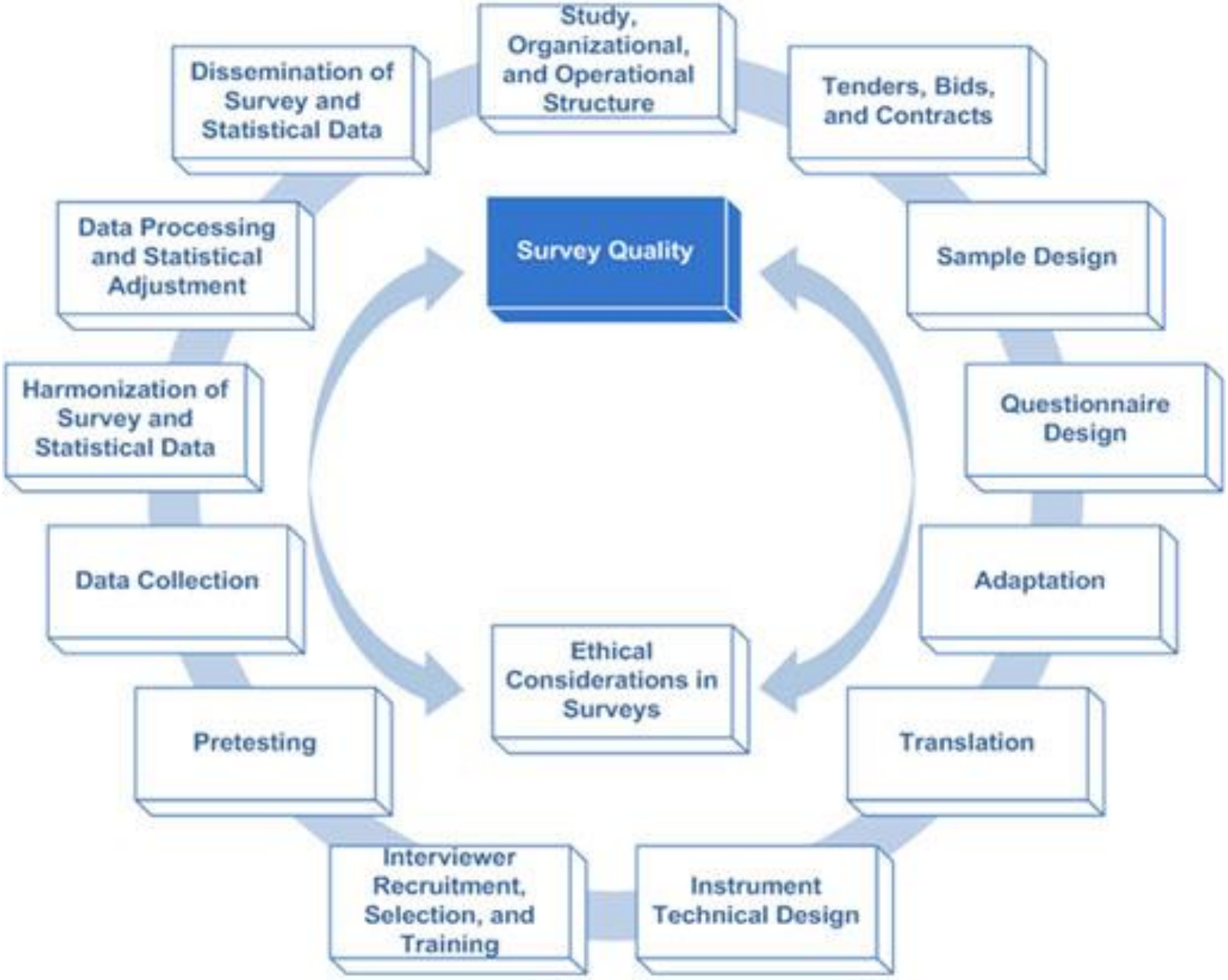
Assessing Quality of Survey Data: Overview



ccsg.isr.umich.edu/quality.cfm

Survey quality framework

‘‘Fitness for use’’ paradigm (Juran & Gryna 1980):

Survey quality as understood by both data producers and by data users.

Hence,

2 distinct elements of quality (as general concept):

(a) freedom from deficiencies;

(b) Responsiveness to users’ needs

↓

Survey quality as a multidimensional concept

Common Dimensions of Survey Quality

Dimension	Description
Accuracy	Total survey error is minimized
Credibility	Data are considered trustworthy by the survey community
Comparability	Demographic, spatial, and temporal comparisons are valid
Usability/Interpretability	Documentation is clear and metadata are well-managed
Relevance	Data satisfy users needs
Accessibility	Access to the data is user friendly
Timeliness/Punctuality	Data deliveries adhere to schedules
Completeness	Data are rich enough to satisfy the analysis objectives without undue burden on respondents
Coherence	Estimates from different sources can be reliably combined

Source: Biemer 2010, p. 109

Total Survey Error (TSE) as part of the Accuracy dimension

CSDI guidelines for quality & examples of indicators of quality (adapted from Eurostat's standard quality indicators): ccsg.isr.umich.edu/quality.cfm Appendix A

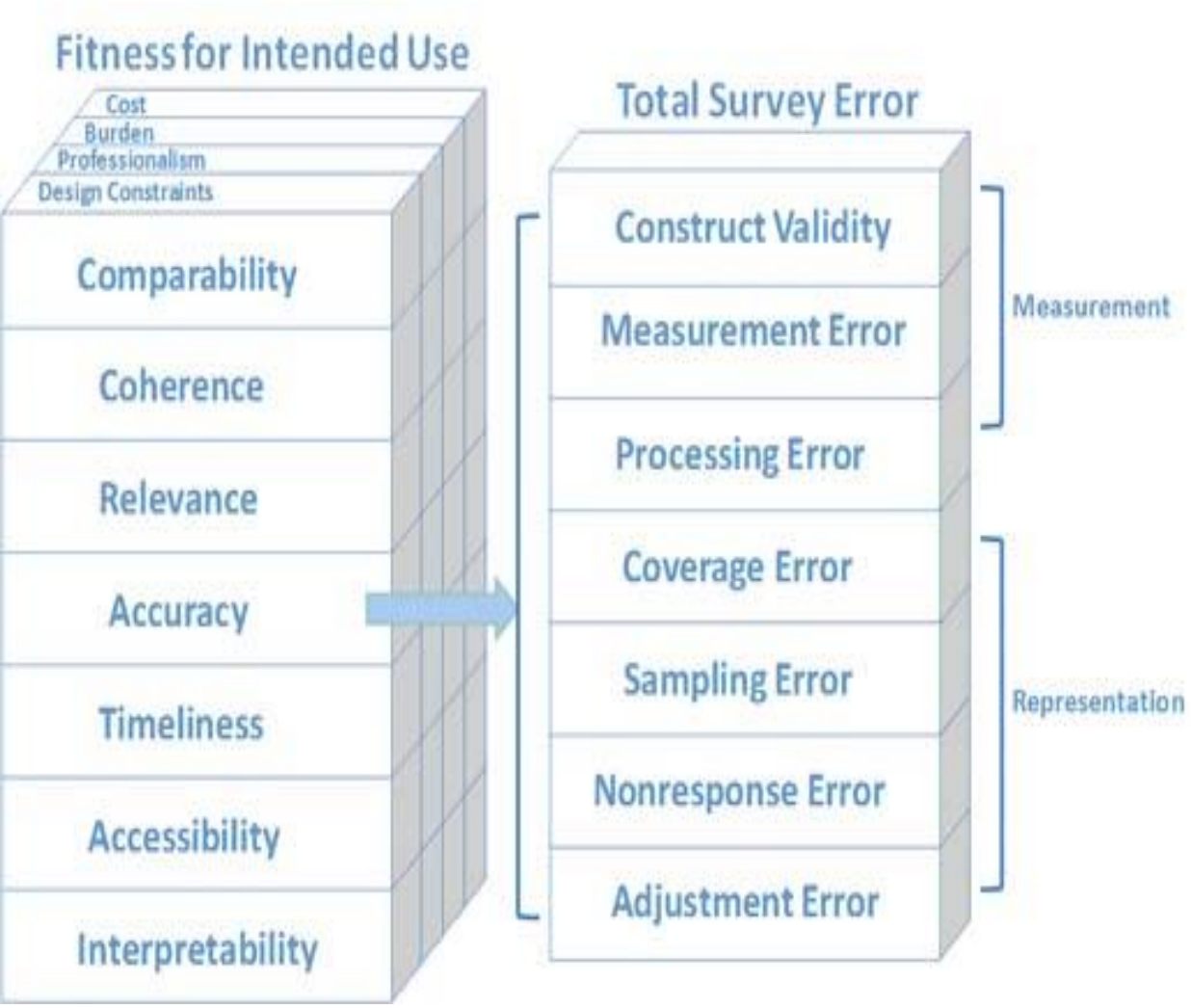
Total Survey Error (TSE)

TSE = Sampling Error (SE) + Non-sampling Error (NE)

- **SE = due to selecting a sample instead of the entire population**
- **NE = due to mistakes or system deficiencies**

SE = even if the sample is well constructed (& does not needs repairs), there may me a considerable difference btw. estimated values and true (population) values of the distribution properties

Fitness of use & TSE



TSE - NE

- **Specification**
  - **Concepts**
  - **Objectives**
- **Frame error**
  - **Omissions**
  - **Erroneous inclusions**
- **Non-response error**
  - **Whole unit**
  - **Item**
- **Measurement error**
  - **Respondent**
  - **Interviewer**
  - **Instrument**
- **Processing error**
  - **Data entry**
  - **Coding**
  - **Weighting**

The Survey Process Quality Management framework

To obtain quality products, quality processes are necessary. The latter require quality management at (a) the overall study level; and (b) the national organization level.

Survey production process quality assessment requires:

- use of quality standards;
- collection of standardized study metadata, question metadata, and process paradata

Metadata = information that describes data.

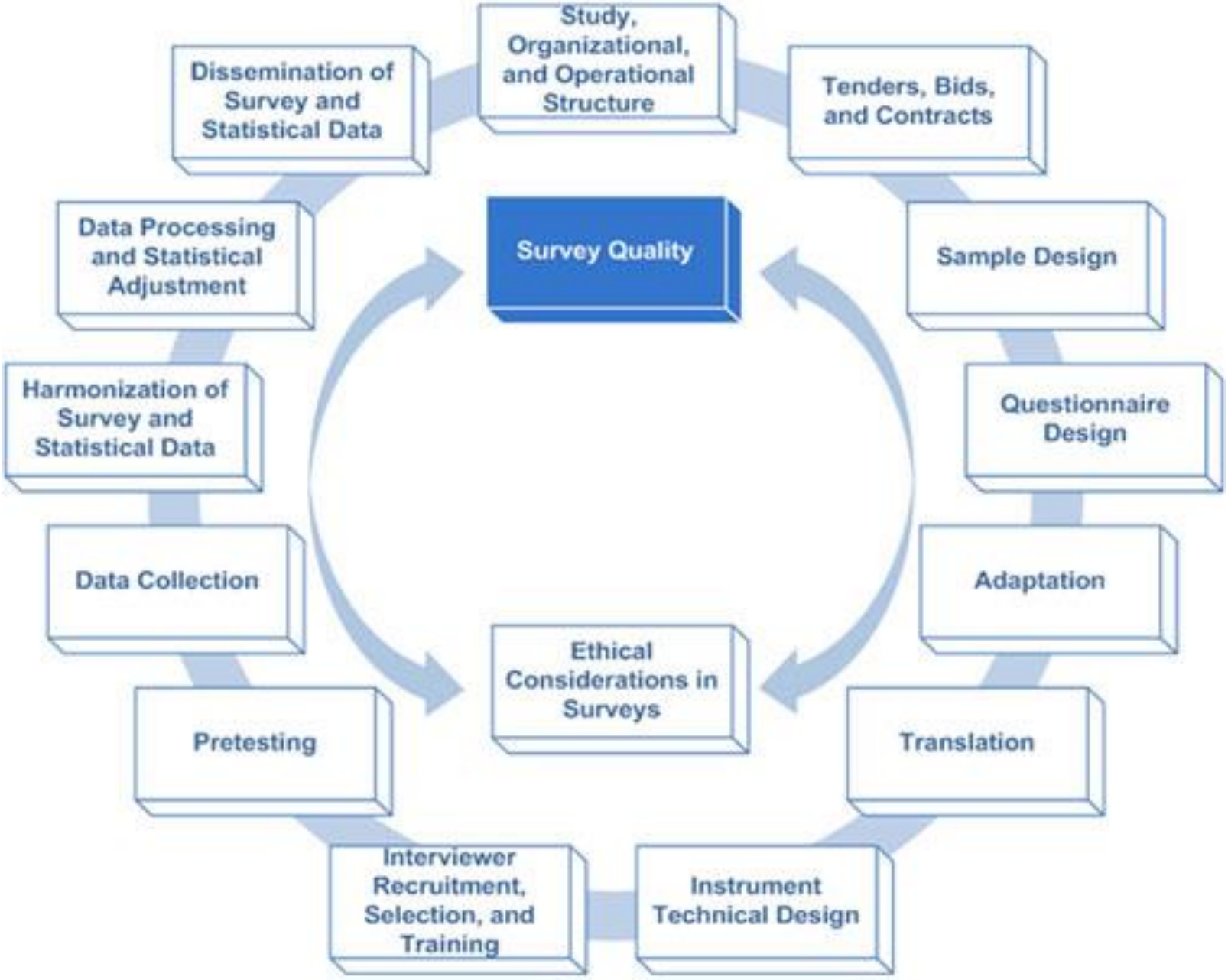
In our project: info. about sample, response rates, translation, pretesting, control of fieldwork to control for the quality of the survey as reflected in the survey documentation.

Paradata = empirical measurements about the process of creating survey data themselves:

visual observations of interviewers, administrative records about the data collection process, computer-generated measures about the process of the data collection, external supplementary data about sample units, observations of respondents themselves about the data collection.

CSDI -> recommended elements of process quality management relevant to each element of the survey lifecycle

ccsg.isr.umich.edu/quality.cfm Appendix B



ccsg.isr.umich.edu/quality.cfm

Survey quality-control indicators in SDR

- **Survey documentation**
- **(In)consistencies btw. the resources defining variables and their values (e.g. codebooks and questionnaires)** on one hand, **and data records** in the computer file on the other
- **Computer data records**

General Survey Documentation: How is the quality of national surveys reflected in data documentation	Answers
Does the survey documentation provide information on the response rate?	Yes = 1 No = 0
Was the questionnaire back-translated or translation checked in some other way?	Yes = 1 Else = 0
Is there any evidence that the questionnaire was pre-tested?	Yes = 1 Else = 0
Does the documentation show that the fieldwork was controlled? Marta and Matt's presentation today Effect of item value = 0 : Reduction of confidence in the data	Yes = 1 Else = 0

Specific Data Description: How have the data been defined?	Answers
Do variable values in the codebook correspond to values in the data file?	**Yes = 0** **No = 1**
Eight binary variables describing discrepancies between data description and the data file	
(Ilona and Olena's presentation tomorrow)	
Effect of negative answers (No = 1): Decrease of interpretability of the data	

Computer Data File: Are the data formally correct?	Answers
Do survey cases (respondents) have unique identification numbers (IDs)?	Yes = 0 No = 1
Are survey weights free of formal errors? (Marcin's presentation yesterday)	Yes = 0 No = 1
Is the proportion of missing values for gender and age within the standard limits (< 5%)?	Yes = 0 No = 1
Is the data file free from repeated cases (duplicates)? (Przemek's presentation tomorrow)	Yes = 0 No = 1

Effect of negative answers (No = 1) : Possible distortion of the research results based on the data