



LUND UNIVERSITY

Transnational research partnerships: leveraging big data to enhance US health.

Crump, Casey; Sundquist, Kristina; Winkleby, Marilyn

Published in:
Journal of Epidemiology and Community Health

DOI:
[10.1136/jech-2015-205451](https://doi.org/10.1136/jech-2015-205451)

2015

[Link to publication](#)

Citation for published version (APA):
Crump, C., Sundquist, K., & Winkleby, M. (2015). Transnational research partnerships: leveraging big data to enhance US health. *Journal of Epidemiology and Community Health*, 69(11), 1029-1030.
<https://doi.org/10.1136/jech-2015-205451>

Total number of authors:
3

General rights

Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

**TRANS-NATIONAL RESEARCH PARTNERSHIPS:
LEVERAGING BIG DATA TO ENHANCE U.S. HEALTH**

Casey Crump, M.D., Ph.D.,¹ Kristina Sundquist, M.D., Ph.D.,^{2,3} Marilyn A. Winkleby, Ph.D.³

¹Department of Medicine, Stanford University, Stanford, California, USA

²Center for Primary Health Care Research, Lund University, Malmö, Sweden

³Stanford Prevention Research Center, Stanford University, Stanford, California, USA

Corresponding author: Casey Crump, M.D., Ph.D., Stanford University, Department of Medicine, 211 Quarry Road, Suite 405, MC 5985, Palo Alto, California USA 94304-1426. Tel: (650) 498-9000. Fax: (650) 498-7750. Email: kccrump@stanford.edu

Key words: cohort studies, biomedical research, global health

Word count: 1,000

In the current era of big data and small research budgets, new strategies are needed for more cost-effective leveraging of big data to enhance our nation's health. One strategy is to promote trans-national partnerships to tap into the rich, extensive databases available in other countries, particularly in Europe. The National Institutes of Health (NIH) has increasingly recognized that new collaborations that bring together multiple data sources will play a critical role in advancing our knowledge of disease causation, improving patient care, and promoting healthier communities. However, given cuts in research funding and fierce competition for U.S. grants, some question whether U.S. dollars should be diverted to fund "foreign" studies. In this commentary, we argue that trans-national research partnerships offer significant advantages for enhancing the health of both the U.S. population as well as the broader global community.

In the U.S., the collection of population-wide health data has been hampered by the inherent difficulties in linking patients across many different health care delivery systems. As a result, the availability of big data for health research has been limited mainly to a few large organizations such as Kaiser Permanente, Group Health, the Mayo Clinic, and VA hospitals. The data collected by such organizations are rich resources but have significant limitations. They include only a selected patient population, which is often poorly representative of the broader population in terms of socioeconomic, ethnic, or health factors, thus limiting generalizability. Their patient populations also fluctuate over time due to changes in insurance plan enrollments, making long-term outcomes more difficult to track. In addition, their patient care data are often not linkable with broader information such as census, neighborhood, and multi-generation data that would allow examination of more complex pathways affecting health. In contrast, these limitations do not exist in certain countries that have universal health care with electronic medical records linked to national health registries. In the Nordic countries (Denmark,

Finland, Iceland, Norway, and Sweden), for example, the entire national population is essentially a cohort. By embedding data collection within the national health care infrastructure, extensive clinical and epidemiologic data are prospectively obtained that are unparalleled in completeness, quality, and size by those available in the U.S., enabling robust testing of hypotheses that are also relevant to the U.S. population.

Trans-National Research Partnerships

NIH-funded partnerships with countries that already have comprehensive population-based datasets are cost-efficient investments, allowing U.S. investigators and their partners to answer important research questions that are not possible using U.S. data. Leveraging of registry data from the Nordic countries (with a combined population of ~25 million), for example, provides numerous advantages compared with U.S. sources alone. They include nearly 100% complete, high-quality nationwide data on inpatient and outpatient clinical diagnoses, prescription records, birth and death records, sociodemographic characteristics, and (in some countries) highly detailed neighborhood-level social and physical environment characteristics, and a national biobank for genetic studies. All data sources are mutually linkable using a confidential, anonymous version of a unique personal identification number assigned to each person at birth or immigration. Many of these data sources have already been prospectively collected for decades, enabling large-scale studies of long-term temporal trends, life course analyses, and extensive family-based designs to disentangle genetic and environmental influences on disease. Universal health coverage also facilitates more complete and equitable ascertainment of health conditions across different social groups, allowing more rigorous studies in high-risk subpopulations. Large national cohort studies based on these data can provide more

robust findings that avoid selection and ascertainment biases commonly affecting other observational study designs. Generalizability of biologic findings to the U.S. and other Western countries is high because of similar underlying mechanisms across these populations, and is enhanced for sociodemographic findings by high immigration rates that have increased social and ethnic diversity over the past few decades (e.g., ~26% of the Swedish population are 1st- or 2nd-generation immigrants). Because these rich data sources have already been constructed, large-scale studies can be conducted at remarkably low direct costs (typically well below one U.S. dollar per subject), as well as much lower indirect costs than most U.S.-based studies.

International Data to Enhance U.S. Health

NIH-funded studies based on these data are making vital contributions to health research and policy in the U.S.—including new knowledge about disease mechanisms, clinical translation, and prevention—that would otherwise be logistically and financially infeasible. For example, Nordic prescription databases have enabled numerous landmark studies of the health effects, safety, and cost-effectiveness of medications that are commonly used in the U.S.[1] Such studies make vital contributions beyond those of clinical trials because of their larger sample sizes, more diverse patients, and longer exposures, enabling robust assessments of how medications work in the real world. In other studies, the unique ability to follow a large national cohort from birth into adulthood has enabled novel investigations of early life influences on chronic disease and mortality in later life. This led to the first-ever identification of increased mortality risks in adulthood associated with preterm birth,[2] as well as many other long-term chronic disease sequelae. These discoveries have advanced our knowledge of early life origins of chronic disease, and will help inform long-term clinical care for the growing number of adult

survivors of preterm birth, which currently affects nearly 12% of U.S. births and costs more than \$26 billion annually in U.S. health care expenditures and lost productivity.[3] Similarly, the first study of long-term mortality associated with early term birth (37-38 weeks of gestation) has supported a re-definition of full term birth,[4] and can potentially influence the timing of deliveries to enhance maternal and infant health outcomes for the nearly 30% of U.S. births that occur at early term. These and many other seminal studies of health issues with large population impacts are expanding the knowledge base needed to develop better-targeted preventive interventions and health policy in the U.S.

Federal Mandates Calling for Greater Collaboration

The NIH has called for increasing collaboration across different agencies to expedite the translation of research findings into knowledge that improves human health. Several recommendations can be made for developing more cost-effective partnerships to facilitate the use of big data for translational research. First, trans-national partnerships to leverage big data that are unavailable in the U.S. are vital for enabling larger population-based studies that are highly robust, generalizable, and cost-efficient. Additional projects that incorporate such data are needed to explore new hypotheses that will benefit important health problems in the U.S., including better prevention and treatment of chronic diseases, mental disorders, and maternal and child health issues. These collaborative projects are a “win-win” for U.S. investigators and their partners, and most importantly for the health of their respective populations. Second, using a similar strategy in U.S. settings, further integration of clinical trials and observational studies into existing health care delivery systems is needed to generate larger study samples at lower costs per subject, by creatively re-using existing data and infrastructure. Third, continued

investments in translational research are needed to ensure that the scientific discoveries from these efforts are translated into new interventions that reach the target U.S. population groups.[5] Such efforts in both the U.S. and abroad can make major contributions toward enhancing our nation's health through cost-effective use of more comprehensive health data. Future NIH partnerships that incorporate these strategies are worthy investments and need broader replication in other health care settings and populations.

ACKNOWLEDGMENTS

Author contributions: All authors (CC, KS, and MAW) made substantial contributions to the drafting or revising of the manuscript, and approved the final version.

Funding: None.

Conflicts of interest: None.

REFERENCES

1. Furu K, Wettermark B, Andersen M, et al. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic & clinical pharmacology & toxicology* 2010;**106**(2):86-94.
2. Crump C, Sundquist K, Sundquist J, et al. Gestational age at birth and mortality in young adulthood. *JAMA* 2011;**306**(11):1233-40.
3. Behrman RE, Butler AS. *Preterm Birth: Causes, Consequences, and Prevention*. Washington, DC: National Academies Press, 2007.
4. Crump C, Sundquist K, Winkleby MA, et al. Early-term birth (37-38 weeks) and mortality in young adulthood. *Epidemiology* 2013;**24**(2):270-6.

5. Woolf SH. The meaning of translational research and why it matters. JAMA 2008;**299**(2):211-3.