

Extending the Total Survey Error Perspective to Multiple-Surveys and Big Data

Tom W. Smith
NORC at the University of Chicago

AAPOR Webinar
September, 2016

Quality Dimensions

- Relevance
- Accuracy
- Timeliness
- Accessibility
- Comparability
- Coherence
- Completeness

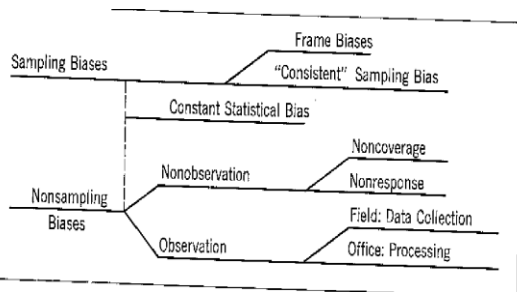
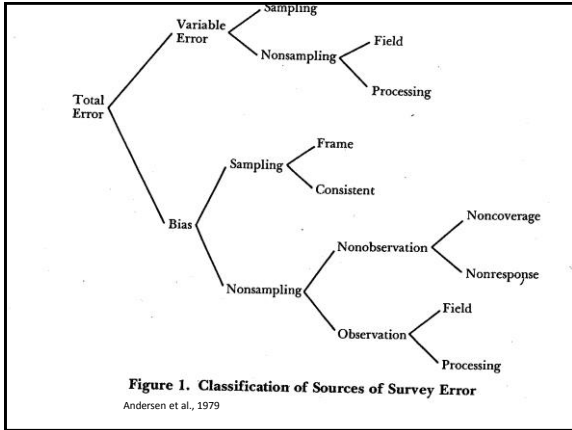
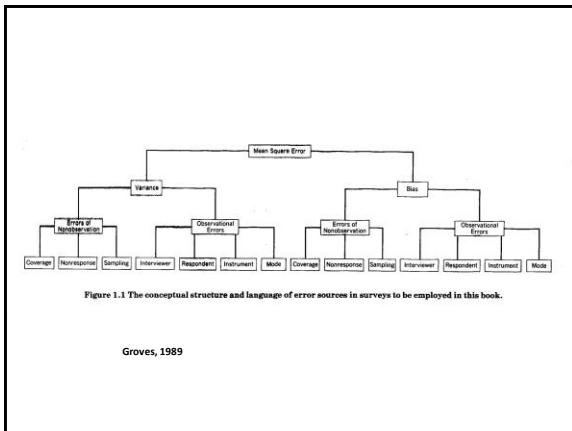
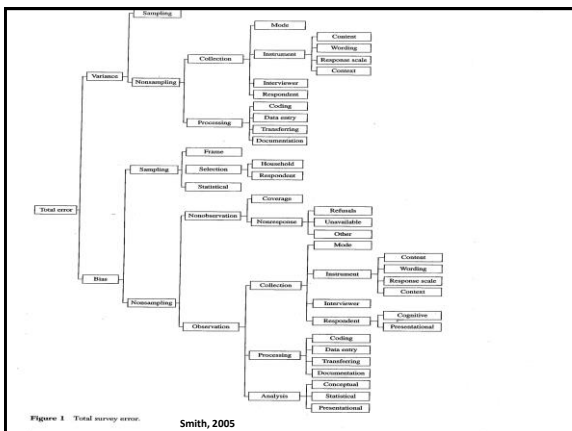


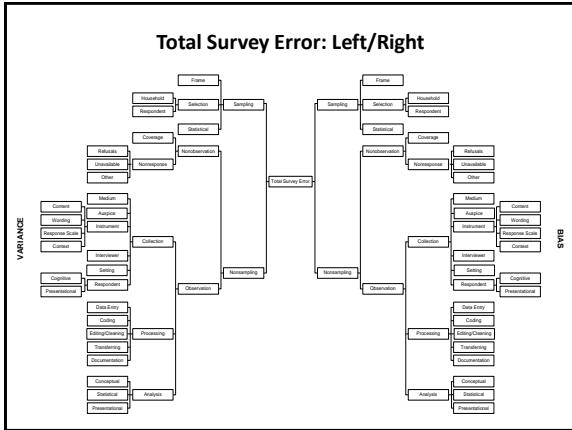
FIGURE 13.2.1 Classification of Sources of Survey Biases.

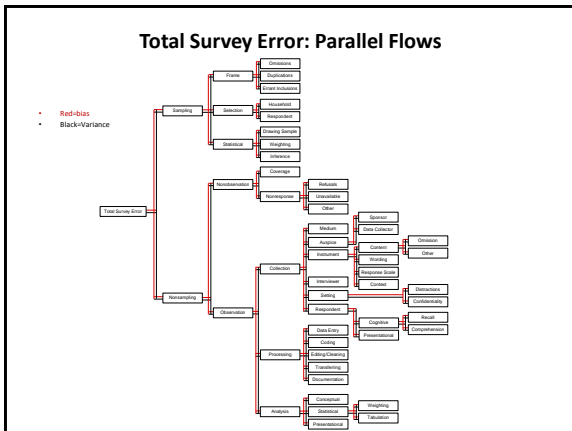
Kish, 1965

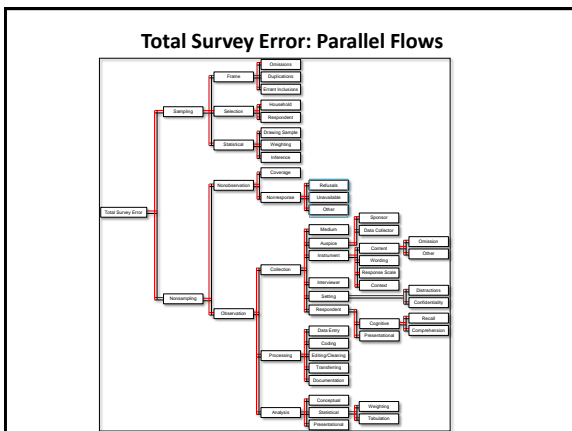








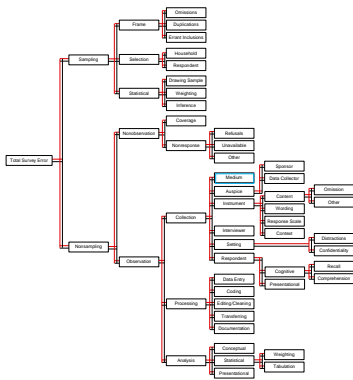




Categorizing Nonresponse Error

		Level of Nonresponse	
	Unit	Supplement/ SAQ	Item
Refusal	Refuse survey	Refuse supp./ SAQ	Refuse question
Unavailable	Noncontact	Null	Null
Other	Illness, Lost case	Illiterate, Poor eyesight	Cognitively unable

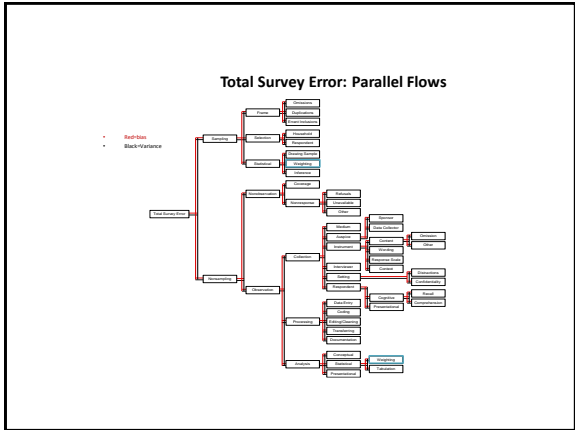
Total Survey Error: Parallel Flows

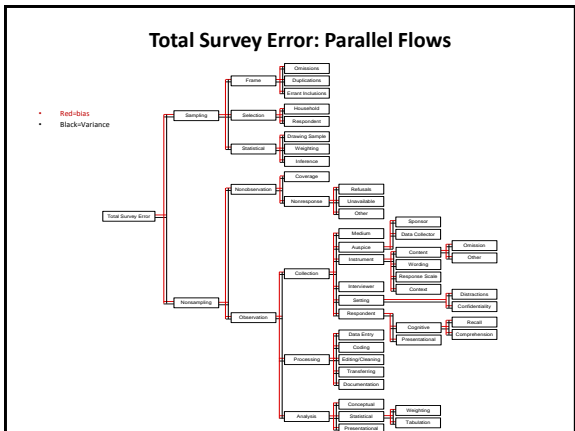


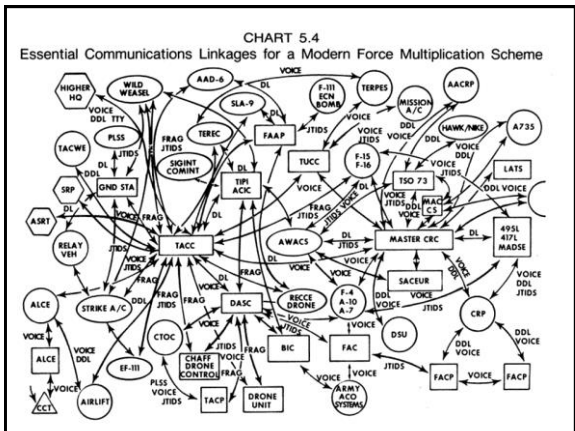
Typology of Surveys by Medium

	Computer		No Computer	
	Interviewer	Self-Admin.	Interviewer	Self-Admin.
Visual	Null	Internet CASI	Null	Postal Classroom handout
Audio	CAPi CATI	Automated voice+ voice recognition	PAPI TI	Null
Mixed	CAPi+show cards	Automated voice+ touchtone response cards ACASI	PAPI+show	TI+show cards TI+diary PAPI+SAQ

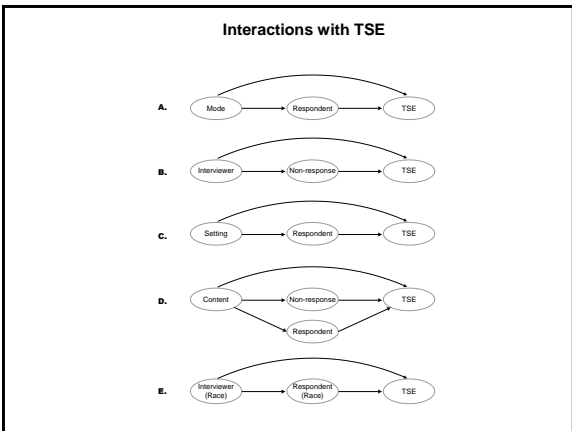
ACASI=audio computer-assisted self-interview
 CAPi=computer-assisted personal interview
 CASI=computer-assisted self-interview
 CATI=computer-assisted telephone interview
 Null=rare or non-existent
 PAPI=paper and pencil interview
 SAQ=self-administered questionnaire
 TI=telephone interview











TSE and Multiple Surveys

Times Series

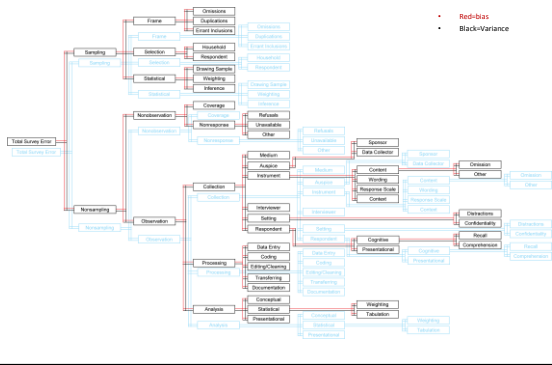
Panel Waves

Comparative

Cross-National/Cross-Cultural

Other

Total Survey Error: Comparison Error



Uses of TSE in Comparative Perspective

The TSE paradigm is a valuable approach for comparative studies for several reasons.

First, it is a blueprint for designing studies. Each component of error can be considered with the object of minimizing comparison error.

Second, it is a guide for evaluating error after the surveys have been conducted. One can go through each component and assess the level and comparability of the error structures.

Third, it can set a methodological research agenda for study error and for the design of experiments and other studies to fulfill that agenda.

Fourth, it goes beyond examining the separate components of error and provides a framework for the combining of the individual error components into their overall sum.

Fifth, by considering error as an interaction across surveys, it establishes the basis for a statistical model for the handling of error across surveys.

Combining TSE with Traditional Functional Equivalence Approach

- Focusing on most important cases of comparison error
- Focusing on comparability
- Etic vs. Emic
- Comparability of scales

Survey Data		
	<u>Yes</u>	<u>No</u>
Big Data	a. Linked	Alternative/New
	b. Unlinked, Used	
	<u>No</u>	Traditional
		Other/Nothing

Survey Data		
	<u>Yes</u>	<u>No</u>
Big Data	a. Linked	Alternative/New
	b. Unlinked, Used	
	<u>No</u>	Traditional
		Other/Nothing

Most Visited Websites, March 2016			
	<u>Alexa</u>	<u>SimilarWeb</u>	<u>eBizMBA</u>
1	Google	Facebook	Google
2	YouTube	Google	YouTube
3	Facebook	YouTube	Facebook
4	Baidu	VK	Yahoo
5	Yahoo	Yahoo	Amazon
Amazon	6	14	--
Baidu	--	16	Not top 15
VK	21	--	Not top 15

Survey Data		
	<u>Yes</u>	<u>No</u>
	<u>Yes</u>	Alternative/New
Big Data	a. Linked	
	b. Unlinked, Used	
	<u>No</u>	Other/Nothing
	Traditional	

Using Survey and Big Data Together

- Macro – post-harmonization error
- Micro – linkage error
 (common names; apartments; rural addresses; typos)
 (levels – respondent, household, neighborhood, community)

TSE + TADE + LE = TDE
TSE + TBDE + LE = TDE

Developing TBDE Models

- Biemer, Paul, "Errors and Inference," in Big Data and Social Science: A Practical Guide to Methods and Tools, edited by Ian Foster et al. New York: Routledge, 2016.
- Hsieh, Yuli Patrick, "When #THC Meets #TSE on @Twitter," RTI SurveyPost Blog, 2015.
- Hsieh, Yuli Patrick and Murphy, Joe, "Total Twitter Error? A Discussion of Surveys and Twitter for Examining Attitudes towards Marijuana Legalization and Abortion Rights," Paper presented to the American Association for Public Opinion Research, Hollywood, Florida, May, 2015.
- Japiec, Lilli et al., AAPOR Report on Big Data. AAPOR Big Data Task Force, February 12, 2015.

Otis Dudley Duncan's Outrigger Principle

"We therefore suggest the 'outrigger principle'...in which there are two supports for the main vessel. The main vessel is the program for repeated cross-sectional surveys to ascertain changes in response levels and differential changes for population subgroups. It is supported on one side, by studies of media output, organizational activities, political debate, and so forth that provide indicators of the everchanging complex of stimuli that may move public opinion and commitment. On the other side, the main vessel is supported by special-purpose, in-depth surveys of leadership cadres, members of relevant groups, communities where effects may be greater or appear sooner, and of panels, which provide supplementary data on the process of change, calibrated by items shared with the main survey. Further supplements to traditional surveys ought to be considered."
