

PERSPECTIVE OPEN



Linking complex disease and exposure data—insights from an environmental and occupational health study

Cataia Ives¹✉, Huaqin Pan¹, Stephen W. Edwards¹, Mark Nelms¹, Hannah Covert², Maureen Y. Lichtveld², Emily W. Harville³, Jeffrey K. Wickliffe⁴, Wilco Zijlmans⁵ and Carol M. Hamilton¹

© The Author(s) 2022

The disparate measurement protocols used to collect study data are an intrinsic barrier to combining information from environmental health studies. Using standardized measurement protocols and data standards for environmental exposures addresses this gap by improving data collection quality and consistency. To assess the prevalence of environmental exposures in National Institutes of Health (NIH) public data repositories and resources and to assess the commonality of the data elements, we analyzed clinical measures and exposure assays by comparing the Caribbean Consortium for Research in Environmental and Occupational Health study with selected NIH environmental health resources and studies. Our assessment revealed that (1) environmental assessments are widely collected in these resources, (2) biological assessments are less prevalent, and (3) NIH resources can help identify common data for meta-analysis. We highlight resources to help link environmental exposure data across studies to support data sharing. Including NIH data standards in environmental health research facilitates comparing and combining study data, and the use of NIH resources and adoption of standard measures will allow integration of multiple studies and increase the scientific impact of individual studies.

Keywords: Epidemiology; Health studies; Child exposure/health; Children's health; Vulnerable populations

Journal of Exposure Science & Environmental Epidemiology (2023) 33:12–16; <https://doi.org/10.1038/s41370-022-00428-7>

INTRODUCTION

Environmental epidemiological studies collect a variety of data, including both survey and exposure data, on environmental exposures and associated health outcomes. A systematic review of studies on barriers to public health data sharing identified a lack of standard protocols for data collection across studies as an important technical obstacle to translating research findings into public health interventions [1]. This analysis examines the use of common exposure measures across selected National Institutes of Health (NIH)-funded environmental epidemiologic studies and provides recommendations to promote the utility of environmental epidemiologic studies by advancing the use of standard protocols, particularly those focused on examining exposures to chemical and non-chemical stressors.

The rationale for our analysis is multi-pronged. First, statistical power and scientific efficiency would be maximized by combining data from multiple studies and thus increasing their effect size, but standardizing the various environmental exposure data collection measurement protocols and linking similar variables is a fundamental challenge [2].

Second, environmental health research relies on data-driven semantic standards for exposure science, centered on characterizing the interactions of a receptor (i.e., an individual or human population) with one or more environmental chemical and non-chemical stressors [3].

Third, exposure data may be expensive to collect, and studies may evaluate exposures in ways that can be difficult to combine across studies. The field of exposure science has identified the need to link information on toxicity to real-world outcomes and to use exposure data for chemical prioritization.

Environmental exposures are complex, encompassing a variety of domains and study types. The “exposome” concept, defined as the totality of exposures experienced by an individual during their life and the health impacts of those exposures [4], strives to capture the diversity and range of ecosystem, social, physical, chemical, and lifestyle exposures [5]. The complexity of exposome data and the need to increase the scale of exposome studies require data homogeneity to allow data merging and integration. The public health exposome represents a further refinement of the original exposome concept, integrating exogenous and endogenous exposures across the lifespan [6]. For environmental epidemiological studies, which typically include exposure and survey data, standards for annotating and identifying common data elements promote data reuse [7]. Therefore, establishing and adopting common data elements makes more data sources available for data integration using analysis and fusion models [8].

NIH environmental health resources

To assess the use of environmental exposures, we compiled several NIH resources relevant to environmental exposures to

¹RTI International, Research Triangle Park, NC, USA. ²School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA. ³Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA. ⁴School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. ⁵Faculty of Medical Sciences, Anton de Kom University of Suriname, Paramaribo, Suriname. ✉email: cives@rti.org

Received: 29 November 2021 Revised: 9 March 2022 Accepted: 10 March 2022

Published online: 28 March 2022

assess the potential impact of using standard measures. We chose the following resources because they contain substantial environmental health study data or measures, providing clear opportunities to link study variables and support collaborative analyses.

Children's/Human Health Exposure Analysis Resources (CHEAR/HHEAR) are centralized networks of exposure analysis tools, services, and expertise to support NIH-funded researchers studying human health [9]. The HHEAR Data Center (<https://heardatacenter.mssm.edu/>) provides a repository of data and laboratory analysis results for CHEAR and HHEAR studies and provides statistical and data analysis services to external researchers.

NIH's **Environmental Influences on Child Health Outcomes (ECHO)** is one of the largest environmental health programs in terms of outcome and exposure measures and number of participants [10]. ECHO supports multiple longitudinal studies using existing study populations to investigate environmental exposures—including physical, chemical, biological, social, behavioral, and natural exposures and the effects of built environments—on child health and development [11]. The ECHO-wide Cohort Data Collection Protocol includes assessments of a rich set of environmental exposures, which can serve as a resource for environmental health research [12].

PhenX (consensus measures for Phenotypes and Exposures) is an NIH common data elements project driven by the research community. It provides tools to help investigators incorporate recommended measurement protocols into their studies (<https://www.phenxtoolkit.org/>). PhenX contains standard measures across 29 research domains to date, including survey questionnaires, clinical examination, medical records abstraction, and bioassay protocols. The PhenX Environmental Exposures domain covers such topics as residential and occupational history, early-life exposures, environmental contaminants, and specific sample collections, providing a useful starting framework of standard measures to investigate environmental contributors to complex diseases [13].

The **Database of Genotypes and Phenotypes (dbGaP)** is a public data repository for a variety of NIH-funded studies, including genotype and phenotype data (<https://www.ncbi.nlm.nih.gov/gap/>) [14]. Currently, it hosts more than 1800 studies from 22 NIH institutes and centers, including 15 studies from National Institute of Environmental Health Sciences (NIEHS). As one of the largest public data repositories, studies in dbGaP have traditionally focused on genetic associations with human diseases and conditions but have recently included exposure data.

METHODS

Assess environmental exposure measures among NIH resources

We compiled environmental exposure measures from four NIH resources (ECHO, CHEAR/HHEAR, PhenX, and dbGaP) to identify overlap and gaps. Exposure measures were differentiated by mode of collection into "Environmental Assessment" (e.g., "Pets in Household") and "Biological Assessment" (e.g., "PAHs") groups. Here, *environmental assessments*, such as "Characteristics of Current Residence," refer to direct exposures as measured by interviewer-administered, self-report, or laboratory analysis protocols. *Biological assessments*, such as "Biomarker of exposure to nicotine-containing products—Urine," refer to chemical assays correlated with specific environmental exposures.

The four resources used different environmental assessment categories, which were harmonized for this analysis. We used CHEAR/HHEAR study descriptions from the data center [9] and classes in the HHEAR Ontology (<http://purl.bioontology.org/ontology/HHEAR>) to annotate the HHEAR studies to exposure categories (see Supplementary Table S1 for detail). We then used keywords extracted from PhenX environmental exposure

protocols to map ECHO data elements. We used the dbGaP Advanced Search feature to identify ECHO environmental assessments present in dbGaP studies. Buckley et al. described existing or planned chemical biomonitoring in either ECHO mother or child cohorts for 15 broad chemical classes, including both well-characterized and emerging chemicals [10]. Using these biological assessments from ECHO and keywords extracted from PhenX environmental exposure protocols, we searched the dbGaP [14] data repository. Biological assessments from ECHO were also mapped to HHEAR analytes, which are targeted analyses assessed by HHEAR Lab Hubs [9].

Comparison between CCREOH and ECHO studies' exposure measures

We compiled the clinical measures and exposure assays from the Caribbean Consortium for Research in Environmental and Occupational Health (CCREOH) study to compare them with those in the ECHO-wide Data Collection Protocol [15]. We conducted mapping between CCREOH health assessments for topics such as health history, depression, perceived stress, and exposure history [15] and ECHO-wide Cohort Data Collection Protocols from the ECHO project site [12]. To identify exposures measured, we conducted mapping between ECHO chemical assays and CCREOH biomarkers.

DISCUSSION

Assessment of environmental exposure measures in NIH resources

In this study, we compiled environmental exposure measures and compared their presence in four resources: ECHO, CHEAR/HHEAR, PhenX, and dbGaP. Figure 1 shows environmental exposure measures from ECHO and compares their presence and absence in four NIH resources. There are a number of environmental assessments in dbGaP studies (Fig. 1a) and biological assessments in ECHO (Fig. 1b); however, biological assessment data are quite limited in other resources (see more details in the Supplementary Information).

Many environmental health studies include more survey data than environmental measures because environmental data are costly to collect, and statistical concerns regarding small effect sizes further hamper comprehensive environmental epidemiologic study designs. Given these key challenges, using resources to link with other data sets becomes invaluable because it helps individual studies find additional environmental data to enhance their statistical power for hypothesis testing. Creating linkages among measures from PhenX, ECHO, HHEAR, and dbGaP makes it easier to identify similarities and differences among exposure categories in the four resources. Individual environmental health studies such as CCREOH could leverage these NIH resources to identify study data sets for comparative analysis. For example, seven ECHO cohorts collected metals or metalloids assay data [16], corresponding with eight HHEAR analytes, 5 PhenX protocols, and 33 dbGaP studies that collected metals or metalloids (Table S2). This approach provides a benchmark for use of environmental exposure measures in NIH-funded research and potentially serves as a framework for standard environmental exposure measures for prospective studies and data harmonization with retrospective studies, both types of which are present in ECHO. This analysis demonstrates a possible approach to combining data from different but conceptually related studies.

Opportunities for increasing CCREOH study impact

We chose existing studies that provided clear opportunities for linking study variables and that would support collaborative analysis of cross-cutting research questions related to environmental exposures, human health, and disease. We first examined the studies published by the NIEHS Deepwater Horizon Research

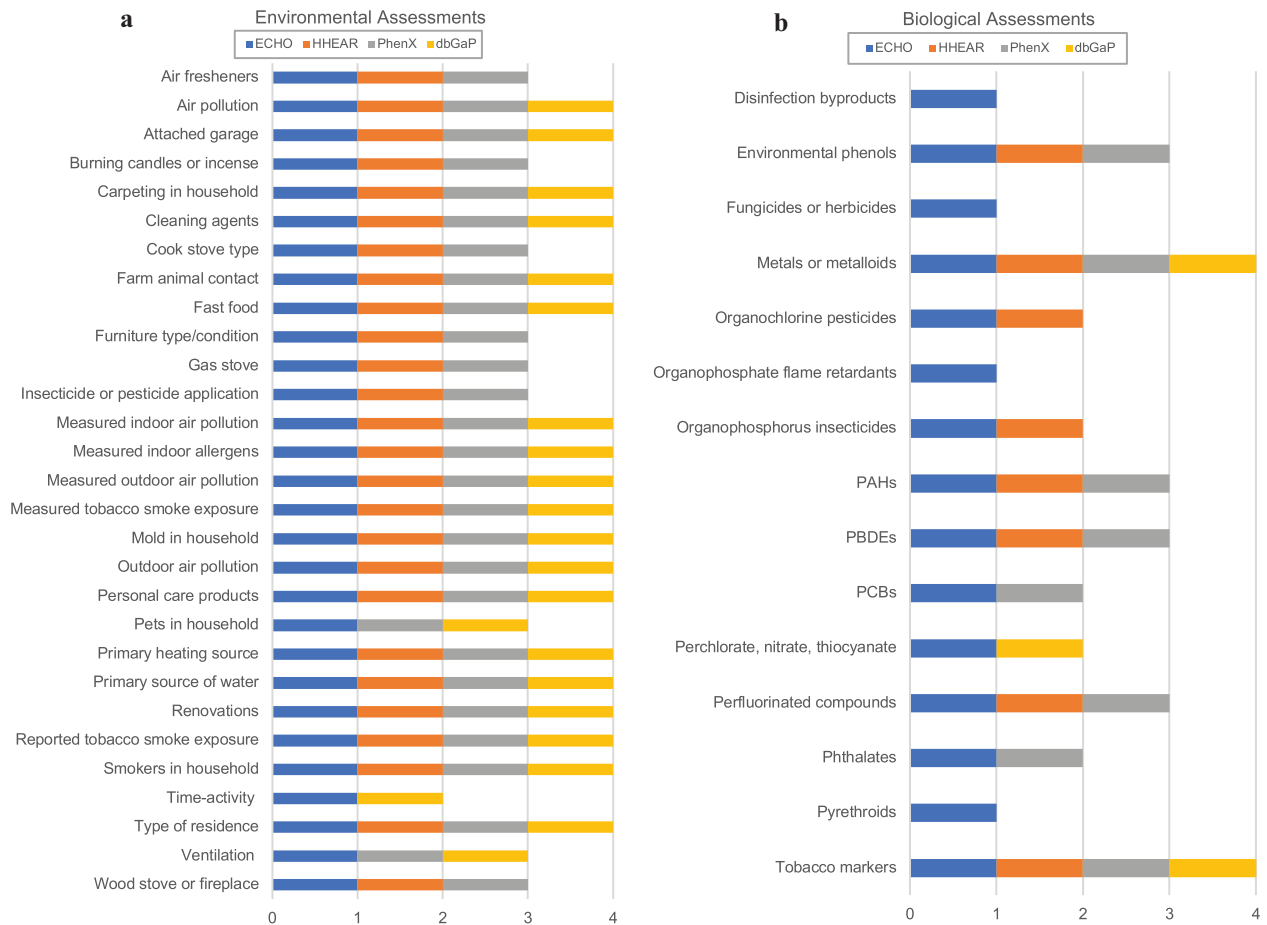


Fig. 1 Presence of environmental exposures in four community resources. **a** Environmental assessments and **b** biological assessments.

Consortia [17]. Cross-consortia investigator groups focused on commonalities across environmental exposures (fish), epidemiologic data, and resilience measures, the latter resulting in a resilience framework [18]. Overall, a dominance of study-specific measures limited the use of common data elements across the consortia studies [2]. Although study-specific measures are necessary and valuable to address local community concerns and specific research questions, standard measures facilitate the ability to combine data from conceptually related studies. We then continued the assessment with another study, the CCREOH Cohort Study, funded by the NIH Fogarty International Center [15]. CCREOH was selected as a good use case as a cohort that could benefit from collaboration with investigators who used similar data collection protocols. The study is a longitudinal follow-up of pregnant women ($N=1143$) and their children from birth to 48 months ($N=992$), providing several timepoints to assess birth and neurodevelopmental outcomes linked to environmental health exposures [15], including biospecimen samples for chemical exposure analysis and non-chemical stressors assessment.

CCREOH and ECHO share health measures. Mapping between CCREOH health assessments and ECHO-wide data collection protocols resulted in five common instruments and ten different instruments between the two studies, with some overlapping variables (Table 1). CCREOH assessments included assessments of depression and perceived stress administered prenatally; neurodevelopmental assessments, including the Bayley Scales of Infant Development for assessment of infant cognitive and motor development at 12–27 months; assessments of cognitive and social-emotional development at 36 months; and planned

assessments of executive function at 48 months [15]. This partial overlap of identical and alternative instruments used for the same data element concept presents opportunities for meta-analysis for some data elements, and challenges for harmonization for other data elements, with the ECHO study data across child life stages [11].

CCREOH and ECHO share biological assessments. ECHO collected essential biospecimens in mothers and children across a broad range of life stages, from prenatal to adolescence [12], whereas biological samples in CCREOH were collected in mothers during the first/second and third trimester of pregnancy; at birth; and in children at 12 and 36 months of age, with planned collection at 48 months [15]. ECHO survey modules measured biomarkers of 15 broad chemical classes of interest [10]. CCREOH and ECHO shared biomarkers indicative of specific biological assessments, as shown in Table 2. This data set, with additional planned chemical assays, aligns well with the 4 to 34 ECHO cohorts with neurodevelopmental assessments and assays for the four chemical groups [16]. Telomere length was also an assessment conducted by both ECHO and by CCREOH that was collected via maternal and child buccal swab [15].

CCREOH study data could be enriched with ECHO cohort data. ECHO includes 69 cohorts consisting of existing data collected prior to the implementation of ECHO and new data to be collected using the ECHO-wide Cohort Data Collection protocol [19]. The existing data collection has a large sample size containing (1) a range of 4 to 34 cohorts that collected 21 neurodevelopmental assessments and assays for 12 chemical groups [16], (2) a range of

Table 1. CCREOH and ECHO share standard questionnaire measures with some collected using the same instruments and others with different instruments.

CCREOH instruments	ECHO instruments
Modified Checklist for Autism in Toddlers	Same
Child Behavior Checklist	Same
Bayley Scales of Infant and Toddler Development	Same
Edinburgh Depression Scale	Same
Cohen Perceived Stress Scale	Same
Alcohol, Smoking and Substance Involvement Screening Test V3.0	Maternal Medical Record Abstraction, Youth Risk Behavior—Substance Use
Brief Trauma Interview	Childhood Trauma Questionnaire
Child Development Review	Ages and Stages Questionnaire
Exposure History Form	Household Chemical Exposure; Household Exposure to Secondhand Smoke
Dietary Questionnaire	NCI Diet History Questionnaire Third Edition (DHQ III); Dietary Screener Questionnaire—Self Report; Block Questionnaire; Dietary Screener Questionnaire
General Health and Demographic Questionnaire	Various Health and Demographic Questionnaires
Generation R Questionnaire (for 36 months)	Various Health Questionnaires
Prenatal Life Events Scale	Crisis in the Family Systems—Revised (CRISYS-R); Stressor Checklist
SF 36 Health Survey	Perceived Stress Scale
Social Support List	PROMIS v2.0 - Emotional Support; Informational Support; Instrumental Support

1 to 69 cohorts that collected the 47 measures of family environment data [11], and (3) 25,526 subjects with descriptive characteristics of mothers of singleton live births by gestational age at birth category [20]. The ECHO Program provides opportunities for potentially combining and comparing data analyses across up to 69 cohorts collecting data using one study protocol and published using the existing and new data [12, 19].

Furthermore, in its strategic plan for 2020–2024, ECHO aims to “make its data and biospecimens accessible to a wide scientific community, in forms that are (1) compatible with other data sets and (2) suitable for use in multiple analyses” [21]. ECHO data sets will become an even more valuable resource for the environmental health research community, and utilization of common data elements will enable further data integration, expanding the potential for meta-analysis and extending the impact from the original individual studies to elucidate subtle and complex interactions between environmental exposures and patient outcomes.

CCREOH provides a rich data set of outcome measures and biospecimens from mother/child dyads. CCREOH measures have a substantive overlap with the environmental exposure measurements and health outcomes in ECHO; therefore, we assessed the standard measures to identify opportunities for meta-analysis to increase statistical power or to validate findings at different geographic locations. Given the degree of overlap of chemical measures, CCREOH has the potential to serve as a highly feasible pilot for ECHO to develop with a process for integrating data from external studies as part of the ECHO collaboration process open to the environmental health research community.

CONCLUSIONS

This assessment and analysis of the prevalence of and commonalities among environmental exposure measures demonstrates the need for standard measures for combining data from conceptually related resources. As indicated by the CCREOH study, the use of standard measures among studies increases opportunities for meta-analysis and may be used to validate findings across geographic locations. With standard measures, NIH resources may expand the potential of individual studies by enabling cross-study analysis, contributing to the body of knowledge on environmental health outcomes. Individual environmental health

Table 2. CCREOH and ECHO share biomarkers indicative of specific biological assessments.

ECHO chemical assays	CCREOH analyzed (+in progress)
Metals/metalloids (Pb, Hg, Cd, Mn)	400 (+600 women)
Herbicides	200 women
Organophosphorus insecticides	200 women
Pyrethroids	200 women
Telomere length	786 Children (+1033 women, 60 children)
Metabolomics	(336 women)

studies can use NIH resources to identify study data sets for opportunities for comparative analysis while preserving some flexibility for community- and culturally tailored inquiry and maintaining scientific rigor.

DISCLAIMER

The views expressed are those of the authors and do not necessarily represent the official position of NIEHS.

REFERENCES

1. Van P, Willem G, Paul P, Emerson C, Grefenstette J, Wilder R, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health*. 2014;14:1144. <https://doi.org/10.1186/1471-2458-14-1144>.
2. Pan H, Edwards SW, Ives C, Covert H, Harville EW, Lichtveld MY, et al. An assessment of environmental health measures in the deepwater horizon research consortia. *Curr Opin Toxicol*. 2019;16:75–82. <https://doi.org/10.1016/j.cotox.2019.07.003>.
3. Mattingly CJ, McKone TE, Callahan MLA, Blake JA, Hubal EAC. Providing the missing link: the exposure science ontology ExO. *Environ Sci Technol*. 2012;46:3046–53. <https://doi.org/10.1021/es2033857>.
4. Sillé FCM, Karakitsios S, Kleensang A, Koehler K, Maertens A, Miller GW, et al. The exposome—a new approach for risk assessment. *Altex*. 2020;37:3–23. <https://doi.org/10.14573/altex.2001051>.
5. Vermeulen R, Schymanski EL, Barabási AL, Miller GW. The exposome and health: where chemistry meets biology. *Science*. 2020;367:392–6. <https://doi.org/10.1126/science.aay3164>.

6. Juarez PD, Matthews-Juarez P, Hood DB, Im W, Levine RS, Kilbourne BJ, et al. The public health exposome: a population-based, exposure science approach to health disparities research. *Int J Environ Res Public Health*. 2014;11:12866–95. <https://doi.org/10.3390/ijerph111212866>.
7. Mattingly CJ, Boyles R, Lawler CP, Haugen AC, Deary A, Haendel M. Laying a community-based foundation for data-driven semantic standards in environmental health sciences. *Environ Health Perspect*. 2016;124:1136–40. <https://doi.org/10.1289/ehp.1510438>.
8. Meeuws S, Yue JK, Huijben JA, Nair N, Lingsma HF, Bell MJ, et al. Common data elements: critical assessment of harmonization between current multi-center traumatic brain injury studies. *J Neurotrauma*. 2020;37:1283–90. <https://doi.org/10.1089/neu.2019.6867>.
9. Balshaw DM, Collman GW, Gray KA, Thompson CL. The children's health exposure analysis resource: enabling research into the environmental influences on children's health outcomes. *Curr Opin Pediatr*. 2017;29:385–9. https://journals.lww.com/co-pediatrics/Fulltext/2017/06000/The_Children_s_Health_Exposure_Analysis_Resource__22.aspx.
10. Buckley JP, Barrett ES, Beamer PI, Bennett DH, Bloom MS, Fennell TR, et al. Opportunities for evaluating chemical exposures and child health in the United States: the Environmental influences on Child Health Outcomes (ECHO) Program. *J Exposure Sci Environ Epidemiol*. 2020;30:397–419. <https://doi.org/10.1038/s41370-020-0211-9>.
11. Bush NR, Wakschlag LS, LeWinn KZ, Hertz-Picciotto I, Nozadi SS, Pieper S, et al. Family Environment, Neurodevelopmental Risk, and the Environmental Influences on Child Health Outcomes (ECHO) initiative: looking back and moving forward. *Front Psychiatry*. 2020;11. <https://www.frontiersin.org/article/10.3389/fpsy.2020.00547>.
12. Gillman MW, Oken E, Cella D, Gershon R, Parker CB, Jacobson LP, et al. Environmental influences on Child Health Outcomes (ECHO)-wide Cohort Data Collection Protocol. 1.2 ed. 2018.
13. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, et al. The PhenX Toolkit: get the most from your measures. *Am J Epidemiol*. 2011;174:253–60. <https://doi.org/10.1093/aje/kwr193>.
14. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 2007;39:1181–6. <https://doi.org/10.1038/ng1007-1181>.
15. Zijlman W, Wickliffe J, Hindori-Mohangoo A, MacDonald-Ottevanger S, Ouboter P, Landburg G, et al. Caribbean Consortium for Research in Environmental and Occupational Health (CCREOH) Cohort Study: influences of complex environmental exposures on maternal and child health in Suriname. *BMJ Open*. 2020;10:e034702. <https://bmjopen.bmj.com/content/bmjopen/10/9/e034702.full.pdf>.
16. Schantz SL, Eskenazi B, Buckley JP, Braun JM, Sprowles JN, Bennett DH, et al. A framework for assessing the impact of chemical exposures on neurodevelopment in ECHO: Opportunities and challenges. *Environ Res*. 2020;188:109709. <https://www.sciencedirect.com/science/article/pii/S0013935120306022>.
17. Lichtveld M, Sherchan S, Gam KB, Kwok RK, Mundorf C, Shankar A, et al. The deepwater horizon oil spill through the lens of human health and the ecosystem. *Curr Environ Health Rep*. 2016;3:370–8. <https://doi.org/10.1007/s40572-016-0119-7>.
18. Abramson DM, Grattan LM, Mayer B, Colten CE, Arosemena FA, Bedimo-Rung A, et al. The resilience activation framework: a conceptual model of how access to social resources promotes adaptation and rapid recovery in post-disaster settings. *J Behav Health Serv Res*. 2015;42:42–57. <https://doi.org/10.1007/s11414-014-9410-2>.
19. Jacobson LP, Lau B, Catellier D, Parker CB. An environmental influences on child health outcomes viewpoint of data analysis centers for collaborative study designs. *Curr Opin Pediatr*. 2018;30:269–75. <https://doi.org/10.1097/mop.0000000000000602>.
20. Dunlop AL, Essalmi AG, Alvalos L, Breton C, Camargo CA, Cowell WJ, et al. Racial and geographic variation in effects of maternal education and neighborhood-level measures of socioeconomic status on gestational age at birth: findings from the ECHO cohorts. *PLoS ONE*. 2021;16:e0245064. <https://doi.org/10.1371/journal.pone.0245064>.
21. Environmental Influences on Child Health Outcomes (ECHO) Program. ECHO Program Office Strategic Plan 2020–2024 [Internet]. Bethesda: National Institutes of Health; 2020 [cited 2021 Nov 29]. 23 p. Available from: <https://www.nih.gov/sites/default/files/research-training/initiatives/echo/ECHO-Program-Office-Strategic-Plan-2020.pdf>.

ACKNOWLEDGEMENTS

The authors would like to thank Ashna Hindori-Mohangoo, Shellice Sairras, and Anisma Gokoel for their contribution to data collection in the Caribbean Consortium for Research in Environmental and Occupational Health (CCREOH) project.

AUTHOR CONTRIBUTIONS

MN, HP, and CLI compiled data from NIH resources and conducted mapping analyses. HC and HP conducted mapping between CCREOH and ECHO exposure measures. HC, SWE, CMH, EWH, MYL, HP, JKW, WZ, and CLI drafted and revised the manuscript. All authors reviewed and approved the manuscript.

FUNDING

This project is funded by National Institute of Environmental Health Sciences (NIEHS) award 1R24ES028479-01 and Fogarty International Center awards U01TW010087 and U2RTW010104 of the National Institutes of Health (NIH).

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41370-022-00428-7>.

Correspondence and requests for materials should be addressed to Cataia Ives.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022