Qualitative Assessment of Administrative Records: The Case of State Prisons' Records

Anna Y. Chan

Center for Survey Measurement
U.S. Census Bureau
Paper prepared for FCSM 2012 conference

Key words: Group Quarters, Administrative Records, Data Quality.

Abstract

This study examines data quality of administrative records (ARs) obtained from state prisons. It assesses five data quality indicators: relevance, timeliness, coverage, item non-response, and data consistency of the ARs. Analyses were based on ethnographic data, information provided by ARs, and self-responded survey with inmates in state prisons. Results from this study show that ARs delivered by state prisons are relevant, timely and may have excellent coverage if a complete list of ARs is obtained. The demographics information provided by the ARs for individual inmates was complete with no item non-response but the race/ethnicity data did not correspond well to those reported by inmates. Classification of race and ethnicity is inaccurate using ARs that used a combined race and ethnicity question with single response option. Implications for future research on ARs include (1) the examination of potential estimate biases in the Hispanic origin and race data generated using ARs; (2) the need to use multiple qualitative methods to identify and triangulate study issues systematically; and (3) the need to understand prison culture to design proper communication methods and procedures to obtain a complete list of ARs for inmates in state prisons.

Qualitative Assessment of Administrative Records: The Case of State Prisons' Records

Anna Y. Chan, Ph.D.

Center for Survey Measurement
U.S. Census Bureau

1. Introduction

This study utilizes a qualitative evaluation research method to examine Administrative Records (ARs) from correctional facilities, in particular, state prison records. The evaluation goal of this research is exploratory and formative in nature (Trochim 2000). It measures the data quality of ARs from state prisons using five data quality indicators. It examines the *relevancy* of ARs on how well the records match the data needs of the decennial census and the *timeliness* of such data. It assesses the *coverage* of ARs by comparing different sets of ARs from the same prisons and examines the *consistency* of data by comparing information provided by ARs to self-responded survey data collected from the same inmates in three state prisons. It explores the underlying reasons why ARs may have coverage error and are inconsistent with self-responded data. The results provide insights on relevancy of ARs and how well the records match the data needs of the decennial census. Findings for this study are based on data collected from ethnographic studies in 2010, an inmates" self-response survey study conducted in 2010 and 2011 in state prisons from two western states and ARs data from the same prisons. Findings from this study will have implications for future research on usage of administrative data from state prisons.

2. Background

2.1 Decennial Group Quarters Enumeration (GQE) and the Usage of Administration Records

Once every ten years the U.S. Census Bureau counts the entire U.S. population. Although Administrative-record based census has replaced the „self-reported" or „enumerator-administered" census in many Nordic countries for decades (Holmberg 2011; Mulalic 2011; Thygesen 2011), in 2010, the U.S. census continued to us a self-administered census form. The paper census form was mailed to each address classified as housing unit such as a house, apartment, mobile home or trailer. The census form collects household information and demographics for every person residing in the housing unit on Census Day – April 1, 2010. The U.S. Census Bureau enumerates the „group quarters" population in a separate operation that is conducted differently from the enumeration of populations residing in housing units. Group Quarters (GQs) are places where people live or stay in a group living arrangement that is owned or managed by an entity or organization that provides housing and/or services for residents (U.S. Census Bureau 2009). GQs include places such as correctional facilities (jails and prisons), skilled nursing facilities, college or university student housing, military barracks, juvenile facilities, living quarters for migrant workers, homeless shelters, group homes and religious living quarters (U.S. Census Bureau 2009). Given GQ residents are mostly unrelated to one another; each GQ resident receives an Individual Census Report (ICR) form, which collects personal information about the person only. There are three acceptable methods to complete the ICR form: (a) self-administered by the GQ residents; (b) enumerator administered face-to-face interview; or (c) proxy interview or information provided by staff, family members, or ARs.

Almost 52 percent of decennial census data collected from 7.1 million ICR forms came from ARs (Jonas 2003) and about a quarter of the ICRs were classified as non-response. Among the 1.9 million people residing in correctional institutions, 56.3 percent of the ICRs were completed by using ARs, 15.3 percent were completed by the residents, and another 4.4% were interviewed in person. The remaining 24 percent of ICRs distributed to correctional facilities in Census 2000 ended up as unit non-response (Jonas 2003).

Enumerating and accessing these facilities and their residents is a challenge. It is difficult to conduct in-person enumeration or to distribute self-administered census forms in some facilities. It could be operationally infeasible for some facilities to accommodate the census takers without adding cost (overtime wage for staff) and burden

(interruption of daily operation). Many large GQs choose to provide administrative records (ARs) to enumerators to complete the census forms. Hence, the use of ARs as primary source of data for generating census statistics is common and often necessary when the target population resides in GQs.

Although ARs have been used as a source of decennial census group quarters data for quite some time and may become the major source of data for the 2020 group quarters' enumeration (U.S. Census Bureau 2011), knowledge regarding the quality of administrative data usage for group quarters and its impact on the statistics is limited. Given the production of high quality statistics depends on the quality of the input data, it is important to systematically assess the quality of administrative data with some standard metrics. This study is the first of a planned series of qualitative evaluation research of AR collected from different types of GQs. This research will start with examining records from correctional facilities; in particular, it focuses on state prison"s records.

2.2 Ethnographic Studies on Correctional Facilities - 2010 Census

The research on ARs described in this paper was part of a larger ethnographic study of different types of GQ facilities during the 2010 Census (Chan 2009). The emphasis in ethnography is to study an entire culture. The ethnographers become immersed in the culture as an active participant and record extensive field notes. The focus of that study is to identify contextual and social factors that may affect the Census Bureau's ability to conduct a Census Coverage Measurement (CCM) survey in GQs that yields quality data. Of relevance to this study, two of the seven ethnographers from the GQs CCM study chose correctional facilities where they have security clearance and ongoing work relationships to conduct the observations and interviews with staff and inmates. The researchers obtained copies of ARs printed for census enumerators during the 2010 Census. Staff from two prisons printed an additional alternate roster on Census Day for the researchers.

A major issue with the ARs used for statistical purposes is that the data provided by the records are collected and maintained by facilities or organizations, in this case, correctional facilities. The quality and content of the records are determined by the record-keeping systems in the correctional facilities and the researchers need to understand the source of such data and its respective record-keeping systems. The ethnographic data suggests that state prison record systems are decentralized and not uniform between and within each state. The observed facilities kept detailed daily records that they submit to the state agency. An important function of qualitative assessment of administrative data is that the results can help assess the need for exclusion of certain types of GQs for the 2020 CCM program, in particular, those that are enumerated with high quality AR.

2.3 Quality Assessment for Group Quarter Administrative Records

To understand the quality of group quarters administrative data, quality frameworks and metrics are needed to measure the quality of the ARs. In contrast to surveys, the tools for evaluating the quality of register-based data[3] are lacking (Berka et. al. 2011). Recent work by Daas and colleagues (Daas et. al. 2009; Nordholt, Ossen, and Daas 2011) has developed a comprehensive multi-dimensional quality framework to standardize the determination of various quality components of administrative data sources in the Netherlands. The framework consists of three high level views on quality referred to as hyper dimensions: Source, Metadata, and Data (for more details regarding the full framework, see Daas et al. 2009). In this exploratory assessment study, we propose a modified framework to guide the conceptualization and measurement of GQ ARs data quality.

According to Daas et al, (2009), a quality indicator is measured or estimated by one or more methods that can be qualitative or quantitative. In this study the measures for the quality indicators utilized both qualitative and quantitative methods. Four dimensions and their respective quality indicators are identified for this assessment study. Table 1 provides the definitions of data quality indicators chosen for this study based on the quality framework developed for on register-based data[2] and other health record studies (Nordholt, Ossen, and Daas 2011; Pipino, Lee and Wang 2002). The five data quality indicators include: fitness of purpose, timeliness, coverage, item non-response and consistency of the data.

Table 1: Quality Framework for State Prisons' Administrative Records

Dimensions	Quality Indicators	Definitions
Relevance	Fitness of purpose and Usefulness	meet the purpose of data collection and to the extent to which data is applicable and helpful for the statistical goal at hand
Time-related	Timeliness, Updates & Time lag	point in time data; the extent to which the data is sufficiently up-to-date for the task at hand
Accuracy	Coverage (i.e., over-coverage and under-coverage)	the extent to which the data cover the target population and are of sufficient breadth and depth for the task at hand
	Item Non-response	the extent to which the data for each record is complete
Reliability	Consistency	the extent to which data from different sources are comparable

3. Research Questions

This exploratory research study assesses the data quality of state prison records by addressing the following research questions.

1. Are the data provided by the state prisons" ARs relevant and met the census data needs?
2. Are the state prisons" ARs data timely and up-to-date?
3. Are the state prisons" ARs data accurate in terms of coverage and item responses?
4. Are the state prisons" ARs data consistent with other sources of data?

4. Methods

4.1 Data Sources

This study uses three sources of data to triangulate the assessment of the data quality of ARs. The sources of data came from (1) ethnography, (2) alternate rosters/ARs; and (3) self-administered survey with selected inmates. The next three sections describe each of these data source respectively.

4.1.1 Ethnographic Data

An ethnographic approach is ideal for study the correctional facilities' culture. This research utilized two qualitative ethnographic data collection methods: (1) participant observation[4] and (2) unstructured interview. Two ethnographers collected observation data prior to, during, and after the 2010 decennial census. Staff and census takers provided feedback and information regarding the census operations, enumerations, and administrative record keeping system during the unstructured interviews. Ethnographers obtained these data from three prisons during 2010 (one male prison and two female prisons.) See ethnographic data in Table 2 below.

Table 2. Data Sources

	Ethnographic Data Participant Observation and Unstructured Interviews in 2010	**ARs for Census Takers** Printed on Census Day (April 1st, 2010)	**Alternative ARs** Inmate Rosters printed at the Time of the Prison Survey in 2011	**Survey Instrument** Self-Administered ICR-Liked Form (Year of interview)
State 1				
Male Prison 1	X	X		X (2010)
State 2				
Female Prison 1	X	X	X (2011)	X (2011)
Female Prison 2	X	X		
Male Prison 2			X (2011)	X (2011)

4.1.2 Administrative Records

All ARs obtained for this study were printed in 'real-time' where the records were up-to-date with minimal lag time between movements and record updates. The two ethnographers obtained two different types of ARs in 2010: 1) ARs for census takers; and 2) alternate rosters for researchers. The ethnographers received three copies of ARs printed for the census enumerators during the 2010 Census from the same three prisons that were observed in the ethnographic study (labeled as Male Prison 1, Female Prison 1 & 2 in Table 2 above).

The 2010 ARs obtained from the Male Prison 1, which has inmates participated in the prison survey study, were used to evaluate the data consistency of ARs. Alternative ARs were obtained from a fourth prison (Male prison 2) and Female Prison 2 in 2011 when inmates from those facilities participated in the prison survey study (see section 4.1.3 below). Information from these ARs was included in the data consistency evaluation of ARs.

4.1.3 Inmates' Self-Response Survey Study

Three correctional facilities from two western states participated in a prison survey study in 2010 and 2011 (Male Prison 1, Female Prison 2 and Male Prison 2). A total of 155 inmates participated in the study. Respondents were asked to complete a self-administered census-like paper instrument, which contained items identical to those used in the 2010 ICR form for group quarters' enumeration. For inmates who had literacy issues, could not read without glasses, or were locked inside their cells, these questions were administered in-person by the researchers. The instrument collected demographic information of the inmates and their alternate addresses if they were not residing at the prison during Census Day. All respondents knew that their participation was voluntary and that the information they provided would be confidential. Below is a brief description of each of the study sites and the sample selection methods. ARs containing the name, date of birth, race and or ethnicity were obtained at the time of the interviews and were compared to the self responses of the inmates to examine the consistency of the ARs data.

4.1.3.1 Survey Study Site One: Male Prison in State one

During the 2010 Census, inmates from a maximum security male prison from a north western state participated in the prison study. A total of 87 inmates participated in the self-administered questionnaire study. On April 2nd 2010, the ethnographer in that state, equipped with the ARs of the selected inmates, distributed the self-administered questionnaire to the sample inmates and collected demographic information from these inmates while verifying their information on the records. A sample of 50 inmates was randomly selected from the general population by the facility administrator. Forty-eight of the 50 inmates participated in the survey. Another 12 inmates from the general population who volunteered to participate in the study were working on the activities floor at the time when the interview was conducted. Another 14 interviews were conducted with inmates at their cell-front in three special housing units (disciplinary segregation unit –DSU[1], the intensive management unit -IMU , and administrative segregation unit) where the researcher filled out the form and several of the selected inmates had refused to participate). Another 13 inmates were briefly interviewed at a voluntary meeting of the inmates" Chicano Club.

4.1.3.2. Survey Study Site Two and Three: Male and Female Prison in State 2

In the spring of 2011, inmates from one maximum security male prison and one female prison participated in the prison study. The inmates in the male facility are often segregated by race in this state, resulting in the populations of some units of buildings belonging to one racial and ethnic group. Thus, individual building is unrepresentative of the prison in terms of its race and ethnic component. Forty prisoners were randomly selected using cluster sample from each of the two prisons. Each prison was divided into "yards", which are clusters of buildings separated by walls and fencing. Only yards housing general populations (inmates that are allowed to move around freely without staff escort) were eligible for the study. During the initial selection phase, one eligible yard was selected from each prison. Housing units with a concentration of one single racial or ethnic group were excluded from the cluster sampling phase. One eligible building was then randomly selected from the clusters of buildings in the identified yard. Thirty of the 40 selected male inmates (75%) participated in the study while thirty-eight of the selected 40 female inmates (95%) participated in the study. ARs of all participants were collected at the time of the interview. The female prison had been observed in 2010 ethnographic study.

4.1.3.3 Sample Demographics

The basic demographic characteristics of the inmates who participated in the self-administered survey study are shown in Table 3. Three quarters of them were male (117/155). About one third of the respondents self-identified as White (34.8%), and the remaining two thirds of the respondents self identified as non-white or Hispanics (race unknown) or of more than one race. The self-reported mean age at the time of the interview was 41.0 (s.d. =13.09, range 19-84). Mean age based on administrative record was 40.7 (s.d. 13.41, 19-84).

Table 3. Self- Reported Demographics Characteristics of Survey Participants by Sample Sites.

	2010 Interviews	2011 Interviews		All Interviews
	Male Prison 1 (N=87)	Male Prison 2 (N=30)	Female Prison 1 (N=38)	Respondents (N=155)
Age	39.4 (s.d.=11.9) (21-80)	47.0 (s.d.=17.2) (24-84)	39.8 (s.d. 10.71) (19-63)	41.0 (s.d. 13.09) (19-84)
Male	100%	100%	NA	75.5%
Female	NA	NA	100%	24.5%
Race				
White	26 (31.0%)	14 (46.7%)	14 (36.8%)	54 (34.8%)
Black	9 (10.3%)	6 (20.0%)	15 (39.5%)	30 (19.4%)
AIAN	10 (11.5%)	2 (6.7%)	2 (5.3%)	14 (9.0%)
Asian	10 (11.5%)	1 (3.3%)	1 (2.6%)	12 (7.7%)
Birace/Multiple	10 (11.5%)	1 (3.3%)	0	12 (7.7%)
Other	21 (24.2%)	6 (20.0%)	6 (15.8%)	33 (21.3%)
Hispanic Origin				
Hispanics	27 (31.0%)	8 (26.7%)	12 (31.6%)	47 (30.3%)
Non-Hispanics	60 (69.0%)	22 (73.3%)	26 (68.4%)	108 (69.7%)

4.3 Measures and Assessment

Based on the framework presented in Table 1, five data indicators are measured. They include:

(1) Relevancy – whether the data matches the data need for the Decennial Census;

(2) Timeliness –whether the data can be provided „point-in-time" with minimal time lag;

(3) Accuracy of the records, which consists of:

(a) coverage (whether there is any evidence of coverage error) and

(b) item response (whether information for each inmates are complete on the ARs); and

(4) Reliability – consistency between ARs data and self-responded survey data.

4.3.1 Processing Data from Administrative Records

The data format of the collected records varies, but all followed the Office of Management Budget"s (OMB) guideline (OMB 1997). The OMB standard provides two formats that may be used for data on race and ethnicity. Self-reporting or self-identification using two separate questions is the preferred method for collecting data on race and ethnicity[6]. In situations where self-reporting is not practicable or feasible, the combined format may be used[7]. The census form followed the first standard while the prison records utilized the second format, mostly due to situations where inmates" race/ethnicity were not self-reported. Moreover, the records collected from state 2 allow detailed reporting on the race/ethnicity variable that mirror those of the race question on the census form while the records from state 1 only have five categories instead of six (it did not follow the OMB combined format guideline and did not have the Native Hawaiian or other Pacific Islander category). Due to the difference in formats, both the survey and administrative record data were re-coded into the combined format for ease of comparison.

5. Results

5.1 Relevancy and Timeliness of Records

Our first research question "Are the data provided by the state prisons" ARs relevant and met the census data needs?" was addressed by using ethnographic data. The staff at the observed facilities provided a copy of the prison records printed on the same day for the census enumerators containing the full name, age or date of birth, race/ethnicity, and sex of inmates. A different staff printed the alternate rosters requested by the research team on Census Day (April 1st, 2010) containing the same type of information but in a different format. These records were relevant for the data needs of the census.

5.2 Timeliness of Records

The ethnographic observation data indicated that given the custodial nature of correctional facilities, detailed and timely records of inmates were kept. These data addressed our second research question: Are the state prisons" ARs data timely and up-to-date? The ARs were updated in a timely manner with minimal lag time and the records can be printed in real time. These records were accessible mostly in printed forms. One of the two states in this study can provide electronic format but delivered the records in paper format as per census takers" request. They met the purpose of data collection and statistical goal demanded by the mandated Decennial Census. These records were printed in a few hours. It would have taken enormous effort and considerable time from the facility to have the inmates complete the ICR form. Using ARs to enumerate prison population minimized burden on the prison staff and caused no respondent burden on the inmates.

5.3 Accuracy

In this section, the third research question "Are the state prisons" ARs data accurate in terms of coverage and item responses?" is addressed. The assessment includes (1) information from the ethnographic data; (2) validating the completeness of the ARs printed for census takers versus those printed for the researchers and the final census population count; (3) exploring contextual factors affecting coverage accuracy; and (4) examining item non-response.

5.3.1 Within GQ Coverage Error

Ethnographic Observation Data: State Prisons" Administrative Records and Daily Count

ARs referred to as „rosters" were kept and updated frequently by all observed facilities in this study. The staff conducted „daily census" at least four times a day. During each of the daily counts, the prison is „locked down" meaning everyone on the roster has to be accounted for before any activities can resume. During an observation for the American Community Survey with a state prison in the Midwest, Chan (2010) observed that all inmates, staff, and visitors had to „stay put" during a routine count when the staff tried to reconcile a discrepancy between the actual head count and the known population count for that day. Hence, in terms of counts/coverage, the ethnographic results suggested that the coverage of the ARs should be accurate if a comprehensive listing was obtained.

Validating ARs with and Alternative Rosters

The key information required for the Decennial Census count is the number of inmates that reside at the facility on Census day. The two female prisons in the study provided two different sets of ARs. One set was printed for the census enumerator and the other was printed for the ethnographer by different staff. For Women Prison 1, the two different reported counts matched perfectly. In Women Prison 2, however, a discrepancy of 679 inmates was discovered between the two ARs. The omitted 679 were inmates residing in the reception center waiting to be classified. They were omitted on the administrative record printed for the census enumerators on April1st, 2010. A careful comparison showed that these omitted inmates were all in the reception center waiting to be classified and assigned to different buildings. During unstructured interviews, it was clear that the records for population in the reception center were kept in a separate file. Hence, in order to obtain a complete listing, records have to be printed from two separate computer files. The author verified with the 2010 census operation that the final population count recorded for that prison was in fact closer to the one provided by the ethnographer. There were extra 15 inmates in the final count than the „correct" one printed for the ethnographer. The additional counts were for inmates staying at

the out-patient housing unit on Census Day. We suspect that the quality control process of the regional office noted the much lower count when checked against the initial daily count provided by the facility a month before the actual census. However, it is also possible that the ethnographer was given an incomplete listing. The validation suggests that the ARs is rather complete except for inmates that are being moved or in transit.

Contextual Factors Relating to Records Coverage Error: Movements and Record Keeping Systems

As noted above, inmate movement, particularly during transit, is a key factor that can affect the coverage of the state prison records. Based on ethnographic observation and interview data, inmates are separated into several major populations: general population, unclassified population in the reception area, and inmates in administrative segregation. Inmates are often transferred to other prisons (sometimes across the state) for disciplinary or gang-related reasons, or transferred to minimum-security prisons as their release date draws closer. Sometimes they are moved to county jails for court hearings, taken to hospitals for illness or operations that could not be handled within the prison or a hospice, or they may be in transit to the prison after their sentences. There are also inmate movements within the prisons between cells and workplaces, dining halls, recreation yards, classrooms, activity areas, hobby shops, and visiting rooms. Some inmates are moved to Disciplinary Segregation and the Intensive Management Unit due to disciplinary action. These units are considered „jail" inside the prisons.

5.3.2 Item Nonresponse

The information provided for each inmate listed on the ARs of the facilities was complete. There were no item non-response for any of the four variables that were assessed (names, date of birth (age), gender and race/ethnicity). The lack of missing items suggests that ARs provide complete data for the purpose of census if every record is collected.

5.4 Reliability: Information Agreement

The final research question "Are the state prisons" ARs data consistent with other sources of data?" is addressed in this section. The four major variables examined in this study included: names, date of birth, race and Hispanic origin. The names and date of birth information are presented in section 5.4.1. The agreement rates between the ARs information and the survey data on race and Hispanic origin are discussed in details in section 5.4.2.

5.4.1. Names and Date of Birth

Although there were no missing data from the name field and the date of birth information listed on the records, the comparison shows that there were a number of small discrepancies. There were nine discrepancies (5.8%) in the name variable. Some had spelling errors, some used initials as their first names, and some used their middle name as their first name. These discrepancies are expected. Four inmates provided discrepant dates of birth but three of these discrepancies varied less than a year. The remaining one was an error by an inmate who wrote the interview year instead of his birth year. There were three other discrepancies when comparing inmates self-reported age and the record. In all three incidents, the inmates reported an incorrect age.

5.4.2 The Measurement of Race and Ethnicity

All obtained ARs used a combined race and Hispanic original format to record „ethnicity". The combined racial/ethnic categories used by each facility varied and some of the categories are not mutually exclusive. For instance, the ARs from one of the prisons has „Hispanic" as a ethnicity category, but detailed Hispanic origins such as „Mexican" or „Cuban" are also available as racial/ethnic categories.

The ICR-like form collects a person"s race information and Hispanic origin separately. The prison records used the combined format. Despite the OMB guidelines for allowing multiple race responses for combined race format, the prison records did not contain any multiple responses to the race/ethnicity question. Twelve inmates, however, (7.7%) reported more than one race on the survey and all but one of them was a respondent from male prison 1.

5.4.2.1 Hispanic Origin

Table 4 summarizes the number of agreements between the information on Hispanic origin provided by the ARs and the self-reported survey. Given the ARs provided only one racial/ethnic category for inmates, those inmates who were listed as Hispanics did not have a designation for a racial category. Nine of the 47 respondents (19%) who identified themselves as Hispanics were not listed as Hispanics on the administrative record. Hence, it is likely that the proportion of inmates who are of Hispanic descent would be underestimated if records are used to generate

statistics for state prison inmates. In order to adjust for the potential bias, a probability sample survey of coverage measurement is needed to adjust for such estimates.

Table 4. Agreement Between Administrative Data and Self-Report Data on Hispanic Origin

	Self-Reported Hispanic Origin	
Administrative Records Data Combined Race/Ethnicity Format	Non-Hispanics (N=108, 69.7%)	Hispanics (N=47, 30.3%)
Non-Hispanics (N=116, 74.8%)	**98.1% (n=106)**	21.3% (n=10)
Hispanics (N=39, 25.2%)	1.9% (n=2)	**78.7% (n=37)**
	100%	100%

5.4.2.2 Racial/Ethnic Background

Table 5 shows the agreement rate (in bolded numbers) between the self-response and the ARs data for race. Overall, 33 of the 155 inmates (21.3%) provided a self-identified race(s) that did not match the ARs. The agreement rate was calculated for each race/ethnicity group based on the administrative data. The agreement rates range from 71.8% to 90.9%. The agreement rate was highest among Asian inmates and 90.9% of inmates listed as Asian also reported that they were Asian only. The rates were relatively high among white (84.6%) and black (82.4%) inmates. The agreement rate between the records and the survey for inmates listed as American Indian, Alaskan Native or Hispanics on the ARs is lower than the other racial/ethnic groups.

Table 5. Percentage of Agreement between Self-Reported and Administrative Data on Race/Ethnicity

	Self-Reported Race (N=155)					
Administrative Records Data - Combined Race/Ethnicity Format	White (N=54, 34.8%)	African American (N=30, 19.4%)	American Indian & Alaskan Native (N=14, 9.0%)	Asian (N=12, 7.7%)	*Blank/Other (N=33, 21.3%)	One or More Race (N=12, 7.7%)
White (N=52, 33.5%)	**81.5% (n=44)**	-	14.3% (n=2)	-	12.1% (n=4)	16.7% (n=2)
African American (N=34, 21.9%)	1.9% (n=1)	**93.3% (n=28)**	7.1% (n=1)	-	-	33.3% (n=4)
American Indian & Alaskan Native (N=15, 9.7%)	-	-	**78.6% (n=11)**	-		33.3% (n=4)
Asian (N=11, 7.1%)	-	-	-	**83.3% (n=10)**		8.3% (n=1)
Other (N=4, 2.6%)	-	6.7% (n=2)	-	8.3% (n=1)	**3.0% (n=1)**	-
Hispanic (N=39, 25.2%)	16.7% (n=9)	-	-	8.3% (n=1)	84.8%** (n=28)	8.3% (n=1)
	100%	100%	100%	100%	100%	100%

*16 were „blank“ and 17 were „other“; **16 were „blank“ and 12 were „other“

Discussion and Conclusions

Based on the assessments, the data provided on the state prison records were relevant to the data needs of the decennial census and were provided to the enumerators and researchers in a timely and efficient manner. Use of State Prison records can reduce operation cost for the Census Bureau, minimize respondent burden on the facilities and poses no burden on the inmates. Observation and in-depth interview data suggest that the two sets of ARs reviewed for this study has excellent coverage error. Important steps and procedures could be taken to help state prison facilities avoid providing incomplete lists, in particular, leaving those that are in transit or were moved to other special unit off the ARs. Information for each inmate listed on the ARs was complete - there were no missing data on names, date of birth or race/ethnicity for all counted inmates. However, there was no uniformity in the data format provided by prisons between states and within each state. While prisons records collected from one of the two states used only five racial/ethnic categories (one less than the OMB recommended six categories), the other state used more categories than the census forms. The combined race/ethnicity data provided by ARs are inconsistent with the inmates‟ self-identification. Using ARs to complete the ICR imply that there will be classification errors of race and Hispanic origin. Using the combined race and Hispanic origin variable kept by the state prisons‟ record system could lead to bias estimation of Hispanic in correctional facilities. There may also be under-representation of inmates who consider themselves more than one race.

The findings in this study imply that ARs can provide complete and rather accurate answers to the key questions on the ICR form if a complete set of ARs were obtained. The completeness of the state prison records suggest that state prisons can be excluded in the 2020 CCM program as long as steps are taken to ensure the delivered ARs include all inmates residing in the facilities. However, estimates for Hispanic origin and race may potentially be biased. A probability sample coverage measurement survey study will be needed to adjust for such estimates. Given this study uses probability sampling within the facility, the findings are only generalizable within the facility and not beyond the case studies. A future study should replicate the findings with a probability sample.

Future Research and Recommendation

Researchers who want to understand the quality of ARs should examine the sources of the records; in particular, at the time the organization collects or records such data. Future research on correctional facilities‟ records should pay special attention at the intake procedures for Departments of Corrections and Juvenile agencies. This is an important data collection point where ARs are finalized for individual inmates, and this stage determines the data quality of ARs in the correctional facilities. The knowledge on the data quality will identify the most cost efficient use of the ARs data. As ARs continue to be an important source of census data, it is important that statistical agencies at both state and federal level work together to design uniform categories to minimize data processing cost and error. As suggested from our results, the differences in data format from different prison record systems require substantive data editing efforts (Fellegi and Holt 1976; Herzog, Scheuren and Winkler 2007) prior to using the ARs for statistical purpose.

More research is needed to explore the ideal way to obtain the most complete records from state prisons. For instance, although the census mandates only basic information of all its residents, when using ARs from correctional facilities, it is important to consider collecting extra information to help determine the location of prisoners. For those that are residing in jails where the length of sentence may be short, their length of sentence expected date for release and parole will help analysts make sound judgment on duplicated person records. This knowledge will help identify prisoners who are reported in more than one GQ, or both GQ and housing units.

Endnotes

1. The disciplinary segregation unit (DSU) is described as the jail within the prison – the place to confine inmates who simply cannot get along and follow the rules when in general population.
2. The 2010 GQ Enumeration Operation did not have procedures for enumerators to collect administrative data records electronically. Privacy Protection was a key concern for electronic data.
3. Register-based data are data adapted and processed from already recorded information such as those collected and managed by agencies and businesses, e.g., administrative data on patients in hospitals, employees of an organization or drivers‟ at the Department of Motor Vehicles.

4. Participant Observation is one of the most common methods for qualitative data collection in many disciplines including anthropology, sociology and communication studies. It requires that the researcher become a participant in the culture or context being observed and aims to gain a close and intimate familiarity with a given group of individuals and their practices through an intensive involvement with people in their natural environment, usually over an extended period of time.
5. Prison and jails typically call the buildings where inmates live "housing units." This definition of the term "housing unit" is in conflict with how the Census Bureau uses this term to denote individual "free world" residential units. Prisons and jails are considered "group quarters".
6. The OMB Two-question format guideline

   To provide flexibility and ensure data quality, separate questions shall be used wherever feasible for reporting race and ethnicity. When race and ethnicity are collected separately, ethnicity shall be collected first. If race and ethnicity are collected separately, the minimum designations are:
   Race:
   -- American Indian or Alaska Native
   -- Asian
   -- Black or African American
   -- Native Hawaiian or Other Pacific Islander
   -- White
   Ethnicity:
   -- Hispanic or Latino
   -- Not Hispanic or Latino
   When data on race and ethnicity are collected separately, provision shall be made to report the number of respondents in each racial category who are Hispanic or Latino.
   When aggregate data are presented, data producers shall provide the number of respondents who marked (or selected) only one category, separately for each of the five racial categories. In addition to these numbers, data producers are strongly encouraged to provide the detailed distributions, including all possible combinations, of multiple responses to the race question. If data on multiple responses are collapsed, at a minimum the total number of respondents reporting "more than one race" shall be made available.
7. OMB Combined race/ethnicity format guideline

   The combined format may be used, if necessary, for observer-collected data on race and ethnicity. Both race (including multiple responses) and ethnicity shall be collected when appropriate and feasible, although the selection of one category in the combined format is acceptable. If a combined format is used, there are six minimum categories:
   -- American Indian or Alaska Native
   -- Asian
   -- Black or African American
   -- Hispanic or Latino
   -- Native Hawaiian or Other Pacific Islander
   -- White
   When aggregate data are presented, data producers shall provide the number of respondents who marked (or selected) only one category, separately for each of the six categories. In addition to these numbers, data producers are strongly encouraged to provide the detailed distributions, including all possible combinations, of multiple responses. In cases where data on multiple responses are collapsed, the total number of respondents reporting "Hispanic or Latino and one or more races" and the total number of respondents reporting "more than one race" (regardless of ethnicity) shall be provided.

Reference

Chan, Anna Y. 2009. "2010 Census Evaluations, Experiments and Assessments (CPEX) Study Plan: Investigation of Methods to Evaluate Coverage of the Group Quarters (GQs) Population. Unpublished Internal Study Plan. Center for Survey Measurement. U.S. Census Bureau.

Chan, Anna Y. 2012. Report on In-Depth Interviews with Group Quarters" Staff. Unpublished Internal Report. Center for Survey Measurement. U.S. Census Bureau.

Daas, Piet J.H., Judit Arends-Tóth, Barry Schouten, Léander Kujivenhoven. 2008. "Quality Framework for the Evaluation of Administrative Data." Proceedings Paper for the 2008 European Conferences on Quality in Official Statistics. Access on December 14, 2011. http://www.pietdaas.nl/beta/pubs/pubs/21Daas.pdf.

Fellegi, I.P. and D. Holt. "A Systematic Approach to Automatic Edit and Imputation." Journal of the American Statistical Association, 71 (353), 17-35.

Holmberg, Anders 2011 "Combing registers into a fully register-based census – some methodological issues. Paper presented at the U.S. Census Bureau Conference on Utilizing Administrative Data: Technical, Statistical and Research Issues. Washington, D.C. October 27 & 28, 2011.

Inderbitzen, Michelle and Anna Chan. 2010. "Ethnographic Study of Group Quarters in the 2010 Census: Prisons for Audlts and Juvenile Correctional Facilities." Internal Unpublished Report. Center for Survey Measurement. U.S. Census Bureau. Washington, D.C.

Jonas, Kimball. 2003. Group Quarters Enumeration. Final Report. Census 2000 Evaluation E.5, Revision 1. U.S. Census Bureau. Accessed on October 21, 2011. http://www.census.gov/pred/www/rpts/E.5%20R.pdf.

Mulalic, Lada 2011. Dutch Virtual Census. Paper presented at the U.S. Census Bureau Conference on Utilizing Administrative Data: Technical, Statistical and Research Issues. Washington, D.C. October 27 & 28, 2011.

Nordholt, Eric Schulte, Saskia J.L. Ossen and Piet J. H. Daas. 2011. "Research on the Quality of Registers to make Data Decisions in the Dutch Virtual Census." ISI 2001 Proceeding Paper.

Owen, Barbara and Anna Chan. 2010. "Using Ethnography to Investigate Coverage of Group Quarters" Population. Jails and Prisons." Internal Unpublished Report. Center for Survey Measurement. U.S. Census Bureau. Washington, D.C.

OMB 1997. "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity". *Federal Register* Notice (October 30, 1997). Accessed on December 12, 2011. http://www.whitehouse.gov/omb/fedreg_1997standards

Pipino, Leo L., Yang W. Lee and Richard Y. Yang. 2002. "Data Quality Assessment." *Communications of the ACS.* V 45: 211-218.

Thygesen, Lars 2011. Use of Administrative Sources for Censuses and Demographic and Social Statistics. Paper presented at the U.S. Census Bureau Conference on Utilizing Administrative Data: Technical, Statistical and Research Issues. Washington, D.C. October 27 & 28, 2011.

Trochim, W. (2000). The Research Methods Knowledge Base, 2nd Edition. Atomic Dog Publishing, Cincinnati, OH.

U.S. Census Bureau, 2009. 2010 Census Group Quarters Definitions and Code List.

U.S. Census Bureau. 2010. Group Quarters Population, Table B26001. 2010 American Community Survey – 1 year estimates. Assessed on December 14, 2011.

U.S. Census Bureau. 2011. Business Case for Using Administrative Records to Capture Decennial Data in Group Quarters. Internal document prepared by the Business Case Development Team.

Acknowledgement

The author wishes to acknowledge Professor Barbara Owen (University of California, Fresno) and Professor Michelle Inderbitzen (Oregon State University) for their contribution to this research effort; Magdalena Ramos, Dr. Yuling Pan, and Dr. Young Chun for reviewing earlier drafts of this paper, their comments have made this manuscript better; Sarah Wilson, Aiden Ettlinger, and Egan Jackson for editing the paper.