European Commission

JRC TECHNICAL REPORT

# JRC Statistical Audit of the 2020 Gender Equality Index

Papadimitriou, Eleni

Norlén, Hedvig

Del Sorbo, Maria

2020

Joint Research Centre

EUR 30423 EN

Contact information
European Commission
Joint Research Centre

Directorate for Competences
Monitoring, Indicators and Impact Evaluation Unit
Competence Centre on Composite Indicators and Scoreboards
E-mail: jrc-coin@ec.europa.eu

Competence Centre on Composite Indicators and Scoreboards
https://ec.europa.eu/knowledge4policy/composite-indicators

EU Science Hub
https://ec.europa.eu/jrc

JRC122232
EUR 30423 EN

PDF ISBN 978-92-76-24507-0 ISSN 1831-9424 doi:10.2760/39645

Luxembourg: Publications Office of the European Union, 2020



How to cite this report: Papadimitriou, E.; Norlén, H.; Del Sorbo, M. *JRC Statistical Audit of the 2020 Gender Equality Index*, EUR 30423 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-24507-0, doi:10.2760/39645, JRC122232.

## Contents

**Abstract**

The 2020 Gender Equality Index ranks the 27 EU member states and the UK on 31 performance indicators measuring gender equalities in the six domains of work, money, knowledge, time, power, health and additionaly the domain of violence, as well as intersecting inequalities. By providing relevant statistics, data and measures, all essential components for evidence-based policymaking and successful gender mainstreaming, it supports the assessment of policy outcomes on women and men. Since 2013 the Gender Equality Index is released biannually by the European Institute for Gender Equality (EIGE) while from 2019 and on, the releases will be on annual basis. The European Commission's Competence Centre on Composite Indicators and Scoreboards (COIN) at the Joint Research Centre (JRC) was invited by the European Institute for Gender Equality to audit the 2020 edition of the index. The statistical audit presented herein aims to contribute to ensuring the transparency of the methodology and the reliability of the results. The report touches upon data quality issues, the conceptual and statistical coherence of the framework and the impact of modelling assumptions on the results. The analysis suggests that meaningful inferences can be drawn from the Gender Equality Index. It confirms that the 2020 Index meets the quality standards for statistical soundness and acknowledges it as a reliable composite indicator to measure gender equality in the European Union.

# 1 Introduction

Equality between women and men is and has always been at the core values of the European Union (EU). It goes back to the beginning of European Communities, in 1957, when the principle of equal pay for equal work became part of the Treaty of Rome (Article 157) . Over the last 63 years, European legislation and changes to the Treaties have reinforced this core value and its implementation in the EU; "In all its activities, the Union shall aim to eliminate inequalities and to promote equality, between men and women" (Article 8).

The Gender Equality Index, developed by the European Institute for Gender Equality (EIGE), assesses the levels of gender equality across the Member States of the EU based on the EU policy framework. It is the benchmark index of gender equality in the EU.

The statistical assessment of the fifth edition of the index (version 2020) was performed by the European Commission's Competence Centre on Composite Indicators and Scoreboards at the Joint Research Centre (JRC) and was conducted upon invitation of the index developers. The first version of the index, published in 2013 (European Institute for Gender Equality, 2013), and the following four editions were published biannually, the last being in 2019 (European Institute for Gender Equality, *Gender Equality Index 2019 Work-life balance*, 2019). From 2019, the index will be updated annually. The JRC pre-audited an earlier version of the index in 2017. The present statistical audit is assessing the last available data where most data points are deriving from 2018.

The index is hierarchically structured and is composed of six core domains (d) (work, money, knowledge, time, power and health) and each core domain is sub-divided into two or three sub-domains (subd), in total there are 14 sub-domains. Further, each sub-domain is divided into one to three indicators. In total, there are 31 indicators across the sub-domains in the last version of the index. There are also two satellite domains, violence and intersecting inequalities, likewise important domains of gender equality but due to data unavailability, they are left apart from the index calculation.

The Gender Equality Index provides a comprehensive framework of gender equality, in line with the EU's framework on gender equality for both women and men. No distinction is made as to the direction of the gender gap, meaning that the gender approach takes into account the situation of women and men in various domains of economic and social life, including those where men are in disadvantaged situations. The target is the equality point, and a given Member State is equally treated whether a gap is to the advantage of women, or men. The JRC audit of the Gender Equality Index 2020 focuses on two main issues: the statistical coherence of the hierarchical structure of indicators and the impact of key modelling assumptions on the index ranking. The JRC analysis complements the reported country rankings for the index with confidence intervals in order to better appreciate the robustness of these ranks to the calculation methodology (in particular construction of the metric with or withought the correctin coeficient, weights and aggregation formula at the domain level).

## 2 Conceptual framework, data analysis and rationale for the choices supporting the Gender Equality Index construction

The Gender Equality index is based on a hierarchical structure of six domains and 14 sub-domains (Table 1). The six domains —work, money, knowledge, time, power and health — constitute the core index.

The index is aggregated at sub-domain level with an arithmetic mean of the metrics of the indicators, while at domain level a simple geometric mean is used and at index level a weighted geometric mean, with weights derived from expert opinions (analytic hierarchy process AHP).

Table 1. Conceptual framework of the Gender Equality Index.

| Domain | Sub-domain | Indicator | Code |
|---|---|---|---|
| Work | Participation | Full-time equivalent employment rate | Fte |
| | | Duration of working life | Dwl |
| | Segregation and quality of work | Employed people in Education, Human Health and Social Work activities | Seg_W |
| | | Ability to take an hour or two off during working hours to take care of personal or family matters | Flexibility |
| | | Career Prospects Index | Prospects |
| Money | Financial resources | Mean monthly earnings | Earnings |
| | | Mean equivalised net income | Income |
| | Economic situation | Not-at-risk-of-poverty, ≥60% of median income | Poverty |
| | | S20/S80 income quintile share | S20/80 |
| Knowledge | Attainment and participation | Graduates of tertiary education | Grad |
| | | People participating in formal or non-formal education and training | Part |
| | Segregation | Tertiary students in the fields of Education, Health and Welfare, Humanities and Art (tertiary students) | Seg_E |
| Time | Care activities | People caring for and educating their children or grandchildren, elderly or people with disabilities, every day | Care |
| | | People doing cooking and/or housework, every day | Cooking |
| | Social activities | Workers doing sporting, cultural or leisure activities outside of their home, at least daily or several times a week | Leisure |

| Domain | Sub-domain | Indicator | Code |
|---|---|---|---|
| | | Workers involved in voluntary or charitable activities, at least once a month | Voluntary |
| Power | Political | Share of ministers | Min |
| | | Share of members of parliament | Parl |
| | | Share of members of regional assemblies | Reg |
| | Economic | Share of members of boards in largest quoted companies, supervisory board or board of directors | Boards |
| | | Share of board members of central bank | Banks |
| | Social | Share of board members of research funding organisations | Res |
| | | Share of board members in publically owned broadcasting organisations | Media |
| | | Share of members of highest decision making body of the national Olympic sport organisations | Sport |
| Health | Status | Self-perceived health, good or very good | SelfPerc |
| | | Life expectancy in absolute value at birth | Life ex |
| | | Healthy life years in absolute value at birth | HLY |
| | Behaviour | People who don't smoke and are not involved in harmful drinking | Risk |
| | | People doing physical activities and/or consuming fruits and vegetables | Behav |
| | Access | Population without unmet needs for medical examination | Medical |
| | | People without unmet needs for dental examination | Dental |

The data sources of the index include several European institutions and agencies, such as the Eurostat, Eurofound, and the European Institute of Gender Equality (EIGE). The source of the data is a key element when building composite indicators; the quality and adequacy of the index lies not only on its development, but also on gathering reliable data.

The dataset already included –very few – imputations for missing data made by the developers; these concern the values of BE, NL and FR for two indicators only (People who don't smoke and are not involved in harmful drinking and People doing physical activities and/or consuming fruits and vegetables) that were estimated with the Expectation–

maximization[1] (EM) algorithm (European Institute of Gender Equality, 2017, p11). In other cases, when data were not available, the last available data were used from the available time series. Then, for some variables, the EU-28 average was not available. In those cases the un-weighted average of the 28 values was used instead.

For this edition, the developers still consider UK in the analysis, and use EU+UK as aggregate for comparison, along with the EU2020 average.

With regard to the direction, adjustments have been made, so that all variables move towards the same direction which means that they have a positive sign, i.e. higher value would indicate the 'desirable situation'. For example, variables measuring 'healthy life years' have a positive direction, as it is desirable to live a long healthy life. Quite the opposite, the variable measuring 'being at risk of poverty' entails a negative sign or interpretation, which means that for the Index the indicator was changed to 'not being at risk of poverty'.

Each of the 31 indicators present in the framework is converted into a metric that combines the women and men figures in a single measure. The metric measures gender gaps by taking into account the relative position of women and men. An important element of it, is the correcting coefficient that makes possible to take into account the country context by comparing the levels achieved in all Member States for each indicator. In this way, an indicator with a good score is the reflection of both low gender gaps and high levels of achievement. The purpose of the correcting coefficient is to compare the performance of each country with the best performer in the EU-28. This actually normalises the indicators allowing for comparability across them.

The data analysis presented herein uses the 'normalised' by the metric dataset (1-100 scale) for the 27 European countries, and the UK, included in the Gender Equality Index.

Table 2 shows summary statistics for the 31 indicators included in the Gender Equality Index using the 'normalised' indicator data (final metric) and highlights the cases in which specific issues were found, in terms of data coverage and presence of outliers.

Table 2. Summary statistics of the metric scores included in the Gender Equality Index (2018 data)

| Code | Countries | % Missing data | Mean | Min. value | Max. value | Range | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| Fte | 30 | 0 | 78.7 | 62.1 | 93.8 | 31.8 | -0.4 | 0.4 |
| Dwl | 30 | 0 | 87.9 | 75.1 | 97.8 | 22.7 | -0.2 | -0.7 |
| Seg_W | 30 | 0 | 33.3 | 20.8 | 47.0 | 26.1 | -0.1 | -1.3 |
| Flexibility | 30 | 0 | 63.6 | 42.6 | 92.6 | 50.0 | 0.5 | -0.5 |
| Prospects | 30 | 0 | 92.5 | 83.4 | 98.3 | 14.9 | -0.9 | 0.1 |

[1] The Expectation-Maximization (EM) algorithm (Little & Rubin, 2002) is an iterative procedure that finds the maximum likelihood estimates of the parameter vector by repeating two steps. Step 1: The expectation E-step: Given a set of parameter estimates, such as a mean vector and covariance matrix for a multivariate normal distribution, the E-step calculates the conditional expectation of the complete-data log likelihood given the observed data and the parameter estimates. Step 2: The maximization M-step: Given a complete-data log likelihood, the M-step finds the parameter estimates to maximize the complete-data log likelihood from the E-step. The two steps are iterated until the iterations converge.

| Code | Countries | % Missing data | Mean | Min. value | Max. value | Range | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| Earnings | 30 | 0 | 69.8 | 47.4 | 96.0 | 48.7 | -0.1 | -1.0 |
| Income | 30 | 0 | 71.1 | 46.5 | 98.6 | 52.1 | -0.1 | -0.9 |
| Poverty | 30 | 0 | 94.1 | 87.0 | 98.4 | 11.4 | -0.8 | 0.0 |
| S20/80 | 30 | 0 | 82.8 | 65.3 | 99.3 | 34.0 | 0.0 | -0.9 |
| Grad | 30 | 0 | 77.7 | 58.4 | 95.0 | 36.6 | -0.1 | -1.0 |
| Part | 30 | 0 | 62.7 | 45.7 | 80.9 | 35.3 | 0.3 | -0.7 |
| Seg_E | 30 | 0 | 53.9 | 39.7 | 69.0 | 29.3 | 0.3 | -1.0 |
| Care | 30 | 0 | 81.6 | 69.6 | 97.6 | 28.0 | 0.3 | -0.9 |
| Cooking | 30 | 0 | 58.3 | 31.0 | 86.7 | 55.7 | -0.1 | -1.2 |
| Leisure | 30 | 0 | 62.6 | 31.6 | 98.0 | 66.4 | 0.0 | -1.2 |
| Voluntary | 30 | 0 | 51.6 | 25.5 | 85.8 | 60.3 | 0.3 | -0.1 |
| Min | 30 | 0 | 55.5 | 9.7 | 96.7 | 87.0 | 0.1 | -0.8 |
| Parl | 30 | 0 | 54.3 | 21.2 | 92.6 | 71.4 | 0.2 | -0.8 |
| Reg | 30 | 0 | 55.8 | 22.6 | 95.4 | 72.8 | 0.5 | -0.7 |
| Boards | 30 | 0 | 44.0 | 14.7 | 83.0 | 68.3 | 0.1 | -1.1 |
| Banks | 30 | 0 | 39.0 | 1.0 | 89.8 | 88.8 | 0.4 | -0.5 |
| Res | 30 | 0 | 64.6 | 1.0 | 96.5 | 95.5 | -0.9 | 0.2 |
| Media | 30 | 0 | 67.1 | 1.0 | 96.5 | 95.5 | -0.9 | 0.9 |
| Sport | 30 | 0 | 29.8 | 7.1 | 90.1 | 83.1 | 1.4 | 2.5 |
| SelfPerc | 30 | 0 | 86.3 | 66.9 | 99.7 | 32.8 | -1.0 | 0.5 |
| Life ex | 30 | 0 | 94.9 | 89.2 | 97.8 | 8.6 | -0.8 | -0.8 |
| HLY | 30 | 0 | 90.3 | 82.2 | 98.1 | 16.0 | 0.1 | -0.8 |
| Risk | 30 | 0 | 77.1 | 57.5 | 89.6 | 32.2 | -0.6 | -0.1 |
| Behav | 30 | 0 | 70.7 | 27.5 | 93.4 | 65.9 | -0.7 | 0.5 |
| Medical | 30 | 0 | 97.5 | 88.4 | 99.9 | 11.5 | -1.8 | 3.9 |

| Code | Countries | % Missing data | Mean | Min. value | Max. value | Range | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|---|
| Dental | 30 | 0 | 97.6 | 89.9 | 100.0 | 10.0 | -2.1 | 4.6 |

Notes: Data refer to year 2018, N=30 (27 EU countries, the UK, EU 2020 average, EU + UK average). Values in red colour indicate absolute skewness greater than 2.0 and kurtosis greater than 3.5.
Source: European Commission, Joint Research Centre, 2020.

Data coverage is excellent since the very few missing values were imputed beforehand, as already discussed.

Potentially problematic indicators that could bias the overall index results are usually identified on the basis of two measures related to the shape of the distributions: the skewness and kurtosis. A practical rule is that an indicator should be considered for treatment if it has an absolute skewness greater than 2.0 and kurtosis greater than 3.5 (Groeneveld & Meeden, 1984).

Overall, indicator scores do not seem to have outliers or skewed distributions, except for the indicator 'People without unmet needs for dental examination' (Dental) which is at the limits of this threshold without that consisting a problem. **Our suggestion would be to monitor its distribution in the future editions and intervene in case the distribution becomes more skewed.**

The developer's choices on the framework, the use of the specific metric, weights and aggregation methods are very well justified. Nevertheless in the section 4 of this report we challenge those choices, prerforming a robustness analysis to show how stable in these choices the index is.

# 3 Statistical coherence of the Gender Equality Index framework

The reliability of the Gender Equality Index depends - among other things – on the degree of coherence between the conceptual framework and the statistical structure of the data. The more the conceptual framework supports the statistical structure, the higher the reliability of the index will be. The coherence of the Gender Equality Index framework was assessed using two tests: (a) firstly, the analysis of the extent to which the indicators can explain a sufficient amount of variation in the aggregated scores (be those sub-domains, domains or the overall index). This first test entails correlation, cross-correlation and principal component analysis. Secondly, (b) the analysis of the impact on the country ranks with no least influential indicators (as identified in the first test) in the Gender Equality Index framework. The validity of the framework relies on the combination of both statistical and conceptual soundness.

Given that the present statistical analysis of the Gender Equality Index is in part, though not exclusively, based on correlations, the correspondence of the index to a real-world phenomenon needs to be critically addressed by experts in the field because 'correlations need not necessarily represent the real influence of the individual indicators on the phenomenon being measured'. (OECD and JRC, 2008). In a nutshell, the argument is that the validity of the framework relies on the combination of both statistical and conceptual soundness.

## 3.1 First statistical coherence test for the Gender Equality Index framework

The first coherence test consists of correlation and cross-correlation analyses, which was used to assess to what extent the data collected support the index's conceptual framework. There is no redundancy of information in the Gender Equality Index framework given the lack of highly collinear (i.e. Pearson correlation coefficients greater than 0.92) pairs of indicators within the same sub-domain. Indicators within most sub-domains exhibit modest to strong correlations between each other, see Table 3.

Table 3. Correlation matrix of the 31 indicators.

| Domain | Indicator | Fte | Dwl | Seg | Flexibility | Prospects | Earnings | Income | Poverty | S20/80 | Grad | Part | Seg | Care | Cooking | Leasure | Voluntary | Min | Parl | RegParl | Boards | Banks | Funding | Media | Sport | SelfPerc | Life ex | HLY | Risk | Behav | Medical | Dental |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Work | Fte | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Dwl | 0.70 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Seg | -0.15 | -0.15 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Flexibility | -0.13 | 0.22 | 0.50 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Prospects | 0.39 | 0.35 | -0.28 | -0.16 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Money | Earnings | -0.16 | 0.07 | 0.85 | 0.48 | 0.28 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | | |
| | Income | 0.01 | 0.26 | 0.81 | 0.48 | 0.32 | 0.91 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | | |
| | Poverty | -0.11 | -0.09 | 0.35 | -0.05 | 0.09 | 0.44 | 0.37 | 1.00 | | | | | | | | | | | | | | | | | | | | | | | |
| | S20/80 | 0.11 | 0.04 | 0.14 | 0.01 | 0.14 | 0.29 | 0.28 | 0.81 | 1.00 | | | | | | | | | | | | | | | | | | | | | | |
| Knowledge | Grad | 0.12 | 0.32 | 0.64 | 0.45 | 0.26 | 0.63 | 0.67 | 0.19 | 0.15 | 1.00 | | | | | | | | | | | | | | | | | | | | | |
| | Part | 0.25 | 0.56 | 0.64 | 0.38 | 0.36 | 0.65 | 0.76 | 0.39 | 0.33 | 0.63 | 1.00 | | | | | | | | | | | | | | | | | | | | |
| | Seg | 0.20 | -0.24 | 0.54 | 0.39 | 0.35 | 0.50 | 0.42 | 0.30 | 0.20 | 0.20 | 0.26 | 1.00 | | | | | | | | | | | | | | | | | | | |
| Time | Care | 0.45 | 0.55 | 0.18 | 0.18 | 0.20 | 0.23 | 0.29 | -0.20 | -0.10 | 0.11 | 0.47 | -0.05 | 1.00 | | | | | | | | | | | | | | | | | | |
| | Cooking | 0.37 | 0.59 | 0.36 | 0.47 | 0.54 | 0.38 | 0.46 | -0.13 | -0.04 | 0.47 | 0.60 | 0.19 | 0.47 | 1.00 | | | | | | | | | | | | | | | | | |
| | Leasure | 0.08 | 0.43 | 0.60 | 0.60 | 0.42 | 0.69 | 0.73 | 0.24 | 0.29 | 0.60 | 0.81 | 0.42 | 0.46 | 0.66 | 1.00 | | | | | | | | | | | | | | | | |
| | Voluntary | 0.16 | 0.43 | 0.46 | 0.43 | 0.33 | 0.56 | 0.55 | 0.37 | 0.51 | 0.33 | 0.65 | 0.22 | 0.41 | 0.50 | 0.74 | 1.00 | | | | | | | | | | | | | | | |
| Power | Min | 0.17 | 0.55 | 0.35 | 0.32 | 0.24 | 0.35 | 0.41 | 0.18 | 0.05 | 0.21 | 0.51 | 0.16 | 0.27 | 0.38 | 0.56 | 0.43 | 1.00 | | | | | | | | | | | | | | |
| | Parl | 0.07 | 0.45 | 0.53 | 0.40 | 0.21 | 0.59 | 0.60 | 0.25 | 0.15 | 0.33 | 0.66 | 0.43 | 0.32 | 0.46 | 0.66 | 0.44 | 0.78 | 1.00 | | | | | | | | | | | | | |
| | RegParl | 0.16 | 0.51 | 0.58 | 0.45 | 0.32 | 0.54 | 0.55 | 0.18 | 0.16 | 0.52 | 0.64 | 0.31 | 0.23 | 0.53 | 0.72 | 0.45 | 0.79 | 0.84 | 1.00 | | | | | | | | | | | | |
| | Boards | 0.04 | 0.30 | 0.42 | 0.27 | 0.18 | 0.52 | 0.50 | 0.38 | 0.29 | 0.18 | 0.45 | 0.37 | 0.22 | 0.33 | 0.55 | 0.55 | 0.68 | 0.74 | 0.69 | 1.00 | | | | | | | | | | | |
| | Banks | 0.14 | 0.25 | 0.31 | 0.22 | 0.04 | 0.24 | 0.25 | -0.08 | -0.31 | 0.22 | 0.28 | 0.20 | 0.39 | 0.33 | 0.37 | 0.07 | 0.60 | 0.49 | 0.57 | 0.42 | 1.00 | | | | | | | | | | |
| | Funding | 0.15 | 0.18 | 0.38 | 0.45 | 0.32 | 0.52 | 0.48 | -0.03 | -0.10 | 0.31 | 0.28 | 0.32 | 0.14 | 0.42 | 0.44 | 0.20 | 0.57 | 0.54 | 0.57 | 0.40 | 0.52 | 1.00 | | | | | | | | | |
| | Media | 0.33 | 0.45 | 0.17 | 0.39 | 0.37 | 0.20 | 0.21 | -0.26 | -0.17 | 0.24 | 0.28 | -0.04 | 0.56 | 0.38 | 0.34 | 0.35 | 0.45 | 0.29 | 0.38 | 0.27 | 0.44 | 0.43 | 1.00 | | | | | | | | |
| | Sport | 0.28 | 0.53 | 0.51 | 0.40 | 0.36 | 0.42 | 0.45 | 0.08 | -0.06 | 0.47 | 0.61 | 0.34 | 0.38 | 0.58 | 0.58 | 0.52 | 0.68 | 0.66 | 0.70 | 0.62 | 0.62 | 0.45 | 0.49 | 1.00 | | | | | | | |
| Health | SelfPerc | -0.40 | -0.29 | 0.58 | 0.36 | 0.05 | 0.66 | 0.55 | 0.50 | 0.33 | 0.51 | 0.27 | 0.51 | -0.19 | 0.13 | 0.41 | 0.40 | 0.14 | 0.28 | 0.25 | 0.26 | 0.17 | 0.27 | 0.02 | 0.30 | 1.00 | | | | | | |
| | Life ex | 0.29 | -0.02 | 0.80 | 0.49 | -0.06 | 0.84 | 0.79 | 0.52 | 0.38 | 0.54 | 0.57 | 0.49 | 0.09 | 0.18 | 0.58 | 0.53 | 0.32 | 0.53 | 0.44 | 0.41 | 0.24 | 0.35 | 0.04 | 0.34 | 0.77 | 1.00 | | | | | |
| | HLY | -0.32 | -0.43 | 0.50 | 0.20 | -0.11 | 0.40 | 0.28 | 0.18 | 0.02 | 0.17 | 0.00 | 0.60 | -0.19 | -0.06 | 0.14 | 0.07 | 0.02 | 0.17 | 0.16 | 0.11 | 0.35 | 0.27 | -0.17 | 0.25 | 0.65 | 0.55 | 1.00 | | | | |
| | Risk | -0.36 | -0.23 | 0.33 | 0.25 | -0.39 | 0.27 | 0.21 | 0.42 | 0.33 | 0.06 | 0.14 | 0.39 | -0.06 | -0.34 | 0.17 | 0.19 | 0.08 | 0.21 | 0.16 | 0.33 | 0.11 | -0.17 | -0.25 | 0.13 | 0.31 | 0.54 | 0.43 | 1.00 | | | |
| | Behav | 0.26 | 0.52 | 0.65 | 0.39 | 0.42 | 0.67 | 0.78 | 0.36 | 0.39 | 0.58 | 0.76 | 0.27 | 0.40 | 0.54 | 0.69 | 0.62 | 0.28 | 0.46 | 0.46 | 0.40 | 0.12 | 0.19 | 0.18 | 0.43 | 0.31 | 0.59 | 0.08 | -0.01 | 1.00 | | |
| | Medical | 0.18 | -0.22 | 0.36 | 0.34 | -0.06 | 0.42 | 0.41 | 0.38 | 0.23 | 0.16 | 0.09 | 0.33 | -0.16 | -0.25 | 0.10 | 0.06 | 0.18 | 0.17 | 0.12 | 0.24 | 0.07 | 0.43 | -0.03 | 0.08 | 0.35 | 0.43 | 0.45 | 0.41 | 0.12 | 1.00 | |
| | Dental | 0.18 | -0.22 | 0.29 | 0.16 | 0.31 | 0.42 | 0.44 | 0.38 | 0.31 | 0.30 | 0.20 | 0.42 | -0.23 | -0.08 | 0.33 | 0.31 | 0.11 | 0.11 | 0.12 | 0.10 | -0.10 | 0.10 | -0.11 | 0.10 | 0.56 | 0.37 | 0.38 | 0.24 | 0.23 | 0.48 | 1.00 |

*Notes: Numbers represent the Pearson correlations coefficients between the index indicators. Correlations that are not significant at the significance level of a = 0.10 are left light grey. Negative significant correlations are in red. Blue boxes show the conceptual grouping of the indicators into subdomains and black contour boxes the grouping of indicators into domains. Source: European Commission, Joint Research Centre, 2020.*

Table 4 illustrates the index correlation structure within and across the 14 sub-domains and it confirms the expectation that most indicators, apart from two, are more associated with their own sub-domain than to any of the other sub-domains. The two conflicting indicators are both in sub-domain Quality of work and Segregation (subd2): indicator Segregation is more related to sub-domain Financial resources (subd3) and indicator Prospects is marginally higher correlated to sub-domain Care activities (subd7).

At the next level, the majority of the indicators are more associated with their pertaining domain but seven indicators are in fact correlating better with other domains (in five cases only marginally). Indicator Segregation (Seg_W) is more related to the domains Money (d2), Knowledge (d3) and Health (d6), than to its pertaining domain Work (d1). Indicator Participation (Part) is more related to the domains Money (d2) and Time (d4) than to its domain Knowledge (d3). Five out of the seven indicators in domain Health (d6) are marginally more related to either the Money (d2), or Knowledge (d3) domains and the related indicators are Self-perceived health (SelfPerc), Life expectancy (Life_ex), Healthy life years (HLY), Population without unmet medical examination needs (Medical) and Population without unmet dental examination needs (Dental). For most indicators these results suggest that the allocation of the indicators to the specific sub-domain, and allocation of sub-domains to domains, is consistent both from conceptual and statistical perspectives.

The majority of the indicators (24 out of 31) also are positively and significantly correlated with the overall index. However, seven indicators are found to be not sufficiently related to the overall index: Full-time equivalent employment (Fte) in the Work domain (d1), Not at-risk-of-poverty (Poverty) and Income distribution S20/S80 in the Money domain (d2), Healthy life years (HLY), Risk (Risk), Population without unmet medical examination needs (Medical) and Population without unmet dental examination needs (Dental) in the Health domain (d6). These results are in line with the earlier editions of the index. However, given that, these seven indicators are influential at the first and second aggregation levels (sub-domains and domains), their inclusion in the framework is partially supported by the analysis. Nevertheless, as mentioned above, it is noted that the three indicators of the Health domain (d6) were found to be correlating better with other domains.

Table 4. Correlations between indicators and other Gender Equality Index components.

| Domain | Subdomain | Indicator | subd1 | subd2 | subd3 | subd4 | subd5 | subd6 | subd7 | subd8 | subd9 | subd10 | subd11 | subd12 | subd13 | subd14 | d1 | d2 | d3 | d4 | d5 | d6 | Index |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Work | Participation (subd1) | Fte | **0.94** | -0.06 | -0.09 | 0.06 | 0.20 | -0.20 | 0.45 | 0.12 | 0.15 | 0.07 | 0.31 | -0.40 | 0.08 | -0.21 | 0.46 | -0.06 | -0.01 | 0.25 | 0.18 | -0.07 | 0.17 |
| | | Dwl | 0.91 | 0.28 | 0.17 | 0.01 | 0.48 | -0.24 | 0.66 | 0.46 | 0.54 | 0.32 | 0.47 | -0.31 | 0.35 | -0.25 | 0.72 | 0.16 | 0.12 | 0.57 | 0.48 | 0.16 | 0.49 |
| | Quality of work & Segr. (subd2) | Seg | -0.02 | 0.79 | 0.85 | 0.19 | 0.71 | 0.54 | 0.34 | 0.58 | 0.51 | 0.43 | 0.43 | 0.67 | 0.67 | 0.38 | 0.61 | 0.78 | 0.78 | 0.55 | 0.51 | 0.75 | 0.72 |
| | | Flexibility | 0.03 | 0.90 | 0.50 | 0.00 | 0.47 | 0.39 | 0.43 | 0.57 | 0.42 | 0.28 | 0.52 | 0.38 | 0.42 | 0.29 | 0.72 | 0.43 | 0.54 | 0.56 | 0.43 | 0.46 | 0.61 |
| | | Prospects | 0.40 | 0.44 | 0.31 | 0.13 | 0.34 | 0.35 | 0.49 | 0.41 | 0.28 | 0.12 | 0.44 | -0.02 | 0.21 | 0.13 | 0.57 | 0.29 | 0.44 | 0.48 | 0.28 | 0.15 | 0.44 |
| Money | Financial resources (subd3) | Earnings | -0.06 | 0.72 | **0.98** | 0.33 | 0.71 | 0.50 | 0.38 | 0.68 | 0.52 | 0.43 | 0.48 | 0.69 | 0.67 | 0.48 | 0.54 | **0.93** | 0.75 | 0.63 | 0.52 | 0.77 | 0.77 |
| | | Income | 0.12 | 0.71 | **0.98** | 0.31 | 0.79 | 0.42 | 0.46 | 0.70 | 0.55 | 0.42 | 0.47 | 0.58 | 0.74 | 0.49 | 0.63 | **0.92** | 0.74 | 0.68 | 0.52 | 0.80 | 0.78 |
| | Economic situation (subd4) | Poverty | -0.11 | 0.13 | 0.41 | 0.89 | 0.31 | 0.30 | -0.17 | 0.31 | 0.21 | 0.14 | -0.10 | 0.46 | 0.46 | 0.44 | 0.04 | 0.63 | 0.39 | 0.16 | 0.11 | 0.54 | 0.26 |
| | | S20/80 | 0.09 | 0.09 | 0.29 | **0.99** | 0.26 | 0.20 | -0.07 | 0.41 | 0.12 | -0.06 | -0.14 | 0.28 | 0.46 | 0.31 | 0.12 | 0.57 | 0.29 | 0.27 | -0.02 | 0.49 | 0.20 |
| Knowledge | Attainment & Partic. (subd5) | Grad | 0.23 | 0.61 | 0.67 | 0.17 | 0.91 | 0.20 | 0.40 | 0.52 | 0.37 | 0.24 | 0.41 | 0.47 | 0.51 | 0.26 | 0.61 | 0.62 | 0.67 | 0.53 | 0.37 | 0.56 | 0.60 |
| | | Part | 0.42 | 0.59 | 0.72 | 0.35 | 0.89 | 0.26 | 0.63 | 0.79 | 0.64 | 0.42 | 0.47 | 0.28 | 0.69 | 0.16 | 0.69 | 0.73 | 0.69 | 0.81 | 0.56 | 0.64 | 0.78 |
| | Segregation (subd6) | Seg | -0.23 | 0.55 | 0.47 | 0.23 | 0.25 | **1.00** | 0.13 | 0.36 | 0.31 | 0.32 | 0.25 | 0.59 | 0.38 | 0.43 | 0.30 | 0.47 | 0.83 | 0.31 | 0.33 | 0.49 | 0.50 |
| Time | Care activities (subd7) | Care | 0.54 | 0.23 | 0.27 | -0.13 | 0.31 | -0.05 | 0.73 | 0.47 | 0.29 | 0.38 | 0.44 | -0.16 | 0.32 | -0.22 | 0.48 | 0.18 | 0.15 | 0.60 | 0.42 | 0.19 | 0.45 |
| | | Cooking | 0.51 | 0.57 | 0.43 | -0.06 | 0.58 | 0.19 | **0.94** | 0.64 | 0.48 | 0.39 | 0.57 | 0.09 | 0.32 | -0.20 | 0.73 | 0.35 | 0.47 | 0.81 | 0.51 | 0.24 | 0.66 |
| | Social activities (subd8) | Leasure | 0.26 | 0.73 | 0.73 | 0.29 | 0.77 | 0.42 | 0.68 | **0.95** | 0.69 | 0.53 | 0.56 | 0.41 | 0.65 | 0.24 | 0.71 | 0.71 | 0.74 | **0.94** | 0.65 | 0.65 | 0.87 |
| | | Voluntary | 0.31 | 0.54 | 0.57 | 0.49 | 0.53 | 0.22 | 0.54 | 0.91 | 0.47 | 0.33 | 0.43 | 0.37 | 0.59 | 0.21 | 0.59 | 0.65 | 0.46 | 0.84 | 0.45 | 0.58 | 0.67 |
| Power | Political (subd9) | Min | 0.37 | 0.40 | 0.39 | 0.08 | 0.38 | 0.16 | 0.39 | 0.54 | **0.93** | 0.75 | 0.70 | 0.15 | 0.26 | 0.17 | 0.53 | 0.36 | 0.34 | 0.53 | 0.87 | 0.25 | 0.74 |
| | | Parl | 0.26 | 0.52 | 0.61 | 0.18 | 0.54 | 0.43 | 0.47 | 0.61 | **0.93** | 0.71 | 0.61 | 0.33 | 0.47 | 0.17 | 0.56 | 0.59 | 0.60 | 0.61 | 0.82 | 0.47 | 0.83 |
| | | RegParl | 0.35 | 0.60 | 0.56 | 0.17 | 0.64 | 0.31 | 0.49 | 0.65 | **0.94** | 0.73 | 0.68 | 0.29 | 0.45 | 0.14 | 0.67 | 0.54 | 0.59 | 0.65 | 0.86 | 0.44 | 0.85 |
| | Economic (subd10) | Boards | 0.12 | 0.38 | 0.52 | 0.32 | 0.34 | 0.37 | 0.33 | 0.59 | 0.75 | 0.79 | 0.52 | 0.27 | 0.46 | 0.21 | 0.38 | 0.56 | 0.45 | 0.54 | 0.77 | 0.46 | 0.73 |
| | | Banks | 0.20 | 0.28 | 0.25 | -0.26 | 0.27 | 0.20 | 0.40 | 0.27 | 0.60 | 0.89 | 0.65 | 0.26 | 0.14 | -0.01 | 0.34 | 0.14 | 0.30 | 0.34 | 0.80 | 0.17 | 0.59 |
| | Social activities (subd11) | Funding | 0.18 | 0.52 | 0.51 | -0.08 | 0.33 | 0.32 | 0.38 | 0.36 | 0.60 | 0.55 | 0.81 | 0.32 | 0.10 | 0.32 | 0.51 | 0.41 | 0.41 | 0.39 | 0.70 | 0.18 | 0.65 |
| | | Media | 0.41 | 0.40 | 0.21 | -0.20 | 0.29 | -0.04 | 0.50 | 0.37 | 0.41 | 0.43 | 0.81 | -0.03 | 0.06 | -0.08 | 0.55 | 0.12 | 0.14 | 0.45 | 0.59 | 0.01 | 0.50 |
| | | Sport | 0.43 | 0.54 | 0.44 | -0.03 | 0.59 | 0.34 | 0.59 | 0.59 | 0.73 | 0.73 | 0.77 | 0.33 | 0.41 | 0.11 | 0.67 | 0.37 | 0.58 | 0.64 | 0.82 | 0.40 | 0.79 |
| Health | Status (subd12) | SelfPerc | -0.38 | 0.48 | 0.62 | 0.38 | 0.44 | 0.51 | 0.03 | 0.43 | 0.23 | 0.24 | 0.24 | **0.96** | 0.38 | 0.52 | 0.17 | 0.65 | 0.60 | 0.32 | 0.27 | 0.60 | 0.45 |
| | | Life ex | -0.19 | 0.63 | 0.84 | 0.43 | 0.61 | 0.49 | 0.17 | 0.60 | 0.45 | 0.37 | 0.30 | 0.84 | 0.71 | 0.47 | 0.40 | 0.85 | 0.69 | 0.50 | 0.42 | 0.84 | 0.65 |
| | | HLY | -0.40 | 0.31 | 0.35 | 0.06 | 0.10 | 0.60 | -0.12 | 0.12 | 0.12 | 0.29 | 0.14 | 0.81 | 0.23 | 0.49 | 0.02 | 0.32 | 0.47 | 0.05 | 0.21 | 0.45 | 0.26 |
| | Behaviour (subd13) | Risk | -0.33 | 0.21 | 0.24 | 0.36 | 0.11 | 0.39 | -0.28 | 0.19 | 0.15 | 0.25 | -0.14 | 0.43 | 0.56 | 0.39 | -0.01 | 0.33 | 0.34 | 0.03 | 0.13 | 0.62 | 0.19 |
| | | Behav | 0.41 | 0.61 | 0.74 | 0.40 | 0.73 | 0.27 | 0.56 | 0.71 | 0.42 | 0.28 | 0.32 | 0.33 | **0.93** | 0.20 | 0.70 | 0.76 | 0.61 | 0.73 | 0.38 | 0.85 | 0.67 |
| | Access (subd14) | Medical | -0.21 | 0.35 | 0.42 | 0.27 | 0.14 | 0.33 | -0.25 | 0.09 | 0.17 | 0.17 | 0.22 | 0.44 | 0.26 | 0.88 | 0.16 | 0.44 | 0.31 | -0.03 | 0.20 | 0.42 | 0.25 |
| | | Dental | -0.22 | 0.28 | 0.44 | 0.34 | 0.28 | 0.42 | -0.15 | 0.35 | 0.12 | -0.02 | 0.03 | 0.53 | 0.29 | 0.84 | 0.10 | 0.47 | 0.45 | 0.20 | 0.04 | 0.46 | 0.24 |

*Notes: Numbers represent Pearson correlations coefficients.. Correlations that are not significant at the significance level of a = 0.10 are left light grey. Negative significant correlations are in red. Blue boxes show the conceptual grouping of the indicators into subdomains and black contour boxes the grouping of indicators into domains. Source: European Commission, Joint Research Centre, 2020.*

All sub-domains correlate moderately, or strongly with the respective domains (correlation coefficients from 0.51 to 0.96, Table 5). Nevertheless, sub-domain Access (subd14) is slightly more associated with the Money domain (d2) than to its Health domain (d6) which is in alignment with previous results showing that these indicators were higher correlated with the Money domain (d2). Further, the six domains strongly correlate and in a relatively balanced way with the index (correlation coefficients ranging between 0.66 and 0.89, Table 5). Further, it is noticed that the Power domain (d5) has the highest correlation with the index, and the Health domain (d6) has the lowest correlation and at the same time has the lowest experts' weight.

Table 5. Correlations between sub-domains, domains and index.

| GEI components | subd1 | subd2 | subd3 | subd4 | subd5 | subd6 | subd7 | subd8 | subd9 | subd10 | subd11 | subd12 | subd13 | subd14 | d1 | d2 | d3 | d4 | d5 | d6 | GEI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Participation | 1.00 | | | | | | | | | | | | | | 0.63 | 0.04 | 0.05 | 0.43 | 0.34 | 0.04 | 0.34 |
| Quality of work & Segr. | 0.10 | 1.00 | | | | | | | | | | | | | 0.84 | 0.66 | 0.76 | 0.69 | 0.55 | 0.64 | 0.79 |
| Financial resources | 0.03 | 0.74 | 1.00 | | | | | | | | | | | | 0.60 | **0.95** | 0.76 | 0.67 | 0.54 | 0.80 | 0.79 |
| Economic situation | 0.04 | 0.11 | 0.33 | 1.00 | | | | | | | | | | | 0.10 | 0.61 | 0.32 | 0.25 | 0.02 | 0.52 | 0.22 |
| Attainment & Partic. | 0.35 | 0.67 | 0.77 | 0.28 | 1.00 | | | | | | | | | | 0.72 | 0.74 | 0.75 | 0.73 | 0.50 | 0.67 | 0.76 |
| Segregation | -0.23 | 0.55 | 0.47 | 0.23 | 0.25 | 1.00 | | | | | | | | | 0.30 | 0.47 | 0.83 | 0.31 | 0.33 | 0.49 | 0.50 |
| Care activities | 0.59 | 0.52 | 0.43 | -0.10 | 0.57 | 0.13 | 1.00 | | | | | | | | 0.74 | 0.34 | 0.41 | 0.84 | 0.55 | 0.25 | 0.68 |
| Social activities | 0.30 | 0.69 | 0.71 | 0.40 | 0.72 | 0.36 | 0.66 | 1.00 | | | | | | | 0.71 | 0.74 | 0.66 | **0.96** | 0.61 | 0.67 | 0.85 |
| Political | 0.35 | 0.54 | 0.55 | 0.15 | 0.55 | 0.31 | 0.48 | 0.64 | 1.00 | | | | | | 0.63 | 0.52 | 0.54 | 0.64 | 0.91 | 0.41 | 0.86 |
| Economic | 0.20 | 0.38 | 0.44 | -0.01 | 0.36 | 0.32 | 0.44 | 0.48 | 0.78 | 1.00 | | | | | 0.42 | 0.38 | 0.43 | 0.51 | **0.93** | 0.35 | 0.77 |
| Social activities | 0.41 | 0.61 | 0.49 | -0.13 | 0.49 | 0.25 | 0.60 | 0.54 | 0.71 | 0.70 | 1.00 | | | | 0.71 | 0.37 | 0.45 | 0.60 | 0.87 | 0.23 | 0.80 |
| Status | -0.39 | 0.52 | 0.65 | 0.34 | 0.42 | 0.59 | 0.01 | 0.42 | 0.27 | 0.31 | 0.25 | 1.00 | | | 0.19 | 0.66 | 0.65 | 0.31 | 0.31 | 0.67 | 0.48 |
| Behaviour | 0.22 | 0.59 | 0.72 | 0.48 | 0.66 | 0.38 | 0.37 | 0.67 | 0.42 | 0.33 | 0.21 | 0.45 | 1.00 | | 0.59 | 0.77 | 0.65 | 0.63 | 0.37 | **0.96** | 0.64 |
| Access | -0.25 | 0.37 | 0.50 | 0.35 | 0.24 | 0.43 | -0.24 | 0.24 | 0.17 | 0.10 | 0.15 | 0.56 | 0.32 | 1.00 | 0.16 | 0.53 | 0.44 | 0.09 | 0.15 | 0.51 | 0.28 |

Notes: Numbers represent the Pearson correlations coefficients between the index components. Correlations that are not significant at the significance level of α = 0.10 are left light grey. Black contour boxes show the conceptual grouping of the indicators into sub-domains and domains. Very strong correlations (*i.e.* Pearson correlation coefficients greater than 0,92) are marked in bold. Negative significant correlations are in red.
Source: European Commission, Joint Research Centre, 2020.

Table 6. Correlations between domains and the index.

| | Work | Money | Knowledge | Time | Power | Health | GEI |
|---|---|---|---|---|---|---|---|
| Work | 1.00 | | | | | | 0.81 |
| Money | 0.55 | 1.00 | | | | | 0.75 |
| Knowledge | 0.63 | 0.75 | 1.00 | | | | 0.78 |
| Time | 0.78 | 0.66 | 0.64 | 1.00 | | | 0.86 |
| Power | 0.63 | 0.47 | 0.52 | 0.64 | 1.00 | | 0.89 |
| Health | 0.52 | 0.85 | 0.72 | 0.58 | 0.38 | 1.00 | 0.66 |

Notes: Numbers represent the Pearson correlations coefficients.
Source: European Commission, Joint Research Centre, 2020.

### 3.1.1 Principal component analysis

Next, the principal component analysis (PCA) was used to confirm whether there is a single statistical dimension in each index component, which would give the "statistical justification" for aggregating indicators into one number. Technically, the expectation here is that there is only one principal component with eigenvalue greater than 1.0 (or explained variance of around 70%).

Indeed, PCA results confirm the presence of a single latent dimension within each of the 6 domains that captures between 55% in the Work domain (d1) to 83% in the Time domain (d4) of the total variance in the underlying sub-domains. The results of the six PCAs performed in the sub-domains of the six domains are summarised in Table 7.

Table 7. Results of the PCA (first PC) within each Gender Equality Index domain

| First principal component (PC1) | Eigenvalue | Percentage of variance |
|---|---|---|
| Work | 1.1 | 0.55 |
| Money | 1.33 | 0.66 |
| Knowledge | 1.25 | 0.63 |
| Time | 1.66 | 0.83 |
| Power | 2.47 | 0.82 |
| Health | 1.89 | 0.63 |

Notes: Information is shown only for the first principal component (PC) of each of the analysis performed individually for each domain. The analysis was done at sub-domain level for each domain.
Source: European Commission, Joint Research Centre, 2020.

Finally, as can be seen in Table 8, the six domains share a single statistical dimension that summarizes 69% of the total variance. This latter result supports the aggregation of the six domains into one number that would represent the common dimension that they share.

Table 8. Results of the Principal component analysis at index level

| | Eigenvalue | Percentage of variance | Cumulative percentage of variance |
|---|---|---|---|
| PC1 | 4.1 | 68.5 | 68.5 |
| PC2 | 0.8 | 14.3 | 82.8 |
| PC3 | 0.4 | 6.6 | 89.4 |
| PC4 | 0.3 | 4.9 | 94.3 |
| PC5 | 0.2 | 3.5 | 97.8 |
| PC6 | 0.1 | 2.2 | 100 |

Notes: Information for all principal components (PC) of the analysis performed at domain level.
Source: European Commission, Joint Research Centre, 2020.

Concluding, the first statistical coherence test and the principal component analysis, corroborated the three-level structure of the Gender Equality Index framework and the unidimensionality of all index components (sub-domains, domains and index). Furthermore, all 31 indicators were found to be influential at the first and second aggregation level (sub-domains and domains). However, two indicators were slightly more associated with other sub-domains with respect to their own sub-domains and seven indicators (in five cases only marginally) were more associated with other domains. At the last aggregation level, 24 out of 31 indicators are significantly correlated with the overall index while seven indicators explain a not full sufficient amount of variation in the index scores. Nevertheless, the information content is kept at all levels of aggregation in the index framework for the majority of the underlying indicators.

## 3.2 Second statistical coherence test in the Gender Equality Index framework

A second coherence test aims at assessing whether the seven indicators found to be not significantly correlated with the overall index, are important in a different way in the overall index, for instance by influencing the overall index ranking.

The seven indicators that are found not to be sufficiently related to the overall index are: Full-time equivalent employment (Fte), Not at-risk-of-poverty (Poverty), Income distribution S20/S80, Healthy life years (HLY), Risk (Risk), Population without unmet medical examination needs (Medical) and Population without unmet dental examination needs (Dental). However, given that these seven indicators are influential at the first and second aggregation levels (sub-domains and domains), their inclusion in the Gender Equality Index framework is

supported to a certain extent by the analysis. The test consists of assessing how country ordering changes when these indicators are omitted one-at-a-time from the calculation.

Table 9. Number of countries with a rank change when a selected indicator is omitted.

| Indicator | Number of rank changes when omitted |
|---|---|
| Ftc (ind1) | 4 |
| Poverty (ind8) | 8 |
| S20/80 (ind9) | 4 |
| HLY (ind27) | 0 |
| Risk (ind28) | 8 |
| Medical (ind30) | 0 |
| Dental (ind31) | 0 |

Source: European Commission, Joint Research Centre, 2020.

Two indicators are found to be somewhat more influential: Not at-risk-of-poverty (Poverty) and Risk (Risk). When omitting either of the indicators Not at-risk-of-poverty or Risk, eight Member States either lose, or gain one position in ranking. When omitting the Full-time equivalent employment indicator (Fte), or Income distribution S20/S80 indicator, four Member States change ranks with one position. Omitting either one of the remaining indicators: Healthy life years (HLY), Population without unmet medical examination needs (Medical) and Population without unmet dental examination needs (Dental) has no impact on the index ranks, and as already discussed, these can explain only a small (negligible) amount of variation in the scores. The results are summarized in Table 9. Although these three indicators are conceptually enriching the framework and their statistical impact contribute to the first and second aggregation levels, **it is recommended to carefully monitor their performance in the coming releases of the index.**

### 3.3 JRC recommendations based on the statistical coherence tests

Generally, the results of the two statistical coherence tests (correlation analysis and impact of excluding from the Gender Equality Index framework the least influential indicators one-at-a time) suggest that the conceptual grouping of the indicators into 14 sub-domains and six domains is statistically confirmed, and that the index is in general influenced by most but not all of the underlying indicators.

According to the first coherence test, 24 out of the 31 indicators are positively and significantly correlated with the overall index. The remaining seven indicators that are found to be the least influential at the index level are: Full-time equivalent employment (Fte), Not at-risk-of-poverty (Poverty), Income distribution S20/S80, Healthy life years (HLY), Risk (Risk), Population without unmet medical examination needs (Medical) and Population without unmet dental examination needs (Dental). However, given that these seven indicators are influential at the first two aggregation levels (sub-domains and domains), their inclusion in the framework is to a certain extent corroborated by the analyses. Although conceptually enriching and statistically informative up to the second aggregation level, the information content of these indicators does not show sufficiently up at the index level. The second

coherence test offers an additional perspective on the impact of these seven indicators, showing that three of them have a low impact on the country ranks and can explain only a small (negligible) amount of variation in the scores. These three indicators follow Healthy life years (HLY), Population without unmet medical examination needs (Medical) and Population without unmet dental examination needs (Dental). Moreover, these three indicators were "puzzling" as they also are more associated with other domains with respect to their own domains. The JRC recommendation to the Gender Equality Index development team is to monitor carefully how these three indicators perform in the coming releases of the index, and eventually fine-tune the framework in this respect.

# 4 Impact of modelling assumptions on the Gender Equality Index results

A fundamental step of the statistical analysis of a composite indicator is to assess the effect of different modelling assumptions among reasonable alternatives. The development of a composite indicator, like any model, involves assumptions and subjective decisions. The Gender Equality Index is the outcome of a number of choices concerning, among other things, the theoretical framework, the indicators selected, the metric construction, the weights assigned and the aggregation method. Some of these choices may be based on expert opinion or other considerations, driven by statistical analysis or by the need for ease of communication or draw attention to specific issues.

This section aims to test the impact of varying some of these assumptions within a range of plausible alternatives in an uncertainty analysis. The objective is therefore to try to quantify the uncertainty in the ranks of the Gender Equality Index, which can demonstrate the extent to which countries can be differentiated by their scores.

As suggested in the relevant literature on composite indicators (M Saisana, Tarantola, & Saltelli, 2005; Michaela Saisana & Saltelli, 2011), the robustness assessment is based on Monte Carlo simulations and multi-modelling approaches. In particular, the key modelling issues considered in the assessment of the Gender Equality Index are the use of the correcting coefficient *a*, the aggregation formula and the weights at domain level, as outlined in Table 10. These were chosen as plausible alternative pathways in the construction of the index, which can be relatively easily investigated.

Metric construction. The metric used in the index measures gender gaps by taking into account the relative position of women and men. An important element of it, is the correcting coefficient that makes possible to take into account the country context by comparing the levels achieved in all Member States for each indicator. In this way, an indicator with a good score is the reflection of both low gender gaps and high levels of achievement. This construction has been an important methodological choice of the developers based on evidence and experts' opinion. It is important though to test its effect in the countries overall ranking and understand its contribution to a possible volatility of results.

Aggregation formula. Regarding the aggregation formula, the Gender Equality Index team opted for the geometric average of the sub-domains into the domains and then again the geometric average of the six domains. This is a rather non-compensatory approach, since the geometric mean tends to penalize the existence of a low value, particularly when the other values are not so low. This choice is in line with relevant literature that challenges the use of simple arithmetic averages at higher aggregation levels because of their fully compensatory nature, in which a comparative high advantage on a few indicators can compensate a comparative disadvantage on many indicators (Munda, 2008). To assess the impact of this choice, a comparison with the arithmetic mean is included in the analysis. The arithmetic mean allows outstanding performance in some aspects to balance for weaknesses in others and vice-versa. The comparison of the two aggregation approaches should be able to highlight the individual countries with unbalanced profiles.

Weights. The weighting system of the index at domain level is based on experts' opinion, using the analytic hierarchy process (AHP). The effect of randomly varying these weights by +/-25% is investigated: Monte Carlo simulation comprised 1,000 runs of different sets of weights for the six domains constituting the index. The weights are the result of a random extraction based on uniform continuous distributions centred in the reference value plus or minus 25% of this value.

Four models were tested comparing the different aggregation formulas and the presence or not of the correcting coefficient $\alpha$ in the metric, which resulted in a total of 4,000 runs of simulations, as can be seen in Table 10.

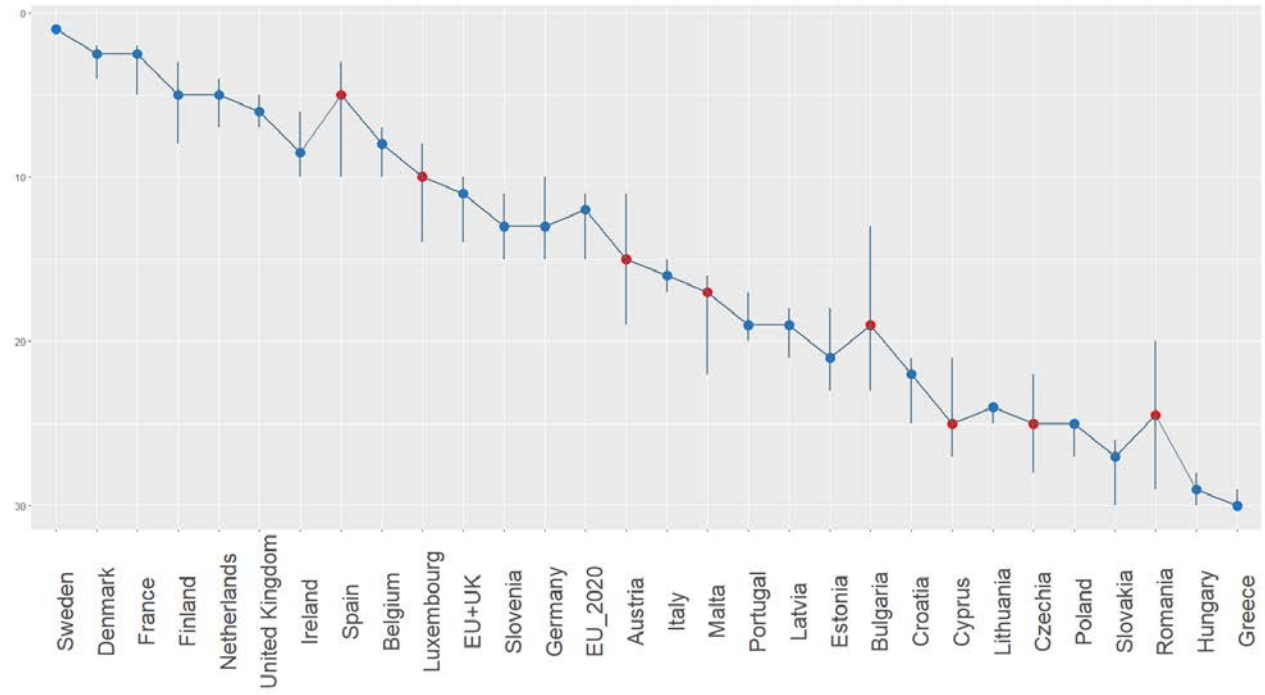
Table 10. Sources of uncertainty examined in the robustness analysis.

| Assumption | Reference | Alternative |
|---|---|---|
| I. Uncertainty in the use of the correcting coefficient | Correcting coefficient | Metric without the correcting coefficient $\alpha$ |
| II. Uncertainty in the aggregation method at domain and index level | Geometric mean (at both levels) | Arithmetic mean (at both levels) |
| III. Uncertainty intervals for the domain weights | Work: 0.19<br>Money: 0.15<br>Knowledge: 0.22<br>Time: 0.15<br>Power: 0.19<br>Health: 0.10 | U[0.14,0.24][2]<br>U[0.12,0.19]<br>U[0.16,0.27]<br>U[0.11,0.18]<br>U[0.14,0.24]<br>U[0.07,0.12] |

The main results obtained from the robustness analysis are shown in Figure 1, with median ranks and 90% intervals computed across the 4,000 Monte Carlo simulations. Countries are ordered from best to worst according to their original index rank where the blue dots represent the median rank among the iterations. For each country, the error bars represent the 90% interval across all simulations, that is, from the 5th to the 95th percentile. Countries for which the intervals have a range larger than 5 are highlighted with a red point.

[2] For the calculations, weights with 15 digits were used. In the table, they are rounded for simplification.

Figure 1.Robustness analysis on ranks (CRII rank vs median rank and 90% intervals)




Source: European Commission, Joint Research Centre, 2020.

Gender Equality Index ranks are shown to be representative of a plurality of scenarios and robust to changes in the absence or presence of the correcting coefficient, aggregation method and domain weights. If one considers the median rank across the simulated scenarios as being representative of these scenarios, then the fact that the index rank is very close to the median rank (less than four positions away) **for all the countries**[3] included in the analysis, suggests that the Gender Equality Index is a suitable summary measure. Furthermore, the reasonable narrow intervals for the majority of the countries' ranks (less than 6 positions for 22 out of 30 countries[3], including the EU averages) imply that the index ranks are also, for most countries, robust to the applied changes. However, it is also true that four country ranks, those of Bulgaria, Romania, Austria and Spain, vary significantly (seven and more positions) with the changes in the assumptions and four other (Cyprus, Malta, Luxemburg and Czechia) have interval ranges of six positions.

That could be due to the lack of balance among their values on the six domains as in the case of Austria that shows a rather large range of different scores in the domains, varying from 44 in the Power domain to 92 in the Health domain. In fact, when a country shows unbalanced values, it is particularly penalised by the geometric mean.

The reasons of the presence of this larger volatility in the four countries are investigated further using sensitivity analysis.

[3] In the uncertainty and sensitivity analysis all 27 EU MS, the UK and the averages EU2020 and EU+UK were used, giving in total 30 "countries".

Overall, country ranks in the Gender Equality Index are fairly robust to changes in the domain weighs, the aggregation formula and the presence or not of the correcting coefficient in the metric for the majority of the countries considered, enough to allow for meaningful inferences to be drawn. For full transparency and information, Table 11 reports the original index country ranks together with the median rank and the simulated intervals (central 90 percentiles observed among the 4,000 scenarios) in order to better appreciate the robustness of these ranks to the computation methodology, and to ease the analysis of the behaviour of specific countries respect to perturbations.

Table 11. Median rank of the countries and 90% confidence intervals.

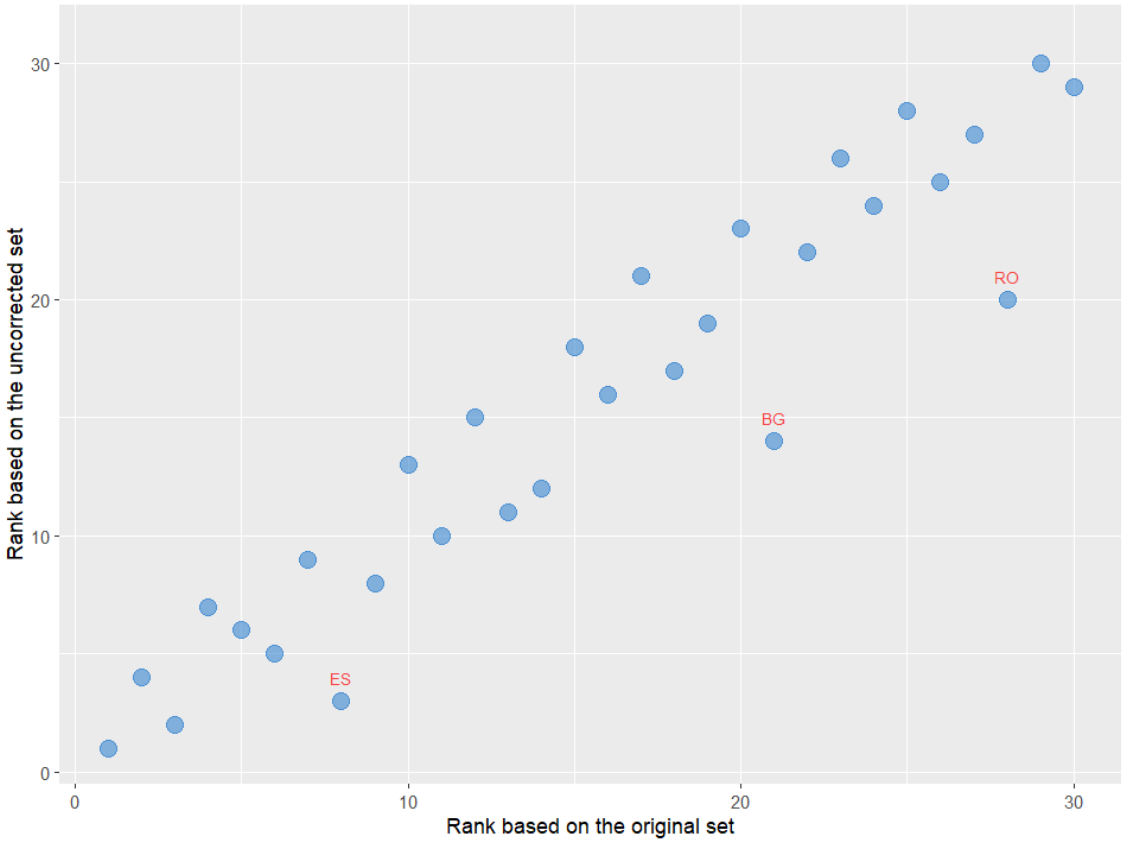
| Country | Gender Equality Index rank | Median rank | 90% Confidence Interval over the 4000 scenarios |
|---|---|---|---|
| SE | 1 | 1 | [1,1] |
| DK | 2 | 2.5 | [2,4] |
| FR | 3 | 2.5 | [2,5] |
| FI | 4 | 5 | [3,8] |
| NL | 5 | 5 | [4,7] |
| UK | 6 | 6 | [5,7] |
| IE | 7 | 8.5 | [6,10] |
| ES | 8 | 5 | [3,10] |
| BE | 9 | 8 | [7,10] |
| LU | 10 | 10 | [8,14] |
| EU+UK | 11 | 11 | [10,14] |
| SI | 12 | 13 | [11,15] |
| DE | 13 | 13 | [10,15] |
| EU_2020 | 14 | 12 | [11,15] |
| AT | 15 | 15 | [11,19] |
| IT | 16 | 16 | [15,17] |
| MT | 17 | 17 | [16,22] |
| PT | 18 | 19 | [17,20] |
| LV | 19 | 19 | [18,21] |
| EE | 20 | 21 | [18,23] |
| BG | 21 | 19 | [13,23] |
| HR | 22 | 22 | [21,25] |
| CY | 23 | 25 | [21,27] |
| LT | 24 | 24 | [24,25] |
| CZ | 25 | 25 | [22,28] |
| PL | 26 | 25 | [25,27] |
| SK | 27 | 27 | [26,30] |
| RO | 28 | 24.5 | [20,29] |
| HU | 29 | 29 | [28,30] |
| EL | 30 | 30 | [29,30] |

Notes: Confidence intervals with range larger than six positions are highlighted in red.
Source: European Commission, Joint Research Centre, 2020.

The uncertainty analysis is complemented by a sensitivity analysis to identify which of the modeling assumptions presented in Table 10 have the highest impact on the countries with the most volatile Gender Equality Index ranks.

In Figure 2, it is possible to compare the ranks derived from the Gender Equality Index with the ranks which would have been obtained by changing the metric and excluding the correcting coefficient *α* from it. The use of the correcting coefficient has a very strong conceptual meaning and it can affect strongly the rank of some countries. The reason it was included in the JRC modelling assumptions to be tested, was to understand how drastically in can affect them and how much it contributes in the volatility of some of those ranks. From the figure though we see that although we do notice some rank changes, in most cases they are less than five positions. Only three countries stand out and these are Bulgaria, Romania and Spain. These countries have a better ranking by eight, seven and five positions respectively, when the correcting coefficient is not in place. This would mean that in some specific indicators, the corr. coeff. *α* of these countries should be particularly low; and in this way it penalises them.

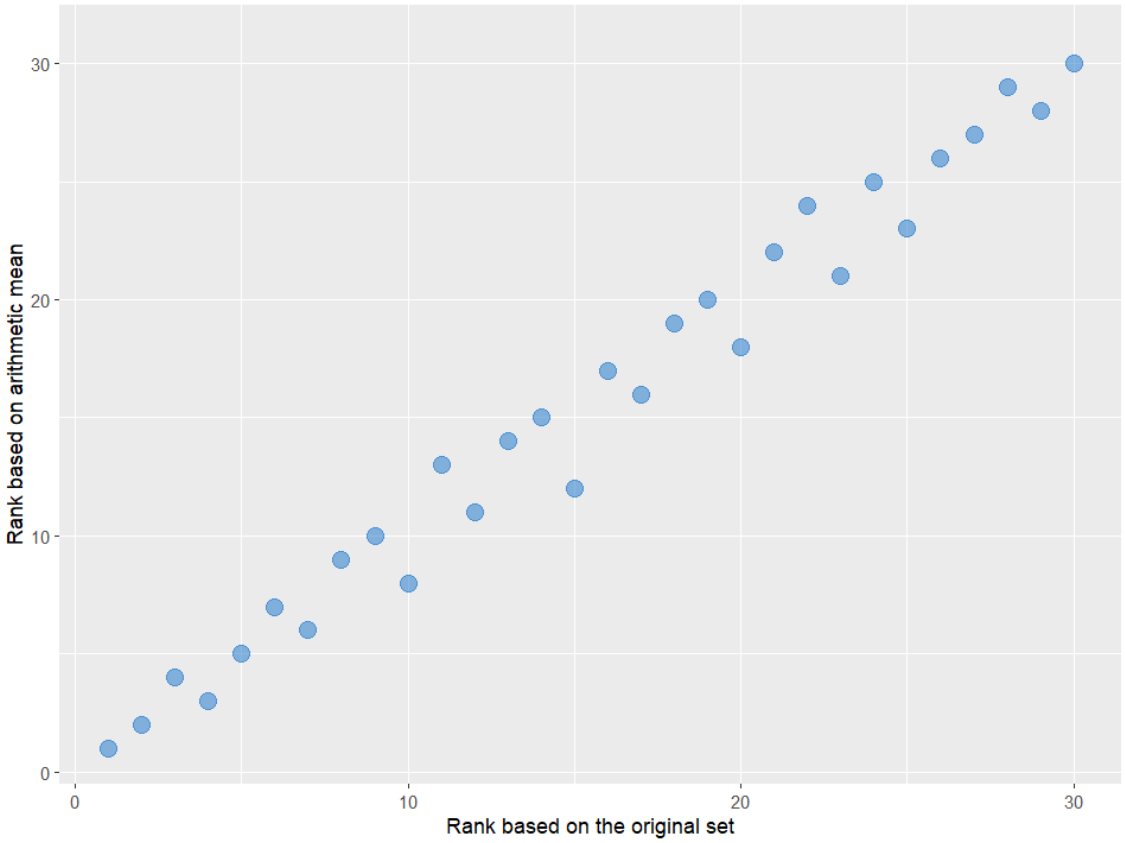
Figure 2. Comparison of ranks according to the use of a metric with and without the correcting coefficient α.




Notes: Countries with rank changes more than four positions are highlighted in red.
Source: European Commission, Joint Research Centre, 2020.

Next, we compare the original Gender Equality Index ranks (geometric aggregation) with the ranks computed using arithmetic aggregation at all levels. In Figure 3 we notice that the changes are even more subtle. In fact, no country has rank changes for more than three positions, which lead us to the conclusion that although the geometric mean do not really allows for compensations in the scores and it does affect the results; the aggregation method does not contribute significantly in the high volatility of the ranks of some countries. From the four countries identified at the uncertainty analysis, only Austria has a rank change of 3 positions, confirming again that its large range in the domain scores is penalised by the geometric average.

Figure 3. Comparison of ranks according to the use of geometric mean (default) and arithmetic mean at domain and index level.




Source: European Commission, Joint Research Centre, 2020.

Last, we looked at the median rank produced from 1,000 stimulations where the domain weights where randomly varying ±25% from their nominal values, keeping unchanged all other developers' choices (metric, geometric aggregation at both top levels). The results show no actual changes in the ranking for any country.

The overall implications of the uncertainty analysis are that the uncertainty in the rankings is mostly manageable; only the ranks of three countries are explicitly affected by the assumption made on the metric (RO, BG, ES). Nevertheless, the index allows meaningful conclusions to be drawn from the index.

# 5 Conclusions

The JRC statistical audit delves into the extensive work carried out by the developers of the Gender Equality Index with the aim of suggesting improvements in terms of data characteristics, structure and methods used. The analysis aims to ensure the transparency of the Gender Equality Index methodology and the reliability of the results.

This report focused on the assessment of the statistical coherence of the Gender Equality index by carrying out a multilevel analysis of the correlations within and across the indicators, sub- domains and domains as well as by an assessment of the impact of key modelling assumptions on the index ranking.

The analysis suggests that the Gender Equality Index is statistically well balanced with respect to its structure. The results of the two statistical coherence tests suggest that the conceptual grouping of the indicators into 14 sub-domains and six domains is statistically confirmed, and that the index is in general influenced by most - though not all - underlying indicators. Seven indicators are found to be not so influential at the index level and for three of those - Healthy life years, Population without unmet medical examination needs and Population without unmet dental examination needs – it was shown that they also have a low impact on the index country ranks. Moreover, these three indicators are more associated with other domains with respect to their own domains. Although they are conceptually enriching, the JRC recommendation is to monitor carefully their performance in the future releases of the index, and eventually fine-tune the framework if needed.

The results of the uncertainty analysis reveal that for most countries, the confidence intervals are narrow enough for meaningful inferences to be drawn from the index: there is a shift of fewer than six positions for 22 of the countries included in the Gender Equality Index. Nevertheless, there are four countries with 90% confidence interval widths of more than six positions, and thus their ranks vary significantly with changes in metric construction, weights and aggregation formula. The four countries are: Bulgaria, Romania, Austria and Spain. The sensitivity analysis that followed confirmed that Austria's volatility is due to the uneven distribution of scores at domain level that is penalised with the use of geometric mean while the rest three countries have a significant rank change due to the use or not of the correcting coefficient in the metric. That volatility could be raising an issue in some cases, but given the fact that the metric with the addition of the correcting coefficient is based in a strong conceptual choice, it is accepted. At the same time it should be aknowledged that a change in the construction of the metric could result in significant changes in the ranking of the specific countries.

In general, the present audit confirms that the 2020 Gender Equality Index meets the quality standards for statistical soundness and acknowledges it as a reliable composite indicator to measure gender equality in the EU context, tailored to fit the EU's policy goals. By looking beyond the overall index scores, the Gender Equality Index allows to provide insights on its underlying categories where the real essence of a composite indicator lies.

## References


European Institute for Gender Equality. (2013). *Gender Equality Index Report*. https://doi.org/10.2839/9948

European Institute for Gender Equality. (2019). *Gender Equality Index 2019 Work — life balance European Institute for Gender Equality*. https://doi.org/10.2839/44985

European Institute of Gender Equality. (2017). *Gender Equality Index 2017, Methodological report* (Vol. 1). https://doi.org/10.2839/3514

Groeneveld, R. A., & Meeden, G. (1984). Measuring Skewness and Kurtosis. *The Statistician*, *33*, 391–99.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data Second Edition*.

Munda, G. (2008). *Social Multi-Criteria Evaluation for a Sustainable Economy*. https://doi.org/10.1007/978-3-540-73703-2_2

OECD and JRC. (2008). *Handbook of Constructing Composite Indicators - Methodology and user guide*. https://doi.org/10.1787/9789264043466-en

Saisana, M, Tarantola, S., & Saltelli, A. (2005). Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. *Journal of the Royal Statistical Society*, *168*(2), 307–323.

Saisana, Michaela, & Saltelli, A. (2011). Rankings and ratings: Instructions for use. *Hague Journal on the Rule of Law*, *3*(2), 247–268. https://doi.org/10.1017/S1876404511200058

**GETTING IN TOUCH WITH THE EU**

**In person**

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

**On the phone or by email**

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

**FINDING INFORMATION ABOUT THE EU**

**Online**

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

**EU publications**
You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-30423-EN-N

## The European Commission's science and knowledge service

Joint Research Centre

### JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

**EU Science Hub**

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/39645

ISBN 978-92-76-24507-0