

## Discussion: Uses of Administrative Records Benchmarking in Modern Census Taking

Amy O'Hara<sup>1</sup>

Center for Administrative Records Research and Applications, U.S. Census Bureau

Abstract: A discussion of the session “Uses of Administrative Records Benchmarking in Modern Census Taking” is provided.

Key Words: Administrative Records, Census, Data Quality

### Introduction

The three papers in this session reveal how Norway, Sweden, and the Netherlands use administrative records data to conduct their censuses. Before discussing the papers from each country, I will point out the challenges of using administrative records in the United States (U.S.), particularly in a Census of Population and Housing context. The U.S. has no system of registers; we do have a Master Address File that includes all housing units, a Business Register that includes all establishments and companies, and an employer-employee database. The United States has no Statistics Act; we do have statutory authority in the Census Act<sup>2</sup> to ask federal, state, local, and private entities for data but cannot compel entities to share data.

### Input Data Quality in Register-Based Statistics: The Norwegian Experience

Norway takes a proactive stance on input data quality for their register-based statistics. Rather than repairing errors, they seek to avoid errors in the input data. The Three C's – cooperation, communication, and coordination- are improved by having templates for agreements, data quality reports, and better relationships with data providers. The U.S. is engaged in similar work through the Federal Committee on Statistical Methodology Administrative Records Working Group. Norway and the U.S. share a difficulty with missing address data and accurately reflecting mobile populations. Two major differences exist between Norway and the U.S. involving the frequency of data updates and discussions with the data provider. Regarding frequency, Norway receives daily data updates, while deliveries are quarterly at best in the U.S. For disclosure, Statistics Norway can discuss micro-level anomalies with the data provider to improve data quality, while the U.S. cannot disclose any information back to the source on individual cases. Based on the Norway model the U.S. can strive to professionalize contacts with data providers, creating guidelines for follow-up to improve input data quality. The U.S. continues efforts to produce quality reports on administrative records source data.<sup>3</sup>

### Doing a Register-Based Census for the First Time: The Swedish Experiences

---

<sup>1</sup> Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

<sup>2</sup> The Census Act, 13 U.S.C. § 6, directs the Census Bureau to use administrative records data to the maximum extent possible instead of conducting direct inquiries.

<sup>3</sup> The Federal Committee on Statistical Methodology has been investigating quality measurement, and will be presenting “New Perspectives on the Quality of Administrative Data” at a Policy Seminar in December 2012.

Statistics Sweden conducted its first fully register based census in 2011 using three base registers (Business Register, Total Population Register, and Real Property Register) along with a Dwelling Register created through cooperation with property owners, residents and the Tax Authority. The Dwelling Register includes 96.8 percent of the population who registered their dwelling id-key with the Tax Authority. Sweden noted difficulties with missing and false dwelling ids, and conducted a follow-up study to assess quality. The Swedish census experience parallels recent research in the U.S. – difficulties with people who do not live where they are registered (or appear in administrative data in the U.S. context), and lower coverage of administrative records for young adults. Looking ahead, the U.S. can monitor how Swedish evaluation studies assess the accuracy of their register data, especially regarding dwellings that have no people registered.

#### Evaluation of the Quality of Administrative Data Used in the Dutch Virtual Census

The Dutch paper rated the inputs to the Dutch Virtual Census to determine the effect of data quality by source on the combined result. Fitness of use is critical when assessing data quality; the paper describes the quality frameworks for the Source, Metadata, and Data hyperdimensions. Checklists are used to apply these frameworks to investigate the quality of multiple registers. The paper demonstrated an application of the checklist, scoring each register as good, reasonable, and poor across Source and Metadata Hyperdimensions. In Figure 1, I summarize the evaluation results for the five Dutch registers in the paper by color-coding good as green, reasonable as yellow, and poor as red. The Dutch panel of data is classified as follows:

ER = Education Register  
 UR = Unemployment Benefit Register  
 SR = Social Security Register  
 HR = Housing Register  
 PR = Population Register

Next, I provide a personal assessment of administrative records sources in the U.S. according to the Dutch typology of hyperdimensions. While U.S. administrative records data lack register-level coverage, I offer opinions on six data sources as follows:

AR = Address Register (i.e., the Master Address File at the U.S. Census Bureau)  
 PL = Person List of individuals with Social Security Numbers  
 TL = Tax List of persons, spouses and dependents on individual income tax returns and persons with information returns  
 ML = Medicare List of persons in the public health insurance program for persons 65 and older and some with disabilities  
 FSL = Food Stamps List of persons in the financial assistance program for low- and no-income persons to buy food  
 CL = Commercial Lists including person data purchased from private companies including mailing lists and credit header data

In the Dutch Evaluation panel, only the Education Register scored below reasonable in the Delivery and Comparability dimensions. The poor Delivery score was due to low frequency of delivery (often only once a year). The poor Comparability score was due to

difficulty aligning time period variables to time points used by Statistics Netherlands. I took a more liberal, but consistent, approach when grading the U.S. sources. For the U.S., the Address Register (AR) fares the best in the checklist, followed by the Person List (PL) of Social Security registrants. The Tax List (TL) rates poor for delivery in our current arrangement with the revenue agency, as we receive data in two deliveries not timed well to census purposes. The Medicare List (ML) also rates poor for delivery, as we receive an annual delivery instead of more frequent updates. The Food Stamps List (FSL) rates poor on Supplier, Relevance and Delivery dimensions. The Supplemental Nutrition Assistance Program (Food Stamps) is administered at the state-level, requiring the Census Bureau to initiate contacts and negotiate with many individuals. This affects the Supplier and Delivery dimensions, since the states are often only able to deliver annual extracts with considerable lags. I scored Relevance low because we have only acquired the data from a small number of states; the usefulness and ability to satisfy our data requirements are unknown at this time. The Commercial Lists (CL) fare the worst, with six poor scores. With commercial data, it is necessary to deal with multiple vendors with varying procedures on database development, variable classification schemes, a lack of comparability across vendors and with Census Bureau concepts, poor completeness and quality of unique keys for record linkage and unduplication, and scarce information on the quality control employed by each vendor.

Figure 1: Checklist for Dutch Registers and U.S. Administrative Records  
Sources by Hyperdimensions  
(Good = Green, Reasonable = Yellow, Poor = Red)

	Dutch Evaluation						U.S. Speculation					
	ER	UR	SR	HR	PR		AR	PL	TL	ML	FSL	CL
Supplier	Green	Yellow	Yellow	Green	Green		Green	Green	Yellow	Yellow	Red	Red
Relevance	Yellow	Green	Green	Yellow	Green		Green	Yellow	Yellow	Green	Red	Yellow
Privacy and Security	Green	Green	Green	Green	Green		Green	Green	Green	Green	Green	Green
Delivery	Red	Green	Green	Green	Green		Green	Yellow	Red	Red	Red	Green
Procedures	Yellow	Yellow	Yellow	Green	Green		Green	Yellow	Yellow	Yellow	Yellow	Red
Clarity	Green	Green	Green	Green	Green		Yellow	Green	Green	Green	Yellow	Red
Comparability	Red	Yellow	Yellow	Green	Green		Green	Yellow	Yellow	Yellow	Yellow	Red
Unique Keys	Green	Green	Green	Yellow	Green		Yellow	Green	Green	Green	Green	Red
Data Treatment	Green	Green	Green	Green	Green		Yellow	Yellow	Yellow	Yellow	Yellow	Red

Note: It is a typology adapted from the Dutch Checklist for quality of administrative data

Despite the challenges involved with U.S. administrative records data, particularly dealing with conflicting information in the data, different definitions and classifications across agencies, and incomplete identifiers, recent research indicates that administrative

records may hold promise for decennial census uses. The Census Bureau acquired administrative records data from federal agencies and commercial vendors to conduct the 2010 Census Match Study (Rastogi and O'Hara 2012). In this study, administrative records matched to 92.6 percent of all housing units in the 2010 Census, and to 88.6 percent of all persons in the 2010 Census. The 2010 Census Match Study was a proof of concept for using administrative records data in a U.S. decennial census context. Results indicate that using administrative records may help maintain data quality while reducing costs. Administrative records are being tested to: update and validate the residential address frame, provide new modes (email and text messaging) to connect with households, model response propensities, and to enhance imputation strategies.

### Summary

The Swedish, Norwegian, and Dutch papers are aspirational examples of how to use register/administrative records data in censuses. The U.S. is committed to designing and conducting a 2020 Census that costs less per housing unit than the 2010 Census, while maintaining high quality data. Administrative records are part of the plan to achieve these goals. However, like our European colleagues, the U.S. must investigate and document data quality, integrate the administrative records lists appropriately, and assess coverage. Beyond these technical details, the U.S. needs to build awareness and confidence in the use of these data. The Census Bureau is addressing these challenges through robust research within the agency, dialogue within the Federal Statistical System, and engagement with the statistical community and academia. The research underway will build evidence to inform decisions on how to conduct an efficient modern census, bringing us closer to the European standards yet tailored to the U.S. conditions, as discussed in this session.