

Synthesized Group-Quarters Agents as Part of a U.S. Synthesized-Agent Database

By Bernadette Chasteen,¹ Bill Wheaton, Philip C. Cooley,
Laxminarayana Ganapathi, Diane Wagener
RTI International

ABSTRACT

The Models of Infectious Disease Agent Study (MIDAS)² seeks to develop and refine models for the spread of infectious disease to test containment and intervention strategies to help guide policy makers on the best approach for preventing disease spread and mortality. Some of these models depend on synthesized-agent databases that contain records representing the people who might contract and spread diseases. To date, these databases have represented only people living in households, not group quarters such as military bases, prisons, college dormitories, or nursing homes. There has long been concern about how such populations might affect models of infectious disease outbreak because these people live close to others, but these populations are not easily quantified due to Census reporting methods designed to protect individual privacy. RTI has developed methods to synthesize group-quarters agents that match overall U.S. Census group-quarters data and provide adequate geographic resolution for MIDAS.

INTRODUCTION / BACKGROUND

MIDAS relies on synthesized-agent databases to represent the people who might contract and spread diseases. To date, these databases contain records that represent individual people who live in households, based on Census data. They have not included records to represent people living in group quarters such as military bases, prisons, college dorms, and nursing homes. There has long been concern about how such populations, which make up almost 3 percent of the U.S. population, might affect models of infectious disease outbreak because the people live in close proximity to each other and may have an increased chance of contracting and spreading infectious diseases. The Census reports these populations in a way that protects individuals' privacy, making it difficult to construct a synthesized group-quarters population from Census data alone. This paper presents methods RTI has developed to synthesize group-quarters agents that match overall U.S. group-quarters Census data, based on Census data and additional data sources.

Ideally, the data used to generate synthesized-agent records should be scalable to reflect the full population, be presented at a sufficiently fine level of geographic resolution to model individual interactions that affect disease spread, and provide information that informs those interactions. Geographically, synthesized agents need to be located more specifically than a Census block group, which can range in size from 65 to 21,700 acres. No single data source exists that meets these criteria without manipulation.

¹ RTI International, P.O. Box 12194, Research Triangle Park, NC, 27709. Phone: 919-541-7167. Fax: 919-541-7155. E-mail: bmc@rti.org. Web site: www.rti.org.

² Supported by the National Institute of General Medical Science of the National Institutes of Health; see <http://www.nigms.nih.gov/Initiatives/MIDAS/>.

The synthesized household population (i.e., those not living in group quarters) is based on the 2000 Census Public Use Microdata Samples (PUMS) file (U.S. Census Bureau, 2005a). This file contains a 5-percent sample of the housing units in the United States and the people living in them. These housing units are identified as a housing unit (i.e., a household), institutional group quarters, and non-institutional group quarters. For households, each record contains a housing unit weight that, when applied to the individual records, expands the 5-percent sample to the total population. The PUMS data consists of specific information about each household, as well as specific information about the people in each household. Those data include data relevant to models of disease spread (e.g., age, gender, occupation, method of transportation to work, relationship to householder). However, The PUMS data reflect an even coarser geographic resolution than Census block groups. To achieve the needed resolution for MIDAS in the household data, households were randomly located using Census Summary File 1 (SF1) (U.S. Census Bureau, 2005b) block group household counts, and the PUMS data were then assigned to those randomly located households (Wheaton and Chasteen, 2005). Thus, for households, the PUMS data supplemented by the SF1 data and random location of households within block group provides the information needed to construct the synthesized agents needed for MIDAS.

By contrast, the PUMS data cannot be used to generate synthesized group-quarters populations. Although the PUMS data contains 391,377 records coded as group quarters (either institutional or non-institutional), the housing unit weight is set to 0 for all group-quarters records, making it impossible to expand the sample to the full population. In addition, each group-quarter housing unit record has only one associated person record (as contrasted with the household housing unit records, which have multiple Person records reflecting the number of persons in the household).

Therefore, we turned instead to data from the SF1, which provides the number of people in some types of group quarters by gender, age, and group-quarters type on a block group basis. **Table 1** shows the types of group quarters for which the SF1 reports number of residents.

Table 1. Group-Quarters Types with Population by Block Group, Gender, and Age

Type of Group Quarters	Census category	SF1 field
Military Quarters	Non-institutional	P038009
Correctional Institutions	Institutional	P038003
College Dormitories (on and off campus)	Non-institutional	P038010
Nursing Homes	Institutional	P038006
Other Institutions ^a	Institutional	P038007

Source: Census SF1, Table P38.

^a "Other Institutions" includes facilities for the chronically ill, mentally ill, or physically handicapped; facilities for drug and alcohol abuse; wards in general hospitals or military hospitals for patients with no usual home; and juvenile institutions.

Using these data and data on the actual locations of correctional institutions, nursing homes, military quarters, and college dormitories, we propose to generate points within each block group for those types of group quarters and assign the associated synthesized persons (by age and sex) to the appropriate group-quarters point. We did not generate group-quarters records for the Other Institutions category: it is too broad to be of much use for MIDAS studies because it contains many different types of group quarters with presumably different types of social interactions between members.

This paper describes the creation of a synthesized group-quarters-agent data set for use with the synthesized-agent database being used for the MIDAS models. There are anomalies with some model runs which researchers have suspected of being caused by the presence of group quarters which were not accounted for in the synthetic population. The North Carolina Metro Study has been designed to incorporate the synthesized population and associated data sets into a model for the spread of infectious disease. These group-quarters records were created as a pilot study to test the approach. The study area consists of three metropolitan regions in North Carolina (defined by the 2000 U.S. Census), one of which extends into South Carolina. Once we successfully incorporate the group quarters into the model runs, we can begin to assess the effects of this population on disease transmission. Ultimately, our goal is to apply the method for creating group-quarters populations to the entire United States.

METHODS

The goal was to generate records for group quarters and associated records for the people within each group-quarters unit that mimic the existing synthesized population data structure as much as possible so that the data sets can be used together with minimal difficulty. The person records do not include all the detailed information contained in the synthesized household population generated from the PUMS data, but they do include sex and age.

Tables 2 and 3 show an example group-quarters record and associated person record. The GQ_ID is an identifier constructed from the letter G (for group quarters), the block group number (here, 370190203021), a letter specifying the type of group quarters (M for military, N for nursing home, C for college or university dorm, or P for prisons/correctional institutions), and a sequential group quarters unit number (here, 01). The Person ID adds a sequential person number to the GQ_ID. The State and County FIPS (Federal Information Processing Standard) codes are additional location identifiers.

Table 2. Example Group-Quarters Table

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

Table 3. Example Group-Quarters Synthesized Person Table

Person_ID	GQ_ID	Person_Num	Age	Sex	State FIPS	County FIPS
G370190203021M01_1	G370190203021M01	1	20	M	37	019

To establish locations at the desired level of geographic resolution, we looked for the best available databases that included locations for prisons, nursing homes, military bases, and college dormitories, and used them to distribute the SF1 group-quarter populations to specific group quarters. **Table 4** lists the data sources we identified for locations. We were able to locate national data sources for location.

Table 4. Data Sources for Group-Quarters Locations

Type of Group Quarters	Data Source	Number of Group Quarters	
		NC	SC
Military Quarters	HSIP Gold 2005 Database ^{a,b}	60	37
Federal, State and Local Prisons	HSIP Gold 2005 Database ^a	82	29
U.S. Colleges, Community Colleges, & Universities	HSIP Gold 2005 Database ^a	143	82
Nursing Homes	HSIP Gold 2005 Database ^a Nursing Home Atlas ^c	21	8

^a The Homeland Security infrastructure Program (HSIP) Gold Database is a unified homeland infrastructure geospatial data inventory assembled by the National Geospatial Intelligence Agency (NGA) in partnership with the Department of Homeland Security (DHS) for common use by the Homeland Security and Homeland Defense (HLS/HD) Community. Data retrieved June 2007.

^bBases are shown using the block group centroid for security reasons.

^cThe Nursing Home Atlas is an RTI data set prepared for the Centers for Medicaid and Medicare Services (CMS) of the U.S. Department of Health & Human Services (AHRQ, 2007). Data retrieved June 2007.

The SF1 provides the number and gender of group-quarters populations, and breaks out population by gender into the following age groups:

- Under 18 years
- 18–64 years
- Over 64 years.

To refine the age data, we needed additional data on the age distributions of these group-quarters populations for the year 2000. These additional data will allow us to generate person records with associated age and gender information. **Table 5** lists the data sources we identified for age distribution. We were able to locate national data sources for age distribution.

Table 5. Data Sources for Group-Quarters Age Distribution

Type	Spatial	Scope	Source
Military Population	Yes	Nationwide	U.S. DOD (2005)
Prisons	Yes	Nationwide	U.S. DOJ (2000a)
Colleges/Universities	Yes	Nationwide	U.S. Census SF1 data (U.S. Census Bureau, 2005b) and PUMS data (U.S. Census Bureau, 2005a)
Nursing Homes	Yes	Nationwide	Nursing Home Atlas ^a

^aThe Nursing Home Atlas is an RTI data set prepared for the Centers for Medicaid and Medicare Services (CMS) of the U.S. Department of Health & Human Services (AHRQ, 2007). Data retrieved June 2007.

We used the location data to generate points reflecting more accurate locations for group quarters. We then assigned the count of residents of that type (e.g., military residents) for a specific block group from the SF1 data to the group-quarters points in that block group. Each point becomes a group quarter of a specific type. Where accurate locations could not be found for group quarters and we could demonstrate that it was rare for more than one group quarters of a specific type (e.g. military bases) to occur within a block group, we generated a single point at the block group centroid for that group-quarters type and assigned all population for that type of group quarters in that block group to that point.

We then refined the age distributions using the additional age data. We used the SF1 data to obtain, for example, the number of males under 18 in the military population, and then used the age distribution of males found in the population for military personnel to assign specific person records for the block group.

The following sections describe how we used the SF1, location, and age distribution data for each type of group quarters.

Military Quarters

Location and Number of Residents

We found no case where more than one military base was in the same block group for North Carolina or South Carolina. Therefore, we associated all military personnel with one point for each block group, placed at the block group centroid. As a quality assurance check, we used the HSIP military-base layer and found that all military populations enumerated in the 2000 Census SF1 file were in or near military bases. We found some block groups with fewer than 50 military personnel, but all of these were in or near larger bases or Coast Guard stations. We assigned these populations to military group-quarters units in the block group indicated by the SF1 data.

Age Distribution

We used the age distribution for military personnel published by the Defense Manpower Data Center (U.S. DOD, 2005) to derive the age distribution for military personnel shown in **Table 6**. The original table presented some ages in ranges (e.g., 22–24, 24–29). In those cases, we distributed the population in an age range equally among the component ages.

Table 6. Age Distribution of Military Personnel

Age	Male (%)	Female (%)	Age	Male (%)	Female (%)
17	0.25	0.48	34	3.00	2.43
18	3.42	3.88	35	3.17	2.43
19	5.90	7.28	36	3.25	2.43
20	7.02	8.25	37	3.25	2.43
21	6.94	8.25	38	2.91	1.94
22	6.25	7.77	39	2.48	1.94
23	5.56	6.31	40	1.37	1.068
24	4.79	5.83	41	1.37	1.068
25	4.46	4.85	42	1.37	1.068
26	4.11	4.37	43	1.37	1.068
27	3.85	3.88	44	1.37	1.068
28	3.60	3.40	45	0.462	0.388
29	3.60	2.91	46	0.462	0.388
30	3.34	2.91	47	0.462	0.388
31	3.00	2.91	48	0.462	0.388
32	3.00	2.43	49	0.462	0.388
33	2.91	2.43	50	0.04875	0.06125

Age	Male (%)	Female (%)
51	0.04875	0.06125
52	0.04875	0.06125
53	0.04875	0.06125
54	0.04875	0.06125
55	0.04875	0.06125
56	0.04875	0.06125
57	0.04875	0.06125
58	0.04875	0.06125

Age	Male (%)	Female (%)
59	0.04875	0.06125
60	0.04875	0.06125
61	0.04875	0.06125
62	0.04875	0.06125
63	0.04875	0.06125
64	0.04875	0.06125
65	0.04875	0.06125
Total	100.00	100.00

Prisons

Location and Number of Residents

Based on the HSIP prisons layer, many block groups contain more than one correctional facility. We found 43 cases where the HSIP data showed more than one facility for a block group and the Census data showed prison populations. In North and South Carolina, we found 1 block group with eight facilities, 4 block groups with three facilities, and 38 block groups with two facilities.

If there was only one correctional facility in a block group, we assigned all prison population from the SF1 data to that facility. If there was more than one correctional facility in a block group, we assigned the total group-quarters prison population to individual facilities based on certain decision rules.

First, because prison populations are typically housed in separate facilities based on age (adult vs juvenile) and gender, we identified four prison types for assigning populations:

- Male under 18
- Male over 18
- Female under 18
- Female over 18.

Based on the gender and age characteristics of the populations reported in the SF1 data for a specific block group, we identified which of those prison types should be found in the block group. For example, if the data included males under 18 and females over 18 (but no males over 18 and no females under 18), we would expect to find at least two facilities in that block group (and possibly more if there are more than one of the same type). Depending on how the number of facilities in a block group compared to the number of prison types suggested by the SF1 data, we split the population among the facilities as shown in **Table 7**. Note that with the available data, it was not possible to identify specific facilities with specific types (male or female, adult or juvenile). Therefore, the assignments of the needed types of persons to the specific facilities in a block group was random.

Facility locations in HSIP appear to be based on address geocoding, which can be unreliable because it defaults to the ZIP Code centroid if the street address is insufficient to definitively identify the location. As a result, we encountered some mismatches between the Census prison population data and the HSIP prison facility data.

Of the 305 prisons in the HSIP data set for the study area, 52 were located in block groups for which the Census reports no group-quarters prison population. These could be closed facilities, but it's more likely that they reflect inaccuracies in the facility locations. These facilities were removed from the data set.

We also found 65 block groups with reported prison populations in the Census data that had no prison facilities in the HSIP data set. After reviewing these 65 block groups, we were able to locate associated facilities in adjoining block groups for 45. Only 9 of the remaining 20 block groups had prison populations over 100. In those cases, we placed a facility at the block group centroid.

Table 7. Rules for Allocating Population Among Prisons Within a Block Group

Number of Prison Types vs Facilities in Block Group	Assignment Rules																												
More prison types than facilities	Generate additional facilities located at the block group centroid as needed to assign population from each population type to its own facility.																												
Same number of prison types and facilities	Assign all population of a specific type (e.g., males over 18) to one (randomly selected) facility.																												
1 prison type and 2 facilities	Split population evenly between the two facilities.																												
Fewer prison types than facilities	<p>If a population type has less than 10 people, assign that population to one (randomly selected) facility. Split the remaining population type(s) between the remaining facilities as follows: split the population for the largest population type between two facilities; split the next largest population type between the next two facilities. Continue like this until the number of remaining population types matches the number of remaining facilities. For example: A block group with all four facility type populations as follows: Male <18 (94), Male ≥18 (263), Female <18 (31), Female ≥18 (178) would be distributed as follows:</p> <table border="1" data-bbox="500 1209 1286 1402"> <thead> <tr> <th>Facility</th> <th>Type</th> <th>Number</th> <th>Explanation</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Male ≥18</td> <td>132</td> <td>Largest population split</td> </tr> <tr> <td>2</td> <td>Male ≥18</td> <td>131</td> <td>between facilities 1 and 2</td> </tr> <tr> <td>3</td> <td>Female ≥18</td> <td>89</td> <td>Second largest population split</td> </tr> <tr> <td>4</td> <td>Female ≥18</td> <td>89</td> <td>between facilities 3 and 4</td> </tr> <tr> <td>5</td> <td>Male <18</td> <td>94</td> <td>Last two populations assigned</td> </tr> <tr> <td>6</td> <td>Female <18</td> <td>31</td> <td>to last two facilities</td> </tr> </tbody> </table>	Facility	Type	Number	Explanation	1	Male ≥18	132	Largest population split	2	Male ≥18	131	between facilities 1 and 2	3	Female ≥18	89	Second largest population split	4	Female ≥18	89	between facilities 3 and 4	5	Male <18	94	Last two populations assigned	6	Female <18	31	to last two facilities
Facility	Type	Number	Explanation																										
1	Male ≥18	132	Largest population split																										
2	Male ≥18	131	between facilities 1 and 2																										
3	Female ≥18	89	Second largest population split																										
4	Female ≥18	89	between facilities 3 and 4																										
5	Male <18	94	Last two populations assigned																										
6	Female <18	31	to last two facilities																										

Age Distribution

To refine the prison population age distribution, we used the age distribution reported by the U.S. Dept of Justice Bureau of Justice (U.S. DOJ, 2000a) for adults, shown in **Table 8**.

This distribution does not address juveniles (under 18). U.S DOJ (2000a) also lists 6,126 juvenile prisoners held as adults and 1,489 prisoners held as juveniles, or 7,615 total, but this age group is not broken out by gender. The percentage of men among prisoners over 18 decreases from age 18 until age 34 and then increases again after age 34. Therefore, we used the percentage at age 18 (95.42% males) to estimate the percentage of male and female juveniles, as that seemed likely to be the most representative. This results in 7,266 males and 349 females for the under 18 prisoners. As a QA step, we compared another U.S. DOJ source containing older data on juvenile prisoners (*Profile of State Prisoners under Age 18 1985–97*; U.S. DOJ, 2000b), which

lists the percentage of male prisoners under 18 as 97%. However since those data cover 12 years, all prior to 2000, we gave preference to the estimate based on the 2000 data.

Table 8. DOJ Age Distribution of Adult Prisoners (Number of Prisoners)

Age	Male	Female
18-19	81,300	3,900
20-24	310,100	19,600
25-29	329,900	30,000
30-34	334,000	39,100
35-39	294,100	30,700
40-44	198,300	17,000
45-54	164,500	12,100
>55	51,300	2,700
Total	1,763,500	155,100

The DOJ data lump all prisoners over 55 together, a rather broad range given that the age of the prisoners would have a large effect on the disease transmission rate. Therefore, we refined the distribution of prisoners 55 and over using the distribution of the general population over 55 from the Census (U.S. Census, 2000). Those data are summarized in **Table 9**.

Table 9. Age Distribution of the General Population Over Age 55

Age	Male (Number)	Male (Fraction)	Female (Number)	Female (Fraction)
55-59	6,508,729	0.2498	6,960,508	0.2096
60-64	5,136,627	0.1971	5,668,820	0.1707
65-69	4,400,362	0.1689	5,133,183	0.1546
70-74	3,902,912	0.1498	4,954,529	0.1492
75-79	3,044,456	0.1168	4,371,357	0.1316
80-84	1,834,897	0.0704	3,110,470	0.0937
85-89	876,501	0.0336	1,913,317	0.0576
≥90	350,497	0.0135	1,099,272	0.0331
Total	26,054,981	1.00	33,211,456	1.00

This spreading out of the over 55 data results in a small (less than 0.5 percent) changes in the total number of male prisoners due to rounding.

The final age distribution for prisoners, including juveniles and the expanded age distribution for prisoners over 55 is shown in **Table 10**. Where population was reported in age ranges (e.g., 18–19), we distributed the population in the age range equally among the component ages.

Table 10. Age Distribution Used for Prison Populations

Age	Males (#)	Males (%)	Females (#)	Females (%)
15	2,422	0.14	116	0.07
16	2,422	0.14	116	0.07
17	2,422	0.14	116	0.07
18	40,650	2.30	1,950	1.25
19	40,650	2.30	1,950	1.25
20	62,020	3.50	3,920	2.52
21	62,020	3.50	3,920	2.52
22	62,020	3.50	3,920	2.52
23	62,020	3.50	3,920	2.52
24	62,020	3.50	3,920	2.52
25	65,980	3.73	6,000	3.86
26	65,980	3.73	6,000	3.86
27	65,980	3.73	6,000	3.86
28	65,980	3.73	6,000	3.86
29	65,980	3.73	6,000	3.86
30	66,800	3.77	7,820	5.03
31	66,800	3.77	7,820	5.03
32	66,800	3.77	7,820	5.03
33	66,800	3.77	7,820	5.03
34	66,800	3.77	7,820	5.03
35	58,820	3.32	6,140	3.95
36	58,820	3.32	6,140	3.95
37	58,820	3.32	6,140	3.95
38	58,820	3.32	6,140	3.95
39	58,820	3.32	6,140	3.95
40	39,660	2.24	3,400	2.19
41	39,660	2.24	3,400	2.19
42	39,660	2.24	3,400	2.19
43	39,660	2.24	3,400	2.19
44	39,660	2.24	3,400	2.19
45	16,450	0.93	1,210	0.78
46	16,450	0.93	1,210	0.78
47	16,450	0.93	1,210	0.78
48	16,450	0.93	1,210	0.78
49	16,450	0.93	1,210	0.78
50	16,450	0.93	1,210	0.78
51	16,450	0.93	1,210	0.78
52	16,450	0.93	1,210	0.78
53	16,450	0.93	1,210	0.78

Age	Males (#)	Males (%)	Females (#)	Females (%)
54	16,450	0.93	1,210	0.78
55	2,563	0.14	113	0.07
56	2,563	0.14	113	0.07
57	2,563	0.14	113	0.07
58	2,563	0.14	113	0.07
59	2,563	0.14	113	0.07
60	2,022	0.11	92	0.06
61	2,022	0.11	92	0.06
62	2,022	0.11	92	0.06
63	2,022	0.11	92	0.06
64	2,022	0.11	92	0.06
65	1,733	0.10	83	0.05
66	1,733	0.10	83	0.05
67	1,733	0.10	83	0.05
68	1,733	0.10	83	0.05
69	1,733	0.10	83	0.05
70	1,537	0.09	81	0.05
71	1,537	0.09	81	0.05
72	1,537	0.09	81	0.05
73	1,537	0.09	81	0.05
74	1,537	0.09	81	0.05
75	1,198	0.07	71	0.05
76	1,198	0.07	71	0.05
77	1,198	0.07	71	0.05
78	1,198	0.07	71	0.05
79	1,198	0.07	71	0.05
80	722	0.04	51	0.03
81	722	0.04	51	0.03
82	722	0.04	51	0.03
83	722	0.04	51	0.03
84	722	0.04	51	0.03
85	345	0.02	31	0.02
86	345	0.02	31	0.02
87	345	0.02	31	0.02
88	345	0.02	31	0.02
89	345	0.02	31	0.02
90	692	0.04	89	0.06
Total	1,770,758	100.01	155,447	99.97

College Dormitories

Location and Number of Residents

We used the HSIP data to locate colleges and universities. In cases where the HSIP indicated there was a college but we had no college-dormitory group-quarters population from the SF1 data, we eliminated the college point from our data set. These were either non-resident colleges or missing or incorrect data points in the HSIP. In cases where the Census SF1 data showed group-quarters college population but we had no location identified by HSIP, we generated a point at the block group centroid.

Age Distribution

The SF1 data set gives the number of students in college dormitories (on and off campus) for each block group. We were not able to locate an external (non-Census) sources to refine the age distribution, so we used additional data from the SF1 and PUMS.

In addition to the data on college students living in group quarters, the SF1 data set also includes data on students aged 15 and over enrolled in college, regardless of whether they live in households or group quarters, broken out into smaller age ranges than the SF1 group-quarters data (15–17, 18–24, 25–34, 25–65, over 65).

To split these into students living in group quarters and not living in group quarters, we used data from PUMS to determine the number of students in college living in households by gender and age.³ (As noted before, the PUMS data are a sample, and weights are not available to weight up the group-quarters populations from PUMS data, only the household data). Then, for each more detailed age range, we subtracted the number of students living in households from PUMS from the total number of college students (regardless of living arrangement) from SF1 to estimate the number of students living in group quarters by age. **Table 11** shows the results of those calculations. The resulting age distribution of students who live in college group quarters shows that most of the students living in group quarters are in the 18–24 age group, as expected.

Table 11. Calculation of College Dormitory Population by More Detailed Age Ranges and Gender

Age ^a	Males			Females		
	All Students (SF1)	Household Students (PUMS)	Group-Quarters Students (Derived)	All Students (SF1)	Household Students (PUMS)	Group-Quarters Students (Derived)
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+	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

^a The SF1 data report only 105 college students over 65 (48 males and 57 females) in the entire United States; therefore, we combined that age category with 35–64 for this calculation. The effect should be negligible, given the relative sizes of the populations.

³ We counted as college students 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) of 7 (graduate or professional school).

To test the assumption that we could combine the enumerated SF1 data set and the data derived from the PUMS sample and associated weights as described, we looked at all college students (regardless of age or gender), took the total college students in households from PUMS, added the total college students in group quarters from SF1, and compared that total to the total of all college students from SF1. The two totals compared favorably, with a difference was only 2.3%, suggesting 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, we distributed the population in each age range equally among the component ages. For the 65+ age group, we assigned one person to each age beginning at 65 within each block group.

Table 12. Age Distribution for College Dormitories

Age	Males (#)	Males (%)	Female (#)	Female (%)
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
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
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	1,194,309	100.15	177,453	100.13

Nursing Homes

Location and Number of Residents

We used two data sources to locate nursing homes and allocate population from SF1 to the nursing homes. The Nursing Home Atlas (AHRQ, 2007) contains both location information and information on bed capacity for nursing homes. However, the Nursing Home Atlas data include only facilities receiving Medicaid, not all facilities, and the difference is clearly significant: the total number of beds reported in the Nursing Home Atlas for North Carolina for 2005 is only 58 percent of the total nursing home population reported in SF1 for 2000 for North Carolina; similarly, the total number of beds reported for South Carolina is only 63 percent of the SF1 population. In addition, the Nursing Home Atlas data had 405 facilities for North and South Carolina, while SF1 reported nursing home residents in 815 separate block groups. Thus, the Nursing Home Atlas data are not adequate to reliably locate nursing homes and ensure that the points are in the correct block groups.

The HSIP data on nursing homes includes a larger, more geographically complete set. However, the HSIP data do not include facility size. Therefore, we used the more geographically complete HSIP data to establish locations. We used the data in the Nursing Home Atlas to calculate minimum, average, and maximum facility sizes for nursing homes by state, and used those to allocate population when we had a block group with nursing home population from SF1 but no facility from HSIP. For each such block group, if the total nursing home population was less than the maximum nursing home size for the state, we created one point for the entire population at the block group centroid. If the population was greater than the maximum nursing home size, we divided the total nursing home population for the block group by the average nursing home size for the state to determine the appropriate number of facilities, and placed those at the block group centroid. We placed population equal to the average facility size at all but one of these, and placed any remaining population at the last facility.

Age Distribution

The age distribution assignments for nursing homes have not yet been completed.

RESULTS AND CONCLUSION

The MIDAS North Carolina Metro Study is an example of an application that incorporates the synthesized population and associated data sets into a model for the spread of infectious disease. One of the goals of the North Carolina Metro Study is to create an experimental setting and tools to explore relevant issues (e.g., do prisons influence and sustain flu transmission?).

The North Carolina Metro Study has been using just the synthetic household population; however, the synthesized group-quarters population described in this paper has now been incorporated into the Study. Initial simulation runs using the group-quarters data combined with the synthesized household data have been done using transmission rules analogous to the household and workplace transmission rules to simulate contact and determine if transmission occurs. If more information becomes available on how specific group quarters types interact (e.g.; inmates in a prison), it can be incorporated into the rules and tested. Until then, initial runs indicate the model runs appropriately and promise interesting results.

REFERENCES

- AHRQ (Agency for Healthcare Research and Quality). 2007. *Emergency Preparedness Atlas: U.S. Nursing Home and Hospital Facilities*. AHRQ Publication No. 07-0029-2, April 2007. Agency for Healthcare Research and Quality, Rockville, MD. Available at <http://www.ahrq.gov/prep/nursinghomes/atlas/>
- U.S. Census Bureau. 2000. *QT-P1. Age Groups and Sex: 2000; Data Set: Census 2000 Summary File 1 (SF 1) 100-Percent Data, Geographic Area: United States*. Available at <http://factfinder.census.gov/servlet/DatasetMainPageServlet>. Accessed June 2007.
- U.S. Census Bureau. 2005a. *2000 Census of Population and Housing, Public Use Microdata Sample 2000*. PUMS/15-US (RV). Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. Issued December 2005. Available at http://ftp2.census.gov/census_2000/datasets/PUMS/FivePercent. Accessed August 12, 2005. Documentation available at <http://www.census.gov/prod/cen2000/doc/pums.pdf>.
- U.S. Census Bureau. 2005b. *2000 Census of Population and Housing, Summary File 1*. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. Issued March 2005 SF1/13 (RV). Available at <http://www.census.gov/Press-Release/www/2001/sumfile1.html>. Accessed July 1, 2005. Documentation available at <http://www.census.gov/Press-Release/www/2001/sumfile1.html>.
- U.S. DOD (Department of Defense). 2005. *Selected Manpower Statistics for Fiscal Year 2005: Defense Manpower Data Center Report*. Tables 2-17 for Year 2000 and 2-17A for year 2001, Age Distribution of Male/Female Military Personnel Strength (in Thousands) and Percent Figures. Available at <http://siadapp.dmdc.osd.mil/personnel/M01/fy05/m01fy05.pdf>.
- U.S. DOJ (Department of Justice). 2000a. *Prison and Jail Inmates at Midyear 2000*. DOJ Statistics Bulletin. Table 12—Number of Inmates in State or Federal Prisons or Local Jails by Gender, Race Hispanic Origin, and Age. June 30, 2000. Available at <http://www.ojp.usdoj.gov/bjs/pub/pdf/pjim00.pdf>.
- U.S. DOJ (Department of Justice). 2000b. *Profile of State Prisoners under Age 18 1985–97*. Bureau of Justice Statistics Special Report, NCJ 176989. U.S. Department of Justice, Office of Justice Programs. February. Available at <http://www.ojp.usdoj.gov/bjs/pub/pdf/pspa1897.pdf>.
- Wheaton, B. and B. Chasteen. 2005. *Generating Synthetic Populations Using Public Use Microdata (PUMS)*. Presented at the ESRI Health GIS Conference 2005, Chicago, IL, October 23–26.

AUTHOR BIOGRAPHY

Ms. Chasteen works for RTI International and has 17 years of experience developing GIS applications and databases for modeling the spread of infection diseases, for health-related survey project management and assessment, and for environmental assessment and protection. She is developing a nationwide synthetic population for use by modelers on MIDAS.