-time surveillance data from EHR systems.

DEVELOPMENT AND EVOLUTION OF RISKScape

In 2006, the Department of Population Medicine at Harvard Medical School and Harvard Pilgrim Health Care Institute obtained funding from the Centers for Disease Control and Prevention, via their Centers of Excellence in Public Health Informatics Program, to develop an automated reporting platform for notifiable diseases using EHR data. Working closely with MDPH, we developed the Electronic medical record Support for Public health (ESP; <http://esphealth.org>) surveillance platform. ESP is an open-source software suite that clinical practices can populate with EHR data by using a common data model (i.e., a standard data structure with data elements to which all sites map their underlying data); ESP analyzes these data for notifiable diseases and chronic conditions, generating individual case reports for notifiable disease and aggregate summaries of nonnotifiable conditions to the state health depart-

ment.^{3,4} Selected Massachusetts practice groups use ESP for automated notifiable disease reporting. We have since added further functionality to ESP to enable MDPH to query ESP data for aggregate counts of notifiable and nonnotifiable conditions via a Web-based user interface, in a secure, transparent, and controlled fashion using a system called MDPHnet.^{5,6} MDPHnet data are also aggregated and deidentified to support the RiskScape data visualization platform.

RiskScape is a Web-based interactive data aggregation and visualization tool that allows users to generate timely, tailored, high-level summaries of specific health measures and conditions of interest on an in-care population. It enhances public health surveillance by enabling policymakers and public health managers to easily review data on numerous conditions of interest, both notifiable (e.g., chlamydia, HCV infection) and nonnotifiable (e.g., asthma, obesity, hypertension).

Because RiskScape draws on EHR data, it can provide data on denominators (i.e., patients in care during a specified period of time), care patterns, case counts, and estimates of various conditions' prevalence. Denominators are important because they allow one to calculate and compare rates of disease and care patterns rather than just counts. Users interested in chlamydia, for example, can evaluate testing and coinfection rates as well as disease prevalence, while users interested in hypertension can examine diagnosed hypertension and controlled hypertension in addition to total hypertension counts and prevalence rates. Users have the option to select among multiple outcomes; filter down to populations of interest; stratify by demographics, comorbidities, and certain treatments; and compare conditions between locations or across time.

By providing public health officials the capacity to rapidly and easily work with surveillance data, RiskScape makes it possible for users to explore their evolving hypotheses about disease distribution, disparities, and the impact of public health interventions in near real time.

RISKScape IN MASSACHUSETTS

In Massachusetts, RiskScape currently draws upon EHR data from 3 clinical practice groups. Atrius Health serves a population of about 720 000 individuals in eastern Massachusetts, the majority of whom have health insurance. Cambridge Health Alliance serves about 140 000 individuals and is a safety net provider for vulnerable populations in eastern Massachusetts including Cambridge and greater Boston. The Massachusetts League of Community Health Centers data include approximately 400 000 people at federally qualified community health centers throughout the state. Taken together, these clinical practice groups represent approximately 20% of the state population and include people of all age groups, races, and ethnicities. Participation by additional sites that provide care, particularly in the central and western parts of the state, is currently being considered. Of note, patients who seek care at multiple sites in the network are not currently linked or de-duplicated.

Because the data in RiskScape are from patients in care at participating sites, they are not a random sample and do not necessarily reflect the general population, though Massachusetts has a very high percentage of the population with health insurance, likely meaning generalizability is less of a concern compared with a state with low insurance coverage. However, we do not have

geographic coverage across the state. (This issue and other important considerations are discussed in the Limitations section later.)

We have previously compared estimates of various chronic conditions from the RiskScape source data to those from the Massachusetts BRFSS data and observed comparable estimates of disease prevalence, particularly at the state level; for small-area estimates we observed correlations by condition and locale after adjustment for MDPHnet versus census demographics but with some variability and outliers.⁶ Although comparing these 2 distinct systems has limitations, this analysis suggested that we have reasonable capacity to estimate some conditions on the local level but need to devote more attention to areas where coverage is currently lacking (i.e., the central and western parts of the state).

RiskScape utilizes an individual-level, deidentified data set that is automatically generated monthly by each participating site's ESP installation. The extract transmitted to RiskScape includes 1 row per patient in the practice and includes dichotomous (e.g., gender, type 2 diabetes status, influenza vaccination), categorical (e.g., age group, race, ethnicity, smoking status, body mass index grouping), and continuous (e.g., number of medical encounters in the last year, blood pressure, hemoglobin A1C) variables. Geographical data are based on each patient's most recent zip code of residence. The underlying data at each site are assessed approximately quarterly for data quality and consistency. We review patterns in patient visits, prescriptions, immunizations, and other measures to identify anomalies for detailed investigation and rectification. In addition, all of MDPHnet's key users, including MPDH epidemiologists, participating site representatives,

and those implementing and maintaining the system, meet regularly to share and discuss forthcoming updates to the system (and potential new sites).

Participating sites populate their ESP systems using standardized daily extracts from their EHRs that include structured data on all patient encounters from the preceding 24 hours. The extracts include demographics, diagnosis codes, prescriptions, laboratory tests (all are included in the extract, but we only map and clean the subset pertinent to the conditions we assess), vaccinations, and social history (e.g., tobacco use). ESP analyzes these data nightly to detect chronic conditions and notifiable diseases using custom algorithms designed to maximize sensitivity, positive predictive value, or both depending upon the condition.^{3,4,6-9} The algorithms integrate vital signs, laboratory tests, prescriptions, and diagnosis codes from both current and previous encounters to detect conditions of public health interest. For example, the prevalent hypertension algorithm evaluates diagnosis codes, blood pressure measures, and medication prescriptions to assess whether a person meets our definition of hypertension (2 or more elevated blood pressure readings within a year, diagnosis codes for hypertension, or normal blood pressure readings but prescribed an antihypertensive). Note that users with programming expertise can adapt ESP's existing algorithms or develop new algorithms to redefine existing conditions in new ways or identify different conditions to meet their specific needs.

ESP system data are stored on dedicated servers managed within sites' data centers per local policy and procedure. Access to the ESP servers is managed by site. All communication between ESP

and RiskScape, and between RiskScape and users, is encrypted in transit.

RiskScape does not maintain personal health information data, but the application and data are maintained on a dedicated server. The RiskScape database is configured for access from the application only. All remote access to the server and the RiskScape application is via whitelisted and authorized permission. Further information about ESP is available at <http://esphealth.org>, including technical details and links to download the algorithms used in Massachusetts.

USING RISKScape

Authorized users log into the RiskScape Web site to review estimates of disease and conditions. There are 4 ways of examining the data: heat maps of disease prevalence by zip code, bar graphs and pie charts to evaluate demographic and clinical characteristics, time series to evaluate changes over time, and continuum-of-care tabular reports to evaluate care cascades. The dashboard (Figure 1) allows the user to review and select a condition, specify the population of interest, and designate the favored analysis (e.g., heat map, demographic description). These capabilities are further described herein and shown in Figures A through C (available as supplements to the online version of this article at <http://www.ajph.org>).

To generate prevalence estimates, users first select a condition of interest. The conditions in RiskScape are defined by algorithms that have been developed and validated within the system: type 1 diabetes, type 2 diabetes, prediabetes, gestational diabetes, categories of body mass index, hypertension, smoking status, asthma, treated depression, influenza-like illness, Lyme disease,

RiskScape

MDPHNET



FIGURE 1— The Dashboard of the RiskScape User Interface

Notes. BMI = body mass index (defined as weight in kilograms divided by the square of height in meters); MDPH = Massachusetts Department of Public Health. The dashboard allows the user to review and select a condition, specify the population of interest, and designate the favored analysis (e.g., heat map, demographic description).

vaccination status for several vaccines (influenza, Tdap), chlamydia, gonorrhea, opioid prescription, benzodiazepine prescription, and cardiovascular risk score. Users can then select among various denominator options; in our RiskScape instance, these are predominantly outpatient or ambulatory encounters. The default option is “patients with ≥1 encounter in the past two years.”

Users have the option, however, to select the denominators’ minimum encounter count (≥ 1 encounter or ≥ 2 encounters), look-back period (past 1 year or past 2 years), and minimum number of lifetime encounters within the participating site. Clinical encounter counts for the purpose of estimating denominators (i.e., persons at risk) are defined broadly and include any

interaction in the EHR with at least 1 vital sign (i.e., blood pressure, height, weight, or temperature), diagnosis code, prescription, laboratory test, or immunization; multiple encounters on the same day are treated as a single encounter. The rationale for these different denominator options and their impact on disease prevalence estimates has been previously described.¹⁰

In the heat map capability (Figure A), we can review, for example, the relative prevalence of pediatric asthma, with each outlined area representing a zip code. The taupe zip codes are those with inadequate or no data included in the system (RiskScape will only provide data on disease prevalence in a zip code if there are data on at least 100 residents in the zip code). A user can click on a zip code and a pop-up window with the following information will display: the prevalence of the outcome in that zip code, the number of patients in the numerator and denominator, and RiskScape's coverage rate for the chosen zip code (i.e., number of people with the user's selected demographic characteristics in that zip code within RiskScape vs the count of people with those demographic characteristics in the zip code per the 2010 US Census [any zip code-based population estimates can be used]).

The bar charts and pie graphs that RiskScape can generate allow users to explore the demographic and clinical characteristics of patients with a chosen outcome. The bar graph in Figure B depicts the prevalence of obesity (defined as a body mass of ≥ 30 kg/m²) among adults aged 20 years or older while the pie chart shows the age distribution of people with obesity. Users can specify target towns and neighborhoods for analysis, compare 2 locations side by side, or compare disease prevalence in the chosen location to the state as a whole. Neighborhoods are currently only available for the City of Boston.

RiskScape can also generate time series and regression statistics to help users assess trends and changes over time. The denominator is calculated each month based on the number of patients who meet the user's chosen

denominator criteria (e.g., those with at least 1 encounter in the last 1 year; this automatically adjusts for temporal changes in the population of patients in care). Figure C shows the prevalence of hypertension among adults from January 2012 through July 2020, stratified by race. Users can select a "trend line summary" to receive statistics on a trend for a particular group based on generalized least squares regression. Users can specify an inflection point to assess for changes in disease prevalence and trends before versus after a specific point in time. This feature can be used to obtain a rapid sense of the impact of new programs or policy changes on processes of care (such as hemoglobin A1C testing or gonorrhea screening) or prevalence (such as gonorrhea cases).

An additional capability within RiskScape is a set of "continuum of care" summary reports for HIV, HIV risk, HCV, diabetes, and cardiovascular risk score. For these reports, users can select the clinical site of interest, the time period, age groups, gender, race, and ethnicity for the analysis. These reports provide users with data on the fraction of patients with key diagnoses who are retained in care, receive recommended processes of care, and success rates for disease control.

For HCV infection, RiskScape reports the number and percentage of individuals tested for HCV, the number among them who test positive, the number with an HCV viral load test, and whether the latest test had detectable virus. The number of individuals who have acute HCV are reported separately from those who have chronic HCV, as defined by internally validated algorithms. The report provides the number of HCV cases who have been treated, their recent viral load results, and the number of patients

with HCV who spontaneously cleared their infection without treatment.

For individuals with HIV, the care cascade starts with the number of patients with HIV and then reports the number and percentage of those with the following: an encounter after diagnosis, a prescription for HIV medications, being retained in care, a measured viral load, viral suppression, and diagnosis with an opportunistic infection. That same cascade is reported separately for those who are newly diagnosed with HIV during a specified time period.

There is also a care cascade designed to track uptake of HIV preexposure prophylaxis. ESP calculates an estimated risk of HIV acquisition in the forthcoming year for every person in the system using a validated EHR-based prediction rule.^{11,12} It then stratifies the population into high-, medium-, and low-risk categories and summarizes HIV testing rates, preexposure prophylaxis prescribing, and HIV acquisition per strata.

The diabetes continuum-of-care report starts with individuals with at least 1 clinical encounter in the specified year(s) of interest and then provides the number and percentage of those patients with a hemoglobin A1C test, those with diabetes, the number on treatment, and patients' outcomes by hemoglobin A1C strata.

Finally, we recently created a report to provide information on risk factors and preventive care for patients at risk for cardiovascular disease using the American College of Cardiology's Atherosclerotic Cardiovascular Disease risk score algorithm.¹³ This score is calculated for every member of the population aged 20 to 60 years, divides the population into strata of risk (low, medium, high, established cardiovascular disease), and then for each strata characterizes the fraction of the population screened and

treated for hypertension, diabetes, hypercholesterolemia, and smoking. This analysis provides a unique population-level perspective on risk for cardiovascular disease and where opportunities to improve preventive practices might lie.

WHO CAN USE RISKScape?

RiskScape in Massachusetts is accessible only to authorized members of MDPH and participating sites via logins and passwords. However, RiskScape source code is open source and freely available to developers under a 3-clause Berkeley Source Distribution license. Source code is available from <http://esphhealth.org>.

In Massachusetts, clinical practice groups' participation in RiskScape and the underlying MDPHnet system is voluntary. Staff from each of the participating sites are informed of new capabilities added to the system and weigh in on prioritization and development of the platform. Stakeholders from MDPH, participating sites, the informatics developer (Commonwealth Informatics Inc), and the coordinating center (Harvard Pilgrim Health Care Institute) have biweekly conference calls to discuss updates, address any technical issues, and confer on plans. Within Massachusetts, users are trained and provided with background information on RiskScape and the underlying ESP system. Documentation is embedded in the platform, including algorithm definitions and major data interpretation issues. Data that can be queried via MDPHnet could be made available to external researchers, with permission and appropriate institutional review board oversight, but this has not occurred. To date, any research conducted using data from the underlying system has been limited to MDPHnet collaborators.

LIMITATIONS

Data from EHR systems must be interpreted appropriately, with understanding of the limitations inherent to the data type. The population is people in care and may not be representative of the general population, and diagnoses may be recorded that are differential or suspect only. The prevalence estimates generated by RiskScape must be interpreted with the same caution as with any data leveraged from clinical databases developed for clinical care or billing rather than for public health surveillance. The accuracy and completeness of EHR data vary, and disease detection frequency of a system like RiskScape is only as complete as the underlying source EHR data. Variations in the frequency of patients seeking care; differences between clinicians and practices in testing, diagnosing, and treatment practices; variations and changes in the completeness and accuracy of coding; and the total amount of time an individual has been affiliated with a given site are challenges inherent to the use of EHR data for surveillance. The data in RiskScape may be incomplete for individuals who divide their care between clinical sites contributing to RiskScape and other health care institutions outside of the system. Patients who seek care at multiple sites in the network are not currently de-duplicated, potentially leading to inflation of numerators, denominators, or both depending on the query. The major limitations of the system are documented within RiskScape and are actively discussed with MDPH users to facilitate their interpretation of data drawn from the platform.

It is technically feasible to link data from MDPHnet with data from other sources such as vital statistics, disease registries, claims databases, and other

EHR repositories and then enable RiskScape to display data integrated across multiple sources, but such work has not yet been undertaken. Governance issues as well as the technical and logistical aspects of that work have been discussed with MDPH and linkage with other sources may be pursued at some later time.

While RiskScape does not currently provide an option to generate prevalence adjusted by age or other demographics that could account for differences between clinical sites' patient populations and the Massachusetts census data, we have found that crude disease prevalences tend to be very similar to those adjusted for age, race/ethnicity, and gender, particularly at the state level. This is presumably a reflection of the size of the RiskScape population as well as the diversity of the contributing practices in Massachusetts.⁶

IMPLICATIONS

RiskScape enables epidemiologists, other public health professionals, and site staff focused on population health to quickly examine patterns and trends in various conditions or measures of interest. The ability to generate estimates of chronic disease and other non-notifiable conditions or measures on a monthly basis, stratified by site, allows users to follow trends in disease prevalence and care patterns, with increased frequency and timeliness relative to most existing public health surveillance systems for chronic conditions.

At this time, sites can review their own data individually and compare their data with data from other sites, enabling them, for example, to develop community needs assessments as well as to better understand health status, needs, and opportunities of the populations in their catchment areas. The demographic and

geographic stratifications provide insight into the epidemiology of conditions and measures that are hard to obtain elsewhere. For example, patterns or trends in health disparities are difficult to find elsewhere because of lack of data or incomplete data on race/ethnicity in other systems. While race and ethnicity data are not complete in RiskScape, they are more complete than in other data sources routinely used for public health surveillance (e.g., notifiable disease case report forms or electronic laboratory data), and the system is larger and more timely than other routine surveillance systems (e.g., BRFS).

The aggregate nature of the system means we can examine data on measures not otherwise available to MDPH. For example, MDPH does not have access to data on the number of people tested for HIV outside of sites that they fund. RiskScape's continuum-of-care reports allow MDPH to see patterns of care and prevention for a general patient population across numerous types of clinical sites. In addition, it can be readily adapted for new conditions, making otherwise inaccessible or hard-to-access data available to public health agencies. For example, we have developed pilot definitions for COVID-19 laboratory-based and syndromic surveillance criteria via ESP.

Over time, RiskScape has become an increasingly important tool in MDPH's planning and evaluation of chronic disease efforts. Examples of its use include identifying local hot spots of chronic disease and affected populations for targeted intervention, exploring population-level prevalence of risk factors for chronic disease to inform program design, and evaluating program impact, especially for state-wide infrastructure grants. In addition, RiskScape serves an essential role in

demonstrating need and burden for MDPH's applications for funding, particularly through the identification of inequitably burdened populations. As such, RiskScape has become an indispensable tool to support data-driven public health practice. That being said, there are numerous considerations for a jurisdiction or entity to plan for when preparing to implement a system like RiskScape. Governance, initial and ongoing funding, maintenance (e.g., monitoring of data quality), and expansion (e.g., creation and incorporation of new conditions) of the system are some of the major issues. It is also imperative for each stakeholder to fully understand what their participation includes. RiskScape is currently being adapted and implemented by multiple jurisdictions outside of Massachusetts under the umbrella of the National Association of Chronic Disease Directors' Multistate EHR-based Network for Disease Surveillance (<http://chronicdisease.org/page/MENDSInfo>).

In conclusion, RiskScape quickly and easily enables users to identify novel patterns and trends, get a rapid sense of the impact of new interventions, inform the design of program evaluations, provide data for new funding applications, generate hypotheses, and help plan for future analyses. *AJPH*

ABOUT THE AUTHORS

Noelle M. Cocoros, Aileen Ochoa, John T. Menchaca, and Michael Klompas are with the Department of Population Medicine at Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA. Chaim Kirby, Bob Zambarano, Karen Eberhardt, and Catherine Rocchio are with Commonwealth Informatics, Waltham, MA. W. Sanouri Ursprung, Victoria M. Nielsen, and Natalie Nguyen Durham are with Massachusetts Department of Public Health, Boston. Mark Josephson, Diana Erani, and Ellen Hafer are with Massachusetts League of Community Health Centers, Boston. Michelle Weiss and Brian Herrick are with Cambridge Health Alliance, Cambridge, MA. Myfanwy Callahan and Thomas Isaac are with Atrius Health, Boston.

CORRESPONDENCE

Correspondences should be sent to Noelle M. Cocoros, 401 Park Drive, Suite 401 E, Boston, MA 02215 (e-mail: noelle_cocoros@harvardpilgrim.org). Reprints can be ordered at <http://www.ajph.org> by clicking the "Reprints" link.

PUBLICATION INFORMATION

Full Citation: Cocoros NM, Kirby C, Zambarano B, et al. RiskScape: a data visualization and aggregation platform for public health surveillance using routine electronic health record data. *Am J Public Health*. 2021;111(2):269–276.

Acceptance Date: September 2, 2020.

DOI: <https://doi.org/10.2105/AJPH.2020.305963>

CONTRIBUTORS

N. M. Cocoros, C. Kirby, B. Zambarano, K. Eberhardt, C. Rocchio, J. T. Menchaca, and M. Klompas designed the system described. M. Josephson, D. Erani, E. Hafer, M. Weiss, B. Herrick, M. Callahan, and T. Isaac contributed data and information to the system. N. M. Cocoros and M. Klompas initiated the article. N. M. Cocoros drafted, wrote, and revised the article. All authors reviewed and provided feedback on the article.

ACKNOWLEDGMENTS

This work was funded in part by the Massachusetts Department of Public Health.

CONFLICTS OF INTEREST

The authors have no conflicts to disclose.

HUMAN PARTICIPANT PROTECTION

Institutional review board (IRB) review has not been required for the installation and implementation of RiskScape for the Massachusetts Department of Public Health given that the system is used for public health surveillance using aggregated data along with small cell suppression rules. Work on the system conducted by the Harvard Pilgrim Health Care Institute as the coordinating center, however, has been reviewed and approved by the Harvard Pilgrim Health Care IRB, per Harvard Pilgrim's internal policies. Practices participate on behalf of their patients; patient consent is not required.

REFERENCES

1. Newton-Dame R, McVeigh KH, Schreibstein L, et al. Design of the New York City MacroScope: innovations in population health surveillance using electronic health records. *EGEMS (Wash DC)*. 2016; 4(1):1265. <https://doi.org/10.13063/2327-9214.1265>
2. Bacon E, Budney G, Bondy J, et al. Developing a regional distributed data network for surveillance of chronic health conditions: the Colorado Health Observation Regional Data Service. *J Public Health*

- Manag Pract.* 2019;25(5):498–507. <https://doi.org/10.1097/PHH.0000000000000810>
3. Centers for Disease Control and Prevention. Automated detection and reporting of notifiable diseases using electronic medical records versus passive surveillance—Massachusetts, 2006–2007 June–July. *MMWR Morb Mortal Wkly Rep.* 2008; 57(14):373–376.
 4. Klompas M, McVetta J, Lazarus R, et al. Integrating clinical practice and public health surveillance using electronic medical record systems. *Am J Public Health.* 2012;102(suppl 3):S325–S332. <https://doi.org/10.2105/AJPH.2012.300811>
 5. Vogel J, Brown JS, Land T, Platt R, Klompas M. MDPHnet: secure, distributed sharing of electronic health record data for public health surveillance, evaluation, and planning. *Am J Public Health.* 2014; 104(12):2265–2270. <https://doi.org/10.2105/AJPH.2014.302103>
 6. Klompas M, Cocoros NM, Menchaca JT, et al. State and local chronic disease surveillance using electronic health record systems. *Am J Public Health.* 2017;107(9):1406–1412. <https://doi.org/10.2105/AJPH.2017.303874>
 7. Klompas M, Haney G, Church D, Lazarus R, Hou X, Platt R. Automated identification of acute hepatitis B using electronic medical record data to facilitate public health surveillance. *PLoS One.* 2008;3(7): e2626. <https://doi.org/10.1371/journal.pone.0002626>
 8. Allen-Dicker J, Klompas M. Comparison of electronic laboratory reports, administrative claims, and electronic health record data for acute viral hepatitis surveillance. *J Public Health Manag Pract.* 2012;18(3):209–214. <https://doi.org/10.1097/PHH.0b013e31821f2d73>
 9. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. *Diabetes Care.* 2013;36(4):914–921. <https://doi.org/10.2337/dc12-0964>
 10. Cocoros NM, Ochoa A, Eberhardt K, Zambarano B, Klompas M. Denominators matter: understanding medical encounter frequency and its impact on surveillance estimates using EHR data. *EGEMS (Wash DC).* 2019;7(1):31. <https://doi.org/10.5334/egems.292>
 11. Gruber S, Krakower D, Menchaca JT, et al. Using electronic health records to identify candidates for human immunodeficiency virus pre-exposure prophylaxis: an application of super learning to risk prediction when the outcome is rare. *Stat Med.* 2020;39(23):3059–3073. <https://doi.org/10.1002/sim.8591>
 12. Krakower DS, Gruber S, Hsu K, et al. Development and validation of an automated HIV prediction algorithm to identify candidates for pre-exposure prophylaxis: a modelling study. *Lancet HIV.* 2019; 6(10):e696–e704. [https://doi.org/10.1016/S2352-3018\(19\)30139-0](https://doi.org/10.1016/S2352-3018(19)30139-0)
 13. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines [erratum in *J Am Coll Cardiol.* 2014;63(25 Pt B):3026. *J Am Coll Cardiol.* 2014;63(25 Pt B):2935–2959. <https://doi.org/10.1016/j.jacc.2013.11.005>