Opportunities and Challenges for Developing Syndromic Surveillance Systems for the Detection of Social Epidemics

David Scales[1]*

[1]Joan and Sanford I. Weill Medical College of Cornell University

Abstract

This commentary explores the potential and challenges of developing syndromic surveillance systems with the ability to more rapidly detect epidemics of addiction, poverty, housing instability, food insecurity, social isolation and other social determinants of health (SDoH). Epidemiologists tracking SDoH heavily rely on expensive government surveys released annually, delaying for months if not years the timely detection of social epidemics, defined as sudden, rapid or unexpected changes in social determinants of population health. Conversely, infectious disease syndromic surveillance is an effective early warning tool for epidemic diseases using various types of non-traditional epidemiological data from emergency room chief complaints to search query data. Based on such experience, novel social syndromic surveillance systems for early detection of social epidemics with health implications are not only possible but necessary. Challenges to their widespread implementation include incorporating disparate proprietary data sources and database integration. Significantly more resources are critically needed to address these barriers to allow for accessing, integrating and rapidly analyzing appropriate data streams to make syndromic surveillance for social determinants of health widely available to public health professionals.

Key Words: Syndromic surveillance; social determinants of health; social epidemics; public health informatics

Introduction

The past two decades have seen parallel but separate developments of renewed interest in social epidemiology—the social, economic, cultural and political forces that impact, and at times determine, patient health outcomes [1]—and syndromic surveillance for infectious diseases, defined as an investigational approach where staff, assisted by automated data acquisition and statistical alerts, monitor disease indicators in real-time or near real-time to detect outbreaks earlier than would otherwise be possible with traditional methods (adapted from [2]).

Decades of research have convincingly demonstrated how the social environments in which patients live are risk factors that predict disease risk and outcomes [3]. Increasingly experts recommended incorporating ways to screen for and address social determinants of health (SDoH) within clinical care delivery. A number of health systems functioning in an accountable care model, such as some insurance companies and Medicaid in certain states, have embarked on aligning the various community stakeholders required to effectively pursue this path, and a number of these initiatives are actively being rolled out and evaluated [4-6].

Despite these welcome advances, there has been less focus on the necessary surveillance systems that can improve epidemiological monitoring of SDoH (for example, see [7]), though leveraging social data to benefit communities does appear to be on the rise [8]. Specifically, if we view individual SDoH as contributors to health issues, then we should also invest in epidemiological systems that provide insight into how SDoH prevalence in a given area changes over time plus sentinel systems for early detection of rapid changes as harbingers of socioeconomic crises.

We are currently far from this epidemiological ideal. SDoH surveillance data lags months to years behind clinical surveillance data, creating a relatively static retrospective tableau of prevalence that misrepresents the real-time environment. This causes significant delays in both detection and response to rapid changes in SDoH, like epidemics of housing instability or food insecurity.

This paper will examine a number of challenges to translating the success of syndromic surveillance to the detection of social epidemics, defined as sudden, rapid or unexpected changes in social determinants of population health (definition derived from [9]). These challenges are surmountable, however, and are likely within reach for teams developing integrated frameworks with local partners for SDoH screening at the point of care. With foresight and planning, such partnerships can help ensure the data infrastructure to make future SDoH syndromic surveillance possible as funding becomes more available.

The pressing need for more rapid detection of social epidemics

Data sources measuring SDoH are not released with sufficient frequency or local granularity to provide much insight into real-time trends. Yet, as risk-based payment models increasingly require addressing SDoH issues, such interventions must detect and respond to social epidemics as early as possible for the interventions to be economically viable. There is therefore a pressing need for rapid detection via cheaper mechanisms than costly annual surveys.

The maximum release frequency of much sociological, demographic and economic data is yearly, a rate far too infrequent to identify trends or detect rapid changes in local SDoH. This is the case for census American Community Survey (ACS) estimates, the Behavioral Risk Factor Surveillance System (BRFSS), an annual telephone survey collecting state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services, and the National Health and Nutrition Examination Survey (NHANES), which assesses health and nutritional status by combining interviews and physical examinations. Such annual data releases are further limited by their large geographies, like entire cities or, at best, county-level data. The sub-county, micro-level detail, such as neighborhood, zip code, census tract

or block group, that would be required to signal local emerging trends may be lost as geographical data is aggregated to reduce sampling error. ACS data releases provide more geographic granularity at the expense of precision over time, as, for example, sub-county-level geographies are only widely released to cover a 5-year timespan. A notable exception is US unemployment estimates, released monthly at various geographic levels as small as counties or cities of 25,000 people, and derived from the costly Current Population Survey done by the Bureau of Labor and Statistics.

Social syndromic systems offer the potential of early detection through less costly mechanisms. These systems, which have been developed at numerous public health departments in the US [10] and in countries around the world since [11], have proven their utility by detecting epidemics of certain diseases much sooner than traditional reporting methods. Syndromic surveillance systems have been successfully developed for various epidemic diseases, relying on varied data sources including emergency department chief complaints [12,13], ambulance dispatch records [14], hospital admission diagnosis [13], outpatient clinic visits [15], school/work absenteeism [16], over the counter medication sales [17] and web search queries [18].

There is high potential to use similar tools to detect social epidemics more quickly and at less cost. Social scientists have successfully used syndromic surveillance principles to retrospectively detect behaviors that are otherwise challenging to track, like consumer trends [19], tax avoidance [20], and e-cigarette uptake [21]. Others have found links to psychological illness [22] and unemployment [23].

While some of these examples demonstrate success at large geographic scales, such as state- or nationwide, more granular local data has also shown promise that these systems can be developed at smaller scales. Indeed, initial studies suggest that social service-related query data can be used for effective syndromic surveillance at local levels. For example, 311 data was found to be a better predictor of local crime rates than the "broken windows" theory of crime [24]. Though social scientists continue to debate the challenges and opportunities of 311 data [25], it has been shown to be useful especially when combined with other locally contextualized data [26]. Neighborhood disorder has also been found to correlate to geographic-specific data from 911 dispatches [27].

Integrating and validating existing social data streams therefore has significant potential to provide real time surveillance of SDoH. Such integration is an important complement to scholarship proposing three applications for the integration of SDoH into population health analytics: individual patient care, predictive risk modeling and community engagement [28].

Individual patients may be reluctant to offer potentially embarrassing socioeconomic information, like housing insecurity, to providers doing SDoH screening at the point of care [29]. More rapid and localized social syndromic surveillance can help inform SDoH screening by providing a rough pre-test probability in patient encounters where privacy and disclosure issues might inhibit point of care data collection. Indeed, localized surveillance data has successfully been shown to accurately revise pre-test probability in group A streptococcus pharyngitis, where higher proportions of local positive throat cultures change the threshold at which a clinician should offer empiric antibiotics [30].

At a community level, an early warning system based on SDoH can assist community advocates and clinical partners in mobilizing for increased resources and response well before such epidemics reveal themselves in data released less frequently. Similarly, data aggregated at flexible geographies can be helpful in raising hypotheses at the community level, allowing discussion of socioeconomic issues and to design early interventions. Moreover, risk-based payment models will likely push payees toward early detection and intervention of social epidemics related to SDoH to help prevent the accumulation of problems that later become both more difficult and expensive to address.

Despite the successes in infectious disease and bioterrorism epidemiology, there two key interrelated challenges to developing and deploying syndromic surveillance systems for SDoH: data challenges and associated costs.

Data Challenges

There are multiple potential data sources for SDoH syndromic surveillance, a number of which are publicly available. For example, Google search query data [31], public transportation or bike share program ridership data [32,33], 311 data and other government administrative data [34] are often publicly available. Social service-related query data offers probably the highest potential but presents a significant risk because proprietary data, held by private entities who can change their data access policies at will, already presents a challenge for infectious disease syndromic surveillance [35].

Still, communities and clinics are partnering with social resource databases to help link patients to resources during clinical SDoH screening, making these entities lynchpins in the framework to integrate SDoH into population health analytics [28]. As a result, query data from social resource databases, like Purple Binders, Aunt Bertha, Health Leads or 2-1-1 have significant potential if leveraged into a social syndromic surveillance system.

For example, the United Way's 2-1-1 program [36] has a mandate to provide financial, domestic, health or disaster-related information to callers free of charge. With over 240 systems in North America, the proliferation of city and state-wide 211s and other similar resource databases has created a wealth of data that can be used to combine the speed of syndromic surveillance with social epidemiology in an effort to provide a much more rapid indication of social epidemics. The Community Information Exchange in San Diego, which leverages 2-1-1 services, is an innovative example of the cross-sector collaboration required to leverage sufficient data for a SDoH surveillance system, and one that could, with modest additional effort, leverage their existing system for social epidemic detection [37].

Despite the applicability of such query data for detecting or predicting social epidemics, research is lacking. A major challenge has been that many of these data remain hidden in proprietary systems, limiting empirical studies. Additionally, different social resource database companies often outsource their database to different management companies with different architecture, making it time consuming and expensive to integrate data from various sources into one system. For example, different 2-1-1 systems contract with a handful of different private database

companies, making it challenging to integrate data for joint projects. The Alliance of Information and Referral Systems (AIRS) has set guidelines for resource database structures, which should facilitate data integration, but the current incentives for such private enterprises do not yet facilitate data sharing for public health or research purposes [38].

Data quality can also be an issue, especially when working with small geographies. Infectious disease syndromic surveillance presents some important examples. For example, Google Flu Trends, a tracking system designed to predict influenza, was initially heralded as a success before it was found to have significant flaws and was suspended [39]. However, these systems are responsive to improvements [40]. Google Flu Trend's problem may stem from using such generalized query data, known as a 'feature selection' problem [41] that requires researchers to assume a query for "influenza" means that someone may be ill rather than something else, like someone is doing a school report. The more specialized users and queries in resource databases avoids the "feature selection" problem, which can improve signal-to-noise ratios, suggesting that syndromic surveillance will be most successful when appropriate data sources are be found [42].

Moreover, concerns first raised about screening for SDoH also apply to syndromic surveillance, such as concerns about increasing stigma, the ecological fallacy, and the potential to reinforce inequalities [29,43].

Cost

Syndromic surveillance for SDoH faces at least two cost-related barriers. First, there is the cost to developing the system, including gathering the rights to relevant data and integrating and harmonizing disparate data sources into a unified platform. Second, there are potential costs to responding to the signals detected to the system that currently remain unknown.

Syndromic surveillance systems for infectious diseases often require access to proprietary data, and similar systems for SDoH will likely face the same challenge. For example, data from queries of resource databases like Purple Binder, Aunt Bertha, HealthLeads or the United Way's 2-1-1 has high potential to form a backbone in syndromic surveillance of social epidemics. However, most resource databases charge tens of thousands of dollars per clinic per year for their services, regardless of whether they are private or non-profit businesses. Gathering search query or use data from these organizations is likely to be expensive.

As medical providers and payers, as well as public health departments, take an increasing interest in SDoH surveillance, the cost of integrating disparate data sources to inform the project has been well documented [28]. Integrating data sources with particular relevance for syndromic surveillance will have a cost, but these costs can be kept low if these efforts can dovetail with the current efforts to build SDoH surveillance networks and leverage existing syndromic surveillance systems at public health departments.

The costs associated with epidemic response presents a different challenge. The USPSTF does not recommend clinical screening for depression unless there are sufficient regional resources to enable treatment [44]. The concern is that it may be unethical for a physician to screen for a disease

that they cannot treat. Similarly, creating syndromic surveillance systems for social determinants of health presents a similar quandary of who bears the ethical burden of responding to syndromic alerts, and what budgetary pressure will that put on already strained clinical and public health systems [4].

Despite the budgetary constraints, concerns are overblown. We do not yet know enough about social epidemics to preclude studying them. Much more empirical work needs to be done to set activation thresholds and discern which types of social epidemics are most amenable to detection through these systems and at which geographic levels – limitations that will be particular to the different data sets used to construct the time series. Such detection thresholds cannot be discerned in advance but must be empirically tested and revised in real time based on surveillance objectives, practical budgetary constraints, geographic limitations of the data and ethical concerns.

But learning lessons from infectious disease syndromic surveillance, such systems were similarly uncertain in their ability to accurately detect epidemic and bioterrorism events, with similar questions about alarm thresholds and response capacity [45]. Public health officials overcame these challenges by recognizing that the cost of responding late to an epidemic were likely higher than the cost of setting up early detection and response systems. Now such systems are nimble and flexible, allowing for near real-time algorithmic adjustments based on surveillance objectives (i.e. novel disease, bioterrorism, or natural disaster) and signal-to-noise thresholds [46,47]. However, similar analyses have not yet been undertaken for SDoH [48]. Now syndromic surveillance is an integral component of public health departments, with constant maintenance and testing to ensure data quality, integration, appropriate alarm thresholds and integration of new data sources – an iterative strategy that should guide the approach for SDoH [10].

The Way Forward

Clearly all of these challenges are interrelated. A lack of easily accessible data makes it difficult for researchers to combine sufficient time series data for empirical testing. Sufficient funds will likely overcome those barriers, but without promising empirical data, it is difficult to procure sufficient funds to surmount the data interoperability and ownership issues. In these settings, cross-sector collaboration has already shown much promise, but it needs much more support [49]. With current financial support, progress is likely to be slow, as effective research can only be carried out in environments with data that can be integrated easily into existing database systems.

Enough evidence already exists from the syndromic surveillance systems developed for infectious diseases and bioterrorism to suggest the capital required to build syndromic surveillance systems designed to rapidly detect changes in SDoH will be worth the investment. Therefore funders with mandates to address SDoH should consider early investments in creating data-sharing infrastructure and collaborations. For screening for SDoH is not enough; we can do more to detect social epidemics as they are starting to ensure the most rapid response to them as possible.

References

1. Berkman LF, Kawachi I, Glymour MM. Social epidemiology: Oxford University Press; 2014.

2. Henning KJ. 2004. What is syndromic surveillance? *MMWR Suppl*. 53, 5-11. PubMed

3. Braveman P, Egerter S, Williams DR. 2011. The social determinants of health: coming of age. *Annu Rev Public Health*. 32, 381-98 PubMed https://doi.org/10.1146/annurev-publhealth-031210-101218.

4. Garg A, Boynton-Jarrett R, Dworkin PH. 2016. Avoiding the unintended consequences of screening for social determinants of health. *JAMA*. 316(8), 813-14 PubMed https://doi.org/10.1001/jama.2016.9282.

5. Gottlieb L, Ackerman S, Wing H, Manchanda R. 2017. Understanding Medicaid managed care investments in members' social determinants of health. *Popul Health Manag*. 20(4), 302-08 PubMed https://doi.org/10.1089/pop.2016.0092.

6. Alderwick H, Hood-Ronick CM, Gottlieb LM. 2019. Medicaid Investments To Address Social Needs In Oregon And California. *Health Aff (Millwood)*. 38(5), 774-81 PubMed https://doi.org/10.1377/hlthaff.2018.05171.

7. Moore K, Belanger P. 2016. Surveillance Systems that Include Deprivation Indices & Social Determinants of Health. *Online J Public Health Inform*. 8(1), e182. PubMed https://doi.org/10.5210/ojphi.v8i1.6601.

8. Gamache R, Kharrazi H, Weiner JP. 2018. Public and population health informatics: the bridging of big data to benefit communities. *Yearb Med Inform*. 27(01), 199-206 PubMed https://doi.org/10.1055/s-0038-1667081.

9. Krieger N. 2001. A glossary for social epidemiology. *J Epidemiol Community Health*. 55(10), 693-700 PubMed https://doi.org/10.1136/jech.55.10.693.

10. Buehler JW, Sonricker A, Paladini M, Soper P, Mostashari F. 2008. Syndromic surveillance practice in the United States: findings from a survey of state, territorial, and selected local health departments. *Advances in Disease Surveillance.* 6(3), 1-20.

11. van Noort SP, Muehlen M, Rebelo de Andrade H, Koppeschaar C, Lima Lourenco J, et al. 2007. Gripenet: an internet-based system to monitor influenza-like illness uniformly across Europe. *Euro Surveill*. 12(7), E5-6. https://doi.org/10.2807/esm.12.07.00722-en

12. Irvin CB, Nouhan PP, Rice K. 2003. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity for bioterrorism and influenza surveillance. *Ann Emerg Med*. 41(4), 447-52 PubMed https://doi.org/10.1067/mem.2003.104.

13. Peterson D, Perencevich E, Harris A, Novak C, Davis S. 2003. Using existing electronic hospital data for syndromic surveillance. *J Urban Health*. 80, i122-23 https://doi.org/10.1007/BF02416908.

14. Greenko J, Mostashari F, Fine A, Layton M. 2003. Clinical evaluation of the Emergency Medical Services (EMS) ambulance dispatch-based syndromic surveillance system, New York City. *J Urban Health*. 80(Suppl 1), i50. PubMed

15. Marsden-Haug N, Foster VB, Gould PL, Elbert E, Wang H, et al. 2007. Code-based syndromic surveillance for influenzalike illness by International Classification of Diseases, Ninth Revision. *Emerg Infect Dis*. 13(2), 207 PubMed https://doi.org/10.3201/eid1302.060557.

16. Besculides M, Heffernan R, Mostashari F, Weiss D. 2005. Evaluation of school absenteeism data for early outbreak detection, New York City. *BMC Public Health*. 5(1), 105 PubMed https://doi.org/10.1186/1471-2458-5-105.

17. Vergu E, Grais RF, Sarter H, Fagot J-P, Lambert B, et al. 2006. Medication sales and syndromic surveillance, France. *Emerg Infect Dis*. 12(3), 416 PubMed https://doi.org/10.3201/eid1203.050573.

18. Hulth A, Rydevik G, Linde A. 2009. Web queries as a source for syndromic surveillance. *PLoS One*. 4(2), e4378 PubMed https://doi.org/10.1371/journal.pone.0004378.

19. Choi H, Varian H. 2012. Predicting the present with Google Trends. *Econ Rec*. 88(s1), 2-9 https://doi.org/10.1111/j.1475-4932.2012.00809.x.

20. Ayers JW, Ribisl K, Brownstein JS. 2011. Using search query surveillance to monitor tax avoidance and smoking cessation following the United States' 2009 "SCHIP" cigarette tax increase. *PLoS One*. 6(3), e16777 PubMed https://doi.org/10.1371/journal.pone.0016777.

21. Ayers JW, Ribisl KM, Brownstein JS. 2011. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *Am J Prev Med*. 40(4), 448-53 PubMed https://doi.org/10.1016/j.amepre.2010.12.007.

22. Ayers JW, Althouse BM, Allem J-P, Childers MA, Zafar W, et al. 2012. Novel surveillance of psychological distress during the great recession. *J Affect Disord*. 142(1-3), 323-30 PubMed https://doi.org/10.1016/j.jad.2012.05.005.

23. Baker SR, Fradkin A. The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data. Review of Economics and Statistics. 2014(0).

24. O'Brien DT, Sampson RJ, Winship C. 2015. Ecometrics in the age of big data: Measuring and assessing "broken windows" using large-scale administrative records. *Sociol Methodol*. 45(1), 101-47 https://doi.org/10.1177/0081175015576601.

25. White A, Trump K-S. 2018. The promises and pitfalls of 311 data. *Urban Aff Rev*. 54(4), 794-823 https://doi.org/10.1177/1078087416673202.

26. O'Brien DT. 2016. Using small data to interpret big data: 311 reports as individual contributions to informal social control in urban neighborhoods. *Soc Sci Res*. 59, 83-96 PubMed https://doi.org/10.1016/j.ssresearch.2016.04.009.

27. O'Brien DT, Sampson RJ. 2015. Public and private spheres of neighborhood disorder: Assessing pathways to violence using large-scale digital records. *J Res Crime Delinq*. 52(4), 486-510 https://doi.org/10.1177/0022427815577835.

28. Predmore Z, Hatef E, Weiner JP. 2019. Integrating social and behavioral determinants of health into population health analytics: a conceptual framework and suggested road map. *Popul Health Manag*. 22(6), 488-94 PubMed https://doi.org/10.1089/pop.2018.0151.

29. Gottlieb LM, Francis DE, Beck AF. 2018. Uses and Misuses of Patient- and Neighborhood-level Social Determinants of Health Data. *Perm J*. 22, 18-078. PubMed

30. Fine AM, Nizet V, Mandl KD. 2011. Improved Diagnostic Accuracy of Group A Streptococcal Pharyngitis With Use of Real-Time Biosurveillance. *Ann Intern Med*. 155(6), 345-52 PubMed https://doi.org/10.7326/0003-4819-155-6-201109200-00002.

31. Mohebbi M, Vanderkam D, Kodysh J, Schonberger R, Choi H, Kumar S. Google correlate whitepaper. 2011.

32. Fishman E. 2016. Bikeshare: A review of recent literature. *Transp Rev*. 36(1), 92-113 https://doi.org/10.1080/01441647.2015.1033036.

33. Welch TF, Widita A. 2019. Big data in public transportation: a review of sources and methods. *Transp Rev*. 39(6), 795-818 https://doi.org/10.1080/01441647.2019.1616849.

34. Connelly R, Playford CJ, Gayle V, Dibben C. 2016. The role of administrative data in the big data revolution in social science research. *Soc Sci Res*. 59, 1-12 PubMed https://doi.org/10.1016/j.ssresearch.2016.04.015.

35. Salathé M. 2018. Digital epidemiology: what is it, and where is it going? *Life Sci Soc Policy*. 14(1), 1 PubMed https://doi.org/10.1186/s40504-017-0065-7.

36. United Way. 2-1-1 Program 2017 [Available from: https://www.unitedway.org/our-impact/featured-programs/2-1-1.

37. Christenson C, Grounds K, Peter J. Collaboration and Cross-Sector Data Sharing to Create Healthier Communities. 2018.

38. Alliance of Information and Referral Systems. AIRS Standards and Quality Indicators for Professional Information and Referral 11240 Waples Mill Road Fairfax, Virginia 22030: AIRS; 2016.

39. Lazer D, Kennedy R, King G, Vespignani A. 2014. The parable of Google Flu: traps in big data analysis. *Science*. 343(6176), 1203-05 PubMed https://doi.org/10.1126/science.1248506.

40. Santillana M, Zhang DW, Althouse BM, Ayers JW. 2014. What can digital disease detection learn from (an external revision to) Google Flu Trends? *Am J Prev Med*. 47(3), 341-47 PubMed https://doi.org/10.1016/j.amepre.2014.05.020.

41. Antenucci D, Cafarella MJ, Levenstein M, Ré C, Shapiro M, eds. Ringtail: Feature Selection For Easier Nowcasting. WebDB; 2013.

42. Santillana M, Nsoesie EO, Mekaru SR, Scales D, Brownstein JS. 2014. Using clinicians' search query data to monitor influenza epidemics. *Nephrol Dial Transplant*. 59(10), 1446. PubMed

43. Gottlieb LM, Alderwick H. 2019. Integrating Social and Medical Care: Could it Worsen Health and Increase Inequity? *Ann Fam Med*. 17(1), 77-81 PubMed https://doi.org/10.1370/afm.2339.

44. Siu AL, Force UPST. 2016. Screening for depression in adults: US Preventive Services Task Force recommendation statement. *JAMA*. 315(4), 380-87 PubMed https://doi.org/10.1001/jama.2015.18392.

45. Buehler JW, Berkelman RL, Hartley DM, Peters CJ. 2003. Syndromic surveillance and bioterrorism-related epidemics. *Emerg Infect Dis*. 9(10), 1197-204 PubMed https://doi.org/10.3201/eid0910.030231.

46. Lall R, Abdelnabi J, Ngai S, Parton HB, Saunders K, Sell J, et al. Advancing the use of emergency department syndromic surveillance data, New York City, 2012-2016. Public Health Reports. 2017;132(1_suppl):23S-30S.

47. Thomas MJ, Yoon PW, Collins JM, Davidson AJ, Mac Kenzie WR. 2018. Evaluation of syndromic surveillance systems in 6 US state and local health departments. Journal of public health management and practice. *J Public Health Manag Pract*. 24(3), 235. PubMed

48. Chisolm DJ, Brook DL, Applegate MS, Kelleher KJ. 2019. Social determinants of health priorities of state Medicaid programs. *BMC Health Serv Res*. 19(1), 167 PubMed https://doi.org/10.1186/s12913-019-3977-5.

49. Mattessich PW, Rausch EJ. 2014. Cross-sector collaboration to improve community health: a view of the current landscape. *Health Aff (Millwood)*. 33(11), 1968-74 PubMedhttps://doi.org/10.1377/hlthaff.2014.0645.