

DATA COLLECTION STRATEGIES

Quantitative Data: Part II

Amílcar Moreira

ISEG - Institute of Economics and Management
University of Lisbon

April 16, 2024

**OPEN
MINDS.
GRAB
THE FUTURE.**

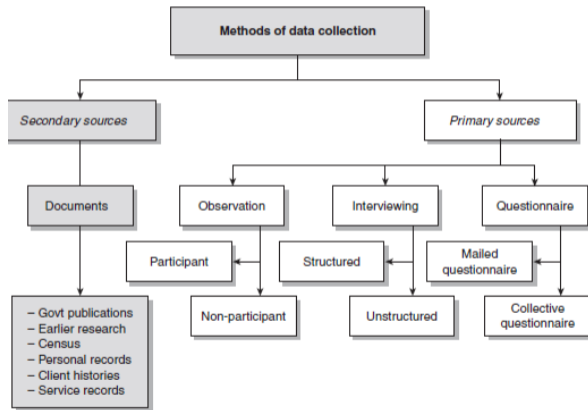
Overview

1. Data Collection: Primary vs. Secondary Sources

2. Exploring the Potential of Secondary Data with Composite Indicators

Data Collection: Primary vs. Secondary Sources

Primary vs. Secondary Sources



Types of Quantitative Secondary Data

- **Microdata**

- Census;
- Large-scale surveys;
- Cohort and other longitudinal studies;
- Administrative Records

- **Aggregate Data**

- Statistical Series;

- **Other**

- Geospacial data;
- Online data;
- Comercial data.

Source: Smith, 2008.

Doing Secondary Data Analysis

- **Saisana and Tarantola (2002):** Mathematical combination (or aggregation) of a set of indicators that represent the different dimensions of a phenomenon to be measured.

Source: Smith, 2008.

Choosing the Data

Key Questions

- **What is the purpose of original the research?**
- How timely is the data?
- Do you have the resources to retrieve and analyse the data?
- It is important to get some idea about the concepts that motivated the original research and in which way this may influence its use.
- Example: Was the research commissioned or undertaken by an advocacy group?

Source: Smith, 2008.

Choosing the Data

Key Questions

- What is the purpose of original the research?
 - **How timely is the data?**
 - Do you have the resources to retrieve and analyse the data?
- Need to determine when the data was collected and if it is data still relevant for today.

Source: Smith, 2008.

Choosing the Data

Key Questions

- What is the purpose of original the research?
- How timely is the data?
- **Do you have the resources to retrieve and analyse the data?**
- Need to consider if the data is easily accessible and if requires complex and time-intensiev procedures to access it;
- Need to consider if you have the necessary skills, resources and time to deal effectively with the scale and complexity of the data.

Source: Smith, 2008.

Assessing the Quality of the Data

Key Questions

- **Are concepts/categories comparable?**
- How was the data collected?
- Are there any missing data?
- Need to assess if the concepts adopted in the original study are comparable with the definitions adopted for the current study;
- If not, you need to assess to what degree this will impact on your results.

Source: Smith, 2008.

Assessing the Quality of the Data

Key Questions

- Are concepts/categories comparable?
- **How was the data collected?**
- Are there any missing data?
- Broadly, you need to consider how was the data collected and whether it sufficiently representative to allow generalisations to be made to the wider population;
- In particular, you need to assess if there are there any groups who have been excluded from the data;
- For survey data, you need to look at adequacy of sampling strategy and response rates.

Assessing the Quality of the Data

Key Questions

- Are concepts/categories comparable?
 - How was the data collected?
 - **Are there any missing data?**
- ‘Missing’ data, which can include ‘not known’, ‘information refused’, ‘information not yet sought’, and ‘other’ non-completed, often account for a large proportion of the responses;
 - Need to assess the types of missing data and their prevalence;
 - Need to consider if data imputation is a viable option for dealing with missing data.

Ethical Considerations when Using Secondary Data



- Need to assess if ethical approval was requested/granted for original data;
- Need to assess if declarations of 'informed consent' allow for the use of data for other purposes besides the original study.

Source: Smith, 2008.

Exploring the Potential of Secondary Data with Composite Indicators

What is a Composite Indicator

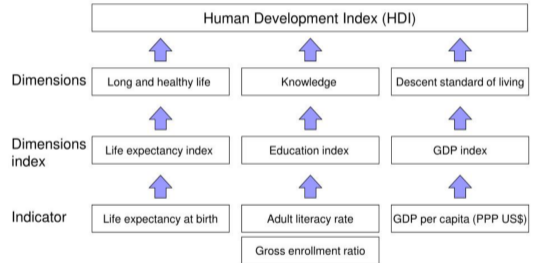
- **Saisana and Tarantola (2002):** Mathematical combination (or aggregation) of a set of indicators that represent the different dimensions of a phenomenon to be measured.
- Can involved the use of statistical indicators, aggregate indicators based on micro-data, and (in some cases) policy-rules (based on legislation).

Source: Kumar, 2012.

Composite Indicator as Tools for Monitoring and Comparing Countries and Policies

Examples

- **Human Development Index**
- European Innovation Scorecard
- Green Budgeting Index

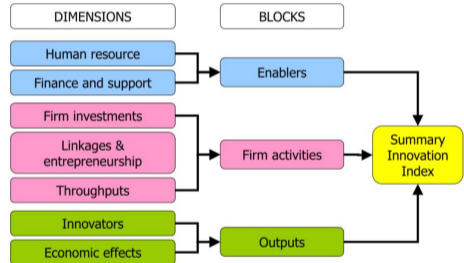


Source: XXXX, XXXX.

Composite Indicator as Tools for Monitoring and Comparing Countries and Policies

Examples

- Human Development Index
- **European Innovation Scorecard**
- Green Budgeting Index



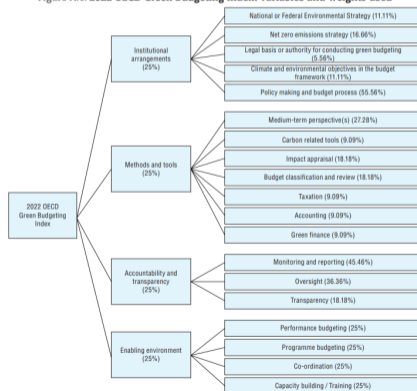
Source: XXXX, XXXX.

Composite Indicator as Tools for Monitoring and Comparing Countries and Policies

Examples

- Human Development Index
- European Innovation Scorecard
- **Green Budgeting Index**

Figure A.1. 2022 OECD Green Budgeting Index: Variables and weights used



Composite Indicator: Pros and Cons

Advantages

- Summarise complex, multi-dimensional realities.
- Easier to interpret than a battery of multiple indicators.
- Enable users to compare complex dimensions effectively.
- Allows to assess progress over time.
- Facilitate communication with general public and promote accountability.

Disadvantages

- May invite simplistic, or wrong policy conclusions.
- May send misleading policy messages if poorly constructed or misinterpreted.
- May be subject to manipulation due to lack of transparency.
- The selection of indicators and weights can be the subject of dispute.

Source: OECD/JRC, 2008.

Key Steps in Building a Composite Indicator

Key Steps

- **Analytical Framework**
 - Selection of Indicators
 - Normalization of Indicators
 - Weighting and Agregation
 - Validation
- This involves the identification of a 'Measurement Model', which specifies the relationship between the phenomenon to be measured (latent variable) and its measures (individual indicators).

Source: OECD/JRC, 2008; Mazziotta and Pareto, 2017.

Key Steps in Building a Composite Indicator

Key Steps

- Analytical Framework
 - **Selection of Indicators**
 - Normalization of Indicators
 - Weighting and Agregation
 - Validation
- Indicators should be selected according to their relevance, analytical soundness, timeliness, accessibility, etc.
 - A statistical approach to the choice of indicators involves calculating the correlation between potential indicators, and including the ones that are less correlated in order to minimize redundancy.

Source: OECD/JRC, 2008; Mazziotta and Pareto, 2017.

Key Steps in Building a Composite Indicator

Key Steps

- Analytical Framework
 - Selection of Indicators
 - **Normalization of Indicators**
 - Weighting and Agregation
 - Validation
- Normalization is required beacuse:
 - Indicators might be measured using different measurement units and ranges;
 - Some indicators may be positively correlated with the phenomenon, whereas others may be negatively correlated.

Source: OECD/JRC, 2008; Mazziotta and Pareto, 2017.

Key Steps in Building a Composite Indicator

Key Steps

- Analytical Framework
 - Selection of Indicators
 - **Normalization of Indicators**
 - Weighting and Agregation
 - Validation
- Typical normalization techniques:
 - Ranking;
 - Standardization (or Z-scores);
 - Re-scaling (or Min-Max);
 - Distance from a Reference (ex. Distance from best performer)

Source: OECD/JRC, 2008; Mazziotta and Pareto, 2017.

Key Steps in Building a Composite Indicator

Key Steps

- Analytical Framework
 - Selection of Indicators
 - Normalization of Indicators
 - **Weighting Agregation**
 - Validation
- The literature offers a wide variety of aggregation methods, each with its pros and cons.
 - They range from the simple arithmetic o geometric mean to multivariate statistical methods.

Source: OECD/JRC, 2008; Mazziotta and Pareto, 2017.

Key Steps in Building a Composite Indicator

Key Steps

- Analytical Framework
 - Selection of Indicators
 - Normalization of Indicators
 - **Weighting Agregation**
 - Validation
- The literature offers a wide variety of aggregation methods, each with its pros and cons.
 - They range from the simple arithmetic o geometric mean to multivariate statistical methods.

Source: OECD/JRC, 2008; Mazziotta and Pareto, 2017.

Key Steps in Building a Composite Indicator

Key Steps

- Analytical Framework
 - Selection of Indicators
 - Normalization of Indicators
 - **Weighting and Agregation**
 - Validation
- Different weithing approaches:
 - No weights = Equal Weights;
 - Expert weighting;
 - Principal Component Analysis (PCA).
 - No explicit weighting should be the norm. Need to justify the need for alternative approach.

Source: OECD/JRC, 2008; Mazziotta and Pareto, 2017.

Key Steps in Building a Composite Indicator

Key Steps

- Analytical Framework
 - Selection of Indicators
 - Normalization of Indicators
 - Weighting Agregation
 - **Validation**
- The creation of a composite indicator has numerous sources of uncertainty (choice of dimensions/indicators; choice of normalisation techniques, choice of agregation systems, choice of weighting approach or of weighting values).
 - Validadtion exercises are meant to a identify and quantify sources of uncertainty.

Source: OECD/JRC, 2008; Mazziotta and Pareto, 2017.

Key Steps in Building a Composite Indicator

Key Steps

- Analytical Framework
- Selection of Indicators
- Normalization of Indicators
- Weighting Agregation
- **Validation**
- **Uncertainty Analysis (UA):** Estimating the uncertainty of the results of the composite indicator by reference to the uncertainties in the inputs, usually represented by confidence intervals;
- **Sensitivity Analysis (UA):** Assess how much the uncertainty in the composite indicator for would be reduced if that particular input source of uncertainty were removed.



Amílcar Moreira

ISEG - Institute of Economics and Management
University of Lisbon

April 16, 2024