COMPARATIVE ERROR AND TOTAL SURVEY ERROR

BRAD EDWARDS, WESTAT
AND TOM W. SMITH, NORC
AT THE UNIVERSITY OF CHICAGO

TALK STRUCTURE

- tse framework

- 3mc influence

- recent literature

- examples, and thoughts on quantifying comparability

- outside the box

- conclusion, future research

TSE FRAMEWORK

- difference between a true value and a survey estimate for a construct; MSE

- Deming (1944), Hansen et al (1953), Kish (1965), Brown (1967), Lessler (1984), Groves (1989), Smith (1996), Biemer & Lyberg (2003), Alwin (2007), Biemer & Lyberg (2011)

- two error domains
  - variance or variable error, random, no effect on means
  - bias, systematic error, directional, affects means

- two error branches
  - samplIng
  - non-sampling

- interactions, tradeoffs between error sources

MAJOR TSE COMPONENTS

1. construct validity
2. measurement error
3. processing error
4. coverage error
5. sampling error
6. nonresponse error
7. adjustment error

3MC INFLUENCE

MULTI-NATIONAL, -REGIONAL, -CULTURAL

- concerns about equivalence, errors, lack of a framework (Verba, 1950s), greater complexity

- most 3mc work done by naïve (one-country) researchers

- weak theoretical underpinnings of general survey research

- Harkness "professionalized" 3mc work, starting with translation and instrument design in the 1990s

- 2008 Berlin conference and 2010 monograph

- today
  - general survey research more mature
  - well-conceived 3mc sessions and papers abound in survey conferences

COMPARATIVE ERROR IN TSE PERSPECTIVE

- Smith (IJPOR 2011): Refining the Total Survey Error Perspective

    "Concept of comparison error should be used to extend TSE to cover multiple survey types."

- dimensions: time, data collection organization, population

- comparative studies (designed to compare 2 or more groups) by definition involve different populations, and most cross-national surveys involve different organizations

- the greater the difference on these dimensions = the greater the potential problems with comparability

- TSE can be a design "blueprint," inform design of each component to maximize comparability

COMPARABILITY IS RELATIVE

- Mohler & Johnson (2010) argue "equivalence" in all its guises is not a useful term for describing the relationship between data from different populations in comparative surveys. It implies an absolute that is impractical, never found in the real world

- "Comparability" is their preferred term; it implies degrees rather than an absolute.

- In TSE context, comparative error applies to every error source across population surveys

EXAMPLES

CALIFORNIA HEALTH INTERVIEW SURVEY

- largest Asian population in the US

- CHIS is a very large RDD phone survey designed to collect data in a number of Asian languages, as well as English and Spanish, in order to compare Asian groups with each other and with other populations (Hispanic, black, other...)

- Sunghee Lee conducted analysis showing impact on survey estimates if interviews had not been conducted in any Asian languages

- non-response bias based on age, health status, health conditions would have occurred

- comparison of Asians with other groups : comparative error reduced by conducting interview in respondent's preferred language

- cost/quality trade-off

WORLD HEALTH ORGANIZATION SURVEYS

- King (2004) developed vignettes to directly assess the incomparability of response categories in a culture

- a tool to calibrate respondents' self-assessments with their assessment of others' opinions in a culture, by asking all respondents their opinions about a few standard, brief, hypothetical scenarios

- China, Mexico, Slovakia

- reduced comparability error by "correcting" self-reported data using a direct measure of the incomparability of response categories

BEHAVIORAL RISK FACTOR SURVEILANCE SURVEY (BRFSS)

- state-wide surveys conducted in each of the US states by phone

- common core instrument

- comparison of response rates across states; proxy for survey quality, crude attempt to understand "house effects"

- BUT failed to account for population differences

- comparability problems so severe that premise of analysis was undermined

HOW CAN WE MEASURE COMPARABILITY?

1. quantification
   1. sometimes can obtain reliability and validity data from respondents or records. Differences between estimate and "true" value by population can inform quantitative estimate of comparability.
   2. sometimes analysis can quantify how much an estimate is affected by a design feature used for one but not all comparison groups.
   3. new tools provide more data on error properties for specific components (e.g., CARI samples to detect magnitude of question error, interviewer error, nonresponse error; subject to process quality control -- Hicks et al 2011)

2. direction
   1. if no way to quantify, can look at least for direction of error that might lead to comparability problems

OUT OF THE BOX

FITNESS FOR INTENDED USE AND PROCESS QUALITY CONTROL

CAN TSE CHANGE TO FIT 3MC?

- fitness for intended use (timeliness, relevance, etc.)

- "mindful" approach to comparative survey design takes advantage of 3mc surveys' unique ability to highlight comparability and definitional problems with concepts and constructs, which seems to go a bit beyond TSE framework

- concepts => constructs => indicators => questions

- Does the concept exist in all comparison groups?

- How does the concept relate to constructs in each culture? To indicators?

- advantages of adaptation, ADQ

FITNESS FOR INTENDED USE AND TSE

Comparability	Construct validity
Coherence	Measurement
Relevance	Processing
Accuracy	Coverage
Timeliness	Sampling
Accessibility	Nonresponse
Interpretability	Adjustment

3MC QA AND QC

- Process Quality Control

- Lyberg and Stukel (in 2010 monograph from Berlin conference):

    So far there has been very little research associated with controlling the quality of cross-national surveys. One possible explanation is that most resources have gone toward developing concepts, survey materials, and in targeting equivalence.

CONCLUSIONS AND FUTURE RESEARCH

- TSE and 3mc both coming of age, it's time they talked to each other more

- Comparative error is a unique aspect of TSE, layered on top of all error sources

- Comparative error has potential for broadening TSE and increasing its usefulness

- Carry comparative error concept through quality profile for well-documented 3mc survey program

- Quantify comparative errors for multiple 3mc measurement error sources

- Trade off two sources of comparative error on a 3mc survey

Thank you

bradedwards@westat.com