American Journal of Epidemiology

Vol. 174, No. 3
DOI: 10.1093/aje/kwr191
Advance Access publication:
July 11, 2011

Response to Invited Commentary

Hamilton et al. Respond to ''Consolidating Data Harmonization''

Carol M. Hamilton*, Lisa C. Strader, Joseph G. Pratt, Deborah Maiese, Tabitha Hendershot, Richard K. Kwok, Jane A. Hammond, Wayne Huggins, Dean Jackman, Huaqin Pan, Destiney S. Nettles, Terri H. Beaty, Lindsay A. Farrer, Peter Kraft, Mary L. Marazita, Jose M. Ordovas, Carlos N. Pato, Margaret R. Spitz, Diane Wagener, Michelle Williams, Heather A. Junkins, William R. Harlan, Erin M. Ramos, and Jonathan Haines

* Correspondence to Dr. Carol M. Hamilton, RTI International, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle Park, NC 27709 (e-mail: chamilton@rti.org).

Initially submitted February 15, 2011; accepted for publication May 17, 2011.

Abbreviations: DataSHaPER, Data Schema and Harmonization Platform for Epidemiological Research; dbGaP, database of Genotypes and Phenotypes; GuLF STUDY, Gulf Long-term Follow-up Study; LOINC, Logical Observation Identifiers Names and Codes; NEIGHBOR, National Eye Institute Glaucoma Human Genetics Collaboration; PhenX, consensus measures for Phenotypes and eXposures.

We appreciate the comparison of the consensus measures for Phenotypes and eXposures (PhenX) Toolkit (1) with the Data Schema and Harmonization Platform for Epidemiological Research (DataSHaPER) provided by Fortier et al. (2). The PhenX Toolkit is most relevant for prospective data collection efforts, such as designing new studies or expanding existing studies. For example, the Gulf Long-term Follow-up Study (GuLF STUDY) is a large prospective cohort study investigating the health of Deepwater Horizon oil-spill cleanup workers. These investigators recognized the value of the PhenX Toolkit for incorporating standard measures, and they selected critical PhenX measures when designing the GuLF STUDY (http://www.niehs.nih.gov/about/od/programs/gulfspill/gulfstudy/index.cfm). These investigators also selected PhenX measures for covariates, such as education, marital status, alcohol and tobacco use, and reproductive history. We agree with Fortier et al. (2) that the PhenX Toolkit is ''stringent'' and provides detailed protocols to ensure that data collected by different investigators are comparable. Including common PhenX measures in multiple studies makes it possible to directly compare data across studies. This should improve the sensitivity and specificity of the study measures during cross-study analysis. In the GuLF STUDY, investigators selected PhenX measures and incorporated Toolkit protocols verbatim.

However, the PhenX Toolkit is not restrictive, because investigators can select any PhenX measures in their primary research area or beyond. In the GuLF STUDY, investigators selected the measures that best suited their needs, including some outside of the main area of focus. Because PhenX measures were selected by domain experts and vetted by the scientific community, Toolkit users can be confident that they are incorporating high-priority, high-quality measures. Investigators will find that the Toolkit addresses the most important conditions affecting public health and includes a variety of exposures (e.g., environmental, psychosocial, and social). Many different diseases have common risk factors, and relying on PhenX measures may help researchers combine studies of divergent research focus, such as cardiovascular and gastrointestinal studies. The National Eye Institute Glaucoma Human Genetics Collaboration (NEIGHBOR) is a consortium of clinicians and geneticists using harmonized clinical definitions for glaucoma phenotypes and identical genotyping platforms to facilitate their data analysis (http://www.nei.nih.gov/news/statements/glaucoma_initiatives.asp). The NEIGHBOR consortium needed to collect data describing alcohol and tobacco use extracted from medical records. This consortium of investigators decided to use PhenX measures to collect the data (Dr. Louis R. Pasquale, Massachusetts Eye and Ear Infirmary, personal communication, 2011). In this case, PhenX measures helped investigators assemble data into a standard format. Although we would expect PhenX measures to be collected using standard PhenX protocols, the use of PhenX measures to ''collect'' data retrospectively may be an option and could be considered data harmonization.

The PhenX Toolkit may also help users identify common measures in existing data. PhenX measures are linked to standards and resources, thus making it easier for investigators to identify comparable data. Currently, all PhenX measures have associated Cancer Biomedical Informatics Grid common data elements. Logical Observation Identifiers Names and Codes (LOINC) codes are being developed for measures, protocols, and variables; and some codes have been released in trial status. To date, PhenX has mapped to 12 studies included in the database of Genotypes and Phenotypes (dbGaP), identifying ''comparable'' and ''related'' variables. One of the key goals of the PhenX–dbGaP collaboration is to highlight PhenX measures and variables in dbGaP, so investigators can identify common measures across studies (Dr. Michael Feolo, National Center for Biotechnology Information, personal communication, 2011).

We applaud the efforts of the DataSHaPER investigators to provide a resource harmonizing data retrospectively, which should help collaborators share existing data. DataSHaPER currently provides links to the PhenX Toolkit, and the PhenX investigators plan to develop a cross-reference to other standards and resources, including DataSHaPER. Both PhenX and DataSHaPER serve the research community by addressing different aspects of the same problem: the need to combine studies. Both resources support collaborative biomedical and translational research, which may lead to more effective interventions and treatments for common complex diseases.

ACKNOWLEDGMENTS

Author affiliations: RTI International, Research Triangle Park, North Carolina (Carol M. Hamilton, Lisa C. Strader, Joseph G. Pratt, Deborah Maiese, Tabitha Hendershot, Jane A. Hammond, Wayne Huggins, Dean Jackman, Huaqin Pan, Destiney S. Nettles); National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland (Heather A. Junkins, Erin M. Ramos); Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland (Terri H. Beaty); Departments of Medicine, Neurology, Ophthalmology, and Genetics and Genomics, School of Medicine, Boston University, Boston, Massachusetts (Lindsay A. Farrer); Departments of Epidemiology and Biostatistics, School of Public Health, Boston University, Boston, Massachusetts (Lindsay A. Farrer); Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts (Peter Kraft); Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania (Mary L. Marazita); Human Nutrition Research Center on Aging, Tufts University, Boston, Massachusetts (Jose M. Ordovas); Zilkha Neurogenetic Institute, University of Southern California, Altadena, California (Carlos N. Pato); M. D. Anderson Cancer Center, University of Texas, Houston, Texas (Margaret R. Spitz); RTI International, San Diego, California (Diane Wagener); Department of Epidemiology, School of Public Health and Community Medicine, University of Washington, Seattle, Washington (Michelle Williams); Office of the Director, National Institutes of Health, Chevy Chase, Maryland (William R. Harlan (retired)); Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee (Jonathan Haines); and Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina (Richard K. Kwok, Destiney S. Nettles).

Conflict of interest: none declared.

REFERENCES

1. Hamilton CM, Strader LC, Pratt JG, et al. The PhenX Toolkit: get the most from your measures. *Am J Epidemiol*. 2011; 174(3):253–260.
2. Fortier I, Doiron D, Burton P, et al. Invited commentary: consolidating data harmonization—how to obtain quality and applicability? *Am J Epidemiol*. 2011;174(3):261–264.