

An Open Letter on Expanding Access to Administrative Data for Research in the United States

The aim of this letter is to express our support for the development and expansion of direct and secure access to administrative data at government agencies in the United States for scientific research.

Administrative data such as tax records and the earnings and benefits records in the Social Security and Medicare systems are an extraordinarily important source of information for research on economic policies. While researchers have long used traditional survey data such as that provided by the Current Population Survey or the Panel Study of Income Dynamics to study a range of questions, these data sources have substantial limitations relative to administrative data files. Administrative records offer much larger sample sizes for analysis, as full population files are available; include a longitudinal structure that suffers less from attrition problems; and provide higher quality information with a correspondingly reduced rate of measurement error.

The key challenges in permitting researchers outside the government agencies that maintain administrative records to access these data are confidentiality and security concerns. Micro-level administrative data cannot be publicly disclosed. There are three approaches to addressing confidentiality concerns while using the data for research. These approaches are differentially successful in delivering the best conditions for high quality research and, correspondingly, offer different degrees of benefits to society.

(1) The most effective way to support research with administrative data is to provide secure, direct access to micro-data records stripped of individual identifiers, for example through the statistical office of the administrative agency itself. Under a direct access arrangement, researchers can analyze administrative data records directly in the secure environment, but only aggregate statistical results can be taken out and publicly disclosed. Experience in a number of European nations, and limited experience with pilot efforts in various government agencies in the United States, suggests that such secure access is workable. This approach is most fruitful when three conditions are met: (a) fair and open competition for access based on scientific merit, (b) sufficient bandwidth to accommodate many outstanding projects simultaneously, (c) inclusion of younger scholars and graduate students, who have the most time to devote to data work, in the research teams that are given access to the data.

(2) An alternative approach is to disclose synthetic data, i.e., micro-data simulated to match a limited set of statistics from the actual data. This approach is much less attractive than providing direct access to the full administrative data set because in practice it is virtually impossible for the researchers to fully specify the contents of the ideal synthetic dataset in advance. Moreover, synthetic data make it difficult and in some cases impossible to study sub-sets of the population, unless a distinct synthetic data set is specifically created for this purpose. In particular, such data are not well suited to the analysis of income distribution and related subjects that require the full joint distribution of the variables of interest, and not just a limited set of moments. This approach consequently is a very poor substitute for authorized secured access to actual administrative record micro-data files.

(3) Still a third option is to create a researcher-to-agency interface that allows researchers to send computer programs to agency employees, who will run those programs on the administrative data records. This option, while providing some data access, is also substantially inferior to direct data access because it does not allow for the iterative, interactive data work that is critical for most empirical projects. Top flight empirical researchers learn from examining and analyzing data records directly, and from simple summary statistics for subsets of data records. Without such direct access, it is much more difficult for this element of the empirical process to move forward.

Some government agencies in the United States have already made valuable efforts to provide direct secured access to administrative data for research. But the United States lags behind other countries in creating opportunities for researcher access to these data. A number of European countries even offer *remote* access to confidential micro-data through secure connections. As a result, frontier micro-economic research is in many cases shifting away from the United States and toward European countries purely because of data access issues.

In order to maintain and spark cutting-edge research on critical policy questions in the U.S. economy, and to provide reliable evidence on key economic parameters that underlie economic policy design, we believe that it is extremely important to expand direct access to administrative micro-data in the U.S. We strongly support efforts at various government agencies to enhance and promote such access.

David Card, Professor of Economics, University of California, Berkeley, and Director of the Labor Studies Program, National Bureau of Economic Research

Raj Chetty, Professor of Economics, Harvard University, and Co-Director of the Public Economics Program, National Bureau of Economic Research

David Cutler, Professor of Economics, Harvard University, and Co-Editor of the Journal of Health Economics

Steven Davis, Professor of Economics, University of Chicago Booth School of Business, and Editor of the American Economic Journal: Macroeconomics

Martin Feldstein, Professor of Economics, Harvard University, and National Bureau of Economic Research President Emeritus

William Gale, Senior Fellow, Brookings Institution

Jonathan Gruber, Professor of Economics, MIT, and Director of the Health Care Program, National Bureau of Economic Research

Michael Greenstone, Professor of Economics, MIT, and Director of the Hamilton Project, Brookings Institution

Caroline Hoxby, Professor of Economics, Stanford University, and Director of the Economics of Education Program, National Bureau of Economic Research

Lawrence Katz, Professor of Economics, Harvard University, and Editor of the Quarterly Journal of Economics

Emmanuel Saez, Professor of Economics, University of California, Berkeley, and Editor of the Journal of Public Economics

July 27, 2010