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.

**Total Survey Error**

- **Sampling Error**
  - Sampling scheme
  - Sample size
  - Estimator choice

- **Nonsampling Error**
  - Specification
  - Frame
  - Nonresponse
  - Measurement
  - Data processing
  - Modeling
  - Revision

**Mean Squared Error (MSE)**

\[
\text{MSE} = \text{Bias}^2 + \text{Variance}
\]
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

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Verification</td>
<td>Verify that an selected establishment is eligible</td>
</tr>
<tr>
<td>2. Screening</td>
<td>Determine whether the establishment employs the occupations of interest</td>
</tr>
<tr>
<td>3. Recruiting</td>
<td>Obtain cooperation of the POC to proceed with sampling and data collection</td>
</tr>
<tr>
<td>4. Sampling</td>
<td>Sample the employees from lists provided by the POC</td>
</tr>
<tr>
<td>5. Response</td>
<td>Obtain a completed questionnaire from the sample member</td>
</tr>
</tbody>
</table>
## Decomposition of Nonresponse Bias

\[
\text{Bias}(\hat{\mu}_5) = \pi_0(\mu_5 - \mu_0) + \pi_1(\mu_5 - \mu_1) + \pi_2(\mu_5 - \mu_2) + \pi_3(\mu_5 - \mu_3) + \pi_4(\mu_5 - \mu_4)
\]

<table>
<thead>
<tr>
<th>Expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\pi_0(\mu_5 - \mu_0))</td>
<td>Bias of an estimate based only upon sample members completing Stage 5</td>
</tr>
<tr>
<td>(\pi_1(\mu_5 - \mu_1))</td>
<td>Did not complete Stage 1</td>
</tr>
<tr>
<td>(\pi_2(\mu_5 - \mu_2))</td>
<td>Completed Stage 1, but not Stage 2</td>
</tr>
<tr>
<td>(\pi_3(\mu_5 - \mu_3))</td>
<td>Completed Stage 2, but not Stage 3</td>
</tr>
<tr>
<td>(\pi_4(\mu_5 - \mu_4))</td>
<td>Completed Stage 3, but not Stage 4</td>
</tr>
<tr>
<td></td>
<td>Completed Stage 4, but not Stage 5</td>
</tr>
</tbody>
</table>
## Preliminary Results

<table>
<thead>
<tr>
<th>Stage</th>
<th>Average Contribution to Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Verification</td>
<td>2%</td>
</tr>
<tr>
<td>2. Screening</td>
<td>14%</td>
</tr>
<tr>
<td>3. Recruiting</td>
<td>59%</td>
</tr>
<tr>
<td>4. Sampling</td>
<td>25%</td>
</tr>
<tr>
<td>5. Response</td>
<td>Not yet evaluated</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Survey Products</th>
<th>Error Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreign Trade of Goods Survey (FTG)</td>
<td>Specification error</td>
</tr>
<tr>
<td>Labour Force Survey (LFS)</td>
<td>Frame error</td>
</tr>
<tr>
<td>Annual Municipal Accounts (RS)</td>
<td>Nonresponse error</td>
</tr>
<tr>
<td>Structural Business Survey (SBS)</td>
<td>Measurement error</td>
</tr>
<tr>
<td></td>
<td>Data processing error</td>
</tr>
<tr>
<td></td>
<td>Sampling error</td>
</tr>
<tr>
<td></td>
<td>Model/estimation error</td>
</tr>
<tr>
<td></td>
<td>Revision error</td>
</tr>
<tr>
<td>Registers</td>
<td>Error Sources</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>Business Register (BR)</td>
<td>Specification error</td>
</tr>
<tr>
<td>Total Population Register (TPR)</td>
<td>Frame: Overcoverage</td>
</tr>
<tr>
<td></td>
<td>Undercoverage</td>
</tr>
<tr>
<td></td>
<td>Duplication</td>
</tr>
<tr>
<td></td>
<td>Missing Data</td>
</tr>
<tr>
<td></td>
<td>Content Error</td>
</tr>
<tr>
<td>Compilations</td>
<td>Error Sources</td>
</tr>
<tr>
<td>National Accounts (NA)</td>
<td>Specification error</td>
</tr>
<tr>
<td>Consumer Price Index (CPI)</td>
<td>Missing Data</td>
</tr>
<tr>
<td></td>
<td>Content error</td>
</tr>
<tr>
<td></td>
<td>Sampling error</td>
</tr>
<tr>
<td></td>
<td>Model/estimation error</td>
</tr>
<tr>
<td></td>
<td>Revision error</td>
</tr>
</tbody>
</table>
Quality Criteria were Applied to Each Error Source

<table>
<thead>
<tr>
<th>Criteria by Error Source</th>
<th>Ratings by Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Knowledge of risks</td>
<td>Poor ( - )</td>
</tr>
<tr>
<td>2. Communication with users</td>
<td>Fair (○)</td>
</tr>
<tr>
<td>3. Compliance with standards and best practices</td>
<td>Good (○)</td>
</tr>
<tr>
<td>4. Available expertise</td>
<td>Very Good (●)</td>
</tr>
<tr>
<td>5. Achievement toward risks mitigation and/or improvement plans</td>
<td>Excellent (●●)</td>
</tr>
</tbody>
</table>

Risks to Data Quality by Error Source
High, Medium, Low
An Example of the Rating Guidelines – Knowledge of Risks

<table>
<thead>
<tr>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
<th>Very Good</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal program documentation does not acknowledge the source of error as a potential factor for product accuracy.</td>
<td>Internal program documentation acknowledges error source as a potential factor in data quality.</td>
<td>Some work has been done to assess the potential impact of the error source on data quality.</td>
<td>Studies have estimated relevant bias and variance components associated with the error source and are well-documented.</td>
<td>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.</td>
</tr>
</tbody>
</table>

**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
<table>
<thead>
<tr>
<th>Error Source</th>
<th>Average score</th>
<th>Knowledge of Risks</th>
<th>Communication to Users</th>
<th>Available Expertise</th>
<th>Compliance with standards &amp; best practices</th>
<th>Plan towards mitigation of risks</th>
<th>Risk to data quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification</td>
<td>46</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>○</td>
<td>○</td>
<td>M</td>
</tr>
<tr>
<td>Frame</td>
<td>62</td>
<td>#</td>
<td>○</td>
<td>○</td>
<td>#</td>
<td>#</td>
<td>M</td>
</tr>
<tr>
<td>Nonresponse</td>
<td>74</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>○</td>
<td>○</td>
<td>M</td>
</tr>
<tr>
<td>Measurement</td>
<td>50</td>
<td>○</td>
<td>#</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>H</td>
</tr>
<tr>
<td>Data proc.</td>
<td>52</td>
<td>○</td>
<td>#</td>
<td>#</td>
<td>○</td>
<td>○</td>
<td>H</td>
</tr>
<tr>
<td>Sampling</td>
<td>80</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>○</td>
<td>#</td>
<td>M</td>
</tr>
<tr>
<td>Model/est’n</td>
<td>60</td>
<td>○</td>
<td>○</td>
<td>#</td>
<td>○</td>
<td>#</td>
<td>H</td>
</tr>
<tr>
<td>Revision</td>
<td>58</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>H</td>
</tr>
<tr>
<td>Total score</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Summary of Results for All Products

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

**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
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.