

Quality Assessment of Data Collected from Non-English Speaking Households in the American Community Survey

Pamela D. McGovern and Deborah H. Griffin
U.S. Census Bureau

2003 Joint Statistical Meetings, San Francisco, CA
August 3, 2003

Overview

- Background
- Terminology
- Research Objectives
- Methodology
- Results
- Conclusions and Other Research

Background

Based on:

- National tests of operational feasibility of American Community Survey (ACS) in 2000 and 2001.
- Annual sample of 700,000 addresses that collected detailed demographic, socioeconomic, and housing data about the nation.

Background (cont.)

- Housing data collected:
Physical characteristics, utilities, mortgages, food stamps, taxes, etc.
- Population data collected:
Sex, age, date of birth, race, marital status, Hispanic origin, education, income, labor force status, etc.

Background (cont.)

Three mode data collection operation:

- Mail
- Telephone
- Personal Visit

Background (cont.)

Mail:

- Paper questionnaire available only in English.
- Spanish language assistance.
- Telephone follow-up if necessary.

Background (cont.)

Telephone and Personal Visit:

- Automated survey instrument.
- Available in English and Spanish.
- Bilingual field representatives if needed.

Terminology

- Household Language Spoken
- Linguistically Isolated (LI) Households

Household Language Spoken

- Assigned when one or more people age 5+ speak a language other than English.
- Specific language assigned to household based on data collected during interview.
- Variable created at the household level.
- Further divided into LI and non-LI.

Linguistically Isolated Households (LI)

- No household member age 14 years or over reports speaking English “very well”.
- Collected from questionnaire.
- Variable created at the household level.
- Research focuses on these households since face greatest challenges in survey response.

Research Objectives

To determine:

- Which languages have greatest numbers of linguistically isolated households in the ACS?
- How were linguistically isolated and non-linguistically isolated households interviewed in the ACS?
- How complete are the ACS data collected from linguistically isolated and non-linguistically isolated households?

Methodology – Objective 1

Produced 2-year average estimates of total occupied households by:

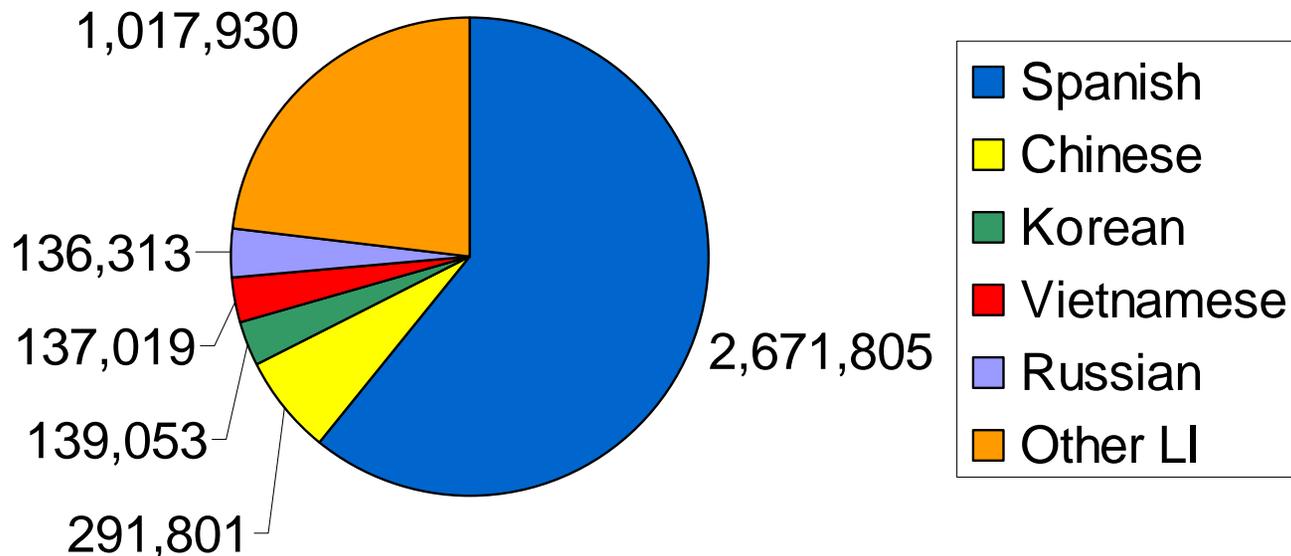
- language spoken (40 language groups)
- linguistic isolation

Total Number of Households Speaking Language at Home

Household Language	# Speaking Lang.	Percent
All	105,623,930	
English	86,655,932	82.0
Spanish	10,375,325	9.8
Chinese	798,276	0.8
Korean	384,168	0.4
Vietnamese	318,074	0.3
Russian	316,151	0.3

Which languages have the greatest numbers of linguistically isolated households?

Estimated Total Number of LI = 4,393,921



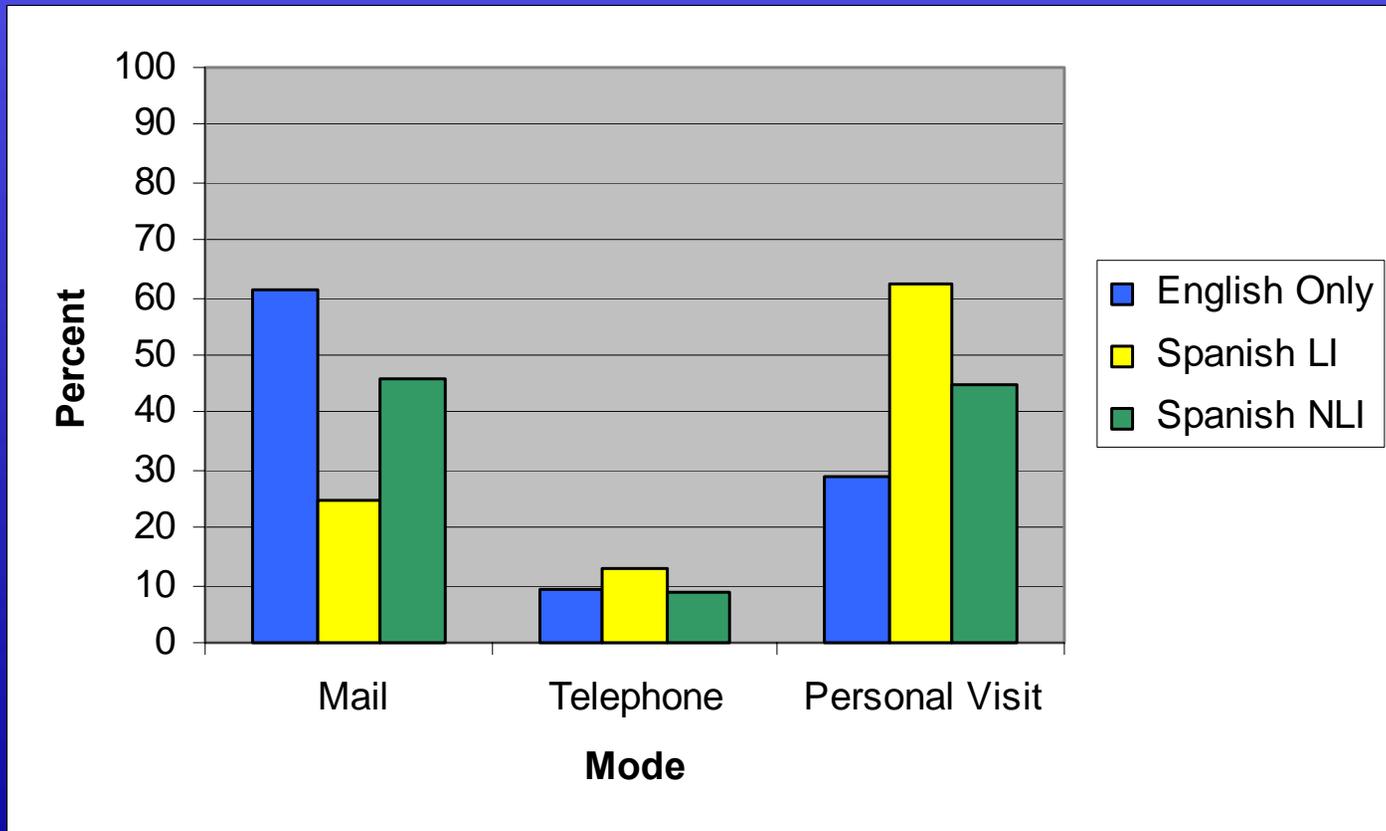
Methodology – Objective 2

- Used operational data on mode (mail, telephone, personal visit), household language, and linguistic isolation.
- Produced distribution of interviews across modes for various language groups.

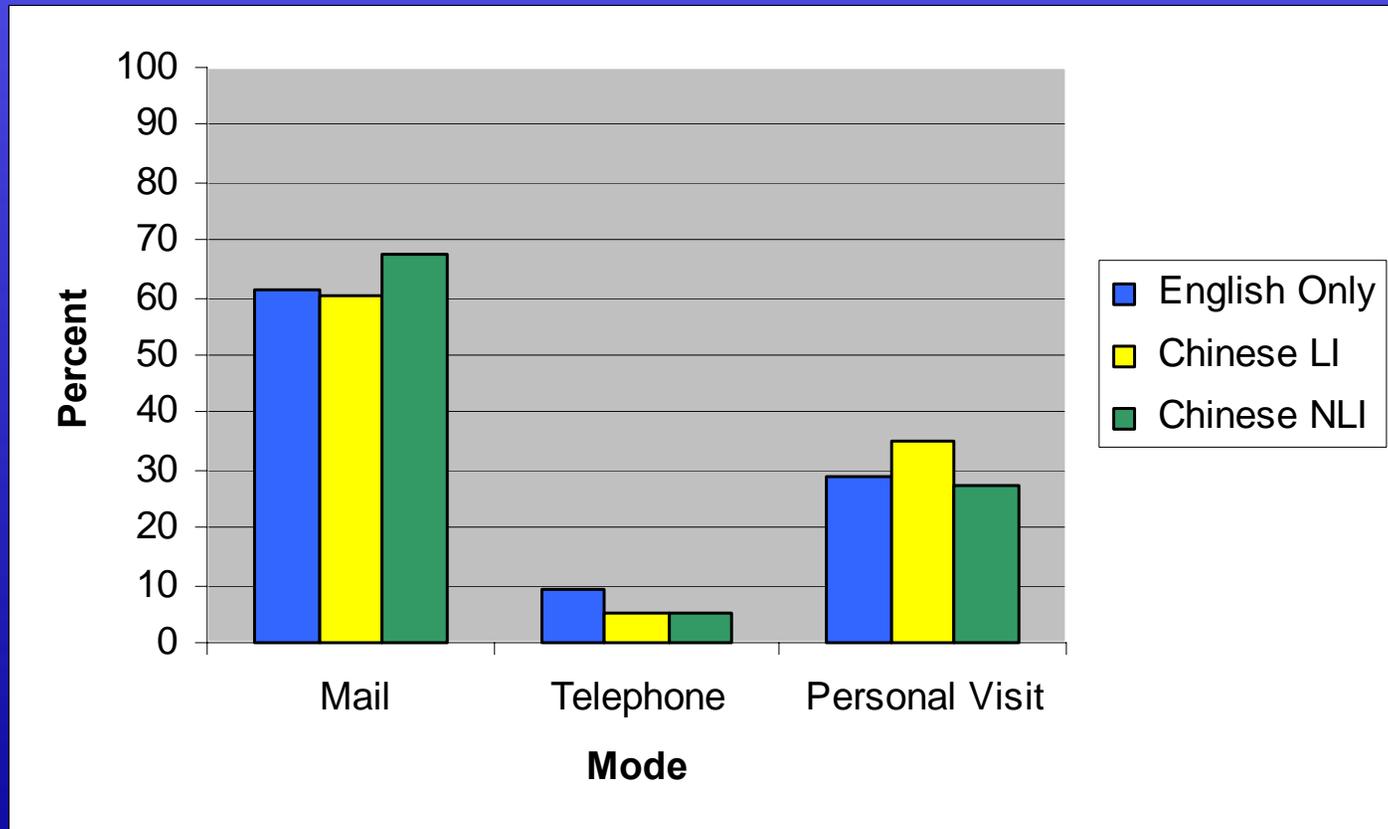
How were LI and non-LI households interviewed?

- Linguistically isolated households of all language groups have lower percentages of response by mail (relative to English only).
- Mixed results of non-linguistically isolated households (relative to English only).

How were Spanish LI and non-LI households interviewed?



How were Chinese LI and non-LI households interviewed?



Methodology – Objective 3

- Measured completeness using item nonresponse.
- Missing data compensated in allocation procedures.
- Allocation rates: measure of final data quality that quantify how frequently allocation was the source of data.

Methodology - Objective 3 (cont.)

- Produced combined allocation rates across all population and housing items.
 - by mode
 - by language group
 - two-year averages
- Ratio of the total # of population (housing) items for which a value was allocated to the total # of population (housing) items for which response was required.

How complete were population data collected by mail?

Household Lang.	Allocation Rates (%)			
	LI	diff.	Non-LI	diff.
English Only	6.4			
Spanish	11.7*	5.3	9.0*	2.6
Chinese	7.5*	1.1	7.2*	0.8
Korean	9.1*	2.7	7.7*	1.3
Vietnamese	9.5*	3.1	11.2*	4.8
Russian	9.5*	3.1	7.6*	1.2

* Significantly different from English Only at the $\alpha=.10$ level.

How complete were population data collected by personal visit?

Household Lang.	Allocation Rates (%)			
	LI	diff.	Non-LI	diff.
English Only	4.8			
Spanish	4.1*	-0.7	4.1*	-0.7
Chinese	7.4*	2.6	7.7*	2.9
Korean	7.1*	2.3	6.4*	1.6
Vietnamese	4.8	0.0	5.7	0.9
Russian	4.4	-0.4	5.3	0.5

* Significantly different from English Only at the $\alpha=.10$ level.

How complete were the population data across all modes?

Household Lang.	Allocation Rates (%)			
	LI	diff.	Non-LI	diff.
English Only	5.7			
Spanish	5.5	-0.2	6.5*	0.8
Chinese	7.3*	1.6	7.3*	1.6
Korean	7.9 *	2.2	7.1*	1.4
Vietnamese	7.3 *	1.6	9.0*	3.3
Russian	7.0*	1.3	6.4*	0.7

* Significantly different from English Only at the $\alpha=.10$ level.

Conclusions

- Spanish largest non-English language group in the U.S.
- ACS interviews more linguistically isolated households by personal visit.
- ACS successful in obtaining complete data from linguistically isolated households with three modes of data collection.

Other Research

- Testing and refining a Spanish questionnaire.
- Conducting qualitative quality assessments.
- Reviewing item-level rates to identify problem questions for linguistically isolated households.
- Continuing to monitor rates using 2002 data and beyond.

Contact Information

Pamela McGovern:

E-mail: Pamela.D.McGovern@census.gov

(301)763-5951