Survey Quality Evaluation for Business Surveys

Paul P. Biemer
RTI International and
University of North Carolina at Chapel Hill

Outline

- Survey quality and total survey error (TSE)
- Overview of business surveys
- Business vs. household surveys
- TSE evaluation in business surveys
- Two illustrations
  - Decomposition of nonresponse bias
  - Survey level evaluations of TSE

Survey Quality

Total survey quality is multi-dimensional

- Producer vs. user dimensions
- Producers focus on *accuracy* (data quality) and *credibility* (reputation as a survey organization)
- Users are often more concerned about
  - *Timeliness* of the data
  - Its *relevance* to their needs
  - That it is *accessible* and clearly documented (*interpretability*), and
  - That the *comparability* of the data across time, space, and demographic domains is preserved

User and Producer Perspectives of Survey Quality

- Producers place high priority on
  - **Accuracy** – total survey error is minimized
  - **Credibility** – credible methodologies; trustworthy data

- Users place higher priority on
  - **Timeliness** – data deliveries adhere to schedules
  - **Relevance** – data satisfy user needs
  - **Accessibility** – access to data is user friendly
  - **Interpretability** – documentation is clear; meta-data are well-managed

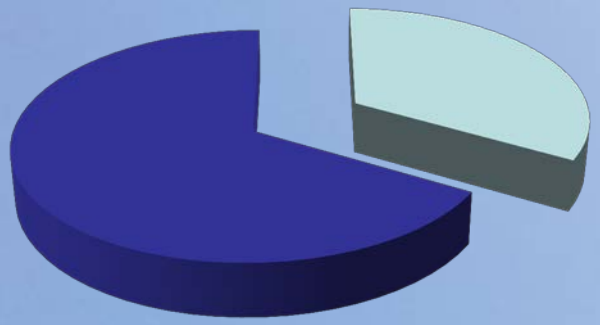
Other Important User Dimensions of Survey Quality

- **Comparability** – valid demographic, spatial and temporal comparisons
- **Coherence** – estimates from different sources can be reliably combined
- **Completeness** – data are rich enough to satisfy the analysis objectives without undue burden on respondents

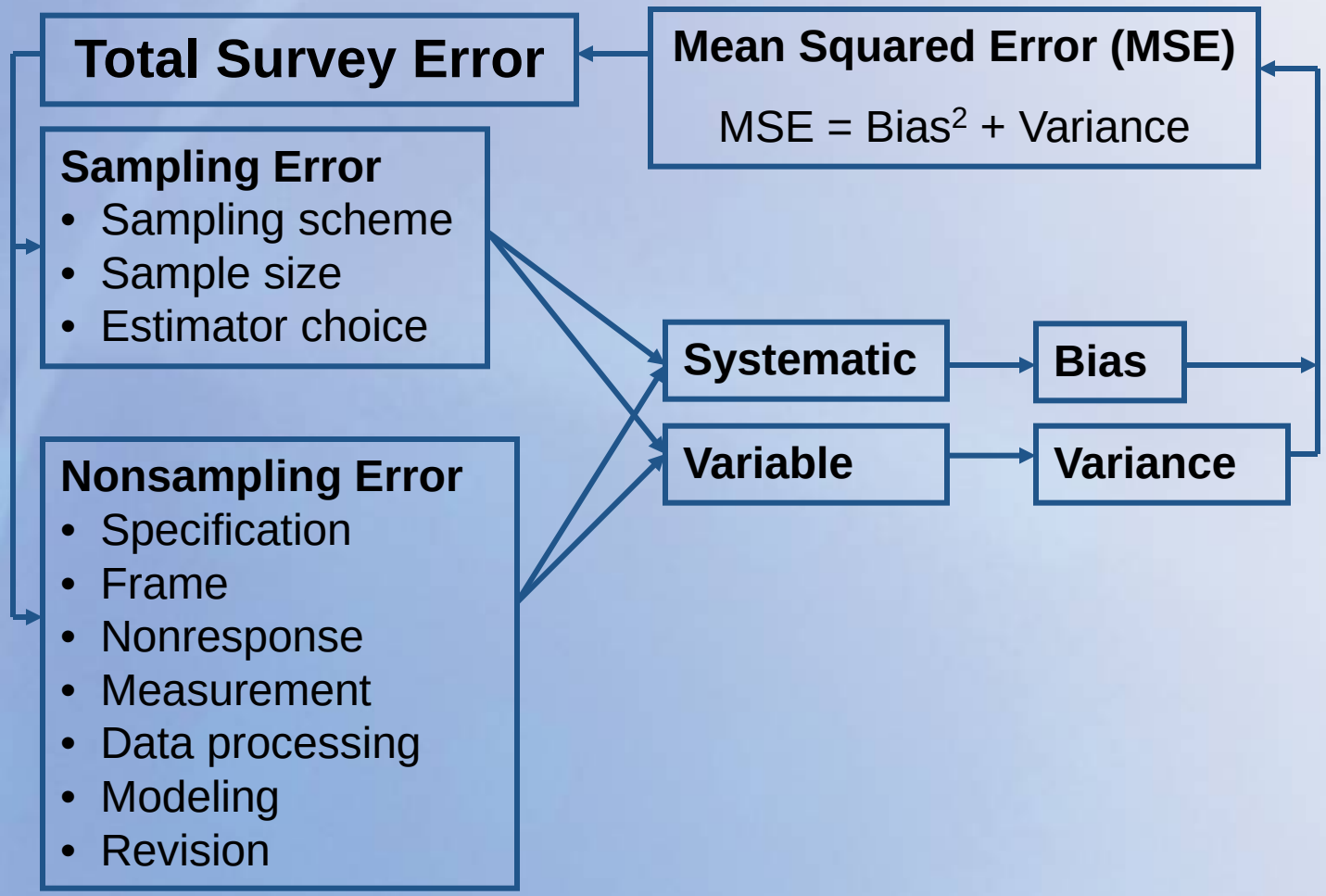
Optimal survey design balances these perspectives

- Identifies measurable and achievable objectives for each user-defined dimension of quality
- Determines costs and resources required to achieve these objectives
- Maximizes survey accuracy with remaining budget

Survey Budget = Cost of Accuracy
+ Cost of User-Defined Quality



Accuracy is maximized by minimizing total survey error within the available budget



Types of Business Surveys

- Enterprise surveys
  - Economic census
  - Survey of Business Owners
- Establishment surveys
  - Current Employment Survey
  - National Compensation Survey
- Employee surveys
  - Occupation Information Network (O*NET) Survey
  - Federal Employee Viewpoint Survey
- Customer surveys
  - Consumer Assessment of Healthcare Providers and Systems
  - Airport Service Quality Survey

Design Features Unique to Business Surveys

Sample

- Target populations are highly skewed by size
- Frame units are complex (one-to-many, many-to-one)
- Stratifiers (e.g. industry) can be unreliable

Response

- Decision to participate is not the informant's
- Time to respond costs sample members real money
- Gatekeepers block access to decision makers
- Multiple respondents needed per questionnaire
- Data transmitted electronically in format of choice

Design Features Unique to Business Surveys (cont'd)

Questionnaire/Form

- Information retrieved from records
- Responses to some questions generated by models
- Accounting concepts are complex requiring explicit definitions

Estimation

- Estimates may be compiled from multiple surveys and other data sources
- Published estimates may be preliminary and revised
- Small errors in the published estimates can have enormous consequences to national statistics

Types of Errors that Arise in Business Surveys

Many errors are similar to household surveys

- Specification error
- Frame error
- Nonresponse error
- Measurement error
- Data processing error
- Modeling error

However, they may have very different causes and frequencies. In addition, we can encounter two additional error sources:

- Revision error
- Compilation error

Total Survey Error Evaluation

- Addresses several dimensions of total survey quality.
- Essential for optimizing resource allocations to reduce the errors.
- In experimentation, needed to compare the quality of alternative methods.
- Provides valuable information on data quality for gauging uncertainty in estimates, interpreting the analysis results, and building confidence and credibility in the data.

Key Methods for Evaluating Nonresponse Bias

- Comparing surveys to external data
  - frame data, benchmarking against other surveys, nonresponse followup studies
- Analysis of response propensity
  - response rates by subgroups, representativeness indicators, level of effort analysis, incentive experiments
- Contrasting alternative post-survey adjusted estimates
  - “missing not at random” (calibration or callback model weighting) vs. “missing at random” (logistic regression) adjustments

Key Methods for Evaluating Measurement Error

- External consistency analysis
  - record check and other gold standards studies, interview/reinterview analysis
- Internal consistency analysis
  - correlations with replicate or similar measures
- Experimental designs
  - split-ballot, interpenetration, mode comparisons studies
- Model-based approaches
  - structural equation modeling, latent class analysis, other latent variable methods

Some Methods for Other Error Sources

- Cognitive lab methods (comprehension/recall error, data sensitivity, questionnaire issues)
- Subject matter expert reviews of concepts vs. question meaning (specification error)
- Multiple frame comparisons (frame error)
- Code/re-code consistency analysis (coding error)
- Key/re-key consistency analysis (keying error)
- Pre- and post-editing comparisons (editing error)
- Revision comparisons (revision error)

Illustrations of Two New Methods

- Nonresponse decomposition analysis
  - Illustrated by application to O*NET
- Total survey error evaluation
  - Illustrated by application to Stat Sweden surveys

O*NET Survey: Background

- Occupation Information Network Survey (O*NET)
- Provides descriptive ratings on 800+ U.S. occupations
- Target population is all U.S. employees in these 800+ occupations
- Continuing survey since 2001
- Telephone contacts to establishments to select sample
- PAPI or Internet questionnaires completed by selected employees
- Two-stage sample design
  - Establishments and employees within establishments in selected occupations
  - ~125,000 participating establishments (76% RR)
  - ~162,000 employee respondents (65% RR)

Response Stages

Stage	Description
1. Verification	Verify that an selected establishment is eligible
2. Screening	Determine whether the establishment employs the occupations of interest
3. Recruiting	Obtain cooperation of the POC to proceed with sampling and data collection
4. Sampling	Sample the employees from lists provided by the POC
5. Response	Obtain a completed questionnaire from the sample member

Decomposition of Nonresponse Bias

$$\text{Bias}(\hat{\mu}_5)$$

$$= \pi_0(\mu_5 - \mu_0)$$

$$+\pi_1(\mu_5 - \mu_1)$$

$$+\pi_2(\mu_5 - \mu_2)$$

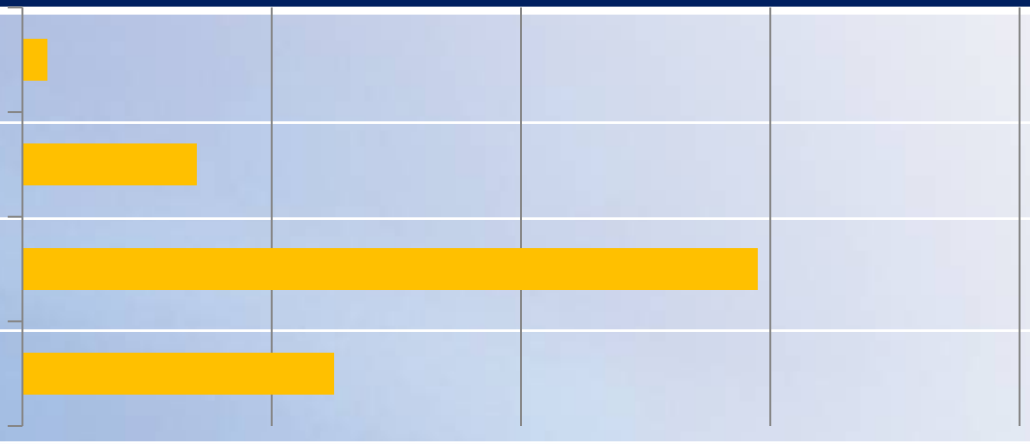
$$+\pi_3(\mu_5 - \mu_3)$$

$$+\pi_4(\mu_5 - \mu_4)$$

Bias of an estimate based only upon sample members completing Stage 5
Did not complete Stage 1
Completed Stage 1, but not Stage 2
Completed Stage 2, but not Stage 3
Completed Stage 3, but not Stage 4
Completed Stage 4, but not Stage 5

Preliminary Results

Stage	Average Contribution to Bias
1. Verification	2%
2. Screening	14%
3. Recruiting	59%
4. Sampling	25%
5. Response	Not yet evaluated



Next Steps

- Extend decomposition to all five stages
- Incorporate selection weights
- Develop strategies to address the major contributors of nonresponse bias
- Incorporate decomposition approach in adaptive total design (ATD) dashboards and interventions

A General System for Evaluating TSE
The Case of Statistics Sweden

- Background
  - Need for a quality evaluation system and process for Statistics Sweden
  - Ministry of Finance will use results to monitor quality improvements over time
- Survey quality must be assessed for many surveys, registers, and programs within the agency
- The process must be thorough, the reporting must be simple, and the results must be credible
- Paul Biemer and Dennis Trewin asked to develop and implement this system

Products to be Reviewed

Survey Products	Error Sources
Foreign Trade of Goods Survey (FTG) Labour Force Survey (LFS) Annual Municipal Accounts (RS) Structural Business Survey (SBS)	Specification error Frame error Nonresponse error Measurement error Data processing error Sampling error Model/estimation error Revision error

Products to be Reviewed (cont'd)

Registers	Error Sources
Business Register (BR) Total Population Register (TPR)	Specification error Frame: Overcoverage Undercoverage Duplication Missing Data Content Error
Compilations	**Error Sources**
National Accounts (NA) Consumer Price Index (CPI)	Specification error Missing Data Content error Sampling error Model/estimation error Revision error

Quality Criteria were Applied to Each Error Source

Criteria by Error Source

1. Knowledge of risks
2. Communication with users
3. Compliance with standards and best practices
4. Available expertise
5. Achievement toward risks mitigation and/or improvement plans

Ratings by Criterion

Poor ( )

Fair ( )

Good ( )

Very Good ( )

Excellent ( )

Risks to Data Quality by Error Source

High, Medium, Low

An Example of the Rating Guidelines – Knowledge of Risks

Poor	Fair	Good	Very Good	Excellent
Internal program documentation does not acknowledge the source of error as a potential factor for product accuracy.	Internal program documentation acknowledges error source as a potential factor in data quality.	Some work has been done to assess the potential impact of the error source on data quality.	Studies have estimated relevant bias and variance components associated with the error source and are well-documented.	There is an ongoing program of research to evaluate all the relevant MSE components associated with the error source and their implications for data analysis. The program is well-designed and appropriately focused, and provides the information required to address the risks from this error source.
	But: No or very little work has been done to assess these risks	**But:** Evaluations have only considered proxy measures (example, error rates) of the impact with no evaluations of MSE components	**But:** Studies have not explored the implications of the errors on various types of data analysis including subgroup, trend, and multivariate analyses	

The Evaluation Process

- Pre-interview activities
  - Background reading by the two evaluators
  - Self-assessments by each program area
- The Quality Interview
  - ½ day sessions involving 4-5 key product owners
  - Overview discussions of product processes
  - Detailed assessment of each of the 5 criteria
- Post-interview activities
  - Review of and comment on ratings by product owners
  - Ratings adjustments by evaluators to achieve equity

Example of Rating Results Structural Business Survey

Error Source	Average score	Knowledge of Risks	Communi-cation to Users	Available Expertise	Compliance with standards & best practices	Plan towards mitigation of risks	Risk to data quality
Specification	**46**						**M**
Frame	**62**						**M**
Nonresponse	**74**						**M**
Measurement	**50**						**H**
Data proc.	**52**						**H**
Sampling	**80**						**M**
Model/est'n	**60**						**H**
Revision	**58**						**H**
Total score	**59**						

Summary of Results for All Products

Error Source	RS	CPI	FTG	LFS	NA	SBS	BR	TPR	Avg
Specification	**74**	**68**	**62**	66	**56**	**46**	62	44	60
Frame	36	**42**	**62**	58		**62**			49
Overcov.							**48**	**52**	
Undercov.							**40**	34	
Duplication							46	64	
NR/Miss. data	**62**	36	**62**	**66**	64	**74**	40	**60**	57
Meas/Content	**52**	**40**	**54**	**50**	**58**	**50**	**42**	50	50
Data proc.	**46**	**70**	**46**	**54**	**44**	**52**			52
Sampling		**54**		**72**	**44**	**80**			64
Model/est'n	**54**	**64**	**66**	**46**	**44**	**60**			56
Revision	74		62		**62**	**58**			64
Total	**57**	**56**	**59**	**58**	**51**	**59**	**45**	**52**	55

Red Bold = High Risk, Black Bold = Medium Risk, No Bold = Low Risk

Strengths and Weaknesses of the Process

Strengths

- Comprehensive approach
- Easily understood by management
- Identifies important areas to improve within and across products
- Can be updated periodically to assess improvement

Strengths and Weaknesses of the Process

Weaknesses

- Does not really reflect total MSE
- Can be somewhat subjective
- Highly dependent on knowledge and skills of the external evaluators
- Requires thorough documentation of processes and improvements (e.g., quality profiles)

Summary

- Although similar in some ways, business surveys and household surveys have many differences
- These differences affect the error components and methods for evaluating them
- Editing error may be particularly problematic
- In addition, more studies of measurement error in business surveys are needed