Received: 9 May 2024 | Revised: 24 June 2024 | Accepted: 25 July 2024

DOI: 10.1002/lrh2.10445

COMMENTARY

Learning Health Systems

US public health surveillance, reimagined

Elina Guralnik

Department of Health Administration and Policy, College of Public Health, George Mason University, Fairfax, VA, USA

Correspondence
Elina Guralnik, Department of Health Administration and Policy, College of Public Health, George Mason University, Fairfax, VA, USA.
Email: eguralni@gmu.edu

Abstract

Introduction: This study presents two novel concepts for standardizing electronic health records (EHR)-based public health surveillance through utilization of existing informatics methods and data platforms.

Methods: Drawing from the collective experience in applied epidemiology, health services research and health informatics, the author presents a vision for an alternative path to public health surveillance by repurposing existing tools and resources, such as (1) computable phenotypes which have already been created and validated for a variety of chronic diseases of interest to public health and (2) large data platforms/collaboratives, such as All of Us Research Program and National COVID Cohort Collaborative. Opportunities and challenges are discussed regarding EHR-based chronic disease surveillance, as well as the concept of phenotype definitions and large data platforms reuse for public health needs.

Results/Framework: Reusing of computable phenotypes for EHR-based public health surveillance would require secure data platforms and nationally representative data. Standardization metrics for reuse of previously developed and validated computable phenotypes are also necessary and are currently being developed by the author. This study presents a reimagined Learning Health System framework by incorporating Public Health and two novel concept sets of solutions into the healthcare ecosystem.

Conclusion/Next Steps: Alternative approaches to limited resources and current infrastructure of the US Public Health System, especially as applied to disease surveillance, are needed and may be possible when repurposing the resources and methodologies across the Learning Health System.

KEYWORDS
computable phenotypes, electronic health records, public health surveillance

1 | INTRODUCTION

This study aims to introduce yet another perspective on how to advance population health surveillance, which is the cornerstone of public health, beyond traditional national surveillance surveys, which, most would agree, is beyond antiquated. Two novel concepts will address repurposing of existing informatics tools, such as computable phenotypes, and data platforms, such as All of Us Research Program and National COVID Cohort Collaborative (N3C) as alternatives to building new infrastructure or spending resources and funds on projects that may not be sustainable in the long run.

Both concepts were inspired by the idea of "cross-pollination" of methods and resources across multiple subdisciplines within the healthcare ecosystem, and adopting those methods to the public

health needs, desperately requiring modernization and repair. Furthermore, the concepts presented herein aim to expand our understanding of public health as another integral component, rather than a separate entity, of an already complex healthcare ecosystem, which we have come to define as a Learning Health System.

Between 2015 and 2023, close to 30 US studies, reviewed by the author, explored the idea of and validated the feasibility and function of EHR-based chronic disease surveillance systems, as either an alternative to or an extension of the traditional national surveillance surveys such as American Community Survey or National Health and Nutrition Examination Survey (NHANES).[1–13] Some studies were able to continue past the pilot phase, with the support and funding through their respective states and federal agencies (i.e., Massachusetts—MDPHnet), while others ceased to exist due to a variety of factors, ranging from the lack of continuous funding (i.e., NYC Macroscope) to depletion of resources, including valuable workforce, following COVID-19 pandemic, causing a pause or compete dissolution of seemingly successful surveillance programs (i.e., Colorado's CHORDS).[14,15] Most recent initiatives—MENDS, a CDC-funded pilot focusing on providing a framework for an establishment of a national syndromic disease surveillance system for chronic diseases, and MedMorph, another CDC initiative, funded by the National Patient-Centered Clinical Research Network (PCORnet), which focuses more on EHR data interoperability and standardization of extraction and exchange across EHR—deserve a special mention.[14,16] Both have addressed the needs and ways to modernize US public health infrastructure; one by utilizing the already existing infrastructure of more resource-rich health departments, while other delving into the interoperability across EHR themselves.[14,16] While each of these initiatives is valuable in their own ways, both would continue to involve significant investment in infrastructure, resources, and time to bring them to fruition beyond short-term initiatives. Most importantly, all past and current pilot programs and feasibility studies have exclusively relied on data marts, health information exchanges or other data partnerships which require ongoing organizational commitments for adequate funding and infrastructure support, both technical and workforce-related, from data providers. For example, a more established clinical research data network, such as PCORnet, could also be considered for use in public health surveillance on the national and state levels, with some of the PCORnet partner organizations already demonstrating successful implementation of EHR-based disease surveillance in select states, such as in Florida.[17–19] However, access for and governance of those entities which would serve less as data contributors and more as data users, that is, local health departments, would have to be modified to meet the needs of local jurisdictions and other prospective users of the network for public health purposes.

2 | PUBLIC HEALTH SURVEILLANCE—CHALLENGES AND OPPORTUNITIES

Public health surveillance is defined as "the regular collection, analysis, use, and sharing of data to prevent and control disease and injury."[16] Although public health includes three major components in its mission—Assessment, Assurance and Policy Development—disease surveillance of newly emerging and prevalent infectious and chronic diseases in a population is the cornerstone of public health.[20] The definition of public health surveillance remained essentially unchanged over the years.[21] However, the aspect of timeliness in dissemination of collected data for public health action remained limited by the resources and technology. Each year since the 1990s, tremendous resources have been allocated for national surveillance surveys and programs, such as NHANES and Behavioral Risk Factor Surveillance System, to name a few.[21] National health surveillance surveys have long been the gold standard for collection, analysis, and use of population health data to inform public and population health programs, discover new associated risk factors, and predict potential health outcomes.[22] However, the national health data collected through those surveys take at least 2–3 years to disseminate and act on through public health programs. While all surveys are routinely administered on nationally representative samples and present a valuable resource by design, breadth, and depth for learning about the health of the population, the data are not collected or analyzed in real time or near real time. Furthermore, with the advancement of smart phone technology and disappearance of phone land lines, one of the standard modes of conducting such surveys, the validity and robustness of data collection for large-scale national surveys must be called into question.[23] All this presents challenges to timeliness in identifying emerging health conditions, to having a meaningful and impactful public health response, and, consequently, to providing appropriate and timely allocation of resources. The adoption and availability of electronic health records (EHR), offering a massive accumulation of clinical, laboratory and sociodemographic data collected and stored within EHR systems, present an opportunity to finally consider EHR data as a valid source for public health surveillance.[1–14,24]

There are clear advantages of using EHR data for surveillance. Some of the main advantages which have not been adequately addressed with the traditional national surveys are: (1) *timeliness* in identifying health trends, disease burden and emerging patterns of concern in population health; (2) *cost-efficiency*—traditional surveys are expensive due to implementation and administrative costs; (3) *more targeted resource allocation* for public health and healthcare programs; (4) *timeliness in assessing* the effectiveness of public health interventions; and (5) *efficiency in identifying* disease incidence and prevalence estimates as opposed to traditional surveillance methods.

EHR-based disease surveillance also presents with many challenges. In a recent systematic review, Aliabadi et al. identified six distinct categories of challenges in development and implementation of EHR-based disease surveillance: policy and regulatory, technical, management, standardization, financial resources, and data quality.[25] The top challenges within each categories included (1) privacy, confidentiality and data security issues (policy and regulatory); (2) difficulties in accessing, cleaning, and analyzing unstructured EHR data (technical); (3) inadequate population coverage at the EHR system (management); (4) lack of or not widespread use of interoperability standards (standardization); (5) the need for considerable investment, time and resources (financial resources); and (6) poor data quality in EHRs (data quality).[25]

Similar to traditional survey data, EHR data are subject to selection biases. Whereas in EHR-based surveillance only those patients who are able to receive care would have their records available in the system, traditional population health surveys have raised increasing concerns with representativeness of data and external validity due to participants' selection biases as well as declining participation in such surveys in past decades.[26] Data quality issues, such as duplicate data and, especially, missing data, perhaps presents the bigger issue with EHR-based surveillance. Since the data collection from EHRs is less standardized than traditional survey methods, it could negatively impact the interpretation of EHR data and threaten the validity of the surveillance estimates and sensitivity of measured health outcomes. From the interoperability perspective, EHR systems that are developed and supported by different vendors (i.e., Epic, Cerner) are usually incompatible with one another and have no capacity for interoperability to share patient records across a spectrum of providers unless the providers are part of the same health network which operate one type of EHR system.[27] Additionally, inadequate population coverage in a given EHR system used for surveillance would challenge the representativeness of the population sample and could present yet another issue.[28]

Nonetheless, there is enough evidence pointing to the untapped opportunities to enhance and improve public health practice through more efficient surveillance methods with acquired timeliness, reduced cost and increased volume of data that has become available. Of course, more data do not necessarily mean more complete data or higher quality data. In fact, traditional surveys sometime offer more breadth of data in terms of diversity of topics and variables included in a survey, especially for conditions that are nonurgent or not as severe in manifestation.[29] However, with more outpatient and primary care physicians using EHRs and the availability of unstructured clinical history texts, the depth and breadth of data collected in EHR will continue to improve. Routine monitoring of newly emerging health trends, with targeted monitoring and collection of key health attributes of interest, as well as timely analyses and interpretation of collected data, all utilizing EHR data, present the way of the future.[30] Some would argue that EHR data must be augmented with other data sources, such as insurance claims data or even traditional survey data.[14,25] Validation of surveillance methods using a variety of data sources is needed to establish further advantages to augmentation of EHR data with other data. Nonetheless, the efficiency and effectiveness of EHR-based disease surveillance would largely depend on the quality of data, soundness of analytical tools used and clear sampling design to ensure representativeness of target population, among other factors. While the quality of EHR data is beyond one's control, significant progress can be made in advancing EHR-based surveillance with development and validation of novel analytical tools and appropriate sampling techniques.

3 | FRAMEWORK: PUBLIC HEALTH AS PART OF LEARNING HEALTH SYSTEM

Those of us trained in public health can attest to two old adages: "If it ain't broke, don't fix it" and "Why reinvent the wheel?" We are often trained to use what we have at our disposal, almost always strapped for budget or time to figure out the best solutions. However, against our best judgment and common sense, we often find ourselves in silos when it comes to innovation.[31] Inspired by the practical and cost-saving mindset and recognizing that healthcare research and practice, as an ecosystem, should build bridges across subdisciplines and schools of thoughts, the author proposes an expansion of the current Learning Health System framework to include Public Health. The expanded framework also introduces two novel concepts of repurposing tools and systems that have already been designed and used for the purposes other than public health. Furthermore, this framework builds on and draws from the work of Richesson et al. and, most recently, Rasmussen et al. and Brandt et al. for their championing ideas of shared resources, from creating searchable libraries of "explicitly defined phenotype definitions" to facilitate effective sharing and re-using of computable phenotype definitions to developing portability framework for successful implementation of phenotyping algorithms sharing across multiple data networks and domains.[32–34]

Figure 1 presents a diagram of a reimagined Learning Health System which incorporates Public Health as an integral part of the system, while showing the interconnectedness between five major elements of the health ecosystem—Patient-Consumer, Healthcare Practice, Health Data, Research and Public Health—and two novel concept sets of actionable solutions for knowledge exchange and sharing of already established methods (i.e., computable phenotypes) and data platforms (i.e., All of Us Research Program and/or N3C): (1) Research and Public Health—to identify, catalog and re-purpose previously developed and validated computable phenotypes for use in disease surveillance and (2) Patient-Consumer and Health Data—to identify, refocus and re-purpose existing data platforms, that is, All of Us Research Program and/or N3C, to use for public health surveillance. The reimagined framework, depicted in Figure 1 also incorporates 10 Essential Public Health Services[20] to highlight the connection between Public Health services and Learning Health System ecosystem.

3.1 | Repurposing existing informatics tools: Computable phenotypes and public health surveillance

The concept of electronic phenotyping has been widely used in bioinformatics for a long time, especially since the development of genomic research. Drawn from genetics terminology, the term "phenotype" simply refers to individual's observable characteristics "resulting from the interaction of one's genotype with the environment."[35] Phenotypes ascertained from EHR are sometimes referred to as "computable phenotypes" which are patient characteristics, clinical or others, which could be queried from the EHR as well as other electronic health data sources (i.e., insurance claims, mobile data) to identify patients with conditions of interest.[32,36] With the adoption of administrative codes for use in EHR data, such as the International Classification of Diseases and the Current Procedural Terminology codes for use in health insurance claims data, phenotyping originally heavily relied on the creation of rule-based algorithms, primarily based on diagnostic and billing

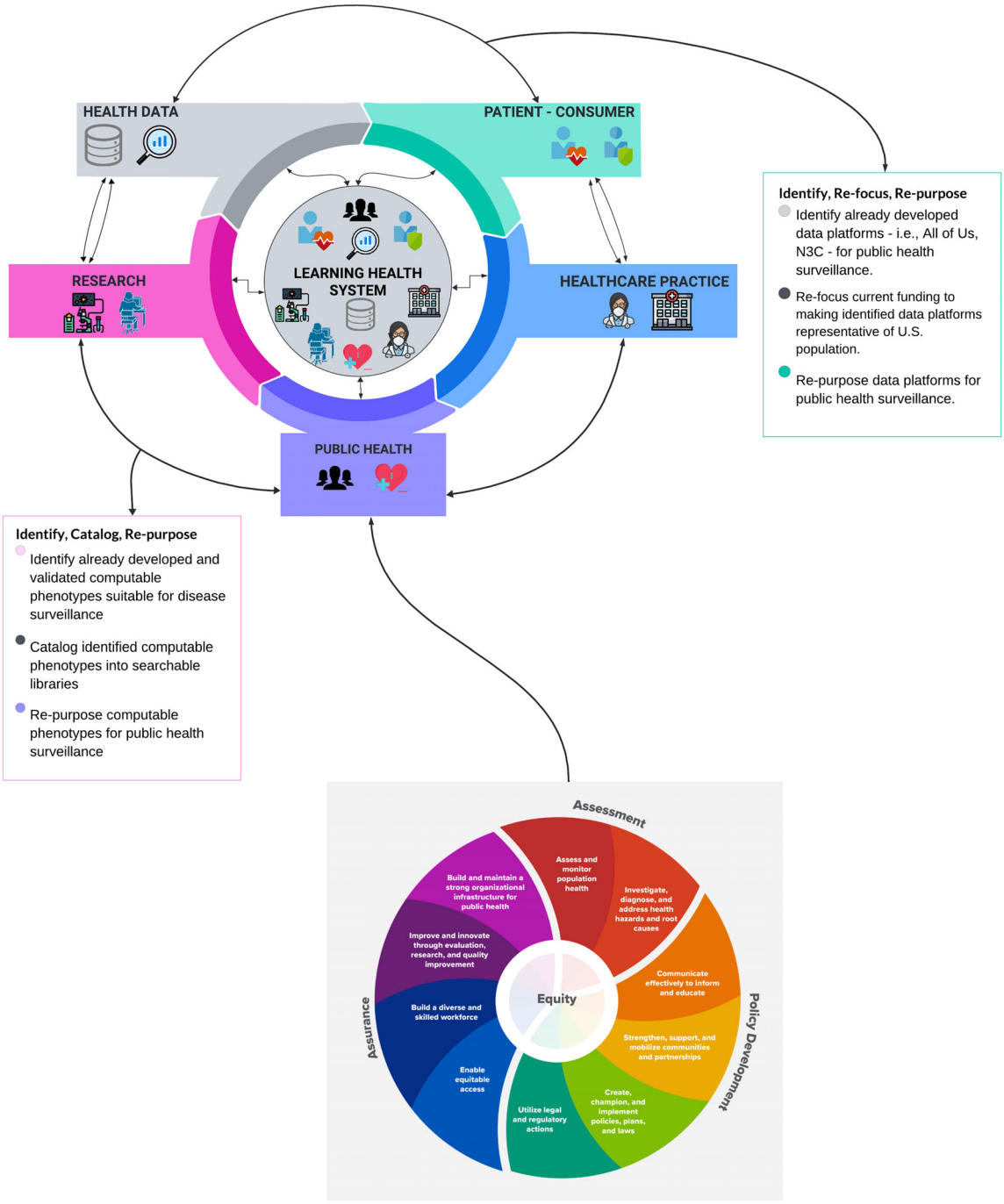

FIGURE 1 Learning Health System, reimagined.

codes, which in time evolved into including more complex data (i.e., unstructured clinical text data, laboratory results, medications, vital signs) and the use of machine learning, natural language processing, and deep learning methods.[35,37–41]

Electronic phenotyping allows for patient cohort identification based on common observable characteristics inferred from EHRs and has many applications. Rule-based or manual queries can identify cohort of patients for association studies, clinical trials, and large-scale pragmatic clinical trials, in which standardization measures, reliability, and reproducibility of developed phenotypes have been thoroughly addressed.[32,36] Cohort identification also finds its purpose in observational and interventional research as well as population management and quality measurement.[32,36] Whereas manual and rule-based phenotyping algorithms produce traditional binary estimates or cohort classifications, machine learning approaches to phenotyping affords probability estimates (i.e., probability of having the disease or disease progression), estimation of related parameters or even an estimation of severity of a disease.[42] Regardless of the approach, computable

phenotypes, as a means for cohort identification, have yet another application which has not been previously explored—EHR-based public health surveillance—for which one of the fundamental goals is to obtain disease estimates in a population. Thus, rule-based computable phenotypes can be used to find estimates of already diagnosed conditions, estimating the disease burden in a population, as the traditional surveillance surveys do, while prognostic phenotyping algorithms would identify expected disease burden based on population's health profile, using biomarkers and/or diagnostic codes of preexisting conditions, an approach more suitable for syndromic surveillance systems.

3.2 | Data standardization, phenotype validation and portability

The Common Data Model (CDM), championed by Observational Health Data Sciences and Informatics (OHDSI) collaborative, was conceived out of necessity to standardize observational health data, such as contained in EHRs, across multiple sources, which is often necessary for validation purposes, statistical power, and mitigation of missing and incomplete data.[43–45] Aside from standardizing the formatting and encoding of data content across institutions, the CDM was also designed to mitigate bias from local to external institutions or data sources as well as to improve the protection of patient health data by harmonizing it into a common standard.[45]

The role of CDM in implementing public health surveillance tasks is of particular importance, especially in the context of validation of computable phenotypes for disease surveillance purposes, to be able to sustain reproducibility of results, that is, disease prevalence estimates, and provide meaningful comparisons between results and concordance measures.[45] Presently, the validity of a rule-based phenotype is defined as the ability of the phenotype to correctly detect individuals with a condition of interest as well as those that do not have the condition of interest.[32,46,47] Estimation of a phenotype validity also requires concordance measures between disease estimates obtained from the existing gold standard, that is, traditional surveillance surveys, and those derived from applying cohort-identifying or phenotyping algorithms. Whereas in pragmatic clinical trials such gold standard would require a resource- and time-intensive chart reviewing of individual patient data to ascertain current and historical estimates of a cohort with a specific condition, the estimates from the traditional national surveillance surveys would serve as gold standard for accurate cohort identification in the context of disease surveillance. Though offering a definite advantage in validation of computable phenotypes, by cutting down on manual labor and time spent on chart reviewing, one has to remain cautious and acknowledge the vast differences between traditional survey and EHR data, whenever possible. This further warrants the necessity to continue testing how disease prevalence estimates, derived from the EHR using computable phenotypes would compare with those from the national surveillance surveys that US public health, clinical and healthcare delivery entities have long considered the gold standard.

Disease definitions as well as patient physiological states are not constant which makes it a challenge to assess disease estimates and trends with any accuracy after some time.[41,48] Hence, a modular approach to rule-based phenotyping, in which most constant data elements of the algorithms are identified and sustained within an algorithm, while other, more interchangeable elements that are likely to change over time, are replaced as needed, is essential for phenotyping to apply in disease surveillance. The same holds true for computable phenotypes in general,[32,34] so incorporating ways to preserve prior phenotype development and validation with an opportunity to adjust the algorithms based on new disease definitions or other factors is going to be crucial in application of phenotyping algorithms for public health surveillance purposes. Tracking changes of patient physiological states over time on a population level is outside the scope of this study but is, nonetheless, another important aspect in ensuring phenotype portability and robustness over time.

Locally and externally validated algorithms which could be transferrable from one data source to another, would have the potential for use in any disease surveillance system. Furthermore, the standardization of clinical data with the use of CDM would further improve external validation and portability of phenotyping algorithms, though CDM alone does not currently solve the interoperability issues since it is not yet widely adopted.[32,33] Nonetheless, the establishment of EHR-based disease surveillance systems on a national as well as local levels would largely depend on those three conditions—data standardization, phenotype validation, and phenotype transferability/portability. Moreover, standardization of computable phenotypes themselves, which could be used for disease surveillance, is necessary. Recent works by He et al. and Hohman et al. have acknowledged the gaps in standard definition of hypertension to be used for EHR-based disease prevalence estimates.[49,50] Both studies also discussed the lack of transparency in analytical decisions behind phenotype development.[49,50] In contrast to the aforementioned studies in which hypertension phenotypes were newly constructed[49] or refined[50] for EHR-based disease surveillance purposes, the author proposes to consolidate our collective knowledge and resources by evaluating phenotypes already in existence for their suitability for disease surveillance and to promote the use of standard disease definitions across the Learning Health System landscape.

3.3 | Practical considerations for re-purposing of existing data platforms: All of Us Research Program and National COVID Cohort Collaborative (N3C) reimagined

Two research data consortia were formed for specifically intended use—All of Us Research Program, to build one of the largest biomedical data sources, with special emphasis on research participants from historically underrepresented communities and backgrounds, and N3C, to establish a network of data collaborations during COVID-19, breaking the barriers of data access across institutions and to, collectively, be able to tackle the pandemic along with its immediate and long-term aftermath.[51,52] Both initiatives are unique in what they

tried to accomplish and how far they have come. All of Us focused on "direct to consumer" outreach, in which recruitment and participation have been extended to research participants with a special focus to empower people participating in research by giving back to them through informing them of study outcomes. With current volume of research, presently at 325 publications and over 10 000 active studies under way, All of Us is a promising data platform.[51] Nonetheless, the recent announcement about budget cuts of 2016 law, the 21st Century Cures Act, which initially secured the funding for All of Us for at least the next 10 years, is currently being reduced to 34% of last year's budget.[53] While it is believed that the participant recruitment will be the one most affected by the 30%–40% funding cuts in the next year, what if the recruitment of additional participants became more targeted beyond the original scope of collecting a large national sample of mostly traditionally underrepresented communities? For example, what if we dared to dream that one day this data platform, and others like it, may be the closest we could ever come to its participant EHR data representative of the US population. With infrastructure already in place, including different layers of protected data tiers, All of Us could potentially be the perfect solution to a number of impediments holding us back in our quest for a fully functional, nationally representative data repository with its EHR data being used for public health purposes, such as disease surveillance, among other things.

In contrast, the National COVID Cohort Collaborative (N3C) has accomplished something that could only be achieved in time of dire need, such as pandemic, in forging data collaborations and transcending institutional and governance barriers for more impactful and timely research. Furthermore, N3C was able to incorporate best practices from All of Us Research Program, already up and running for 2 years at the time, which points to one of many examples in which best practices of one initiative were successfully integrated into a new initiative.[52] With 98 sites contributing data to the N3C Enclave, 392 participating researcher organizations, 574 currently active projects, and 22.6 million persons contributing individual patient data, the number of EHR available from the past 4 years, since N3C's formation, and the enthusiasm with which this platform came to life warrants keeping that momentum going. Introducing additional initiatives and functions for the Enclave would not necessarily mean extra funding or resources. On the contrary, the author proposes to look at the Enclave holistically, from the perspective of a long-term public health and disease surveillance use beyond COVID-19. To some extent, the work on Long COVID already ventured out to chronic disease domain, employing machine learning models to predict Long COVID as well as using computable phenotyping for cohort identification of patients with post-COVID sequalae.[54–57]

The idea of redirecting the tremendous amount of work and resources that have already been put into All of Us and N3C, despite the original intent of these research platforms, is one of the potential avenues to explore for a more rapid and cost-efficient approach to EHR-based chronic disease surveillance. The expansion of these platforms into the public health arena, such that of the chronic disease surveillance using computable phenotyping, especially with data security, harmonization, and other important aspects already in place, makes them perfect candidates for a new mission. Moreover, the proposed harmonization of resources and tools will further contribute to the Learning Health System, in which every single element, including Public Health, is interconnected and, ideally, utilized to its full potential.

4 | CONCLUSION/NEXT STEPS

This study aimed to reimagine how we can use existing tools and resources to expand our definition and understanding of the Learning Health System, while advancing Public Health, depleted from the recent pandemic, the lack of proper infrastructure, and years of fiscal instabilities, to a stronger footing within the healthcare ecosystem.

Active research is underway by the author to establish standardization metrics for reusing previously developed and validated computable phenotypes to obtain chronic disease prevalence estimates of select chronic conditions using EHR data. If either or both of the data platforms, All of Us and N3C, redirected their efforts to achieve a nationally representative data sample suitable for surveillance and other needs, it would offer low-resource health departments an opportunity to obtain timely population disease estimates without significant investment in infrastructure and workforce. There is a real need to simplify the way we conduct EHR-based public health surveillance and utilize what has already been put in place. Though detailed analysis of suitability of All of Us and N3C for public health surveillance is outside the scope of this study, the amount of funds, partnerships and human capital infused into those data platforms merit an attempt to continue using them for additional applications, aside from clinical and biomedical research. Additionally, in lessons learned from the last pandemic, which will likely not be the last one we experience, novel pathogens are prone to wreak havoc with our detection, diagnosing, treatment and understanding of chronic health conditions as we have witnessed with Long COVID. So, it is imperative to be adequately prepared and work with what we have already built.

CONFLICT OF INTEREST STATEMENT

The author declares no conflicts of interest.

ORCID

Elina Guralnik https://orcid.org/0000-0002-7624-0620

REFERENCES

1. Guralnik E. Utilization of electronic health records for chronic disease surveillance: a systematic literature review. *Cureus*. 2023;15(4):e37975. doi:10.7759/cureus.37975
2. Chan PY, Perlman SE, Lee DC, Smolen JR, Lim S. Neighborhood-level chronic disease surveillance: utility of primary care electronic health records and emergency department claims data. *J Public Health Manag Pract*. 2022;28(1):E109-E118. doi:10.1097/phh.0000000000001142
3. Choi YG, Hanrahan LP, Norton D, Zhao YQ. Simultaneous spatial smoothing and outlier detection using penalized regression, with application to childhood obesity surveillance from electronic health records. *Biometrics*. 2022;78(1):324-336. doi:10.1111/biom.13404

4. Lucero-Obusan C, Oda G, Mostaghimi A, Schirmer P, Holodniy M. Public health surveillance in the U.S. Department of Veterans Affairs: evaluation of the Praedico surveillance system. *BMC Public Health*. 2022;22(1):272. doi:10.1186/s12889-022-12578-2
5. Mardon R, Campione J, Nooney J, et al. State-level metabolic comorbidity prevalence and control among adults age 50-plus with diabetes: estimates from electronic health records and survey data in five states. *Popul Health Metr*. 2022;20(1):22. doi:10.1186/s12963-022-00298-z
6. Salinas JJ, Sheen J, Shokar N, Wright J, Vazquez G, Alozie O. An electronic medical records study of population obesity prevalence in El Paso, Texas. *BMC Med Inform Decis Mak*. 2022;22(1):46. doi:10.1186/s12911-022-01781-1
7. Singh A, Dasgupta M, Retherford D, Baker M, Hulihan M, Brandow AM. Surveillance for the rare condition of sickle cell disease in Wisconsin. *WMJ*. 2022;121(4):297-300.
8. Frelinger C, Gardner RM, Huffman LC, Whitgob EE, Feldman HM, Bannett Y. Detection of speech-language delay in the primary care setting: an electronic health record investigation. *J Dev Behav Pediatr*. 2023;44(3):e196-e203. doi:10.1097/dbp.0000000000001167
9. Gu W, Tagg NT, Panchal NL, Brown-Bickerstaff CA, Nyman JM, Reynolds ME. Incidence of optic neuritis and the associated risk of multiple sclerosis for service members of U.S. armed forces. *Mil Med*. 2023;188(3–4):e697-e702. doi:10.1093/milmed/usab352
10. Kerchberger VE, Peterson JF, Wei WQ. Scanning the medical phenome to identify new diagnoses after recovery from COVID-19 in a US cohort. *J Am Med Inform Assoc*. 2023;30(2):233-244. doi:10.1093/jamia/ocac159
11. Negriff S, Lynch FL, Cronkite DJ, Pardee RE, Penfold RB. Using natural language processing to identify child maltreatment in health systems. *Child Abuse Negl*. 2023;138:106090. doi:10.1016/j.chiabu.2023.106090
12. Schumm M, Hu MY, Sant V, et al. Automated extraction of incidental adrenal nodules from electronic health records. *Surgery*. 2023;173(1):52-58. doi:10.1016/j.surg.2022.07.028
13. Singleton J, Li C, Akpunonu PD, Abner EL, Kucharska-Newton AM. Using natural language processing to identify opioid use disorder in electronic health record data. *Int J Med Inform*. 2023;170:104963. doi:10.1016/j.ijmedinf.2022.104963
14. Hohman KH, Martinez AK, Klompas M, et al. Leveraging electronic health record data for timely chronic disease surveillance: the multi-state EHR-based network for disease surveillance. *J Public Health Manag Pract*. 2023;29(2):162-173. doi:10.1097/phh.0000000000001693
15. Colorado Health Institute. Colorado Health Observation Regional Data Service (CHORDS). Accessed May 8, 2024. https://www.coloradohealthinstitute.org/research/CHORDS%20Institute
16. Michaels M, Syed S, Lober WB. Blueprint for aligned data exchange for research and public health. *J Am Med Inform Assoc*. 2021;28(12):2702-2706. doi:10.1093/jamia/ocab210
17. Filipp SL, Cardel M, Hall J, et al. Characterization of adult obesity in Florida using the OneFlorida clinical research consortium. *Obes Sci Pract*. 2018;4(4):308-317. doi:10.1002/osp4.274
18. Smith SM, McAuliffe K, Hall JM, et al. Hypertension in Florida: data from the OneFlorida clinical data research network. *Prev Chronic Dis*. 2018;15:E27. doi:10.5888/pcd15.170332
19. Cooper-Dehoff RM, Fontil V, Carton T, et al. Tracking blood pressure control performance and process metrics in 25 US health systems: the PCORnet blood pressure control laboratory. *J Am Heart Assoc*. 2021;10(21):e022224. doi:10.1161/jaha.121.022224
20. CDC Public Health Professionals Gateway. 10 Essential Public Health Services. Accessed May 5, 2024. https://www.cdc.gov/publichealthgateway/publichealthservices/essentialhealthservices.html
21. Thacker SB, Berkelman RL. Public health surveillance in the United States. *Epidemiol Rev*. 1988;10:164-190.
22. Paul MM, Greene CM, Newton-Dame R, et al. The state of population health surveillance using electronic health records: a narrative review. *Popul Health Manag*. 2015;18(3):209-216. doi:10.1089/pop.2014.0093
23. Hu SS, Balluz L, Battaglia MP, Frankel MR. The impact of cell phones on public health surveillance. *Bull World Health Organ*. 2010;88(11):799. doi:10.2471/BLT.10.082669
24. Kraus EMBB, Hohman KH, Baker EL. New directions in public health surveillance: using electronic health records to monitor chronic disease. *J Public Health Manag Pract*. 2022;28(2):203-206. doi:10.1097/PHH.0000000000001501
25. Aliabadi A, Sheikhtaheri A, Ansari H. Electronic health record-based disease surveillance systems: a systematic literature review on challenges and solutions. *J Am Med Inform Assoc*. 2020;27(12):1977-1986. doi:10.1093/jamia/ocaa186
26. Czajka JL, Beyler A. *Declining Response Rates in Federal Surveys: Trends and Implications (Background Paper)*. Accessed June 15, 2016. https://aspe.hhs.gov/sites/default/files/private/pdf/255531/Decliningresponserates.pdf
27. Snyder K. Cerner vs. epic EHR (2024 comparison). *Forbes Business*, Jan 5, 2024: https://www.forbes.com/advisor/business/software/cerner-vs-epic/#:~:text=However%2C%20both%20services%20do%20integrate,patient%20records%20across%20different%20providers
28. McVeigh KH, Newton-Dame R, Chan PY, et al. Can electronic health records be used for population health surveillance? Validating population health metrics against established survey data. *EGEMS*. 2016;4(1):1267. doi:10.13063/2327-9214.1267
29. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health*. 2016;37(1):61-81. doi:10.1146/annurev-publhealth-032315-021353
30. Groseclose SL, Buckeridge DL. Public health surveillance systems: recent advances in their use and evaluation. *Annu Rev Public Health*. 2017;38(1):57-79. doi:10.1146/annurev-publhealth-031816-044348
31. Acharya JC, Staes C, Allen KS, et al. Strengths, weaknesses, opportunities, and threats for the nation's public health information systems infrastructure: synthesis of discussions from the 2022 ACMI Symposium. *J Am Med Inform Assoc*. 2023;30:1011-1021. doi:10.1093/jamia/ocad059
32. Richesson R, Smerek M, Cameron CB. A framework to support the sharing and re-use of computable phenotype definitions across health care delivery and clinical research applications. *EGEMS*. 2016;4(3):2. doi:10.13063/2327-9214.1232
33. Rasmussen LV, Brandt PS, Jiang G, et al. Considerations for improving the portability of electronic health record-based phenotype algorithms. *AMIA Annu Symp Proc*. 2019;2019:755-764.
34. Brandt PS, Kho A, Luo Y, et al. Characterizing variability of electronic health record-driven phenotype definitions. *J Am Med Inform Assoc*. 2022;30(3):427-437. doi:10.1093/jamia/ocac235
35. Linder JE, Bastarache L, Hughey JJ, Peterson JF. The role of electronic health records in advancing genomic medicine. *Annu Rev Genomics Hum Genet*. 2021;22(1):219-238. doi:10.1146/annurev-genom-121120-125204
36. Richesson RL, Hammond WE, Nahm M, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *J Am Med Inform Assoc*. 2013;20(e2):e226-e231. doi:10.1136/amiajnl-2013-001926
37. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med Inform*. 2019;7(2):e12239. doi:10.2196/12239
38. Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. *J Allergy Clin Immunol*. 2020;145(2):463-469. doi:10.1016/j.jaci.2019.12.897
39. Zeng Z, Deng Y, Li X, Naumann T, Luo Y. Natural language processing for EHR-based computational phenotyping. *IEEE/ACM Trans*

Comput Biol Bioinform. 2019;16(1):139-153. doi:10.1109/tcbb.2018.2849968
40. Yang S, Varghese P, Stephenson E, Tu K, Gronsbell J. Machine learning approaches for electronic health records phenotyping: a methodical review. *J Am Med Inform Assoc*. 2023;30(2):367-381. doi:10.1093/jamia/ocac216
41. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc*. 2013;20(1):117-121. doi:10.1136/amiajnl-2012-001145
42. Hripcsak G, Albers DJ. High-fidelity phenotyping: richness and freedom from bias. *J Am Med Inform Assoc*. 2018;25(3):289-294. doi:10.1093/jamia/ocx110
43. Brown JS, Mendelsohn AB, Nam YH, et al. The US Food and Drug Administration sentinel system: a national resource for a learning health system. *J Am Med Inform Assoc*. 2022;29(12):2191-2200. doi:10.1093/jamia/ocac153
44. Kent S, Burn E, Dawoud D, et al. Common problems, common data model solutions: evidence generation for health technology assessment. *Pharmacoeconomics*. 2021;39(3):275-285. doi:10.1007/s40273-020-00981-9
45. The Book of OHDSI. Defining cohorts: chapter 4. Accessed March 11, 2024. https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html
46. Designing with Implementation and Dissemination in Mind. Introduction. *Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials*. NIH Health Care Systems Research Collaboratory; 2017. https://rethinkingclinicaltrials.org/chapters/conduct/electronic-health-records-based-phenotyping/evaluating-phenotype-definitions/
47. Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Inform Assoc*. 2013;20:e147-e154. doi:10.1136/amiajnl-2012-000896 PMID: 23531748.
48. National Library of Medicine. Finding and using health statistics. Health data sources: surveillance. Accessed March 11, 2024. https://www.nlm.nih.gov/oet/ed/stats/03-500.html
49. He S, Park S, Kuklina E, et al. Leveraging electronic health records to construct a phenotype for hypertension surveillance in the United States. *Am J Hypertens*. 2023;36(12):677-685. doi:10.1093/ajh/hpad081
50. Hohman KH, Zambarano B, Klompas M, et al. Development of a hypertension electronic phenotype for chronic disease surveillance in electronic health records: key analytic decisions and their effects. *Prev Chronic Dis*. 2023;20:E80. doi:10.5888/pcd20.230026
51. National Institutes of Health, All of Us Research Program: All of Us Research Hub. Accessed May 8, 2024. https://www.researchallofus.org/
52. Haendel MA, Chute CG, Bennett TD, et al. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. *J Am Med Inform Assoc*. 2021;28(3):427-443. doi:10.1093/jamia/ocaa196
53. Kaiser J. Major budget cuts to two high-profile NIH efforts leave research reeling: drop in 21st century cure act funding will slow BRAIN and all of us projects. *Science*, April 30, 2024. Accessed May 9, 2024 https://www.science.org/content/article/major-budget-cuts-two-high-profile-nih-programs-leave-researchers-reeling
54. Pfaff ER, Girvin AT, Bennett TD, et al. Identifying who has long COVID in the USA: a machine learning approach using N3C data. *Lancet Digit Health*. 2022;4(7):e532-e541. doi:10.1016/s2589-7500(22)00048-6
55. Hill EL, Mehta HB, Sharma S, et al. Risk factors associated with post-acute sequelae of SARS-CoV-2: an N3C and NIH RECOVER study. *BMC Public Health*. 2023;23(1):2103. doi:10.1186/s12889-023-16916-w
56. Reese JT, Blau H, Casiraghi E, et al. Generalisable long COVID subtypes: findings from the NIH N3C and RECOVER programmes. *EBioMedicine*. 2023;87:104413. doi:10.1016/j.ebiom.2022.104413
57. Pungitore S, Olorunnisola T, Mosier J, Subbian V. Computable phenotypes for post-acute sequelae of SARS-CoV-2: a national COVID cohort collaborative analysis. *AMIA Annu Symp Proc*. 2023;2023:589-598.

How to cite this article: Guralnik E. US public health surveillance, reimagined. *Learn Health Sys*. 2024;8(4):e10445. doi:10.1002/lrh2.10445