

# **Administrative Records in the U.S. Census for Group Quarters Frame: Merits and Drawbacks<sup>1</sup>**

Asaph Young Chun and Jessica Gan<sup>2</sup>

## **Abstract**

The purpose of this paper is to present the use of administrative records in the U.S. Census for Group Quarters, or known as collective dwellings elsewhere. Group Quarters enumeration involves collecting data from such hard-to-access places as correctional facilities, skilled nursing facilities, and military barracks. We discuss benefits and constraints of using various sources of administrative records in constructing the Group Quarters frame for coverage improvement. This paper is a companion to the paper by Chun and Gan (2014), discussing the potential uses of administrative records in the Group Quarters enumeration.

Key words: Administrative records, collective dwellings, group quarters, frame, census

## **1. Introduction**

### **1.1 Background**

Administrative records (AR) are data collected primarily for the purpose of carrying out administrative programs (Federal Committee on Statistical Methodology<sup>3</sup>, 1980). Researchers have used AR in creating alternative population or sampling frames, refining the estimation process by the use of auxiliary variables extracted from an AR database, and making nonresponse follow-up more cost effective<sup>4</sup>. Though the collection and use of AR is as old as the period of Egyptian pharaohs, major advances of integrating administrative records into census taking have occurred only in the past few decades (Nordbotten, 2009; Chun and Scheuren, 2011). Arguably, by leveraging the existing, centralized administrative records system and for controlling accelerated census cost and sustaining data quality, an administrative-record-based census taking has replaced a traditional census taking in several European countries, led by Nordic countries (Wallgren and Wallgren, 2007; Sundgren, 2010).

The U.S. Census is, in principle, conducted to enumerate people in typical households and group quarters (GQs), or known as collective dwellings elsewhere. The GQ enumeration involves collecting demographic information from places where people who are not usually related to each other stay together in a living arrangement that is owned or managed by an entity providing housing or services for the residents (Williams, De Vos, Russell, and Barrett, 2013). These GQs include facilities, such as college and university residence halls, correctional institutions, military barracks, nursing facilities, and service-based locations, such as emergency and transitional shelters, soup kitchens, and regularly scheduled mobile food vans. Given collective nature of group dwelling, a roster or administrative data about people in GQs tends to be maintained by an organizational entity that provides housing services. However, the dearth of published research is obvious when one pays attention to the application of administrative records particular to GQs. In the seminal 1980 conference the Federal Committee on Statistics Methodology hosted on the use of administrative records for statistical purposes, none of papers addressed GQs. The International Methodology Symposium Statistics Canada devoted in 2010 to the interplay of administrative data with census and surveys did not include any papers addressing challenges or opportunities in the use of administrative records in GQs or similar cases.

---

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

<sup>2</sup> Asaph Young Chun, U.S. Census Bureau, Room 4K273, Washington, DC 20233. Jessica Gan, Rice University.

<sup>3</sup> The Statistical Policy Working Paper 6 - Report on Statistical Uses of Administrative Records prepared by Federal Committee on Statistical Methodology (1980), Washington DC. is the first comprehensive U.S. report discussing various ways administrative records can be used to supplement surveys and censuses.

<sup>4</sup> See the FCSM (1980) Working Paper 6 to find examples illustrating a variety use of administrative records.

A few recent studies have started to address this issue in the GQ context. In particular, ethnographic studies conducted for the 2010 Census isolated challenges and opportunities of using administrative data for measuring census coverage in jails and prisons, group homes for women, military barracks, and skilled nursing facilities (Chan, 2012). The Census Bureau's assessment staff revealed the merit of using administrative data in constructing the frame of GQs (Williams, Barrett and Williams, 2013) and GQs enumeration (Chun and Gan, 2014) in the 2010 Census. These recent findings support findings Jonas (2003) reported for the 2000 GQs Enumeration that administrative records were the primary source of data used by almost half of GQs questionnaires (i.e., 3.7 million out of 7.5 million GQs questionnaires). Jonas further pointed out that at most types of GQs, administrative data was the primary means of completing the Individual Census Reports, the GQs questionnaire for each person living in a GQ – 73 percent of nursing homes; 66 percent of hospitals; 60 percent of group homes; 56 percent of correctional institutions; and 49 percent of juvenile institutions. His evidence-based recommendation to further test and tailor the use of administrative data for targeted GQs supports the hypothesis that administrative records could help design a cost effective and high quality GQs operation. Chun, Chan et al (2012)<sup>5</sup> explored a total GQs error and cost model to identify optimal trade-off between various sources of GQs coverage errors and associated cost components. Yet these recent research efforts are limited in their scope and impacts. For example, a total survey error and cost paradigm Chun, Chan and their associates extended to GQs is preliminary without GQs error and cost data. The National Research Council (2011) encouraged the Census Bureau to carry out studies that explore the use of AR to supplement and improve the traditional GQs data collection. Our paper is an attempt to address this scientific concern with our eyes especially fixed on the GQs frame construction<sup>6</sup>.

In this paper, we demonstrate the utility of AR particularly in constructing the GQs frame for improvement in coverage. We analyze the results of the 2010 Census to investigate the extent to which AR were used for the GQ frame and discuss merits and limitations of various sources of AR used for validating GQ units.

## 1.2 Group Quarters Validation Operation Overview<sup>7</sup>

The 2010 GQs Validation (GQV) was the GQs list-building operation where all potential GQs were followed up. The Census Bureau staff identified Other Living Quarters (OLQs) as potential GQs through the listing field work, partnership programs such as Local Update of Census Addresses (LUCA), and the existing list of Census 2000 GQs, among other sources. The GQV listers were sent to each of OLQs to verify that an address had the correct census geography, and validate whether or not it was a valid GQs for the 2010 Census. If the lister validated the address as a GQ, then they determined the type of GQ based on information collected in the GQV interview, and collected additional information about the GQ. Listers conducted this work using the GQV Questionnaire designed to facilitate personal interview.

## 2. Analytical Approach

The main research question we address is to assess whether AR is useful to GQ frame construction, and if so, to appraise the extent to which AR can be instrumental to supplementing the GQ frame, and investigate the contribution of various sources of AR to the frame of GQs.

### 2.1 Analysis of 2010 GQV Frame Sources

The 2010 Census GQV frame consisted of the pre-identified addresses that were included in the initial workload as well as the number of addresses added during GQV. The 2010 Census GQV operation received all addresses identified as OLQs during AC and addresses identified as potential GQs by other selected sources, including the sources below. This provided a list of potential GQs that is subject to individual validation in the GQV operation. The focus of our analysis is to discuss the contribution of administrative records in the last three groups to the GQs frame relative to the contribution by non-administrative records.

- Census 2000 GQs (Pre-2010 Census) - GQ addresses from Census 2000 plus any mid-decade updates
- OLQs identified during the 2010 AC operation
- GQs identified during the American Community Survey Time of Interview (ACS TOI) operation

---

<sup>5</sup> See Chun, Chan, Barrett, Williams, and Belton (2012). "What Counts as Group Quarters? A Glimpse of Total Coverage Errors and Costs Model in Census." Paper presented in the 2012 Joint Statistical Meetings in the Topic-Contributed Session organized by Chun, Chan and Scheuren. San Diego, CA.

<sup>6</sup> One may find background and details of key findings discussed in this paper in the 2010 Census Group Quarters Validation Assessment Report prepared by Williams, Barrett and Williams (2013).

<sup>7</sup> One may find details of GQV operation in Williams, Barrett and Williams (2013).

- GQs from *administrative records*, such as the Federal-State Cooperative Program for Population Estimates (FSCPE) - a GQ Report file from participating representatives to the FSCPE, as well as business files for the nation and the state.
- GQs addresses provided by government participants of the Local Update of Census Addresses (LUCA) operation, which we consider as *administrative records* in this paper
- GQs identified by the National Processing Center's (NPC) internet research. Staff at the NPC searched the internet for Service Based Enumeration (SBE) and Group Homes as well as solicited information from the Highest Elected Official at tribal and national advocacy organizations for people experiencing homelessness to update the address list of SBE locations (e.g., shelters, soup kitchens). We consider this source as *administrative records*.

## 2.2 Data: 2010 GQV and 2010 Enumeration Frame

The 2010 GQV frame file included addresses identified as OLQs during AC and addresses existing in the Master Address File identified as potential GQs by other selected sources. These addresses were sent to the GQV operation for confirming the OLQ's status as a GQ, Housing Unit, and Transitory Location, among others. The 2010 enumeration frame file was used to identify all the records that were in the GQV frame and whether or not they were also in AC. In addition, this file had values for the variables used to determine the AC status of the OLQ, and was also used to identify the enumeration path of all records in the initial GQV frame<sup>8</sup>. There were multiple sources of administrative and non-administrative records that occasionally provided the same address to the GQV frame. When there was overlap, we were not able to conduct refined analysis by different sources' contributions. However, we were able to identify single vs. multiple sources of the same address. This method allowed us to isolate the single source address and identify those addresses provided by sources that were already listed on the Master Address File from 2000 (i.e., Pre2010).

## 3. Findings and Discussion

### 3.1 GQV Frame by Source

We examined the GQV outcome by the various sources that provided potential GQ addresses to the initial frame of GQs as well as the final frame of validated GQs prior to GQ Enumeration (GQE). This analysis assesses the impact of several list-building operations.

Table 1 shows the number of addresses provided by each source to the GQV frame. In those instances where there were multiple sources for an address, each source was credited with contributing the address to the GQV frame. Thus, the number of addresses provided by each source sums to more than the number of units in the GQV frame because of the overlap when multiple sources provided the same addresses. Approximately 58 percent of the addresses in the GQV frame were already on the frame (i.e., Pre-2010 Census source). Administrative records, including LUCA, FSCPE and Internet Research of SBE locations and group homes, provided a total of about 17 percent of addresses in the GQV frame. AC Adds and GQV Adds contributed a total of 16 percent and 20 percent, respectively, in the GQV frame.

---

<sup>8</sup> The final statuses we present in this paper are as of the GQV status. This status may change later in the remaining operation of the Census.

**Table 1.  
Number of Addresses Provided to the GQ Initial Frame by Source -  
Includes Overlap**

Source	Count	Percent of Units in GQV Frame
Total Units in GQV Frame.....	2,551,447 <sup>+</sup>	100.00
<i>Administrative Records Sources</i>		
LUCA.....	339,206	13.29
FSCPE.....	45,678	1.79
NPC/SBE and GH.....	55,435	2.17
<i>Other Sources</i>		
Pre-2010 Census.....	1,467,211	57.51
ACS Time of Interview.....	4,006	0.16
AC Adds.....	418,779	16.41
GQV Adds.....	511,094	20.03
Notes: Counts and percentages are unweighted. <sup>+</sup> Counts and percentages do not add up to the total since multiple sources provided the same address. Data Sources: 2010 GQV Initial Universe MTdb Extract, 2010 Enumeration MTdb Extract, and the 2010 National ADDUP file.		

### 3.2 GQV Initial Frame by Source for Valid Living Quarters

Table 2 provides an overall summary of the GQV address outcomes identified as valid living quarters (GQs, vacant GQs, HUs, and TLs) and non-valid living quarters (duplicates, deletes, and non-residential) by source. In Table 2, where there were multiple sources for the same address, each source was credited with contributing the address to the GQV frame. As in Table 1, the number of addresses provided by each source sums to more than the number of units in the GQV frame because of the overlap when multiple sources provided the same address. Percentages reported for each source in the following sections are those that accounted for such an overlap.

#### LUCA (Administrative Records)

As shown in Table 2, about 12 percent of the addresses provided by LUCA were validated as GQs during GQV. Approximately 46 percent of the addresses from LUCA were duplicates, deletes or non-residential addresses, approximately 39 percent were HUs, and 3 percent were TLs.

#### FSCPE and NPC's Internet Research for SBE Places (Administrative Records)

The FSCPE and Internet Research of SBE locations and group homes were part of the Administrative Records Updating Operation. Geographic Programs developed procedures to update the GQ frame by using administrative records from tribal and local governments, national and state advocacy organizations and other administrative records. They also used the Internet to update SBE locations and group homes. Most of the addresses provided by FSCPE and NPC's Internet Research were validated as GQs during GQV. As Table 2 shows, approximately 61 percent of the 45,678 addresses provided by FSCPE and about 57 percent of the 55,435 addresses provided by NPC's Internet search for SBE locations and group homes were found to be GQs (including vacant GQs). Of the GQs provided by NPC's Internet research for SBE locations and group homes, over half (51.1 percent) were validated as group homes during GQV. Of these, there were 14,593 Group Homes Intended for Adults and 1,563 Group Homes for Juveniles.

#### Pre-2010, 2010 AC Adds and 2010 GQV Adds<sup>9</sup>

Also shown in Table 2, of the almost a million and a half addresses with a Pre-2010 source in the GQV frame, about nine percent were validated as GQs. Of the 418,779 OLQs added by the 2010 Address Canvassing operation, seven percent were GQs (includes vacant GQs). Of the over half a million units added by GQV listers, almost four percent were GQs, including vacant GQs.

#### GQV Initial Frame by Single vs. Multiple Sources

An additional analysis discussed by Williams, Barrett, and Williams (2013) - not shown here due to limited space - presents each source that contributed to the initial GQV frame by analyzing the number of addresses that were

<sup>9</sup> We remind the reader that AC added potential GQs, and GQV added GQs and had to add, by design, all the housing units if they turned out to be regular housing.

provided as follows: a single source; multiple sources where at least one of the sources was Pre-2010 Census; and multiple sources where all of the sources were added after the 2010 operations (i.e., non Pre-2010 sources). This is the approach taken to addressing the issue of the overlap when multiple sources were present to contribute to the same address in the GQV frame.

For all of the non Pre-2010 sources, most of the addresses they provided were already on the frame. The percentages ranged from approximately 59 percent to 73 percent. Of the 339,206 addresses provided by LUCA, approximately 59 percent were already on the frame. However, LUCA was the sole source for approximately 37 percent of their addresses that contributed to the GQV frame, the highest percentage among all sources that were categorized as the sole source<sup>10</sup>. Over 70 percent of the addresses provided by administrative records from FSCPE (32,584 of 45,678) and the Internet research of SBE locations and group homes (SBE/GH 40,591 of 55,435) had multiple sources where at least one was a Pre-2010 source. Administrative records from FSCPE and SBE/GH was the sole source for about 14 percent and 15 percent, respectively.

**Table 2.  
Initial GQV Address Status Outcomes by Source – Includes Overlap**

Source of Unit Addresses	GQs and Vacant GQs		Housing Units		Transitory Locations		Duplicates, Deletes, Non-residential			
	Count*	Percent of Total <sup>+</sup>	Count*	Percent of Total <sup>+</sup>	Count*	Percent of Total <sup>+</sup>	Count*	Percent of Total <sup>+</sup>		
<b>Total</b>	2,551,447	100	205,406	8.05	1,391,477	54.54	80,483	3.15	874,081	34.26
<i>Administrative Record Sources</i>										
LUCA	339,206	100	41,512	12.24	131,672	38.82	10,577	3.12	155,445	45.83
FSCPE	45,678	100	27,792	60.84	5,923	12.97	129	0.28	11,834	25.91
SBE/GH	55,435	100	31,636	57.07	11,697	21.10	529	0.95	11,573	20.88
<i>Other Sources</i>										
Pre-2010 Census	1,467,211	100	133,222	9.08	747,368	50.94	40,896	2.79	545,725	37.19
ACS Time of Interview	4,006	100	326	8.14	1,402	35.00	403	10.06	1,875	46.80
AC Adds	418,779	100	28,768	6.87	98,583	23.54	33,794	8.07	257,634	61.52
GQV Adds	511,094	100	19,352	3.79	490,051	95.88	1,045	0.20	646	0.13

\*Counts and percentages are unweighted. Counts may not add up to the total since multiple sources provided the same address. Percentages may not sum to 100 due to rounding. Data Sources: 2010 GQV Initial Universe MTdb Extract, 2010 Enumeration MTdb Extract, and 2010 National ADDUP file.

For the sources providing GQ addresses, Williams, Barrett, and Williams (2013) report the percentage of those addresses that were already on the list ranged from almost 60 percent to over 76 percent. Administrative records, such as FSCPE and SBE/GH sources, contributed the highest percentages (over 70 percent respectively) of addresses that were already on the frame. LUCA provided about 60 percent of GQ addresses that were already on the frame. Although LUCA as a sole source provided more valid living quarters addresses overall compared with ACS TOI, FSCPE, and the NPC Internet Research alone, they also provided more non-valid addresses (39 percent) as a sole source compared with the other sources alone.

When compared among sole sources, LUCA as a sole source of administrative records also provided more addresses

<sup>10</sup> We remind the reader that the LUCA records in this analysis are not the complete set of OLQs submitted by LUCA participants.

of housing units (39 percent, or 50,877 of 131,672 and of TLs 42 percent, or 4,429 of 10,577 TL addresses). Over seven percent (9,186) of the 125,183 addresses with LUCA as a sole source in the GQV frame were validated as GQs, including vacant GQs.

#### 4. Contribution of Each Source to the Post-GQV Frame

As shown in Table 4 below, there were a total of 205,406 validated GQs - including 44,120 vacant GQs at the time of GQV interview - found during the 2010 GQV operation. Among the different sources, the Pre-2010 made the greatest contribution to the frame of GQs in that they accounted for approximately 65 percent of the 205,406 GQs found during the GQV operation. LUCA also provided a large number of addresses that GQV confirmed as GQs with over 20 percent of such addresses. The FSCPE and internet research of SBE and Group Homes - the two other sources of administrative records - were also found as a notable contribution to the frame of validated GQs: 14 percent and 15 percent, respectively. The percentages include overlap with other sources and thus do not sum to 100 percent.

**Table 4.**  
**Number of Addresses in the Post-GQV Frame by Source –Includes the Overlap**

Source of Unit Addresses	Count*	Percent of Total <sup>+</sup>	GQs		Vacant GQs	
			Count*	Percent of Total <sup>+</sup>	Count*	Percent of Total <sup>+</sup>
Total in Post-GQV Frame	205,406	100.00	161,286	100.00	44,120	100.00
<i>Administrative Records Sources</i>						
LUCA	41,512	20.21	33,338	20.67	8,174	18.53
FSCPE	27,792	13.53	25,628	15.89	2,164	4.90
SBE/GH	31,636	15.4	27,999	17.36	3,637	8.24
<i>Other Sources</i>						
Pre-2010 Census	133,222	64.86	103,604	64.24	29,618	67.13
ACS TOI	326	0.16	189	0.12	137	0.31
AC Adds	28,768	14.01	19,212	11.91	9,556	21.66
GQV Adds	19,352	9.42	18,471	11.45	881	2.00

\*Counts and percentages are unweighted.

<sup>+</sup>Counts and percentages do not add up to the total since multiple sources provided the same address.

Data Sources: 2010 GQV Initial Universe MTdb Extract, 2010 Enumeration MTdb Extract, and the 2010 National ADDUP file.

#### 5. Conclusions and Future Research Agenda

Administrative records were found to be useful to improve coverage of the GQ frame in the 2010 Census. It has been verified that the three primary sources of administrative records - LUCA, FSCPE, and NPC's internet search of SBE and groups homes - contributed considerably to supplementing the GQ frame, yet at a varying rate. LUCA as a sole source provided more valid living quarters addresses - approximately 37 percent of the addresses that contributed to the GQV frame - when compared with ACS Time of Interview, FSCPE, and the NPC Internet Research alone. However, LUCA also provided more non-valid living quarters addresses (39 percent) as a sole source compared with the other sources alone. Administrative records from FSCPE and SBE/GH was the sole source of the GQ frame at 14 percent and 15 percent, respectively. Observing merits and limitations of administrative records relative to their effective use for GQ validation and frame construction, we offer the following research be

considered for implementation in the near future:

First, although there were identified GQs types that made substantial contributions to the frame of GQs, about 22 percent (44,120 of the total 205,406 GQs) of the units were vacant GQs that had an unknown GQs type. To minimize the number of GQs with an unknown GQs type, we recommend researching the availability of administrative records that could be used to assign possible GQs type codes to subsequent GQs operations. Second, it is recommended to conduct studies to identify GQs types where the GQ Validation rate with administrative records can be optimized and cost-effective. The Census Bureau is encouraged to continue to obtain administrative records from federal, state, tribal and local governments. The combined sources of administrative records (FSCPE and the Internet research of SBE locations and group homes) contributed to the frame of GQs at 29 percent. Furthermore, we reiterate that the majority of the addresses provided by administrative records (i.e., 61 percent of the 45,678 addresses provided by FSCPE and 57 percent of the 55,435 addresses provided by the NPC Internet research) were validated as GQs during the 2010 GQV operation. Third and lastly, we may explore innovative ways to use the Business Register developed by the Census Bureau for the economic census to help identify certain GQs on an ongoing basis. We also should explore a more effective and cost efficient method to build upon the Internet research that was conducted in 2010 to update SBE locations and group homes, and include other group quarter types. For example, websites of colleges and universities may offer current information about student housing and college dorms. The NPC Internet research of SBE locations and group homes found 3,900 GQs. These addresses were not already on the Master Address File, nor did another source submit the address to be included in the 2010 GQV frame.

## References

- Chan, A. (2012), "2010 Census Program for Evaluations and Experiments, Center for Survey Measurement", *Investigation of methods to evaluate the coverage of group quarters populations report*.
- Chun, Y., Chan, E., Barrett, D., Williams, A., and Belton, J. (2012), "What Counts as Group Quarters? A Glimpse of Total Coverage Errors and Costs Model in Census". Paper presented in the 2012 Joint Statistical Meetings in the Topic-Contributed Session organized by Chun, Chan and Scheuren. San Diego, CA.
- Chun, Y, and Gan, J. (2014), "Analysis of the Sources of Group Quarters Enumeration Data in the 2010 Census" Paper presented in Joint Statistical Meetings Survey Research Methods Session, Boston, MASS.
- Chun, Y, and Scheuren, F. (2011), "Pandata Systems to Enhance Survey and Census Systems: Employing Administrative Data to Better Inform Public Policy", Invited paper presented in the 2011 Joint Statistical Meetings.
- Jonas, K. (2003), *Group quarters enumeration. Census 2000 Evaluation E.5, Revision I.*, Extracted from <https://www.census.gov/pred/www/rpts/E.5%20R.pdf>
- National Research Council of the National Academies. (2011), *Change and the 2020 Census: Not Whether but How*.
- Nordbotten, S. (2009), "Use of Administrative Data in Official Statistics – Past, Present and Future – With Special Reference to the Nordic Countries", Official Statistics in Honour of Daniel Thorburn, pp 205-223.
- Russell, D., & Barrett, D. F. (2012). *2010 census service-based enumeration operation assessment report*. United States Census Bureau, Decennial Management Division/Decennial Statistical Studies Division, <http://www.census.gov/2010census/pdf/2010%20Census%20Service-Based%20Enumeration%20Assessment%20Report.pdf>
- Sundgren, B. (2010), "Towards a system of official statistics based on a coherent combination of data sources, including surveys and administrative data.", Paper presented at the European Survey Research Association annual conference, Lausanne, Switzerland.
- Wallgren, A., & Wallgren, B. (2007), *Register-based statistics: Administrative data for statistical purposes*. John Wiley & Sons, Ltd.
- Williams, J., Barrett, D., Williams, A. (2013), *2010 census group quarters validation assessment report*. U.S. Census Bureau, Suitland, MD. [http://www.census.gov/2010census/pdf/2010%20Census%20Group%20Quarters%20Validation%20Operation%20Assessment%20\(Reissue\).pdf](http://www.census.gov/2010census/pdf/2010%20Census%20Group%20Quarters%20Validation%20Operation%20Assessment%20(Reissue).pdf)

Williams, J., De Vos, B., Russell, D., Williams, A., and Barrett, D. (2013), *2010 Census group quarters enumeration assessment report*. United States Census Bureau, Extracted from <http://www.census.gov/2010census/pdf/2010%20Census%20Group%20Quarters%20Enumeration%20Assessment%20Report.pdf>