Evolution of the Modern Post-Enumeration Survey: How did we get here and where should we go next?

Howard Hogan, Ph.D.

Chief Demographer

Outline

- Evolution in coverage measurement
- Evolution in coverage measurement goals
- Where to next?

Any views expressed on the statistical and methodological issues in this presentation are those of the author and not necessarily those of the U.S. Census Bureau.

.

Outline

- **Evolution in coverage measurement**
- Evolution in coverage measurement goals
- Where to next?

The foreign phenomenon of deliberately lying to the census enumerator and of hiding individuals or whole families is almost totally absent, even in the slums of our large cities.

William Lane Austin,

Director of the Census

March 22, 1939

Getting Started

- Early cohort analysis (e.g. Young 1901)
- Comparisons with Draft Registration
- 1949 Chandrasekaran & Deming
- 1950 PES
- 1955 Demographic Analysis
- 1960 Further evolution of DA
  - Good estimates for the Black population

Early Evolution of PES

- 1950 “Do it again but better”
  - Assumption was that most errors were do to random interviewer errors
- 1960 Try everything
  - Reverse Record check
  - Housing unit check
- 1970 CPS-Census Match
  - Complete failure
- First widely accepted estimates of Black/Non-black differential from DA

1980

- P Sample: April & August CPS Match
- E Sample: Separate sample
- Motivation: Strong evidence of a racial bias in the census based on DA
- Done in shadow of adjustment lawsuits
- Important result: First serious evaluation of the evaluation
- Led to serious funding for 1990 PES Research

Experiments & Research

- Pre-1980
  - PES A vs B; List vs Area
  - Triple system: CPS-IRS-Census
- Post-1980
  - IRS-Census Match
  - RRC Research
    - Forward Trace Study
    - CPS-Census Retrospective Match and Trace

1990 PES

- Computer and computer assisted matching
- Over-lapping P & E Samples
- Still pencil & paper interviewing
- “PES – B”
- Universe: Household and non-institutional GQ.

Errors in the PES

- Matching error
- Reporting census-day address
- Fabrication in the PES interview
- Missing data
- Error in measuring erroneous enumerations
- Balancing gross overcounts and undercounts
- Correlation bias
- Random error

See Wolter & Hogan 1988

Matching error

- Great progress on precision
- Underlying accuracy limited by quality of census enumerations.
- Strict application of “sufficient information has helped.
  - However, excluding more enumerations risks further correlation bias.
- At the margin, there is still level of professional judgement hard to quantify.

Reporting census-day address

- Increased probes have helped some, but there is a limit.
  - Suggestion: Use CAPI for targeted probes
- Nation wide duplicate search has helped a lot.
- Restrictions on specific follow-up probes has made the problem worse.

Fabrication in the PES interview

- Greatly reduced due to CAPI.

Missing data

- Greatly reduced due to CAPI
- Still dependent on cooperation of public.

Error in measuring erroneous enumerations

- Nation wide duplicate search has helped a lot.
- Proving an enumeration as fictitious still hard.
- For Components analysis: Still no consensus on what defines an erroneous enumeration.

Balancing gross overcounts and undercounts

- Overlapping P & E samples greatly reduced this problem
- TIGER and other improvements in Census geography reduced it further.

Correlation bias

- Sex ratio adjustment reduces this for some adult black males
  - Increased more than one race
  - Increased percent of blacks foreign born.
- No adjustment (yet) for Hispanics & young children
- Could get worse in the future.

Random error

- Probably at the limit of balancing large sample sizes with well-trained and supervised field and office staffs.
- Adequate at the state level and for important groups.
- However, most users would like information about their city.
- Increased modelling might help
  (cf. UK approach)

Outline

- Evolution in coverage measurement
- **Evolution in coverage measurement goals**
- Where to next?

Evaluation Goals

- Inform users
- Help plan the next census
- Adjust/correct the current census

PES Specific Goals

- Estimates for specific demographic groups
- State and local estimates
- Components of error

Inform users

Limited evidence of census data users explicitly factoring in coverage error into their analysis.

However, perhaps this is because the evaluations show errors are small enough to ignore.

Estimates for specific demographic groups

Success!

Until recently, provided the only estimate of coverage of Hispanics

Still only estimate for AIAN*, Asian, NHPI

* Except Remote Alaska

State and local estimates

Mixed

Good estimates for states

Synthetic estimates for local of limited value

Components of error

Success

- Able to demonstrate the off-setting nature of large errors to produce small net error.
- Able to demonstrate role of duplication in census process.
- Still, nothing on the non-household population

Evaluation Goals

- Inform users
- **Help plan the next census**
- Adjust/correct the current census

Help plan the next census

- Coverage evaluation has played a big role in, for example, motivating increased Decennial budget
- Credit must be shared with Demographic Analysis estimates
- Hard to document connection between PES components of error analysis and modified census procedures
  - E.g. Level of Census Duplication

Adjust/correct the census results

- USA: Very limited success
  - 1970s adjustments (not strictly PES)
  - Use in 1990s CPS controls
  - Decisions predicated on relative accuracy of PES and Census
- Greater success in other nations
  - UK
  - Australia
  - Canada (using RRC)

Outline

- Evolution in coverage measurement
- Evolution in coverage measurement goals
- **Where to next?**

Past performance does not necessarily indicate future results.

Census

- Net national undercount has been very low in the last few census and generally decreasing.
- However, census success is always constrained by:
  - Ability to recruit an able temporary field staff
  - The cooperation of the population
  - Unknown unknowns

PES

- Everything that can be controlled has (largely) been controlled
  - Matching error
  - PES Fabrication
  - Geographic balancing
  - Etc.

PES

- What cannot be controlled might get worse
  - Missing data in Census
  - Missing data in PES
  - Mis-reporting of census day residence
  - Correlation bias

My Advice: Be modest

Realistic bounds around results rather than point estimate with (only) sampling error.

Shift emphasis from

‘what might be,’

to

‘what we know cannot be.’"

Does not require full blown total error model

See Tukey’s Sunset Salvo

Remember the Real Goal

The goal is **not**

"a great PES,"

but rather

"useful information about census coverage."

Work with DA to provide the users and planners with the best information possible.