TOTAL SURVEY ERROR

Roots and Evolution

ITSEW, Bergamo, June 10, 2019

Lars Lyberg, PhD

Inizio

INIZIO

My Background

- Statistics Sweden 1966-2010
- Head of R&D, Statistics Sweden 1983-1999
- JOS Chief Editor 1985-2010
- SU 2003-2015 Associate professor survey methodology
- 2010- Consultant
- 2014- Inizio

INIZIO

Outline

- The total survey error (TSE) paradigm
- Historical backdrop
- Quality management philosophies (QM) and Total Survey Quality (TSQ)
- How to study survey quality
- How we must change to better serve our users
- The new survey landscape

INIZIO

Types of Survey

- One-time
  - Attitudes, opinions
- Repeated or continuing including panel studies
  - Official statistics (short term indicators, agriculture, living conditions, crime)
  - Other (drug use, consumer research, behaviors)
- International and comparative
  - Official statistics (European Statistical System, poverty, water supply)
  - Student achievement, literacy, values, happiness, marketing, attitudes

INIZIO

Total Survey Error as a Concept TSE

- Purports to describe statistical properties of survey estimates incorporating ”all” error sources, not just sampling
- TSE is a planning criterion
- Among a set of alternative designs, choose the design that gives the smallest TSE, i.e. the highest accuracy for an estimate

Total Survey Error Paradigm

- Embodies a set of principles, methods, and processes that minimize total survey error within the budget allocated for accuracy.

- Other dimensions of total survey quality can viewed as constraints – timeliness and comparability constrain the design; accessibility, relevance, and completeness constrain the budget; and so on.

INIZIO

Key Statistical Developments in Total Survey Error 1

- Errors of observers can be correlated (1902), Karl Pearson
- Deming's survey error typology (1944)
- Interpenetrating samples (1946), Mahalanobis
- Criteria for true values (1951), Hansen, Hurwitz, Marks and Mauldin
- Essential survey conditions, correlated response variance (1959), Hansen-Hurwitz-Bershad (H-H-B)
- U.S. Census Bureau survey model "mixed-error model"(1961), H-H-B

INIZIO

Key Statistical Developments in Total Survey Error 2

- Interviewer effects using ANOVA (Kish 1962)
- Simple response variance via reinterviews (1964), H-H-Pritzker
- Relaxed assumptions of zero covariance of true values and response deviations (1964, 1974), Fellegi
- Errors of Measurement (1968), Cochran
- Estimating model components via basic study schemes using replication, interpenetreation and combinations of the two (1969), Bailar and Dalenius
- Estimating nonsampling variance using mixed linear models (1978), Hartley and Rao
- “Error Profile” of Current Population Survey (1978), Brooks and Bailar

INIZIO

Key Statistical Developments in Total Survey Error 3

- Measurement of imputation error variance through multiple imputation (1987), Rubin
- Total error model for PES (1991), Mulry and Spencer
- Measurement Errors in Surveys (1991) (Biemer and Stokes; O'Muircheartaigh; Fuller; and others)
  - Attempts to juxtapose psychometric notions with survey statistical notions of measurement error
- Latent class model applications to survey errors (late 1990's), Biemer; Tucker; and others
- Measurement error effects on analysis (1997), Biemer and Trewin

INIZIO

Key Statistical Developments in Total Survey Error 4

- Certain components understudied
  - coverage error variance
  - nonresponse error variance
  - all processing errors
  - biases in general
- Simultaneous treatment of more than one error source

$$MSE = Bias^2 + Variance$$

$$= (B_{Spec} + B_{Fr} + B_{Mea} + B_{NR} + B_{DP} + B_{Mod})^2 +$$

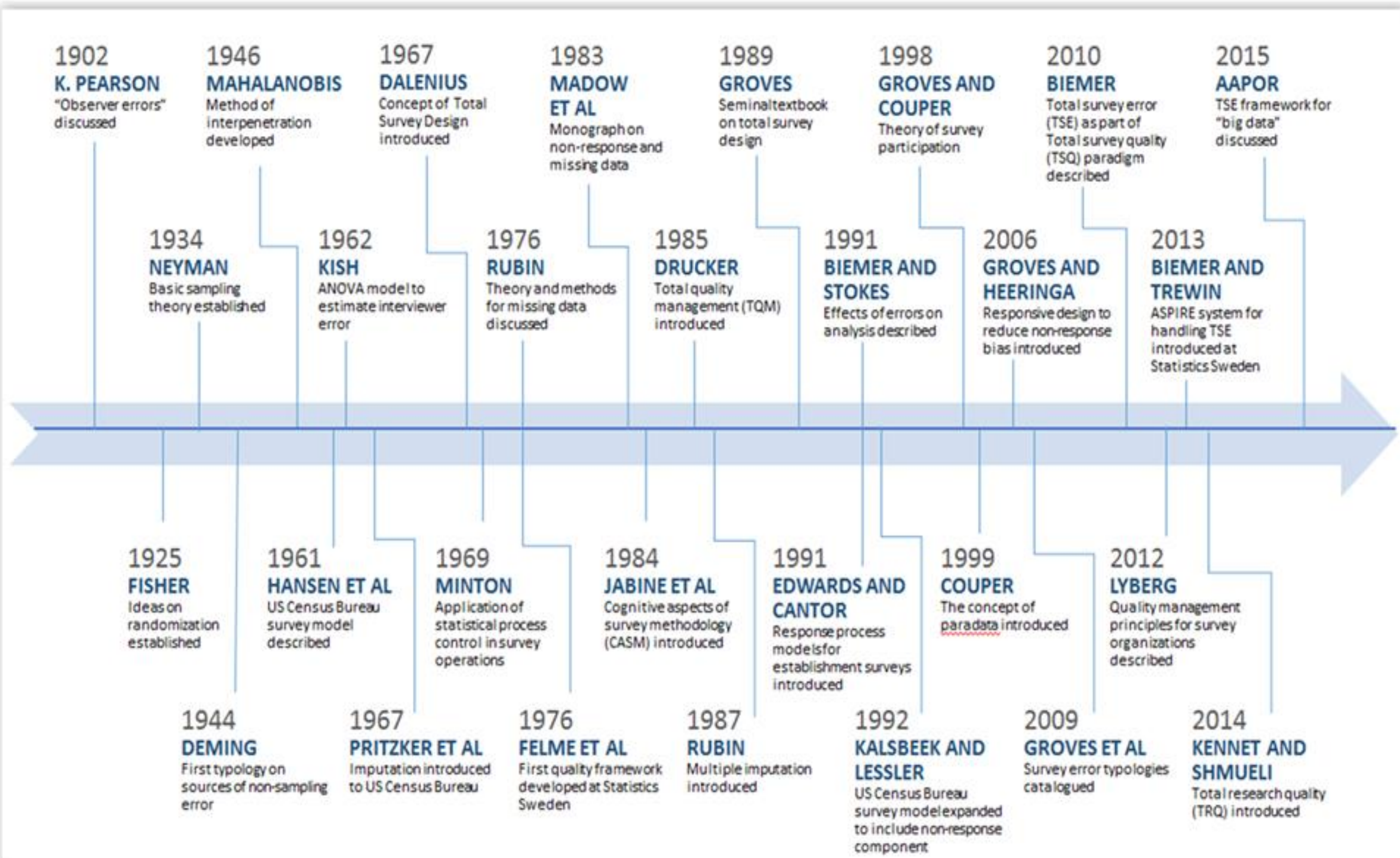
$$Var_{Sam} + Var_{Mea} + Var_{DP} + Var_{Mod}$$

+ covariance terms

- Rather than measuring the MSE, develop stable processes via CBM and SOP, thereby getting a situation where Var becomes an approximation of MSE (error-free processes)
- ITSEW
- Many parallel frameworks

INIZIO

TSE Timeline



Summary of the Evolution of "Total Survey Error"

- Roots in cautioning against sole attention to sampling error
- Key omissions in research and all MSE models are in some sense incomplete
- A user perspective is missing and the complexity does not invite user scrutiny of accuracy
- No producer measures error components routinely
- Survey error research compartmentalized
- Root causes of error often still missing
- Most users think that accuracy is the responsibility of the service provider

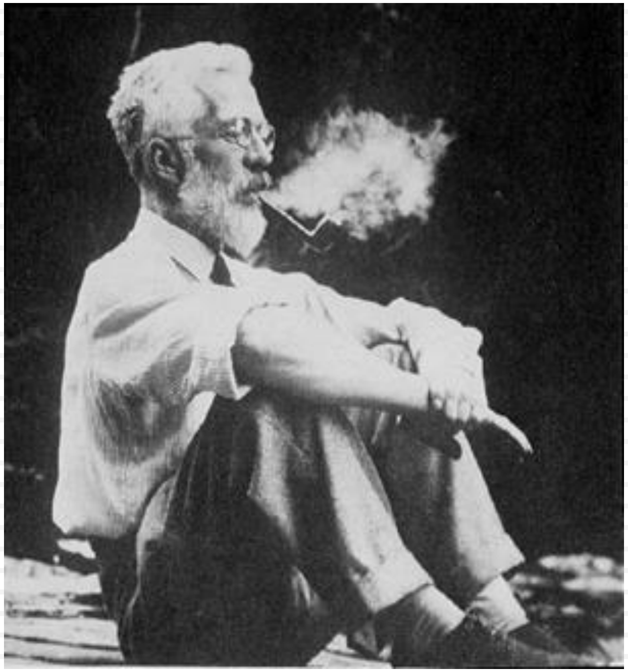
INIZIO

Summary Continued

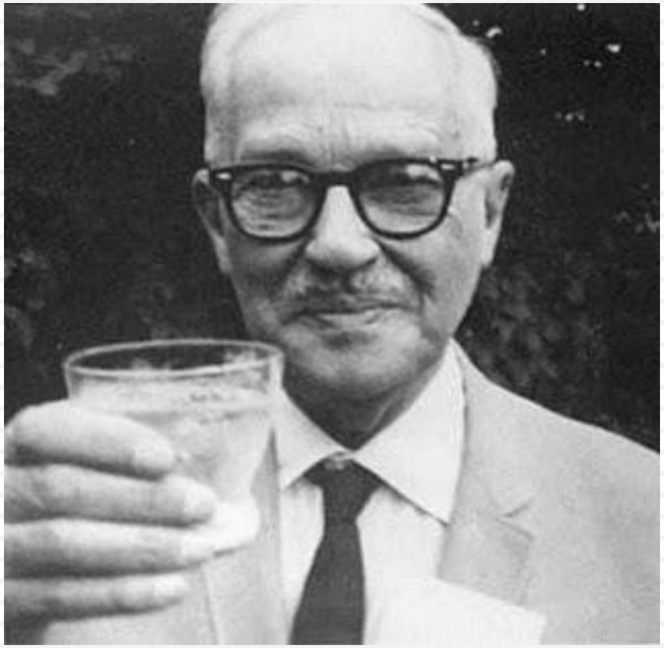
- Great conceptual foundation
- Not known how to allocate resources on TSE improvement versus TSE measurement
- Not known how to allocate resources on prevention, quality control and evaluation
- If fitness for use predominates as a conceptual base how can we study error variation associated with different uses?
- Are standards and design principles a way forward?

A Couple of Giants

Sir Ronald Fisher



Jerzy Neyman



Neyman

- Sampling variance properties of descriptive statistics from probability samples of finite populations in Neyman's landmark paper from 1934
- Inferential properties of his theory required strong assumptions about nonsampling errors, i.e., they had to be very small
- Assumptions were questioned very early but the idea of probability samling had to be promoted

INIZ!O

Morris Hansen

Prasanta Mahalanobis

Tore Dalenius

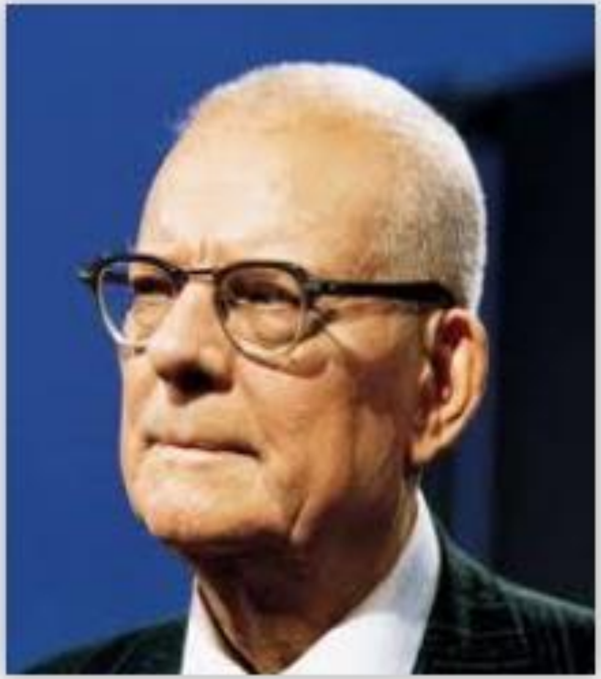
Barbara Bailar

Leslie Kish

Don Rubin

Bill Cochran

Ed Deming



Judy Lessler

Bob Groves

INIZIO

Paul Biemer

INIZIO

What Were the Issues in the Past?

- Nonsampling errors
- Balancing errors and costs
- Design criteria
- The limitations of sampling theory
- Standards
- Similarities between survey implementation and the assembly line

Deming's 13 Factors

The 13 factors that affect the usefulness of a survey

-To point out the need for directing effort toward all of them in the planning process with a view to usefulness and funds available

-To point out the futility of concentrating on only one or two of them

-To point out the need for theories of bias and variability that correlate accumulated experience

1. Variability in response;
2. Differences between different kinds and degrees of canvass;
   (a) Mail, telephone, telegraph, direct interview;
   (b) Intensive vs. extensive interviews;
   (c) Long vs. short schedules;
   (d) Check block plan vs. response;
   (e) Correspondence panel and key reporters;
3. Bias and variation arising from the interviewer;
4. Bias of the auspices;
5. Imperfections in the design of the questionnaire and tabulation plans;
   (a) Lack of clarity in definitions; ambiguity; varying meanings of same word to different groups of people; eliciting an answer liable to misinterpretation;
   (b) Omitting questions that would be illuminating to the interpretation of other questions;
   (c) Emotionally toned words; leading questions; limiting response to a pattern;
   (d) Failing to perceive what tabulations would be most significant;
   (e) Encouraging nonresponse through formidable appearance;
6. Changes that take place in the universe before tabulations are available;
7. Bias arising from nonresponse (including omissions);
8. Bias arising from late reports;
9. Bias arising from an unrepresentative selection of date for the survey, or of the period covered;
10. Bias arising from an unrepresentative selection of respondents;
11. Sampling errors and biases;
12. Processing errors (coding, editing, calculating, tabulating, tallying, posting and consolidating);
13. Errors in interpretation;
   (a) Bias arising from bad curve fitting; wrong weighting; incorrect adjusting;
   (b) Misunderstanding the questionnaire; failure to take account of the respondents' difficulties (often through inadequate presentation of data); misunderstanding the method of collection and the nature of the data;
   (c) Personal bias in interpretation.

Comments on Deming (1944)

- Does include nonresponse, sampling, interviewer effects, mode effects, various other measurement errors, and processing errors
- Omits coverage errors
- Includes nonstatistical notions (auspices)
- Includes estimation step errors (wrong weighting)
- “Total survey error” not used as a term
- Lots of other typologies followed eventually

INIZIO

A Difficult Position

- They had to promote Neyman’s theory
- But his theory basically assumes very small nonsampling errors
- They were in a first-things-first situation
- They promoted vigorous controls hopefully leading to small biases
- They discussed what a Bayesian approach might offer

Lines of Thought I

- “There is as yet no universally accepted ‘survey design formula’ that provides a solution to the design problem (Dalenius 1967)
- That’s why textbooks devote little space to design
- Important to control specific error sources

INIZIO

Lines of Thought II

- The U.S. Bureau of the Census is a statistical factory. The main product is statistical tables (Deming and Geoffrey 1941)
- Concentration on QC of error sources, evaluation, and survey models
- Disentangling the design process

Lines of Thought III

- Hansen-Hurwitz-Pritzker 1967
  - Take all error sources into account
  - Minimize all biases and select a minimum-variance scheme so that Var becomes an approximation of (a decent) MSE
  - The zero defects movement that later became Six Sigma
- Dalenius 1969
  - Total survey design

Kish's Contributions

- The neo-Bayesian view
  - Appreciates the literature by Schlaifer, Ericson, Edwards, Lindman and Savage on Bayesian methods in survey sampling and psychometrics
  - For instance, judgment estimates of measurement biases may be combined with sampling variances to construct more realistic estimates of the total survey error

More From Kish

- Experiments and sample surveys might not be sufficient. Other investigations “collecting data with considerable care and control” but without randomization and probability sampling might be necessary.

Kish's View on Design

- Multipurpose is great from an economical point of view.
- If one principal statistic can be identified that alone can decide the design
- If a small number of principal statistics can be identified a reasonable design compromise is possible
- If statistics are too disparate a joint design might not be possible

Kish Summed Up

- Get a good balance between different error sources
- We need to know how error structures behave under different design alternatives
- Relevant information should be recorded during implementation (paradata)
- Many practical constraints
- The multipurpose nature calls for a compromise

Hansen, Dalenius and Colleagues Summed Up

- One should be guided by common sense, experience and theory
- Design and execution is a management and systems analysis problem
- A survey is an economic production process
- Survey goals must be identified
- Standards must be dynamic
- End the practice that sampling error is viewed as the total error
- They predicted the CASM movement

INIZIO

What Happened?

- Still no “design formula”
- General design principles exist for some areas
- Still a concentration on some error sources more than others
- CASM happened
- We got standards
- The TSE paradigm accepted but has some promotional problems
- Many of the early thoughts were just that, very little practice, but still useful

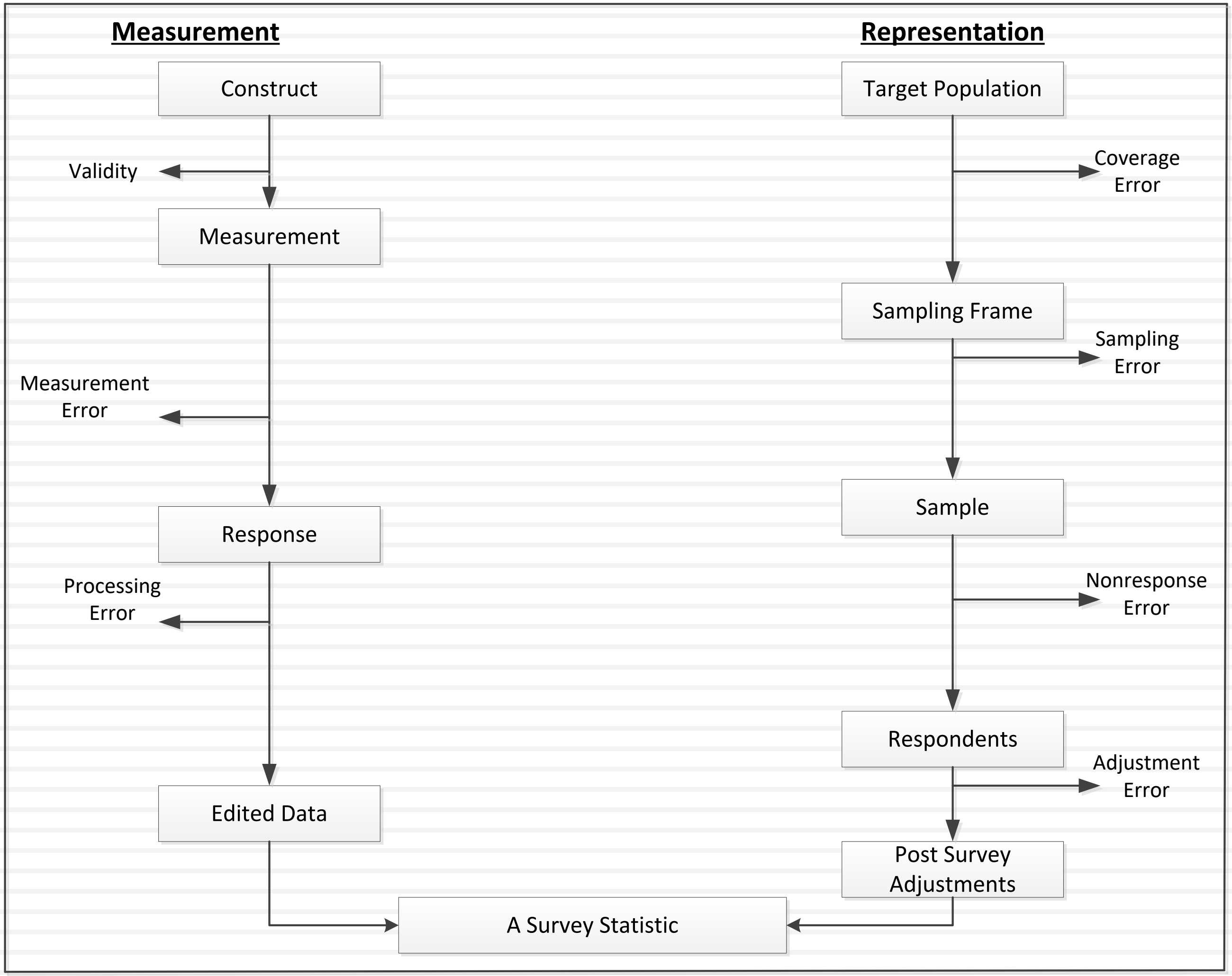
INIZIO

The Rise and Fall and Rise of the TSE Perspective

- Estimating MSE was complicated
- Basic reinterview schemes did not work as intended (Biemer and Forsman 1992)
- Dillman’s concerns about the lack of innovation in the U.S. Federal statistical system, 1996
- Platek and Särndal’s discussion paper in JOS Can a Statistician Deliver? Raising their concerns about survey quality, 2001
- ITSEW starts in 2005 and here we are in Bergamo

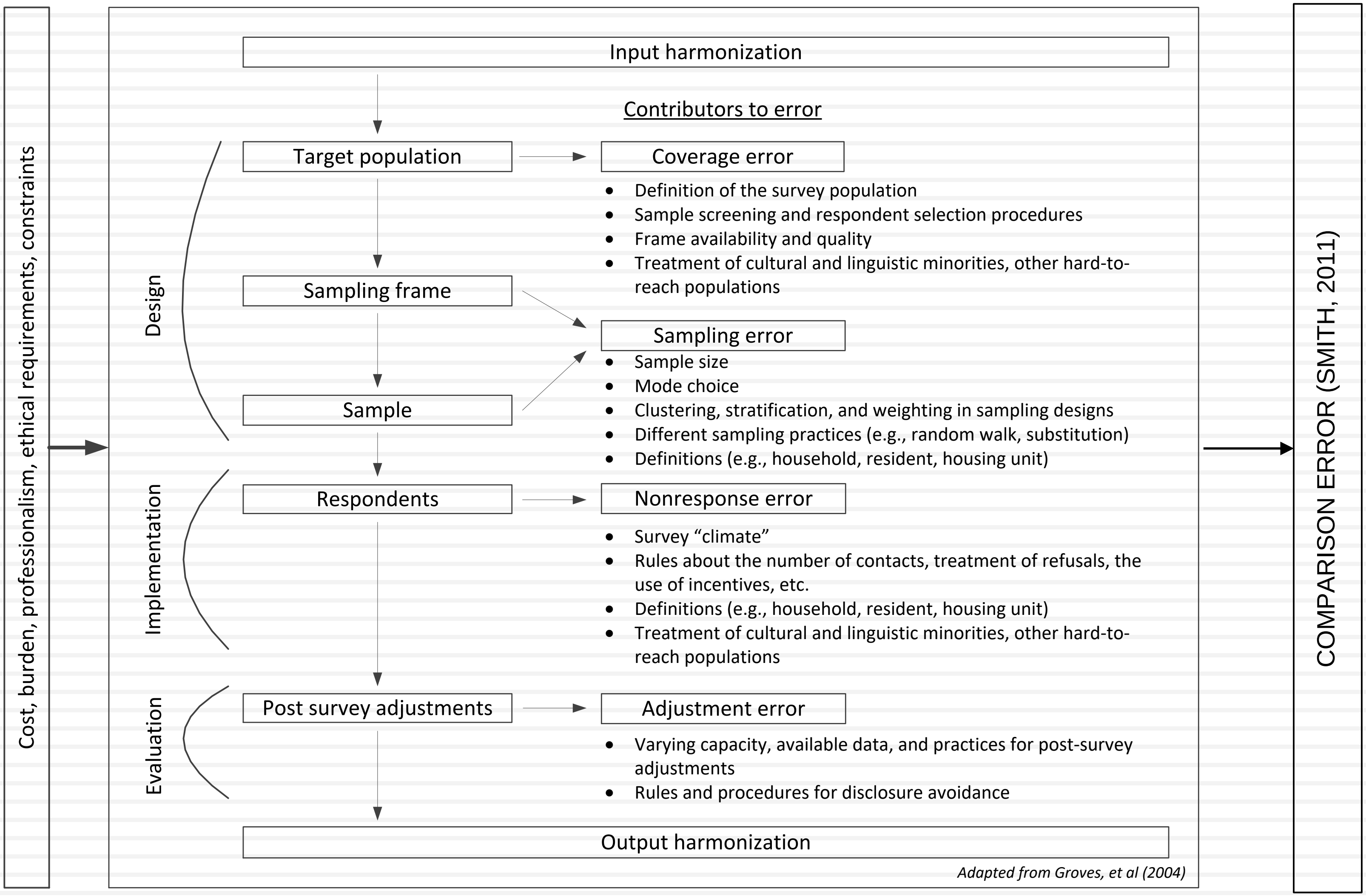
INIZIO

Traditional Total Survey Error (Groves et al., 2009)

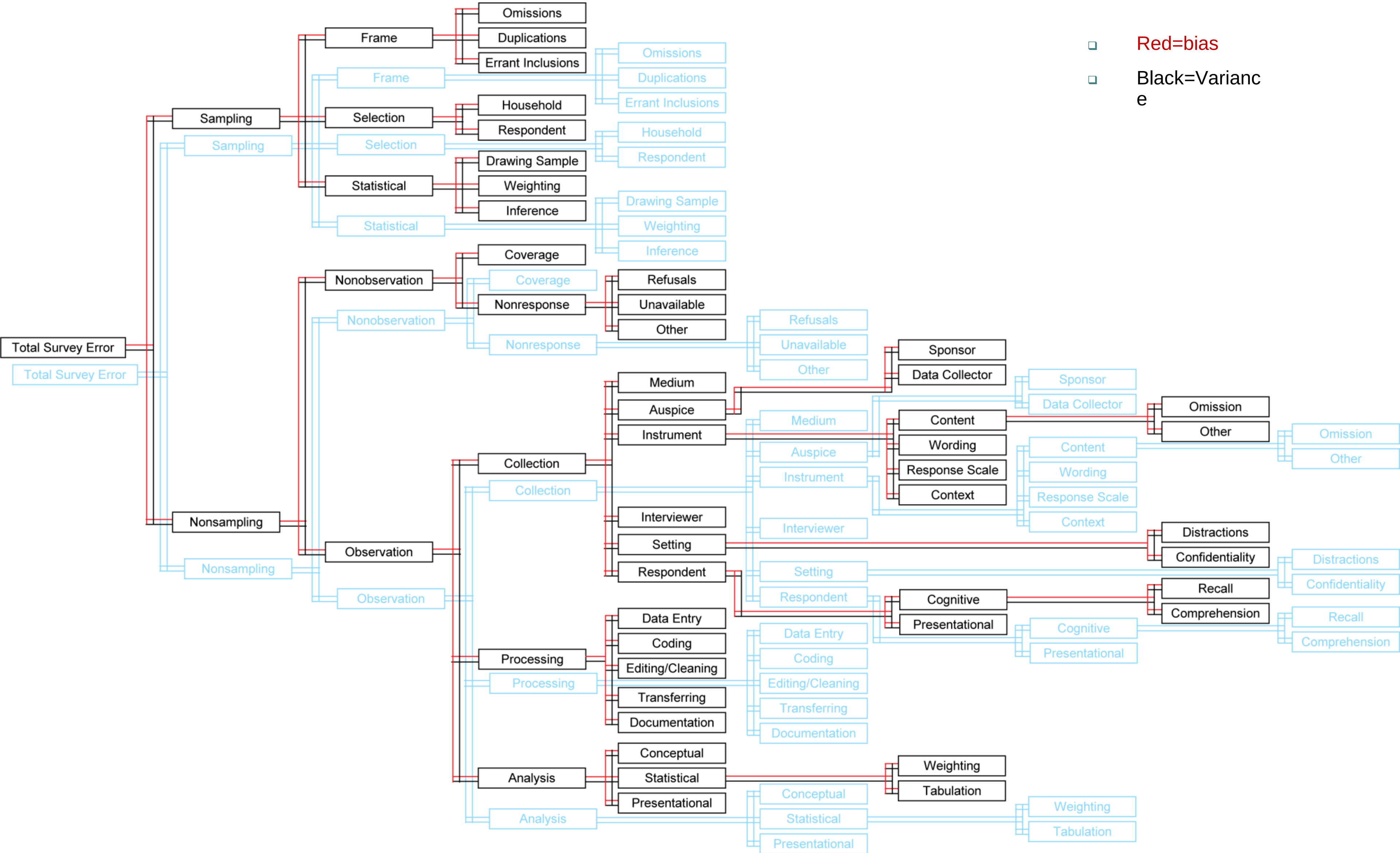


Modified from Groves et al. (2004)

Total Survey Error Representation in a Cross-cultural Context



Total Survey Error: Comparison Error



- Red=bias
- Black=Variance

Smith (2011)

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.

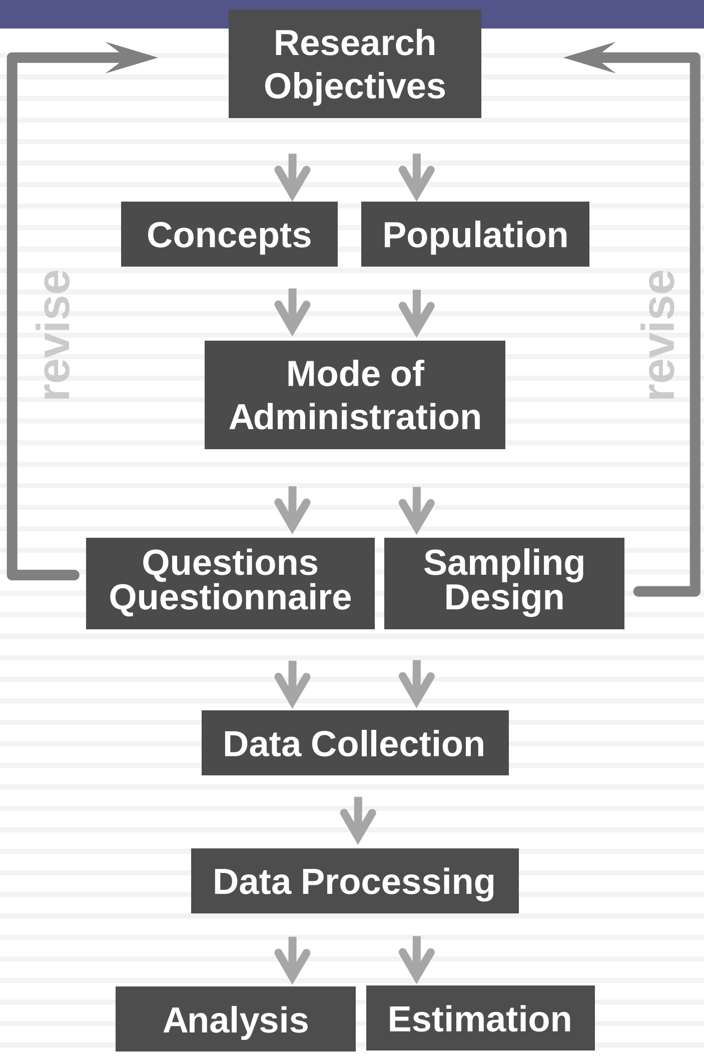
Third, it can set a methodological research agenda for study error and for the design of experiments and other studies to fulfill that agenda.

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.

Smith (2011)

The Survey Process



Specification Error

Frame Error

Measurement Error

Nonresponse Error

Sampling Error

Data Processing Error

Modeling/Estimation Error

INIZIO

Generalized TE Framework

Total Error = Sample Recruitment Error
+ Data Encoding Error

Sample Recruitment Error is a generalization of the concept of representation error

Data Encoding Error is a generalization of the concept of measurement error

Biemer (2019)

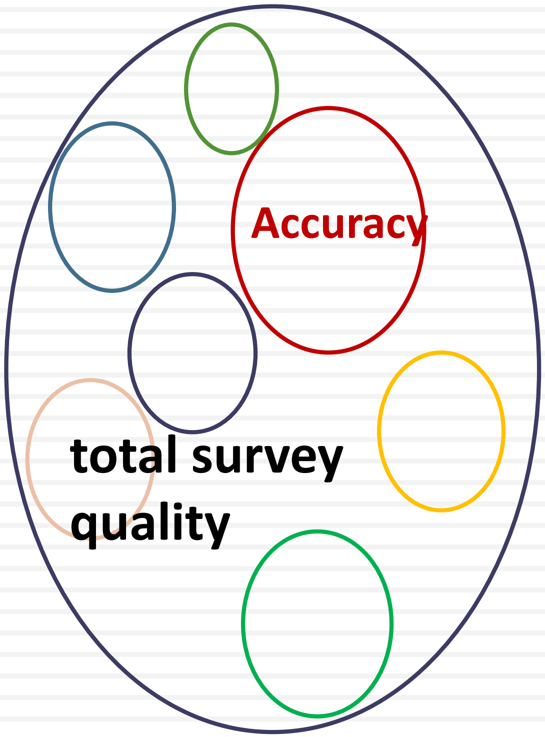
InfoQ

- Total study quality
  - Given a stated goal, InfoQ can be assessed at the design stage, at the release stage or before embarking on secondary analyses
  - Can discover faulty translation from statistics to domain
  - Useful when integrating official statistics with other data sets
- InfoQ is a step away from the one-size-fits-all frameworks
- Great description in JRSS, A 2014

INIZ!O

Focus on Accuracy → The Absence of Error

Accuracy is but one dimension of the larger concept of ***total survey quality****



*Statistics Canada, Statistics Sweden, ABS, IMF, Eurostat, OECD and others

INIZIO

Other Dimensions of Total Survey Quality Include, for Instance, …

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 is clear; meta-data are well-managed

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

The Quality Concept in Surveys

- Accuracy
- Framework (accuracy, relevance, timeliness, accessibility, coherence, comparability, completeness, and other dimensions)
- A framework can be seen as a quality vector

Problems:

- No users involved
- A quality ”index” is not realistic
- Dimensions are in conflict
- Accuracy is still number one

INIZIO

Quality Frameworks

- Statistics Canada, Statistics Sweden, ABS, IMF, Eurostat, OECD and more
- Typical dimensions include relevance, accuracy, timeliness, coherence, comparability, accessibility
- Any client requirement can be part of a quality vector
- Dimensions are in conflict
- Accuracy is difficult to beat as *the* main dimension (two exceptions are exit polls and international surveys)

INIZIO

A SURVEY'S DEADLY ERRORS

Total Survey Error

Sampling Error

Due to selecting
a sample instead of
the entire population.

Nonsampling Error

Errors due to
mistakes or system
deficiencies.

Sampling Error

Sampling scheme

Sample size

Estimator choice

- **Stratification**
- **Clustering**
- **Selection probabilities**
- **Sampling phases**

Sampling Error

Sampling scheme

Sample Size

Estimator choice

- **Overall *n***
- **Effective *n***
- **Sample size allocation**

Sampling Error

Sampling scheme

Sample size

Estimator choice

- Simple
- Use of auxiliary information
- Model-based
- Model-assisted

Nonsampling Error

Specification Error

Frame Error

Measurement Error

Nonresponse Error

Data Processing Error

Modeling/Estimation Error

- **Concept vs. operational definition**
- **Concept vs. question**
- **Concept vs. information system**

Nonsampling Error

- Specification Error
- **Frame Error**
  - Omissions
  - Inclusions
  - Duplications
  - Content Errors
- Measurement Error
- Nonresponse Error
- Data Processing Error
- Modeling/Estimation Error

Nonsampling Error

Specification Error

Frame Error

Measurement Error

Nonresponse Error

Data Processing Error

Modeling/Estimation Error

- **Respondent**
- **Interviewer**
- **Questionnaire**
- **Mode of Data Collection**
- **Information System**
- **Setting**

Nonsampling Error

Specification Error

Frame Error

Measurement Error

Nonresponse Error

- **Unit**
- **Within Unit**
- **Item**
- **Panel attrition**

Data Processing Error

Modeling/Estimation Error

Nonsampling Error

Specification Error

Frame Error

Measurement Error

Nonresponse Error

Data Processing Error

- **Data Entry**
- **Editing**
- **Coding**
- **Weighting**

Modeling/Estimation Error

Nonsampling Error

Specification Error

Frame Error

Measurement Error

Nonresponse Error

Data Processing Error

Modeling/Estimation Error

- **Weight adjustments**
- **Imputation**
- **Error reduction models**
- **Analytic models**

Risk of Variance and Bias by Error Source for a Typical Survey

MSE Component	Variance	Bias
Sampling error	High	Low
Specification error	Low	High
Frame error	Low	High
Measurement error	High	High
Nonresponse error	Medium	High
Data processing error	High	High
Modeling error	High	High

Applications

- ASPIRE evaluations
- survey redesign
- quality reporting and profiling
- error mitigation, e.g.
  - adaptive total design
  - nonresponse bias reduction
- data analysis – sampling error, missing data, measurement error, complex design

Two Routes to Handling Survey Errors

1. Get an estimate of MSE so that we get confidence or other intervals that we can trust
2. Try to develop and use methods that are almost error-free so that the estimated variance becomes an approximation of the MSE

This is one justification for quality improvement work

Quality Assurance and Quality Control

- QA is defined as a set of activities whose purpose is to demonstrate that an entity meets all quality requirements

- QC is defined as a set of activities whose purpose is to ensure that all quality requirements are met

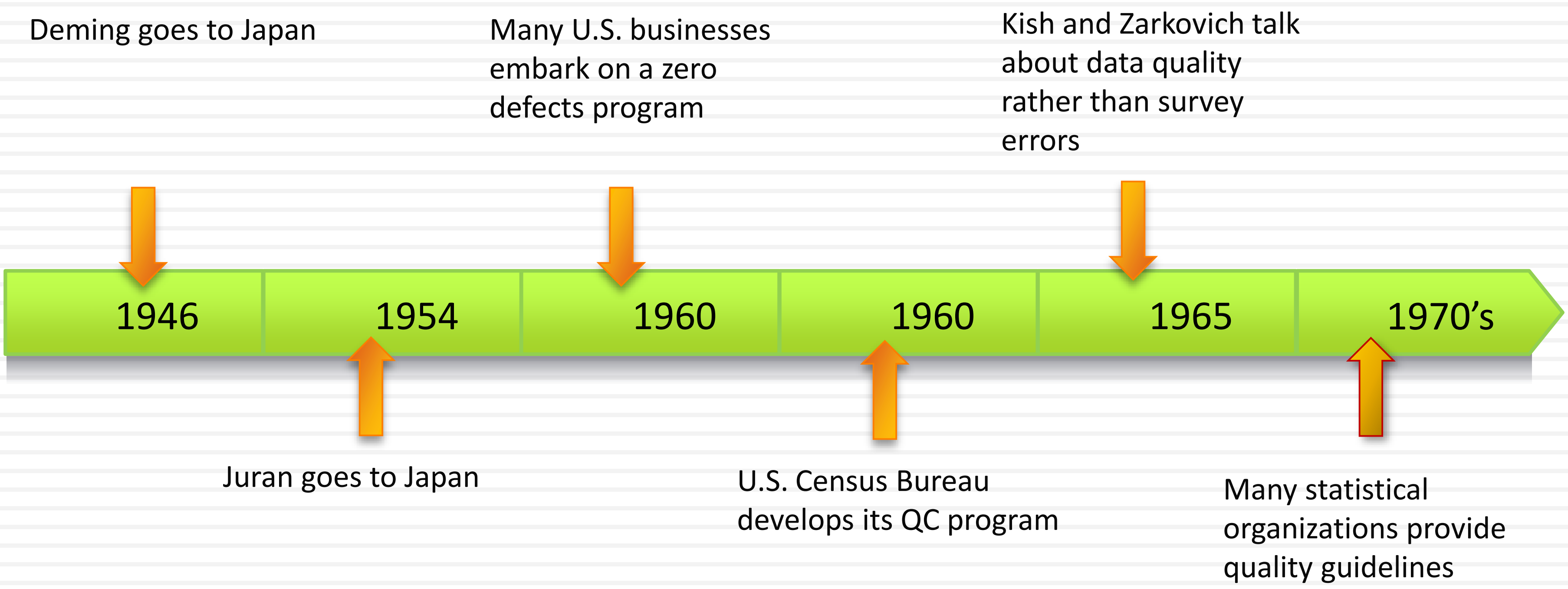
Quality Management Philosophies in Statistical Organizations

- Main triggers in the 90's were
  - the need for greater effectiveness and enhanced user contacts
  - the general hype around TQM
  - the Deming and Juran factors
  - new or rediscovered tools found in the QM toolbox
- A gradual merging of QM and TSE ideas has taken place

Timeline

- 1875 – Taylor introduces scientific management
- 1900-30 – Taylor's ideas used at Ford and Benz
- 1920 – Fisher develops experimental design
- 1924 – Shewhart develops the control chart
- 1940 – The U.S. war department develops a guide for analyzing process data
- 1944 – Dodge and Romig's theory for acceptance sampling

Timeline



Deming goes to Japan
1946
1954
Juran goes to Japan
Many U.S. businesses embark on a zero defects program
1960
1960
U.S. Census Bureau develops its QC program
Kish and Zarkovich talk about data quality rather than survey errors
1965
1970's
Many statistical organizations provide quality guidelines

Timeline

- **1975** – The TQM framework is launched
- **1976** – First published quality framework for a statistical organization that goes beyond relevance and accuracy
- **1987-89** – Launching of ISO9000, Malcolm Baldrige Award, Six Sigma, and EFQM
- **1990's** – Many statistical organizations start working with excellence models
- **2001** – Eurostat leadership group on quality organizes the first conference on QM in official statistics
- **2007** – Business architecture ideas enter the survey world

The Quality Movement's Impact on Statistical Organizations

- An extended quality concept
- The recognition of a user/customer/client
- Awareness of competition
- The importance of adopting a process view
- Business excellence models (EFQM, Malcolm Baldrige) and other models (TQM, Six Sigma, Lean, BPR, Balanced Scorecard)
- Tools and metrics for handling process variability, risks, priorities, and waste
- Efficient work processes
- Continuous quality improvement

INIZIO

The Process View

- Process is a series of actions or steps towards achieving a particular end
- Process quality is an assessment of how far each step meets defined criteria
- Process variables are factors that can vary with each repetition of the process
- Key process variables are factors that have a large effect on process end result
- Observations of process variables result in paradata

INIZIO

Assuring and Controlling Quality on Three Levels

Quality Level	**Main stake-holders**	**Control instrument**	**Measures and indicators**
Product	User, client	Product specs, SLA, evaluation studies, frameworks, standards	Frameworks, compliance, MSE, user surveys
Process	Survey designer	SPC, control charts, acceptance sampling, risk analysis, CBM, SOP, paradata, checklists, verification	Variation via control charts, other paradata analysis, outcomes of evaluation studies
Organization	Agency, owner, society	Excellence models, ISO, CoP, reviews, audits, self-assessments	Scores, strong and weak points, user surveys, staff surveys

The User

In Place:

- The principle of openness (OMB 1978)
- Responsibility to inform users (many agencies in the 70’s)
- Dissemination procedures
- Customer satisfaction and image surveys
- Councils and service level agreements

Problems:

- How should quality information be communicated?
- What quality information should be communicated?
- How do we distinguish between different kinds of users?
- How do users and producers use quality information and metadata?
- How do producers and users collaborate on fitness for use? (ABS)

INIZIO

Understanding Variation (1)

Common cause variation

- Common causes are the process inputs and conditions that contribute to the regular, everyday variation in a process
- Every process has common cause variation
- Example: Percentage of correctly scanned data, affected by people's handwriting, operation of the scanner…

INIZIO

Understanding Variation (2)

Special cause variation

- Special causes are factors that are not always present in a process but appear because of particular circumstances
- The effect can be large
- Special cause variation is not present all the time
- Example: Using paper with a colour unsuitable for scanning

Action

- Eliminate special cause variation
- Decrease common cause variation if necessary
- Do not treat common cause as special cause

Roots of Paradata

- Traditional global ones such as error rates (since 1940)
- Hansen-A signal system 1960's
- U.S. Census Bureau's process control in the 60's and 70's
- Keystroke files in CATI
- The 1998 ASA session in Dallas, Mick Couper
- Rapid development last 15 years

INIZIO

Importance of Paradata (1)

- Continuous updates of progress and stability checks (monitoring)
  - Control charts, standard reports, dashboards
  - Managers choose to act or not to act
  - Early warning system
- Input to long-run process improvement of product quality
  - Analysis of special and common cause variation
- Input to methodological changes
  - Finding and eliminating root causes of problems
  - Research

Importance of Paradata (2)

- Responsive designs
  - Simultaneous monitoring of paradata and regular survey data to improve efficiency and accuracy
- Input to organizational change
  - E.g., centralization, decentralization, standardization
- Quality profiles, client communication, public use paradata files, inference, picturing quality over time

Plan for Continuous Improvement (of a Product) Marker and Morganstein 1997

- Identify critical product characteristics
- Develop a process flow map
- Determine key process variables
- Evaluate measurement capability
- Determine stability of critical processes
- Determine process capability
- Establish a system for continuous process monitoring

INIZIO

Standards and Guidelines

- Early standards on presentation of survey errors
- Standards for quality reporting (ESS, Statistics Canada and others)
- European statistics Code of Practice
- Quality indicators
- UN fundamental principles
- ISO 20252
- US OMB standards

INIZ!O

Statistical Business Process Models

- Generic model for statistics production
- Based on architectural principles
  - Increased centralization
  - Smallest possible number of processes and systems
  - Minimal corporate toolkit
  - Staff proficiency
  - Elimination of rework
  - Focus on core business
  - No overlapping or unclear mandates

CTQs and Metrics for Costs, Production, and Timeliness

CTQs	Process Metrics
• Maximizing interviewing efficiency • Maximize effectiveness of refusal conversions attempts • Complete call histories accurately and completely • Minimize hours per completed screener • Minimize hours per completed interview • Maintain planned costs per quarter • Maintain planned schedule for sample completion per quarter	• Cost per interview • Dollars spent vs. dollars budgeted by Interviewer • Cost breakdown (by phase and overall) • Number of cases interviewed (actual vs. budgeted) • Calls per hour (actual vs. expected) • Refusal conversion rates by interviewer • Hours charged (actual vs. expected) • Level of effort per case by interviewer and overall • Hours per completed screener • Hours per completed interview

From P. Biemer, QMMS II Design Seminar, 2010

Audits and Self-assessments

- Audits require an excellence model, a standard and clear organizational goals as benchmarks
- Self-assessments require clear quantifiers and objective reviewers

INIZIO

The 2010 Marriage Following a 20 year Engagement

- TSE+QM=Total Survey Quality (TSQ)
- Biemer defined the TSE paradigm as part of a larger design strategy that seek to optimize TSQ
- The paradigm’s four pillars are design, implementation, evaluation and the effects of errors on the analysis

INIZIO

Measuring Quality

- Direct estimates of TSE are difficult to obtain
- Smaller error component evaluations are more realistic and also necessary
- General indicators calculated directly from the survey data are few and sometimes less useful
- Specific metrics for process characteristics that are critical to quality
- Audits and self-assessments
- Quality profiles
- ASPIRE, TSQ, InfoQ

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

Quality Criteria Applied to Each Error Source

Criteria by Error Source	Ratings by Criterion
1. Knowledge of risks	Poor (●)
2. Communication with users	Fair (◓)
3. Compliance with standards and best practices	Good (○)
4. Available expertise	Very Good (◒)
5. Achievement toward risks mitigation and/or improvement plans	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

INIZIO

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

INIZIO

Example: LFS Accuracy Ratings for 2012

Error source	Score round 1	Score round 2	Knowledge of Risks	Communication to Users	Available Expertise	Compliance with standards & best practices	Plans towards mitigation of risks	Risk to data quality
Specification error	66	70	◒	[illegible]	[illegible]	◒	◒	**L**
Frame error	58	58	◒	[illegible]	[illegible]	[illegible]	[illegible]	**L**
Non-response error	56	52	○	○	○	[illegible]	[illegible]	**H**
Measurement error	50	56	○	○	○	[illegible]	[illegible]	**H**
Data processing error	54	62	○	○	◒	◒	◒	**M**
Sampling error	70	78	◒	◯	◒	◒	◯	**M**
Model/estimation error	50	60	○	○	○	◒	◒	**M**
Revision error	N/A	N/A	N/A	N/A	N/A	N/A	N/A	**N/A**
Total score	**56,4**	**60,9**						

Blue indicates deterioration in quality

Pink indicates improvement in quality

INIZIO

ASPIRE As We Know It

- Response to severe pressures from the Swedish government whether quality is going up or down for some specific key products
- Previous evaluation exercises have used internal teams resulting in underreporting of problems
- ASPIRE results can be understood by users
- Combines evaluation, risk thinking and CQI
- It’s been a hit (quality up, awareness up, interest in improvement up, government happy)
- The reviewers have been able to teach product workers about error structures and how to deal with them

INIZIO

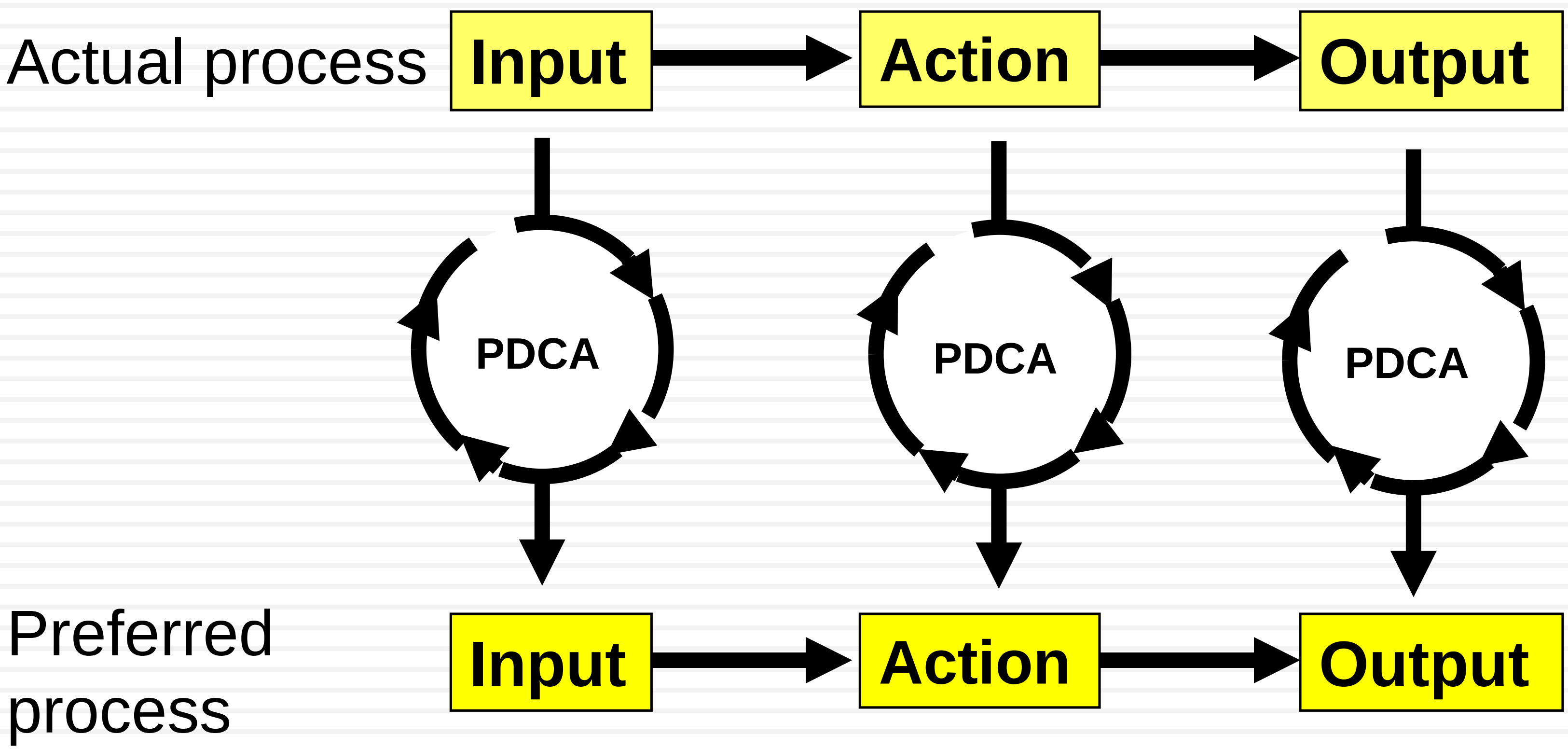
Cautions Regarding ASPIRE

- Reviewers need to be very skilled and experienced
- The ASPIRE process is impressive but can hardly be exported to other less key products
- Can we find a model that can help also the other 100 products continuously improve?
- But do not develop a purely objective ASPIRE light
- And do not believe in the myth about the good example being spread like ripples
- Adding dimensions beyond accuracy means a new set of tradeoffs that need to be handled

INIZIO

How Is Survey Quality Achieved?

Continuous quality improvement (CQI)



Current Status of Survey Quality

- The survey community is moving in a more streamlined and cost-effective direction
- The pace is slow and extensive collaboration is rare (compare with air travel)
- A majority of users think that the service provider is responsible for accuracy
- Many service providers ignore or play down certain error sources leading to overstated confidence levels and biased estimates
- Training is lagging behind
- Quality reporting a la one-size-fits-all is not working for most users
- Competition from new actors cannot be ignored

INIZIO

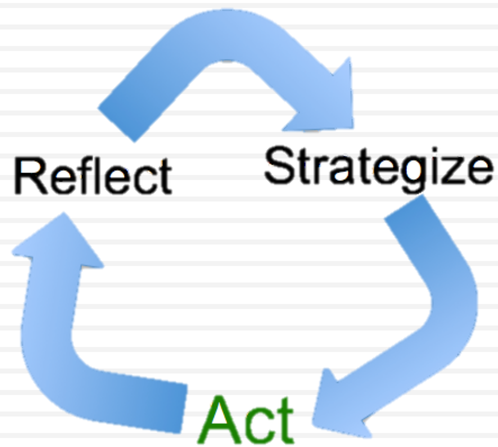
Where To Go

- Use excellence models to improve organizations and processes
- Use the modern quality management tools and principles that are based on statistical methods and therefore especially suitable for statistical organizations
- Adopt the philosophy of continuous quality improvement with a user perspective
- Strive for error-free processes so that Var approximates MSE
- Innovative TSE assessments
- Find a model for international collaboration that can generate common standards for statistics production
- Develop a structured international competence development program for service providers
- Enter the new survey landscape

INIZIO

STRATEGIES FOR APPLYING THE TSE PARADIGM

Lecture 4



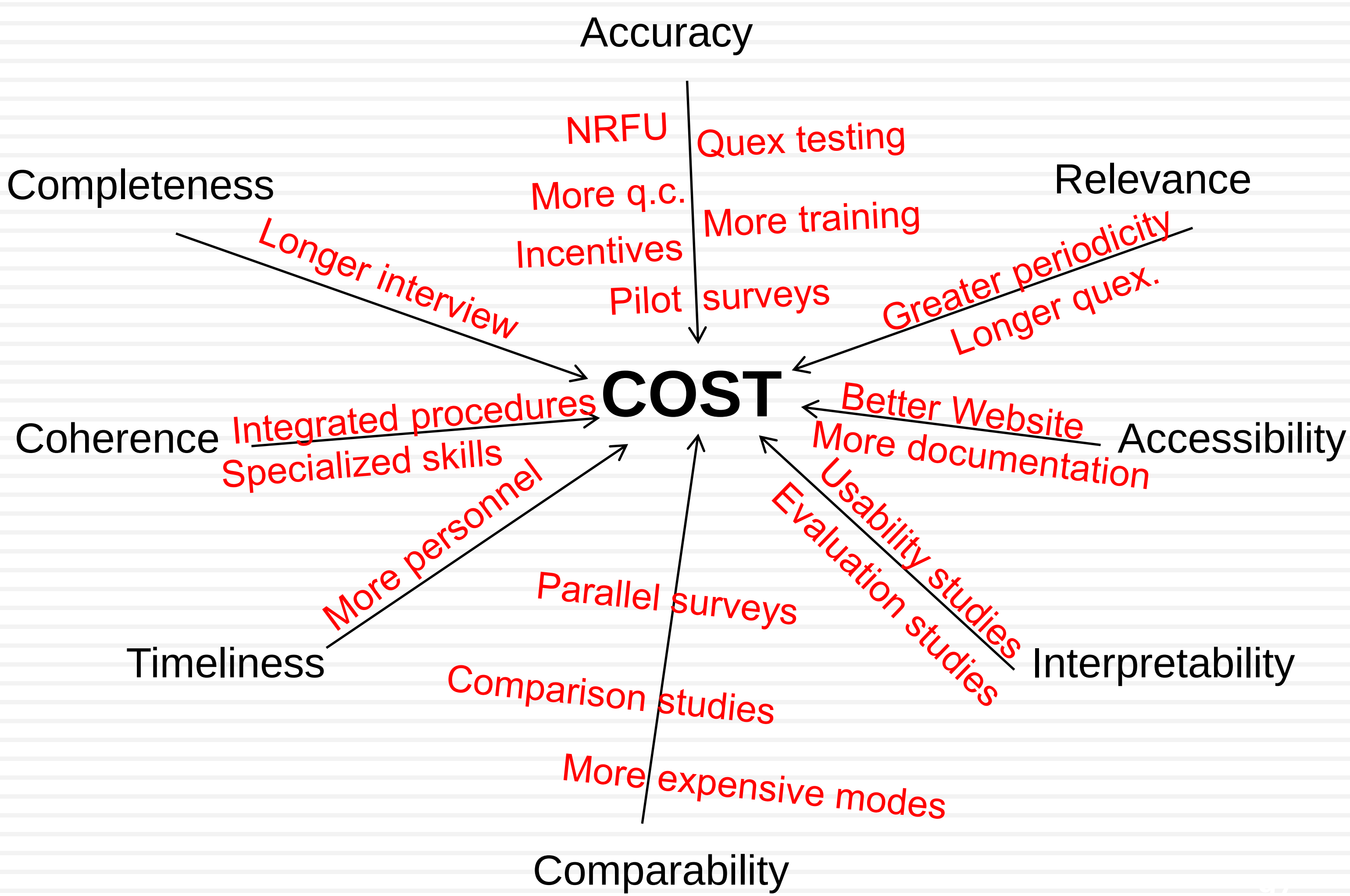
Goals of the TSE Paradigm

- Surveys should be designed and implemented to maximize estimator accuracy within budget constraints.
- But how when…
  - survey budgets are severely constrained,
  - data must be produced and disseminated in a timely fashion,
  - public interest in participating in surveys has been declining world-wide for years, and
  - even when participation is obtained, responses may be inaccurate.
- This is the challenge for survey research in the 21st century

Some Decisions in the Design of Surveys

- How should costs be allocated across the quality dimensions?
- How should costs be allocated across the stages of the survey process to maximize accuracy?
  - Which components of the MSE should drive the design?
- Which estimate(s) should be optimized in the design?
- What MSE and cost components should be monitored during data collection?

Quality Impacts on Costs



Strategies for Identifying Largest MSE Components

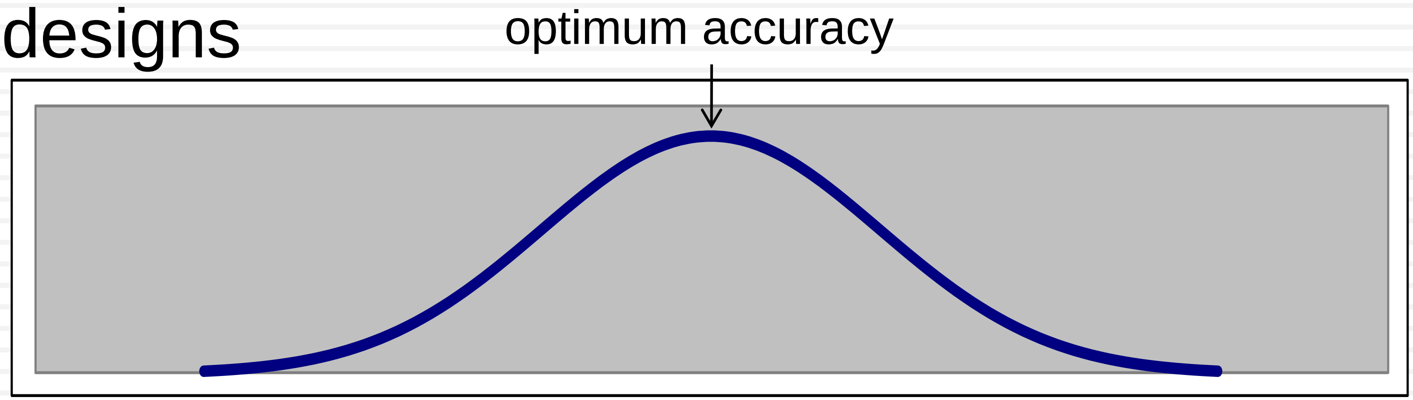
- Consult
  - The survey methodology literature
  - Results of previous survey implementations
  - Prior experience and expert opinions
- Rules of thumb for gauging large or small variances and bias
  - Compare relative bias for coverage and nonresponse with relative standard error of the key estimates
  - Include interviewer deff in the sampling deff when optimizing sampling designs

Design Optimization Strategy

- Optimize for key survey variable if one exists
- Otherwise, optimize for key statistics using a compromise design

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

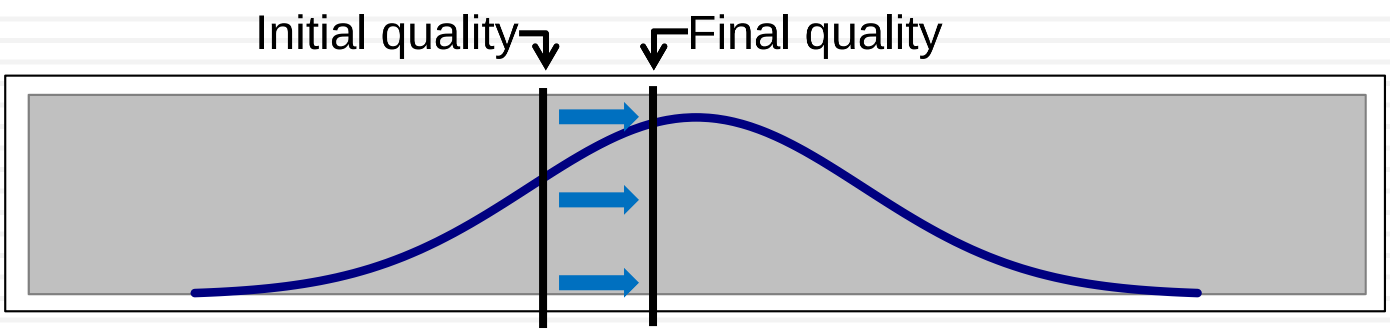
- Compile information on TSE (e.g., quality profiles)
- Identify major contributors to TSE
- Allocate resources to control these errors
  - Use results from the literature and other similar surveys to guide the design
- Develop an effective process for modifying the design during implementation to achieve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

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
- Responsive design
- Six Sigma
- 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's DMAIC Strategy

- ***D**efine* the problem.
- ***M**easure* key aspects of the process and collect relevant data.
- ***A**nalyze* the data to determine root causes of the problem.
- ***I**mprove* the process based upon results from the data analysis.
- ***C**ontrol* the process by continuously monitoring metrics from the process.

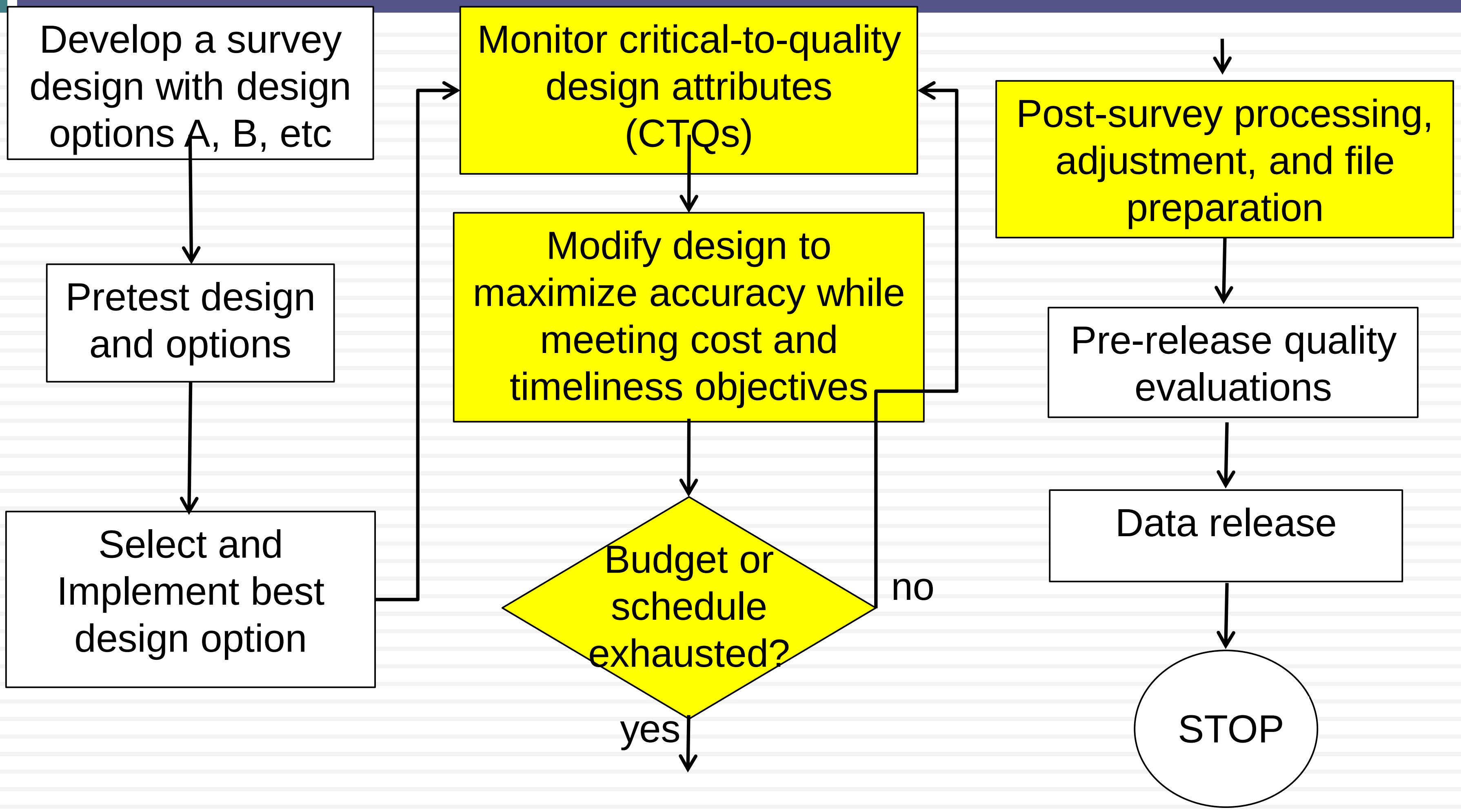
Strengths of Six Sigma

- Provides a systematic, highly effective approach for quality improvement (DMAIC).
- Focuses on attributes of a process that are most important to the client.
- Emphasizes decision making based on data analysis.
- Strives for verifiable and sustainable improvements for both costs and quality.
- Contains a rich set of techniques and tools for monitoring, controlling, and improving a process.

Weaknesses of Six Sigma

- Can be expensive to implement.
- Achieving 3.4 defects per million opportunities is an impossible goal for many survey processes.
- Often requires data that do not exist and cannot be obtained affordably.
- Terminology and some techniques are too business and manufacturing oriented. This obscures its applicability to survey work.
- Uses a lot of jargon.

Six Sigma Focuses Primarily on These Activities



Definition: Responsive Design

Survey designs that:

1. Pre-identify a set of alternative features potentially affecting costs and errors of statistics
2. Identify a set of indicators of the cost and error properties of those feature
3. Monitor indicators in initial stages of data collection
4. Alter the active features of the survey based on cost/error tradeoff decision rules
5. Combine data from separate phases into a single estimator

INIZIO

Adaptive Total Design and Implementation

- 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

CTQs and Metrics for Observation Quality

CTQs

- Detect/control post-survey measurement errors
- Identify/repair problematic survey questions
- Detect/control response errors
- Minimize interviewer biases and variances

Process Metrics

- CARI results by interviewer and overall
- Interviewer exception report
- Missing data item frequency by interviewer
- Replicate measurement analysis summary
- Interview length by interviewer
- CARI refusal rate by FI, by phase

Special Versus 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

Chart of Screening Response Rates by County

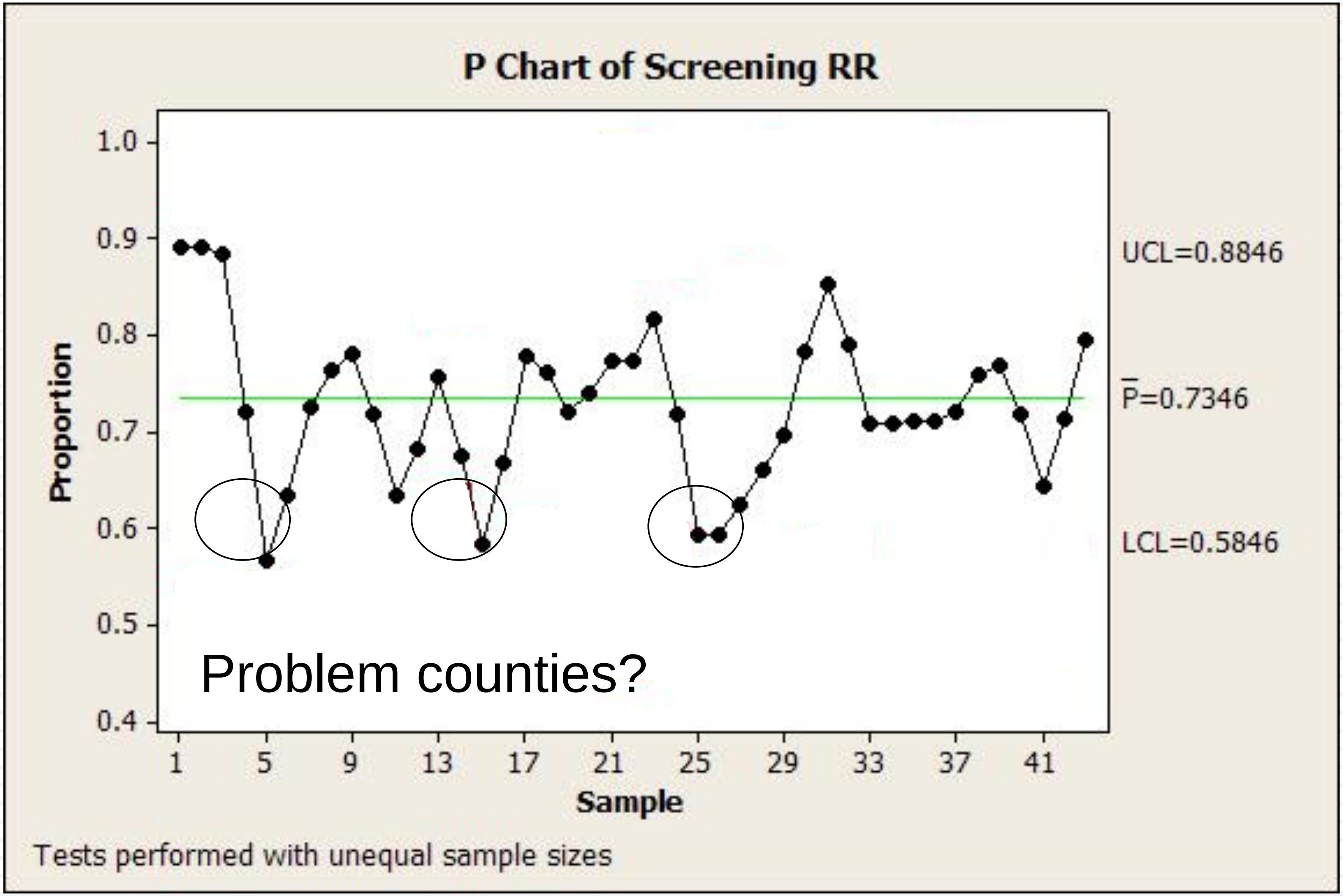
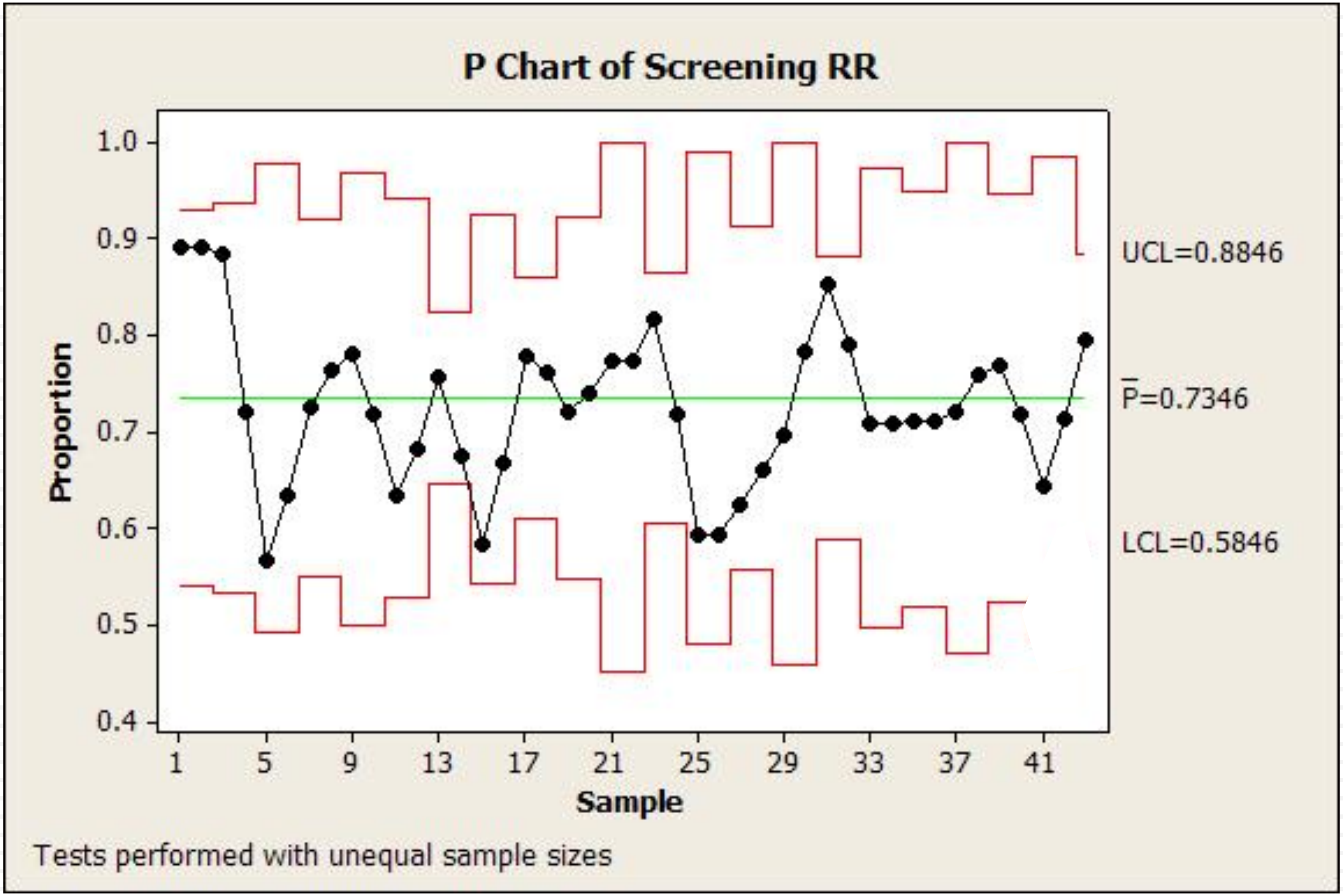
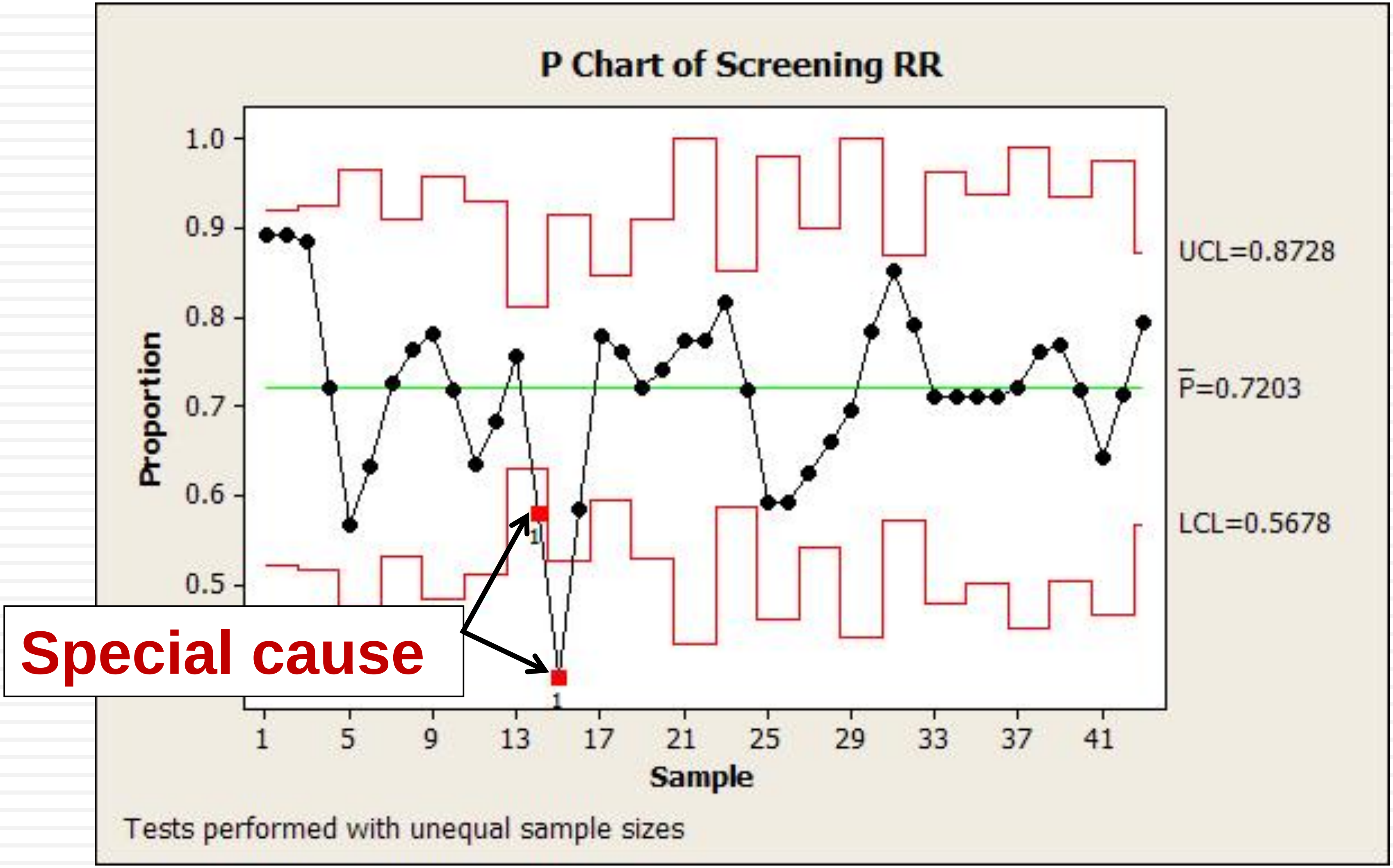


Chart of Screening Response Rates by County

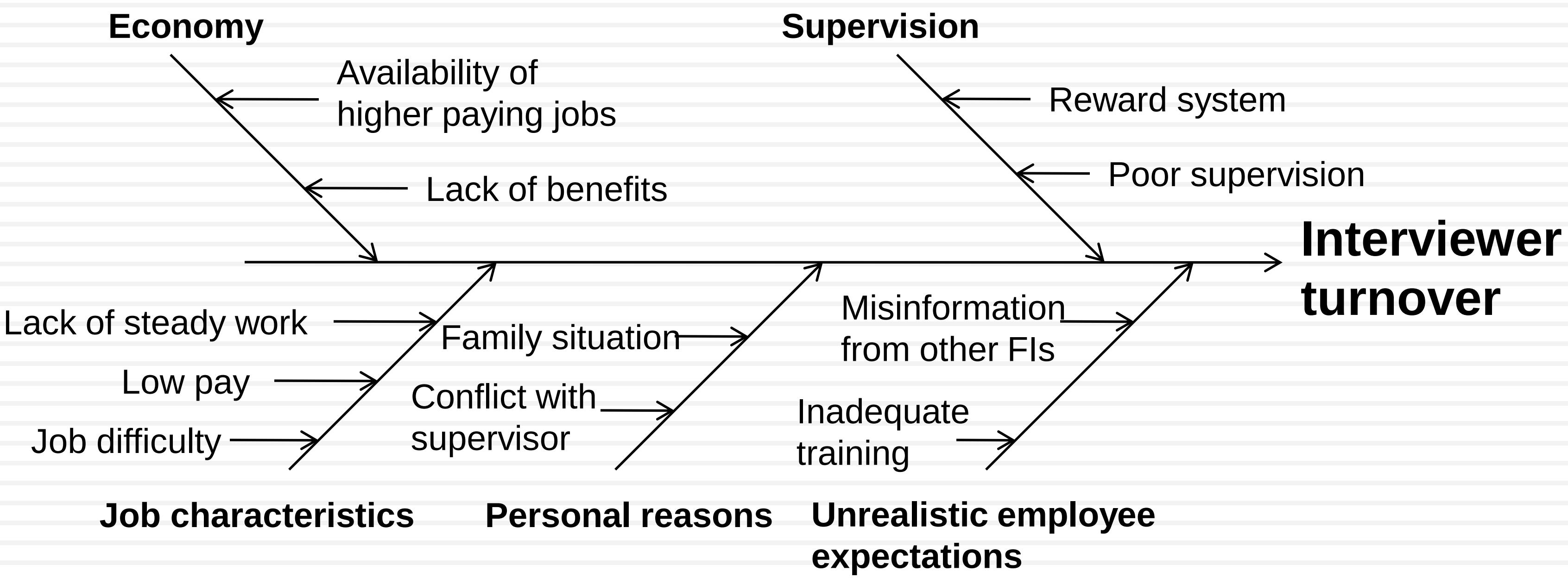


Process Control Chart with More Extreme Values



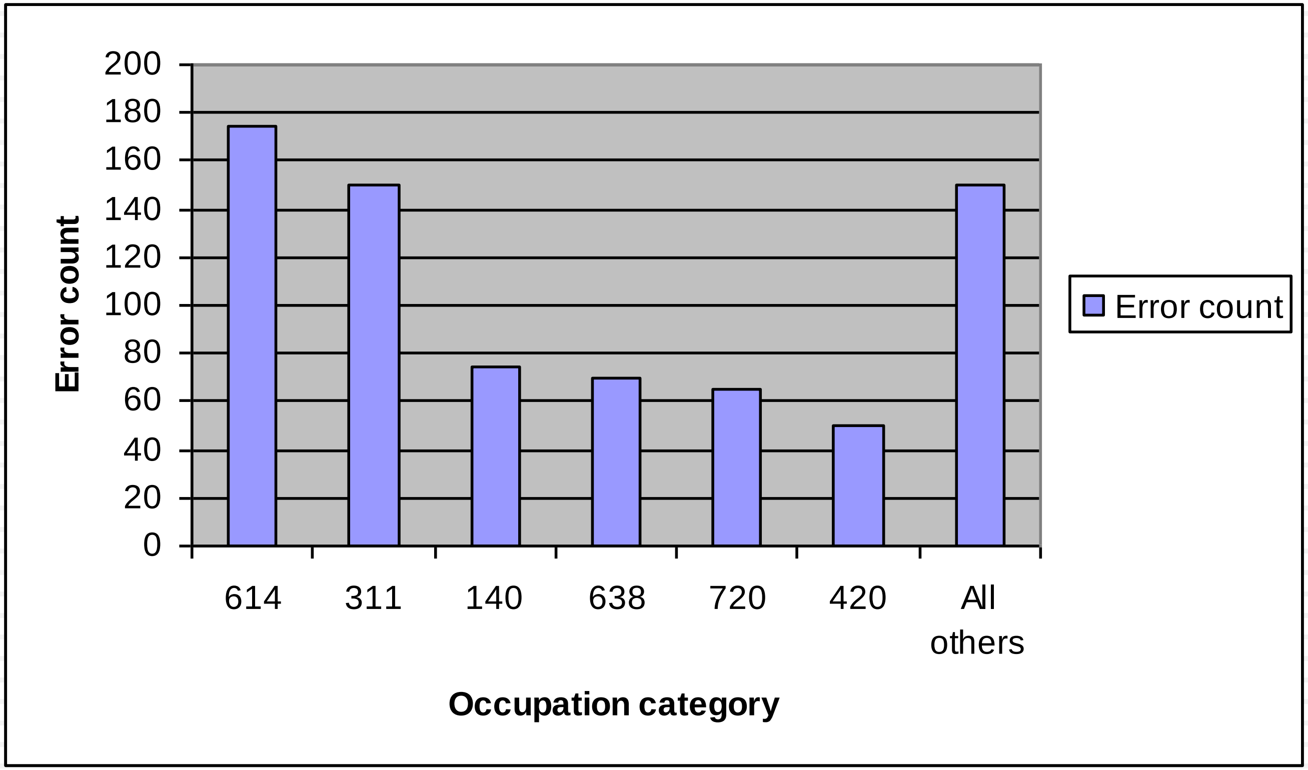
A Useful Tool

- Cause and effect (fishbone) diagrams
  - Helps to identify all possible root causes of a problem
  - An important component of the measure stage of DMAIC.



Another Useful Tool

- Pareto chart
  - Useful for identify the “vital few” sources of process deficiencies



The Golden Age of Survey Research 1960-1975

- Probability sampling well established in most countries (Neyman 1934)
- Nonresponse rate above 10-15% suggested that the survey was bad
- Cost situation was reasonable
- A limited number of modes available
- Things were allowed to take time
- Some error sources were unknown

Kalton (2018)

Drivers of Major Developments Since 1975

- More advanced computers and software
- More sophisticated users
- New modes such as CATI and Voice Recognition
- Increasing nonresponse
- Increasing costs
- The cognitive revolution (CASM)
- Model-assisted surveys
- The quality movement

Kalton (2018)

A Changing Survey Landscape

- Users want more timely and richer data
- Increased nonresponse and costs in surveys
- Demands for reduced respondent burden
- New data sources, new actors and new technology
- Combining different data sources
- Nonprobability samples and inferential issues
- New data collection modes and mixed modes
- New storage requirements

INIZIO

Why Did Data Become “BIG”

- Technological advances associated with data science and computational tools and methods.
- Information-based Decision Making
  - “Evidence-based”, “Data-Driven”, “Analytics”, “Machine Learning”
- Focus on short-run prediction
  - Business decision making
  - Health risks (e.g. Google Flu)
  - Financial markets
  - Political processes
- Style points: “Tail Fins”

Source: Heeringa 2018

Some Concepts

- Artificial Intelligence-machines being able to carry out tasks in a smart way
- Machine Learning-application of AI where we give machines access to data and let them learn for themselves via neural networks and natural language processing
- Data Mining-builds intuition about what is really happening in some data
- Data Science-combines the application of computer science, statistics, programming and business management

Happiness and Well-being

The common survey question: How satisfied are you with your life?

BD alternative

- 10 million tweets that are coded for happiness (rainbow, love, beauty, hope, wonderful, wine…) and non-happiness (damn, boo, ugly, smoke, hate, lied,…)
- Happiest states: Hawaii, Utah, Idaho, Maine, Washington
- Saddest states: Louisiana, Mississippi, Maryland, Michigan, Delaware

The Potential Use of Big Data in Statistics Production

- Produce statistics based on BD that can replace surveys
- Combine BD with admin data, sample surveys, and nonprobability sources in order to improve statistics
- Explore new topics and concepts
- Data mining to identify new patterns and models

Examples of Sources of Data

- Censuses
- Other survey programs
- Administrative data systems
- Medical records systems
- Commercially compiled data
- Financial data
- Satellite imagery
- GPS and GIS
- Social media
- Mobile devices
- Wearable measurement devices
- Sensors (Internet of Things)
- Visual data: pictures and video
- Genetic profile data
- Transactional data systems

Adapted from Heeringa 2018

User Demands

- More timely data
- Not just estimates
- Advanced products such as interactive features
- Surveys not great for complex concepts (religion, loneliness, happiness, existential issues, health, hypothetical questions)

INIZIO

Current Issues

- Nonresponse is up
- Costs are up
- Face-to-face almost gone
- Margin of error understated
- Missing at random not a very realistic assumption
- In Sweden only SCB is ISO 20252 certified
- New version of ISO released with access panels included
- New ideas involving responsive design have had limited success
- Quality management is struggling
- Lots of bad surveys out there

Where to Go

- Strive for error-free processes
- Find a model for international collaboration that can generate common standards for statistics production
- Develop a structured international competence development program for service providers
- Adopt the philosophy of continuous quality improvement with a user perspective
- Modern quality management principles are based on statistical methods and are therefore especially suitable for statistical organizations
- Adapt to a changing survey landscape

A Useful Quote

Until the purpose is stated, there is no right or wrong way of going about the survey

Deming, W. E. (1944)

Over and Out

Lars Lyberg
lyberg.lars@gmail.com
Inizio