Total Survey Error:
Design, Implementation, and Evaluation

Paul P. Biemer
RTI International
University of North Carolina

RTI INTERNATIONAL

Outline

- Total survey error (TSE)
- TSE paradigm: principles and strategies
  - Design
  - Implementation
  - Evaluation
  - Analysis
  - Cost considerations
- Final remarks

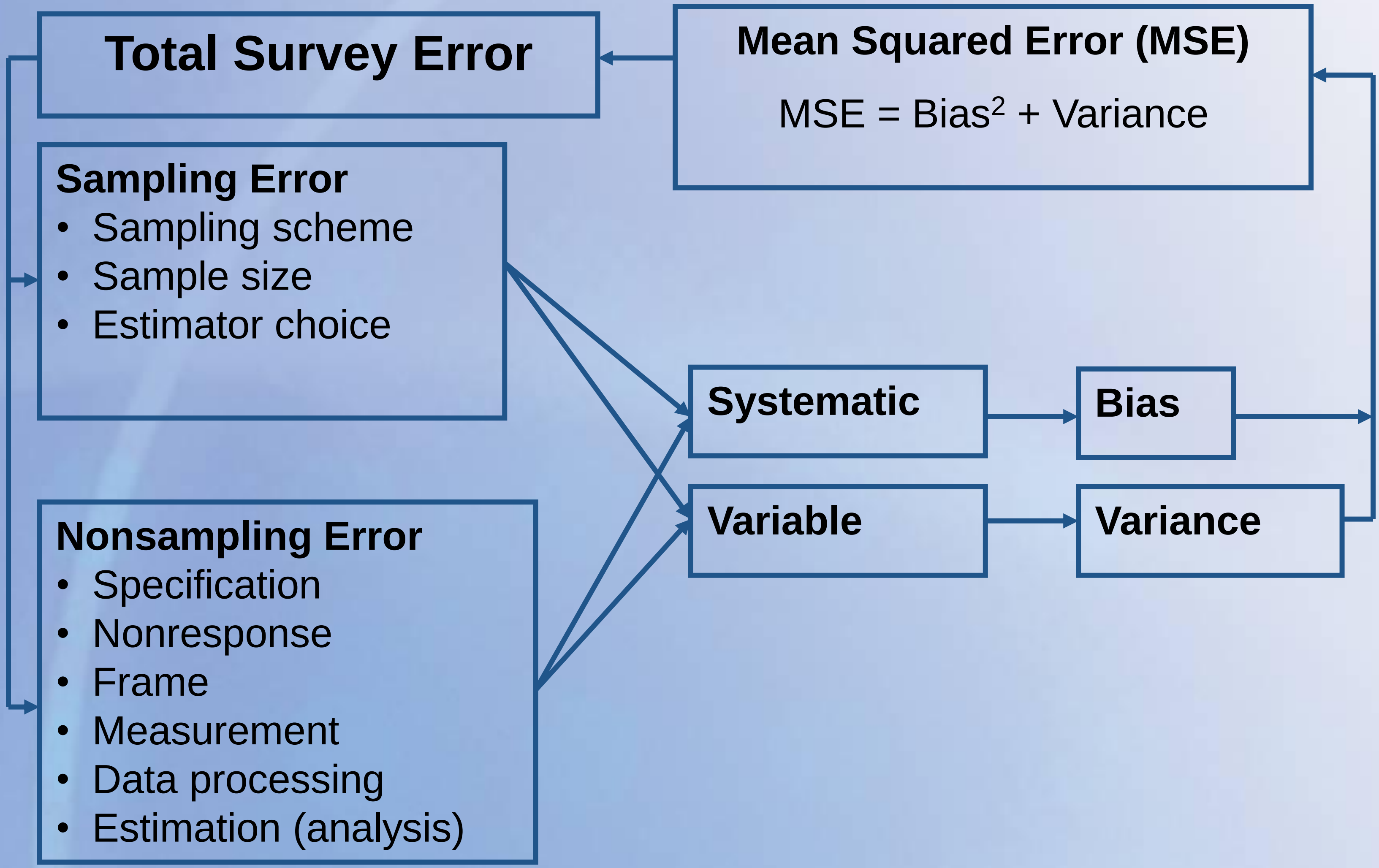
Total Survey Error

- The totality of error that can arise in the design, collection, processing, and analysis of survey data.
- References an estimator, say $\hat{\theta}$
- Measured by the ***total*** mean square error (MSE)

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}) &= E(\hat{\theta}-\theta)^2 \\ &= \mathrm{Bias}^2 + \mathrm{Variance}\end{aligned}$$

- "Total" MSE includes all error sources – both sampling error and nonsampling error

Total Survey Error



Total Survey Error Paradigm

- Concepts have evolved over the years
- ITSEWs have further refined the concept
- Currently, the TSE paradigm comprises “ideals” for
  - designing surveys,
  - implementing all aspects of the survey design,
  - evaluating the main components of TSE, and
  - analyzing survey results

within prespecified costs, timing, and other constraints.

TSE Paradigm: **Survey Design**

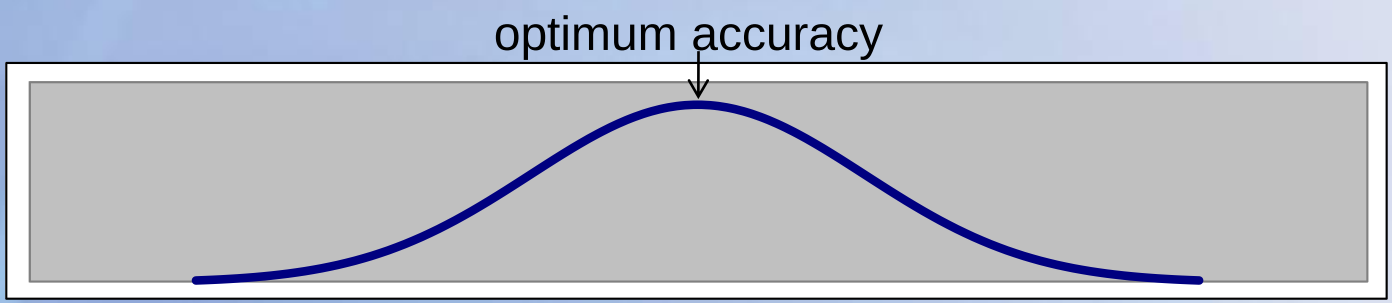
Surveys should be design (or redesigned) to the maximize data accuracy subject to budgetary constraints by minimizing the cumulative effects of error from all sources

Examples:

- Cost reduction by cutting sample size vs. alternatives that achieve equivalent reductions with less impact on TSE
- Effects of increasing response rates on measurement error
- Allocating resources to maximize frame coverage vs. applying same resources to minimize TSE

Key Design Principles

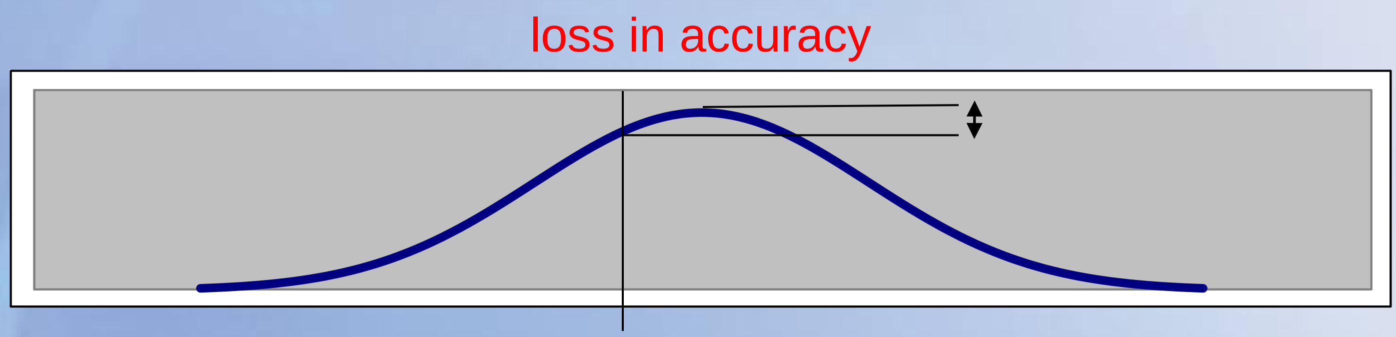
- *Design robustness* – accuracy does not change appreciably as the survey design features change; i.e. optimum is “flat” over a range of alternate designs



- *Effect generalizability* – design features found to be optimal for one survey are often generalizable to other similar surveys

Key Design Principles

- *Design robustness* – accuracy does not change appreciably as the survey design features change; i.e. optimum is “flat” over a range of alternate designs



- *Effect generalizability* – design features found to be optimal for one survey are often generalizable to other similar surveys

Implications for Design

- Compile information on TSE (e.g., quality profiles)
- Identify major contributors to TSE
- Allocate resources to control these errors
  - Use results from the literature and other similar surveys to guide the design
- Develop an effective process for modifying the design during implementation to achieve optimality
- Embed experiments and conduct studies to obtain data on TSE for future surveys

TSE Paradigm: **Implementation**

Adaptive designs should monitor all major error sources and should seek to minimize TSE through real-time interventions

Examples:

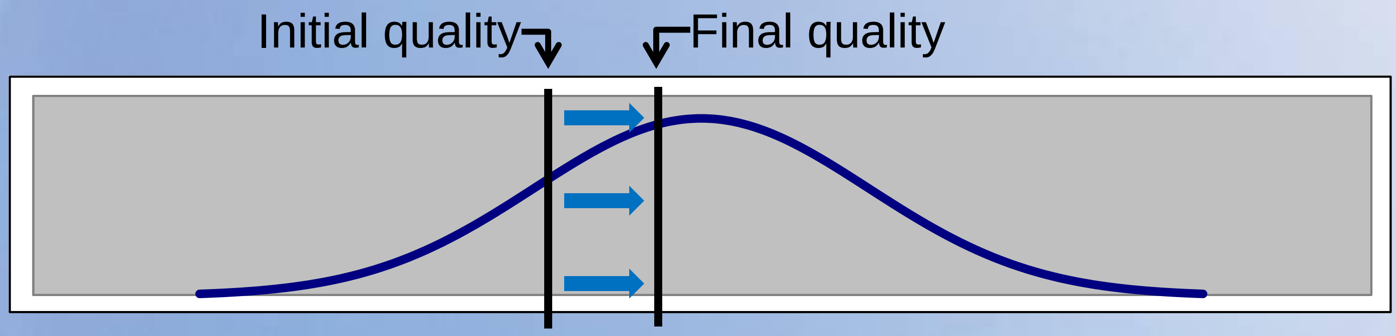
- CARI monitoring of interviewer performance on nonresponse followup cases.
- Strategies for reducing weight variation using two-phase sampling for nonresponse followup
- Reducing interviewer burden and error using ABS in areas where list coverage is adequate
- Model-aided reduction of nonresponse bias

Design Implementation Strategies

The initial survey design must modified or *adapted* during implementation to control costs and maximize quality.

Four strategies for reducing costs and errors in real-time:

- Continuous quality improvement
- Six sigma
- Responsive design
- Adaptive total design and implementation



Continuous Quality Improvement (CQI)

1. Prepare a workflow diagram of the process and identify key process variables.
2. Identify characteristics of the process that are critical to quality (CTQ).
3. Develop real-time, reliable metrics for the cost and quality of each CTQ.
4. Continuously monitor costs and quality metrics during the process.
5. Intervene as necessary to ensure that quality and costs are within acceptable limits.

Six Sigma

- Developed by Motorola in the 1980’s
- Definition (from Pande, et al, 2000, p. xi)

  “A comprehensive and flexible system for achieving, sustaining and maximizing business success,…uniquely driven by a close understanding of customer needs, disciplined use of facts, data, and statistical analysis, and diligent attention to managing, improving, and reinventing business processes.” – Pande, et al (2000, p. xi)

- Extends ideas of Total Quality Management (TQM) and continuous quality improvement (CQI)
- Has mostly been applied in business and manufacturing.

Special vs. Common Cause Variation

- Special causes – assignable to events and circumstances that are extraordinary, rare and unexpected
  - e.g., frame was not sorted prior to sampling
  - Addressed by actions specific to the cause leaving the design of the process essentially unchanged
- Common causes – naturally occurring random disturbances that are inherent in any process and cannot be avoided.
  - e.g., normal fluctuations of response across regions and months
  - Actions designed to address a common cause is neither required nor advisable; this lead to process “tampering”

Responsive Design Strategy

- Developed for nonresponse bias reduction, especially face to face data surveys (Groves & Heeringa, 2006)
- Similar to CQI but includes three phases:
  - Experimental phase – tests major design options
    - For e.g., split sample designs to test incentive levels
  - Main data collection phase – implements design selected in first phase
    - Continues until "phase capacity" is reached
  - NRFU phase – special methods implemented to reduce nonresponse bias and control data collection costs
    - NR double sampling, higher incentives, more intensive followup
- Phase capacity – point at which efforts to reduce NR bias under current protocol are no longer cost effective

Adaptive Total Design

- An approach for continuously monitoring survey processes to control errors, improve quality, and reduce costs.
- *Adaptive* in that it combines the real-time error control features of CQI, responsive design, and Six Sigma strategies.
- *Total* in that it simultaneously monitors multiple sources; for e.g.,
  - Sampling frame and sampling
  - Response quality
  - Nonresponse bias reduction
  - Field production
  - Costs and timeliness

TSE Paradigm: Survey Error Evaluation

Survey error evaluations should consider the major sources of error and their interactions

Examples:

- Effects of nonresponse followup on nonresponse bias and other error sources; for example, measurement error
- Correlations between response propensities and misclassification propensities
- Contributions of interviewer-related nonresponse to interviewer variance
- Effects on TSE of nonresponse followup, especially under subsampling designs

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

Basic Methods

- Nonresponse bias studies (required by OMB for some surveys)
  - Evaluates differences between respondents and nonrespondents for key survey items
  - Frame data or prior waves provides data on nonrespondents
  - Model-based approaches for nonignorable nonresponse bias
- Measurement bias studies
  - Record check studies
  - Reconciled reinterviews
  - Internal and external consistency checks
  - Test-retest reinterview approaches
  - Embedded repeated measures analysis (e.g. structural equation modeling, latent class analysis)

Basic Methods (continued)

- Other methods
  - Frame undercoverage evaluations
  - Editing error (pre- and post-editing comparisons)
  - Cognitive methods for detecting comprehension errors, recall problems, data sensitivity, etc.
  - Subject matter expert reviews of concepts vs. question meaning
  - Process data summaries
    - Response rate analysis
    - Data entry error rates
    - Edit failure rates
    - Missing data rates
    - Post-survey adjustment factors

TSE Paradigm: **Data Analysis**

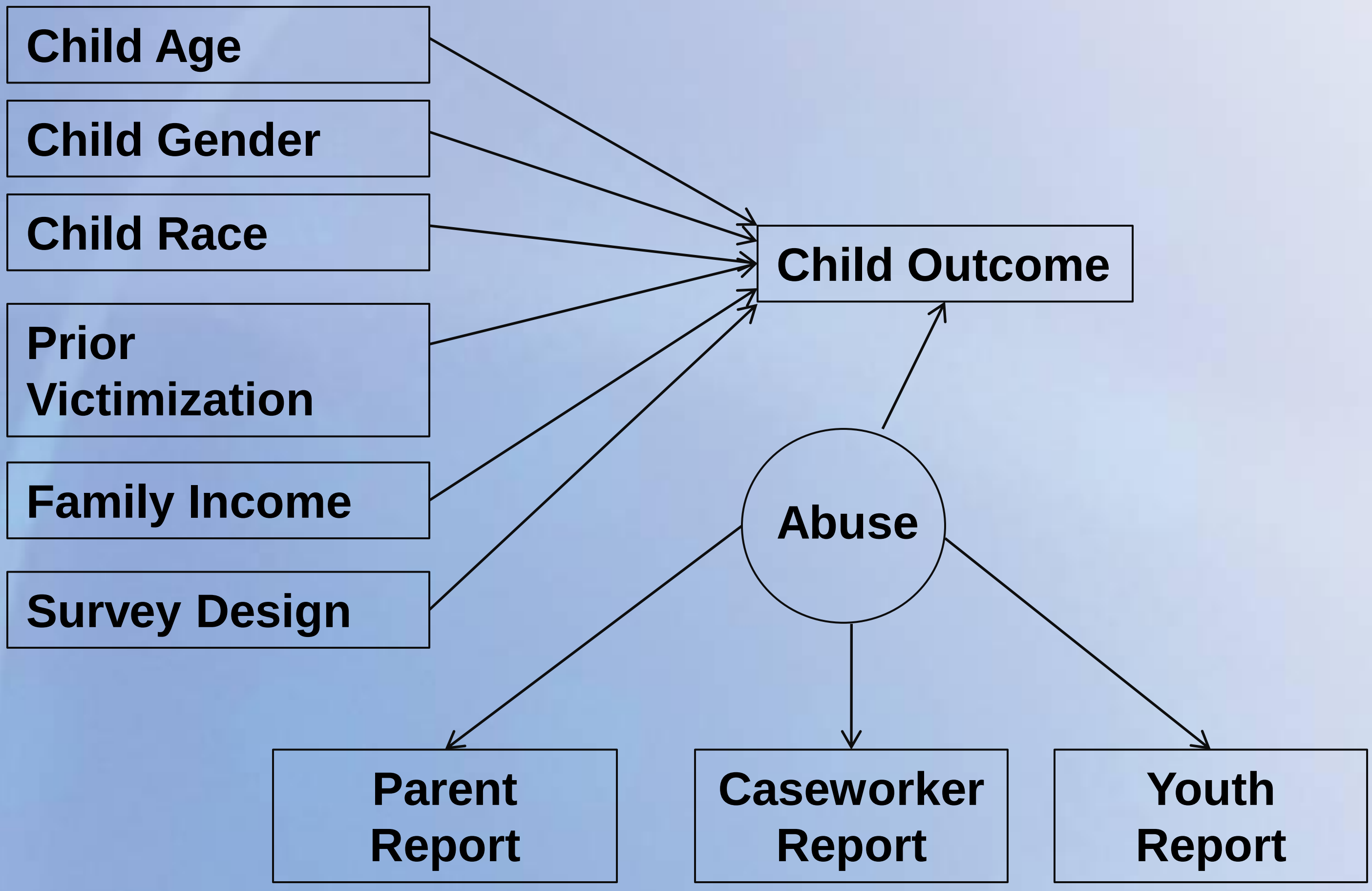
Data analysis should consider the effects

- Unequal probability sampling and clustering
- Nonresponse (both item and unit)
- Measurement error

Examples:

- Software to appropriately handling the complex survey design
- Methods to compensate for unit and item nonresponse
- Methods for dealing with measurement errors; for example, latent variable methods (Structural Equation Models, Latent Class Analysis, etc.)

Example: Correcting for Measurement Error in Reported Abuse in an Analysis of Child Outcomes (Izzo, et al, 2010)

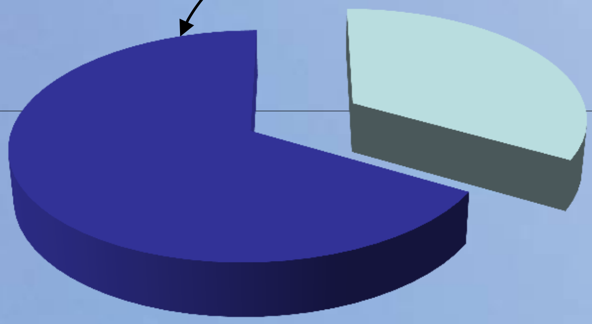


TSE Paradigm: **Costs**

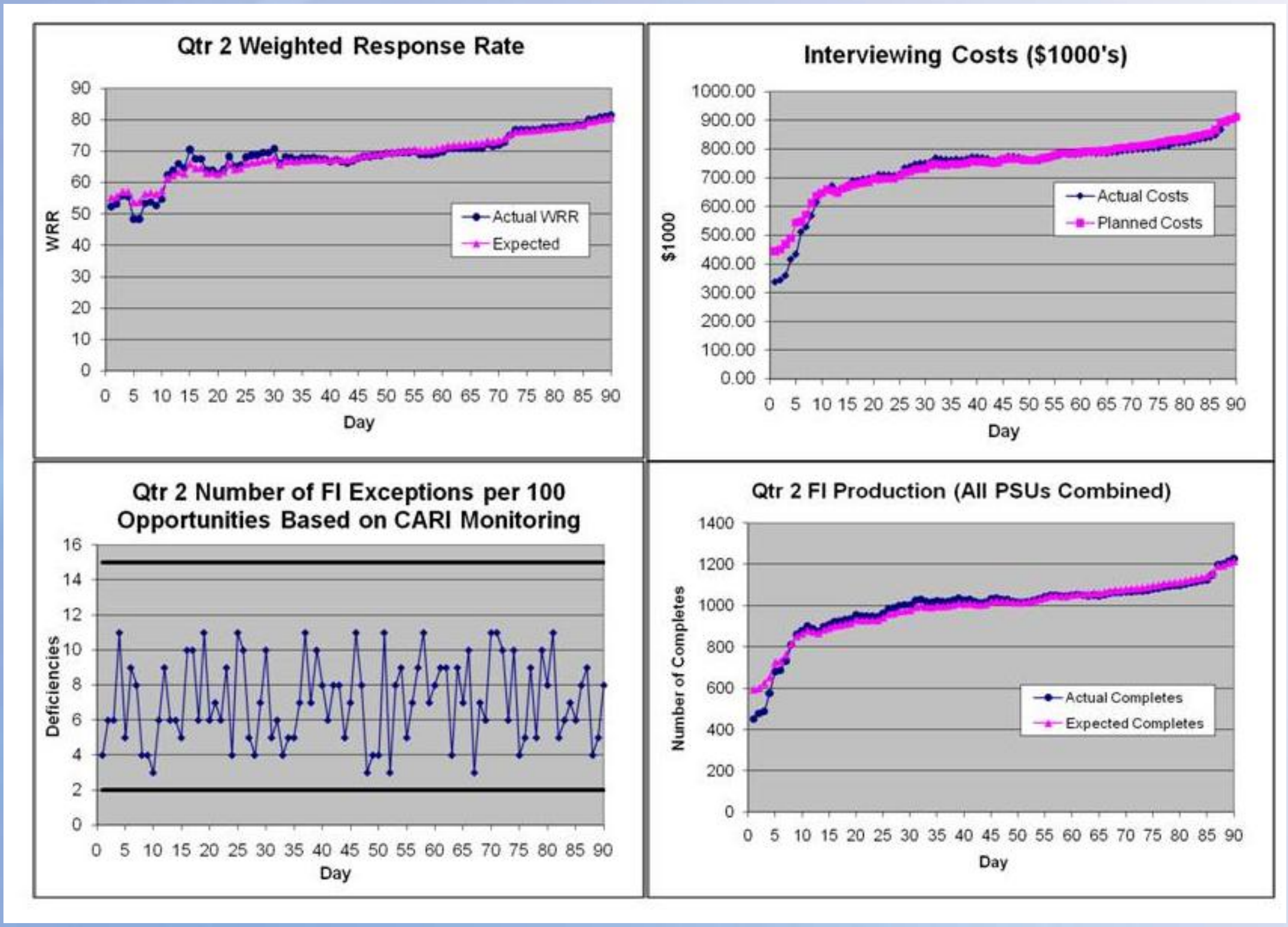
- Identify measurable and achievable objectives each dimension of survey quality
- Estimate the costs and resources required to achieve these objectives
- Maximize survey accuracy within the budget set aside for the control of TSE

Survey Budget = Cost of Accuracy

+ Cost of User-Defined Quality



Dashboard Showing Weighted Response Rates, Interview Costs, Interviewer Exceptions and Production



Final Remarks

The TSE Paradigm

- recognizes that survey quality is more than data accuracy (*total survey quality*),
- comprises a set of principles, optimal strategies, and best practices for survey design, implementation, evaluation, and data analysis
- produces survey results which are as accurate as possible within cost, schedule, and other constraints of the project
- seeks an optimal balance of accuracy with other key dimensions of survey quality

For a copy of these slides, contact me at: ppb@rti.org

Dimensions of Survey Quality

- **Accuracy** – total survey error is minimized
- **Credibility** – credible methodologies; trustworthy data
- **Timeliness** – data deliveries adhere to schedules
- **Relevance** – data satisfy user needs
- **Accessibility** – access to data is user friendly
- **Interpretability** – documentation/metadata are clear
- **Comparability** – valid demographic/spatial/temporal comparisons can be made
- **Coherence** – estimates can be reliably combined across sources
- **Completeness** – data are adequately detailed and complete