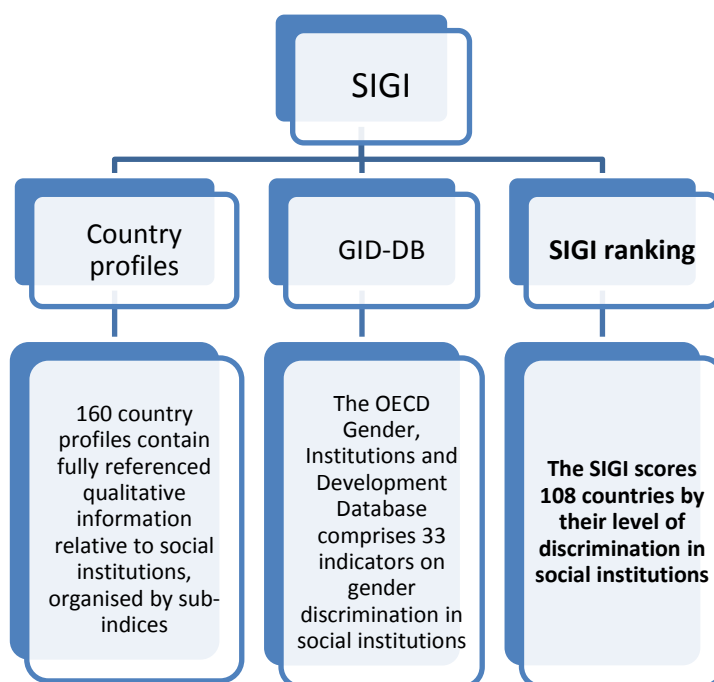# The construction of the SIGI and its sub-indices

The Social Institutions and Gender Index (SIGI) measures discriminatory social institutions that restrict women's access to opportunities, resources and power. The SIGI is based on comprehensive qualitative and quantitative information on formal and informal laws, social norms and practices that discriminate against women and girls in their overall life cycle.

```
                        ┌──────────────────┐
                        │       SIGI       │
                        └──────────────────┘
        ┌───────────────────┬───────────────────┐
┌───────────────┐   ┌───────────────┐   ┌───────────────┐
│    Country    │   │     GID-DB    │   │  SIGI ranking │
│    profiles   │   │               │   │               │
└───────────────┘   └───────────────┘   └───────────────┘
┌───────────────┐   ┌───────────────┐   ┌───────────────┐
│  160 country  │   │   The OECD    │   │               │
│profiles contain│  │    Gender,    │   │  The SIGI     │
│     fully      │   │ Institutions  │   │  scores 108   │
│   referenced   │   │ and           │   │ countries by  │
│  qualitative   │   │ Development   │   │  their level  │
│  information   │   │  Database     │   │      of       │
│ relative to    │   │ comprises 33  │   │discrimination │
│    social      │   │  indicators   │   │  in social    │
│ institutions,  │   │  on gender    │   │ institutions  │
│organised by sub-│  │discrimination │   │               │
│    indices     │   │  in social    │   │               │
│                │   │ institutions  │   │               │
└───────────────┘   └───────────────┘   └───────────────┘
```

While existing gender-specific measures focus on key economic and social indicators, such as employment or education, the SIGI instead captures what drives gender inequalities in economic and social outcomes by examining a neglected aspect of gender inequality: discriminatory social institutions. This technical note explains how the sub-indices are built and aggregated to compute the SIGI.

Selection of variables

SIGI variables were selected on the following criteria, based on the 2012 conceptual framework:

Conceptual relevance: The variable should be closely related to the conceptual framework of discriminatory social institutions and measure what it is intended to capture.
Underlying factor of gender inequality: The variable should capture an underlying factor that leads to unequal outcomes for women and men.
Data quality, reliability and coverage: The variable should be based on high quality, reliable data. Ideally the data should be standardised across countries and have extensive coverage across countries.
Distinction: Each variable should measure a distinct discriminatory institution and should add new
Information not measured by other variables.
Statistical association: Variables included in the same sub-index should be statistically associated, and thereby capture similar dimensions of social institutions without being redundant.

Why is the 0-1 coding used?

The variables, sub-indices and composite indicator are scaled between 0 and 1 for easy interpretation. The best possible performance, i.e. lowest level of inequality, is assigned the value 0, and the worst possible performance, i.e. highest level of inequality, the value 1.

Hence, a value of zero can be considered the goal, and the distance from zero indicates the extent of gender discrimination.

Step 1: Building the Gender, Institutions and Development Database

Truncating quantitative data at the equality benchmark and inverting the scale

The SIGI and its sub-indices range from 0 for low discrimination to 1 for very high discrimination. For some variables, equality is reached at 0.5 instead of 1. Equality in political representation for example is achieved when 50% of parliament members are women. Hence, women share in parliament is truncated at 50% for countries having more than 50% of women.

According to the variable, the scale from low to very high discrimination is inverted to fit with the 0-1 scale. For female political representation for example, a higher share of women in parliamentary seats means lower discrimination against women. Moreover, as equality in political representation is achieved when 50% of parliament members are women, countries having 50% of women have a score of 0.

Figure 1. The composition of the SIGI

Assigning a score to qualitative variables

The qualitative information detailed in the SIGI country profiles are quantified using the following coding manual:

- 0: Women and men enjoy the same rights in the legal framework and practice.
- 0.25: The legislation is not well implemented.
- 0.5: The customary laws and practices discriminate against women.
- 0.75: The legislation is misleading.
- 1: Women and men do not enjoy the same rights in the legal framework.

In cases where no or insufficient information exists, variables are not assigned a value. The legal indicators are assessed based on all applicable legal frameworks, including civil law, religious law, customary law and traditional law.

Constructing indicators

Some indicators are based on one variable while others on several. For example:

$$\text{Parental authority} = \frac{1}{2} Parental\ authority\ during\ marriage + \frac{1}{2} Parental\ authority\ after\ divorce$$

When the indicator aggregates categorical and continuous variables, the latter are the object of discretisation. In this process, the quantitative values of a variable are transformed into qualitative categories.

Where data is available for only one variable of an indicator, the score is based solely on that variable.

Standardising the indicators

Standardisation of the original variables is done by subtracting the mean and then dividing by the standard deviation for continuous variables, using results of an ordered probit model for ordinal categorical variables.

Step 2: Aggregating indicators to build the sub-indices

Measuring association between indicators

Each sub-index combines indicators that are assumed to belong to one dimension of discrimination in social institutions. The statistical association between the indicators is tested using a Kendall Tau b rank correlation analysis and a multiple joint correspondence analysis (MCA). These tests assess the correlation between the indicators combined in each sub-index.

Kendall Tau b

Kendall Tau b ranks correlation statistics and checks for this correlation. It is a non-parametric hypothesis test for statistical dependence based on the tau coefficient. This statistic ranges from zero, meaning no association, to a value of 1 or -1, meaning perfect positive and negative correlation, respectively.

Constructing the sub-indices

The sub-indices aim to provide a summary measure of each dimension of gender discrimination in social institutions. Constructing a sub-index consists of aggregating the indicators with a reasonable weighting scheme through a polychoric principal component analysis (PCA). The first principal component (FPC) is used as a proxy for the common information contained in the indicators: it is the weighted sum of the standardised indicators that captures as much of the variance in the data as possible.

Why are the sub-indices equally weighted in the SIGI?

Equal weights for each sub-index offer two benefits:
- Each dimension of discriminatory social institutions has equal value.
- No dimension is more important than another in terms of the deprivation experienced by women.

Why use polychoric PCA to build sub-indices?

Polychoric PCA has numerous advantages for constructing a sub-index:

- It allows for aggregating continuous and categorical indicators.
- By grouping individual indicators according to their degree of correlation, using endogenous weights, it does not assume a specific underlying structure of the data.
- It summarises the underlying trend and common information captured by raw variables by correcting for statistical bias and redundancy and by preserving the maximum possible proportion of the total variation in the original data set.

The sub-indices are calculated using the following transformation for the country x after generating fictitious *best* and *worst* countries (respectively corresponding to the minimum and the maximum of each indicator):

$$Subindex_x = \frac{FPC_x - FPC\_best}{FPC_{worst} - FPC\_best}$$

Step 3: Computing the SIGI

The SIGI is a composite indicator built as an unweighted average of a non-linear function of the sub-indices:

$$SIGI = \frac{1}{5} \, Discriminatory \; family \; code^2 + \frac{1}{5} \, Restricted \; physical \; integrity^2$$
$$+ \frac{1}{5} \, Son \; bias^2 + \frac{1}{5} \, Restricted \; resources \; and \; assets^2$$
$$+ \frac{1}{5} \, Restricted \; civil \; liberties^2$$

Why square each SIGI sub-index?

The quadratic form is inspired by poverty measures:

- The partial compensation means that very high inequality in one dimension can be only partially offset by low inequality in another dimension.
- It makes for a more egalitarian performance in the sub-indices, which is preferable to a more uneven one.

For more details, refer to the 2014 edition of the SIGI methodological background paper at:
www.genderindex.org