International Experience in Journey-to-Work Data from National Censuses

Ram M. Pendyala
Amlan Banerjee
Dept of Civil and Environmental Engineering
University of South Florida, Tampa

TRB Conference on Census Data for Transportation Planning
May 2005

Outline

- Changing U.S. census
  - American Community Survey (ACS)
  - Impact on CTPP
- Review of international experience (focus on journey-to-work data)
  - Australia and New Zealand
  - Canada
  - France
  - Germany
  - The Netherlands
  - The United Kingdom
- Findings and conclusions

U.S. Census 2000: Overview

- 22nd Census in U.S. decennial census history conducted on April 1, 2000
- Counted 281 million people and 115.9 million households
- Tabulated data prepared for 9 million census blocks
- Questionnaire format
  - Short form: Household and member demographic characteristics
  - Long form: Detailed socio-economic and journey-to-work characteristics
- 1/6th of households receive long form

Changing U.S. Census

- Issues with traditional decennial census format
  - Rapidly changing community characteristics – long form data obsolete within few years
  - Large expense every 10 years
- Goals for future U.S. census (2010 and beyond)
  - Provide timely and relevant data cost-effectively
  - Improve coverage
- Solution: Continuous Measurement Approach

American Community Survey (ACS)

- Continuous survey approach
  - Annual and multi-year estimates of population characteristics
  - Small area characteristics updated every year
- Annual national sample of about 3 million addresses (250,000 addresses per month)
  - Approx equivalent to 2.5% sampling rate per year
- Full implementation initiated in 2005
  - Annual estimates for communities of 65,000 or more
  - 3 year cumulations for communities of 20,000-65,000
  - 5 year cumulations for communities of <20,000

Features of ACS

- Differences with traditional decennial census (TDC)
  - Five year sample fraction: ~12.5% ACS to ~17% TDC
  - TDC estimates based on ~18 million housing units; ACS 5-year estimates based on ~11 million housing units
  - ACS samples every year and spreads sample over 12 mo
  - ACS subsamples for personal visit follow-up
- ACS estimates have higher sampling error
- Preliminary indications: ACS estimates have lower potential non-sampling error (non-response)

Census Transportation Planning Package

- Three sets of standard tabulations make up CTPP
  - Part 1: Residence based tabulations
  - Part 2: Work-place based tabulations
  - Part 3: Residence – Work (journey-to-work) flows
- Used extensively in transportation planning
  - Develop zonal socio-economic and demographic data
  - Analyze socio-economic and demographic characteristics
  - Validate travel demand models using flow tables
- Census 2000 CTPP subjected to disclosure avoidance procedures and rules
  - Rounding and Thresholds

Disclosure Avoidance Rules for CTPP 2000

- Part 1: Residence based tables
  - All tables rounded
    - Zero = 0; 1 through 7 = 4; 8 and above = nearest multiple of 5
- Part 2: Work-place based tables
  - All tables rounded (same rules)
- Part 3: Worker flows
  - All tables rounded (same rules)
  - Some tables with thresholds
    - Any cell with 3 or less records (flows) is suppressed
- Christopher and Srinivasan (2005) discuss adverse implications of these procedures on CTPP

Disclosure Avoidance for PUMS Data

- PUMS data is most disaggregate data from census
  - Individual records: 5% state files and 1% national file
  - Detailed individual records useful for constructing joint distributions needed for synthetic population generation
  - Increasing importance in context of activity-based microsimulation models
- Disclosure avoidance methods:
  - Data swapping: edit data or exchange records
  - Top-coding: Grouping cases above a certain value
  - Geographic population thresholds
  - Age perturbation in large households
  - Collapsing categories that do not meet a threshold

Issues and Challenges

- ACS format has important implications for CTPP
  - Smaller sampling rates and larger sampling error
  - Geographic resolution for reporting data
  - Work place geocoding errors and allocation inaccuracies
  - Implications of rounding and thresholds – many worker flows suppressed
- What are other countries doing and what is their experience in resolving these issues?
  - Identify methods, techniques, lessons, etc.

Geographic Resolution

U.S.	**U.K.**	**Canada**	**Australia/ New Zealand**	**France**
Nation	Nation	Nation	Nation	Nation
State	Region	Province	State	Province
County	County	Census Division	Statistical Div	Metropolitan Area
		Census Sub Div		Urban Unit
Place (City)	District	Cen.Agglomeration	Statistical Sub Div	
Census Tract		Census Tract		Municipality/ Commune
Block-Group	Ward	Dissemination Area	Stat. Local Area	
TAZ	Output Area	Block	Cen Collection Dist	
Block	Post Code	Block-face	Mesh Blocks	

Australia: General Information

- New Zealand very similar to Australia
- Australian Bureau of Statistics (ABS) conducts census once every five years (2001-2006)
- Journey-to-work data used extensively by state transport authorities
- Respondents provide work place address; two-stage geocoding process
  - Based on respondent record
  - Based on facility/business name index
  - Geocoded to DZN (workplace destination zone)
  - Elaborate work place geocoding procedure

Australia: Reporting Geography

- Until 2001, SLA (Statistical Local Area) was smallest geography at which data was reported
  - SLA is aggregation of DZN
- In 2006, census data reported for new smaller geography called Mesh Block (20-50 households)
  - More homogeneous geographic units
- Developed G-NAF (Geocoded National Address File) in 2004 and updated quarterly
  - Primary source of geocoding in 2006 and beyond
  - Extremely accurate – multi-agency collaborative effort

Australia: Disclosure Procedures

- Confidentiality of tabular data maintained
- Assessing size of table
  - Compare number of cells to total population in table; if difference is small, table is suppressed
- Introducing random error
  - Randomly adjust cell values with small values; detailed methodology not released
  - Tables are internally consistent
  - Value of tables as a whole not impaired
  - Allows releasing tables with small cell values

Canada: General Information

- Statistics Canada conducts census once every five years (2001-2006)
- Questionnaire format:
  - Short form: 80% of households
  - Long form: 20% of households
- Long form includes all short form questions plus 52 additional questions
- JTW questions asked for all persons 15 years or older who worked any time since Jan 1, 2000

Canada: JTW Data Details

- Information collected
  - Work status, employer address, nearest landmark/street intersection (if address unknown), mode to work
- Typical two-step work place geocoding procedure
  - Automated system (computerized)
  - Computer-assisted clerical coding
  - Uses National Geographic Base as reference file
- Systematic 3-step imputation technique for missing JTW data
  - Canadian Census Edit and Imputation System (CANCEIS) to impute JTW variables
  - Additional modules to impute work place location

Canada: Disclosure Procedures

- Confidentiality of tabular data maintained
- Data suppression based on population living or working in an area
  - Standard areas: Threshold = 40 (weighted)
  - User-defined areas: Threshold = 100 (weighted)
  - All areas: Threshold = 250 (weighted) if income included
  - Rounding to the nearest 5 except for counts below 10 (rounded to zero or 10)
- No formal CTPP, but similar tabulations produced for provinces and municipal governments

France: General Information

- French Rolling Census closely parallels ACS concept
  - Last traditional census in 1999
- Goals of French Rolling Census
  - To spread burden over a longer period
  - Meet demand for more timely and fresh data
  - Improve data quality by exploiting technical advances
- Budget allocation: 1/7th of traditional census budget each year
  - Implies a 1/7th sampling rate each year (~14%)

France: Sampling Strategy

- Key geographic unit is "commune" (37,000 total communes)
  - Large and small communes defined by population of 10,000
  - Total population equally split between large and small communes
  - Small communes visited once every 5 years (sampled at rate of 20 percent)
  - Large communes visited every year (sampled at rate of 8%)
  - Total sampling rate = 20% x 50% + 8% x 50% = 14%

France: Sampling Strategy

- Small commune: Five rotating groups
  - Rotating samples of communes over a 5 year period
  - ~ 30 million inh × 1/5 × 100% = 6 Million per year
- Large commune: Five rotating groups
  - Based on a building register
  - 40% households drawn from each group every year
  - 8% drawn/yr → 40% of all households in 5 years
  - ~ 30 million inh × 1/5 × 40% = 2.4 Million per year
- Total: 8.4 M per year or 60 M in 7 years

France: Data Reporting

- Data collection methodology
  - Collect information over a five year period cycle
  - Produce every year statistically reliable/significant data for the medium year
- Let current year = Y
  - Produce statistically reliable data for year "Y-2" using data from years "Y-4", "Y-3", "Y-2", "Y-1", and "Y"
- No special information about journey-to-work or work place based data
- Smallest geographical resolution of published data not clear

France: Rolling Census

- Merits
  - Timely data that is maximum of 3 years old
  - More detailed data at same expenditure
  - Improved quality of data even in large communes
  - Updated sampling base of households
- Issues
  - Quality of building register
  - Precision of estimates for small geography(?)

Germany: General Information

- Last traditional census in 1987
- New German census is combination of administrative registers and survey data
  - Population registers
  - Employee registers
  - Housing census (postal survey)
  - Sample survey
- Test surveys conducted to test effectiveness of new system
  - Check accuracy of population register
  - Check for duplicate entries in population register

Germany: JTW Data and Disclosure

- Some journey-to-work questions included in census:
  - Name and address of work and school location
  - Means of transport to work or school
  - Travel time to work or school
- Disclosure protection
  - All personal and identifiable information deleted
  - Data published/released only for "parts of municipalities"
  - Some individual data (excluding names and addresses) may be transmitted to municipal governments only

Germany: New Microcensus

- Microcensus after 1987 conducted every year on 1% of all households in Germany
  - 370,000 households (820,000 persons)
- All households have same probability of selection
- One-stage stratified area sampling scheme
  - Sampled areas are sampling districts
  - Every year, 1/4th of households are rotated off; every household stays in sample for four years
- Several programs
  - Annual Program: Person and household characteristics
  - Annual Supplement: Employment and training
  - Four-year Additional Program: Commuting, housing, health

The Netherlands: General Information

- Dutch census in 2001 is integration of microdata from registers and surveys
- Registers
  - Population register
  - Job files
  - Fiscal administration
  - Social security administration
- Surveys
  - Employment and earnings survey
  - Labor force survey
- Innovative data linkage and integration strategies

The Netherlands: JTW Data

- Household members asked to report trips for one day
- Origin and destination address information collected
- Workplace address information extracted from trip survey records
- Missing trips imputed; follow-up with respondents where possible

The Netherlands: Confidentiality

- Published tables subjected to confidentiality protection rules
  - Table cells with less than 10 persons always suppressed
  - Table cells with 25 or more persons always published
  - Table cells with 10-24 persons published only if they form part of a cross-classification (e.g., age by sex) in which no cells contain less than 10 entries
  - Also, 50% of cells in cross-classification should have 25 or more persons
  - Threshold of 25 persons corresponds to an estimated relative inaccuracy of at most 20 percent

U.K.: General Information

- U.K. Office of National Statistics conducts decennial census in U.K. and Wales
  - Other agencies for other parts of U.K.
- Last census in 2001
  - Single census form delivered to all households
- Journey-to-work questions asked of all persons aged 16-74 years
- Census JTW questions:
  - Home address one year ago
  - Commuting destination
  - Means of travel to work or study

U.K.: JTW Data Tables

- Work place data in Census 2001
  - Standard tables and theme tables published down to the "Ward" level contain a range of JTW data tables
  - Census Area Statistics tables based on daytime work place population; less information but finer level of geography
  - Census Area Statistics published for output area (~125 households)
  - Special Workplace Statistics (SWS) tables include employment and JTW information down to "Ward" level
- Workplace data capture and coding involved multi-step process to assign work locations to post codes
- Samples of Anonymized Records: 3% of persons and 1% of households

U.K.: Imputation Procedure

- Elaborate imputation procedures applied to three data sets
  - Migrant origin, workplace and study address
- Methodology based on donor imputation of postcodes
- Identify the optimum combination of variables on which a potential donor matches an intended recipient
- Technique maximizes the accuracy of the imputation
- Preserves joint and marginal distribution of the data

U.K.: Disclosure Control

- Small cell adjustment
  - Small counts randomly adjusted
  - Totals and subtotals calculated based on adjusted data
  - Tables independently adjusted; counts of same population in two different tables may not be same
  - Tables of higher geographic levels not necessarily sum of tables of lower geographic levels
- Record swapping
- Thresholds
  - Standard tables: At least 1000 residents and 400 households
  - Census Area Statistics tables: At least 100 residents and 40 households
  - Summary Profiles: At least 50 residents and 20 households
- Design of Table
  - Average cell count in a table greater than or equal to one

U.K.: O-D Flow Data Disclosure

- O-D table cells with small counts adjusted using disclosure control techniques
- Count Adjustments
  - Cells with small values adjusted independently upwards or downwards based on prescribed probabilities
  - Does not introduce systematic biases into the count
  - More cells adjusted, larger variation from the true values
  - Other sources of variation: coverage error, respondent error, processing error, record swapping
- Rounding
  - Small cell values rounded to multiples of 3
- Suppression of data on industry at the ward level and below
  - Problem in using data for trip attraction analysis

Conclusions

- Moving away from traditional decennial Census format
- Common goals for this transition
  - Cost
  - Timeliness and quality of census data
- Methodological difference in new Censuses
  - Administrative registers + survey-based
  - Continuous measurement or rolling census approach
  - Mid-decade census
- Common Issues
  - Data dissemination, accuracy, and disclosure control

Conclusions

- Workplace geocoding
  - Accuracy of workplace geocoding of major concern
  - Australia uses separate zonal structures for residence and workplace – capture most O-D flows
  - At least a two-stage process: automated followed by more manual geocoding procedures
  - Development of nationwide geocoding reference address file
    - TIGER (U.S.)
    - G-NAF (Australia)
    - National Geographic Base (Canada)

Conclusions

- Disclosure avoidance techniques
  - Rounding small cell values to multiples of 3 (U.K., Australia, and New Zealand)
  - Data swapping commonly applied to microrecords in U.S. and U.K.
  - Use of thresholds applied to both tabular data and release of data for small geographical units
  - Random data perturbation applied in U.K. and Australia; allows release of tables with small cell values
- Accuracy
  - France also using five year cumulations for small geographies, but with larger sampling rates