MATHEMATICA Policy Research

Transparency, Quality, Documentation, and Dissemination: International Standards

Presentation to the Committee on Transparency and Reproducibility of Federal Statistics for the National Center for Science and Engineering Statistics

May 21, 2019

John L. Czajka

Mathew Stange

Background

- **Mathematica, under a contract with the Statistics of Income Division, Internal Revenue Service, produced a report for the Interagency Council on Statistical Policy that reviewed information on international standards and guidelines on quality reporting relative to statistical estimates that combine survey data with other types of data**
- **In preparing the report, the authors participated in a Federal Committee on Statistical Methodology Working Group on Transparent Quality Reporting in the Integration of Multiple Data Sources**
- **This presentation is based on the Mathematica report:**

  Transparency in the Reporting of Quality for Integrated Data: A Review of International Standards and Guidelines

  John L. Czajka and Mathew Stange

  April 27, 2018

Guidance from Eurostat

What international standards may offer

- **Administrative data systems more developed outside the U.S.**
- **Decline in survey response rates—at least in Europe—more rapid than in the U.S.**
- **International organizations and several national statistical offices have been particularly active in development of standards**
  - **Eurostat and the European Statistical System; United Nations**
  - **Canada, United Kingdom, The Netherlands, Norway, Finland, Sweden, Australia, New Zealand**
  - **Recent focus on use of administrative records and Big Data for official statistics**

European Union statistical organizations

- **Eurostat is a Directorate General of the European Commission, the executive of the European Union**
  - **Eurostat is the statistical office of the European Union**
  - **Eurostat is charged with the production of official statistics at the level of all Europe for the European Union**
- **European Statistical System (ESS) is a partnership between Eurostat and the statistical authorities of the member states**
  - **ESS Committee charged with providing "professional guidance to the ESS for developing, producing, and disseminating European statistics"**

Key documents from European Union

- **European Statistics Code of Practice for the National and Community Statistical Authorities (2011)**
- **Quality Assurance Framework for the European Statistical System (2015)**
- **ESS Handbook for Quality Reports (2015)**
  - **Includes in an appendix:**
    - ESS Guidelines for the Implementation of the ESS Quality and Performance Indicators

ESS Handbook for Quality Reports

- **Purpose is "to provide guidelines for the preparation of comprehensive quality reports for a full range of statistical processes and their outputs"**
- **Specific objectives of these guidelines:**
  - **To promote harmonized quality reporting across statistical processes and their outputs within a Member State and hence to facilitate comparisons across processes and outputs**
  - **To promote harmonized quality reporting for similar statistical processes and outputs across Member States and hence to facilitate comparisons across countries**
  - **To ensure that reports include all the information required to facilitate identification of statistical process and output quality problems and potential improvements**

Quality reports and quality profiles

- **Comprehensive quality reports addressed by the Handbook bear resemblance to U.S. quality profiles**
- **A survey quality profile summarizes what is known about the sources and magnitudes of errors in a survey (Kasprzyk and Kalton 2001)**
  - **A systematic and comprehensive review across the spectrum of survey activities in which both qualitative and quantitative results are brought together to allow an assessment of the quality of the survey operations and the data**
  - **Relevance, timeliness, and accessibility are dimensions of quality not usually treated in quality profiles in the U.S.**
- **Quality profiles were produced for several federal surveys**
- **They are resource intensive, they require information that may not exist, and their value to the survey producer is questionable**

Handbook guidelines

- **Handbook provides guidelines specific to each of the five dimensions of statistical output quality—relevance, accuracy and reliability, timeliness and punctuality, coherence and comparability, and accessibility and clarity—plus three other Code of Practice principles:**
  - **Confidentiality (principle 5)**
  - **Burden (principle 9)**
  - **Cost (principle 10)**
- **Handbook also includes guidelines on statistical processing, which is not one of the Code principles**
- **Recommendations for quality reporting include 16 quantitative indicators for the five quality dimensions**
- **A general recommendation that whenever multiple data sources were used, a separate quality report be produced for each**

Accuracy and reliability

- **Nearly half of the Handbook's main text is devoted to this quality dimension**
- **Guidelines for quality reporting distinguish among six types of statistical processes: (1) sample surveys, (2) censuses, (3) processes using administrative sources, (4) processes using multiple data sources, (5) processes for generating price and other economic indexes, and (6) statistical compilations**
- **Accuracy is divided into overall accuracy, sampling error, and non-sampling error**
- **Non-sampling error is further divided into:**
  - **Coverage error**
  - **Measurement error**
  - **Nonresponse error**
  - **Processing error**

Other quality dimensions

- **Relevance**
  - **Focus is on users of the statistical outputs and to what extent the data satisfy their needs**
  - **Different groups of users may have different needs**
  - **The one quality and performance indicator is the data completeness rate: the ratio of data cells provided to cells required**
- **Timeliness and punctuality**
  - **Quality and performance indicators include:**
    - Time lag between end of reference period and initial results
    - Time lag between end of reference period and final results
    - Time lag between delivery of data and announced target date

Other quality dimensions cont'd

- **Coherence and comparability**
  - **This dimension is assigned high importance, with extensive information requested for the quality report**
  - **Quality and performance indicators address only "mirror flows" (inflows and outflows that should match) and length of unbroken time series**
  - **A caution not to confound coherence/comparability with accuracy (seeming inconsistency could be due to inaccuracy)**
- **Accessibility and clarity**
  - **User feedback is the best source of information in addressing this dimension in the quality report**
  - **What can more sophisticated and less sophisticated users access?**
  - **Quality indicators include how often users consult tables and metadata and the degree of completeness of the latter**

Other principles

- **Cost**
  - **Quality report should include cost breakdown by major components although difficulty of obtaining this is noted**
- **Burden**
  - **Quality report should include:**
    - Respondent burden in financial terms or hours
    - Targets for reducing burden and recent efforts to reduce burden
    - Whether information collected is limited to what is absolutely necessary and cannot be obtained elsewhere
- **Confidentiality**
  - **Distinction between legal requirements and data treatment**
  - **Not mentioned are measures to assess effectiveness**

Transparency as presented by Statistics Canada

Statistics Canada

Policy on Informing Users of Data Quality and Methodology (approved March 31, 2000) http://statcan.gc.ca/eng/about/policy/info-user

1. **Statistics Canada will make available to users indicators of the quality of data it disseminates and descriptions of the underlying concepts and methodology**
2. **Statistical products will be accompanied by or make explicit reference to documentation on quality and methodology**
3. **Documentation on quality and methodology will conform to such standards and guidelines as shall from time to time be issued under this Policy**
4. **Exemption from the requirements of this Policy may be sought in special circumstances**

Statistics Canada

- **Standards and guidelines on the documentation of data quality and methodology (revised 11/25/02)**
  - **Standards detail mandatory requirements for documentation on data quality and methodology**
  - **Guidelines outline types of information to be included in additional documentation when a broader and more detailed range of documentation is desirable**
- **Six dimensions of data quality for official statistics, where quality is defined in terms of "fitness for use"**
  1. **Relevance**
  2. **Accuracy**
  3. **Timeliness**
  4. **Accessibility**
  5. **Interpretability**
  6. **Coherence**

Statistics Canada

Standards constitute mandatory documentation

1. **Note(s) to users (if applicable)**
2. **Concepts, methodology, and data quality**
   - 2.1 Data sources and methodology
   - 2.2 Concepts and variables measured
     - Key concepts, variables, and classifications used
     - Key indicators, indices, or other key data or results being disseminated
   - 2.3 Data accuracy
3. **Appendices (as necessary) and/or references or links**

Statistics Canada

Data sources and methodology

1. **Introductory paragraph (purpose, objectives and subject matter or content)**
2. **Description of the survey or program population**
3. **Statement on time frame or reference period of the data**
4. **General methodology**
   - Statement on data source(s) and sampling and collection methodology
   - Statement on processing and estimation methodology
5. **Revisions and adjustments (if applicable)**
   - Statement advising what data are subject to revision and why, and an indication of the likely size of the revision (for example, a measure based on past revisions)
   - Description of benchmarking, calendarization or seasonal adjustments made to the data and their impact

Statistics Canada

Data accuracy

- Statement of key data accuracy issues plus acknowledgment that the data are subject to error, which may vary across geography and by characteristic
- For all data types, a measure of coverage
- For sample survey data, estimates of sampling error for key characteristics
- For all data types, a response rate, a statement on how non-response and response error are handled, an imputation rate, and assessment of significant, related accuracy issues
- Descriptions and accuracy indicators for important residual errors (if applicable)
- A statement advising on comparability over time (if applicable)

Statistics Canada

Data accuracy cont'd

- Explanation of similarities and differences between related data sources and results of comparisons with other sources (if applicable)
- For analytical results, a summary of methods, assumptions, and caveats as well as discussion of possible effects of data accuracy, survey concepts, and analytical assumptions on the results (especially their validity)
- Description of any other important issues or events influencing the accuracy, interpretation, or use of the data

Statistics Canada

Guidelines—additional documentation—may include:

- Historical quality trend
- Questionnaire(s)
- Sampling frame—creation, updating, and quality assurance
- Detailed sample design and estimation procedures
- Indicators of the extend of coding errors, data capture errors, impact of edits
- Description of imputation approach and key imputation rules
- Quality control procedures used
- Confidentiality protection requirements and procedures
- Special procedures or steps relevant to product content
- Total variance or its components by source
- Assessment of non-response bias
- Evidence of bias in responses
- Seasonal adjustment methodology and its impact
- Data quality validation and evaluation results

Statistics Canada

Each of the following types of data has special requirements for additional documentation

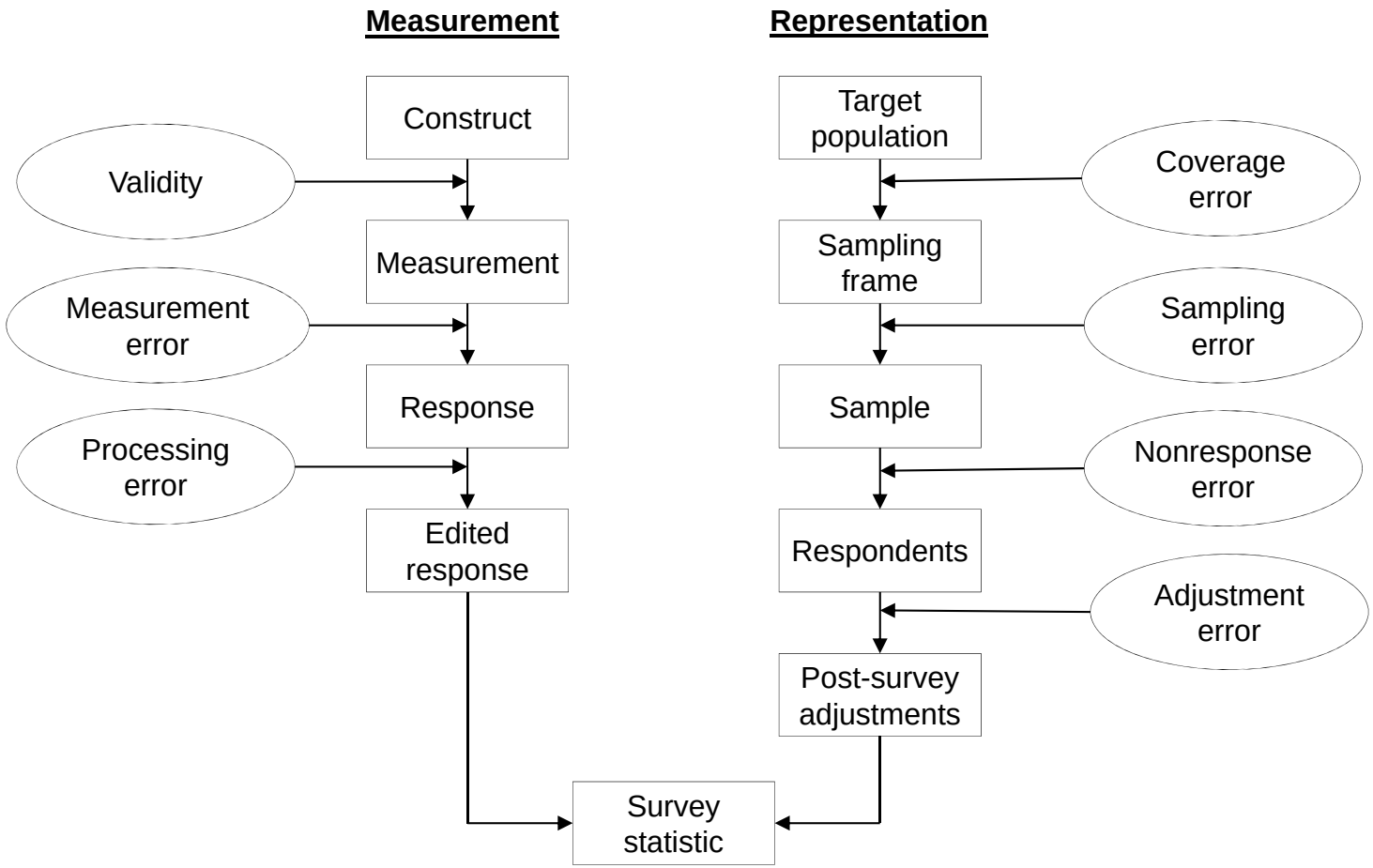
1. **Indices for prices or other quantities**
2. **National Accounts and data resulting from other data integration activities**
3. **Statistics derived from administrative data or other data not collected by Statistics Canada**
4. **Geographic or cartographic data products**
5. **Products including primarily or only analytical results**
6. **Building-block data products (e.g., microdata or low-level aggregations intended for aggregation or analysis)**
7. **Products from a longitudinal survey**

Quality measurement reflecting the life cycle of data production

Overview of the survey life cycle model

- **The Total Survey Error (TSE) model follows the life cycle of a survey from conception to the production of a survey statistic**
- **The model builds on the idea that a sample survey consists of questions administered to a sample drawn from a target population**
- **The model traces the dimensions of measurement and representation from an abstract construct and a target population through the design and implementation of a survey, culminating in a survey statistic**
- **Error may be introduced at each stage as depicted in the figure on the next slide**

Survey life cycle from quality perspective

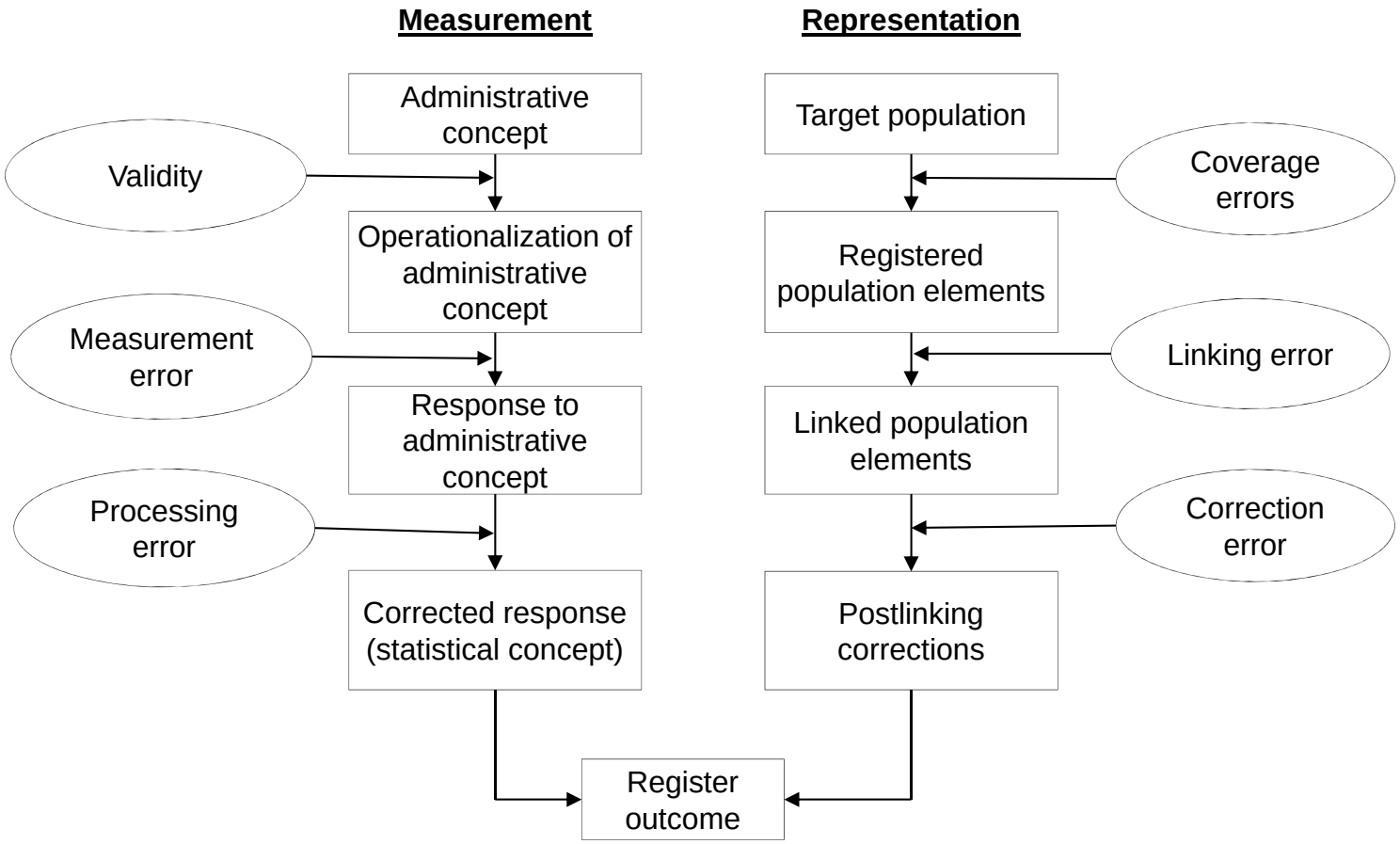


Source: Groves et al. (2009).

Extending TSE to administrative data

- **Bakker (2010) adapted the TSE model for register-based (administrative) data—a step toward the development of a framework for describing error in statistics based on combined sources**
- **The framework for TSE is relevant as a starting point because most registration data are collected with survey techniques**
- **The possible errors common in surveys will also occur in registration data**
- **On the representation side, though, there is no sampling and no unit nonresponse and thus no correction for the latter**
- **But there may be linking of population elements across registers and thus the possibility of linking error (failed and missed links)**

Sources of error in a register life cycle



Source: Bakker (2010).

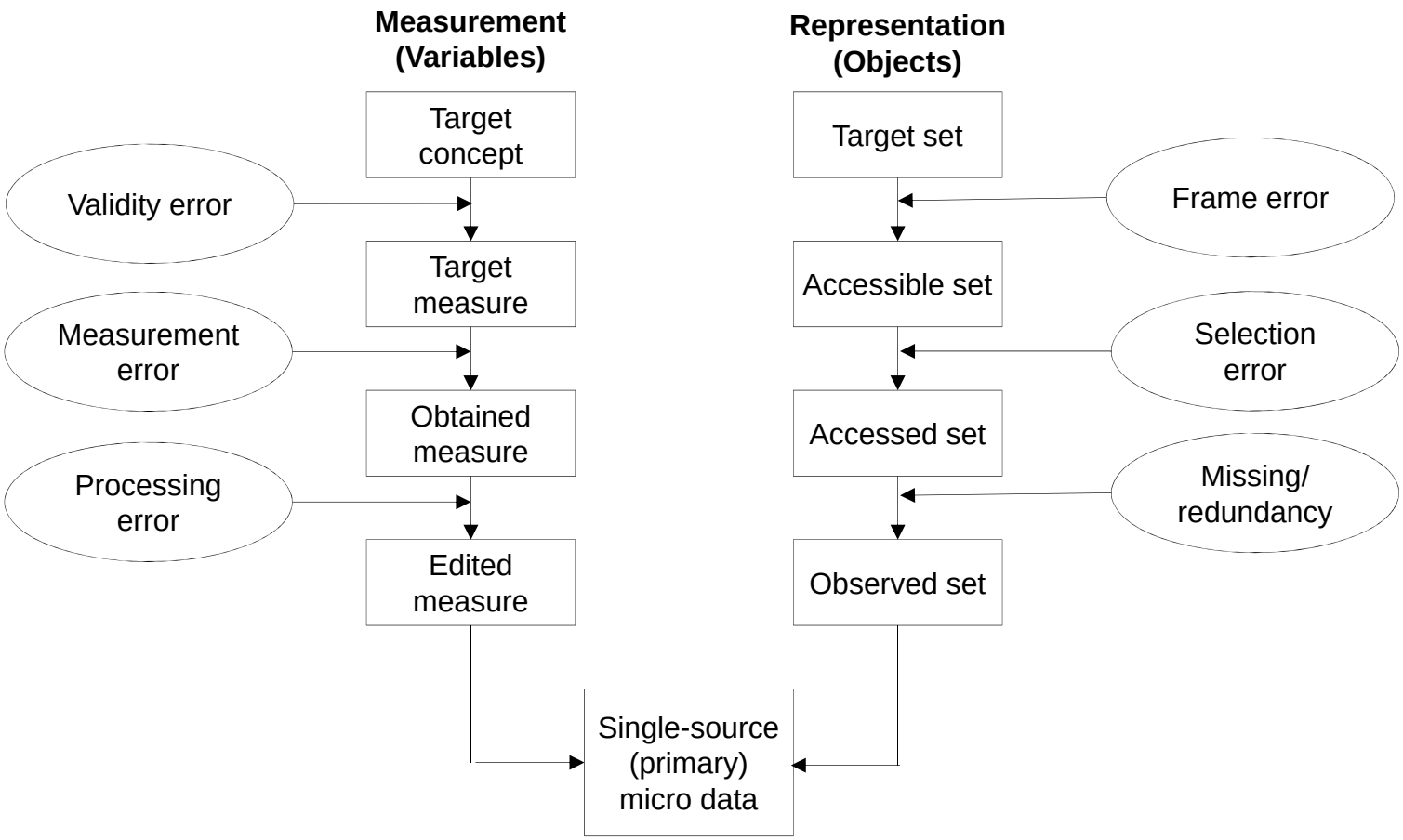
Extending TSE to integrated data

- **Li-Chun Zhang of Statistics Norway proposed a framework for integrated data based on the survey life cycle model of Groves et al. (2009) and its adaptation to register data by Bakker (2010)**
- **Statistics New Zealand has adopted this framework as the basis for its own quality framework for integrated data**
- **Combining data from multiple sources requires additional methods and generates new types of error, both of which should be documented to achieve full transparency**
- **Zhang's framework is helpful in identifying these methods and sources of error**

Two-phase life-cycle model

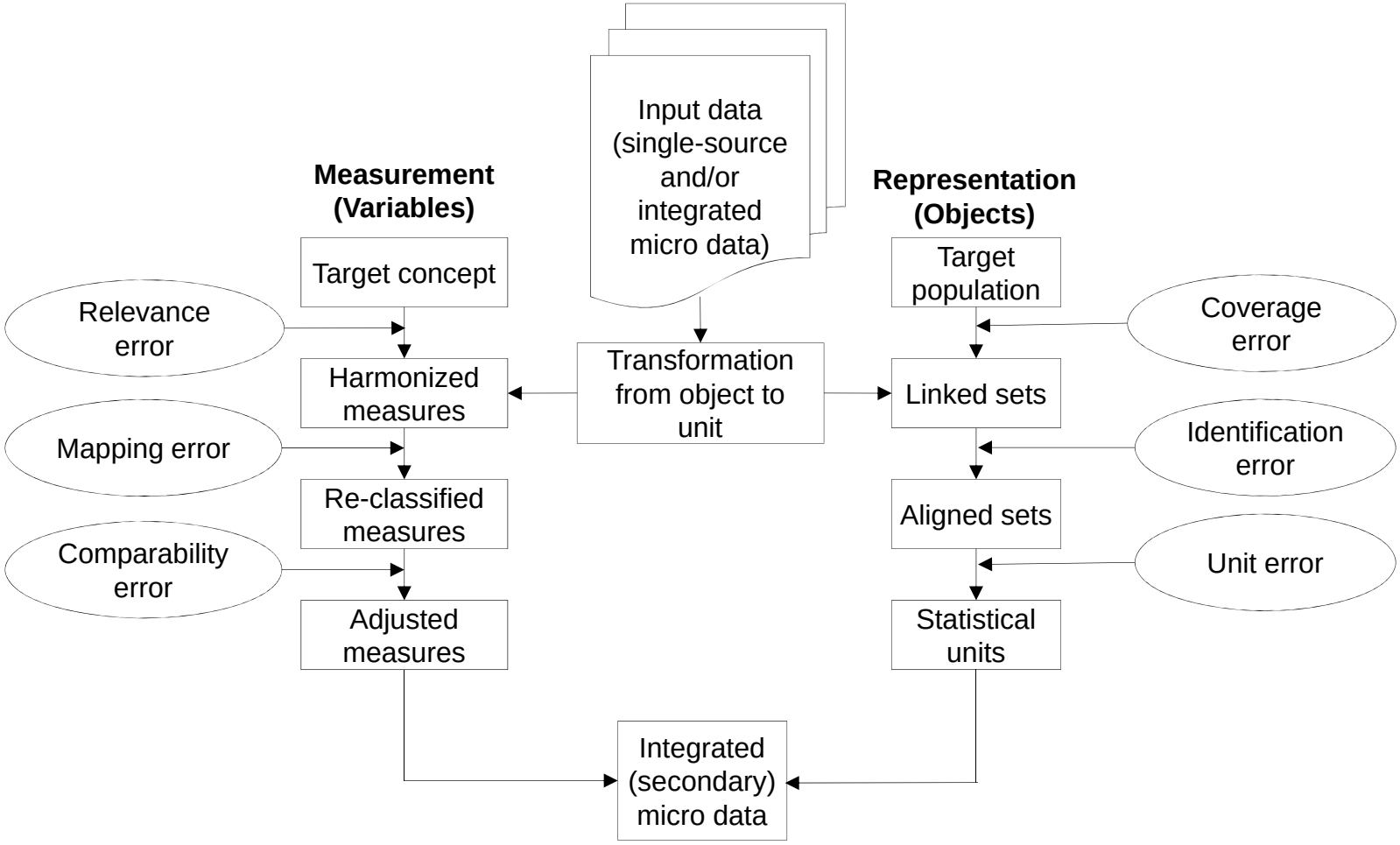
- **In Zhang's two phase model, the end result of each phase is a micro dataset—not a single statistic**
- **In addition, most of the concepts from Groves et al. and Bakker have been renamed to accommodate the inclusion of data from both survey and administrative sources**
  - **The post-survey adjustment, which Bakker removed as unnecessary for register data, is not included here**
  - **Nor is phase one linkage, which Bakker added to accommodate multiple registers**
- **Phase one describes a single microdata source, but each input to the integrated microdata has its own phase one assessment, which addresses the original purpose of the dataset**
- **Phase two shows the multiple inputs and depicts the sources of error for the integrated microdata**

Two-phase life cycle of integrated data (1)



Source: Zhang (2012).

Two-phase life cycle of integrated data (2)



Source: Zhang (2012).

Elements of the two-phase model

- **Zhang observes: "the 20th century witnessed the birth and maturing of sample surveys; the 21st century will be the age of data integration"**
- **Harmonization on the measurement side and linkage on the representation side are steps in phase two**
  - **Harmonization involves a conceptual alignment of measures from different sources; no change is made to the measures at this point**
- **On the representation side, Zhang uses "objects" in phase one and "units" in phase two; the transformation of objects into units is shown in a box in phase two below the input of multiple data sets**
  - **Phase one data may include, for example, jobs while the goal of integration may be data on persons**
  - **Units themselves may have to be combined in some way—for example, persons aggregated to households**

Error at the dataset level

- **Zhang's conceptualization envisions an ideal target integrated dataset—the analog to an error-free survey statistic**
- **Discrepancies between the target dataset and the final integrated dataset are analogous to the concept of TSE in Groves et al.**
- **To assess the accuracy of the final dataset, Zhang develops the concept of empirical equivalence**
  - **Two datasets are empirically equivalent if they generate identical inferences; this does not require micro-level equivalence**
- **Zhang extends empirical equivalence to the assessment of public use data, where error is introduced to protect confidentiality**

Statistics New Zealand (Stats NZ)

- **With a mandate to make administrative data the data source of choice, Stats NZ faces the need to "assess and explain the quality of statistics that use multiple sources, including administrative data" (Holmberg and Bycroft 2017)**
- **Stats NZ issued in 2016 a *Guide to Reporting on Administrative Data Quality*, which uses Zhang's framework**

  http://www.stats.govt.nz/methods/data-integration/guide-to-reporting-on-admin-data-quality.aspx

  - **Includes quality indicators for each of the phase one and phase two error sources**
  - **25 quantitative indicators for phase one and 19 for phase two**
  - **34 qualitative indicators for phase one—mostly descriptive**
  - **No qualitative indicators as yet for phase two**

Quantitative indicators for phase one

Error source and indicator	
Measurement dimension	
Validity error	
1	Percent of items that deviate from target concept definition
2	Percent of items that deviate from StatsNZ/international standards or definitions
3	Percent of inconsistent records
4	Percent of items affected by respondent comprehension of questions asked in collection process
Measurement error	
5	Item nonresponse rate
6	Item imputation rate
7	Percentage of records from proxies
8	Lagged time between reference period and receipt of data
9	Punctuality
10	Overall time lag
11	Percent of units in administrative data which fail checks
12	Stability of variables
Processing error	
13	Percentage of units of a variable with transcription errors
14	Modification rate--frequency of editing changes to a variable
15	Readability

Phase one indicators cont'd

Representation dimension	
Frame error	
16	Lag in updating population changes--delays in registration
17	Undercoverage--units in the target population not in the accessible set
18	Overcoverage--units in the accessible set not in the target population
19	Authenticity--correctness of identifiers
Selection error	
20	Adherence to reporting period
21	Dynamics of births and deaths--changes in rates over time
22	Inconsistent objects/units
Missing/redundancy error	
23	Unit nonresponse rate
24	Percentage of duplicate records
25	Percentage of units that have to be adjusted to create statistical units

Source: Statistics New Zealand (2016).

Quantitative indicators for phase two

Error source and indicator	
Representation dimension	
Coverage error	
1	Undercoverage--proportion of units in the target population missing from the final dataset
2	Overcoverage--proportion of units in the final dataset not in the target population
3	Proportion of units linked from each dataset to a base dataset, or percentage link rates between pairs of datasets
4	Proportion of duplicated records in the linked data
5	False positive and negative link rates
6	Macro-level comparisons of the distribution of linked objects with reference distributions
7	Delay in reporting--time lag between end of reference period and receipt of final data
8	Linking methodology used
Identification error	
9	Proportion of units with conflicting information
10	Proportion of units with mixed or predominance-based classifications
11	Rates of unit change from period to period
Unit error	
12	Proportion of units that may belong to more than one composite unit

Phase two indicators cont'd

Measurement dimension	
Relevance error	
13	Percentage of items that deviate from Statistics NZ/international standards or definitions
Mapping error	
14	Proportion of items that require reclassification or mapping
15	Proportion of units that cannot be clearly classified or mapped
16	Distribution of variables in linked data
17	Indicators and measures of modeling error
Comparability error	
18	Proportion of units failing edit checks
19	Proportion of units with imputed values

Source: Statistics New Zealand (2016).

Adding a third phase to the framework

- **Reid et al. (2017) added a third phase for assessing the quality of final outputs—that is, the statistical estimates derived from the integrated microdata that is the endpoint of phase two**
- **Quality indicators do not yet exist for phase three**
- **Reid et al. provide three case studies that illustrate different approaches to evaluation**
  - **Case study 1: Redesign of the Building Activity Survey**
  - **Case study 2: Evaluating administrative data for personal income**
  - **Case study 3: Population estimation in New Zealand**

For More Information

- **John L. Czajka**
  - JCzajka@mathematica-mpr.com

- **Mathew Stange**
  - MStange@mathematica-mpr.com