PLOS ONE

OVERVIEW

Facilitating cancer systems epidemiology research

Rolando Barajas[1¤a], Brionna Hair[2¤b], Gabriel Lai[1], Melissa Rotunno[1], Marissa M. Shams-White[1], Elizabeth M. Gillanders[1], Leah E. Mechanic[1] *

1 Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences (DCCPS), National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, Maryland, United States of America, **2** DCCPS, NCI, NIH, Bethesda, Maryland, United States of America

¤a Current address: Georgetown University School of Medicine, Washington, District of Columbia, United States of America
¤b Current address: Office of Strategic Coordination, Division of Program Coordination, Planning, and Strategic Initiatives, Office of the Director, NIH, Bethesda, Maryland, United States of America
* mechanil@mail.nih.gov

Check for updates

🔓 OPEN ACCESS

Abstract

Systems epidemiology offers a more comprehensive and holistic approach to studies of cancer in populations by considering high dimensionality measures from multiple domains, assessing the inter-relationships among risk factors, and considering changes over time. These approaches offer a framework to account for the complexity of cancer and contribute to a broader understanding of the disease. Therefore, NCI sponsored a workshop in February 2019 to facilitate discussion about the opportunities and challenges of the application of systems epidemiology approaches for cancer research. Eight key themes emerged from the discussion: transdisciplinary collaboration and a problem-based approach; methods and modeling considerations; interpretation, validation, and evaluation of models; data needs and opportunities; sharing of data and models; enhanced training practices; dissemination of systems models; and building a systems epidemiology community. This manuscript summarizes these themes, highlights opportunities for cancer systems epidemiology research, outlines ways to foster this research area, and introduces a collection of papers, "Cancer System Epidemiology Insights and Future Opportunities" that highlight findings based on systems epidemiology approaches.

Introduction

Epidemiology research has been successful at identifying many risk factors for complex diseases such as cancer, but much of the etiology remains unexplained. This may be due, in part, to the limited focus of many studies on a small number of risk factors or contributors to disease within specific domains or measures. Moreover, many studies fail to evaluate the complexities and interrelations among multiple risk factors on each other and the study outcomes. Each individual risk factor, such as a single dietary component or genetic polymorphism, occurs in a broader biological or societal context that may modulate the effect of individual risk factors on cancer. Many risk factors for cancer are also highly correlated with possible

interactive, additive, synergistic, or attenuating effects. Additionally, many risk factors are dynamic and time-varying: changes over the life course and the timing of exposure may modify cancer risk [1, 2].

Several groups have advocated for a more holistic or comprehensive analytic approach to the study of disease in populations [3–10]. This type of approach may lead to a better understanding of the mechanisms of disease and has been described by investigators using terminologies such as: eco-epidemiology [11], populomics [12], globolomics [13], systems medicine [14], and systems epidemiology [13, 15–18]. The term "systems epidemiology" conceptually borrows from fields such as systems biology, considering epidemiology research in a systems framework. For the purposes of this paper, we define systems epidemiology as an approach to study risk and outcomes that incorporates high-dimensional measurements from multiple domains, assesses the inter-relationships between risk factors, and considers changes over time. Our definition was adapted from Damman (2014) [17], however, we emphasize the importance of dynamism within a systems approach. Systems epidemiology research may leverage advanced computational simulation and modeling techniques to assess these complex networks and perform comprehensive analyses. Importantly, systems approaches are not limited to any single analytical method, but constitute a framework to account for complexity and understand the broader context of disease [7, 17, 19, 20].

Due to the complexity of cancer, a systems epidemiology approach may complement more traditional methods and lead to insights in disease etiology. To facilitate discussion about the application of systems modeling approaches for cancer epidemiology, NCI sponsored a workshop in February 2019 (https://epi.grants.cancer.gov/events/systems-epidemiology/) with presentations and discussions by experts in diverse fields to gain broader perspectives (S1 Table). In this manuscript, we summarize eight major themes from the workshop that will facilitate systems epidemiology research (Table 1) and discuss opportunities for this approach as exemplified by the accompanying papers in this PLOS Collection, "Cancer System Epidemiology Insights and Future Opportunities".

Themes to facilitate systems epidemiology research

1. Transdisciplinary collaboration and a problem-based approach

To more holistically study cancer, collaboration across disciplines is required. Traditionally, there has been a tendency when studying complex diseases for researchers to focus on data from individual disciplines. Focusing on a problem-based approach could bridge scientists across disciplines and integrate unique perspectives to improve understanding [21, 22]. Specifically, systems epidemiology would benefit from building linkages between disease content experts and computational modelers and informaticists who can build informed computational models.

There are several examples of how transdisciplinary collaborations and a problem-based approach can lead to scientific insights. In one transdisciplinary collaboration, researchers developed improved methods of differentiating between benign and aggressive cancer lesions [23]. Transdisciplinary collaboration also allows for methods to be developed and shared across fields. For example, dynamic agent-based modeling was developed for infectious disease modeling but now is commonly used in other complex disease analyses in public health [2, 24].

Several mechanisms, including previous NIH funding initiatives [21, 25–30], have encouraged transdisciplinary collaborations and problem-based approaches, which may serve as models to support this type of work. Another opportunity to foster transdisciplinary collaborations is to bring researchers together prior to applying for research funding. This type of

Table 1. Major themes identified to advance systems epidemiology research.

Theme	Description
1. Transdisciplinary Collaboration and a Problem-Based Approach	The ability to perform systems epidemiology research is contingent on the engagement of experts from varying fields to holistically address a scientific problem. Needs for transdisciplinary collaboration included: encouraging a focus on research problems holistically, bringing researchers together, addressing communication barriers, and sustaining transdisciplinary collaboration.
2. Methods and Modeling Considerations	Whether data-driven or hypothesis-driven, the overall methodology for systems epidemiology must incorporate an iterative approach where models evolve over time based on results. Several methods exist to apply systems modeling. Newer improved methods should incorporate changes over time, bridging multiple scales (e.g., cell, individual, and neighborhood), and dealing with unknown contributions of chance.
3. Interpretation, Validation, and Evaluation of Models	The complexity of systems models results in challenges for interpretation, validation, and evaluation. Comparative modeling, using common datasets or controls, and reproducibility pipelines are possible strategies to address these issues.
4. Data Needs and Opportunities	Despite numerous rich datasets in support of epidemiology research, data gaps remain. These gaps include the need for data from populations underrepresented in biomedical sciences, health behaviors, built environment, and health care provider information. Opportunities exist to leverage data from wearable devices, electronic health records, and large cohorts and initiatives. Challenges were noted regarding combining data from multiple sources and research domains.
5. Sharing of Data and Models	Promotion of systems epidemiology depends on the ability to share models and data. Effective sharing and reuse requires sufficient documentation and mechanisms to assess quality and support findability. Some mechanisms and infrastructures, including existing sharing platforms, could be leveraged to help address these needs.
6. Enhanced Training Practices	The evolving field of systems epidemiology will need to facilitate training for both students and current researchers in systems modeling, transdisciplinary research, data sciences, informatics, and computational modeling.
7. Dissemination of Systems Models	Successful dissemination depends on effective communication with content experts and the non-research community. Through direct engagement of various stakeholders, systems methods are more likely to be translated, utilized, and accepted to inform biological interpretations, interventions, or policies.
8. Building a Systems Community	Sustainability of systems epidemiology may depend on cultivating a systems epidemiology community. This can be facilitated by establishing organizations, interest groups, or other platforms for sharing ideas and discussing models. Specialized funding initiatives and review panels may further support systems epidemiology research.

process was used by the Cancer Research UK-NCI "Sandpit" workshop [31] and National Science Foundation ideas labs [32].

Challenges to be considered in developing transdisciplinary collaborations include sustainability and lack of a shared language among different scientific fields. Individuals with familiarity or training in multiple disciplines can serve as "translators" or "connectors" and facilitate interactions between distinct fields.

2. Methods and modeling considerations

Two complementary strategies were discussed to support systems epidemiology research: hypothesis-driven and data-driven strategies. In a hypothesis-driven strategy, researchers

focus on the data necessary to address a specified scientific hypothesis and analyze the data to test that hypothesis. In a data-driven strategy, the most likely hypothesis is identified based on a more agnostic, algorithm-based data exploration of several hypotheses. If the goal is to understand the overall mechanism, a hypothesis-driven strategy may be preferable. At the same time, given that a hypothesis-driven strategy is limited by current knowledge and assumptions, a data-driven strategy may gain new knowledge by finding unexpected relationships through a more agnostic approach.

The application of a systems approach to epidemiology questions should be considered an iterative process involving several steps, including identifying the problem, determining the model to test, obtaining the data, analyzing results, refining the model, and repeating as necessary based on the results. At times, these steps can occur concurrently. An example of how modeling was used to inform data collection was demonstrated by the Cancer Intervention and Surveillance Modeling Network (CISNET) breast consortium. Using simulation modeling, the CISNET teams examined the need for radiotherapy in women assessed as low risk based on genomic testing. These results were useful for informing the design of clinical trials by identifying those populations where data from the trial would be most informative [33].

Applying a systems approach to epidemiology research may be supported by several types of existing methods including systems dynamics, network analysis, agent-based modeling, and others [19, 24, 34–41] (S2 Table). Regardless of the specific analytical method, the unique aspect of a systems epidemiology approach is accounting for the complexity of the system by considering multiple domains, inter-relationships between risk factors, and dynamism. Needs for additional methods were noted, in particular for models incorporating time and space, dealing with the unknown contribution of chance, and bridging multiple scales (e.g., protein, cell, tissue, individual, neighborhood, community, ecosystem). It is critical for researchers to understand the underlying assumptions in methods and the strengths or weaknesses of particular models for different situations and questions. The variety of methods makes it challenging to interpret and compare results obtained using different approaches. Therefore, an important component of applying systems approaches is the validation and evaluation of models.

3. Interpretation, validation, and evaluation of models

As the complexity of a model increases with additional variables, the sparsity of data increases, thereby reducing the ability to make predictions or classifications. The large number of attributes in a model can also result in overfitting, which leads to biases in a model and makes it difficult to generalize or apply the model to another population. These issues with complex modeling are often referred to as the curse of dimensionality problem [42]. Furthermore, deep epidemiology data, including repeated measurements over time and assessments of multiple domains, is usually only available on smaller populations, limiting generalizability to other populations. Therefore, to adequately interpret the data and assess causality using these models, an iterative process is needed that includes well-designed validation and evaluation steps which will lead to model refinement and attenuation of these issues.

Model validation is the process of checking if all technical aspects (e.g., parameters definitions, coding, etc.) are done adequately or need refining and is preferably performed by an independent party [43]. Model evaluation is the process of assessing the performance and reproducibility of a complete model to discover its likelihood to perform in real world conditions (e.g., training and testing, cross-validation, etc.) [44]. To implement the validation and evaluation of their models, CISNET uses a comparative modeling approach where multiple research groups examine the same research questions using different models and identical predictors, and evaluate the results against real data trends. The consistency across models

provides support for model predictions [45]. When comparing results from complex models, care must be taken in the selection of the evaluation metric or the control used for comparison purposes. The dataset selected as a control could be biased in favor of the model being evaluated based on assumptions inherent in the model and the control dataset. Evaluating these complex models using common control datasets [46] could reduce this potential issue.

Given the complexity of methods for systems approaches, it may be challenging to reproduce all aspects of the analysis. One possible solution suggested at the workshop was to develop a reproducibility pipeline, or clear documentation for other researchers to apply the models on other populations. Notably, lack of reproducibility may also be due to intrinsic differences in the studied populations (e.g., by racial/ethnic or exposures distributions) [46]. Fortunately, there is guidance for validating and evaluating complex models [43, 47]. Further emphasis on the best practices for application of systems models to epidemiology research may help advance the use of these models in epidemiology studies.

4. Data needs and opportunities

Sufficient data (real and simulated) is required to effectively characterize a system [15]. Though workshop participants identified several potential data resources which could be used to support systems epidemiology research (S3 Table), gaps remain. One critical gap is the lack of inclusion of understudied groups, including racial and ethnic, socioeconomic, and geographic diversity and sexual/gender minorities [48, 49]. Insufficient racial and ethnic diversity is also apparent in genomics research [50, 51] and genomic catalogues [52, 53].

Other needs and opportunities identified by participants were for quality information about health behaviors, the built environment, and health care provider data. Systems epidemiology research could be enhanced by improving access or utilization of data sources such as: wearable devices (i.e. Fitbits), electronic health records [54], and large initiatives such as the All of Us cohort [55] and Environmental Influences on Child Health Outcomes (ECHO) Program [56]. As the understanding of a system develops and new hypotheses emerge, data needs may change. Collecting broad and multiple data types may enable the examination of multiple hypotheses without going back to data collection, which is particularly challenging in population-based studies. Such a strategy was used by the Community of Mine study [57]. Moreover, biobanks linked to medical record data provide another potential resource for systems epidemiology research [58] and could be leveraged to estimate risk factor or biomarker distributions in a target population missing that information [59, 60].

Characterizing the system requires combining or integrating several sources of data such as measures from different domains (e.g. genetics and behavioral) or spatial (e.g. cell to tissue) and temporal (e.g. day vs. year) scales. Often data is formatted uniquely, stored with different levels of metadata, or located in diverse databases. In fact, it was suggested that the resources required for integrating diverse, large-scale data types surpasses the resources required for generating these data [61]. For multi-omic data, several software frameworks have been developed to address some of these challenges, including Galaxy, Taverna, KNIME, and bioKepler [62]. Additional work is needed in this area.

5. Sharing of data and models

Improved methods for data linkages and model sharing across disciplines can facilitate systems epidemiology research by enabling a) analyses incorporating information from multiple domains; b) validation and evaluation of models and results; c) efficiency by avoiding duplication of efforts. Effective sharing and reuse of data and models requires adequate documentation (including metadata and descriptors) and mechanisms to assess quality and findability

[63], which can be costly. The NIH has worked to provide additional funding support for data and model sharing [64–67]. Moreover, according to the NIH Genomic Data Sharing Policy (GDS), costs for sharing of data should be included in the project budget [68]. Implementing carefully curated datasets or model resources and standardizing data quality indicators can increase confidence in methods and aid reproducibility and reusability. To address difficulty in finding the appropriate data set or method (i.e. findability), databases or resources that list and describe models are needed, such as the NCI Genetic Simulation Resources [69]. Several platforms and infrastructures were discussed that support sharing data and/or analytical models (Table 2).

6. Enhanced training practices

Participants noted that encouraging a systems framework for epidemiology research will require improvements in training, including more opportunities focused on systems science such as the *Systems Science for Social Impact* program [70]. Meeting participants suggested changes to the current epidemiology academic curriculum to incorporate systems training with more emphasis on complexity, transdisciplinary research, computational modeling, and informatics throughout the training continuum for epidemiologists. The current academic infrastructure is designed to develop scientists that are experts in specific fields/disciplines [22, 71], whereas systems approaches to epidemiology research require breadth of training across disciplines. Several training programs have supported this multidisciplinary model [72–74]. Another important training need is in the areas of data science, informatics, and computational modeling, particularly for population scientists. Training the next generation of data scientists and integrating these researchers into biomedical and public health fields is a priority within the NIH strategic plan for data science [75]. Finally, continuing education programs for epidemiologists [76], along the lines of Continuing Medical Education (CME) course work for physicians, could also broaden use of systems methods. Topics suggested for this type of training included advanced modeling techniques and managing and interpreting uncertainty in models.

7. Dissemination of systems models

For systems epidemiology modeling to be useful for research and policy, models and results using these methods need to be disseminated and accepted.

Stakeholders (e.g., patients, providers, payers, policy makers) should be involved early in model development to inform parameters and priorities. Incorporating stakeholder feedback can improve model quality by better defining the system and increase stakeholders' adoption of such models. Obtaining feedback on models as they are being developed through early publication may also lead to better models. However, journals may be reluctant to publish conceptual models in the absence of application results. A venue allowing for publication of early conceptual models could promote feedback (e.g., the preprint server bioRxiv https://www.biorxiv.org/).

Another key component to dissemination of models is effectively communicating models to the community for researchers, clinicians, policymakers, and the general public. Making the results interpretable regardless of model complexity would build confidence in the model and results [77]. Moreover, it is important to explain that uncertainty in the results remains even though systems models are sophisticated [78]. Effective communication could be enhanced by encouraging media training for scientists.

8. Building a systems epidemiology community

Growth in the application of systems approaches to epidemiology research will require building a community of systems epidemiology researchers. Workshop participants noted that the

Table 2. Example platforms supporting sharing of data and analytical models or methods.

Data Sharing Platforms/Solutions	Description
All of Us Research Workbench Link: https://www.researchallofus.org/workbench/	The All of Us cohort has a goal of enrolling one million participants in the United States to improve health sciences research. With the goal of oversampling for diverse participants, the data being collected includes various measures such as: health questionnaires, electronic health records (EHRs), physical measurements, the use of digital health technology, and biospecimens.
Database of Genotypes and Phenotypes (dbGaP) Link: https://www.ncbi.nlm.nih.gov/gap/	dbGaP stores and distributes results of studies that have investigated the association between genotypes and phenotypes.
NCI Cancer Research Data Commons (CRDC) Link: https://datascience.cancer.gov/data-commons	The NCI Cancer Research Data Commons (CRDC) is a cloud-based data science infrastructure that connects data sets with analytics tools to allow users to share, integrate, analyze, and visualize cancer research data.
Model Sharing platforms/Analysis platforms	**Description**
Bioconductor Link: https://www.bioconductor.org/	Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data.
Galaxy Link: https://usegalaxy.org/	Galaxy is a web-based platform that enables multi-omics data integration and analysis workflows.
Kepler project (and bioKepler) Link: https://kepler-project.org/	Kepler is designed to harmonize data by allowing scientists to create, execute, and share models and analyses across a broad range of scientific disciplines.
KNIME Link: https://www.knime.com/	KNIME is an analytics platform that supports data science workflows and reusable components.
Taverna Link: https://taverna.incubator.apache.org/introduction/	Taverna is a suite of tools used to design and execute scientific workflows.
Combined Analysis and Data platforms	**Description**
Biosphere Link: https://www.databiosphere.org/	Biosphere is an open-source platform developed by the Broad Institute that can operate across several different platforms (e.g., Terra, Gen3, and Dockstore) to create an interoperable data environment for the biomedical community.
Genomic Data Science Analysis, Visualization, and Informatics Lab-space (AnVIL) Link: https://www.genome.gov/Funded-Programs-Projects/Computational-Genomics-and-Data-Science-Program/Genomic-Analysis-Visualization-Informatics-Lab-space-AnVIL#overview	AnVIL is a scalable and interoperable resource for the genomic scientific community. It leverages a cloud-based infrastructure for genomic data access, sharing and computing across various data sets.
NCI Cloud Resources Link: https://datascience.cancer.gov/data-commons/cloud-resources	The NCI Cloud Resources are components of the NCI Cancer Research Data Commons that allow researchers to download, store, and analyze vast datasets in the cloud. The platform gives users access to tools and pipelines already implemented or lets them upload their own data or analytical methods to workspaces.
St. Jude Cloud-based Repository Link: https://www.stjude.cloud/	St. Jude Cloud is a large pediatric genomics dataset that offers a suite of unique analysis tools and visualizations. It supports access to more than 700 paired tumor/germline samples for common and rare pediatric cancers, sequenced as part of the Pediatric Cancer Genome Project (PCGP).
Terra Link: https://terra.bio/	Terra is a scalable and secure platform for biomedical researchers to access Broad hosted data, upload their own data, and combine data to run on analytic tools. The platform also contains functions that promote sharing of data to facilitate collaboration between scientists.

https://doi.org/10.1371/journal.pone.0255328.t002

current workshop was unique and expressed enthusiasm for bringing together researchers from disparate fields to address complex problems. In addition to periodic in-person meetings or workshops such as the one described by this paper, one strategy to build this type of community is to establish organizations, interest groups, or social platforms that can bring many different scientists together to share ideas and discuss and compare models, such as the Interdisciplinary Association for Public Health Science (IAPHS) [79]. Opportunities to promote cross-fertilization of ideas would be a systems epidemiology-focused journal, or a journal collection such as this one, where researchers from different disciplines can publish papers in this arena.

Another strategy to build a systems epidemiology community is through tailored grant reviews and specialized funding opportunities. Some meeting participants suggested that the non-linear design and the multidisciplinarity underlying complex modeling and systems approaches do not easily fit into the traditional three-aim structure of R01 applications, making it more challenging for this type of research to compete for funding. Special funding opportunities that support more complex projects and are amenable to non-linear aims, feedback loops and iterative processes may be helpful for this field. Special review panels for systems epidemiology applications could include reviewers from different disciplines, with at least one reviewer with computational modeling expertise, assigned to review each application.

Opportunities for systems epidemiology research

In addition to the eight themes highlighted which would facilitate systems epidemiology, several research opportunities that may be addressed using a systems epidemiology approach were discussed by workshop participants (Table 3).

One research opportunity that received substantial attention was to use systems approaches to help understand and alleviate health disparities. Complex social, behavioral, environmental, biological, and ecological contributions to disparities vary by context, impact multiple scales, and involve nonlinear and multidirectional associations (or feedback loops). The systems nature of health disparities may explain their persistence across different diseases. A systems approach may thus provide valuable insights into the etiology of disparities to highlight sources of inequities, identify data needs, and improve interventions [4].

These research opportunities and the papers in this "Cancer System Epidemiology Insights and Future Opportunities" collection illustrate the promise of systems epidemiology approaches. However, a portfolio analysis by Shams-White et al. found that despite specific systems and computational modeling funding announcements, the representation of systems epidemiology grants in cancer research remains low [80]. Together the above examples and these results suggest that many cancer-related research questions addressable using a systems approach may therefore benefit from tailored funding opportunities.

Conclusions and next steps

At the outset of the workshop, several participants expressed uncertainty about the definition of systems epidemiology. Nevertheless, there was overall agreement about the need for the general approach. Some participants suggested that it was important to emphasize the time element, or dynamism, within the definition as changes over time are critical to consider and are often missing in traditional studies. Others underlined the importance of data as the availability of high throughput data can help support more systems-based approaches.

To conclude, workshop participants supported a more comprehensive approach to population-based research studies and identified several considerations to facilitate the field of systems epidemiology. The workshop identified several themes or considerations for facilitating

Table 3. Example opportunities for a systems epidemiology approach in cancer research.

Major Opportunity Areas	Example Research Questions
Understanding the complexity of common risk factors: Risk factors that have remained elusive in their contribution to cancer etiology can be studied systematically.	• Study obesity via a systems approach to discover the dynamic (feedback/feedforward) role obesity (both child and adult) has on cancer etiology and survivorship. • The systemic effects of circadian rhythm disruptions on behaviors, organ physiology, and metabolism that could better explain cancer etiology. • Estimate the uncertainty of exposure measurements (i.e., contributed from measurement errors, high background rates for some contaminants, challenges in assessing mixtures) in causal relationships and better understand contribution to disease.
Integration of environmental/behavioral factors: Study the interaction between individual health behaviors and environments that increase cancer risk.	• Behaviors are often assessed individually and in absence of environmental context. However, a systems approach can support the evaluation of how factors like sexual behaviors, nutrition, tobacco usage, physical activity, sedentary behavior, circadian rhythm (sleep) disruptions, social networks, and infectious disease transmission work in tandem and vary within different environments (e.g., rural vs. urban settings) to contribute to cancer.
Health inequalities/health disparities: Evaluate how health inequalities are reflected throughout various biological, socioeconomic, and environmental layers.	• Certain groups may have a higher risk of certain cancers due to many factors such as stress, low access to care, education, environmental exposures, and genetics. • Disentangle the independent and interrelated contributions of genetic ancestry and socially defined race/ethnicity to cancer.
Improvement of cancer therapies: Incorporate social determinants of health and other high-dimensional population-based measurements from multiple domains (e.g., neighborhood pollution and physical activity) to inform differences in treatment response.	• Examine how dynamic social determinants (e.g., diet, lifestyle, environmental exposures, behaviors, etc.) work in conjunction with biological processes (e.g., the microbiome) to influence treatment responses.
Effective interventions and screening (i.e., design, predict, and evaluate interventions): Integrate knowledge about biology and the embedded context of individuals to provide personalized intervention strategies and screening programs.	• Utilize a systems approach to better prioritize at-risk populations and tailor interventions beyond a single behavior. • Evaluate and further improve interventions and screening programs over time in the specific environment/population of interest.
Policy Impacts: Examine the government/policy/institutional systemic impacts on cancer risk, and how policies may be improved via a systems approach.	• Use system modeling approaches to test a policy prior to implementation or examine impact of policy under different conditions. • Examine the impact of a policy change on an outcome, accounting for unforeseen consequences and feedback loops. • Evaluate the long-term impacts of a policy, accounting for changes in components of a system (e.g., changes in behaviors or movement of populations).

https://doi.org/10.1371/journal.pone.0255328.t003

systems epidemiology research and exemplified research opportunities. These themes included: transdisciplinary collaboration and a problem-based approach; methods and modeling considerations; interpretation, validation, and evaluation of models; data needs and opportunities; sharing of data and models; enhanced training practices; dissemination of systems models; and building a systems epidemiology community. As a first step to continue the conversation, several researchers participated in this collection of papers, outlining research opportunities and findings using systems epidemiology approaches. Our intent is that this collection will further spark discussion and foster continued research in this area.

Supporting information

S1 Table. Expertise of workshop participants.
(DOCX)

S2 Table. Example methods applicable for systems epidemiology.
(DOCX)

S3 Table. Example data resources applicable for systems epidemiology.
(DOCX)

References

1. Colditz GA, Wei EK. Preventability of cancer: the relative contributions of biologic and social and physical environmental determinants of cancer mortality. Annu Rev Public Health. 2012; 33:137–56. Epub 2012/01/10. https://doi.org/10.1146/annurev-publhealth-031811-124627 PMID: 22224878; PubMed Central PMCID: PMC3631776.

2. Hiatt RA, Porco TC, Liu F, Balke K, Balmain A, Barlow J, et al. A multilevel model of postmenopausal breast cancer incidence. Cancer Epidemiol Biomarkers Prev. 2014; 23(10):2078–92. Epub 2014/07/16. https://doi.org/10.1158/1055-9965.EPI-14-0403 PMID: 25017248.

3. Burke TA, Cascio WE, Costa DL, Deener K, Fontaine TD, Fulk FA, et al. Rethinking Environmental Protection: Meeting the Challenges of a Changing World. Environ Health Perspect. 2017; 125(3):A43–a9. Epub 2017/03/02. https://doi.org/10.1289/EHP1465 PMID: 28248180; PubMed Central PMCID: PMC5332174.

4. Diez Roux AV. Complex systems thinking and current impasses in health disparities research. Am J Public Health. 2011; 101(9):1627–34. Epub 2011/07/23. https://doi.org/10.2105/AJPH.2011.300149 PMID: 21778505; PubMed Central PMCID: PMC3154209.

5. Hu FB. Metabolic profiling of diabetes: from black-box epidemiology to systems epidemiology. Clin Chem. 2011; 57(9):1224–6. Epub 2011/06/22. https://doi.org/10.1373/clinchem.2011.167056 PMID: 21690202.

6. Lee BY, Bartsch SM, Mui Y, Haidari LA, Spiker ML, Gittelsohn J. A systems approach to obesity. Nutr Rev. 2017; 75(suppl 1):94–106. Epub 2017/01/05. https://doi.org/10.1093/nutrit/nuw049 PMID: 28049754; PubMed Central PMCID: PMC5207008.

7. Lich KH, Ginexi EM, Osgood ND, Mabry PL. A call to address complexity in prevention science research. Prev Sci. 2013; 14(3):279–89. Epub 2012/09/18. https://doi.org/10.1007/s11121-012-0285-2 PMID: 22983746.

8. Mabry PL, Kaplan RM. Systems science: a good investment for the public's health. Health Educ Behav. 2013; 40(1 Suppl):9s–12s. Epub 2013/10/23. https://doi.org/10.1177/1090198113503469 PMID: 24084406.

9. Orr MG, Kaplan GA, Galea S. Neighbourhood food, physical activity, and educational environments and black/white disparities in obesity: a complex systems simulation analysis. J Epidemiol Community Health. 2016; 70(9):862–7. Epub 2016/04/17. https://doi.org/10.1136/jech-2015-205621 PMID: 27083491.

10. Weed DL. Beyond black box epidemiology. Am J Public Health. 1998; 88(1):12–4. Epub 1998/05/16. https://doi.org/10.2105/ajph.88.1.12 PubMed Central PMCID: PMC1508377. PMID: 9584017

11. Susser M, Susser E. Choosing a future for epidemiology: I. Eras and paradigms. Am J Public Health. 1996; 86(5):668–73. Epub 1996/05/01. https://doi.org/10.2105/ajph.86.5.668 PubMed Central PMCID: PMC1380474. PMID: 8629717

12. Gibbons MC. Populomics. Stud Health Technol Inform. 2008; 137:265–8. Epub 2008/06/19. PMID: 18560087.

13. Lund E, Dumeaux V. Systems epidemiology in cancer. Cancer Epidemiol Biomarkers Prev. 2008; 17 (11):2954–7. Epub 2008/11/08. https://doi.org/10.1158/1055-9965.EPI-08-0519 PMID: 18990736.

14. Nielsen J. Systems Biology of Metabolism: A Driver for Developing Personalized and Precision Medicine. Cell Metab. 2017; 25(3):572–9. Epub 2017/03/09. https://doi.org/10.1016/j.cmet.2017.02.002 PMID: 28273479.

15. Cerdá M, Keyes KM. Systems Modeling to Advance the Promise of Data Science in Epidemiology. Am J Epidemiol. 2019; 188(5):862–5. Epub 2019/03/17. https://doi.org/10.1093/aje/kwy262 PMID: 30877289; PubMed Central PMCID: PMC6494667.

16. Cornelis MC, Hu FB. Systems Epidemiology: A New Direction in Nutrition and Metabolic Disease Research. Curr Nutr Rep. 2013; 2(4). Epub 2013/11/28. https://doi.org/10.1007/s13668-013-0052-4 PMID: 24278790; PubMed Central PMCID: PMC3837346.

17. Dammann O, Gray P, Gressens P, Wolkenhauer O, Leviton A. Systems Epidemiology: What's in a Name? Online J Public Health Inform. 2014; 6(3):e198. Epub 2015/01/20. https://doi.org/10.5210/ojphi.v6i3.5571 PMID: 25598870; PubMed Central PMCID: PMC4292535.

18. Haring R, Wallaschofski H. Diving through the "-omics": the case for deep phenotyping and systems epidemiology. Omics. 2012; 16(5):231–4. Epub 2012/02/11. https://doi.org/10.1089/omi.2011.0108 PMID: 22320900; PubMed Central PMCID: PMC3339382.

19. Luke DA, Stamatakis KA. Systems science methods in public health: dynamics, networks, and agents. Annu Rev Public Health. 2012; 33:357–76. Epub 2012/01/10. https://doi.org/10.1146/annurev-publhealth-031210-101222 PMID: 22224885; PubMed Central PMCID: PMC3644212.

20. McGuire S. Institute of Medicine. 2012. Accelerating progress in obesity prevention: solving the weight of the nation. Washington, DC: the National Academies Press. Adv Nutr. 2012; 3(5):708–9. Epub 2012/09/18. https://doi.org/10.3945/an.112.002733 PMID: 22983849; PubMed Central PMCID: PMC3648752.

21. Patterson RE, Colditz GA, Hu FB, Schmitz KH, Ahima RS, Brownson RC, et al. The 2011–2016 Transdisciplinary Research on Energetics and Cancer (TREC) initiative: rationale and design. Cancer Causes Control. 2013; 24(4):695–704. Epub 2013/02/05. https://doi.org/10.1007/s10552-013-0150-z PMID: 23378138; PubMed Central PMCID: PMC3602225.

22. Pohl C, Hirsch Hadorn G. Principles for designing transdisciplinary research. Munich: Oekom Verlag; 2007.

23. Frankhauser DE, Jovanovic-Talisman T, Lai L, Yee LD, Wang LV, Mahabal A, et al. Spatiotemporal strategies to identify aggressive biology in precancerous breast biopsies. Wiley Interdiscip Rev Syst Biol Med. 2020:e1506. Epub 2020/10/02. https://doi.org/10.1002/wsbm.1506 PMID: 33001587.

24. Tracy M, Cerdá M, Keyes KM. Agent-Based Modeling in Public Health: Current Applications and Future Directions. Annu Rev Public Health. 2018; 39:77–94. Epub 2018/01/13. https://doi.org/10.1146/annurev-publhealth-040617-014317 PMID: 29328870; PubMed Central PMCID: PMC5937544.

25. Bian J, Xie M, Topaloglu U, Hudson T, Eswaran H, Hogan W. Social network analysis of biomedical research collaboration networks in a CTSA institution. J Biomed Inform. 2014; 52:130–40. Epub 2014/02/25. https://doi.org/10.1016/j.jbi.2014.01.015 PMID: 24560679; PubMed Central PMCID: PMC4136998.

26. Bures RM, Mabry PL, Orleans CT, Esposito L. Systems science: a tool for understanding obesity. Am J Public Health. 2014; 104(7):1156. Epub 2014/05/17. https://doi.org/10.2105/AJPH.2014.302082 PMID: 24832433; PubMed Central PMCID: PMC4056227.

27. Hammond RA. Complex systems modeling for obesity research. Prev Chronic Dis. 2009; 6(3):A97. Epub 2009/06/17. PMID: 19527598; PubMed Central PMCID: PMC2722404.

28. Llewellyn N, Carter DR, DiazGranados D, Pelfrey C, Rollins L, Nehl EJ. Scope, Influence, and Interdisciplinary Collaboration: The Publication Portfolio of the NIH Clinical and Translational Science Awards (CTSA) Program From 2006 Through 2017. Eval Health Prof. 2020; 43(3):169–79. Epub 2019/03/29. https://doi.org/10.1177/0163278719839435 PMID: 30917690; PubMed Central PMCID: PMC7781230.

29. Mabry PL, Olster DH, Morgan GD, Abrams DB. Interdisciplinarity and systems science to improve population health: a view from the NIH Office of Behavioral and Social Sciences Research. Am J Prev Med. 2008; 35(2 Suppl):S211–24. Epub 2008/08/23. https://doi.org/10.1016/j.amepre.2008.05.018 PMID: 18619402; PubMed Central PMCID: PMC2587290.

30. Schmitz KH, Gehlert S, Patterson RE, Colditz GA, Chavarro JE, Hu FB, et al. TREC to WHERE? Transdisciplinary Research on Energetics and Cancer. Clin Cancer Res. 2016; 22(7):1565–71. Epub 2016/01/23. https://doi.org/10.1158/1078-0432.CCR-14-1214 PMID: 26792261; PubMed Central PMCID: PMC4956346.

31. National Cancer Institute. NCI-CRUK Sandpit Workshops 2019 [cited 2020 5/27/2020]. Available from: https://cancercontrol.cancer.gov/brp/hbrb/sandpit.html.

32. Collins T, Kearney M, Maddison D. The Ideas Lab Concept, Assembling the Tree of Life, and AVAToL. PLoS Curr. 2013; 5. Epub 2013/09/21. https://doi.org/10.1371/currents.tol.0fdb85e1619f313a2a5a2ec3d7a8df9e PMID: 24045602; PubMed Central PMCID: PMC3770768.

33. Jayasekera J, Schechter CB, Sparano JA, Jagsi R, White J, Chapman JW, et al. Effects of Radiotherapy in Early-Stage, Low-Recurrence Risk, Hormone-Sensitive Breast Cancer. J Natl Cancer Inst. 2018; 110(12):1370–9. Epub 2018/09/22. https://doi.org/10.1093/jnci/djy128 PMID: 30239794; PubMed Central PMCID: PMC6292790.

34. Fallah-Fini S, Adam A, Cheskin LJ, Bartsch SM, Lee BY. The Additional Costs and Health Effects of a Patient Having Overweight or Obesity: A Computational Model. Obesity (Silver Spring). 2017; 25

(10):1809–15. Epub 2017/09/28. https://doi.org/10.1002/oby.21965 PMID: 28948718; PubMed Central PMCID: PMC5679120.

35. Jones AP, Homer JB, Murphy DL, Essien JD, Milstein B, Seville DA. Understanding diabetes population dynamics through simulation modeling and experimentation. Am J Public Health. 2006; 96(3):488–94. Epub 2006/02/02. https://doi.org/10.2105/AJPH.2005.063529 PMID: 16449587; PubMed Central PMCID: PMC1470507.

36. Langellier BA, Bilal U, Montes F, Meisel JD, Cardoso LO, Hammond RA. Complex Systems Approaches to Diet: A Systematic Review. Am J Prev Med. 2019; 57(2):273–81. Epub 2019/07/22. https://doi.org/10.1016/j.amepre.2019.03.017 PMID: 31326011; PubMed Central PMCID: PMC6650152.

37. Lee BY, Adam A, Zenkov E, Hertenstein D, Ferguson MC, Wang PI, et al. Modeling The Economic And Health Impact Of Increasing Children's Physical Activity In The United States. Health Aff (Millwood). 2017; 36(5):902–8. Epub 2017/05/04. https://doi.org/10.1377/hlthaff.2016.1315 PMID: 28461358; PubMed Central PMCID: PMC5563819.

38. Lee BY, Ferguson MC, Hertenstein DL, Adam A, Zenkov E, Wang PI, et al. Simulating the Impact of Sugar-Sweetened Beverage Warning Labels in Three Cities. Am J Prev Med. 2018; 54(2):197–204. Epub 2017/12/19. https://doi.org/10.1016/j.amepre.2017.11.003 PMID: 29249555; PubMed Central PMCID: PMC5783749.

39. Li Y, Lawley MA, Siscovick DS, Zhang D, Pagán JA. Agent-Based Modeling of Chronic Diseases: A Narrative Review and Future Research Directions. Prev Chronic Dis. 2016; 13:E69. Epub 2016/05/30. https://doi.org/10.5888/pcd13.150561 PMID: 27236380; PubMed Central PMCID: PMC4885681.

40. Powell-Wiley TM, Wong MS, Adu-Brimpong J, Brown ST, Hertenstein DL, Zenkov E, et al. Simulating the Impact of Crime on African American Women's Physical Activity and Obesity. Obesity (Silver Spring). 2017; 25(12):2149–55. Epub 2017/11/01. https://doi.org/10.1002/oby.22040 PMID: 29086471; PubMed Central PMCID: PMC5705259.

41. Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, et al. A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nat Biotechnol. 2017; 35(8):747–56. Epub 2017/07/18. https://doi.org/10.1038/nbt.3870 PMID: 28714965; PubMed Central PMCID: PMC5568837.

42. Altman N, Krzywinski M. The curse(s) of dimensionality. Nat Methods. 2018; 15(6):399–400. Epub 2018/06/02. https://doi.org/10.1038/s41592-018-0019-x PMID: 29855577.

43. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—7. Value Health. 2012; 15(6):843–50. Epub 2012/09/25. https://doi.org/10.1016/j.jval.2012.04.012 PMID: 22999134.

44. Khurana Y. Medium [Internet]2019. Available from: https://medium.com/yogesh-khuranas-blogs/difference-between-model-validation-and-model-evaluation-1a931d908240.

45. Alagoz O, Berry DA, de Koning HJ, Feuer EJ, Lee SJ, Plevritis SK, et al. Introduction to the Cancer Intervention and Surveillance Modeling Network (CISNET) Breast Cancer Models. Med Decis Making. 2018; 38(1_suppl):3s–8s. Epub 2018/03/20. https://doi.org/10.1177/0272989X17737507 PMID: 29554472; PubMed Central PMCID: PMC5862043.

46. Mechanic LE, Chen HS, Amos CI, Chatterjee N, Cox NJ, Divi RL, et al. Next generation analytic tools for large scale genetic epidemiology studies of complex diseases. Genet Epidemiol. 2012; 36(1):22–35. Epub 2011/12/08. https://doi.org/10.1002/gepi.20652 PMID: 22147673; PubMed Central PMCID: PMC3368075.

47. 2020 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation. [Internet]. 2020 [cited 2020 5/6/2020]. Available from: http://sbp-brims.org/2020/cfp/.

48. Martin DN, Lam TK, Brignole K, Ashing KT, Blot WJ, Burhansstipanov L, et al. Recommendations for Cancer Epidemiologic Research in Understudied Populations and Implications for Future Needs. Cancer Epidemiol Biomarkers Prev. 2016; 25(4):573–80. Epub 2016/05/20. https://doi.org/10.1158/1055-9965.EPI-15-1297 PMID: 27196089; PubMed Central PMCID: PMC4874661.

49. Swerdlow AJ, Harvey CE, Milne RL, Pottinger CA, Vachon CM, Wilkens LR, et al. The National Cancer Institute Cohort Consortium: An International Pooling Collaboration of 58 Cohorts from 20 Countries. Cancer Epidemiol Biomarkers Prev. 2018; 27(11):1307–19. Epub 2018/07/19. https://doi.org/10.1158/1055-9965.EPI-18-0182 PMID: 30018149.

50. Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. Nature. 2011; 475(7355):163–5. Epub 2011/07/15. https://doi.org/10.1038/475163a PMID: 21753830; PubMed Central PMCID: PMC3708540.

51. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016; 538(7624):161–4. Epub 2016/10/14. https://doi.org/10.1038/538161a PMID: 27734877; PubMed Central PMCID: PMC5089703.

52. Iorio A, De Angelis F, Di Girolamo M, Luigetti M, Pradotto LG, Mazzeo A, et al. Population diversity of the genetically determined TTR expression in human tissues and its implications in TTR amyloidosis. BMC Genomics. 2017; 18(1):254. Epub 2017/03/25. https://doi.org/10.1186/s12864-017-3646-1 PMID: 28335735; PubMed Central PMCID: PMC5364715.

53. Spratt DE, Chan T, Waldron L, Speers C, Feng FY, Ogunwobi OO, et al. Racial/Ethnic Disparities in Genomic Sequencing. JAMA Oncol. 2016; 2(8):1070–4. Epub 2016/07/02. https://doi.org/10.1001/jamaoncol.2016.1854 PMID: 27366979; PubMed Central PMCID: PMC5123755.

54. Pendergrass SA, Crawford DC. Using Electronic Health Records To Generate Phenotypes For Research. Curr Protoc Hum Genet. 2019; 100(1):e80. Epub 2018/12/06. https://doi.org/10.1002/cphg.80 PMID: 30516347; PubMed Central PMCID: PMC6318047.

55. Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, Jenkins G, et al. The "All of Us" Research Program. N Engl J Med. 2019; 381(7):668–76. Epub 2019/08/15. https://doi.org/10.1056/NEJMsr1809937 PMID: 31412182.

56. Environmental Influences on Child Health Outcomes (ECHO) Program 2020 [cited 2020 3/16/2020]. Available from: https://www.nih.gov/research-training/environmental-influences-child-health-outcomes-echo-program.

57. Jankowska MM, Sears DD, Natarajan L, Martinez E, Anderson CAM, Sallis JF, et al. Protocol for a cross sectional study of cancer risk, environmental exposures and lifestyle behaviors in a diverse community sample: the Community of Mine study. BMC Public Health. 2019; 19(1):186. Epub 2019/02/15. https://doi.org/10.1186/s12889-019-6501-2 PMID: 30760246; PubMed Central PMCID: PMC6375220.

58. Wang K, Gaitsch H, Poon H, Cox NJ, Rzhetsky A. Classification of common human diseases derived from shared genetic and environmental determinants. Nat Genet. 2017; 49(9):1319–25. Epub 2017/08/08. https://doi.org/10.1038/ng.3931 PMID: 28783162; PubMed Central PMCID: PMC5577363.

59. Franks PW, McCarthy MI. Exposing the exposures responsible for type 2 diabetes and obesity. Science. 2016; 354(6308):69–73. Epub 2016/11/16. https://doi.org/10.1126/science.aaf5094 PMID: 27846494.

60. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vich Vila A, Võsa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet. 2019; 51 (4):600–5. Epub 2019/02/20. https://doi.org/10.1038/s41588-019-0350-x PMID: 30778224; PubMed Central PMCID: PMC6441384.

61. Palsson B, Zengler K. The challenges of integrating multi-omic data sets. Nat Chem Biol. 2010; 6 (11):787–9. Epub 2010/10/27. https://doi.org/10.1038/nchembio.462 PMID: 20976870.

62. Boekel J, Chilton JM, Cooke IR, Horvatovich PL, Jagtap PD, Käll L, et al. Multi-omic data analysis using Galaxy. Nat Biotechnol. 2015; 33(2):137–9. Epub 2015/02/07. https://doi.org/10.1038/nbt.3134 PMID: 25658277.

63. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016; 3:160018. Epub 2016/03/16. https://doi.org/10.1038/sdata.2016.18 PMID: 26978244; PubMed Central PMCID: PMC4792175 Academic Editor and consultant.

64. Department of Health and Human Services. Research Supplements to Promote Data Sharing in Cancer Epidemiology Studies (Admin Supp Clinical Trial Not Allowed) 2018. Available from: https://grants.nih.gov/grants/guide/pa-files/PA-18-748.html.

65. Department of Health and Human Services. Notice of Special Interest (NOSI): Administrative Supplements to Support Enhancement of Software Tools for Open Science 2020 [cited 2020 5/7/2020]. Available from: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-073.html.

66. Department of Health and Human Services. Biomedical Data Repository (U24—Clinical Trials Not Allowed) 2020 [cited 2020 7/8/2020]. Available from: https://grants.nih.gov/grants/guide/pa-files/PAR-20-089.html.

67. Department of Health and Human Services. Biomedical Knowledgebase (U24—Clinical Trials Not Allowed) 2020 [cited 2020 7/8/2020]. Available from: https://grants.nih.gov/grants/guide/pa-files/PAR-20-097.html.

68. Paltoo DN, Rodriguez LL, Feolo M, Gillanders E, Ramos EM, Rutter JL, et al. Data use under the NIH GWAS data sharing policy and future directions. Nat Genet. 2014; 46(9):934–8. Epub 2014/08/28. https://doi.org/10.1038/ng.3062 PMID: 25162809; PubMed Central PMCID: PMC4182942.

69. Peng B, Leong MC, Chen HS, Rotunno M, Brignole KR, Clarke J, et al. Genetic Simulation Resources and the GSR Certification Program. Bioinformatics. 2019; 35(4):709–10. Epub 2018/08/14. https://doi.org/10.1093/bioinformatics/bty666 PMID: 30101297; PubMed Central PMCID: PMC6378936.

70. Washington University in St. Louis. Systems Science for Social Impact: Summer Training Institute 2021 [cited 2021 5/31/2021]. Available from: https://systemsscienceforsocialimpact.wustl.edu/.

71. Golembiewski EH, Holmes AM, Jackson JR, Brown-Podgorski BL, Menachemi N. Interdisciplinary Dissertation Research Among Public Health Doctoral Trainees, 2003–2015. Public Health Rep. 2018; 133 (2):182–90. Epub 2018/02/14. https://doi.org/10.1177/0033354918754558 PMID: 29438623; PubMed Central PMCID: PMC5871143.

72. James AS, Gehlert S, Bowen DJ, Colditz GA. A Framework for Training Transdisciplinary Scholars in Cancer Prevention and Control. J Cancer Educ. 2015; 30(4):664–9. Epub 2014/12/17. https://doi.org/ 10.1007/s13187-014-0771-2 PMID: 25510368; PubMed Central PMCID: PMC4469633.

73. Krause K. 2019. [cited 2020]. Available from: https://asunow.asu.edu/20191010-bold-restructuring-asu-college-health-solutions-results-growth-and-innovation.

74. Realmuto L, Daniel S, Jasani F, Weiss L, Bachrach C. Developing population health scientists: Findings from an evaluation of the Robert Wood Johnson Foundation Health & Society Scholars Program. SSM Popul Health. 2019; 7:100373. Epub 2019/02/28. https://doi.org/10.1016/j.ssmph.2019.100373 PMID: 30809585; PubMed Central PMCID: PMC6374691.

75. National Institutes of Health. NIH Strategic Plan for Data Science. 2018.

76. Bensyl DM, King ME, Greiner A. Applied Epidemiology Training Needs for the Modern Epidemiologist. Am J Epidemiol. 2019; 188(5):830–5. Epub 2019/03/17. https://doi.org/10.1093/aje/kwz052 PMID: 30877297; PubMed Central PMCID: PMC6608580.

77. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018; 15(141). Epub 2018/04/ 06. https://doi.org/10.1098/rsif.2017.0387 PMID: 29618526; PubMed Central PMCID: PMC5938574.

78. Chatterjee N. Towards Data Science [Internet]2020. [cited 2020]. Available from: https:// towardsdatascience.com/transparency-reproducibility-and-validity-of-covid-19-projection-models-78592e029f28.

79. Interdisciplinary Association for Population Health Science (IAPHS). Available from: https://iaphs.org/ about-iaphs/.

80. Shams-White MM, Barajas R, Jensen RE, Rotunno M, Dueck H, Ginexi EM, et al. Systems epidemiology and cancer: A review of the National Institutes of Health extramural grant portfolio 2013–2018. PLoS One. 2021; 16(4):e0250061. Epub 2021/04/16. https://doi.org/10.1371/journal.pone.0250061 PMID: 33857240; PubMed Central PMCID: PMC8049352.