Austrian Journal of Statistics
2023, Volume 52, 82–100.
http://www.ajs.or.at/
doi:10.17713/ajs.v52i5.1539

Quality and Sensitivity of Composite Indicators for Sustainable Development

Laura Güdemann
Trier University

Ralf Münnich
Trier University

Abstract

Composite indicators can be understood as a summary of well-chosen and relevant sub-indicators which are combined into a single number. Their aim is to represent a multi-dimensional construct and map the performance of entities such as countries or companies. These multidimensional constructs are for example sustainability, poverty or well-being. Composite indicators are widely applied in various disciplines such as social or economic research and benefit from their apparent ease of interpretation. In the context of the Sustainable Development Framework a composite indicator over all 17 sustainable development goals, as been proposed. As composite indicators are commonly applied in highly sensitive areas this, urges the need to discuss methodical advantages and disadvantages as well as their adequacy for performance comparisons. In this paper we discuss and illustrate quality issues with regard to aspects of the subjective choices made in the construction process of composite indicators, imputation of missing data and the survey design. As an example we construct a composite indicator on sustainable economic development using data of the Sustainable Development framework. Furthermore, we exemplify and discuss strategies and methods for the quality assessment of a composite indicator.

Keywords: composite indicators, sustainable development goals, quality assessment, survey design, R.

1. Introduction

Composite indicators are a common method used for analysing the development of country performances according to relevant quantitative sub-indicators, especially in economic and social policy support. They benefit from their apparent ease of interpretation and their ability to summarise complex multidimensional issues. In the course of the Agenda 2030 of Sustainable Development, 17 Sustainable Development Goals (SDGs) with 169 specific targets have been adapted by world leaders as the main basis for post-2015 intergovernmental process (see, Sustainable Development Solutions Framework 2015, p. 5). Each of the 17 SDGs can be represented as a composite indicator since they are constructed by more than one target. Additionally, it is possible to build a composite indicator aggregating the target variables over all 17 SDGs to measure the overall achievement of a country. Such a SDG index has been proposed by the Bertelsmann Stiftung and is reported in the Sustainable Development Report published together with the Sustainable Development Solutions Network (see, Sachs, Schmidt-Traub, Kroll, Lafortune, and Fuller 2019, p. 19). Their common use in highly sensitive and important areas urges the need to discuss methodical advantages and disadvantages as well as their adequacy for performance comparisons. Methods of relative comparisons used for policy making need to follow high standards with regard to the data quality and interpretability of the results. Furthermore, they have to be applied carefully in order to avoid misinterpretation and false conclusions.

This paper aims to summarise important statements from the existing literature as well as drawing special attention to survey data quality and its impact on policy use sustainable Development Composite indicators. The studies outlined in this paper are an extension of the report by Güdemann and Münnich (2021). In the framework of a composite indicator, variables from different sources and surveys are combined into one single index. Often this is done without taking into account the data quality of the single sub-indicators or the data gathering process with which the data was taken. This can lead to misinterpretations and false conclusions.

In the following Section, general aspects and problems arising from the use and the construction process of composite indicators also in connection with the SDG framework are described. In Section 3, the explained problems are put in the context of quality principles which have been proposed by several national and international institutions to draw the attention on the importance of data quality for official statistics. One specific quality issue of official data is accuracy which includes sampling and non-sampling errors. The latter concerns for example the existence of missing values in datasets. Missingness and the choice of imputation methods can have a great influence on the results of composite indicators which will be further showcased in Section 5. Additional sources of uncertainty of the construction process are analysed in Section 6 with a variance-based sensitivity and uncertainty analysis. Finally, in Section 7, some important aspects from the study conducted by Münnich and Seger (2014) are presented, focusing on the impact of survey designs on the results of composite indicators. Interesting conclusions about quality problems due to comparing data from surveys with different designs are explained in more detail.

2. General aspects of composite indicators

In general, a composite indicator is as a summary of well-chosen and relevant sub-indicators which are combined or aggregated into a single number to represent a multidimensional construct. Usually, the k sub-indicators ($k = 1, \dots, K$) used to measure the multidimensional construct do not have the same measurement scale and therefore, will be normalised in order to allow for comparability of the outcomes (see, Nhemachena, Matchaya, Nhemachena, Karuaihe, Muchara, and Nhlengethwa 2018, p. 3). The composite indicator is then evaluated for different regions or countries c ($c = 1, \dots, C$) and possibly in different time periods t ($t = 1, \dots, T$). Therefore, the composite indicator $CI_{c,t}$ can be expressed as the function f with

$$CI_{c,t} = f_{c,t}(x^*_{1,c,t}, x^*_{2,c,t}, \dots, x^*_{K,c,t}) \tag{1}$$

where $x^*_{k,c,t}$ denotes the value of the k-th normalised sub-indicator for country c at time period t. In the following sections index t will be omitted, since cross sectional data from the same time period is used for all analyses. The construction of composite indicators encompasses multiple stages in which subjective decisions have to be made. OECD and JRC European Commission (2008) provide an overview for some of the stages and a comprehensive description of different methods within each stage. The selection of the single sub-indicators or sub-indicator variables should be ideally made according to considerations about their relevance, analytical soundness, timeliness and accessibility (see, OECD and JRC European Commission 2008, p. 23). After choosing the sub-indicators for the construction of the composite indicator, it can be necessary to impute in case of missing data. The question of which imputation method should be used cannot be answered in general terms, but depends on the data structure, the missing value scheme and the relation structure between variables in the dataset. An overview of single and multiple imputation methods in connection with data for composite indicators can be found for example in Münnich, Magg, Ohly, and Wiegert (2008a) or OECD and JRC European Commission (2008). Depending on which data is available, it can be distinguished between imputation on micro level or macro level. Micro level data encompasses the data of the individual survey units such as firms in a business survey. Macro level data on the other hand, refers to data of the calculated indices for country c such as for example the manufacturing value added share in GDP. In Section 5 it will be shown how influential the choice of one imputation method can be regarding the variability of the resulting composite indicator values. In this section the imputation will be done on the macro level of the sub-indicators. One reason for missing data can be due to the reporting time of the single sub-indicator values which might be after the composite indicator results are needed. In this case, nowcasting methods can help to close the gap due to time delays in the reporting (see for example, Boudt, Todorov, and Upadhyaya 2008).

After imputing missing records, the values of the sub-indicators have to be normalised if the measurement units of the variables differ. There are construction approaches for composite indicators, for example the benefit of the doubt approach, for which this step is not necessary. Cherchye, Moesen, Rogge, and Puyenbroeck (2007) Choices of normalisation methods can be for example standardisation, min-max normalisation or distance to reference measures (see, OECD and JRC European Commission 2008, pp. 27 and p. 92). Another choice in the construction process concerns the form of the aggregation function f. The normalised data of the sub-indicators can be aggregated using for example linear aggregation methods. Another forms with which the performances of the sub-indicators cannot fully counterbalance each other, are geometric aggregations (see, OECD and JRC European Commission 2008, p. 32). The influence of these choices can be measured using a variance-based sensitivity approach explained for example in Saisana, Saltelli, and Tarantola (2005) or Saltelli, Ratto, Andres, Campolongo, Cariboni, Gatelli, Saisana, and Tarantola (2008). In Section 6 a variance-based sensitivity analysis is outlined for a composite indicator of sustainable economic development. The composite indicator and its construction using data from the SDG framework is explained in Section 4.

Saisana *et al.* (2005) as well as OECD and JRC European Commission (2008) summarise the most important advantages and disadvantages of composite indicators which exceeds the points mentioned below. Due to the aggregation of several sub-indicators into one number they can help to make multi-dimensional issues more understandable and therefore support policy makers to base their decisions on relevant data. Furthermore, they can foster discussion about important topics because the selection of sub-indicators can be subject to political discussion. On the downside, the reduction of complex problems to a single number may lead to a false impression of simplicity and oversimplified policy conclusion. Since the influences of the single dimensions are not reflected anymore in the value of the composite indicator, it can be difficult to determine which policy actions and policy fields should be supported by actions. This could be especially a problem in case of a composite indicator based on the SDG framework, since the topics of the single goals are very diverse (see, OECD and JRC European Commission 2008, p. 13 and Saisana *et al.* 2005, p. 307). Therefore, it could be beneficial to evaluate the influence of every sub-indicator on the value of the composite indicator.

Besides this, in OECD and JRC European Commission (2008) it is explained that composite indicators can be calculated for consecutive years and hence utilised to assess the development of countries. This statement has to be discussed in further research and needs evaluation of additional data. Since the values of the sub-indicators are estimated from a sample, changes in the sub-indicators and the aggregated composite indicators could be only due to sampling errors. An increase in the value of the composite indicator does therefore not necessarily mean a significant improvement of the countries performance from one point in time to the other (see, Münnich and Zins 2011, pp. 26). Further evaluation of the data is needed to draw conclusions about the changes and the development of the country performance. Information about the quality of the sample estimates are needed in order to perform this evaluation. This leads over to the following sections in which data quality problems and their impact on the results of the composite indicators are discussed and illustrated in more detail.

3. Data quality of official statistics

Summarizing information of several variables into a composite indicator can become a critical approach, especially if the data quality of the sub-indicators is considerable different. The issue concerning the need of high-quality official statistical information is summarised in principles which describe the requirements of data quality and concepts of critical steps in the statistical production process. Different institutions propose a framework of several principles to ensure the quality of official statistics and to depict critical steps in the statistical production process. Examples for these frameworks and principles are

- UN (United Nations) Fundamental Principles of Official Statistics (United Nations 2015)
- European Statistics Code of Practice (Eurostat 2017)
- African Charter on Statistics from the African Union (African Union 2009)
- Statistics Canada Quality Guidelines (Statistics Canada 2019)
- South Africa Statistical Quality Assessment Framework (Statistics South Africa 2010)
- Statistics Norway's Dissemination Policy (Statistics Norway 2007).

The 10 UN Fundamental Principles encompass the most important aspects of data quality assurance in the data production process and in United Nations (2015). These principles are explained together with implementation guidelines and connected with principles from the above mentioned frameworks. Here, it will be focused on the UN framework since it is very important for the construction of a composite indicator using the SDG framework. The UN framework encompasses the aspects of data quality in the following principles:

1. Relevance, Impartiality and Equal Access
2. Professional Standards, Scientific Principles, and Professional Ethics
3. Accountability and Transparency
4. Prevention of Misuse
5. Sources of Official Statistics
6. Confidentiality
7. Legislation
8. National Coordination
9. Use of International Standards
10. International Cooperation

Out of the 10 Fundamental Principles, three will be discussed in more detail as they are of special interest for composite indicators and discuss various aspects of data quality. Principle 3 *Accountability and Transparency* requires the users of the statistics to gain access to information which is necessary for the understanding of their characteristics and qualities (see, United

Nations 2015, p. 31). This includes information about the survey design, frame, response rates, editing methods, measurement errors and other methods or procedures used in the data production process. The information about the quality should encompass sampling errors as well as non-sampling errors to cover the total survey error. Sampling errors arise due to the selection mode of a sample, whereas non-sampling errors originate from mistakes or system deficiencies in the data collection process. Sources for non-sampling errors can be the respondent, the interviewer, refusals to participate which will result in non-response, or mistakes in the data-entry process (see, Biemer and Lyberg 2003, p. 36). It is important to report the information on the quality assessment. This will increase the transparency of the production process and will therefore increase the trust and acceptance of the statistical outcome. Rosen (1991) points out that in order to be a helpful tool for discussion and monitoring, composite indicators must gain peer acceptance. Transparency on this note will increase their acceptance. Also, this information is necessary to facilitate correct interpretations of the composite indicator results and to judge its suitability to represent a specific multidimensional construct (see, United Nations 2015, p. 31). Commonly used methods for the quality assessment of construction decisions are variance-based sensitivity and uncertainty analysis, as mentioned before. It is also possible to evaluate the quality with unit-by-unit plausibility checks, unit-by-unit checks with previous records, outlier detection or a comparison of the data at hand with other sources (see, Christiansan and Tortora 1995, p. 251). Another method to assess the quality of the sub-indicators is the so called NUSAP (Numerical, Unit, Spread, Assessment, Pedigree) which helps to clarify if the messages of the sub-indicators are reliable and can be used safely to draw policy conclusion from. Van Der Sluijs, Craye, Funtowicz, Kloprogge, Ravetz, and Risbey (2005) A pedigree matrix is used for every sub-indicator to evaluate each step in the statistical production process with regard to its quality. The modes with which the steps are executed will be summarised as a categorical variable in the matrix. The categories will be rated with numerical scores according to their quality. By doing so, the quality of the process can be easily communicated to the user of the statistics (see, Nardo, Saisana, Saltelli, and Tarantola 2005, p. 14).

Principle 5: *Sources of Official Statistics* can also be discussed in connection with composite indicators. Sources of data for the sub-indicators have to be chosen carefully with regard to the quality, timeliness, costs and the burden on respondents. This principle aims to ensure that data characteristics and the quality are identified beforehand and governed by implemented rules (see, United Nations 2015, p. 31). It is important, because a composite indicator often summarises data from very different types of surveys. For example, a composite indicator based on the SDGs can encompass data form economic surveys, social surveys or environmental surveys. Social surveys with households or individuals as respondents and economic surveys in which businesses, institutions or farms are the units of interest, are likely to have very different data structures and data quality issues. For economic surveys, less standardised designs and survey practices are applied than for social surveys which can make a comparison between statistics from different survey types difficult. Besides this, economic surveys entail specific characteristics which social surveys do not reflect on to the same extent. Some of these issues concern intensively skewed distributions or rapid rate of changes in the data and estimated sub-indicators. Additionally, more data alternatives for economic surveys are available such as for example administrative records (see, Cox and Chinnappa 1995, pp. 2).

Principle 9: For the *use of international standards* it is important to promote the utilisation of commonly agreed methods and classification systems from statistical agencies. This helps to ensure comparability between statistics from different countries or sources within a country and it is a dimension of quality which can be communicated via published metadata. Comparability is an important dimension of quality, especially if the statistics are used to compare the performance of countries or regions. Standard methods in the statistical production process can enhance the explanatory power of the comparisons and improve the efficiency of the process (see, United Nations 2015, p. 80). Though, the level of standardisation might be restricted due to country specific circumstances and particularities. In economic surveys these can for example influence the survey frame. With regard to the data used for a composite indicator based on the SDG framework, the Intern-Agency and Expert Group on SDG indicators (IAEG-SDG) developed criteria and guidelines for regulating the data flow between countries and custodian agencies. These are responsible to collect, analyse and report the SDGs. The guidelines are aimed to ensure the quality of the reported estimates and the harmonisation of the single SDG target variables over the countries. Gennari and Navarro (2019) discuss three main issues of the guidelines on data flows and global data reporting which could lead to inconsistencies in the reporting of the SDG target variables. The authors discuss the problem of inconsistencies in the data validation process between national statistical organisations and the custodian organisation, the problem that specific provisions of the guidelines are not followed in practice and the absence of detailed modalities on the data validation process (see, Gennari and Navarro 2019, p. 738). This could lead to a situation in which the data quality of countries differs and comparisons of SDG indicators between countries might not be fair. Further literature, like Thomas, Silvestre, Salentine, Reynolds, and Smith (2016) and Sarvajayakesavalu (2015) describes the challenges of countries to implement the statistical infrastructure which is needed to report quality indicators in more detail.

4. A composite indicator on sustainable economic development

Composite indicators using the SDG framework could be generally constructed in two different ways. The SDG index or composite indicator proposed in Sachs *et al.* (2019) is constructed using a basket of the targets from the 17 SDGs as sub-indicators and aggregating them with equal weighting. In the case of a composite indicator for the sustainable development of countries, this could mean that countries decide on either the single targets of the 17 goals or on the 17 goals themselves to construct the country specific composite indicator. This composite indicator could be constructed with a combination of targets applicable for every country and targets which are not relevant for all countries (see, Melamed and Bergh 2014, p. 4). On SDG level for example goal 14: Life Below Water might not be relevant for landlocked countries.

Additionally, it is possible to use the SDG framework to construct a composite indicator measuring specific aspects of sustainable development. This has been done for example by Rickels, Dovern, Hoffmann, Quaas, Schmidt, and Visbeck (2016). The authors build a composite indicator with target variables from SDG 14 to measure oceanic development in the European Union. Another example is the composite indicator on agricultural development by Nhemachena *et al.* (2018) which uses target variables from SDG 1, 2, 6, 7, 15. For the industry-related SDG 9, Luken, Saieed, and Magvasi (2022) have used two indices to assess the progress and performance of 20 sub-Saharan African countries.

In alignment with these ideas, the composite indicator used in the following studies is built as a composite indicator on sustainable economic development from a specific set of the 17 SDGs and target variables as sub-indicators. Wu, Guo, Huang, Liu, and Xiang (2018) explain that the 17 SDGs can be split up into three dimensions of sustainable development according to human needs. Each dimension is represented by a different set of SDGs. Sustainable economic development is represented by SDG 1, 2, 3, 8, 9. Additionally, the social dimension of sustainable development is represented by SDG 4, 5, 10, 11, 16 and 17. Lastly, the set of SDG 6, 7, 12, 13, 14, 15 represents the environmental dimension of sustainable development (see, Wu *et al.* 2018, p. 4).

To construct the composite indicator of sustainable economic development, country level data on sub-indicators or target variables from the five corresponding goals (SDG 1, 2, 3, 8 and 9) and from the year 2015 was downloaded from `https://unstats.un.org/sdgs/indicators/database/` [1]. The sub-indicator variables are shown in Table 1 ordered by the sustainable development goal and the corresponding target within each goal. A selection of sub-indicator

[1] Last access: 4 July 2020

variables and the year of the data reporting has taken place based on data availability and suitability of the data structure. Data was taken from 20 countries with different economic profiles. Since the following studies refer to data specific issues and do not try to make any implications about country specific economic profiles, indices rather than country names are going to be used to refer to the countries.

Table 1: Sub-indicators of the composite indicator on sustainable economic development

SDG	Sub-indicators of the composite indicator of sustainable economic development
1 NO POVERTY	1.1.1 Proportion of population below international poverty line (%)
	1.4.1 Proportion of population using basic drinking water services (%)
2 ZERO HUNGER	2.1.1 Prevalence of undernourishment (%)
	2.a1 Agriculture orientation index for government expenditures
3 GOOD HEALTH AND WELL-BEING	3.8.1 Universal health coverage (UHC) service coverage index
	3.c1 Health worker density, by type of occupation (per 10,000 population)
8 DECENT WORK AND ECONOMIC GROWTH	8.1.1 Annual growth rate of real GDP per capita (growth factor)
	8.2.1 Annual growth rate of real GDP per employed person (growth factor)
	8.6.1 Proportion of youth not in education, employment or training (%)
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE	9.2.1 Manufacturing value added as a proportion of GDP (%)
	9.3.1 Proportion of small-scale industries in total industry value added (%)
	9.b1 Proportion of medium and high-tech industry value added (%)
	9.c1 Proportion of population covered by at least a 4G mobile network (%)

As shown in Table 1, the sub-indicator variables $x_{k,c}$ are not all measured using the same measurement unit. In order to make the measurements comparable when building the composite indicator of sustainable economic development a min-max normalisation is applied, using the following formula

$$x^*_{k,c} = \frac{x_{k,c} - min(x_k)}{max(x_k) - min(x_k)} \times 100 \quad . \tag{2}$$

Due to Formula (2) and after the aggregation of the sub-indicators, the resulting composite indicator values will range between 0 and 100, similar to Nhemachena *et al.* (2018) and Sachs *et al.* (2019). In the step of normalisation it is also important to account for the polarisation or direction of the sub-indicators. Greater values of positive sub-indicators will indicate greater sustainable economic development, whereas greater values of negative sub-indicators refer to lower sustainable economic development. From Table 1 it can be understood that the sub-indicators 1.1.1, 2.1.1 and 8.6.1 are cases of negative sub-indicators. Their scale will be reversed for consistency in the interpretation of the resulting composite indicator values, before the aggregation step. Therefore, greater values of the composite indicator will indicate greater sustainable economic development. To construct the composite indicator of sustainable economic development, the sub-indicators are equally weighted and aggregated using the arithmetic mean, similarly to the construction of the composite indicator of agricultural development by Nhemachena *et al.* (2018). A very important step of the construction process is the handling of missing data which actually takes place before normalisation and aggregation of the sub-indicators. So far, it has been assumed that no missing data was present in the dataset. Section 5 aims to visualise how the choice of imputation methods, as a common method to handle missing data, can influence the results of the composite indicators and country comparisons.

5. Impact of missing values and the choice of imputation method

In this section, a small case study showing the impact of imputation on the results of the composite indicator of sustainable economic development is outlined. Figure 1 describes the structure of missing values in the data set of 20 countries used subsequently to calculate the composite indicator of sustainable economic development. The left side of Figure 1 visualises the amount of missing records for each variable in the dataset. Some of the sub-indicator variables show high frequencies of missing records such as for example sub-indicator 1.1.1 with 45% of the values missing. For sub-indicator 9.3.1 40% of the values are missing. Sub-indicator 2.1.1 and 9.b1 for example are recorded completely for all countries in the dataset.

The right side of Figure 1 shows all apparent combinations of missing and non-missing values in the sub-indicator variables. From the missing value frequencies of variable pairs it can be seen that 11 different patterns of missing values can be found in the dataset. Since the variables of sub-indicators 1.1.1 and 9.3.1 have the most missing records, most of the patterns include these two variables. 25% of the cases in the dataset have no missing record at all.

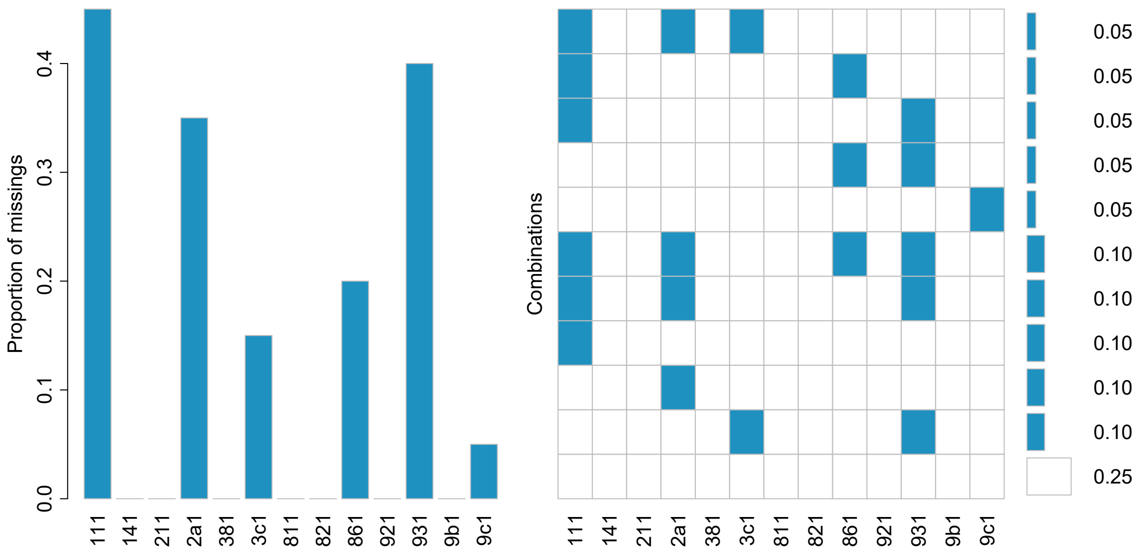

Figure 1: Missing data pattern

OECD and JRC European Commission (2008) introduce different imputation methods for the construction of composite indicators and also propose rules of thumb in order to choose between them. Generally, imputation methods are grouped in single imputation and multiple imputation methods. A further overview of imputation methods can be found in Meinfelder (2014), Zhang (2003) or OECD and JRC European Commission (2008). Even though, there is no definite answer, the rules of thumb for choosing an imputation method by OECD and JRC European Commission (2008) depend on the scale of the variables, the amount of missingness in the dataset and the relationships of the sub-indicator variables with missingness of the corresponding country. The performance of the imputation methods could be examined by applying an *in sample/ out of sample* logic. For this the complete part of the dataset is taken and similar missingness is introduced. After this step, the different considered imputation methods are applied and their performance is compared for example by measuring the correlation between the complete and imputed dataset (see, OECD and JRC European Commission 2008, p. 62).

To show the influence of the imputation method choice on the results of the composite indicator of sustainable economic development mean imputation, hot deck imputation and the predictive mean matching imputation as multiple imputation method are applied using the R package MICE (see, van Buuren and Groothuis-Oudshoorn 2011). 5 imputed datasets are generated by the predictive mean matching imputation. For each of the 5 datasets, the composite indicator is calculated as described above for each country. These results are pooled using an arithmetic mean. This method has the advantage that the imputed values are restricted to the observed values and also that it can preserve non-linear relationships (see, van Buuren and Groothuis-Oudshoorn 2011, p. 18). For a more detailed explanation of the method see Little (1988). This is also true for the mean imputation and the hot deck imputation as single imputation methods.

In Figure 2 the distribution of the composite indicator results for each country calculated from the imputed datasets are shown by boxplots. The boxplots include also the results of the 5 imputed datasets from the predictive mean matching method. In addition, the boxplots are ordered according to the pooled results based on the predictive mean matching method. As explained in Section 2, ranking and comparisons of the observed regions or countries are often times one of the main interests in the construction of composite indicators. Higher values of the composite indicator refer to higher sustainable economic development. Therefore, using the predictive mean matching imputation, country 11 is found to have the highest sustainable economic development. Country 20 shows the lowest sustainable economic development. Results from the three imputation methods are also highlighted with coloured squares to inspect if the results had come to similar conclusions with regards to the composite indicator values and the country rankings. No squares are shown for countries which have complete datasets as no data was imputed.

First of all, it can be seen that country results differ in their range of resulting composite indicator values. Some of this variation can be easily explained by the amount of missingness for each country which is indicated with the percentage numbers above each boxplot. In case of a small amount of missing data the imputation method might not have a great effect on the resulting composite indicators. Some countries with larger amount of missingness such as country 10 also do not show very different results of the composite indicator calculated with different imputation methods. For most of the countries with lager amount of missing data, it can be seen that the results using mean imputation, hot deck imputation, or predictive mean matching imputation differ to a larger extent. This may lead to problems in the interpretation of the output. If the resulting composite indicator values are influenced by the chosen imputation method, country rankings and comparisons will not solely reflect on the country performances. Instead, they will also reflect on possible country specific choices in the construction process of the composite indicator. In this case the choice of imputation method in combination with the amount of missing data. Examples for this in Figure 2 are country 5 and 12 which will change places in the country ranking if the hot deck imputation method is used to calculate the composite indicator rather than predictive mean matching. Another example is country 14 which will change ranking with country 3 if mean imputation will be used to impute the missing data before calculating the composite indicator. Ranking might be less distinct and the performance judgement of a country will highly depend on the imputation method used on the data.

The yellow squares in figure 2 indicate the results in case of leaving the sub-indicators with missingness out of the calculation of the composite indicator for each respective country. Therefore, the country specific calculations will be based on different sets of sub-indicators in case of missingness in the data. These results visualise how the country rankings change if different sets of sub-indicators, for example due to missingness, are used to calculate the composite indicator. Again, we can see that the results differ considerably between the different methods.

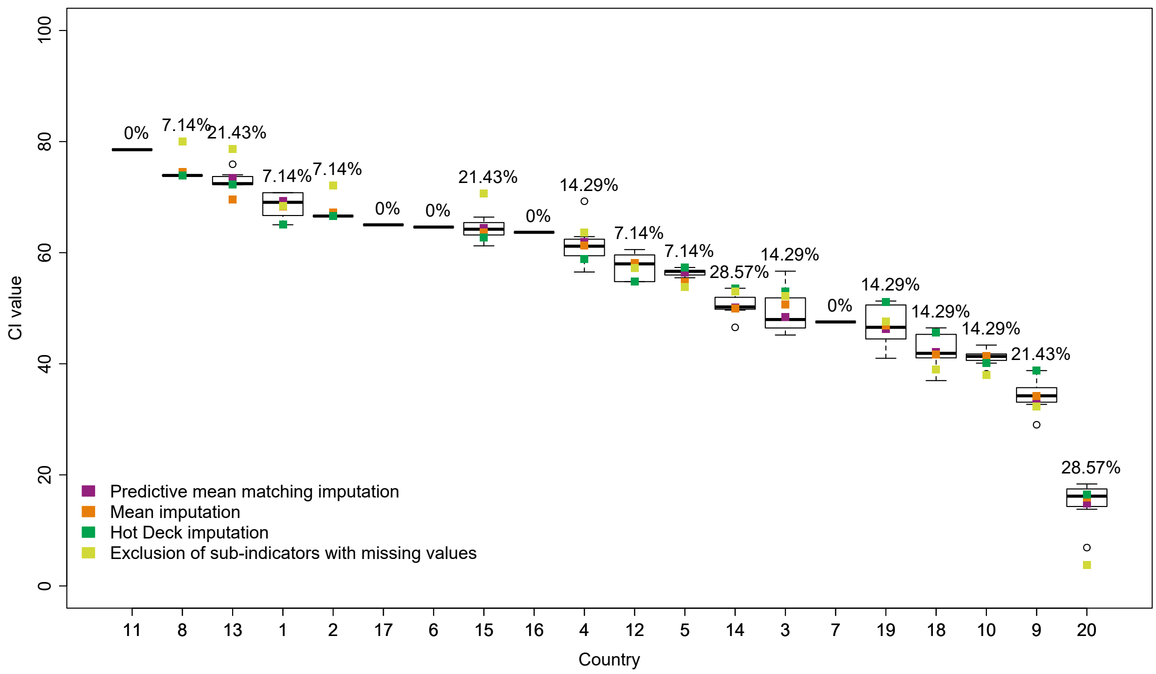

Figure 2: Distribution of composite indicator values with and without imputation

It is questionable how meaningful comparisons of performance developments are, if different imputation methods or a different sets of sub-indicators due to missing values or other considerations are used. The method of imputation should therefore be always chosen due to discussion and based on the availability of additional data and further knowledge of the sub-indicator's relationships. Furthermore, it is important to report the used methods in the metadata to improve the interpretability of resulting country rankings. For this example, data was imputed on macro level because the dataset already included estimates of the sub-indicator variables. In general, it can be helpful to use micro level data, thereby additional information and country characteristics can be used to enhance the imputation model.

In the next section, a sensitivity and uncertainty analysis is used to further show how composite indicator results may be affected by the subjective choices made throughout the construction process. This analysis will look at the choice of the sub-indicator set, the normalisation method and the aggregation method.

6. Sensitivity and uncertainty analysis

The construction process of a composite indicator includes different stages in which more or less subjective choices have to be made. For example choices about the sub-indicator variables, the normalisation method or the aggregation method, are necessary. Saisana *et al.* (2005) or Saltelli *et al.* (2008) describe a variance-based sensitivity analysis and an uncertainty analysis for composite indicators. This analysis evaluates the impact of decisions on the result of an output variable. For example the composite indicator values itself or the corresponding country rank. In this setup, choices in the construction process are referred to as input triggers or input variables. The composite indicator values or country rankings are calculated due to every possible combination of input triggers. To quantify the impact of each input trigger, the first-order sensitivity indices can be calculated. These are based on a decomposition of the total output variance. The first-order sensitivity index describes the direct influence of one input trigger on the output variance. Therefore, it represents how much of the output variance can be reduced by fixing the corresponding input trigger to a certain decision (see, Saltelli *et al.* 2008, p. 21 and Saisana *et al.* 2005, p. 12). In the uncertainty analysis the distributions of output variables are graphically illustrated by boxplots or by visualising the relative frequencies of the achieved ranks by country. Generally, country rankings aim to evaluate a country's performance relative to other countries in the observed population. Therefore, it is important to analyse how distinct the rankings based on a composite indicator are. If a rank position of a country depends strongly on the methods used to construct the composite indicator and changes greatly under the use of different methods, it is questionable how meaningful the country's performance is. This is because the rank results might not solely depend on the country performance itself.

Below, a variance-based sensitivity and uncertainty analysis of the composite indicator of sustainable economic development as described above and in Saisana *et al.* (2005) or Saltelli *et al.* (2008) will be discussed. For the purpose of this analysis a complete dataset will be used which resulted from the application of the hot deck imputation. This sensitivity and uncertainty analysis aims to show how strong the influence of the choice of the sub-indicator set, the normalisation method and the aggregation method on the composite indicator values and the country rankings can be.

For the sensitivity and uncertainty analysis the composite indicator of sustainable economic development is calculated 84 times based on all suitable input trigger combinations. For each calculation either one or none of the 13 sub-indicators are excluded from the considered set. As normalisation method either the min-max normalisation or a standardisation is used by applying the function normalise_ci from the R package Compind (see, Fusco, Vidoli, and Sahoo 2018). The choice of the aggregation method is also strongly connected to the weighting of the single sub-indicators. Not all sub-indicators are weighted equally when applying different aggregation methods. As in Section 5, the first chosen aggregation method is the arithmetic mean, with which all sub-indicators are equally weighted. OECD and JRC European Commission (2008) also suggest the use of a geometric mean, as well as the aggregation of the sub-indicators in connection with a principle component weighting approach. Countries with higher sub-indicator scores will be rewarded when using the geometric mean as aggregation method. Because with this method is more difficult for the countries to compensate low sub-indicator values with higher sub-indicator values. This is not the case for the arithmetic mean as aggregation method (see, OECD and JRC European Commission 2008, p. 32). Additionally to using an equal weighting scheme, the weights of the sub-indicators can be calculated using the principle component approach. Thereby, the weights are calculated due to the correlations of the sub-indicators and the aggregation is done by a weighted arithmetic mean. An in depth description of how to calculate the weights using the principal component approach can be found in OECD and JRC European Commission (2008). The fourth aggregation method used in this sensitivity analysis is explained in De Muro, Mazziotta, and Pareto (2012) and is called the Mazziotta-Pareto Index (MPI). This composite indicator starts with a linear aggregation and introduces a penalty for unbalanced values of sub-indicator sets using the coefficient of variation. Therefore, this approach can be applied if the set of sub-indicators is considered to be non-substitutable (see, De Muro *et al.* 2012, pp. 8). For an in depth explanation of the construction of this composite indicator the reader is referred to De Muro *et al.* (2012). Not all combinations of normalisation methods and aggregation methods are applied in this analysis. The MPI index aggregation is applied using only the standardisation method and the geometric mean aggregation is applied using the min-max method in the step of the normalisation. This will lead to 84 combination possibilities of the trigger values for the sensitivity analysis.

In Figure 3 the results of the first-order sensitivity indices for the three triggers of the sensitivity analysis are shown. They are calculated using the function multisensi from the homonymous R Package (see, Lamboni, Makowski, Lehuger, Gabrielle, and Monod 2009). For an explanation of how to calculate the first-order sensitivity indices see for example Saltelli *et al.* (2008), Chapter 4. In this setting countries are affected by the choices made in the construction process to a different degree, but it is noticeable that for most countries the aggregation method or the normalisation method has the largest influence on the outcome of the composite indicator calculation. It can be also seen that the influence of the input triggers is country specific, as the calculation of the composite indicator for country 5 and 14 depends mostly on the choice of the sub-indicator set.

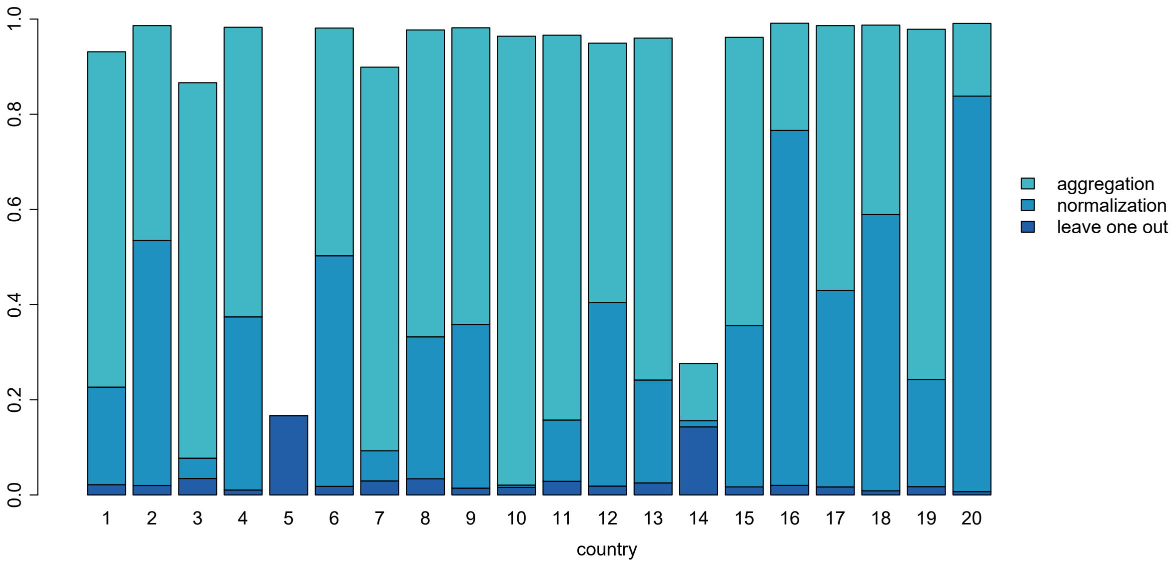

Figure 3: First-order sensitivity indices

In the uncertainty analysis, the resulting country ranking distributions are analysed. They are represented in Figure 4 which visualises the relative frequencies of how often each country takes on each rank position. Deeper colours indicate a higher relative frequency of the corresponding rank. The red sign in the figure points out the mode rank of each country. On the x axis the countries are ranked by this measure for readability of the plot.

It can be concluded from Figure 4 that due to the construction choices a comparison based on country rankings has to be viewed critically. Most of the countries take on a large range of rank positions over all 84 calculations. This can be seen for example for country 1 or country 5 which take on almost all ranks with similar relative frequencies. Therefore, the country rankings are not distinct. Also, most of the ranks are coloured in similar blue shades and their mode ranks are coloured in light blue which implies that they are not achieved with high relative frequencies. No country takes on only one rank in this relative comparison over all calculations. However, most countries take on multiple different ranking positions over all calculations with similar frequencies. The ranking of a country very much depends on the construction decisions and not necessarily on the country's performance on the sub-indicators.

The results of the sensitivity and uncertainty analysis show that using rankings to compare country performances has to be viewed critically. Very different rankings can result due to different methods used to construct the composite indicator of sustainable economic development. It can be concluded that the normalisation and aggregation method have a great impact on the rank positions for most of the countries. But these triggers are only two examples of critical choices in the construction process. The varying results for the composite indicator and the country rankings lead to the problem, that the performances of countries cannot be compared distinctively. It is therefore very important to analyse the composite indicator values in more detail. This includes an evaluation of the results with regard to their sensitivity towards method choices and also the consideration of country performances on single sub-indicators, for example in dashboards. For a detailed analysis of the results it is furthermore necessary to account for additional information on the data quality to avoid

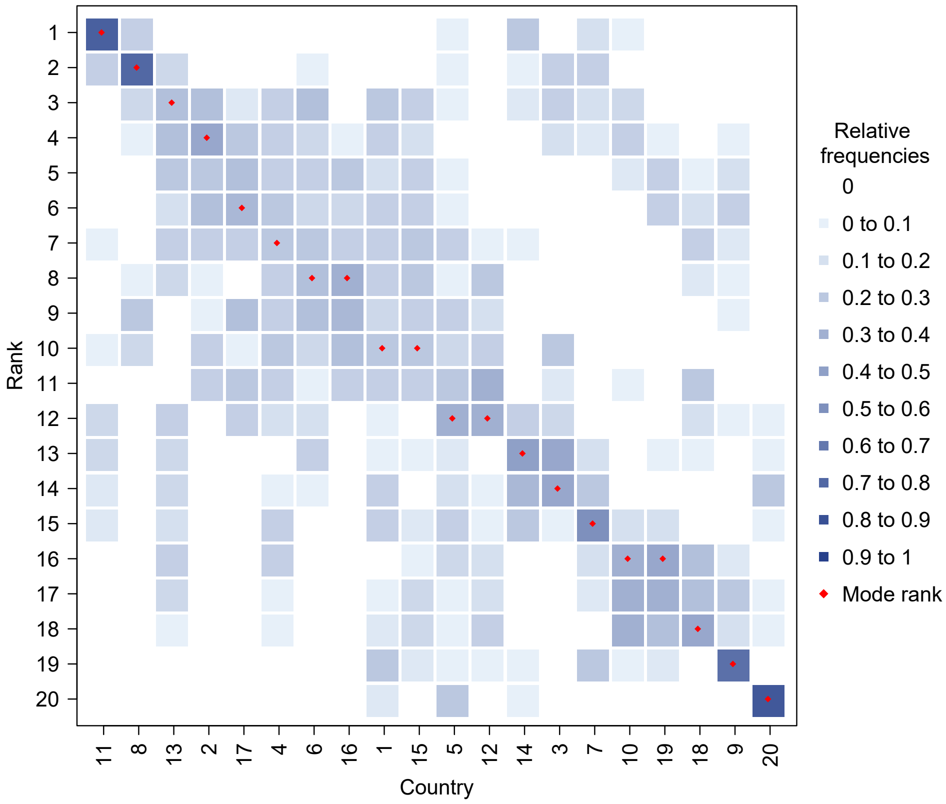

Figure 4: Relative frequencies of country ranks

misinterpretations of the results, such as for example the missing data structure and applied imputation methods.

7. Impact of survey quality on composite indicators

Additionally to the necessary decisions of the construction process which have been discussed before, only little attention of the impact of sampling designs on composite indicator values can be recognized. Due to the lack of additional metadata on the sampling scheme, often the sub-indicator estimators are combined without taking into account that they are based on several surveys with different sampling designs. The study by Münnich and Seger (2014) uses the variance-based sensitivity approach with an input trigger for different sampling designs and shows how sampling designs can influence the outcome of composite indicators and country rankings. Some of the main results are summarized in this section but for further reading the authors refer to the actual paper which is listed in the literature section below. The calculated composite indicator is chosen from the context of poverty and social exclusion measurement in the style of the Mazziotta-Pareto Index described in De Muro *et al.* (2012). It entails the following sub-indicators:

- At-risk-of-poverty rate
- Gini coefficient
- Quintile share ratio
- self-reported unmet need for medical examination or treatment
- medium- and long-term unemployment rate
- education.

The last sub-indicator on education is measured by the proportion of the total population (25 – 64 years old) who have achieved level two or less on the International Standard Classification of Education (ISCED) (see, Münnich and Seger 2014, p. 274). Due to its suitable structure the open dataset AMELIA from the AMELI project was used to execute the analysis. A description of this dataset can be found in Merkle, Burgard, and Münnich (2016). It is an extensive close-to-reality synthetic dataset which is based on the EU-SILC scientific use file of the year 2005 and entails information about regional membership of the households needed for the sampling schemes. As for the sensitivity analysis, four different input triggers are included in the evaluation. The first input trigger relates to the sampling design which is used to draw the data. Three different three stage sampling designs are used which differ in their second stage. Furthermore, 10,000 samples are drawn with each of the sampling designs and hence the resulting high number of 30,000 samples allow for a better fit of the estimation distribution. On the first stage of each sampling design stratified random sampling (STR) is applied with proportional allocation and region as stratification variable. The resulting strata represent the different regions of interest and by method it is ensured that the sample sizes in each region do not vary. On the second stage a sample of households is drawn in every region according to simple random sampling (SRS), stratified random sampling or cluster sampling (CS). The sub-indicators are subsequently estimated based on the information of one person within a household which in this case is considered to be a cluster. For the stratification in the second stage, 20 strata with equal sample sizes in each region are created. A certain level of heterogeneity between the strata is ensured by ordering the households in every region according to the income variable. Then, the three non-monetary sub-indicators are used to split each strata into two parts. One part includes low values of these variables and the other part high values. This procedure results in 20 strata per region with low variation within a strata for all sub-indicators and high variation between the strata. The clustering for the second stage is done based on the 20 strata. Clusters are build by combining 1 to 30 households which are ordered within the strata beforehand. As ordering variables firstly income is used and then a combination of the non-monetary sub-indicators. Neighbouring households are finally combined into clusters. It has to be mentioned that for the STR and SRS design equal sampling sizes of 10,000 can be achieved whereas for the CS the sample size may vary due to the cluster sizes. Therefore, it is drawn with expected sample size of 10,000 (see, Münnich and Seger 2014, pp. 275).

The second input trigger concerns the choice of the normalisation method whereby either standardization, min-max or distance to reference normalisation is applied in this study. As in the analysis above, Münnich and Seger (2014) include a leave one sub-indicator out trigger and as the fourth trigger they incorporate the weighting scheme and a choice between equal weights, principle component weights and random weights (see, Münnich and Seger 2014, p. 278).

As a first step, the design effects of the stratified and cluster sampling are compared over all sub-indicators. It is measured with a ratio between the variance of an estimator under a complex design and the variance of an equivalent estimator under SRS. STR sampling results to out be more efficient than SRS whereas CS on the other hand turns out to be very inefficient over all sub-indicators and regions (see, Münnich and Seger 2014, p. 276). Comparing the four input triggers with their corresponding first-order sensitivity indices the results show, that the choice between normalisation methods has the biggest influence on the output variance even though the impact varied greatly between the regions. Interestingly, the choice of the sampling design represents the second largest influence on the composite indicator values. The leave one sub-indicator out trigger and the weighting trigger have considerably less influence on the outcome.

Since the study aims to focus mainly on the impact of sampling designs, three additional studies are executed in which all samples are drawn separately according to only one design. As a results their impact can be evaluated in more detail and it appears that the direct impact on the output declines with increasing efficiency of the sampling design. Therefore the first-order sensitivity effects are greater for CS than for STR. In fact, the first-order sensitivity indices resulting for STR are similar to the indices of SRS, which can be explained by the design effect. The first-order sensitivity indices can be interpreted as a measure of uncertainty and with total knowledge about the population it would be zero (see, Münnich and Seger 2014, p. 272).

For the uncertainty analysis of the country rankings, calculated with the different construction choices, the results show that each country can take on each rank in at least a few cases. This leads to indistinct country rankings. A comparison of countries based on rankings has to be viewed critically. The same results are concluded from an additional study in which the standardized values and equal weights were chosen and the leave one sub-indicator out trigger was excluded. Based on the relative frequencies of each country and rank position it is shown that again almost every position is taken at least once from all countries. This leads to the conclusion that with extreme samples basically all ranking results are possible regardless of the other choices in the construction process. If rankings are done conditional to only one of the sample designs, the results of a design comparison are as expected. The rank position variation is considerably reduced with SRT, but CS still leads to variation. From the mode ranks it can be concluded that SRS and STR again lead to similar results, whereas CS has some difficulties in ranking some of the regions distinctively. The overall results make it clear how important it is to use efficient estimators if sensible designs are applied. The choice of suitable estimators will help to improve the accuracy of the composite indicators considerably, especially if available auxiliary information is used properly (see, Münnich and Seger 2014, p. 283).

It is known that due to the administrative structures, surveys from different countries are often conducted with different survey designs. In a further analysis, the composite indicator is estimated for different regions with different sampling designs and the results are compared to the true value calculated from the population. This can be done because it is sampled from an artificial dataset which is considered as the population. The results show clearly different levels of accuracy for the estimated composite indicator values in the regions. Some of the results are negatively biased and the regions are evaluated with a lower performance on poverty and social exclusion due to the choice of sampling design. The bias of the estimated indicators and also the very sensitive country rankings due to necessary decisions in the construction process yield substantial problems. Because the bias of the estimates is usually unknown without further research and because the country rankings entail a great degree of sensitivity, political decisions drawn from these results can be based on vague information about the country performances. It is therefore very important to consider additional information about the data production process to ensure that reliable sub-indicators are used for the calculation of the composite indicator (see, Münnich and Seger 2014, pp. 286).

8. Concluding remarks

The present paper highlights and summarizes some problems when drawing conclusions from composite indicators used for informed political decision making. This concerns comparisons of country performances as well as measuring development. The major focus is laid on data quality aspects, such as the impact of missing values and the data gathering process, i.e. the sampling procedure. Missing values and the choice of the imputation method, in general, can have a big influence on the outcome of composite indicators. This is shown in an analysis comparing calculation results with imputed datasets using single and multiple imputation methods. Providing information on the missingness for each sub-indicator and the used imputation method is therefore very important and can help to gain a further insight in the quality of the data. This documentation should go hand in hand with a commonly used process on how to chose the most suitable imputation method over the countries of the comparison in order to ensure consistency.

The sensitivity and uncertainty analysis of the composite indicator of sustainable economic development showed, that the results of the country rankings can be changed greatly by the choice of the normalisation and aggregation method. Different aggregation methods handle substitutability of sub-indicators differently and therefore, countries with balanced or unbalanced sub-indicator profiles will benefit differently from the chosen aggregation method. These choices should be discussed intensively and justified considering the insights the analysts intended to gain from the the composite indicator. Also, the choice of the sub-indicator set can influence the ranking of a country. This circumstance might be of special interest for a composite indicator constructed from the SDG framework. Not all SDGs and their targets are important for all countries, or target variables are not available for each country. The extent to which countries can be compared if the composite indicator is calculated based on country specific sub-indicator sets has to be discussed critically and the influence of this choice should not be overlooked.

Regarding the sample designs used to gather information from the population, it is demonstrated by Münnich and Seger (2014) how the sampling design effects the outcome of composite indicators. Even if this is well-known, the possible amount seems surprising. Further, in mixed design environments which is especially likely for SDG sub-indicator variables coming from very different sources, the possible non-linearities of the composite indicators due to rescaling may lead to unknown biased sub-indicator values. Special attention should be drawn on this effect by using adequate estimation methods, e.g. calibration, to improve survey quality (see, Münnich, Huergo, and Enderle 2008b, p. 39). Indeed, in composite indicator practice, often auxiliary information is not used to improve the accuracy of sub-indicators and, hence, also the overall composite indicator. Modern survey statistics methods, including small area estimation and multiple imputation, may then help to improve the quality of sub-indicators considerably. This automatically fosters better conclusions for policy.

Results of composite indicators and relative rankings of countries should not be interpreted on their own. Additional steps in the interpretation include an in depth look into the construction process of the composite indicator with a sensitivity and uncertainty analysis. Also, the country specific performance on the single sub-indicators using for example dashboards and an analysis of quality of the data with which the sub-indicators were estimated are vital and make the composite indicator results more reliable.

Finally, we want to stress that high quality statistics starts from the scratch of the statistical production process. Hence, indicator selection and according data gathering methods should always be developed commonly amongst area-specific experts, policy support, and (survey) statisticians. As we have learned from the Covid crisis, it is important to include statistics experts, and especially those whose focus is on data gathering, sampling, and data processing, at the very beginning of the process. For the SDG indicators, additional attention has to be paid on the fact that for country-wise comparisons, quality aspects which are stressed for example in the European Statistics Code of Practice (Eurostat 2017) or the other quality frameworks mentioned above, have to be strongly taken into account, especially with regard to independence and robustness.

References

African Union (2009). "African Charter on Statistics." *Technical report*, African Union. URL https://au.int/en/treaties/african-charter-statistics.

Biemer PP, Lyberg LE (2003). *Introduction to Survey Quality*. Wiley Series in Probability and Mathematical Statistics. doi:10.1002/0471458740.

Boudt K, Todorov V, Upadhyaya S (2008). "Nowcasting Manufacturing Value Added for Cross-Country Comparison." *Statistical Journal of the IAOS*, **26**(1), 15–20. doi:10.3233/SJI-2009-0694.

Cherchye L, Moesen W, Rogge N, Puyenbroeck TV (2007). "An Introduction to 'Benefit of the Doubt' Composite Indicators." *Social Indicators Research*, **82**(1), 111–145. doi:10.1007/s11205-006-9029-7.

Christiansan A, Tortora RD (1995). "Issues in Surveying Business: An International Survey." In BG Cox, DA Binder, BN Chinnappa, A Christianson, MJ Colledge, PS Kott (eds.), *Business Survey Methods*, pp. 237 – 256. Wiley Series & Sons: New York et al. doi:10.1002/9781118150504.ch14.

Cox B, Chinnappa B (1995). "Unique Features of Business Surveys." In BG Cox, DA Binder, BN Chinnappa, A Christianson, MJ Colledge, PS Kott (eds.), *Business Survey Methods*, pp. 1 – 17. Wiley Series & Sons: New York et al. doi:10.1002/9781118150504.ch1.

De Muro P, Mazziotta M, Pareto A (2012). "Composite Indices of Development and Poverty: An Application to MDGs." *Social Indicators Research*, **104**(1), 1–18. doi:10.1007/s11205-010-9727-z.

Eurostat (2017). "European Statistics Code of Practice: For the National and Community Statistical Authorities." *Technical report*, European Statistical System. doi:doi:10.2785/798269. URL https://ec.europa.eu/eurostat/web/products-catalogues/-/KS-02-18-142.

Fusco E, Vidoli F, Sahoo BK (2018). "Spatial Heterogeneity in Composite Indicator: A Methodological Proposal." *Omega*, **77**, 1–14. doi:10.1016/j.omega.2017.04.007.

Gennari P, Navarro DK (2019). "Validation of Methods and Data for SDG Indicators." *Statistical Journal of the IAOS Preprint*, **35**(4), 735–741. doi:10.3233/SJI-190519.

Güdemann L, Münnich R (2021). "Quality Assessment of Composite Indicators from the SDG Framework." In United Nations Industrial Development Organization (ed.), *Modernizing Products and Services of Industrial Statistics*, pp. 26–59. URL https://stat.unido.org/content/publications/modernizing-products-and-services-of-industrial-statistics;jsessionid=A051A8451EBD30F2E56804B44C7251EE.

Lamboni M, Makowski D, Lehuger S, Gabrielle B, Monod H (2009). "Multivariate Global Sensitivity Analysis for Dynamic Crop Models." *Field Crops Research*, **113**, 312–320. doi:10.1016/j.fcr.2009.06.007.

Little RJA (1988). "Missing Data Adjustments in Large Surveys (with Discussion)." *Journal of Business Economics and Statistics*, **6**(3), 287–301. doi:10.1080/07350015.1988.10509663.

Luken R, Saieed A, Magvasi M (2022). "Industry-related Sustainable Development Goal-9 Progress and Performance Indices and Policies for Sub-Saharan African Countries." *Environmental Development*, **42**, 100694. doi:10.1016/j.envdev.2021.100694.

Meinfelder F (2014). "Multiple Imputation: An Attempt to Retell the Evolutionary Process." *AStA Wirtschafts- und Sozialstatistisches Archiv*, **8**, 249—-267. doi:10.1007/s11943-014-0151-8.

Melamed C, Bergh G (2014). "Sustainable Development Goals and Targets: Options for Differentiating between Countries." *Technical report*, ODI Report.

Merkle H, Burgard JP, Münnich R (2016). "The Amelia Dataset – A Synthetic Universe for Reproducible Research." *Technical report*, InGRID Deliverable 23.1 (2016). URL https://zenodo.org/record/1306634.

Münnich R, Huergo L, Enderle T (2008b). "Imputation of Knowledge Economy Indicators." *Technical report*, KEI Workpackage 3. Deliverable 3.2. URL https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/KEI-WP3-D3.2.pdf.

Münnich R, Magg K, Ohly D, Wiegert R (2008a). "Accuracy Measurement of Composite Indicators." *Technical report*, KEI Workpackage 3. Deliverable 3.4. URL https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/KEI-WP3-D3.4_01.pdf.

Münnich R, Seger JG (2014). "Impact of survey quality on composite indicators." *Sustainability Accounting, Management and Policy Journal*, **5**(3), 268–291. doi:10.1108/SAMPJ-10-2013-0045.

Münnich R, Zins S (2011). "Variance Estimation for Indicators of Poverty and Social Exclusion." *Technical report*, AMELI Workpackage 3. Deliverable 3.2. URL https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Ameli_Delivrables/AMELI-WP3-D3.2-20110515.pdf.

Nardo M, Saisana M, Saltelli A, Tarantola S (2005). "Input to Handbook of Good Practices for Composite Indicators' Development." *Technical report*, KEI Workpackage 5. Deliverable 5.2. URL https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/KEI-WP5-D5.2.pdf.

Nhemachena C, Matchaya G, Nhemachena CR, Karuaihe S, Muchara B, Nhlengethwa S (2018). "Measuring Baseline Agriculture-related Sustainable Development Goals Index for Southern Africa." *Sustainability*, **10**(3), 849–865. doi:10.3390/su10030849.

OECD and JRC European Commission (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD publishing. URL https://www.oecd.org/sdd/42495745.pdf.

Rickels W, Dovern J, Hoffmann J, Quaas MF, Schmidt JO, Visbeck M (2016). "Indicators for Monitoring Sustainable Development Goals: An Application to Oceanic Development in the European Union." *Earth's Future*, **4**(5), 252–267. doi:10.1002/2016EF000353.

Rosen R (1991). *A Review of Life Itself: A Comprehensive Inquiry into Nature, Origin and Fabrication of Life*. Columbia University Press. doi:10.1080/03081079308945090.

Sachs J, Schmidt-Traub G, Kroll C, Lafortune G, Fuller G (2019). "Sustainable Development Report 2019." *Technical report*, Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN): New York, NY, USA. URL https://www.sdgindex.org/reports/sustainable-development-report-2019/.

Saisana M, Saltelli A, Tarantola S (2005). "Uncertainty and Sensitivity Analysis Techniques as Tools for the Quality Assessment of Composite Indicators." *Journal of the Royal Statistical Society*, **168**(2), 307–323. doi:10.1111/j.1467-985X.2005.00350.x.

Saltelli A, Ratto M, Andres R, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008). *Global Sensitivity Analysis, A Primer*. John Wiley and Sohns, Ltd. doi:10.1002/9780470725184.

Sarvajayakesavalu S (2015). "Addressing Challenges of Developing Countries in Implementing Five Priorities for Sustainable Development Goals." *Ecosystem Health and Sustainability*, **1**(7), 1–4. doi:10.1890/EHS15-0028.1.

Statistics Canada (2019). "Statistics Canada Quality Guidelines." *Technical report*, Statistics Canada. URL https://www150.statcan.gc.ca/n1/en/catalogue/12-539-X.

Statistics Norway (2007). "Statistics Norway's Dissemination Policy." *Technical report*, Statistisk sentralbyrå. URL https://www.ssb.no/en/befolkning/artikler-og-publikasjoner/statistics-norways-dissemination-policy.

Statistics South Africa (2010). "South Africa Statistical Quality Assessment Framework (SASQFA) - Operational Standards and Guidelines." *Technical report*, Statistics South Africa. URL http://www.statssa.gov.za/standardisation/SASQAF_OpsGuidelines_Edition_1.pdf.

Sustainable Development Solutions Framework (2015). "Indicators and a Monitoring Framework for the Sustainable Development Goals: Launching a Data Revolution. Report to the Secretary-General of the United Nations by the Leadership Council of the Sustainable Development Solutions Network." *Technical report*, Paris and New York, N.Y. URL https://sustainabledevelopment.un.org/content/documents/2013150612-FINAL-SDSN-Indicator-Report1.pdf.

Thomas JC, Silvestre E, Salentine S, Reynolds H, Smith J (2016). "What Systems are Essential to Achieving the Sustainable Development Goals and What Will It Take to Marshal Them?" *Health Policy and Planning*, **31**(10), 1445–1447. doi:10.1093/heapol/czw070.

United Nations (2015). "United Nations Fundamental Principles of Official Statistics: Implementation Guidelines." *Technical report*, United Nations. URL https://unstats.un.org/unsd/dnss/gp/Implementation_Guidelines_FINAL_without_edit.pdf.

van Buuren S, Groothuis-Oudshoorn K (2011). "mice: Multivariate Imputation by Chained Equations in R." *Journal of Statistical Software*, **45**(3), 1–67. doi:10.18637/jss.v045.i03. URL https://www.jstatsoft.org/v45/i03/.

Van Der Sluijs JP, Craye M, Funtowicz S, Kloprogge P, Ravetz J, Risbey J (2005). "Combining Quantitative and Qualitative Measures of Uncertainty in Model-based Environmental Assessment: The NUSAP System." *Risk Analysis: An International Journal*, **25**(2), 481–492. doi:10.1111/j.1539-6924.2005.00604.x.

Wu J, Guo S, Huang H, Liu W, Xiang Y (2018). "Information and Communications Technologies for Sustainable Development Goals: State-of-the-art, Needs and Perspectives." *IEEE Communications Surveys & Tutorials*, **20**(3), 2389–2406. doi:10.1109/COMST.2018.2812301.

Zhang P (2003). "Multiple Imputation: Theory and Method." *International Statistical Institute (ISI)*, **71**(3), 581–592. doi:10.2307/1403830.

Affiliation:

Laura Güdemann and Ralf Münnich
Department of Social and Economic Statistics
Trier University
Germany, 54296 Trier, Universitätsring 15
E-mail: muennich@uni-trier.de

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 52 *Submitted:* 2022-05-05
2023 *Accepted:* 2022-10-31