

The Use of Archival Data to Inform Health Policy

This article provides an introductory overview of the origins, benefits, and applications of archival data for informing health policy.

In contrast to investigation that relies on data collected for a specific research project, scientific inquiry using archival data relies upon data that have already been collected. Researchers may seek the archival data that best fit their research question, adapt their research question to the data they have on hand, or choose to combine both methods.¹ However, aligning the relevant research question and the available data often poses a challenge.

Archival data collected as part of *longitudinal research studies* has been used to inform health policy. For instance, decades of information collected from participants in the classic Framingham Heart Study (1948) provided a wealth of information that has been made available to researchers through archived sources.² More recently, Christakis and Fowler analyzed longitudinal data collected as part of the Framingham Study to describe the increase in rates of obesity among large social networks.³ Following in part from the contributions of Christakis and Fowler, there are empirically supported recommendations for developing effective programs aimed at using social networks to reduce the obesity epidemic.⁴

Analyses of archival data from *cross-sectional studies* have also been used to inform health policy. Cross-sectional data are derived from the study of groups of individuals at the same point in time. Ellaway, Macintyre, and Bonnefoy conducted secondary analyses of the Large Analysis and Review of European Housing and

Health Status (LARES) Survey, an international cross-sectional survey administered in Europe.⁵ In part, their results provided empirical support that environmental factors, such as green space, should be included in health policies aimed at promoting physical activity to decrease rates of obesity.

Efforts to store and make available archival data for research also extend to *administrative data*, i.e., pieces of information collected as part of the routine operations of a business or agency. In 1998, The University of Pennsylvania established the *Neighborhood Information System (NIS)*, which makes available in one central repository a vast amount of information routinely gathered for operational or business purposes by public or private agencies. Hillier and colleagues have analyzed these data from multiple administrative sources to inform public policy initiatives regarding housing abandonment in large urban cities such as Philadelphia, PA.⁶

Administrative medical claims data are often used to compare outcomes of interventions. For example, they may be analyzed to determine cost effectiveness, differences in the intensity of care among different patient populations, or both.

Recently, the Einstein Center for Urban Health Policy and Research analyzed inpatient data from persons with diabetes mellitus to better understand the percent of admissions associated with diabetes, and the most common complications in this patient population. This information may be useful in planning the allocation of hospital services and in directing programs aimed at reducing medical complications. Importantly, this information is

readily available for analysis and offers comparable or better rates of follow up without the resource demands of primary data collection.^{7,8} One caveat – the reliability of these data appears to decrease for elements that less directly relate to the primary purpose of the data collection. For instance, the reliability of secondary diagnoses may be lower than for primary diagnoses.⁷

Despite the challenges described above, analysis of archival data from administrative sources has substantial benefits. Typically gathered through processes that present no more than minimal risk to participants, these data may provide information about vulnerable populations such as prisoners, the elderly, and children – information that can be particularly difficult to obtain elsewhere.

Dialog among researchers and policymakers in conjunction with advancements in technology, such as the Internet and personal computers, during the last several decades have coincided with increased accessibility and variety in archival data. Archival data analyses offer significant ethical and feasibility benefits. Results from these studies are directly applicable to illuminating the most prominent health policy issues; developing effective public health interventions; and developing administrative policies related specific health issues.

Shana Stites, MA, MS

*Research Analyst
Einstein Center for Urban Health Policy and
Research*

REFERENCES

1. Elder G, Pavalko E, Clipp E. *Working with Archival data: Studying Lives*. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-088. Newbury Park, CA: SAGE Publications, Inc; 1993.
2. Framingham Heart Study. History of the Framingham heart study. <http://www.framinghamheartstudy.org/>. Updated January 11, 2011. Accessed February 28, 2011.
3. Christakis N, Fowler J. The spread of obesity in a large social network over 32 years. *N Engl J Med*. 2007;357:370-379.
4. Koehly L, Loscalzo A. Adolescent obesity and social networks. *Prev Chronic Dis*. 2009;6(3):1-8.
5. Ellaway A, Macintyre S, Bonnefoy X, Graffiti, greenery and obesity in adults: Secondary analysis of European cross sectional survey. *BMJ*. 2005;331(7517):611-612.
6. Hillier A, Culhane D, Smith T, Tomlin CD. Predicting housing abandonment with the Philadelphia Neighborhood Information System. *J Urban Affairs*. 2003;25(1):91-105.
7. Roos L, Nicol J, Cageorge S. Using administrative data for longitudinal research: Comparisons with primary data collection. *J Chronic Dis*. 1987;40(1):41-49.
8. Roos L, Mustard C, Nicol J, Comm B, McLerran D, Malenka D, et al. Registries and Administrative Data: Organization and Accuracy. *Medical Care*. 1993;31(3):201-212.