Curr Diab Rep (2014) 14:468
DOI 10.1007/s11892-013-0468-7

HEALTH CARE DELIVERY SYSTEMS IN DIABETES (D WEXLER, SECTION EDITOR)

Innovative Uses of Electronic Health Records and Social Media for Public Health Surveillance

Emma M. Eggleston · Elissa R. Weitzman

Abstract Electronic health records (EHRs) and social media have the potential to enrich public health surveillance of diabetes. Clinical and patient-facing data sources for diabetes surveillance are needed given its profound public health impact, opportunity for primary and secondary prevention, persistent disparities, and requirement for self-management. Initiatives to employ data from EHRs and social media for diabetes surveillance are in their infancy. With their transformative potential come practical limitations and ethical considerations. We explore applications of EHR and social media for diabetes surveillance, limitations to approaches, and steps for moving forward in this partnership between patients, health systems, and public health.

This article is part of the Topical Collection on *Health Care Delivery Systems in Diabetes*

E. M. Eggleston (✉)
Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, 133 Brookline Avenue, Boston, MA 02215, USA
e-mail: emortoneggleston@partners.org

E. M. Eggleston
Brigham and Women's Hospital Division of Endocrinology, Hypertension and Diabetes, 122 Longwood Avenue, Boston, MA 02115, USA

E. R. Weitzman
Children's Hospital Informatics Program at the Harvard-MIT Division of Health Sciences and Technology, Boston Children's Hospital, Boston, MA, USA

E. R. Weitzman
Division of Adolescent Medicine, Boston Children's Hospital, Boston, MA, USA

E. R. Weitzman
Department of Pediatrics, Harvard Medical School, Boston, MA, USA

Keywords Diabetes · Public health · Surveillance · Electronic health record · Social media · Social networking · Innovative uses

Introduction

While the human and economic costs of diabetes are profound, effective measures exist to prevent or lessen its complications [1, 2] and, in the case of type 2 and gestational diabetes, to prevent disease onset [3, 4]. In this regard, the public health implications of diabetes coupled with clear opportunities for preventive intervention make it an important target for population-based surveillance. The goals of diabetes surveillance include monitoring disease burden overall and among high risk and emerging populations over time and, increasingly, collecting information on quality of care and treatment harms to inform public health policy and targeted interventions.

Multiple monitoring systems in the US undertake diabetes surveillance. Primary among these are 3 periodic national surveys, the Behavioral Risk Factor Surveillance System (BRFSS), National Health and Nutrition Examination Survey (NHANES), and National Health Information Survey (NHIS). The CDC also collects diabetes information from death registries, hospital claims data, Medicare and Medicaid claims, the US Renal Data systems, the Indian Health Service, and shares data with the National Center for Health Statistics and the National Institutes of Health [5]. For some aspects of the disease, for example, diabetes in youth [6] or diabetes-related eye disease [7], data are available from longitudinal cohort studies. However, even with this array of sources there is a trade-off between breadth of population coverage and depth of data obtained. The BRFSS is employed in all 50 states and Puerto Rico, but is limited to self-report and cross-sectional description of disease. NHANES includes detailed laboratory and physical exam data, but small sample size limits state level investigation or focus on high-risk and emerging populations. Cohort studies provide a source of comprehensive longitudinal data about sub-populations, including disease subtypes, progression over time, and quality of care but their cost and complexity constrain use for routine public health surveillance across populations. Traditional surveillance may also miss data vital to “living well” with a chronic disease [8] including information about treatment adherence and side effects, self-management practices, complementary or nontraditional therapies, and barriers to care. Collection of these data, many of which reside with the patient, may align surveillance with national initiatives to engage patients as collaborators in health research [9].

Electronic health records’ clinically rich data and potential for efficient longitudinal coverage of large populations provide an attractive option for augmenting traditional surveillance and building public health capacity in new and innovative directions [10–12], while patient-facing online social media platforms comprise novel sources for collecting patient-reported and -centered health information and hold the potential for health alerting and sharing of surveillance results.

Diabetes Surveillance Using Electronic Health Records

Data available in EHR’s include demographic, history, exam, laboratory, imaging, pharmacy, insurance status and, in integrated systems, claims data. These data can support case ascertainment of diabetes and its subtypes, assessment of care quality in care (eg, frequency and completeness of recommended tests, medications, screening, and treatment goals), risk factors (eg, family history, demographics, body mass index (BMI), smoking status, history of gestational diabetes), micro- and macrovascular complications, and co-morbidities. This detail, and the longitudinal nature of clinical care recorded in the EHR, in turn allows enhanced capture of disparities in disease risk, prevalence, outcomes, and care as well as information on emerging high risk sub-groups such as pre-existing diabetes in pregnancy. Automated systems that can extract, analyze, organize, and communicate EHR data to public health agencies are emerging. Initially developed for infectious disease surveillance [13] and emergency preparedness [14], EHR based surveillance systems are increasingly being applied to chronic disease [15], including diabetes [16••]. The work to apply and refine these systems is ongoing and is facilitated by (1) the growing use of EHR data for population management of patients with diabetes in medical home and other care delivery models, health system-based quality improvement initiatives, and comparative effectiveness research; (2) advances in security and capabilities of information technologies; and (3) the meaningful use legislation of the Affordable Care Act (Health Information Technology for Economic and Clinical Health – HITECH), which supports advances in consistency and quality of EHR data, requires collection of data on diabetes-specific measures, encourages continued development of health information exchanges to share data between EHR’s, and contains specific goals for support of public health and population health [17, 18••, 19–22].

Algorithm Development and Case Ascertainment

Algorithms that can identify cases and complications with high positive predictive value are critical in EHR surveillance, particularly with diseases like diabetes that have multiple subtypes, varied risk factors, and diverse complications. Several algorithms and prediction tools have been developed for ascertainment of diabetes and pre-diabetes in the EHR for panel management, clinical research, and population level surveillance [23, 24•, 25–27]. Surveillance using diagnostic codes alone is insensitive and often inaccurate as clinicians may not have the time to look up the correct ICD9 code (if not automated) or may use the same diagnostic code for resolved disease (eg, past history of gestational diabetes) and for possible current disease (eg, screening for type 2 diabetes). This can be particularly problematic in diabetes due to disease subtypes, lack of a uniformly applied code to denote “pre-diabetes” and lack of clarity in the application of codes for gestational diabetes vs overt pre-existing diabetes in pregnancy. Combining laboratory test orders and results with current and past ICD9 codes and prescriptions in automated decision rules can address weaknesses of ICD9 alone [12]. However, validation of measures [17, 28, 29] and of the final algorithms [24•, 25] is of critical importance.

ESP Diabetes - EHR Support for Public Health (ESP)

Algorithm development and implementation is a core part of the Electronic Medical Record Support for Public Health (ESP) system (http://esphealth.org), a generalizable surveillance platform that can extract structured data from any EHR, run analyses, and transmit summary data on a condition of interest to public health agencies [16••] (Fig. 1). ESP-diabetes is a CDC-funded project that collects and transmits data on diabetes incidence, complications, and quality (type 1, type 2, pre-diabetes, and gestational diabetes) to the Massachusetts Department of Public Health. The computing activity of ESP is independent from the host EHR to minimize computing burden on existing clinical functions, but is physically deployed within the host practice to maintain practice control over the security of clinical data. ESP investigators (including author EME) develop and validate diabetes detection algorithms [23, 24•] for case ascertainment and work with public health colleagues to define the data of greatest use for their surveillance and intervention goals. The summary data from ESP-DM can also be displayed

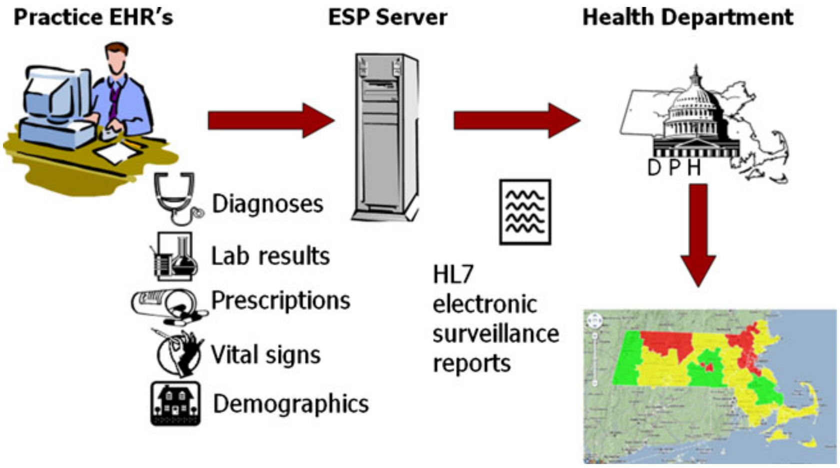

Fig. 1 ESP: Automated detection and reporting of conditions for public health surveillance. Data flow in the automated ESP-diabetes surveillance system. The data extraction module accepts comprehensive encounter data exported from any host EHR, and translates relevant local laboratory, diagnosis, and other codes into standard coding systems and ontologies. Detection algorithms are then applied for case ascertainment, and transmitted to public health partners via aggregate summaries. Summary data are mapped via the interactive mapping tool, RiskScape

on an interactive mapping tool, the RiskScape, to allow for community-specific tailored assessment and targeted intervention. For example, using the RiskScape public health officials can examine referral to a nutritionist by zip code in women with a history of gestational diabetes, and stratify by other factors such as self-reported race/ethnicity or BMI [16••]. The RiskScape and similar mapping modalities offer the possibility of layering EHR data with data on other determinants of health, such as transportation, sidewalk density, or distance from grocery stores for a layered picture of risk and resources across communities.

Limitations of EHR Data for Public Health Surveillance

While there is great potential for the innovative capture, analysis, and communication of population level data from the EHR, these data are primarily collected for the clinical care of individual patients and there are several limitations to their use for surveillance [29–32]. These include limitations in data validity, interpretability (biases), completeness, and privacy concerns. Limitations in validity arise from physician miscoding (eg, type 2 diagnosis codes used in patients with type 1, diagnosis codes used for screening rather than frank disease), EHR miscoding (oral glucose tolerance tests coded as fasting glucose), algorithm programming errors, and mode of entry into EHR (free text vs structured data field) [24•]. Medication and problem lists are generally entered as structured fields, but may not be up to date. Meaningful use requirements for medication and problem list reconciliation may increase the validity of these critical measures over time. Laboratory tests are also extracted from structured fields, but current procedural terminology (CPT) codes for glucose tests can be applied to the same test (fasting and random glucose) and, of central importance to case ascertainment by laboratory tests, fasting tests may not be labeled as such in the EHR. There are several important diabetes-specific parameters that may not be captured in a structured data field. These include hypoglycemia, lifestyle factors (diet and exercise), family history of diabetes, compliance discussions, and presence of co-morbidities that may not be recorded in the problem list (eg, depression or treatment-related anxiety, sexual dysfunction, peripheral neuropathy, fatty liver disease). Advances in natural language processing may help to capture these elements more consistently [33, 34]. However, to reach the full potential of comprehensive and accurate diabetes surveillance, diabetes specific taxonomies and consistent structured data fields for high importance diabetes variables are needed across EHR's. These structures must be developed with forethought and balancing of at times competing goals. As noted by others [35], there is a tension between need for flexibility and expressiveness in the EHR for clinical care—its fundamental purpose—and the need for structured data for research, quality, and surveillance activities. Last, as with other health outcomes [18••, 36], there are important variables of public health interest and impact on diabetes risk and outcomes—such as transportation, violence in the home or community, food security—that are not found in the EHR as they are not routinely collected by health care systems.

Completeness of data is further influenced by missing data due to care or testing received at outside health care providers and lack of integration across systems. Health information exchanges and disseminated networks that allow exchange of data between multiple EHR systems are under development and in some areas, already implemented for both clinical care and public health purposes [11, 37–39], however, considerable barriers in funding and insitutional concerns over privacy and competition remain [40]. The availability of claims data to augment the EHR can improve capture of care delivered outside the host system and, in the case of medications, indicate filling of the prescription by patients—an important measure of adherence [25].

Several biases can be introduced by the use of EHR data for surveillance [31, 32]. For diabetes, confounding by indication when using laboratory results for case ascertainment is a particular concern. Unlike modes of surveillance, in which an entire population is sampled or tested, in the clinical setting testing is performed per clinical judgment on a per patient basis. For example, clinicians are more likely to check an A1C in an asymptomatic patient in whom they have a high suspicion of diabetes. There is also the risk of over, and under, ascertainment of disease based on glycemic testing depending on the test used and the threshold chosen as: (1) A1C, fasting plasma glucose, and oral glucose tolerance tests capture differing proportions, and in some cases differing populations, of patients with dysglycemia [41–46]; (2) glycemic tests may differ by age, race/ethnicity, and pregnancy status [47–49]; and (3) detection algorithms may use differing decision rules in defining a "case" with potentially profound effects on ascertainment due to type, timing, and repetition of test.

Last, privacy concerns are a central factor in the use of EHR data as patients come to their providers for clinical care, not to participate in public health surveillance. Parameters for the use of data for public health surveillance according to the HIPAA privacy rule include aggregate reporting, limited datasets, and de-identification of data [50••]. Meaningful use legislation directly addresses use of EHR data for public health surveillance [22] and supports further development of health exchanges to allow sharing of data and coverage of large populations. Advances in information technology bring the great potential of such exchange closer to reality, but attendant attention to data safety and transparency of use will be critical moving forward, particularly with diabetes—a disease with profound implications for insurance status, employment status, and social stigma.

Social Media-Based Diabetes Surveillance

Population engagement in online social media is generating opportunities to glean information relevant to diabetes surveillance from patient communication. Two thirds (65 %) of adult internet users are in social networking sites (SNSs) [51], many in disease-centered communities and groups [52, 53]. More than 500 diabetes-focused groups exist on Facebook [54], in addition to the many large online diabetes communities operating on other social networking platforms [55••]. Engagement reflects the driving appeal of social connection and—for patients with a health problem—the value placed on community [56], information, and support [52, 57].

Obtaining health information and personalized support from online peers is important to navigating care systems, choosing treatments, and following care plans [59–62]. Participants prefer to draw on personal ties for health-related help. More adults favor communicating with nonprofessional than professional members of their social network for "emotional support in dealing with a health issue" and "a quick remedy for dealing with an everyday health issue", while equal percentages favor tapping informal and professional network contacts for "practical advice for coping with day-to-day health situations" [63]. Hence, discussions in online disease-centered social networks often center on issues related to treatments, symptoms, and side effects in addition to general lifestyle issues [54, 64], topics that may fill gaps in traditional surveillance and guide targeted intervention.

While engaging with online peers through social media may augment public health and clinical services to address rising levels of chronic disease [65–68] and constrained healthcare finances [69, 70], the medium's power to sustain engagement and foster communication could offset cost, labor and processing demands associated with large sample prospective data collection and ameliorate surveillance-related challenges to (a) recruiting and maintaining study cohorts; (b) moving beyond single disease data models that are difficult to modify once in the field to assess co-occurring and interacting conditions; (c) responding rapidly and at scale to emerging health phenomena or findings; and, (d) "closing the loop" between collection and analysis of research data and translation or communication of findings to source populations. The capacity of social media platforms to support bidirectional communication with distributed cohorts may be especially helpful for monitoring noncommunicable diseases (NCDs) such as diabetes that persist over time. Indeed, models of participatory surveillance in diabetes research that build on patients' impulses toward "information altruism" [71] may help foster sustainable relationships between investigators and patients to improve understanding of disease course, correlates and treatment safety/efficacy.

To date, most uses of social media to address diabetes have focused on patient education and support for self-management [72]. Nevertheless, pioneering surveillance efforts that use a range of social media platforms and tools are emerging and encompass *passive approaches*, in which patients neither know nor consent to having their online social media communications used for research, and *active approaches*, in which patients deliberately donate health information. Approaches vary in the extent to which they afford patients choice over whether to participate and what information is shared, and in the extent to which data are collected in structured and standardized formats.

Passive Surveillance of Diabetes through Social Media

In passive surveillance approaches, the content of published posts and reported interests of social media users are abstracted, coded, and categorized for use as indicators of health risk and problem. Because passive approaches can rely on mining of unstructured text, issues of consent, transparency, and validity are raised. Authors of communications may not know their missives are being abstracted and measured [73]. Inferences based on construed meaning of communications may lack context, structure, and standardization—necessitating careful validation. Without unique identifiers to attribute a communication to a given person—counts of persons or events may be incorrect and duplicative. These constraints limit the utility of passive monitoring for cohort or follow-up investigations. Nevertheless the approach can provide important information about levels of problem or concern in a community or network [74], spread of information or awareness [75, 76] and even serve as proxies for more formal measures of health status [77••].

Active Surveillance of Diabetes through Social Media

In active approaches, social media platforms are configured to support collection of structured data entered by patients intentionally for the purpose of sharing information for health research. Engaged populations can be polled about health status, symptoms, treatments, and side effects [78–81]. Patients are willing to share their electronic data for research, especially under models where sharing is transparent and noncommercial [82–84]. Knowledge gleaned from data shared within online communities has produced results comparable, in some circumstances, to randomized trials [85], and is filling gaps in traditional monitoring of healthcare quality and disease status [81, 86]. Patient-led or "citizen science" research falls into the active approach category and may complement traditional top-down efforts [87]. These efforts aim to advance knowledge and accelerate the evidence base by collecting semi-structured information from registered participants. Information is aggregated and fed back to participants to "crowd-source" knowledge and foster peer exchange. Participants may be notified in the terms of use published for a site that their data will be shared; however, participants may not give formal informed consent, in contrast to traditionally constituted research that is governed by institutional review boards. Both PatientsLikeMe.com and CureTogether.com fall within the participant-led rubric of health research and have engaged thousands of patients in contributing health information to support disease discovery and treatment. While these efforts have engaged many patients with diabetes, they have not yet produced diabetes research reports or been used for public health surveillance.

The TuAnalyze Project – Participatory Surveillance of Diabetes in an Online Diabetes Social Network

The TuAnalyze project is a hybrid model, in which investigators (including author ERW), with support from the US Centers for Disease Control and Prevention, partnered with an online international nonprofit diabetes social network called TuDiabetes.org to advance participatory surveillance of diabetes (Fig. 2). The investigators created a privacy-preserving software app called "TuAnalyze" that was launched into the online community as a freely available tool to support participation in health research, initially using it to collect information about glycemic control from participants by diabetes type, filling a gap in traditional health monitoring. The app was engineered to enable safe sharing of personal health information through a software architecture, in which participant surveys and patient-reported outcomes (PROs) were collected and stored in a personally controlled health record [88] created upon user authorization as part of the initial permission and preference setting process established for uploading of the app [86]. English and Spanish language versions of the app were implemented to reach a global membership community. Research data were not stored on the network platform to protect them from unauthorized use. The architecture allowed researchers to poll across consented records. A spatial-temporal mapping feature of the app displayed aggregated member-reported A1C measures for users at state, country, and province levels. The interactive map was implemented to help participants contextualize their measure and to encourage tracking and engagement over time. Participants could complete surveys through the app about their healthcare quality, self-care, experiences of problems and harms, comorbidities, and other topics. Finally, the TuAnalyze program maintained a research blog in the TuDiabetes environment, in which aggregated research data were summarized and fed back to the community for comment and interpretation. Approximately one-fifth of active site members signed on to using TuAnalyze and over 97 % of initially engaged subjects set preferences to allow research recontact; use of the app diffused to almost 4,000 persons, distributed across all 50 US states and internationally. Early adopters and users who selected greater openness in sharing their personal data were, overall, better controlled than late adopters and users who selected less openness in their sharing settings [81]. TuAnalyze data enabled assessment of adherence to preventive care recommendations and of healthcare quality, as well as measurement of comorbidities and challenges among the engaged cohort including by diabetes type [81], filling gaps in traditional health surveillance. An assessment of hypoglycemia and harms related to insulin use was also undertaken, and measures of severe hypoglycemia, frequent recent hypoglycemia, and patient-centered harms reported [89••]. In this fashion, the app served as an exemplar of the use of a patient-centered health information technology employed to quantify drug (insulin) safety, filling a recognized gap [90, 91•].

Limitations of Social Media for Diabetes Surveillance

Social-media based surveillance of diabetes is a promising complement to traditional approaches, but practical and

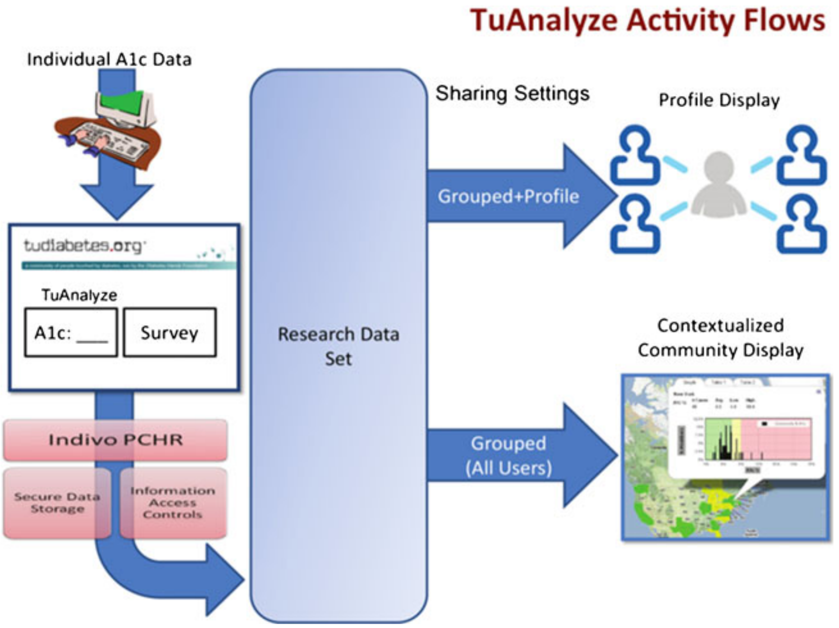

Fig. 2 TuAnalyze model for participatory diabetes surveillance. Using the TuAnalyze app, TuDiabetes members report about their glycemic control, and answer surveys. Data are stored in a personally controlled health record (PCHR), not on the network. Users set permissions for data sharing. Researchers poll across consented PCHRs, for surveillance. Users can see contextualized views of their data on maps/charts and reports about aggregated data in a research blog published on the TuDiabetes site (not depicted). (Adapted from: Weitzman ER, Adida B, Kelemen S, Mandl KD. Sharing data for public health research by members of an international online diabetes social network. PloS One. 2011;6:e19256. doi:10.1371/journal.pone.0019256). [86]

technological progress is needed to address limitations. Efforts to validate social media-sourced data are in an early stage and, in some cases the approach's opt-in, open platform model that operates outside of healthcare systems may preclude validation. Inferences made from engaged samples may not generalize to the population of persons affected by a given condition, and engaged samples may be biased given the opt-in nature of social media activity [92]. Advancing understanding of the composition of self-selecting research cohorts and those defined through passive surveillance is important to interpreting and using data to support policy, prevention and healthcare practice. Ethical standards and practice norms governing health research may imperfectly apply to social media, raising questions about how to balance innovation and openness with principles of patient privacy, autonomy, and consent [93]. New frameworks are emerging for ethics review and oversight of "participant-led research" including research undertaken in collaboration with social networking sites [94••]. Finally, while patients may choose to participate in social media environments to learn about health conditions and participate in research, it is unclear whether they are aware of dangers related to misinformation in these environments, or to unauthorized or unanticipated uses of their health information [55••, 95]. Active surveillance efforts in particular have a special obligation to encourage and engineer for safe sharing of personal health information that may be readily linked to individuals.

Conclusions

The rising incidence of all diabetes subtypes, disproportionate impact of disease on different sociodemographic groups, and wide geographic variations in prevalence, health behaviors, and clinical care speak to the need for timely and comprehensive public health surveillance to inform intervention. Detailed information from the vantage of both care systems and patients is critical to guide use of limited public health resources for disease prevention and to monitor the impact of interventions. Emerging advances in the extraction, analysis, and communication of information from EHR and social media bring the possibility of innovative new public health surveillance approaches to augment and extend surveillance efforts beyond their traditional purview.

Nevertheless, despite their strengths and opportunities, both EHR and social media based approaches may introduce bias and carry important limitations. Considerable work is needed to ensure consistency and validity of data across sources, platforms and systems. In the case of the EHR, distributed data networks are under development to allow separate systems to interface and allow the sharing of data within and between regions. Refinement of coded data and natural language processing may allow improved capture of key measures, but consistent diabetes specific taxonomies are needed. For social media, characterizing self-selecting samples and validating patient reports and/or structuring them to align with standardized measures without losing nuance or the authenticity of peer communication is vital. Both EHR and social media based surveillance require rigorous attention to data security, patient privacy, and consent. Transparency about the intended use of shared data and a commitment to balancing risk and benefit, including through return of research results where possible, are of fundamental importance to these potentially transformative approaches that, together with traditional surveillance systems engage public health, providers, and patients in combating this devastating disease.

Acknowledgments This work was supported by PO1HK000088-01 from the Centers for Disease Control and Prevention (CDC), National Library of Medicine grants 5R01LM007677 and G08LM009778, grant 1U54RR025224-01 from NCRR/NIH, and grant 1R01AA021913-01 from the National Institute of Alcohol Abuse and Alcoholism. The study sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

Emma M. Eggleston was Co-I on the CDC grant that funded the ESP-DM public health surveillance project.

Compliance with Ethics Guidelines

Conflict of Interest Emma M. Eggleston and Elissa R. Weitzman declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med. 1993;329:977–86. doi:10.1056/NEJM199309303291401.
2. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. Lancet. 1998;352:837–53.
3. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med. 2002;346:393–403. doi:10.1056/NEJMoa012512.
4. Ratner RE, Christophi CA, Metzger BE, Dabelea D, Bennett PH, Pi-Sunyer X, et al. Prevention of diabetes in women with a history of gestational diabetes: effects of metformin and lifestyle interventions. J Clin Endocrinol Metabol. 2008;93:4774–9. doi:10.1210/jc.2008-0772.
5. Retnakaran R, Qi Y, Sermer M, Connelly PW, Zinman B, Hanley AJG. Isolated hyperglycemia at 1 hour on oral glucose tolerance test in pregnancy resembles gestational diabetes mellitus in predicting postpartum metabolic dysfunction. Diabetes Care. 2008;31:1275–81. doi:10.2337/dc08-0126.
6. Dabelea D, Pihoker C, Talton JW, D'Agostino Jr RB, Fujimoto W, Klingensmith GJ, et al. Etiological approach to characterization of diabetes type: the SEARCH for Diabetes in Youth Study. Diabetes Care. 2011;34:1628–33. doi:10.2337/dc10-2324.
7. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin epidemiologic study of diabetic retinopathy. III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. Arch Opthalmol. 1984;102:527–32.
8. Institute of Medicine. Living well with chronic illness: a call for public health action. Consensus Report; National Academies Press; 2012.
9. Selby JV. The patient-centered outcomes research institute: a 2013 agenda for "research done differently". Popul Health Manag. 2013;16:69–70. doi:10.1089/pop.2013.1621.
10. Anastasio S, Tabbot P. Independent study rough draft: 1 December 2012. The Public Health Significance of Implementing Electronic Health Records in New Jersey. The Public. 2012.
11. Bailey LC, Milov DE, Kelleher K, Kahn MG, Del Beccaro M, Yu F, et al. Multi-institutional sharing of electronic health record data to assess childhood obesity. PloS One. 2013;8:e66192.
12. Klompas M, Murphy M, Lankiewicz J, McVetta J, Lazarus R, Eggleston E, et al. Harnessing electronic health records for public health surveillance. J Public Health, Inform (online). 2011;3.
13. 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:e2626. doi:10.1371/journal.pone.0002626.
14. Bradley CA, Rolka H, Walker D, Loonsk J. BioSense: implementation of a National Early Event Detection and Situational Awareness System. MMWR MMWR. 2005;54(Suppl):11–9.
15. Baker DW, Persell SD, Thompson JA, Soman NS, Burgner KM, Liss D, et al. Automated review of electronic health records to assess quality of care for outpatients with heart failure. Ann Intern Med. 2007;146:270–7.
16. •• Klompas M, McVetta J, Lazarus R, Eggleston E, Haney G, Kruskal BA, et al. Integrating clinical practice and public health surveillance using electronic medical record systems. Am J Prev Med. 2012;42:S154–62. *Review of EHR use for public health surveillance and detailed description of the Electronic Medical Recor Support for Public Health (ESP) project.*
17. Ahmad FS, Tsang T. Diabetes prevention, health information technology, and meaningful use: challenges and opportunities. Am J Prev Med. 2013;44:S357–63.
18. •• Friedman DJ, Parrish RG, Ross DA. Electronic health records and US public health: current realities and future promise. Am J Pub Health. 2013:e1–8. *Comprehensive reivew of the potential of, and barriers to, the use of EHR data for public health surveillance and proposed steps to ensure that the use of EHR data to support public health reaches its full potential.*
19. Adler-Milstein J, Jha AK. Sharing clinical data electronically a critical challenge for fixing the health care system. JAMA. 2012;307:1695–6.
20. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010;363:501–4. doi:10.1056/NEJMp1006114.
21. Health IT Policy Committee. Meaningful use workgroup request for comments regarding meaningful use stage 2. In: Office of the National Coordinator for Health Information Technology, editor. Washington, DC; 2011.
22. CDC. Meaningful Use Fact Sheet: Syndromic Surveillance. Public Health Information Network. 2012. http://www.cdc.gov/phin/library/PHIN_Fact_Sheets/FS_MU_SS.pdf.
23. Klompas M, McVetta J, Eggleston E, et al. Automated surveillance and public health reporting for gestational diabetes incidence and

care using electronic health record data (Abstract). J Emerg Health Threats. 2011;4:S34.

24.• 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;36:914–21. *Validation of public health diabetes surveillance algorithm to ascertain and distinguish type 1 from type 2 diabetes in the EHR.*

25. Kudyakov R, Bowen J, Ewen E, West SL, Daoud Y, Fleming N, et al. Electronic health record use to classify patients with newly diagnosed vs preexisting type 2 diabetes: infrastructure for comparative effectiveness research and population health management. Popul Health Manage. 2012;15:3–11.

26. Rathmann W, Kowall B, Heier M, Herder C, Holle R, Thorand B, et al. Prediction models for incident type 2 diabetes mellitusin the older population: KORA S4/F4 cohort study. Diabet Med. 2010;27: 1116–23. doi:10.1111/j.1464-5491.2010.03065.x.

27. Makam AN, Nguyen OK, Moore B, Ma Y, Amarasingham R. Identifying patients with diabetes and the earliest date of diagnosis in real time: an electronic health record case-finding algorithm. BMC Med Inform Decis Mak. 2013;13:1–7.

28. Benin AL, Fenick A, Herrin J, Vitkauskas G, Chen J, Brandt C. How good are the data? Feasible approach to validation of metrics of quality derived from an outpatient electronic health record. Am J Med Qual. 2011;26:441–51. doi:10.1177/1062860611403136.

29. Botsis T, Hartvigsen G, Chen F, Weng C. Secondary use of EHR: data quality issues and informatics opportunities. AMIA Summits Transl Sci Proc AMIA Summit Transl Sci. 2010;2010:1–5.

30. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev. 2010;67:503–27. doi:10.1177/1077558709359007.

31. Chiolero A, Santschi V, Paccaud F. Public health surveillance with electronic medical records: at risk of surveillance bias and overdiagnosis. Eur J Pub Health. 2013;23:350–1. doi:10.1093/eurpub/ckt044.

32. Hoffman S, Podgurski A. Big bad data: law, public health, and biomedical databases. J Law Med Ethics. 2013;41(Suppl):56–60.

33. Fischer LR, Rush WA, Kluznik JC, O'Connor PJ, Hanson AM. Abstract C-C1-06: identifying depression among diabetes patients using natural language processing of office notes. Clin Med Res. 2008;6:125–6.

34. Hazlehurst B, Sittig DF, Stevens VJ, Smith KS, Hollis JF, Vogt TM, et al. Natural language processing in the electronic medical record: assessing clinician adherence to tobacco treatment guidelines. Am J Prev Med. 2005;29:434–9. doi:10.1016/j.amepre.2005.08.007.

35. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. JAMIA. 2011;18:181–6. doi:10.1136/jamia.2010.007237.

36. Kukafka R, Ancker JS, Chan C, Chelico J, Khan S, Mortoti S, et al. Redesigning electronic health record systems to support public health. J Biomed Inform. 2007;40:398–409. doi:10.1016/j.jbi.2007.07.001.

37. Brown JS, Holmes JH, Shah K, Hall K, Lazarus R, Platt R. Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care. Med Care. 2010;48(6 Suppl):S45–51. doi:10.1097/MLR.0b013e3181d9919f.

38. Maro JC, Platt R, Holmes JH, Strom BL, Hennessy S, Lazarus R, et al. Design of a national distributed health data network. Ann Intern Med. 2009;151:341–4.

39. Rea S, Pathak J, Savova G, Oniki TA, Westberg L, Beebe CE, et al. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: the SHARPn project. J Biomed Inform. 2012;45:763–71. doi:10.1016/j.jbi.2012.01.009.

40. Lenert L, Sundwall DN. Public health surveillance and meaningful use regulations: a crisis of opportunity. J Inform. 2012;102:e1–7.

41. Cohen RM, Lindsell CJ. When the blood glucose and the HbA1c don't match: turning uncertainty into opportunity. Diabetes Care. 2012;35:2421–3. doi:10.2337/dc12-1479.

42. Mann DM, Carson AP, Shimbo D, Fonseca V, Fox CS, Muntner P. Impact of A1C screening criterion on the diagnosis of pre-diabetes among U.S. adults. Diabetes Care. 2010;33:2190–5.

43. Picón MJ, Murri M, Muñoz A, Fernández-García JC, Gomez-Huelgas R, Tinahones FJ. Hemoglobin A1c versus oral glucose tolerance test in postpartum diabetes screening. Diabetes Care. 2012;35:1648–53. doi:10.2337/dc11-2111.

44. Selvin E, Steffes MW, Gregg E, Brancati FL, Coresh J. Performance of A1C for the classification and prediction of diabetes. Diabetes Care. 2011;34:84–9.

45. Balion CM, Raina PS, Gerstein HC, Santaguida PL, Morrison KM, Booker L, et al. Reproducibility of impaired glucose tolerance (IGT) and impaired fasting glucose (IFG) classification: a systematic review. Clin Chem Lab Med. 2007;45:1180–5. doi:10.1515/CCLM.2007.505.

46. Saukkonen T, Cederberg H, Jokelainen J, Laakso M, Härkönen P, Keinänen-Kiukaanniemi S, et al. Limited overlap between intermediate hyperglycemia as defined by A1C 5.7–6.4 %, impaired fasting glucose, and impaired glucose tolerance. Diabetes Care. 2011;34: 2314–6.

47. Christensen DL, Witte DR, Kaduka L, Jørgensen ME, Borch-Johnsen K, Mohan V, et al. Moving to an A1C-based diagnosis of diabetes has a different impact on prevalence in different ethnic groups. Diabetes Care. 2010;33:580–2. doi:10.2337/dc09-1843.

48. Herman WH, Ma Y, Uwaifo G, Haffner S, Kahn SE, Horton ES, et al. Differences in A1C by race and ethnicity among patients with impaired glucose tolerance in the Diabetes Prevention Program. Diabetes Care. 2007;30:2453–7. doi:10.2337/dc06-2003.

49. Herranz L, Saez-de-Ibarra L, Grande C, Pallardo LF. Non–glycemic-dependent reduction of late pregnancy A1C levels in women with type 1 diabetes. Diabetes Care. 2007;30:1579–80. doi:10.2337/dc06-2568.

50.•• Centers for Disease Control and Prevention. HIPAA Privacy Rule and Public Health: Guidance from CDC and the US Department of Health and Human Services. MMWR 2003;1–20. *Report on the application and interpretation of HIPAA privacy rules for the use of clinical data in public health surveillance.*

51. Madden M, Zickuhr K. 65 % of online adults use social networking sites. Washington: Pew Research Center; 2011.

52. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res. 2009;11:e48.

53. Fox S, Purcell K. Chronic disease and the internet. Washington: Pew Research Center; 2010.

54. De la Torre-Diez I, Diaz-Pernas FJ, Anton-Rodriguez M. A content analysis of chronic diseases social groups on Facebook and Twitter. Telemed J E Health. 2012;18:404–8. doi:10.1089/tmj.2011.0227.

55.•• Weitzman ER, Cole E, Kaci L, Mandl KD. Social but safe? Quality and safety of diabetes-related online social networks. JAMIA. 2011;18:292–7. doi:10.1136/jamia.2010.009712. *Multi-site evaluation of the quality and safety of the 10 most frequently referenced online diabetes social networking sites. A structured measurement protocol was used to rate sites along 28 indicators of quality and safety across 4 domains.*

56. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. J Med Internet Res. 2008;10:e15.

57. Seeman N. Web 2.0 and chronic illness: new horizons, new opportunities. Healthcare Q. 2008;11:104–8. 10, 4.

58. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease.

Cochrane Database Syst Rev. 2005:CD004274. doi:10.1002/14651858.CD004274.pub4.

59. Yang HC, Thornton LM, Shapiro CL, Andersen BL. Surviving recurrence: psychological and quality-of-life recovery. Cancer. 2008;112:1178–87. doi:10.1002/cncr.23272.
60. Bakitas M, Lyons KD, Hegel MT, Balan S, Brokaw FC, Seville J, et al. Effects of a palliative care intervention on clinical outcomes in patients with advanced cancer: the Project ENABLE II randomized controlled trial. JAMA. 2009;302:741–9. doi:10.1001/jama.2009.1198.
61. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med. 2010;26(3):287–92.
62. Kemp V. Use of 'chronic disease self-management strategies' in mental healthcare. Curr Opin Psychiatry. 2011;24:144–8. doi:10.1097/YCO.0b013e3283438014.
63. Kuehn BM. Patients go online seeking support, practical advice on health conditions. JAMA. 2011;305:1644–5. doi:10.1001/jama.2011.509.
64. Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. J Gen Intern Med. 2011;26:287–92. doi:10.1007/s11606-010-1526-3.
65. Hughes SL, Weaver FM, Giobbie-Hurder A, Manheim L, Henderson W, Kubal JD, et al. Effectiveness of team-managed home-based primary care: a randomized multi-center trial. JAMA. 2000;284:2877–85.
66. Bousquet J, Bousquet PJ, Godard P, Daures JP. The public health implications of asthma. Bull World Health Organ. 2005;83:548–54.
67. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. Lancet. 2006;367:1747–57. doi:10.1016/S0140-6736(06)68770-9.
68. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med. 2006;3:e442. doi:10.1371/journal.pmed.0030442.
69. Bodenheimer T, Chen E, Bennett HD. Confronting the growing burden of chronic disease: can the U.S. health care workforce do the job? Health Aff. 2009;28:64–74.
70. Glassman A, Gaziano TA, Bouillon Buendia CP, Guanais de Aguiar FC. Confronting the chronic disease burden in Latin America and the Caribbean. Health Aff. 2010;29:2142–8. doi:10.1377/hlthaff.2010.1038.
71. Kohane IS, Altman RB. Health-information altruists—a potentially critical resource. N Engl J Med. 2005;353:2074–7.
72. Hamm MP, Chisholm A, Shulhan J, Milne A, Scott SD, Given LM, et al. Social media use among patients and caregivers: a scoping review. BMJ Open. 2013;3. doi:10.1136/bmjopen-2013-002819.
73. Bond CS, Ahmed OH, Hind M, Thomas B, Hewitt-Taylor J. The conceptual and practical ethical dilemmas of using health discussion board posts as research data. J Med Internet Res. 2013;15:e112. doi:10.2196/jmir.2435.
74. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PloS One. 2011;6:e19467.
75. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PloS One. 2010;5:e14118. doi:10.1371/journal.pone.0014118.
76. Cassa CA, Chunara R, Mandl K, Brownstein JS. Twitter as a sentinel in emergency situations: lessons from the Boston marathon explosions. PLoS Curr. 2013;5. doi:10.1371/currents.dis.ad70cd1c8bc585e9470046cde334ee4b.
77.•• Chunara R, Bouton L, Ayers JW, Brownstein JS. Assessing the online social environment for surveillance of obesity prevalence. PLoS One. 2013;8:e61373. doi:10.1371/journal.pone.0061373. *This cross-sectional observational study among Facebook users examined the relationship between Facebook use and specific interests posted and prevalence of obesity and overweight in US as measured by the Behavioral Risk Factor Surveillance System (BRFSS) and across New York City neighborhoods as measured by NYC EpiQuery.*
78. Hand E. Citizen science: people power. Nature. 2010;466:685–7.
79. van Noort SP, Muehlen M, Rebelo de Andrade H, Koppeschaar C, Lima Lourenco JM, Gomes MG. Gripenet: an internet-based system to monitor influenza-like illness uniformly across Europe. Eur Surveill. 2007;12:E5–6.
80. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. N Engl J Med. 2009;360:2153–5. 7.
81. Weitzman ER, Kelemen S, Mandl KD. Surveillance of an online social network to assess population-level diabetes health status and healthcare quality. J Pub Health, Inform (Online). 2011;3.
82. Weitzman ER, Kaci L, Mandl KD. Sharing medical data for health research: the early personal health record experience. J Med Internet Res. 2010;12:e14.
83. Zulman DM, Nazi KM, Turvey CL, Wagner TH, Woods SS, An LC. Patient interest in sharing personal health record information: a web-based survey. Ann Intern Med. 2011;155:805–10. doi:10.1059/0003-4819-155-12-201112200-00002.
84. Kaye J, Curren L, Anderson N, Edwards K, Fullerton SM, Kanellopoulou N, et al. From patients to partners: participant-centric initiatives in biomedical research. Nat Rev Genet. 2012;13:371–6. doi:10.1038/nrg3218.
85. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res. 2011;13:e6.
86. Weitzman ER, Adida B, Kelemen S, Mandl KD. Sharing data for public health research by members of an international online diabetes social network. PloS One. 2011;6:e19256. doi:10.1371/journal.pone.0019256.
87. Swan M. Crowd sourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. J Med Internet Res. 2012;14:e46. doi:10.2196/jmir.1988.
88. Adida B, Sanyal A, Zabak S, Kohane IS, Mandl KD. Indivo x: developing a fully substitutable personally controlled health record platform. AMIA Ann Symp Proc / AMIA Symp. 2010;2010:6–10.
89.•• Weitzman ER, Kelemen S, Quinn M, Eggleston EM, Mandl KD. Participatory surveillance of hypoglycemia and harms in an online social network. JAMA Intern Med. 2013;173:345–51. doi:10.1001/jamainternmed.2013.2512. *This cross-sectional survey used a freely available software app to engage members of an online international diabetes social network in reporting about hypoglycemia frequency, severity, and impact.*
90. Pogach L, Aron D. Balancing hypoglycemia and glycemic control: a public health approach for insulin safety. JAMA. 2010;303:2076–7. doi:10.1001/jama.2010.655.
91.• Redberg RF. Online network for adverse event reporting. JAMA Intern Med. 2013;173:352–3. doi:10.1001/jamainternmed.2013.3445. *Editorial on the importance of social media surveillance of hypoglycemia and harms as a novel means for post-marketing drug safety surveillance.*
92. Rhodes SD, Bowie DA, Hergenrather KC. Collecting behavioral data using the world wide web: considerations for researchers. J Epidemiol Community Health. 2003;57:68–73.
93. Li J. Privacy policies for health social networking sites. JAMIA. 2013;20:704–7. doi:10.1136/amiajnl-2012-001500.
94.•• Vayena E, Tasioulas J. Adapting standards: ethical oversight of participant-led health research. PLoS Med. 2013;10:e1001402. doi:10.1371/journal.pmed.1001402. *This policy forum report describes the ethical context for participant-led research and proposes a framework for consenting cohorts and providing ethical oversight.*
95. Quinn E, Corrigan M, McHugh S, Murphy D, O'Mullane J, Hill A, et al. Who's talking about breast cancer? Analysis of daily breast cancer posts on the internet. Breast. 2013;22:24–7.