

## Editorial

---

The theme of this issue is administrative data. Ten papers on this topic, and the related topic of administrative records, are featured. Collectively, these papers offer a glimpse into the status of the use of administrative data and records in the public and private sector. Individually, they represent a variety of case studies that reflect how statisticians and other professionals are exploiting opportunities to explore and evaluate administrative data and administrative records for their research.

Several themes are especially in evidence among the papers:

1. Cost savings and improved operational efficiency are powerful incentives to promote the use of administrative data.
2. Record linkage can yield important advantages to researchers. Linking high-quality data from two different surveys or between a survey and non-survey data source such as commercial data can help create a rich analytical environment.
3. Improving the quality and usability of existing statistics may depend on an ability to find ways to accept the challenges of using administrative data while at the same time working to develop ways to overcome their shortcomings.
4. New and emerging administrative data sources may offer promise for improving surveys in the future. Their exploration is described in a few of the papers, demonstrating that outside-the-box thinking is critical to reaping their benefits.
5. Survey redesign may provide an optimum setting for exploring how administrative data might improve the measurement of demographic and other characteristics if thresholds for data quality and the challenges of integrating these data with other data sources can be met.

The first paper explores how researchers adapted to the decision by Statistics New Zealand (Stats NZ) to adopt an “administrative data first” approach for quarterly financial statistics. Based on this approach, administrative data are the primary source of business information, and further data are collected only when necessary. Craig Liken, Mathew Page, and John

Stewart of Stats NZ describe the transformation process, methods applied, and the continuing commitment required to implement this paradigm shift, both within Stats NZ and externally with customers. The researchers took a cautious approach initially due to concerns about the conceptual fit of the administrative data, quality, and other factors. A small project team established in 2013 developed and refined methods for assessing, transforming, and using administrative data, and the new approach was successfully implemented from the September 2015 quarter onward.

Census-based enumerations of the population and housing of the Netherlands were carried out from 1829 to 2011, when a register-based census was adopted. This development reflected a decline in the willingness of the population to participate in censuses based on privacy and other concerns. Eric Nordholt of Statistics Netherlands describes this transition in a paper that covers the advantages and disadvantages of a register-based census and the growth in its use throughout Europe. The paper includes interesting historical details on the evolution of the Dutch census and explanations of the differences between a population and housing census and a combined or registry-based census.

The *Virginia Plan for Higher Education*, approved in 2014, provides a framework for improving higher education for the state. One of the *Plan's* goals is to increase the rate of postsecondary degree attainment from 51 percent to 70 percent by 2030. Researchers Bianica Pires, Ian Crandell, Madison Arnsbarger, Vicki Lancaster, Sallie Keller, Aaron Schroeder, and Stephanie Shipp of the Biocomplexity Institute of Virginia Tech, Social and Decision Analytics Laboratory, and Wendy Kang and Paula Robinson from the State Council of Higher Education for Virginia used administrative data and other data sets to complete a pilot study of two geographic areas in Virginia to better understand why postsecondary attainment varies across Virginia's communities. Their research demonstrates the perseverance often required to use administrative data, requiring as they put it, “an iterative process between continued data discovery, inventory, and cleaning and transformation processes in addition to feedback from subject-matter experts.”

The largest contributor to the costs of conducting the U.S. 2010 Decennial Census, the Nonresponse Followup (NRFU) operation, has been the subject of several studies showing that modeling based on administrative data can be predictive of NRFU enumeration outcomes in decennial data collection. Authors Melissa Chow, Hubert Janicki, Lawrence Warren, and Moises Yi of the U.S. Census Bureau and Mark Kurtzbach of the Federal Deposit Insurance Corporation compare model predictive power when varying training data sources. They evaluate the extent to which survey data can be used to reduce enumerator workload when combined with available administrative data using 2010 Census and 2014 American Community Survey (ACS) data. They also suggest a broader role for using survey data in NRFU operations of statistical agencies outside of the U.S. if census or administrative data sources have only incomplete coverage of the population.

Joseph Kadane of the Department of Statistics and Data Science at Carnegie Mellon University, and Karl Williams of the Allegheny County, Maryland Office of the Medical Examiner investigate the issue of data quality associated with the system that certifies death in the United States. Under the federal system in the U.S., each state sets its own standards for the qualifications needed to certify deaths. There is substantial underreporting in the case of drug overdoses, which are rising in the U.S. due to the opioid crisis. The current death reporting system follows every single death and determines all drugs that might be associated with that death. The authors argue that this degree of precision is not needed, and that a different way of thinking about the information required should be explored. Challenging existing statistical indicators on the basis of their utility and efficiency is intrinsic to approaches and methods used by others working with administrative data.

National Statistical Offices (NSOs) working to incorporate administrative data into their research agendas are not merely focused on current administrative data sources, but also on new and emerging sources that can produce more robust statistical information. In a study of Australia's road freight industry, Australian Bureau of Statistics researcher Nicholas Husek used data from telematic devices (which record time, coordinates, and speed of a truck, for example) and linked these data with other data sources, such as weather and traffic accident information, to improve the analytical environment for road freight research and better inform infrastructure decisions. This paper was awarded second place in the 2017 Young Statisticians competition.

Record linkage identifies and merges data in two or more sources that refer to the same entity. In a study linking and comparing race and Hispanic origin data of Medicaid participants with their responses in the decennial census, Laticia Fernandez of the U.S. Census Bureau found that missing data in Medicaid records was a much larger issue than was non-matching responses. There was a relatively high overall percent of matching race and Hispanic origin even with differences in how states collected race and Hispanic origin data from Medicaid program participants. Other research showed that minorities are more likely to have non-matching race and Hispanic origin data. The author suggests that there are potentially valuable gains from linking administrative records to Census Bureau data to assess the consistency of demographic data, as well as to supplement information in cases of item non-response.

Quentin Brummet of the National Opinion Research Center at the University of Chicago; Denise Flanagan-Doyle, Joshua Mitchell, and John Voorheis of the U.S. Census Bureau; and Laura Erhard and Brett McBride of the U.S. Bureau of Labor Statistics explore the potential usefulness of linking administrative information on income from the Internal Revenue Service to the Consumer Expenditure Survey (CE), the only federal survey data source that provides a complete profile of the income and spending habits of U.S. households. CE data are used for a number of purposes such as updating the collection of goods and services underlying the Consumer Price Index. As a result of their analysis, the authors gained new insights into potential survey nonresponse bias and measurement error properties of CE income data.

The utility of using commercial property tax data to improve ACS estimates of property tax amounts is explored in a paper by Zachary Seeskin of the National Opinion Research Center at the University of Chicago. Drawing on the results of earlier research and his own analysis, Seeskin identified major challenges with using CoreLogic property tax data. For example, coverage of this commercial data source varies across the country; amounts recorded on property tax records may not reflect the property taxes actually billed. Also, large differences between CoreLogic and ACS property taxes may reflect conceptual differences between what is collected in two data sources for certain counties. While commercial data offers great promise, Seeskin's research shows that challenges emerge when data are collected and maintained by many authorities having different practices.

The goal of a major redesign in 2017 of the Survey of Graduate Students and Postdoctorates in Science and Engineering Survey (GSS) was to improve data quality while reducing reporting burden. Key elements of the redesign included the use of digital file transfers for data reporting, separate collection by degree level of data for graduate students, and the adoption of a more common coding scheme to describe academic disciplines. Researchers Jonathan Gordon, Stephanie Eckman, Peter Einaudi, and Herschel Sanders of RTI International, and Mike Yamaner of the National Center for Science and Engineering Statistics, National Science Foundation, described the development and implementation of the survey, its history, and its challenges. The authors cite the implementation of a survey of institutional coordinators as part of the survey design, site visits to inform the gathering of background information, and the provision of technical and instructional support as important aspects of the redesign experience.

The papers in this issue have provided evidence of the breadth of current applications of administrative data, and also suggested research opportunities for the future. The papers also enabled readers to gain an appreciation of the challenges that must be embraced or overcome to succeed. New research on the topic of administrative data will be welcomed by NSOs worldwide. Further, it seems likely the *Journal of the IAOS* will revisit administrative data as a theme for future issues.

Nancy K. Torrieri, Ph.D.  
Editorial Board

December 2018  
*Statistical Journal of the IAOS*  
E-mail: [nkctorrieri@gmail.com](mailto:nkctorrieri@gmail.com)