

Including the Group Quarters Population in the US Synthesized Population Database

Bernadette M. Chasteen, William D. Wheaton,
Philip C. Cooley, Laxminarayana Ganapathi, and
Diane K. Wagener

February 2011

RTI Press

About the Authors

Bernadette M. Chasteen, MA, is a senior GIS analyst in RTI International's Geospatial Science and Technology program.

William D. Wheaton, MA, is a senior research geographer and director of RTI International's Geospatial Science and Technology program.

Philip C. Cooley, MS, is an RTI Fellow in bioinformatics and high-performance computing at RTI International.

Laxminarayana Ganapathi, PhD, is a programmer/analyst at RTI International.

Diane K. Wagener, PhD, is a senior epidemiologist at RTI International.

RTI Press publication MR-0020-1102

This PDF document was made available from www.rti.org as a public service of RTI International. More information about RTI Press can be found at <http://www.rti.org/rtipress>.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. The RTI Press mission is to disseminate information about RTI research, analytic tools, and technical expertise to a national and international audience. RTI Press publications are peer-reviewed by at least two independent substantive experts and one or more Press editors.

Suggested Citation

Chasteen, B. M., Wheaton, W. D., Cooley, P. C., Ganapathi, L., and Wagener, D. K. (2011). Including the Group Quarters Population in the US Synthesized Population Database. RTI Press publication No. MR-0020-1102. Research Triangle Park, NC: RTI Press. Retrieved [date] from <http://www.rti.org/rtipress>.

This publication is part of the RTI Press Methods Report series.

RTI International
3040 Cornwallis Road
PO Box 12194
Research Triangle Park, NC
27709-2194 USA

Tel: +1.919.541.6000
Fax: +1.919.541.5985
E-mail: rtipress@rti.org
Web site: www.rti.org

©2011 Research Triangle Institute. RTI International is a trade name of Research Triangle Institute.

All rights reserved. Please note that this document is copyrighted and credit must be provided to the authors and source of the document when you quote from it. You must not sell the document or make a profit from reproducing it.

[doi:10.3768/rtipress.2011.mr.0020.1102](https://doi.org/10.3768/rtipress.2011.mr.0020.1102)

www.rti.org/rtipress

Including the Group Quarters Population in the US Synthesized Population Database

Bernadette M. Chasteen, William D. Wheaton, Philip C. Cooley, Laxminarayana Ganapathi, and Diane K. Wagener

Abstract

In 2005, RTI International researchers developed methods to generate synthesized population data on US households for the US Synthesized Population Database. These data are used in agent-based modeling, which simulates large-scale social networks to test how changes in the behaviors of individuals affect the overall network. Group quarters are residences where individuals live in close proximity and interact frequently. Although the Synthesized Population Database represents the population living in households, data for the nation's group quarters residents are not easily quantified because of US Census Bureau reporting methods designed to protect individuals' privacy.

Including group quarters population data can be an important factor in agent-based modeling because the number of residents and the frequency of their interactions are variables that directly affect modeling results. Particularly with infectious disease modeling, the increased frequency of agent interaction may increase the probability of infectious disease transmission between individuals and the probability of disease outbreaks.

This report reviews our methods to synthesize data on group quarters residents to match US Census Bureau data. Our goal in developing the Group Quarters Population Database was to enable its use with RTI's US Synthesized Population Database in the Modeling of Infectious Diseases Agent Study.

Contents

Introduction	2
Methods	2
Creating Group Quarters Facility Records	4
Creating Group Quarters Person Records	7
Results	16
Table Structure of the Group Quarters Facility Records	16
Table Structure of the Group Quarters Person Records	18
Conclusions	20
References	20
Acknowledgments	Inside back cover

Introduction

Agent-based models (ABMs) are computational models that simulate the actions of autonomous “agents” (i.e., model representations of individuals) in a social network and the effect that agent behaviors and interactions have on other agents and the network as a whole. Researchers use ABMs to model important population characteristics and structures. One specific application of ABMs is the modeling conducted for the Models of Infectious Disease Agents Study (MIDAS; www.midasmodels.org), which requires data on agents (e.g., people, animals) and how agents come into close proximity with each other.

The impetus to develop synthesized population databases came from a need for detailed, geospatial data for use in agent-based modeling. In an earlier publication, researchers at RTI International described the development of the US Synthesized Population Database, which provides synthesized population data on US households and household residents, data based on 2000 Census data (Wheaton et al., 2007; Wheaton et al., 2009). However, because the US Census Bureau reports population data in a manner that protects an individual’s privacy (US Census Bureau, 2000a), detailed Census data about the nation’s group quarters population were not available for use in the US Synthesized Population Database.

In 2000, about 3 percent of the total US population lived in group quarters facilities, such as military bases, prisons and other correctional institutions, college dormitories, and nursing homes (US Census Bureau, 2000b). Including data on group quarters residents in a synthesized population database can be important to agent-based modeling because the number of these residents and the frequency of their interactions are variables that directly affect modeling results. Inclusion is important to infectious disease modeling, in particular, because the increased frequency of agent interaction in a group quarters environment may increase the probability of infectious disease transmission between individuals and the probability of disease outbreaks.

Because MIDAS researchers have indicated a need for detailed group quarters population data, we developed methods to synthesize group quarters data and include these data in a Group Quarters Population Database. Development of the Group Quarters Population Database will allow MIDAS researchers to run agent-based modeling under various scenarios—including, for example, scenarios in which

- group quarters residents interact with the general population,
- group quarters residents are quarantined to prevent infection from the general population, and
- group quarters residents are quarantined to prevent reinfection of the general population after an epidemic.

This report describes the methods we developed to synthesize, for use in MIDAS modeling, data on the nation’s group quarters population that matched overall US Census Bureau group quarters population data.

Methods

Our goal was to generate a Group Quarters Population Database that contained records resembling the data structure of RTI’s US Synthesized Population Database (Wheaton et al., 2007; Wheaton et al., 2009). The two databases could then be used together with minimal difficulty for agent-based modeling.

The US Census Bureau segments the group quarters population into the following five categories:

- military bases,
- federal, state, and local prisons and correctional facilities,
- university and college dormitories (“college dormitories” hereafter),
- nursing homes, and
- other institutions.

For each of these categories, the Census data provide information on the population count of group quarters residents per Census Block Group (CBG) and broad-scale age ranges for these residents.

To develop the US Synthesized Population Database, previous RTI researchers used the US Census Bureau's 2000 Public Use Microdata Samples (PUMS) file (US Census Bureau, 2000a), which contains a 5 percent sample of the housing units counted in the 2000 Census. Although the PUMS file contains 391,377 housing units coded as group quarters facilities, the file does not provide data on the specific location of these facilities or the number of facilities per CBG. Nor does it provide detailed information on the gender and age of group quarters residents. Consequently, we could not use the PUMS file to synthesize data for the Group Quarters Population Database and had to obtain this information from other sources.

To synthesize data for the five Census group quarters categories, we had to locate the group quarters facilities more specifically than the Census representation permitted (i.e., by CBG). Moreover, because ABMs require data on the specific gender and age of group quarters residents, we had to obtain data from outside data sources (i.e., non-Census data) that would allow us to synthesize a group quarters population reflective of the underlying population.

To model agent interactions accurately and to mimic the structure of the US Synthesized Population Database, the synthesized Group Quarters Population Database required the following two types of data records:

- Group Quarters Facility Records, which contained
 - the category of group quarters facility,
 - the geospatial location of the facility,
 - the total number of residents per facility, and
 - a unique facility identifier (ID)
- Group Quarters Person Records, which contained
 - the age of group quarters residents,
 - the gender of group quarters residents,
 - a unique ID for each person record, and
 - a group quarters facility ID (GQ_ID) to link the person record to the person's specific facility.

The data used to generate the synthesized data records for the Group Quarters Population Database were expected to satisfy the same criteria as the data used to generate the synthesized records of the US Synthesized Population Database:

- Data had to be scalable to reflect the nationwide population.
- Data had to be presented at a sufficiently fine level of geographic resolution to model individual interactions that affect disease spread.
- Data had to meaningfully inform the agents' interactions.

The last criterion meant that interactions of college dorm residents, for example, would differ from the interactions of nursing home residents.

Unfortunately, we could find no single data source that met all these criteria. Consequently, to develop a synthesized Group Quarters Population Database that contained sufficiently detailed group quarters facility and person records, we reviewed the following national data sources for information on facility locations and the demographic properties of group quarters residents:

- **The 2000 Census of Population and Housing Summary File 1 (SF1).** We used the SF1 (US Census Bureau, 2000b) to obtain demographic information (both facility location and person counts) for the group quarters categories. The SF1 provides data on the number of residents in some types of group quarters categories by age ranges and gender, as segmented by CBG. By using the SF1 data with other sources that provided the facility locations, we were able to generate category-specific group quarters "points" within each CBG. A record for each resident (synthesized person with specific age and gender) was then linked to its associated facility point.

- **Homeland Security Infrastructure Program (HSIP) Gold 2005 Database.** The HSIP database (National Geospatial-Intelligence Agency, 2007) is a US Department of Homeland Security geospatial data product that contains the locations of prisons, military facilities, colleges and universities, and nursing homes.¹
- **Emergency Preparedness Atlas: US Nursing Home and Hospital Facilities.** The nursing home atlas contains information on nursing home facility locations and on bed capacity per facility (Agency for Healthcare Research and Quality, 2007).
- **Group Quarters Age Distribution Data.** To refine the age data for the prison population, we required additional data on the age distributions of group quarters populations for the year 2000. These additional data allowed us to generate person records with associated age and gender information. We used several sources to obtain these data, including sources from the US Department of Justice, Bureau of Justice Statistics (BJS), or data on prison populations and the US Department of Defense (DoD) for data on military populations.

Creating Group Quarters Facility Records

As stated previously, the US Census Bureau SF1 data provide general information, such as population counts, on the group quarters categories by CBG, but these data do not include information on group quarters facility locations or the number of residents per facility. Consequently, we could not use the SF1 data alone to generate data on group quarters facility locations. To address this issue, we identified the HSIP database and the nursing home atlas as additional data sources that provided the most documented and comprehensive data representing locations for the group quarters categories.

¹ HSIP data on group quarters facility locations are often derived from address geocoding. Address geocoding defaults to a zip code centroid when a street address fails to match to any known street. For this reason the spatial location can be inaccurate. This situation occurs because of an error, such as a misspelling, in the address supplied for a facility, or an error or omission in the underlying street data. At times, we encountered mismatches between the SF1 population counts, based on CBGs, and the HSIP database facility location data, which are based, in part, on zip codes.

Table 1 lists the four group quarters categories included in the Group Quarters Population Database as well as any additional data sources used to locate each category's group quarters facilities and to assign resident counts to each facility. We did not generate group quarters facility records for the "other institutions" category because this broad category includes many different kinds of group quarters facilities and social interactions between residents; in effect, the category was not useful for MIDAS.

Table 1. Secondary data sources used to determine group quarters facility locations for the RTI Group Quarters Population Database

Group Quarters Category	Data Source
Military bases	Homeland Security Infrastructure Program Gold 2005 Database (National Geospatial-Intelligence Agency, 2007)
Prisons and correctional facilities	Homeland Security Infrastructure Program Gold 2005 Database (National Geospatial-Intelligence Agency, 2007)
College dormitories	Homeland Security Infrastructure Program Gold 2005 Database (National Geospatial-Intelligence Agency, 2007)
Nursing homes	Homeland Security Infrastructure Program Gold 2005 Database (National Geospatial-Intelligence Agency, 2007) Emergency Preparedness Atlas: US Nursing Home and Hospital Facilities (Agency for Healthcare Research and Quality, 2007)

Note: For security, the points for military bases were not located in their actual position but were located at the center (centroid) of the Census Block Group in which they occurred.

When a data source enabled us to determine a facility location for a specific group quarters category, we used the SF1 data to determine the corresponding CBG that included that location. We then distributed the SF1 group quarters population to specific, geo-referenced group quarters points. As a result, each group quarters point represents one facility for one group quarters category. We repeated this process for each category.

When a data source provided data on a specific facility location, we used that source point to generate the facility record in the Group Quarters Population Database. When a data source did not provide data

on a specific facility location for a category (as with the military bases), and when we could demonstrate that the category rarely had more than one facility location within a CBG, we generated a single group quarters point at the CBG centroid. We then assigned the entire category-specific SF1 population count for that CBG to the point. When a data source indicated that a category had more than one facility location in a CBG, the location and SF1 population count assigned to the facility were distributed according to the category-specific rules delineated in the following subsections.

Military Bases

We used SF1 population data and HSIP location-point data to synthesize group quarters facility records for military bases. We compared the 2000 SF1 population count data with the HSIP military base data layer and found that all military populations enumerated in the SF1 data were located in or near military bases identified in the HSIP database.

We also compared the HSIP data with the SF1 data for several states and found that no CBGs contained more than one military base; therefore, we associated the entire SF1 population count for military personnel with a single point for each CBG. For national security reasons, rather than indicate the actual location of military bases, we placed the points for these group quarters facilities at the CBG centroid. Although the SF1 population data show that some CBGs had fewer than 50 military personnel, all of these CBGs are located near larger bases or US Coast Guard stations; consequently, we assigned the population counts for these CBGs to the facilities indicated by the HSIP data.

Prisons and Correctional Facilities

We used SF1 population data and HSIP location-point data to synthesize group quarters facility records for prisons and correctional facilities. A comparison of SF1 data with HSIP data for federal, state, and local prisons showed that many CBGs contained more than one prison or correctional facility.

Because, typically, inmate populations reside in prisons or correctional facilities separated by gender and age (e.g., adult and juvenile), we first identified four gender-age groups for prison populations:

- men 18 years of age or younger,
- men older than 18 years,
- women 18 years of age or younger, and
- women older than 18 years.

Although, in many cases, inmates of different groups are housed under one roof, these groups are usually separated within the facility; therefore, we decided that for modeling disease transmission scenarios these four gender-age groups would suffice for representing real-world agent interaction.

If a CBG contained only one prison or correctional facility and the SF1 data for the CBG represented one gender-age group, we assigned the entire SF1 population count to that facility location point. If a CBG contained more than one correctional facility or the SF1 data were on more than one inmate group, we distributed the population count to facilities in accordance with the following decision rules:

- We used the SF1 count of inmates by group to determine how many facilities should exist in each CBG. For example, if the SF1 data included men 18 years of age or younger and women older than 18 years (but no men older than 18 years, and no women 18 years of age or younger), then we would expect to find at least two facilities in that CBG, one for each gender-age group.
- On the basis of the HSIP location-point data on the number of prisons or correctional facilities in a CBG and the SF1 data on the gender-age groups, we divided the prison population data among the facilities in accordance with the rules in Table 2. Because of security and confidentiality restrictions, in some cases we could not identify specific facility locations by gender-age group; when this situation occurred, we randomly assigned a facility to locations within a CBG.

Table 2. Rules for allocating population data among prisons and correctional facilities within a Census Block Group

Number of Prison Groups Compared with Number of Facilities	Assignment Rules
More prison groups than facilities	Generate additional points located at the CBG centroid as needed to assign population from each population group to its own facility.
Same number of prison groups and facilities	Assign total population of a specific gender-age group to one randomly selected facility.
One prison group and two facilities	Divide population evenly between the two facilities (points).
Fewer prison groups than facilities	<ol style="list-style-type: none"> 1. If fewer than 10 people in any category, assign all 10 to one facility. 2. If there are still more facilities than groups, take the largest population group and split its population between two facilities. If this split population is larger than the largest facility population for the state, split the population between three or four facilities until the population is less than the largest facility population for the state. 3. If number of groups = remaining number of facilities, assign groups to remaining facilities. 4. If there are still more facilities than groups, return to step 2 and proceed until all facilities have population assigned to them.

CBG = Census Block Group.

Note: For example, a CBG that contains six facilities and all four population groups—with 263 men aged 18 or older, 94 men under age 18, 178 women aged 18 or older, and 31 women under age 18—would be distributed as follows:

Population of Group	Explanation	Number	Facility
Men ≥18: 263	Largest population split between facilities 1 and 2	132 Men ≥18	1
		131 Men ≥18	2
Women ≥18: 178	Next largest population split between facilities 3 and 4	89 Women ≥18	3
		89 Women ≥18	4
Men <18: 94 Women <18: 31	Last two populations assigned to last two facilities	94 Men <18	5
		31 Women <18	6

College Dormitories

We used SF1 data and HSIP location-point data to develop group quarters facility records for college dormitories. In CBGs where the HSIP data had a point for a college, but the SF1 population data had a zero population count for college dormitory residents, we eliminated the HSIP data point. We did so because we assumed either that the data point represented a nonresidential college, or that the location was wrong because of geocoding inaccuracies. In these cases, we did not create a point for the dormitory in the Group Quarters Population Database.

Using the HSIP location-point data, we obtained 1,544 dormitory locations, or points, from the HSIP database. Where the SF1 data showed group quarters residents but the HSIP database did not contain location-point data, we assumed that a college dormitory existed. We then generated a point at

the CBG centroid. Because the data sources never indicated more than one residential college within a CBG, we decided that placing only one college dormitory point in each CBG was appropriate; we placed 2,442 points in this manner. Although we recognized that many colleges have multiple dormitory buildings where students are often separated by gender, we had no way to determine an average dormitory building size; therefore, we placed the entire student dormitory population at a single point in the CBG.

Nursing Homes

We used data from the HSIP database and the nursing home atlas to determine nursing home locations and to allocate SF1 data on nursing home residents to nursing home facility locations. The HSIP data on nursing homes included a larger, more geographically

complete data set than the atlas; however, the HSIP data do not include information on nursing home bed capacity (i.e., facility size). The nursing home atlas provided data on facility locations, together with information on nursing home bed capacity; however, the nursing home atlas data included only facilities receiving Medicaid, not all facilities, and the difference between facility types was significant.

In addition, discrepancies exist between the nursing home atlas data and the SF1 data. The nursing home atlas reports the total number of nursing home beds in 2005 for the United States as 221,630, which is only 12.9 percent of the total nursing home population reported in the SF1 data (1,720,500). The nursing home atlas data also listed 14,955 nursing home facilities for the United States, whereas the SF1 data report nursing home residents in 18,867 separate CBGs.

Because of these discrepancies, we determined that the nursing home atlas data were inadequate for reliably determining nursing home facility locations. Instead, we used the HSIP location-point data to establish facility locations only; we used the nursing home atlas data to calculate minimum, average, and maximum nursing home sizes (i.e., bed capacity) by state. We also used the atlas data to allocate the nursing home resident population when the SF1 data showed a CBG with nursing home residents but the HSIP database did not contain data on facility locations.

When the SF1 nursing home resident count for a CBG was less than the bed capacity for the largest nursing home facility for a state, we created one point for the SF1 population at the CBG centroid. When the SF1 resident count for a CBG was greater than the bed capacity for the largest nursing home, we

- divided the SF1 population count for the CBG by the average nursing home size for the state to determine the appropriate number of facilities that should exist in the CBG,
- generated a point for each facility, and
- placed those points at the CBG centroid.

Then we divided the SF1 count for the CBG by the number of nursing home facilities and placed any remaining residents at the last facility generated

within the CBG. When the SF1 population count for a CBG exceeded the number of residents in the state's largest nursing home but the HSIP location-point data showed only one facility in the CBG, we generated a second facility in order to allocate the population equally between the two facilities.

Half of the US nursing home facilities listed in the HSIP database (18,043 of 36,641) fell into CBGs for which the SF1 data did not indicate nursing home residents. To check whether the HSIP points were placed in the wrong CBG, we selected all CBGs that were adjacent to each CBG having an HSIP point but found that the adjacent CBGs also failed to correspond to the SF1 population count. In response, we decided that the SF1 data were more likely to be correct; therefore, when the SF1 reported no nursing home residents for a CBG, we deleted the nursing home facility data from the HSIP data set.

Creating Group Quarters Person Records

For the purpose of this report, “person(s)” refers to group quarters resident(s). The SF1 data provided information on the population count and gender of group quarters residents, as segmented by three broadly defined age ranges:

- younger than 18 years,
- 18 to 64 years, and
- older than 64 years.

However, ABMs require specific age values for each person. To refine the age data, we reviewed national data sources for the group quarters categories. We also identified additional national data sources that provided data for the age-range values for the year 2000.

Table 3 lists the data sources used to identify specific age values for each of the group quarters categories. Each of these sources was nationwide in scope.

We obtained the overall count of group quarters residents in the SF1 age ranges per CBG for each category (e.g., the number of men in the military population younger than 18 years old). We then proportionally distributed the SF1 population count to specific ages within the age group, using specific age distribution data. This process allowed

Table 3. Data sources used for group quarters age distributions

Group Quarters Category	Data Source
Military personnel	Selected Manpower Statistics for Fiscal Year 2005: Defense Manpower Data Center Report (US Department of Defense, 2005)
Prisons and correctional inmates	Prison and Jail Inmates at Midyear 2000 (Beck & Karberg, 2001)
College dormitory residents	2000 Census of Population and Housing, Summary File 1 (US Census Bureau, 2000b); 2000 Census of Population and Housing, Public Use Microdata Sample 2000 (US Census Bureau, 2000a)
Nursing home residents	1999 National Nursing Home Survey (Centers for Disease Control and Prevention, 1999)

us to generate group quarters person records with specific age and gender information. The following subsections describe, for each category, the data generation methods and data sources used to refine the age distributions.

Military Personnel

We used the national age distribution for military personnel, published by the DoD Defense Manpower Data Center, to derive the age distribution used for generating synthesized military person records (US Department of Defense, 2005). The Defense Manpower Data Center presents population count data for military personnel for specific ages and, in some instances, for specific age ranges (e.g., ages 40–44, 45–49).

When population count data were presented for an age range, we distributed the population count for the age range equally among the component ages of that range. Table 4 shows, as an example, the age distribution for military personnel in 2000 and 2001; we generated enough individual military agents to equal the counts provided by the SF1 data. Then we assigned age and gender values to each agent, in accordance with the ratios shown in Table 4.

Table 4. National percentage distribution of military personnel, by age and gender, for 2005

Age	Males (percent)	Females (percent)
17	0.25	0.48
18	3.42	3.88
19	5.90	7.28
20	7.02	8.25
21	6.94	8.25
22	6.25	7.77
23	5.56	6.31
24	4.79	5.83
25	4.46	4.85
26	4.11	4.37
27	3.85	3.88
28	3.60	3.40
29	3.60	2.91
30	3.34	2.91
31	3.00	2.91
32	3.00	2.43
33	2.91	2.43
34	3.00	2.43
35	3.17	2.43
36	3.25	2.43
37	3.25	2.43
38	2.91	1.94
39	2.48	1.94
40	1.37	1.07
41	1.37	1.07
42	1.37	1.07
43	1.37	1.07
44	1.37	1.07
45	0.46	0.39
46	0.46	0.39
47	0.46	0.39
48	0.46	0.39
49	0.46	0.39
50	0.05	0.06
51	0.05	0.06
52	0.05	0.06
53	0.05	0.06
54	0.05	0.06
55	0.05	0.06
56	0.05	0.06
57	0.05	0.06

(continued)

Table 4. National percentage distribution of military personnel, by age and gender, for 2005 (continued)

Age	Males (percent)	Females (percent)
58	0.05	0.06
59	0.05	0.06
60	0.05	0.06
61	0.05	0.06
62	0.05	0.06
63	0.05	0.06
64	0.05	0.06
65	0.05	0.06
Total	100.00	100.00

Note: Because of rounding, totals do not sum to 100%.

Source: US Department of Defense, 2005. Selected manpower statistics: fiscal year 2005.

Prison or Correctional Facility Inmates

The US Department of Justice, Bureau of Justice Statistics (BJS) provides population count data on the nation's prisoners by age and gender; the data are segmented into two groups—adults and juveniles (Beck & Karberg, 2001). Table 5 shows the BJS data for adult prisoners in 2000.

Table 5. Bureau of Justice Statistics national counts and distribution of adult prisoners, by age and gender, for 2000

Age Group (Years)	Males (N)	Percent of Age Group Who Were Male	Females (N)	Percent of Age Group Who Were Female
18–19	81,300	95.4	3,900	4.6
20–24	310,100	94.0	19,600	6.0
25–29	329,900	91.7	30,000	8.3
30–34	334,000	89.5	39,100	10.5
35–39	294,100	90.5	30,700	9.5
40–44	198,300	92.1	17,000	7.9
45–54	164,500	93.2	12,100	6.8
55 or older	51,300	95.0	2,700	5.0
Total	1,763,500		155,100	

Source: Adapted from Beck and Karberg (2001). Prison and jail inmates at midyear 2000 (NCJ 185989), US Bureau of Justice Statistics (BJS).

As shown in Table 5, BJS provided data for eight age ranges. To generate population counts for each component age in an age range, we distributed the counts for age ranges (e.g., 18- to 19-year-olds) equally among the component ages of the range.

Adjusted Count of Prisoners in the Age Range 55 to 64 Years

The BJS data included a count of prisoners for the age group 55 years or older. Because we needed the BJS distribution of the prison population to be comparable with the SF1 age range of 18 to 64, we calculated the population count of prisoners (by gender) as follows:

$$\begin{aligned} & \text{BJS prison population older than 54 years} \\ & - \text{SF1 prison population older than 64 years} \\ & = \text{prison population 55 to 64 years old.} \end{aligned}$$

Using this formula, we calculated the number of male prisoners aged 55 to 64 years as follows: $51,300 - 14,283 = 37,017$; and we calculated the number of female prisoners as follows: $2,700 - 1,599 = 1,101$. For the adjusted national counts by age and gender and the percentage distribution of adult prisoners 18 to 64 years of age, see Table 6.

Table 6. Adjusted national counts and distribution of adult prisoners, by age and gender

Age Group (Years)	Males (N)	Percent of Total Males in Age Group	Females (N)	Percent of Total Females in Age Group
18–19	81,300	4.6	3,900	2.5
20–24	310,100	17.6	19,600	12.6
25–29	329,900	18.7	30,000	19.3
30–34	334,000	18.9	39,100	25.2
35–39	294,100	16.7	30,700	19.8
40–44	198,300	11.2	17,000	11.0
45–54	164,500	9.3	12,100	7.8
55–64	37,017	2.1	1,101	0.7
65 or older	14,283	0.8	1,599	1.0
Total	1,763,500	100.0	155,100	100.0

Source: Adjusted from Beck and Karberg (2001). Prison and jail inmates at midyear 2000 (NCJ 185989), US Bureau of Justice Statistics (BJS); and US Census Bureau (2000b). 2000 Census of Population and Housing, Summary File 1 (SF1).

Refining Age Distribution for Prisoners 65 Years Old or Older

For the Group Quarters Population Database, we wished to assign a specific age to each agent; however, we were concerned that a flat distribution for this category of older prisoners (65 years of age or older) would not reflect the actual prison population. Therefore, to more accurately approximate the real population count than a flat distribution would, we derived the prison population aged 65 or older to resemble the age distribution of males in the US general population.

The SF1 population data for prisoners older than 64 are grouped into a single broad age category. We refined the age distribution of prisoners 65 years of age or older using the SF1 age distribution data for the national general population aged 64 years or older (US Census Bureau, 2000b). These data are summarized in Table 7.

Table 7. Derived counts and distribution of the US nationwide population over the age of 64, by age and gender

Age Group (Years)	Males (N)	Percent of Age Group Who Were Male	Females (N)	Percent of Age Group Who Were Female
65–69	4,400,362	46.2	5,133,183	53.8
70–74	3,902,912	44.1	4,954,529	55.9
75–79	3,044,456	41.0	4,371,357	59.0
80–84	1,834,897	37.1	3,110,470	62.9
85–89	876,501	31.4	1,913,317	68.6
90 or older	350,497	24.2	1,099,272	75.8
Total	26,054,981		33,211,456	

Source: Adapted from US Census Bureau (2000b). 2000 Census of Population and Housing, Summary File 1 (SF1).

The adjustment that this table displays assumes that the overall US age-gender distribution for persons aged 64 years or older more closely approximates the actual distribution among older prisoners than would a flat distribution. Because of rounding, a small (less than 0.5 percent) change in the total number of male prisoners occurred.

Distribution of Male and Female Inmates 65 Years Old or Older

In the prison and correctional facility population, all age groups have more males than females. The ratio of female inmates to male inmates rises until age 30 to 34 years and then declines with each succeeding age group (see Table 6). However, on average, in the US general population females outlive males, and the trend is for the ratio of females to males to increase with age (see Table 7); therefore, we decided that the age distribution for female inmates older than 64 years old would more closely resemble the US general population distribution for males than it would the US general population distribution for females.

Within each small age range, we again distributed the number of inmates equally among the component ages. The maximum age was assumed to be 92 years old.

Juvenile Detainee Population

For juvenile detainees (juveniles hereafter), the SF1 data provided the population count under one age group—17 years old or younger—with no lower limit. In attempting to establish the age range for the juvenile population, we had to consider the infants of adult female inmates.

At the time of the 2000 Census, only four prisons in the United States had nurseries for the infants of nonviolent offenders whose sentences were very short and whose children were expected to remain with their mothers. One of these prisons had a 30-day limit; another had a capacity of 15 infants. According to the Women's Prison Association (2009), "The overwhelming majority of children born to incarcerated mothers are separated from their mothers immediately after birth and placed with relatives or into foster care." Because of this very small population of infants, we decided not to include infants in the synthesized group quarters population.

To determine a more specific age range for juveniles, we obtained data from the BJS Profile of State Prisoners, which lists the age range for juveniles as any age younger than 18 years (Strom, 2000). A search for a lower age limit produced news media coverage that the nation's youngest juvenile detainee was 12 years old, but this situation appeared to

be quite rare. Therefore, for the synthesized group quarters population data, we assigned the SF1 population data on juveniles to the age group 13 to 17 years.

To assign the SF1 counts to year-specific ages by gender, we did not, as with other categories, assign these counts equally to component ages in the range. Instead, to skew the distribution of juveniles toward the older ages for both boys and girls, we used the formulas shown in Table 8.

The final age distribution for all prisoners—from juveniles to prisoners 65 years of age or older—is shown in Table 9. We generated agents in each CBG to match the SF1 count of prisoners, and we assigned each agent an age and gender based on the values listed in this table.

Table 8. Age distribution method for incarcerated juvenile population

Age (Years)	Formula for Determining Percentage	Percent
17	half of all juveniles	50.00
16	half of remaining 50 percent of juveniles	25.00
15	half of remaining 25 percent	12.50
14	half of remaining 12.5 percent	6.25
13	half of remaining 12.5 percent	6.25

Table 9. National number and percentage age distribution, by age and gender, for prison populations, 2000

Age (Years)	Males (N)	Percent of Total Males at This Age	Females (N)	Percent of Total Females at This Age
13	1,185	0.07	136	0.08
14	1,185	0.07	136	0.08
15	2,370	0.13	271	0.16
16	4,739	0.26	543	0.32
17	9,480	0.52	1,086	0.64
18	41,204	2.28	2,108	1.24
19	41,204	2.28	2,108	1.24
20	62,870	3.48	4,239	2.50
21	62,870	3.48	4,239	2.50
22	62,870	3.48	4,239	2.50

(continued)

Table 9. National number and percentage age distribution, by age and gender, for prison populations, 2000 (continued)

Age (Years)	Males (N)	Percent of Total Males at This Age	Females (N)	Percent of Total Females at This Age
23	62,870	3.48	4,239	2.50
24	62,870	3.48	4,239	2.50
25	66,877	3.70	6,488	3.82
26	66,877	3.70	6,488	3.82
27	66,877	3.70	6,488	3.82
28	66,877	3.70	6,488	3.82
29	66,877	3.70	6,488	3.82
30	67,711	3.75	8,455	4.98
31	67,711	3.75	8,455	4.98
32	67,711	3.75	8,455	4.98
33	67,711	3.75	8,455	4.98
34	67,711	3.75	8,455	4.98
35	59,626	3.30	6,640	3.91
36	59,626	3.30	6,640	3.91
37	59,626	3.30	6,640	3.91
38	59,626	3.30	6,640	3.91
39	59,626	3.30	6,640	3.91
40	40,193	2.23	3,677	2.17
41	40,193	2.23	3,677	2.17
42	40,193	2.23	3,677	2.17
43	40,193	2.23	3,677	2.17
44	40,193	2.23	3,677	2.17
45	16,665	0.92	1,308	0.77
46	16,665	0.92	1,308	0.77
47	16,665	0.92	1,308	0.77
48	16,665	0.92	1,308	0.77
49	16,665	0.92	1,308	0.77
50	16,665	0.92	1,308	0.77
51	16,665	0.92	1,308	0.77
52	16,665	0.92	1,308	0.77
53	16,665	0.92	1,308	0.77
54	16,665	0.92	1,308	0.77
55	3,758	0.21	120	0.07
56	3,758	0.21	120	0.07
57	3,758	0.21	120	0.07
58	3,758	0.21	120	0.07
59	3,758	0.21	120	0.07
60	3,758	0.21	120	0.07

(continued)

Table 9. National number and percentage age distribution, by age and gender, for prison populations, 2000 (continued)

Age (Years)	Males (N)	Percent of Total Males at This Age	Females (N)	Percent of Total Females at This Age
61	3,757	0.21	119	0.07
62	3,757	0.21	119	0.07
63	3,757	0.21	119	0.07
64	3,757	0.21	120	0.07
65	872	0.05	98	0.06
66	872	0.05	98	0.06
67	872	0.05	98	0.06
68	872	0.05	98	0.06
69	872	0.05	98	0.06
70	774	0.04	87	0.05
71	774	0.04	87	0.05
72	774	0.04	87	0.05
73	774	0.04	87	0.05
74	774	0.04	87	0.05
75	604	0.03	68	0.04
76	604	0.03	68	0.04
77	604	0.03	68	0.04
78	604	0.03	68	0.04
79	604	0.03	68	0.04
80	364	0.02	41	0.02
81	364	0.02	41	0.02
82	364	0.02	41	0.02
83	364	0.02	41	0.02
84	364	0.02	41	0.02
85	174	0.01	19	0.01
86	174	0.01	19	0.01
87	174	0.01	19	0.01
88	174	0.01	19	0.01
89	174	0.01	19	0.01
90	32	0.00	4	0.00
91	32	0.00	3	0.00
92	279	0.01	26	0.00
13–92	1,806,261	100.00	169,758	100.00

Note: Not shown is the distribution of the resulting data synthesis of people, but the percentages reported here were used to generate the data at the block group level. The number of people for each age was used to calculate each corresponding percentage.

College Dormitory Residents

The SF1 data provide the number of students in college for each CBG. We were unable to locate any external data sources to refine the age distribution for college dormitory residents; therefore, we used additional data from the SF1 Census data along with the PUMS data file.

The SF1 data include counts on all students aged 15 or older enrolled in college, regardless of whether these individuals lived in households or group quarters facilities. As shown in Table 10, we broke out the SF1 population count data for all college students into narrower age ranges than the SF1 group quarters population age ranges (younger than 18, 18–64, and older than 65).

To distribute the SF1 data between students residing and those not residing in college dormitory group quarters, we used data from the PUMS file (US Census Bureau, 2000a) to determine the number of students in college who were living in households, segmented by gender and age.² For ages 15 to 64, we subtracted the number of students living in households (according to the PUMS file) from the SF1 population count of college students (which were counted without regard to living arrangements) to derive the count for group quarters residents. We then estimated and distributed the group quarters residents by age. Table 11 shows the resulting age distribution. As expected, the distribution of students who lived in college dormitory group quarters shows that most of these students were in the age range 18 to 24 years. Although college students older than 50 are likely not living in dormitories, in the absence of any data to tell us what the maximum age might have been, we assigned students from all age categories to college dormitories.

We used the following method to test the assumption that we could combine the enumerated SF1 data set and the data derived from the PUMS file and

² We counted as college students any person records with an Enroll field value of 2 (public school or college) or 3 (private school or college) and a Grade field value of 6 (college undergraduate) or 7 (graduate or professional school). As noted previously, the PUMS file is a sample of Census data; therefore, weights were not available for weighting the group quarters population count. Only the household data could be weighted.

Table 10. National college dormitory population, by age and gender

Age Group	Males			Females		
	All Students (SF1 ^a)	Household Students (PUMS)	Group Quarters Students (Derived ^b)	All Students (SF1 ^a)	Household Students (PUMS)	Group Quarters Students (Derived ^b)
15–17	32,945	27,700	5,245	41,644	36,200	5,444
18–24	4,241,329	3,235,136	1,006,193	4,961,751	3,806,135	1,155,616
25–34	1,957,404	1,854,521	102,883	2,202,202	2,151,722	50,480
35 or older	1,687,950	1,607,962	79,988	2,358,018	2,292,105	65,913
Total	7,919,628	6,725,319	1,194,309	9,563,615	8,286,162	1,277,453

PUMS = Census Public Use Microdata Sample; SF1 = Census Summary File 1.

^a The SF1 data report only 105 college students aged 65 or older (48 men and 57 women) in the entire United States; therefore, we combined this age category with the 35 or older age category for this calculation. The effect of combining these data should be negligible because of the relative sizes of the populations.

^b All students minus household students.

Table 11. Number and percentage distribution, by age and gender, for college dormitory group quarters residents

Age	Males (N)	Percent of Total Males at This Age	Females (N)	Percent of Total Females at This Age
15	1,748	0.15	1,815	0.14
16	1,748	0.15	1,815	0.14
17	1,749	0.15	1,814	0.14
18	143,742	12.04	165,088	12.92
19	143,742	12.04	165,088	12.92
20	143,742	12.04	165,088	12.92
21	143,742	12.04	165,088	12.92
22	143,742	12.04	165,088	12.92
23	143,742	12.04	165,088	12.92
24	143,741	12.04	165,088	12.92
25	10,289	0.86	5,048	0.40
26	10,289	0.86	5,048	0.40
27	10,289	0.86	5,048	0.40
28	10,288	0.86	5,048	0.40
29	10,288	0.86	5,048	0.40
30	10,288	0.86	5,048	0.40
31	10,288	0.86	5,048	0.40
32	10,288	0.86	5,048	0.40
33	10,288	0.86	5,048	0.40
34	10,288	0.86	5,048	0.40
35	2,581	0.22	2,127	0.17
36	2,581	0.22	2,127	0.17
37	2,581	0.22	2,127	0.17
38	2,581	0.22	2,127	0.17
39	2,581	0.22	2,127	0.17
40	2,581	0.22	2,127	0.17

(continued)

Table 11. Number and percentage distribution, by age and gender, for college dormitory group quarters residents (continued)

Age	Males (N)	Percent of Total Males at This Age	Females (N)	Percent of Total Females at This Age
41	2,581	0.22	2,126	0.17
42	2,580	0.22	2,126	0.17
43	2,580	0.22	2,126	0.17
44	2,580	0.22	2,126	0.17
45	2,580	0.22	2,126	0.17
46	2,580	0.22	2,126	0.17
47	2,580	0.22	2,126	0.17
48	2,580	0.22	2,126	0.17
49	2,580	0.22	2,126	0.17
50	2,580	0.22	2,126	0.17
51	2,580	0.22	2,126	0.17
52	2,580	0.22	2,126	0.17
53	2,580	0.22	2,126	0.17
54	2,580	0.22	2,126	0.17
55	2,580	0.22	2,126	0.17
56	2,580	0.22	2,126	0.17
57	2,580	0.22	2,126	0.17
58	2,580	0.22	2,126	0.17
59	2,580	0.22	2,126	0.17
60	2,580	0.22	2,126	0.17
61	2,580	0.22	2,126	0.17
62	2,580	0.22	2,126	0.17
63	2,580	0.22	2,126	0.17
64	2,580	0.22	2,126	0.17
Total	194,309	100.15	1,277,453	100.13

Note: Because of rounding, totals do not sum to 100%.

associated weights. We added the number of total college students in households (from the PUMS file) to the number of total students in college dormitories reported in the SF1 group quarters data. We compared the summed total with the total number of college students listed in the SF1 data. The summed total from the PUMS file and SF1 group quarters data approximated the SF1 total, with a difference of only 2.3 percent; this small difference suggests that this approach is accurate enough to obtain a representative distribution of age and gender for the college group quarters population.

To create the final age distribution for college students, we divided the total number of students in a specific age range (e.g., 18–24) by the number of specific ages within that range to determine how many agents to assign to each specific age. For the 65 or older age group, we assigned one agent to each age, beginning with age 65. This group is not shown in Table 11 because it was combined with the group of individuals 35 or older to determine the final age distribution.

We had no reliable mechanism for determining the size of any particular dormitory building; therefore, we assigned all students to one dormitory per campus. We found no case in which more than one college had any residential (i.e., dormitory) population existing in a single CBG, so we did not expect that populations from separate campuses were mixed.

Nursing Home Residents

The SF1 data provide information on the population count and gender of nursing home group quarters residents by CBG, Table 12. To create the nursing home population age distribution, we used the age distribution reported by the Centers for Disease Control and Prevention (CDC), National Center for Health Statistics, in the 1999 National Nursing Home Survey (CDC, 1999). An important consideration for the extent of interaction between individuals living in group quarters was the size of the group quarters facility. As mentioned earlier, we used the nursing home atlas to estimate average and maximum nursing home sizes, by state; we used these estimates to determine the number of nursing homes in each CBG and the related sizes for each state.

We generated the number of agents as reported in the SF1 data for each CBG. Table 13 shows the distribution used to assign age and gender information to each agent in the nursing home group quarters population.

Table 12. National nursing home population, by age and gender, 1999

Age Group	Percent Males	Percent Females	Percent Total
Under 65	17.5	6.7	9.8
65–74	18.4	9.5	12.0
75–84	32.7	31.5	31.8
85 or over	31.5	52.4	46.5
All ages	100.0	100.0	100.0
Percentage of Total	28.1	71.9	100.0

Note: Population derived from “Nursing home residents, number, percent distribution, and rate per 10,000, by age at interview, according to gender, race, and region.”

Source: Centers for Disease Control and Prevention, National Center for Health Statistics (1999).

Table 13. Distribution of nursing home residents, by age and gender

Age	Males (percent)	Females (percent)
1	0.00042	0.00000
3	0.00042	0.00000
4	0.00042	0.01706
8	0.00000	0.01706
10	0.00042	0.00000
14	0.00042	0.00000
15	0.00042	0.00000
16	0.00000	0.00000
17	0.00000	0.00000
18	0.00042	0.00000
19	0.00000	0.01706
20	0.00042	0.01706
21	0.00085	0.00000
22	0.00127	0.03412
23	0.00042	0.03412
24	0.00085	0.01706
25	0.00000	0.01706
26	0.00085	0.00000
27	0.00042	0.00000

(continued)

Table 13. Distribution of nursing home residents, by age and gender (continued)

Age	Males (percent)	Females (percent)
28	0.00085	0.00000
29	0.00000	0.08531
30	0.00042	0.01706
31	0.00255	0.05119
32	0.00042	0.01706
33	0.00085	0.01706
34	0.00085	0.10237
35	0.00127	0.01706
36	0.00170	0.03412
37	0.00255	0.01706
38	0.00340	0.08531
39	0.00255	0.10237
40	0.00170	0.05119
41	0.00255	0.06825
42	0.00340	0.08531
43	0.00340	0.13650
44	0.00552	0.15356
45	0.00637	0.18768
46	0.00510	0.10237
47	0.00467	0.13650
48	0.00382	0.10237
49	0.00425	0.17062
50	0.00255	0.15356
51	0.00850	0.32418
52	0.00850	0.22181
53	0.00552	0.20474
54	0.00552	0.22181
55	0.00510	0.40949
56	0.00850	0.39242
57	0.00552	0.25593
58	0.00595	0.34124
59	0.01147	0.32418
60	0.00552	0.35830
61	0.00977	0.35830
62	0.00807	0.32418
63	0.01062	0.69954
64	0.01232	0.47773
65	0.01402	0.44361
66	0.01529	0.59717
67	0.01274	0.64835
68	0.01529	0.87016
69	0.01699	0.85310
70	0.02039	0.75073
71	0.02167	1.07490
72	0.01869	1.10903
73	0.02506	1.48439
74	0.02464	1.72326

(continued)

Table 13. Distribution of nursing home residents, by age and gender (continued)

Age	Males (percent)	Females (percent)
75	0.02379	2.25218
76	0.02719	2.33749
77	0.02846	2.37161
78	0.02889	2.62754
79	0.03738	3.10527
80	0.03568	3.10527
81	0.03398	3.22471
82	0.03738	3.65125
83	0.03611	4.26548
84	0.03696	4.41904
85	0.03738	4.38492
86	0.03781	4.84559
87	0.03696	4.14605
88	0.03016	4.84559
89	0.03314	4.50435
90	0.01954	4.33373
91	0.01954	3.83894
92	0.02506	3.90718
93	0.02082	3.78775
94	0.01189	3.19058
95	0.00935	2.59341
96	0.01274	2.08156
97	0.00467	1.62088
98	0.00595	1.16021
99	0.00127	0.97253
100	0.00042	0.69954
101	0.00127	0.47773
102	0.00085	0.11943
103	0.00000	0.18768
104	0.00085	0.18768
105	0.00000	0.11943
106	0.00000	0.05119
107	0.00000	0.01706
108	0.00000	0.00000
109	0.00000	0.01706
110	0.00000	0.00000
111	0.00000	0.00000
112	0.00000	0.01706
Total	99.99	100.00

Note: Not shown is the distribution of the resulting data synthesis of people, but the percentages reported here were used to generate the synthesized data at the block group level. The number of people for each age was used to calculate each corresponding percentage.

Results

We generated the synthesized Group Quarters Population Database for the entire United States. Figure 1 is a flowchart of the data-generation process.

For 2000, the SF1 data show 7,778,633 group quarters residents in a US population of 281,421,906; therefore, 2.8 percent of the nation's population lived in group quarters facilities in 2000 (US Census Bureau, 2000b). Table 14 gives the number of group quarters facilities and the associated residents for the Group Quarters Population Database, which includes 6,115,802 group quarters residents. Each group quarters facility included in the database has a set of associated person records that represent all the group quarters residents, with associated age and gender characteristics consistent with the marginal distributions available from the various data sources. As explained, the age and gender distributions outlined in the database differ by group quarters category.

The following subsections show that the facility and person records in the synthesized Group Quarters

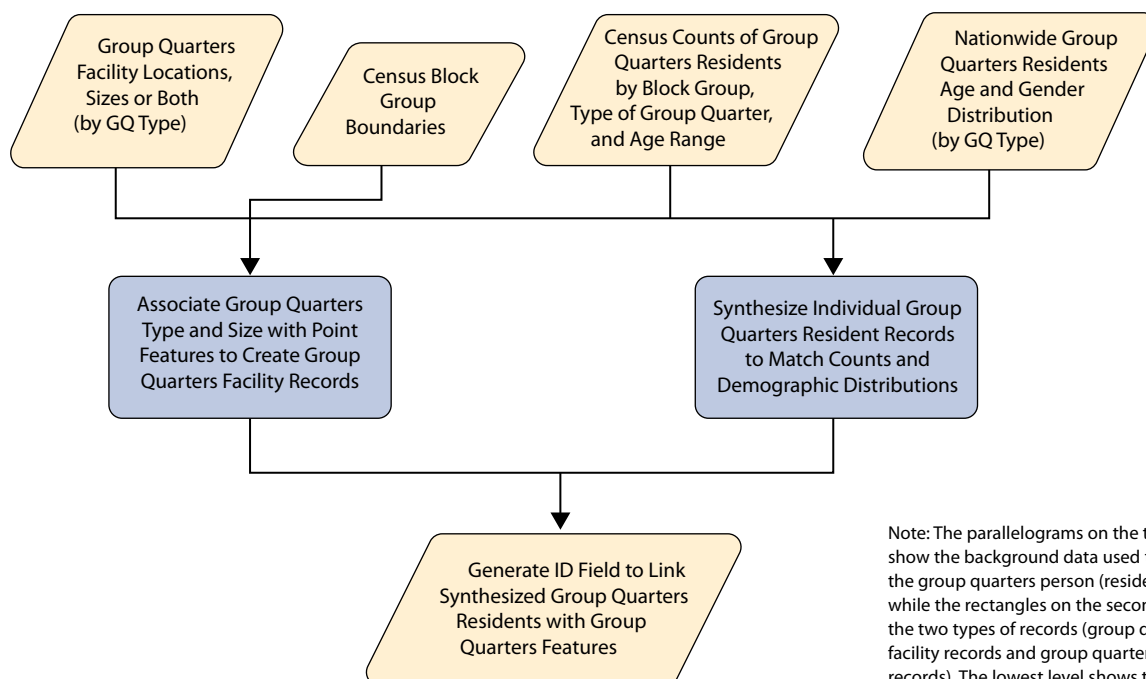
Population Database can be combined with the RTI US Synthesized Population Database to represent the entire household and group quarters population of the United States. As shown in these subsections, the similarities in the design of the two databases assisted with this compatibility.

Table Structure of the Group Quarters Facility Records

Table 15 shows the basic structure of a group quarters facility record from the Group Quarters Population Database. In this table, the GQ_ID field contains the unique identifier we constructed. It uses the letter G to indicate group quarters, followed by

- the CBG number (e.g., 370190203021);
- a letter specifying the category of group quarters, with
 - M for military,
 - P for prison or correctional facility,
 - C for college or university dormitory, or
 - N for nursing home; and
- a sequential group quarters unit number (e.g., 01).

Figure 1. Flow chart of the data-generation process for the Group Quarters Population Database



Note: The parallelograms on the top level show the background data used to develop the group quarters person (resident) records, while the rectangles on the second level show the two types of records (group quarters facility records and group quarters person records). The lowest level shows the process of generating identifiers so that the group quarters residents could be associated with a particular group quarters facility.

As with the data provided in the US Synthesized Population Database, RTI's synthesized group quarters data were of adequate geographic resolution for MIDAS; that is, they were available at the CBG level.

For comparison, Table 16 shows a household record from the US Synthesized Population Database. The household identifier (HH_ID) was constructed with the use of the CBG number and a sequential number for the households within the CGB. The ST_Serialno field links that record back to the corresponding PUMS database record.

The other fields included in the records of both databases are the latitude and longitude coordinates for the household point and the state and county Federal Information Processing Standards (FIPS) codes. The FIPS codes serve as additional location identifiers for the group quarters facility (US Census Bureau, 2009c).

The latitude and longitude coordinates for both databases are used to place the facilities or households (i.e., points) on a map to model the interactions of agents in a simulation. Figure 2 illustrates the manner in which the points of the Group Quarters Population

Table 14. Number of group quarters facilities and associated persons in the Group Quarters Population Database, by group quarters category

Group Quarters Category	Number of Facilities	Population Count	Percentage of Group Quarters Population That Is Female	Percentage of Group Quarters Population That Is Younger Than 18 Years	Percentage of Group Quarters Population That Is Older Than 64 Years
Military Bases	517	355,155	13.5% 47,873	0.6% 2,260	0% 6
Prisons or Correctional Facilities	14,255	1,976,019	8.6% 169,758	1.1% 21,130	0.8% 15,882
College Dormitories	3,808	2,064,128	53.6% 1,106,547	0.5% 10,528	0% 105
Nursing homes	24,874	1,720,500	71.6% 1,232,132	1.1% 18,981	90.5% 1,557,800
Total Included in Group Quarters for United States	43,454	6,115,802	41.8% 2,556,310	0.8% 52,899	25.7% 1,573,793

Note: This data set excludes group quarters from the "other institutions" category, which was too broad to be useful for agent-based modeling. All numbers are consistent with the numbers from the Census Summary File 1 (SF1) data.

Table 15. Structure of group quarters facility record for the Group Quarters Population Database

GQ_ID	GQ_Type	Num_Residents	Latitude	Longitude	State FIPS Code	County FIPS Code
G370190203021M01	Military	1000	Xxxxxx	Xxxxxx	37	019

GQ_ID = group quarters facility identifier; FIPS = Federal Information Processing Standards.

Table 16. Household record for the US Synthesized Population Database

HH_ID	ST_Serialno	Latitude	Longitude	State FIPS Code	County FIPS Code
482917003005_362	48_3993312	Xxxxxx	Xxxxxx	48	291

HH_ID = household identifier; FIPS = Federal Information Processing Standards.

Note: The HH_ID is a unique identifier for each household in the US Synthesized Population Database. The ST_Serialno field is the state code combined with the Census Public Use Microdata Sample (PUMS) serial number; it links to the PUMS household record for number of people, household income, number of vehicles, and many other variables.

Database can be displayed in map format; it presents a map of Allegheny County, Pennsylvania, displaying the resulting locations for each group quarters point. In Allegheny County, 3.2 percent of the population (40,617 of 1,281,666) was living in group quarters facilities at the time of the 2000 Census.

Table Structure of the Group Quarters Person Records

Table 17 shows the structure of a group quarters person record from the Group Quarters Population Database. In this record, the Person_ID (G370190203021M01_1) was an identifier constructed from the GQ_ID (see Table 15), with a sequential ID appended to the end (1–N for the number of people in the associated facility). The GQ_ID was included so that the records might be joined

to the associated group quarters facility record, which was illustrated in Table 15. Other fields include a person identifier, the age and gender of each person, and the state and county FIPS codes to allow for easy selection by these locations.

The group quarters person record shown in Table 17 was designed to work easily with records from the US Synthesized Population Database. A person record for the US Synthesized Population Database includes the HH_ID and a sequential number 1–N for each person in the household. The HH_ID allows analysts to join each record to its associated household record. The ST_Serialno field allows the records to be joined to the associated person record in the PUMS file. Additional fields include a person number and state and county FIPS codes.

Figure 2. Map of Allegheny County, Pennsylvania, showing the point locations for each synthesized group quarters facility

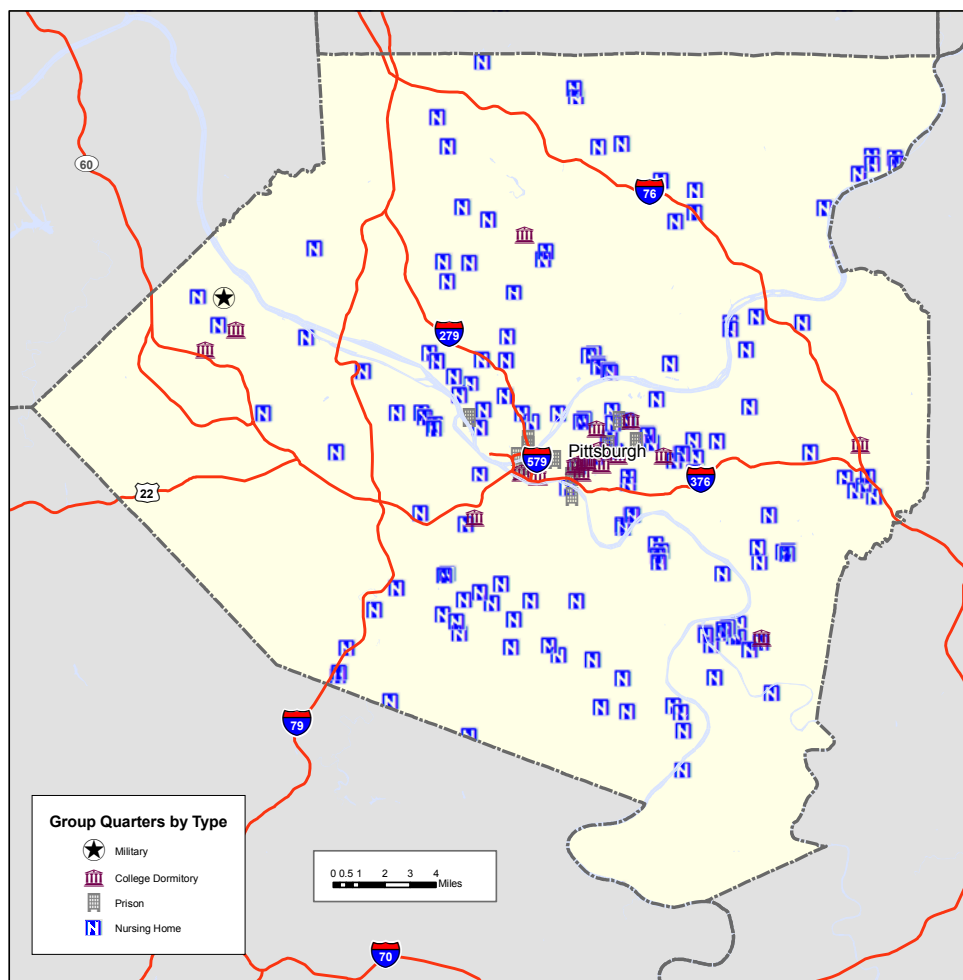


Table 18 shows an example person record from the US Synthesized Population Database. The ST_Serialno field, when combined with the Person_Num field, is used to link the record to the PUMS table, which provides the age and gender of each person.

For example, when considering a model of influenza transmission or mortality rates, researchers must factor in increased transmission rates for children and increased mortality rates for the elderly (Germann, Kadau, Longini, & Macken, 2006). Consequently, including data on nursing home locations and residents may have a considerable effect on modeling results in an influenza mortality model, but less of an effect in an influenza incidence model. In contrast, models of HIV infection rates may be more influenced by data on prison and military populations than by data from other populations (US Department of Health and Human Services, 2005).

We estimated the group quarters distributions from data sources that provide national rather than local data. For specific military bases or nursing homes,

the estimated distributions are unlikely to differ substantially from the actual populations. However, many prison facilities and some colleges are gender-specific. Consequently, the actual population for a specific prison or college dormitory may differ substantially from the synthesized population. For this reason, the Group Quarters Population Database should not be used to model and infer detailed results for small, local populations.

Despite source limitations, we found that the information from the multiple data sources was compatible. For instance, the HSIP data show that category-specific group quarters facilities are generally located in or near a CBG where the SF1 data indicate the presence of group quarters residents for that category. Furthermore, the detailed age distributions obtained from non-Census sources were generally consistent with the total Census counts for broad age ranges. Consequently, we were able to develop a synthesized Group Quarters Population Database that is consistent with data from many sources.

Table 17. Group Quarters Population Database group quarters person record

Person_ID	GQ_ID	Age	Sex	State FIPS Code	County FIPS Code
G370190203021M01_1	G370190203021M01	20	M	37	019

GQ_ID = group quarters identifier; FIPS = Federal Information Processing Standards.

Table 18. US Synthesized Population Database person record

Person_ID	HH_ID	Person_Num	ST_Serialno	State FIPS Code	County FIPS Code
482917003005_362_1	482917003005_362	1	48_3993312	37	019

HH_ID = household identifier; FIPS = Federal Information Processing Standards.

Note: HH_ID is a unique identifier for each household in the US Synthesized Population Database. The ST_Serialno field is the state code combined with the Census Public Use Microdata Sample (PUMS) serial number.

Conclusions

The age and gender characteristics of the group quarters categories differ from those for the total population for the United States. The relative impact of including the group quarters data in an ABM depends on the extent to which a model relies on age-dependent or gender-dependent activities. In addition, it depends on whether the model aims to simulate the interaction between agents who live in households and agents who live in group quarters.

The synthesized data of the Group Quarters Population Database represent, at an individual level, the US population that resides in group quarters. The data developed as part of this process also are compatible with the various reference sources used: the group quarters facility records match the various HSIP database files, and the group quarters person records match the various national data files.³ For agent-based modeling, the database can be used with RTI's US Synthesized Population Database (Wheaton et al., 2007; Wheaton et al., 2009) to provide

³ An independent verification of these source data files was beyond the scope of this project.

detailed data on the household and group quarters populations.

We anticipate that the group quarters data will be used for research and publication of studies that will advance the knowledge of epidemiology and help inform policy makers about the best strategies for responding to specific epidemic scenarios. We expect the group quarters data to be used in

- various MIDAS simulations, including simulations of the spread of various strains of influenza, such as H1N1 and H5N1, or "avian flu";
- models of the spread of HIV/AIDS within and between prison populations and the general population; and
- models of the spread of MRSA (methicillin-resistant *Staphylococcus aureus*) and the relationship between the hospital-acquired and community-acquired strains of the disease.

Development of the synthesized group quarters population data has provided MIDAS and other researchers with the data required for conducting more comprehensive agent-based modeling. These advances will further enhance researchers' ability to predict the outcomes of proposed government interventions in the event of a disease outbreak.

References

- Agency for Healthcare Research and Quality. (2007). Emergency preparedness atlas: US nursing home and hospital facilities. Retrieved November 8, 2010, from <http://www.ahrq.gov/prep/nursinghomes/atlas>
- Beck, A. J., & Karberg, J. (2001). Prison and jail inmates at midyear 2000. (NCJ 185989). US Department of Justice, Bureau of Justice Statistics. Retrieved January 4, 2011, from <http://bjs.ojp.usdoj.gov/content/pub/pdf/pjim00.pdf>
- Centers for Disease Control and Prevention, National Center for Health Statistics. (1999). 1999 National Nursing Home Survey, Table 3, Nursing home residents, number, percent distribution, and rate per 10,000, by age at interview, according to sex, race, and region: United States, 1999. Retrieved November 4, 2010, from http://www.cdc.gov/nchs/data/nnhsd/NNHS99CurrentRes_selectedchar.pdf
- Germann, T. C., Kadau, K., Longini, I. M., & Macken, C. A. (2006). Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences*, 103(15), 5935-5940. doi: 10.1073/pnas.0601266103
- National Geospatial-Intelligence Agency. (2007). Homeland Security Infrastructure Program (HSIP) gold 2005 database. Washington, DC: US Department of Homeland Security. [Available only to qualified users who have signed a contract with the Department of Homeland Security.]
- Strom, K. (2000). Profile of state prisoners under age 18, 1985-97 (NCJ 176989). Washington, DC: US Bureau of Justice Statistics. Retrieved January 4, 2011, from <http://bjs.ojp.usdoj.gov/content/pub/pdf/pspa1897.pdf>

- US Census Bureau. (2000a). 2000 Census of population and housing, Public Use Microdata Sample: 2000 [Data file]. Retrieved November 8, 2010, from http://ftp2.census.gov/census_2000/datasets/PUMS/FivePercent/.
- US Census Bureau. (2000b). 2000 Census of population and housing, summary file 1 (SF1) [Data file], Retrieved January 4, 2011, from <http://www.census.gov/census2000/sumfile1.html>
- US Census Bureau. (2009c). Federal Information Processing Standards (FIPS) Codes [now American National Standards Institute (ANSI) Codes]. Retrieved from <http://www.census.gov/geo/www/fips/fips.html>
- US Department of Defense. (2005). Selected manpower statistics: Fiscal year 2005 [Data file] 2005. Retrieved November 8, 2010, from <http://siadapp.dmdc.osd.mil/personnel/M01/fy05/m01fy05.pdf>
- US Department of Health and Human Services. (2005). HHS pandemic influenza plan. Retrieved November 8, 2010, from <http://www.hhs.gov/pandemicflu/plan/>.
- Wheaton, W. D., Cajka, J. C., Chasteen, B. M., Wagener, D. K., Cooley, P. C., & Ganapathi, L. (2009). Synthesized population databases: A US geospatial database for agent-based models (RTI Press Publication No. MR-0010-0905). Research Triangle Park, NC: RTI Press.
- Wheaton, W. D., Chasteen, B. M., Cajka, J. C., Allpress, J., Cooley, P. C., Ganapathi, L., & Pratt, J. G. (2007). A nationwide geo-referenced synthesized agent database for infectious disease models. *Advances in Disease Surveillance*, 4,19. Retrieved April 19, 2010, from <http://www.isdsjournal.org/article/view/2015/1583>
- Women's Prison Association. (2009). *Mothers, infants and imprisonment: A national look at prison nurseries and community-based alternatives*. Retrieved January 28, 2011, from <http://www.wpaonline.org/pdf/Mothers%20Infants%20and%20Imprisonment%202009.pdf>

Acknowledgments

The project described in this report was supported by grant number U01GM070698 (Models of Infectious Disease Agent Study—MIDAS) from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official view of the National Institute of General Medical Sciences or the National Institutes of Health.

RTI International is an independent, nonprofit research organization dedicated to improving the human condition by turning knowledge into practice. RTI offers innovative research and technical solutions to governments and businesses worldwide in the areas of health and pharmaceuticals, education and training, surveys and statistics, advanced technology, international development, economic and social policy, energy and the environment, and laboratory and chemistry services.

The RTI Press complements traditional publication outlets by providing another way for RTI researchers to disseminate the knowledge they generate. This PDF document is offered as a public service of RTI International. More information about RTI Press can be found at www.rti.org/rtipress.