

Discussion

Alan M Zaslavsky

Harvard University Medical School

Some history

- CM since 1970 (working off various surveys)
 - Focus on net coverage
 - 1980's-1990's: 'adjustment controversy'
 - Legal/constitutional issues, uncertainty over technical feasibility
- 2000-2010: more focus on evaluation

How to measure errors

- Net under/overcount:
 - Assesses bias in census for groups
 - Would be used for synthetic adjustment
 - Does not reveal “balancing” errors
 - Compensate for omissions with duplications? (2000)
 - Balancing error leaves lots of room for error for subgroups

How to measure errors

- Gross errors (inclusion and omission)
 - More errors can hide more subgroup biases
 - Doesn't tell us effect on ultimate outcome
 - Definition depends on processes
 - Geography; how much misplacement changes 'correct' to 'balancing errors'?
 - Outside block search area
- Many fruitless arguments!

Components of census enumeration

- New focus in 2010 CCM
- Census components
 - Correct enumerations
 - Erroneous enumerations
 - Duplicates
 - Others
 - Whole-person imputations
- CCM components
 - Correct enumerations
 - Omissions

Alternative definitions can be accommodated

- Geography
 - Misplacement in nearby block unimportant for most purposes
 - Misplacement in wrong state more serious
- Similar concept could be applied to other characteristics
 - Personal characteristics (e.g. race, age)
 - Count-known whole household imputations

Value of components approach

- Identify problem domains
 - Geographic, demographics
- Attach to populations identified with processes
 - Potential for evaluating and modifying
 - Common categories for evaluation studies
- Flexibility in refined definitions

Synthetic estimation

- Logistic regression is an advance
 - Aho & Mulry
- Allows inclusion of more process variables
 - Cutting across geographic/demographic strata
- More ‘analytical’ models
 - Look at coefficients as well as subgroup predictions

Synthetic estimation

- Error estimation ('synthetic bias')
 - Residual variance of error from synthetic model
 - Estimates by fitting multilevel model
 - 'direct' = truth + sampling error
 - truth = synthetic + 'bias'
 - Constitutes major portion of RMSE
- Use in assessing models

Synthetic estimation

- Synthetic error variance of substantive interest
 - Residual unmeasured SD of coverage $\approx 2\%$ in large counties
 - Only way to summarize variation
- Challenging to estimate and model
 - ~ 1 CCM sample cluster / 50,000 population!
- Model-based small area estimation?

Other ideas

- Assess variance due to imputation
 - Another predictive model with residual at various levels
- Experiments?
 - Limitations: can't randomize in decennial?
 - Maybe at low levels? Random vs systematic variation in methods
 - Other designs for evaluation
- Administrative records look pretty helpful!
- Microdata samples?

Final comments

- Substantial advances
 - Conceptual clarity on CM targets
 - More relevance to process improvement
 - Better memos
- Look forward to evaluation studies
 - More modeling?