European Commission

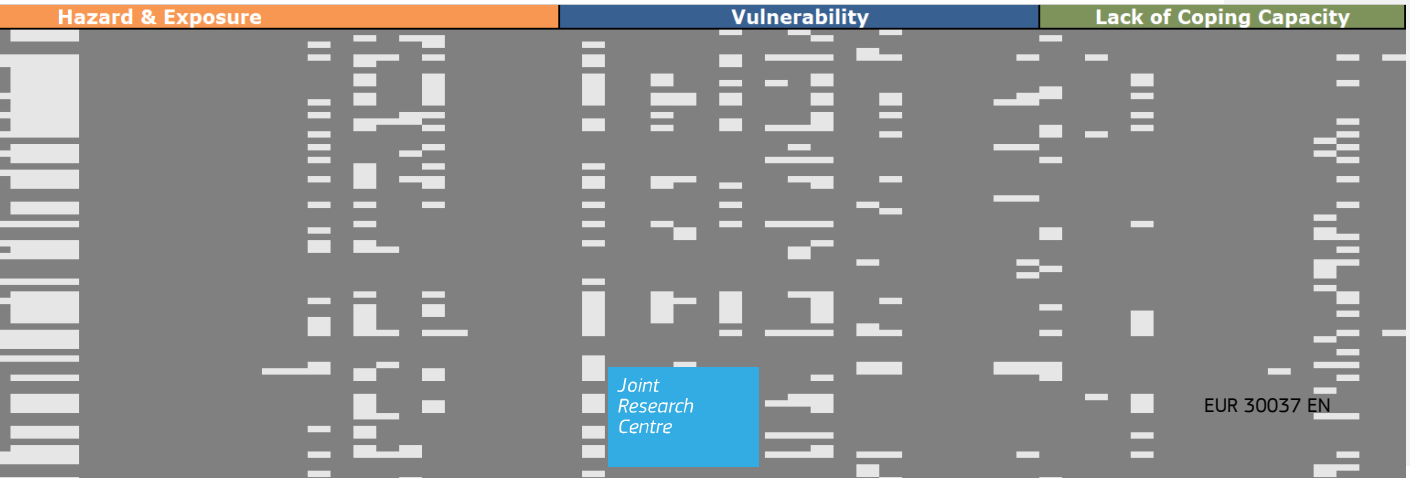
JRC TECHNICAL REPORT

Imputation of missing values in the INFORM Global Risk Index

A multistep approach for imputation of missing values in composite indicators

Poljanšek, K
Vernaccini, L
Nweke, EP
Marin-Ferrer, M

2020



EUR 30037 EN

Contact information
Name: Karmen Poljanšek
Address: European Commission, Joint Research Centre (JRC), Via Fermi, 2479 – Ispra (VA), Italy
Email: karmen.poljansek@ec.europa.eu
Tel.: +39 0332 783650

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

JRC119496

EUR 30037 EN

Print	ISBN 978-92-76-14725-1	ISSN 1018-5593	doi:10.2760/717316
PDF	ISBN 978-92-76-14657-5	ISSN 1831-9424	doi:10.2760/97910

Luxembourg: Publications Office of the European Union, 2020

How to cite this report: Poljansek, K., Vernaccini, L., Nweke, E. and Marin Ferrer, M., Imputation of missing values in the INFORM Global Risk Index, EUR 30037 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-14725-1, doi:10.2760/717316, JRC119496.

Printed in Italy

Authors affiliations:

Karmen Poljanšek, European Commission, Joint Research Centre (JRC), Ispra, Italy
Luca Vernaccini, Fincons SpA external service provider of European Commission, Joint Research Centre (JRC), Ispra, Italy
Eje Philp Nweke, Unisystems Sral exteranl service provider of European Commission, Joint Research Centre (JRC), Ispra, Italy
Marin Ferrer, Montserrat, European Commission, Joint Research Centre (JRC), Ispra, Italy

Contents

Authors

Karmen Poljanšek, as the editor of the report, was responsible for designing the concept and leading the development of the methodology and the preparation of the report.

Luca Vernaccini, as the external consultant, did the study a the status of missing values in INFORM GRI, developed a multi-methods approach for the imputation of missing values applying a variety of scientific and technical solutions and wrote the report.

Philp Eje Nweke, as external consultant, helps to processing the data.

Montserrat Marin Ferrer, as the coordinator of the DRMKC projects, overviewed the whole process.

Abstract

Although they have been selected on the basis of their reliability, consistency, continuity and completeness, most of indicators used in INFORM Global Risk Index do not have global coverage and neither are issued regularly every year. This results in a significant number of missing values, irregularly distributed among countries, time and indicators.

The main motivations for imputing missing values arise from the need to create consistent trends that would otherwise not be possible due to the lack of data in the indicator's time series, and to increase the reliability of the single compound release.

In the presented study we focus on better understanding the patterns and mechanisms of missing values in the INFORM GRI model, and on evaluating their impact on the model's outputs. The scope is to develop a missing data imputation strategy to be implemented in the INFORM GRI that will strongly depend on the reason why data is missing.

1 Introduction

The INFORM Global Risk Index (GRI) is a composite indicator that identifies countries at risk of humanitarian crisis and disaster that would overwhelm national response capacity. The INFORM GRI supports a proactive crisis and disaster management framework.

The INFORM GRI model is based on risk concepts published in scientific literature and envisages three dimensions of risk: hazards & exposure, vulnerability, and lack of coping capacity. The model is split into different levels to provide a quick overview of the underlying factors leading to humanitarian risk and builds up the picture of risk by 73 core indicators.

The initiative to develop the INFORM GRI began in 2012 as a convergence of interests of UN agencies, donors, NGOs and research institutions to establish a common evidence base for global humanitarian risk analysis. Since that time, INFORM has become a multi-stakeholder forum for developing shared analyses to help manage humanitarian crises and disasters. INFORM now has partners from across the UN system, donors, civil society, academic/technical community, and the private sector. The European Commission Joint Research Centre (JRC) is the technical and scientific leader of the model, and responsible for methodological improvements, and their implementation. INFORM has an annual partner conference where strategic developments are discussed among the partners, and JRC will implement the methodological improvements and changes that has been prioritised by the steering committee.

Although they have been selected on the basis of their reliability, consistency, continuity and completeness, most of indicators used in INFORM GRI do not have global coverage and are issued regularly every year. This results in a significant number of missing values, irregularly distributed among countries, time and indicators.

When the INFORM GRI was developed, for the sake of transparency and replicability, it was decided not to explicitly estimate the missing data, a common choice in composite indicators. The only imputation applied, it was using the latest value available as the best proxy for the missing ones.

After 5 years of experience in managing and updating the model, we acquired enough information on the influence that the missing values can have on the INFORM GRI.

The main motivations for imputing missing values arise from the need to create consistent trends that would otherwise not be possible due to the lack of data in the indicator's time series, and to increase the reliability of the single compound release.

JRC presented at the 2018 INFORM annual meeting the intention to work on imputation of the missing values in INFORM GRI. In the same year, JRC published a first report (Marin-Ferrer, 2018), which described a possibility to impute missing values with machine learning techniques by using the Random Forest algorithm. Although the preliminary results were generally promising, the proposed technic did not perform well under certain conditions (i.e. specific type of indicators, or countries). This was mainly due to the large variability of behavioural paths both by country and by indicator. The conclusion was that a sound imputation of the missing values cannot be done without having a prior clusterization of the indicators.

The focus of the presented study is to perform a more comprehensive analysis of the missing values patterns and mechanisms in the INFORM GRI model, evaluate their impact on the final results, with the scope to define a multi-methods approach for handling missing values in the model that will strongly depend on the reason why the values are missing.

The report is organised as follows. The INFORM GRI model and concept is introduced in Section 2. Then, the basic concepts of missing data imputation and the imputations models used in other indexes are introduced and reviewed in Section 3. Sections 4 focuses on analysing the missing data patterns and missingness mechanisms in the INFORM GRI dataset. In the section 5 the influence of the missing values in the current INFORM GRI is assessed along with the preliminary studies on how handle them. Finally, in Section 6 a multi-methods approach for imputing the missing values in INFORM GRI is presented.

2 Short introduction to INFORM Global Risk Index

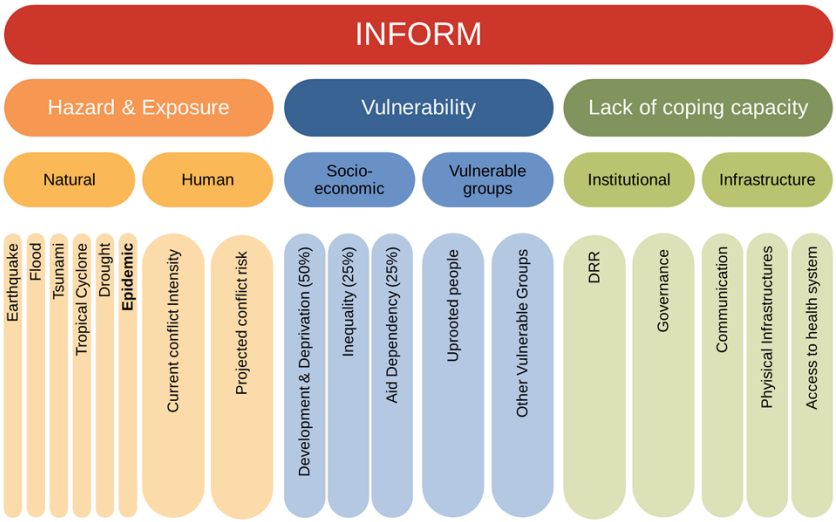
The Index for Risk Management - INFORM - is a way to understand and measure the risk of a humanitarian crisis. The INFORM Global Risk Index (GRI) has been developed to improve the common evidence basis for risk analysis so that all governments, development agencies, disaster risk reduction actors and organisations can work together. INFORM GRI is the first global, open-source, continuously updated, transparent and reliable tool for understanding risk of humanitarian crises and disasters. It covers 191 countries. It use only quantitative indicators provided by the most trusted sources. The methodology is completely transparent and based on scientific concepts and methods. The index results are published twice a year. This year was the sixth edition of the INFORM GRI.

The JRC is the main scientific partner in the INFORM GRI process, and has lead the bottom-up process of building a consensus-based new methodology, taking into account the requirements of participating institutions as well as limitations of data availability.

INFORM GRI is a composite indicator developed by the JRC by combining many indicators into three dimensions of risk (**Figure 1**): hazards (events that could occur) and exposure to them, vulnerability (the susceptibility of communities to those hazards) and the lack of coping capacity (lack of resources that can alleviate the impact). It results in an overall risk score out of 10 for each country, and for each of the dimensions, categories, and components of risk.

INFORM GRI is a widely recognised and valuable tool that supports decision-making of INFORM partners and others. The INFORM risk analysis process and methodology has been extended to the regional and country level and adapted to many scopes and targets.

Figure 1: INFORM GRI Conceptual Framework



3 Imputation of missing values

What is missing data? In statistics, 'missing data (or missing values) is defined as the data value that is not stored for a variable in the observation of interest' (Kang, 2013). Missing data are a common occurrence and can have a significant effect on the conclusions that can be drawn from the data. 'Imputation of missing data on a variable is replacing that missing by a value that is drawn from an estimate of the distribution of this variable' (Dondersa et al., 2006). Understanding the data and the domain from which it comes is very important before imputing it.

3.1 Handling missing values in the composite indicators

Like most statistical series, composite indicators shared the problem of missing values (MV) (OECD, 2008). In many cases, data are only available for a limited number of countries or only for certain indicators. Indicators are not updated at the same time and with the same temporal frequency, leading to an incomplete resulting dataset. MV can render the composite indicator less reliable for the countries for which only limited information is available and can distort the relative standing of all countries in the composite. There are a number of approaches for dealing with MV in composite indicators, which include data deletion, single imputation (hot deck, regression), multiple imputation, machine learning (random forest), or **ignore them** (take the average index of the remaining indicators). The latest approach is particular to the composite indicators, and widely adopted in the first releases of the INFORM GRI (De Groeve, 2014).

3.2 Type of missing values

One of the important issues with MV is to understand the missing data mechanism. It affects how much the MV bias the results, and it needs to consider it when choosing an approach to deal with the MV. How you deal with MV is dependent on the type of missingness.

Rubin (1976) define three qualitatively distinct types of MV, which are based on the relationship between the missing data mechanism and the missing and observed values.

The missing patterns could be (**Table 1**):

1. **missing completely at random (MCAR)**: MV do not depend on the variable of interest or on any other observed variable in the data set. It means that there is no relationship between whether a data point is missing and any values in the data set, missing or observed.
2. **missing at random (MAR)**: MV do not depend on the variable of interest, but are conditional on other variables in the data set.
3. **Not missing at random (NMAR)**: MV depend on the values themselves.

Table 1. Distribution of Missingness

MCAR	$\boldsymbol{P(R\|Y_{com}) = P(R)}$
Missingness does not depend on data	
MAR	$\boldsymbol{P(R\|Y_{com}) = P(R\|Y_{obs})}$
Missingness depends only on observed data	
NMAR	$\boldsymbol{P(R\|Y_{com}) = P(R\|Y_{obs}\|Y_{mis})}$
Missingness depends on observed and missing data	

Source: Rubin, 1976

MV missing at random can be treated using different imputation techniques, while nothing can be done for values structural missing (missing not at random). When there are reasons to assume a non-random missing

pattern (NMAR), the pattern must be explicitly modelled and included in the analysis (OECD, 2008). This could be very difficult and could imply ad hoc assumptions that are likely to influence the result of the entire exercise. The only way to obtain an unbiased estimate of the parameters in such a case is to model the MV but that requires proper understanding and domain knowledge of the missing variable (Kang, 2013).

The type of MV in the indicators used in INFORM GRI are mostly (if even not all of them) missing not at random (chapter 4.2). Therefore, most of the statistical methods available for imputation of MV missing at random would not be helpful for our scope.

In a recent workshop (Ispra, 11/10/2019) with experts from the European Commission Competence Centre on Composite Indicators and Scoreboards (COIN[1]), it was suggested that classifying the MV according to the Rubin's types would be not meaningful for the INFORM GRI model (and in general for the majority of the composite indicators). It was emphasised to get the higher benefit from expert knowledge about the indicators, their meaning and behaviour. This, in combination with a more deep analysis of the statistical characteristics of the MV in the indicators used in INFORM GRI (next chapter), it would enable the identification of a proper strategy on the imputation of the MV in the model.

3.3 Example of other Indexes

It is interesting to know what the other indexes do regarding imputation of MV. This would give to us a stronger confidence on the approach to adopt with our index.

Several indexes with a published methodology show a use of a combination of different methods for estimating the MV. Normally, is the knowledge about the indicators that guides the decision of which estimation method should be adopted.

Table 2 summarises the imputation approaches and methods used by five composite indicators. A part from the Global Innovation Index (GII), all the others indexes use a combination of different imputation methods, quantitative (regression, KNN) and qualitative (expert opinion, qualitative research).

Table 2. Resume table of the imputation strategy applied by other indexes

Index	MV imputation approach	MV imputation methods
Social Progress Index (SPI)	Ad hoc method for each indicator	Linear estimation, qualitative research, regression, regional cohort estimates
Global Peace Index (GPI)	Ad hoc method for each indicator	Manual imputation by expert, most recent data, alternative sources
Ocean Health Index (OHI)	Ad hoc method for each indicator	Most recent data, Zero-filled, regional average, regression
States of Fragility, OECD	Ad hoc method for each indicator	Zero-filled, k-nearest neighbour (KNN) imputation
Global Innovation Index (GII), WIPO	No imputation	

The description of the different imputation strategy adopted by the identified indexes is in the Annex 2.

[1] https://composite-indicators.jrc.ec.europa.eu/

4 Analysis of the MV in the INFORM GRI

The scope of this chapter is to give an overview of the patterns, reasons and distribution of MV in the INFORM GRI model. Where are the missing values located? What indicators have more MV? Which are the countries the more affected ones? Are values missing randomly? Why do we have those MV?

Important is to define what is a MV in the context of the INFORM GRI model (and in general for the composite indicators). In fact, in order to calculate the index, one (year) value for each indicator for each country is needed. Considering that a common pre-imputation method (also used in INFORM GRI) is to use the most recent data, the MV are only the ones where there is not recent data for a particular country for a specific indicator. Instead, if we extend the analysis to the whole time series, the MV are all the ones where the combination of indicator/country/year is not available.

For the purpose of this note, we refer to the former as type A, and the latter as type B:

- **Type A**: MV for indicator / country / most recent
- **Type B**: MV for indicator / country / current year (day) [before first imputation]

The type A are the MV of the dataset we need to produce a single release of the index, which consist of almost 14,000 values (191 countries times 73 indicators).

The type B are the MV of the larger dataset needed for producing the time-series of the INFORM GRI, which includes 10 years of the data and therefore consist of almost 140,000 values (191 countries times 73 indicators times 10 years).

4.1 Missing data patterns

4.1.1 Type A

In the INFORM GRI 2020 model, the percent of MV of type A is relatively low (**11%**, slightly increased from 8% of the previous version of the model due to the inclusion of the 21 new indicators of the epidemic component) (**Table 4**, **Figure 2**); however, it varied substantially among countries and indicators. In general, small countries or countries with autocratic regimes (**Table 3**) from one side, and the indicators in the Vulnerability dimension (13%) on the other, require the most of the gap filling.

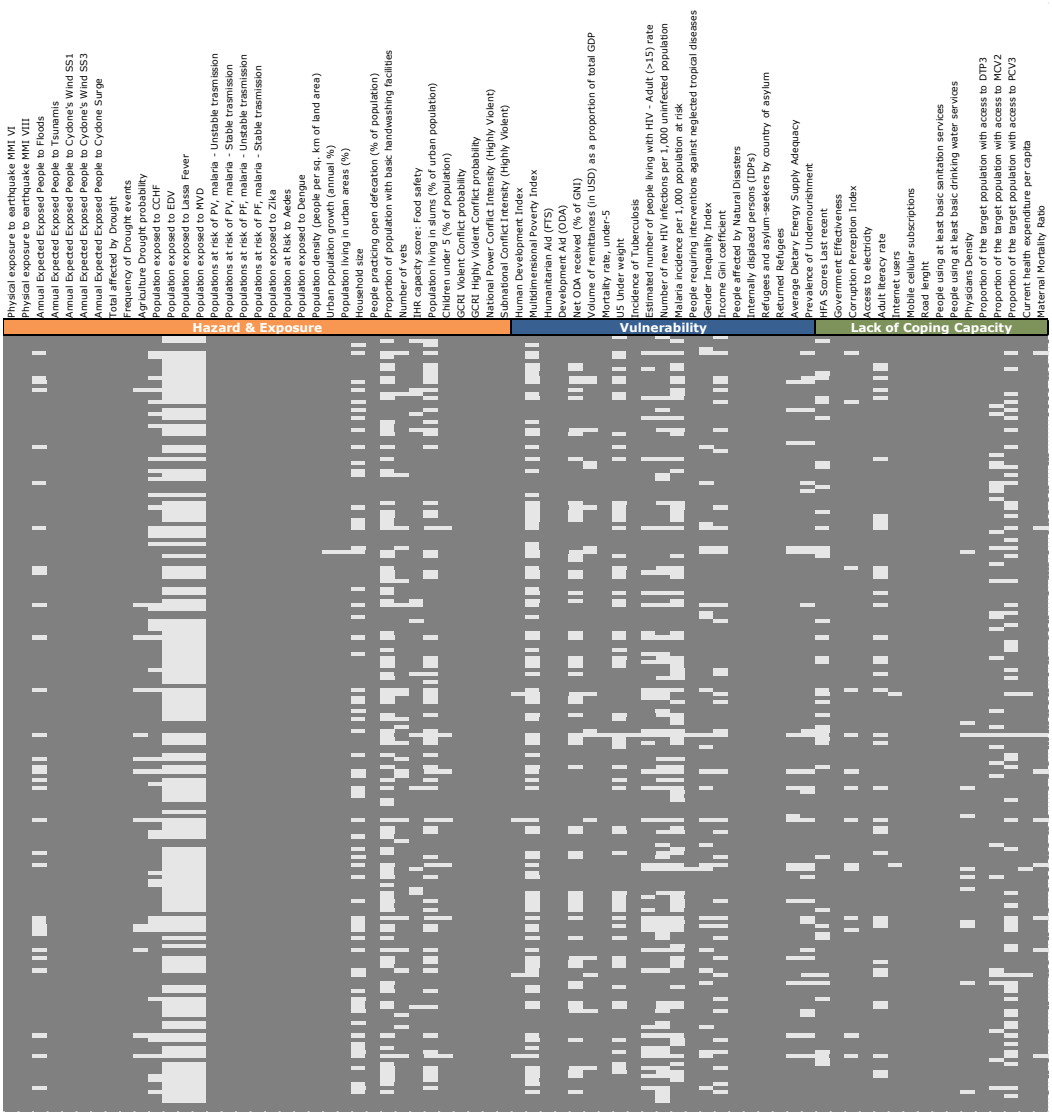
Are the missing values too many?

As rules of thumb, according to the JRC COIN (Damioli, 2018):

- At the country level: not more than 35% of indicators should have MV
- At the indicator level: not more than 35% of countries should have MV

Thresholds have to be considered thoroughly, also reflecting indicator importance and the conceptual framework. A good correlation between indicators supports more MV.

Figure 2. Missing values after the pre-imputation (Type A), INFORM GRI 2020



4.1.1.1 Country analysis

In INFORM GRI 2020, only 14 countries have all data, while 19 countries have more than 20% of MV of type A (Table 3).

Only one country, Liechtenstein, exceed the threshold of 35% of MV (paragraph 4.1.1). However, the set of indicators in the INFORM Risk model was selected to address mainly the situations of developing countries, hence the indicators are not the most suitable ones for countries like Liechtenstein, which is not a relevant country for risk of humanitarian crisis.

Small Island Developing States[2] (SIDS) dominate the ranking with more MV, together with autocratic regimes (Korea DPR, Eritrea).

[2] https://sustainabledevelopment.un.org/topics/sids/list

Table 3. Missing values of type A by country in INFORM GRI 2020 version

Country	Number of MV	% of MV
Liechtenstein	32	43%
Nauru	25	33%
Tuvalu	24	32%
Saint Kitts and Nevis	23	31%
Dominica	21	28%
Antigua and Barbuda	20	27%
Grenada	20	27%
Marshall Islands	20	27%
Saint Vincent and the Grenadines	20	27%
Micronesia	19	25%
Palau	19	25%
Kiribati	18	24%
Iceland	17	23%
Korea DPR	17	23%
Tonga	17	23%
Bahamas	16	21%
Bahrain	15	20%
Palestine	15	20%
Samoa	15	20%
Brunei Darussalam	14	19%
Eritrea	14	19%
Seychelles	14	19%
Singapore	14	19%
Barbados	13	17%
Belgium	13	17%
Finland	13	17%
Ireland	13	17%
Israel	13	17%
Luxembourg	13	17%
Malta	13	17%
New Zealand	13	17%
Qatar	13	17%
Saint Lucia	13	17%
Solomon Islands	13	17%
Switzerland	13	17%

Country	Number of MV	% of MV
United Arab Emirates	13	17%
Austria	12	16%
Canada	12	16%
Cuba	12	16%
Turkmenistan	12	16%
United States of America	12	16%
Cabo Verde	11	15%
Croatia	11	15%
Czech Republic	11	15%
Equatorial Guinea	11	15%
Estonia	11	15%
Fiji	11	15%
Germany	11	15%
Japan	11	15%
Kuwait	11	15%
Latvia	11	15%
Norway	11	15%
Poland	11	15%
Russian Federation	11	15%
Saudi Arabia	11	15%
Somalia	11	15%
Sweden	11	15%
United Kingdom	11	15%
Vanuatu	11	15%
Venezuela	11	15%
Belarus	10	13%
Cyprus	10	13%
Denmark	10	13%
France	10	13%
Korea Republic of	10	13%
Libya	10	13%
Lithuania	10	13%
Maldives	10	13%
Mauritius	10	13%
Netherlands	10	13%
Oman	10	13%

Country	Number of MV	% of MV
Papua New Guinea	10	13%
Slovakia	10	13%
Syria	10	13%
Australia	9	12%
Belize	9	12%
Bhutan	9	12%
Bosnia and Herzegovina	9	12%
Bulgaria	9	12%
Greece	9	12%
Hungary	9	12%
Italy	9	12%
Lebanon	9	12%
Portugal	9	12%
Romania	9	12%
South Sudan	9	12%
Spain	9	12%
Trinidad and Tobago	9	12%
Albania	8	11%
Chile	8	11%
China	8	11%
Malaysia	8	11%
Slovenia	8	11%
Sri Lanka	8	11%
Suriname	8	11%
Timor-Leste	8	11%
Turkey	8	11%
Ukraine	8	11%
Uruguay	8	11%
Argentina	7	9%
Guyana	7	9%
Iran	7	9%
Jamaica	7	9%
Jordan	7	9%
North Macedonia	7	9%
Uzbekistan	7	9%
Afghanistan	6	8%

Country	Number of MV	% of MV
Azerbaijan	6	8%
Bolivia	6	8%
Georgia	6	8%
Iraq	6	8%
Lao PDR	6	8%
Moldova Republic of	6	8%
Montenegro	6	8%
Nicaragua	6	8%
Panama	6	8%
Tajikistan	6	8%
Thailand	6	8%
Brazil	5	7%
Colombia	5	7%
Comoros	5	7%
Costa Rica	5	7%
Djibouti	5	7%
Dominican Republic	5	7%
El Salvador	5	7%
Haiti	5	7%
Honduras	5	7%
Mongolia	5	7%
Peru	5	7%
Philippines	5	7%
Serbia	5	7%
Viet Nam	5	7%
Cambodia	4	5%
Central African Republic	4	5%
Chad	4	5%
Ecuador	4	5%
Gabon	4	5%
Guatemala	4	5%
Guinea-Bissau	4	5%
India	4	5%
Indonesia	4	5%
Kazakhstan	4	5%
Mexico	4	5%

Country	Number of MV	% of MV
Paraguay	4	5%
Yemen	4	5%
Algeria	3	4%
Armenia	3	4%
Bangladesh	3	4%
Botswana	3	4%
Guinea	3	4%
Kyrgyzstan	3	4%
Liberia	3	4%
Madagascar	3	4%
Mauritania	3	4%
Myanmar	3	4%
Nepal	3	4%
Nigeria	3	4%
Pakistan	3	4%
Sao Tome and Principe	3	4%
Tunisia	3	4%
Burundi	2	3%
Congo	2	3%
Congo DR	2	3%
Lesotho	2	3%
Namibia	2	3%
Uganda	2	3%
Angola	1	1%
Benin	1	1%
Cameroon	1	1%
Côte d'Ivoire	1	1%
Egypt	1	1%
Eswatini	1	1%
Ethiopia	1	1%
Mali	1	1%
Morocco	1	1%
Niger	1	1%
Togo	1	1%
Burkina Faso	0	0%
Gambia	0	0%

Country	Number of MV	% of MV
Ghana	0	0%
Kenya	0	0%
Malawi	0	0%
Mozambique	0	0%
Rwanda	0	0%
Senegal	0	0%
Sierra Leone	0	0%
South Africa	0	0%
Sudan	0	0%
Tanzania	0	0%
Zambia	0	0%
Zimbabwe	0	0%

4.1.1.2 Indicator analysis

In INFORM GRI 2020, 33 (45% of the total) indicators have all data, while 15 indicators have more than 20% of MV of type A (**Table 4**).

Seven indicators exceed the 35% threshold of MV (paragraph 4.1.1). Of those, the three indicators of exposure to viral haemorrhagic fever (Population exposed to EDV, Population exposed to Lassa Fever, Population exposed to MVD) are covering only African countries, where such a diseases are mostly concentrated (Pigott and at., 2017). The other four indicators (Proportion of population with basic handwashing facilities on premises, Population living in slums, Multidimensional Poverty Index, Malaria incidence) are only relevant for a subset of countries due to their nature.

Table 4. Missing values of type A by indicators in INFORM GRI 2020 version

Indicator Name	Number of MV	% of MV
Physical exposure to earthquake MMI VI	0	0%
Physical exposure to earthquake MMI VIII	0	0%
Annual Expected Exposed People to Floods	32	17%
Annual Expected Exposed People to Tsunamis	0	0%
Annual Expected Exposed People to Cyclone's Wind SS1	0	0%
Annual Expected Exposed People to Cyclone's Wind SS3	0	0%
Annual Expected Exposed People to Cyclone Surge	0	0%
Total affected by Drought	0	0%
Frequency of Drought events	0	0%
Agriculture Drought probability	13	7%
Population exposed to CCHF	44	23%
Population exposed to EDV	139	73%
Population exposed to Lassa Fever	139	73%

Indicator Name	Number of MV	% of MV
Population exposed to MVD	139	73%
Populations at risk of Plasmodium vivax malaria in 2010 - Unstable transmission	0	0%
Populations at risk of Plasmodium vivax malaria in 2010 - Stable transmission	0	0%
Populations at risk of Plasmodium falciparum malaria in 2010 - Unstable transmission	0	0%
Populations at risk of Plasmodium falciparum malaria in 2010 - Stable transmission	0	0%
Population exposed to Zika	0	0%
Population at Risk to Aedes	0	0%
Population exposed to Dengue	0	0%
Population density (people per sq. km of land area)	0	0%
Urban population growth (annual %)	1	1%
Population living in urban areas (%)	1	1%
Household size	65	34%
People practicing open defecation (% of population)	0	0%
Proportion of population with basic handwashing facilities on premises (% of population)	96	50%
Number of vets	23	12%
IHR capacity score: Food safety	15	8%
Population living in slums (% of urban population)	68	36%
Children under 5 (% of population)	7	4%
GCRI Violent Conflict probability	0	0%
GCRI Highly Violent Conflict probability	0	0%
National Power Conflict Intensity (Highly Violent)	0	0%
Subnational Conflict Intensity (Highly Violent)	0	0%
Hazard & Exposure	**782**	**12%**
Human Development Index	4	2%
Multidimensional Poverty Index	94	49%
Humanitarian Aid (FTS)	0	0%
Development Aid (ODA)	0	0%
Net ODA received (% of GNI)	59	31%
Volume of remittances (in USD) as a proportion of total GDP (%)	16	8%
Mortality rate, under-5	1	1%
U5 Under weight	61	32%
Incidence of Tuberculosis	1	1%
Estimated number of people living with HIV - Adult (>15) rate	40	21%
Number of new HIV infections per 1,000 uninfected population	64	34%

Indicator Name	Number of MV	% of MV
Malaria incidence per 1,000 population at risk	84	44%
Number of people requiring interventions against neglected tropical diseases	2	1%
Gender Inequality Index	29	15%
Income Gini coefficient	38	20%
People affected by Natural Disasters	0	0%
Internally displaced persons (IDPs)	0	0%
Refugees and asylum-seekers by country of asylum	0	0%
Returned Refugees	0	0%
Average Dietary Energy Supply Adequacy	17	9%
Prevalence of Undernourishment	30	16%
Vulnerability	**540**	**13%**
HFA Scores Last recent	40	21%
Government Effectiveness	0	0%
Corruption Perception Index	14	7%
Access to electricity	0	0%
Adult literacy rate	38	20%
Internet users	2	1%
Mobile cellular subscriptions	0	0%
Road length	0	0%
People using at least basic sanitation services (% of population)	0	0%
People using at least basic drinking water services (% of population)	0	0%
Physicians Density	12	6%
Proportion of the target population with access to 3 doses of diphtheria-tetanus-pertussis (DTP3) (%)	1	1%
Proportion of the target population with access to measles-containing-vaccine second-dose (MCV2) (%)	31	16%
Proportion of the target population with access to pneumococcal conjugate 3rd dose (PCV3) (%)	59	31%
Current health expenditure per capita	4	2%
Maternal Mortality Ratio	9	5%
Lack of Coping Capacity	**210**	**7%**
Total	**1532**	**11%**

4.1.2 Type B

Looking at the whole time-series 2009-2018 of the INFORM GRI dataset, the percentage of MV (type B) increases substantially to **50%** (**Table 6**).

4.1.2.1 Country analysis

We couldn't notice any significant patterns in the availability of data by country among the last ten years of data (**Table 5**). The pick in 2015 is most probably depending on the starting of the SDG reporting. While the weak coverage in 2018 is due to the time required to consolidate the data and let them public.

The data availability pattern is very similar among countries. The lack of evidence of decreasing of data availability like for countries in protracting crisis can be explain by the fact that most of the indicators used in INFORM GRI are based on estimation by the data providers.

Table 5. Number of observed indicators by country in the historical series (2009-2018)

Country	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Afghanistan	43	50	48	46	48	51	63	48	43	27
Albania	43	47	46	46	45	45	57	44	41	25
Algeria	43	50	47	47	47	48	62	46	43	26
Angola	42	48	44	44	47	49	62	48	44	26
Antigua and Barbuda	38	42	39	39	41	41	48	38	37	25
Argentina	43	53	46	47	50	51	58	45	42	27
Armenia	45	56	52	49	50	52	62	50	45	28
Australia	42	49	44	44	45	47	58	43	41	27
Austria	40	47	45	43	44	46	57	40	40	26
Azerbaijan	45	51	48	48	50	52	61	47	45	28
Bahamas	38	44	41	41	43	43	53	40	41	25
Bahrain	41	48	42	43	44	45	55	41	40	25
Bangladesh	44	53	51	48	51	55	63	49	47	28
Barbados	42	48	44	46	47	46	55	42	41	24
Belarus	45	48	44	44	47	47	58	43	41	25
Belgium	39	46	43	43	44	46	55	41	39	27
Belize	42	48	46	45	45	48	56	44	42	25
Benin	41	49	46	48	47	50	64	46	42	27
Bhutan	39	47	43	46	45	46	57	41	41	26
Bolivia	43	51	48	48	47	50	59	48	44	26
Bosnia and Herzegovina	41	47	44	45	46	47	57	43	40	27
Botswana	42	48	46	46	48	49	64	45	43	26
Brazil	43	52	48	48	51	50	59	46	44	28
Brunei Darussalam	38	42	40	39	40	40	52	36	34	23
Bulgaria	41	49	46	45	46	47	57	42	41	27
Burkina Faso	45	52	46	46	46	53	64	49	45	28
Burundi	43	54	47	48	50	50	65	50	46	27
Cabo Verde	41	47	44	45	45	45	57	42	40	26
Cambodia	44	53	46	47	48	53	60	46	43	27

Country	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Cameroon	41	51	49	45	46	51	64	46	43	27
Canada	38	46	42	43	45	45	54	41	40	26
Central African Republic	42	51	45	45	46	49	62	44	39	26
Chad	40	49	45	44	47	49	64	46	42	27
Chile	44	47	47	45	49	49	58	44	43	27
China	42	51	46	46	46	47	57	40	39	26
Colombia	45	57	49	48	50	53	63	49	45	28
Comoros	40	45	42	46	45	45	57	42	39	25
Congo	43	49	50	47	47	50	63	46	42	26
Congo DR	42	51	45	47	49	51	64	46	44	27
Costa Rica	45	52	51	48	50	51	60	46	44	26
Côte d'Ivoire	43	51	47	50	47	51	65	47	44	27
Croatia	41	48	44	42	43	44	55	40	37	27
Cuba	40	47	44	45	45	45	54	42	41	26
Cyprus	42	48	46	44	45	47	58	43	42	28
Czech Republic	41	48	45	44	45	45	56	41	40	27
Denmark	43	49	44	45	46	48	58	42	41	27
Djibouti	39	45	41	44	46	48	59	44	43	27
Dominica	36	41	37	37	38	39	47	34	33	24
Dominican Republic	42	54	48	47	51	51	60	46	44	26
Ecuador	45	56	50	50	50	53	62	49	44	27
Egypt	43	52	45	48	48	53	66	47	46	27
El Salvador	44	53	49	48	50	51	60	47	45	26
Equatorial Guinea	39	46	41	41	43	44	57	39	38	24
Eritrea	40	47	42	39	41	42	55	39	34	24
Estonia	40	46	44	43	44	46	57	41	40	27
Eswatini	45	53	47	46	47	50	64	46	39	27
Ethiopia	42	49	47	45	48	51	64	48	45	27
Fiji	40	44	41	41	45	45	55	41	39	26
Finland	39	47	42	43	44	45	54	41	39	27
France	40	48	46	45	46	47	58	43	41	26
Gabon	40	47	43	47	44	46	60	42	40	25
Gambia	41	51	45	46	51	50	65	45	44	27
Georgia	45	50	48	48	49	51	63	47	45	28
Germany	41	48	45	43	45	45	57	42	40	27
Ghana	42	53	46	47	49	52	64	48	45	27

Country	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Greece	40	47	44	44	47	48	59	42	40	25
Grenada	38	43	40	39	41	42	49	39	37	25
Guatemala	43	49	45	46	49	52	63	46	45	28
Guinea	41	48	43	47	44	47	61	46	41	26
Guinea-Bissau	42	49	45	44	46	49	62	45	42	27
Guyana	46	51	46	46	47	51	59	45	43	27
Haiti	41	50	47	49	47	49	58	46	44	29
Honduras	42	50	48	51	50	52	60	48	44	27
Hungary	41	47	46	44	47	48	59	43	41	27
Iceland	40	46	41	43	44	46	53	41	39	27
India	44	50	48	47	48	50	61	47	45	27
Indonesia	47	56	49	51	51	51	63	48	46	28
Iran	42	48	46	45	48	49	59	45	41	25
Iraq	41	50	46	45	46	50	59	44	44	28
Ireland	39	46	45	43	44	46	55	41	40	27
Israel	38	46	41	43	43	46	55	42	39	27
Italy	41	47	45	44	47	47	60	43	42	27
Jamaica	43	51	46	48	48	50	58	43	42	27
Japan	40	49	43	44	44	46	55	42	41	25
Jordan	44	50	43	46	45	47	57	44	41	27
Kazakhstan	45	52	49	49	50	51	62	45	45	27
Kenya	47	51	47	47	49	54	67	47	45	28
Kiribati	39	44	40	38	41	41	49	38	35	24
Korea DPR	33	38	35	37	37	39	49	35	36	24
Korea Republic of	40	49	42	44	44	45	56	41	40	26
Kuwait	41	48	43	45	46	47	57	42	42	28
Kyrgyzstan	49	53	49	52	49	52	61	46	45	27
Lao PDR	40	47	43	45	45	47	56	42	40	25
Latvia	40	46	45	43	44	46	58	42	39	27
Lebanon	42	48	47	45	47	48	59	43	44	28
Lesotho	45	51	45	45	46	50	63	44	42	26
Liberia	43	51	46	45	49	50	62	45	42	26
Libya	39	44	41	38	42	45	57	39	40	26
Liechtenstein	29	34	29	29	30	32	39	28	26	21
Lithuania	40	47	45	43	44	46	57	43	41	27
Luxembourg	40	47	44	44	45	46	56	41	41	26

Country	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Madagascar	46	51	46	46	46	49	62	44	41	28
Malawi	44	55	45	46	47	52	64	48	44	27
Malaysia	41	50	47	46	47	46	60	43	41	26
Maldives	44	46	41	40	42	44	52	41	37	25
Mali	43	50	48	49	49	50	65	46	44	28
Malta	39	46	44	42	43	45	55	39	37	26
Marshall Islands	35	41	38	37	38	39	48	36	35	23
Mauritania	40	47	45	44	46	49	62	43	43	25
Mauritius	42	48	46	46	46	47	57	41	40	27
Mexico	44	55	50	53	51	54	63	50	45	28
Micronesia	38	43	39	39	42	40	49	37	36	23
Moldova Republic of	43	50	47	49	50	51	61	46	43	27
Mongolia	42	52	46	46	47	49	60	46	43	27
Montenegro	42	48	46	44	46	47	56	42	41	27
Morocco	42	49	48	45	48	50	62	44	44	25
Mozambique	46	51	49	46	48	50	63	46	45	28
Myanmar	43	50	47	47	49	51	61	52	46	28
Namibia	42	49	46	45	49	48	63	45	44	27
Nauru	33	40	37	35	36	36	46	35	33	24
Nepal	43	52	50	47	48	51	61	50	46	27
Netherlands	41	48	44	44	46	47	58	43	41	27
New Zealand	39	46	42	43	44	45	55	42	40	27
Nicaragua	43	48	46	47	47	50	57	44	43	28
Niger	43	49	45	48	47	51	63	45	43	27
Nigeria	42	49	45	43	49	48	63	46	43	28
North Macedonia	44	49	47	46	48	48	60	44	42	28
Norway	43	49	45	45	46	47	56	42	41	26
Oman	43	49	44	44	46	49	58	43	43	27
Pakistan	46	55	51	49	53	51	63	46	45	27
Palau	37	44	37	37	40	41	50	36	35	24
Palestine	40	46	41	41	42	45	53	42	38	25
Panama	44	54	48	48	49	50	59	47	44	26
Papua New Guinea	41	48	44	43	46	48	58	43	43	28
Paraguay	44	50	46	50	49	52	60	48	46	28
Peru	48	55	51	53	49	52	61	50	46	27
Philippines	45	52	48	47	51	51	63	46	45	26

Country	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Poland	40	47	45	44	46	47	57	42	38	26
Portugal	41	47	47	44	45	46	57	42	41	26
Qatar	41	47	42	43	44	47	56	42	42	27
Romania	40	47	47	44	44	45	57	43	39	27
Russian Federation	41	50	43	44	45	46	57	42	42	28
Rwanda	41	55	45	48	49	50	67	48	44	27
Saint Kitts and Nevis	36	41	38	38	39	35	45	33	31	24
Saint Lucia	42	45	42	44	43	43	52	42	38	27
Saint Vincent and the Grenadines	39	44	40	40	41	42	50	39	37	26
Samoa	38	45	41	41	44	45	53	40	37	26
Sao Tome and Principe	39	46	41	42	43	46	59	42	40	26
Saudi Arabia	40	47	42	43	44	47	56	42	40	27
Senegal	43	50	50	46	53	54	67	50	46	27
Serbia	44	51	46	44	46	50	58	44	41	27
Seychelles	38	45	41	42	42	41	55	39	36	26
Sierra Leone	42	52	48	46	51	50	63	45	43	27
Singapore	41	49	43	44	45	46	54	43	40	26
Slovakia	43	48	45	44	46	47	58	43	41	27
Slovenia	41	48	44	44	47	48	59	44	41	27
Solomon Islands	38	44	43	42	44	42	52	42	39	26
Somalia	39	44	41	41	43	46	58	42	41	27
South Africa	43	53	49	49	48	52	66	48	45	27
South Sudan	25	35	38	39	42	46	58	40	38	26
Spain	42	48	46	44	45	47	59	43	40	27
Sri Lanka	45	51	46	48	48	48	60	46	45	28
Sudan	38	47	42	44	49	51	63	46	44	28
Suriname	42	49	43	44	46	49	57	44	42	28
Sweden	42	49	44	45	46	47	57	43	39	27
Switzerland	41	48	43	44	45	46	57	41	39	26
Syrian Arab Republic	44	49	44	43	45	49	56	41	40	27
Tajikistan	45	51	47	50	49	49	59	44	42	26
Tanzania	42	53	46	49	47	51	65	47	44	27
Thailand	43	52	46	48	49	50	62	49	45	28
Timor-Leste	41	47	42	43	45	46	55	43	40	26
Togo	44	53	48	46	47	53	66	45	43	26

Country	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Tonga	41	46	41	41	43	43	53	39	37	26
Trinidad and Tobago	42	48	45	43	47	47	57	42	41	26
Tunisia	42	50	45	47	46	49	63	45	41	27
Turkey	45	53	47	48	50	51	60	45	42	28
Turkmenistan	40	48	44	44	46	48	56	42	40	27
Tuvalu	36	41	36	36	37	39	46	36	33	24
Uganda	44	50	48	48	46	51	64	50	44	27
Ukraine	42	49	45	46	47	50	59	44	42	28
United Arab Emirates	38	46	41	42	43	45	57	41	39	26
United Kingdom	40	48	45	44	45	46	57	41	40	26
United States of America	40	49	41	43	44	44	53	42	39	27
Uruguay	42	49	49	46	48	49	58	45	45	27
Uzbekistan	43	50	46	46	48	49	58	44	42	26
Vanuatu	40	47	42	42	44	43	52	39	38	26
Venezuela	43	48	45	44	45	46	56	42	39	26
Viet Nam	45	53	46	47	47	51	60	46	42	27
Yemen	44	50	48	47	52	53	61	45	42	27
Zambia	42	54	46	46	48	51	64	47	43	26
Zimbabwe	40	50	47	45	46	51	64	45	43	26

4.1.2.2 Indicator analysis

Many indicators with a large number of MV (more than 90%), are actually designed not to be updated frequently (**Table 6**). We can notice a large absence of recent data. Most of the indicators don't have yet observed data for 2018, the year before the publication of the last INFORM GRI. This means that in most of the cases we have to deal with data of one or more years old.

As country coverage, there are 1627 (0.9%) combination of indicator/country without any historical values. Of those, 796 (1.2%) are in the Hazard & Exposure dimension, 630 (1.5%) in Vulnerability, and 210 (0.9%) in the Lack of Coping Capacity.

Table 6. Number of MV of type B by indicators the historical series (2009-2018)

Indicator	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	% of MV	Country coverage (*)
Physical exposure to earthquakes of MMI VI	191	191	191	191	191	191	0	191	191	191	90%	0
Physical exposure to earthquakes of MMI VIII	191	191	191	191	191	191	0	191	191	191	90%	0
Physical exposure to flood	191	191	191	191	191	191	32	191	191	191	92%	32

Indicator	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	% of MV	Country coverage (*)
Physical exposure to tropical cyclone of Saffir-Simpson 1	191	191	191	191	191	191	0	191	191	191	90%	0
Physical exposure to tropical cyclone of Saffir-Simpson 3	191	191	191	191	191	191	0	191	191	191	90%	0
Physical exposure to surge from tropical cyclone	191	191	191	191	191	191	0	191	191	191	90%	0
Physical exposure to tsunamis	191	191	191	191	191	191	0	191	191	191	90%	0
Drought affected	191	191	191	191	191	191	191	191	191	0	90%	0
Drought Frequency	191	191	191	191	191	191	191	191	191	0	90%	0
Agriculture Stress Index Probability	191	191	191	191	191	191	191	191	13	191	91%	13
Population density (people per sq. km of land area)	0	0	0	0	0	0	0	0	0	0	0%	0
Population living in slums (% of urban population)	191	141	191	191	191	97	191	74	191	191	86%	68
IHR capacity score: Food safety	191	191	191	191	191	191	191	191	191	15	91%	15
Proportion of population with basic handwashing facilities on premises (% of population)	137	126	115	106	98	97	101	107	113	191	62%	96
People practicing open defecation (% of population)	2	2	1	1	1	2	4	6	11	191	12%	1
Population at Risk to Aedes	191	191	191	191	191	191	0	191	191	191	90%	0
Population exposed to Dengue	191	191	191	191	191	191	0	191	191	191	90%	0
Populations at risk of PF malaria in 2010 - Stable transmission	191	0	191	191	191	191	191	191	191	191	90%	0
Populations at risk of PF malaria in 2010 - Unstable transmission	191	0	191	191	191	191	191	191	191	191	90%	0
Populations at risk of PV malaria in 2010 - Stable transmission	191	0	191	191	191	191	191	191	191	191	90%	0
Populations at risk of PV malaria in 2010 - Unstable transmission	191	0	191	191	191	191	191	191	191	191	90%	0
Population exposed to Zika	191	191	191	191	191	191	0	191	191	191	90%	0
Population exposed to CCHF	191	191	191	191	191	191	44	191	191	191	92%	44
Population exposed to EDV	191	191	191	191	191	191	139	191	191	191	97%	139
Population exposed to Lassa Fever	191	191	191	191	191	191	139	191	191	191	97%	139

Indicator	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	% of MV	Country coverage (*)
Population exposed to MVD	191	191	191	191	191	191	139	191	191	191	97%	139
Household size	167	157	106	171	178	179	178	184	191	191	89%	78
Children under 5 (% of population)	7	7	7	7	7	7	7	7	7	7	4%	7
Urban population growth	1	1	1	1	1	1	1	1	1	1	1%	1
Population living in urban areas	1	1	1	1	1	1	1	1	1	1	1%	1
Number of vets	191	191	191	191	191	41	41	80	51	88	66%	23
Conflict HIIK National Power	0	0	0	0	0	0	0	0	0	0	0%	0
Conflict HIIK Subnational	0	0	0	0	0	0	0	0	0	0	0%	0
GCRI Highly Violent Internal Conflict probability	0	0	0	0	0	0	0	0	0	0	0%	0
GCRI Violent Internal Conflict probability	0	0	0	0	0	0	0	0	0	0	0%	0
Hazard & Exposure												**796**
Human Development Index	10	5	5	5	5	5	5	5	4	191	13%	5
Multidimensional Poverty Index	179	167	172	164	178	173	183	171	189	191	93%	89
Income Gini coefficient	113	107	115	114	116	117	117	152	167	191	69%	30
Gender Inequality Index	190	57	91	43	39	36	32	191	31	191	47%	31
Volume of remittances (in USD) as a proportion of total GDP (%)	21	20	19	19	19	18	19	20	22	25	11%	17
Net ODA received (% of GNI)	54	52	55	55	54	58	62	62	59	191	37%	48
International humanitarian aid by the Financial Tracking Service	0	0	0	0	0	0	0	0	0	0	0%	0
Official Development Assistance	54	55	57	51	51	51	51	51	51	191	35%	0
Internally Displaced People	145	143	141	141	139	135	140	137	140	136	73%	130
Refugees and asylum-seekers by country of asylum	15	15	15	15	15	15	15	15	15	15	8%	0
Refugees (Palestine)	187	187	187	188	187	187	187	187	187	187	98%	0
Returned refugees	15	15	15	15	15	15	15	15	15	15	8%	0
Average dietary supply adequacy	18	18	17	17	17	17	17	17	17	17	9%	17
Prevalence of undernourishment (% of population)	31	31	31	30	30	30	30	30	30	30	16%	30

Indicator	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	% of MV	Country coverage (*)
Estimated number of people living with HIV - Adult (>15) rate	80	79	78	80	76	80	85	57	56	191	45%	38
Number of new HIV infections per 1,000 uninfected population	65	65	65	65	65	65	65	65	64	191	41%	64
Malaria incidence per 1,000 population at risk	191	84	84	84	84	84	84	84	84	191	55%	84
Tuberculosis prevalence	2	2	1	1	1	1	2	1	2	191	11%	1
Number of people requiring interventions against neglected tropical diseases	191	2	2	2	2	2	2	2	2	191	21%	2
Children Under Weight	163	156	167	159	170	169	188	171	191	191	90%	43
Child Mortality	1	1	1	1	1	1	1	1	2	191	11%	1
People affected by Natural Disasters	0	0	0	0	0	0	0	0	0	0	0%	0
Vulnerability												**630**
Hyogo Framework for Action	116	119	76	79	55	191	100	191	191	191	69%	40
Corruption Perception Index	15	18	13	19	21	20	17	18	14	14	9%	11
Government Effectiveness	1	1	0	0	0	0	0	0	1	191	10%	0
Literacy rate, adult total (% of people ages 15 and above)	161	148	139	156	131	138	40	177	182	190	77%	38
Access to electricity	0	0	0	0	0	0	0	0	0	191	10%	0
Mobile cellular subscriptions (per 100 people)	4	1	2	1	2	2	0	2	1	191	11%	0
Internet users (per 100 people)	6	4	3	5	3	3	3	6	191	191	22%	2
People using at least basic drinking water services (% of population)	1	1	0	0	0	1	1	2	5	191	11%	0
People using at least basic sanitation services (% of population)	1	1	0	0	0	1	1	3	6	191	11%	0
Roads, total network (km)	191	191	191	191	191	0	191	191	191	191	90%	0
Proportion of the target population with access to 3 doses of diphtheria-tetanus-pertussis (DTP3) (%)	2	2	1	1	1	1	1	1	1	191	11%	1
Proportion of the target population with access to measles-containing-vaccine second-dose (MCV2) (%)	71	70	67	60	54	47	41	33	31	191	35%	31

Indicator	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	% of MV	Country coverage (*)
Proportion of the target population with access to pneumococcal conjugate 3rd dose (PCV3) (%)	163	147	128	115	102	86	71	63	59	191	59%	59
Physicians Density	102	71	88	85	86	91	111	99	142	184	55%	6
Maternal Mortality	7	7	7	7	7	7	7	7	7	191	13%	9
Current health expenditure per capita	7	6	6	6	7	7	7	10	191	191	23%	4
Lack of Coping Capacity												**201**
TOTAL	**7665**	**6305**	**7045**	**7036**	**6985**	**6672**	**4436**	**7090**	**7322**	**10475**	**50%**	**1627**

(*) Number of countries with no values in the whole time-series.

Figure 3. Number of missing values of type B by indicators the historical series (2009-2018) – Hazard & Exposure

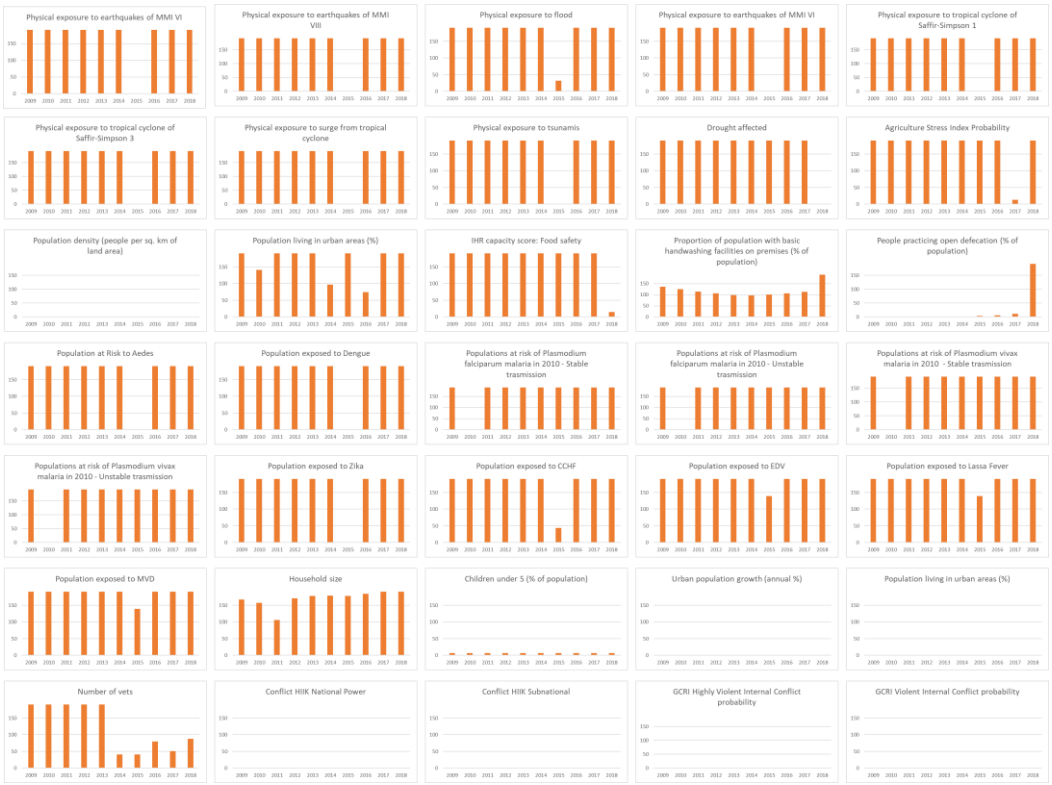


Figure 4. Number of missing values of type B by indicators the historical series (2009-2018) - Vulnerability

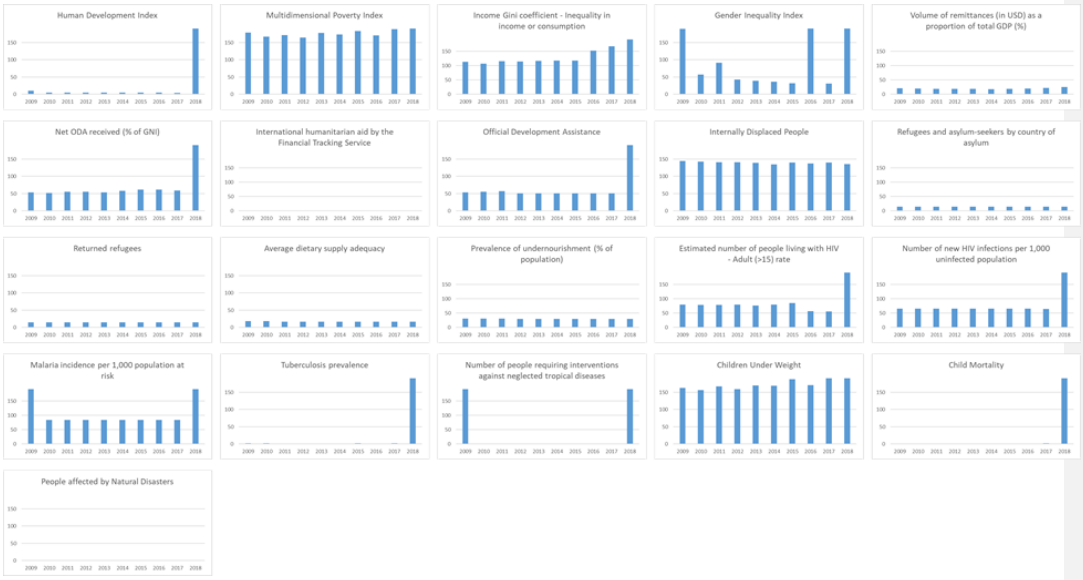


Figure 5. Number of missing values of type B by indicators the historical series (2009-2018) – Lack of coping capacity

4.2 Missing data mechanisms

In the INFORM GRI data set, more than one variable have MV, and they may not all have the same mechanism. It is worthwhile diagnosing the mechanism for each variable with missing data before choosing an approach.

Almost all the missing data are not missing at random NMAR. In fact, for most of the indicators the missingness is specifically related to what is missing (MV depend on the country size, on the region, on the income level, or on the crisis status of the single country), or data are systematically missed for a subset of countries or years (**Table 7**).

Table 7. Example of missing data patterns in the INFORM GRI indicators

Indicator	Reason for missingness
Physical exposure to flood	Depends on the country size (small countries)
Agriculture Stress Index Probability	Depends on the country size (small countries)
Population exposed to EDV	Data available only in Africa
Population exposed to Lassa Fever	Data available only in Africa
Population exposed to MVD	Data available only in Africa
Population living in slums (% of urban population)	Countries without slums + other countries
Human Development Index	Systematically missing for subset of countries
Multidimensional Poverty Index	Developed countries + others
Official Development Assistance	Donors
Net ODA received (% of GNI)	Donors + country not reporting GDP
Malaria incidence per 1,000 population at risk	Countries without malaria + other countries
Average dietary supply adequacy	Systematically missing for subset of countries
Prevalence of undernourishment (% of population)	Systematically missing for subset of countries
Hyogo Framework for Action	Systematically missing for subset of countries
Corruption Perception Index	Systematically missing for subset of countries
Internet users (per 100 people)	Systematically missing for subset of countries
Proportion of the target population with access to measles-containing-vaccine second-dose (MCV2) (%)	Systematically missing for subset of countries
Proportion of the target population with access to pneumococcal conjugate 3rd dose (PCV3) (%)	Systematically missing for subset of countries
Current health expenditure per capita	Systematically missing for subset of countries
Maternal Mortality	Systematically missing for subset of countries

4.3 Patterns in time series analysis

The characteristic of time series is that successive values in the data file represent consecutive measurements taken at equally spaced time intervals.

Identifying the pattern of observed time series data and describe it is helpful for predicting missing values of the time series.

Time series exhibit specific patterns:

1. **Constant** – time series remain at roughly the same level over time.
2. **Trend** - reflects the long-term progression of the series. A trend exists when there is a persistent increasing or decreasing direction in the data. The trend component does not have to be linear.
3. **Cyclic** - reflects repeated but non-periodic fluctuations. The duration of these fluctuations is usually of at least two years.
4. **Seasonal** - reflects seasonality present in the Time Series data, like demand for flip flops, will be highest during the summer season. Seasonality occurs at a fixed period of time could be weekly, monthly, quarterly, etc.
5. **Random** - reflects random or irregular influences.

The INFORM GRI time series have annual frequency, and therefore the Seasonal pattern is not relevant. Also Cyclic patterns are not observed in any indicators used in INFORM GRI, and therefore the time series patterns are either Constant, Trend or Random.

Missing values for Constant and Trend time series patterns are easier to predict, while the imputation is much more difficult with Random patterns. **Table 8** summarises the time series data patterns for all the INFORM GRI indicators.

Table 8. Time series patterns in the INFORM GRI indicators

Indicator	Time series pattern
Physical exposure to earthquakes of MMI VI	Constant
Physical exposure to earthquakes of MMI VIII	Constant
Physical exposure to flood	Constant
Physical exposure to tropical cyclone of Saffir-Simpson 1	Constant
Physical exposure to tropical cyclone of Saffir-Simpson 3	Constant
Physical exposure to surge from tropical cyclone	Constant
Physical exposure to tsunamis	Constant
Drought affected	Constant
Drought Frequency	Constant
Agriculture Stress Index Probability	Constant
Population density (people per sq. km of land area)	Trend
Population living in slums (% of urban population)	Trend
IHR capacity score: Food safety	Trend
Proportion of population with basic handwashing facilities on premises	Trend
People practicing open defecation (% of population)	Trend
Population at Risk to Aedes	Constant
Population exposed to Dengue	Constant
Populations at risk of PF malaria in 2010 - Stable transmission	Constant
Populations at risk of PF malaria in 2010 - Unstable transmission	Constant
Populations at risk of PV malaria in 2010 - Stable transmission	Constant
Populations at risk of PV malaria in 2010 - Unstable transmission	Constant
Population exposed to Zika	Constant
Population exposed to CCHF	Constant
Population exposed to EDV	Constant
Population exposed to Lassa Fever	Constant
Population exposed to MVD	Constant
Household size	Trend
Children under 5 (% of population)	Trend
Urban population growth (annual %)	Trend
Population living in urban areas (%)	Trend

Indicator	Time series pattern
Number of vets	Trend
Conflict HIIK National Power	Random
Conflict HIIK Subnational	Random
GCRI Highly Violent Internal Conflict probability	Trend
GCRI Violent Internal Conflict probability	Trend
Human Development Index	Trend
Multidimensional Poverty Index	Trend
Income Gini coefficient - Inequality in income or consumption	Trend
Gender Inequality Index	Trend
Volume of remittances (in USD) as a proportion of total GDP (%)	Trend
Net ODA received (% of GNI)	Random
International humanitarian aid by the Financial Tracking Service	Random
Official Development Assistance	Random
Internally Displaced People	Random
Refugees and asylum-seekers by country of asylum	Random
Refugees (Palestine)	Random
Returned refugees	Random
Average dietary supply adequacy	Trend
Prevalence of undernourishment (% of population)	Trend
Estimated number of people living with HIV - Adult (>15) rate	Trend
Number of new HIV infections per 1,000 uninfected population	Trend
Malaria incidence per 1,000 population at risk	Trend
Tuberculosis prevalence	Trend
Number of people requiring interventions against neglected tropical diseases	Trend
Children Under Weight	Trend
Child Mortality	Trend
People affected by Natural Disasters	Random
Hyogo Framework for Action	Trend
Corruption Perception Index	Trend
Government Effectiveness	Trend
Literacy rate, adult total (% of people ages 15 and above)	Trend
Access to electricity	Trend
Mobile cellular subscriptions (per 100 people)	Trend
Internet users (per 100 people)	Trend
People using at least basic drinking water services (% of population)	Trend
People using at least basic sanitation services (% of population)	Trend

Indicator	Time series pattern
Roads, total network (km)	Constant
Proportion of the target population with access to 3 doses of diphtheria-tetanus-pertussis (DTP3) (%)	Trend
Proportion of the target population with access to measles-containing-vaccine second-dose (MCV2) (%)	Trend
Proportion of the target population with access to pneumococcal conjugate 3rd dose (PCV3) (%)	Trend
Physicians Density	Trend
Maternal Mortality	Trend
Current health expenditure per capita	Trend

5 Dealing with missing data: impact of missing values in INFORM GRI

Contrary to most of the statistical tools, composite indicators do not necessary require imputation of MV (Paragraph 3.1). A composite indicators exercise is itself a method to estimate MV based on the other indicators values according to the theoretical framework. The consequence of the 'no imputation' choice in an (arithmetic or geometric) average is that it is equivalent to replacing an indicator's missing value for a given country with the respective component score.

However, this choice has limitations. It works only if there is a high cross-correlation between the indicators within the same component. The observed data coverage of the other indicators belong to the same component should be good enough to compensate the MV (if for a country a value is missed for the indicator, there should be at least an observed value for the other indicators in the same component). It also need that all the observed data are updated frequently, otherwise the imputation would be based on old data. In other words, the 'no imputation' choice is not a

5.1 How missing values are currently handle in INFORM GRI

When the INFORM GRI was developed, for simplicity, transparency and replicability, it was decided not to explicitly estimate the missing data (with the exception of a pre-imputation based on the most recent data available), a common choice in composite indicators. Currently in the INFORM GRI, missing values are handled using a combination of simple methods.

Firstly, a pre-imputation process (chapter 4) is performed in order to reduce the amount of MV (**Table 9**). This enables to reduce the number of MV from 50% to 11% (chapter 4).

If data for some countries are not available for a given year, a systematic imputation of missing values is using the data from the most recent year available. The acceptable time-range depends on the characteristics of the indicator (i.e. older values are still representative for structural indicators with low variation in time, while indicators with faster variation need more recent data).

The missing values of the Human Development Index (HDI), which is the key indicator for assessing the level of development in the INFORM GRI model (Development & Deprivation component), are imputed using a linear regression with the GDP per capita (De Groeve, 2014).

The gaps in the two indicators in the Food Security component (Average dietary supply adequacy, Prevalence of undernourishment) are filled using the regional average or a value of a country with similar condition (according to the expert judgment).

For the remaining MV after the pre-imputation, the 'no imputation' is adopted.

Table 9. Overview of MV pre-imputation in the current INFORM GRI

Indicator	**Current pre-imputation methods**
Physical exposure to flood	Fixed value
Agriculture Stress Index Probability	None
Population living in slums (% of urban population)	Most recent
IHR capacity score: Food safety	None
Proportion of population with basic handwashing facilities on premises (% of population)	Most recent
People practicing open defecation (% of population)	Most recent
Population exposed to CCHF	None
Population exposed to EDV	None
Population exposed to Lassa Fever	None
Population exposed to MVD	None

Indicator	Current pre-imputation methods
Household size	Most recent
Children under 5 (% of population)	None
Urban population growth (annual %)	Most recent
Population living in urban areas (%)	Most recent
Number of vets	Most recent
Human Development Index	Linear regression
Multidimensional Poverty Index	Most recent
Income Gini coefficient - Inequality in income or consumption	Most recent
Gender Inequality Index	Most recent
Volume of remittances (in USD) as a proportion of total GDP (%)	Most recent
Net ODA received (% of GNI)	Most recent
International humanitarian aid by the Financial Tracking Service	Zero-filled
Official Development Assistance	Zero-filled
Internally Displaced People	Zero-filled
Average dietary supply adequacy	Regional avg
Prevalence of undernourishment (% of population)	Regional avg
Estimated number of people living with HIV - Adult (>15) rate	Most recent
Number of new HIV infections per 1,000 uninfected population	Most recent
Malaria incidence per 1,000 population at risk	Most recent
Number of people requiring interventions against neglected tropical diseases	Most recent
Children Under Weight	Most recent
Hyogo Framework for Action	Most recent
Corruption Perception Index	Most recent
Literacy rate, adult total (% of people ages 15 and above)	Most recent
Mobile cellular subscriptions (per 100 people)	Most recent
People using at least basic drinking water services (% of population)	Most recent
People using at least basic sanitation services (% of population)	Most recent
Proportion of the target population with access to 3 doses of diphtheria-tetanus-pertussis (DTP3) (%)	Most recent
Proportion of the target population with access to measles-containing-vaccine second-dose (MCV2) (%)	Most recent
Proportion of the target population with access to pneumococcal conjugate 3rd dose (PCV3) (%)	Most recent
Physicians Density	Most recent
Maternal Mortality	Most recent
Current health expenditure per capita	Most recent

5.2 Random forest regression applied to INFORM GRI

In the 2017, JRC started investigating on the MV in the INFORM GRI model, experimenting machine learning methods for imputing them (Marin-Ferrer, 2018).

This new approach is based in the field of supervised learning an area inside artificial intelligence often called non-linear regression (Trevor Hastie, 2009). Among every available model for regression, we have tested different methods like Ridge, Lasso, ElasticNet or Random Forest Regression, finding the latter the best balance between performance and complexity.

Random Forest Regression (RFR) is a non-parametric imputation method applicable to various variable types. It uses multiple decision trees to estimate missing values. These predictions come with an error level to provide a level of accuracy to the model.

The preliminary results confirmed some of the expected advantages, but also highlighted problems and limitation of the use of RFR to impute missing values in a composite indicators if no prior clusterization of indicators is done (**Table 10**).

Table 10. Pros and cons of the Random Forest Regression approach in INFORM GRI.

Pros	**Cons**
Good average performance	Bad performance for country outliers (i.e. in protracting crisis), and where there are not historical values
Room for improvements (clustering, anomalies detection)	Unreliable estimations for indicators with random patterns
Fills the gaps in the time-series	Need statistical methods for further development
Updating of not recent values	Long running-time, difficult to integrate it in the INFORM GRI workflow
Estimation of the uncertainty	Difficult to

There are other techniques to apply in this predictive model not just based on the information provided by the indicator and the correlation between one and other.

Cluster analysis can be used as a method for selecting groups of countries or indicators for the imputation of missing data with a view to decreasing the variance of the imputed values. With this approach is possible to detect hidden structures inside the data like patterns in the countries (some countries have the same behaviour in some indicators).

Anomaly detection techniques detect anomalies in an unlabeled dataset under the assumption that the majority of the instances in the data set are normal by looking for instances that seem to fit least to the remainder of the dataset. This technique could detect patterns in the missing data providing more information about the reason o those gaps in the dataset.

5.3 Impact of missing values

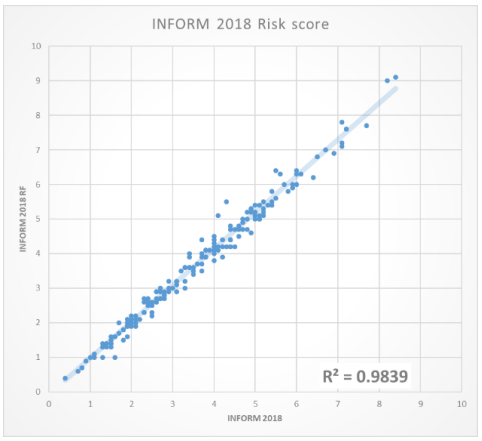
A way to test the impact of missing values in composite indicators is to apply a sensitivity analysis using different imputation methods and evaluate the differences in the results. Larger is the difference, higher is the uncertainty related to the presence of missing data (OECD, 2008).

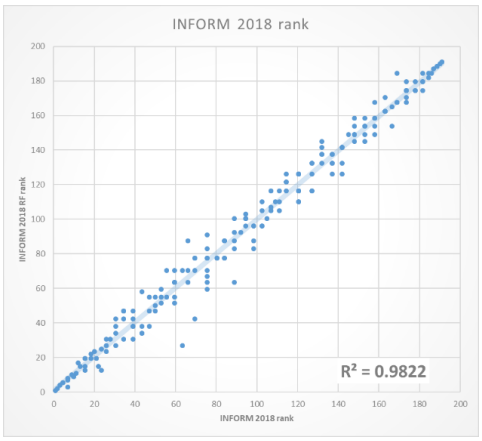
To test the impact of MV on the INFORM GRI, we compared the results of the model with the current imputation strategy (chapter 5.1), with the ones obtained with the estimated missing data using the Random Forest algorithm without applying any prior clusterization (Marin-Ferrer, 2017).

5.3.1 Sensitivity Analysis Results

Sensitivity analysis has been used to test the influence of the missing values in the model, by comparing the INFORM GRI 2018 (defined as 'no imputation' choice), with the results of having imputed the missing data with the RF method. **Figure 6** plots the original INFORM GRI 2018 scores and ranking versus the ones obtained with the RF imputation (INFORM 2018 RF).

Figure 6. Sensitivity analysis: No imputation (INFORM 2018) vs RF imputation (INFORM 2018 RF) – Risk





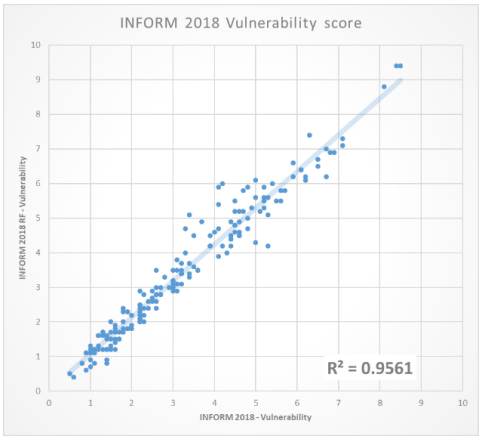
Pearson's correlation[3]: 0.99

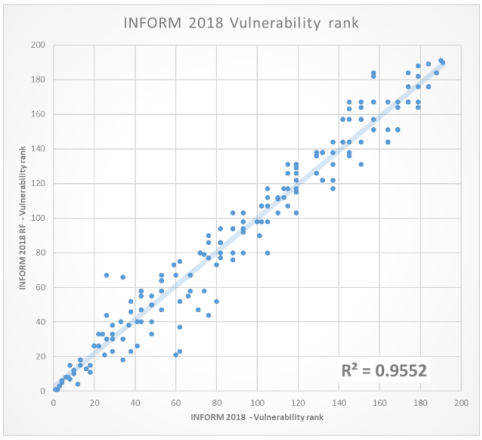
Spearman's correlation[4]: 0.99

Very high values of Spearman's correlation coefficient (0.99), and Pearson's cc (0.99) show that the impact of the missing values on the final risk score is not significant.

However, the variation of the results is not equal among the countries and components. The variation in score and ranking by dimensions is slightly more significant. Anyway, Pearson cc and Spearman cc are still very high for vulnerability (**Figure 7**), and lack of coping capacity (**Figure 8**).

Figure 7. Sensitivity analysis: No imputation (INFORM 2018) vs RF imputation (INFORM 2018 RF) - Vulnerability





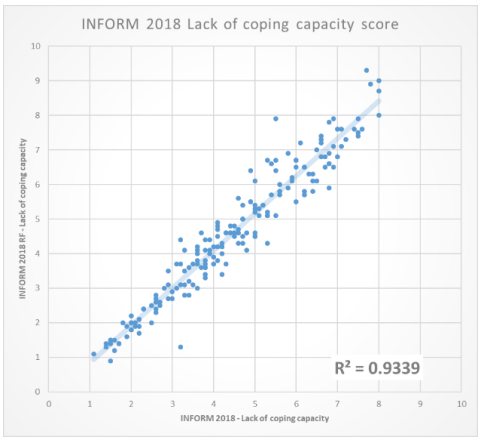
Pearson's correlation: 0.98

Spearman's correlation: 0.98

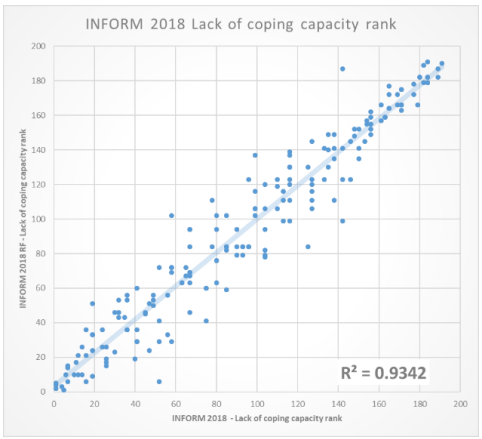
[3] The Pearson's correlation coefficient is a measure of a linear relationship between the scores of the two variables.

[4] The Spearman's correlation coefficient is a nonparametric measure of statistical dependence between two ranked variables.

Figure 8. Sensitivity analysis: No imputation (INFORM 2018) vs RF imputation (INFORM 2018 RF) – Lack of coping capacity



Pearson's correlation: 0.97



Spearman's correlation: 0.97

More evidence in the differences in results between the two imputation approaches can be find in the underlying components. For instance, the Pearson cc of the 'inequality' component in the vulnerability dimension is 0.78, and of the 'physical infrastructures' in the lack of coping capacity dimension is 0.93.

But is at the single country level that the differences are more relevant. There is a strong correlation between the countries with more missing values (**Table 3**) and the countries with larger variation in the results (**Table 11**).

Table 11. TOP 10 countries with large decreasing of Risk (left) after the RF imputation; TOP 10 countries with large increasing of Risk (right) after the RF imputation.

Country	INFORM 2018 RF	Delta
Eritrea	5.5	-1.2
Korea DPR	5.1	-1
Myanmar	6.4	-0.9
South Sudan	9	-0.8
Chad	7.8	-0.7
Haiti	6.3	-0.7
Marshall Islands	4.4	-0.7
Somalia	9.1	-0.7
Tuvalu	4	-0.6
Equatorial Guinea	3.9	-0.5

Country	INFORM 2018 RF	Delta
Liechtenstein	1	0.6
Palestine	4.6	0.3
Vanuatu	3.9	0.3
Iceland	1	0.3
Saint Kitts and Nevis	1.5	0.3
East Timor	4.2	0.3
Kazakhstan	2.2	0.3
South Korea	1.6	0.3
Uzbekistan	3	0.3
Argentina	2.3	0.2

Countries with larger decreasing in score after the RF imputation are generally the ones with higher risk score (**Table 11**). We can generically say that the RF imputation reduce the range of the scores ([0.4-9.1] vs [0.4-8.4], with an average decrease of final score of 0.1).

6 Description of the proposed strategy: a multi-methods approach

'The main question a modeller has to face when dealing with imputation is which method he/she has to use to fill in empty data spaces. To the best of our knowledge there is no definitive answer to this question but a number of rules of thumb (and lot of common sense). The choice principally depends on the dataset available, the number of missing data as compared to the dimension of the dataset (few missing data in a large dataset do not probably require sophisticated imputation methods), and **the identity of the country and the indicator for which the data is missing**' (OECD, 2008).

Selecting proper methods to best treat missing data depends not only on the type of the variables, the missing data rate, the missing data pattern, and the missingness mechanism, but also on the context of the specific analysis (Song, 2007).

In line with the approach used in many others indexes (chapter 3.3), we propose a combination of various imputation methods (interpolation, regional average, most recent, zero-filled, regression, random forest, 'no-imputation'), according to the type of indicator, his missing data rate and pattern. This has been already in part implemented in the INFORM GRI methodology as pre-imputation process (De Groeve, 2014), but it will now be extended to others methods (especially for filling the gaps in the historical data and to update less recent values) and applied systematically to all the indicators.

As we saw in the paragraph 5.1, with pre-imputation process we were able to reduce the MV from 50% to only 11%. The intent is now to improve the pre-imputations process using more proper imputation methods in order to have better estimation and reduce the final number of MV.

Deleted:)

Finally, further investigation is needed on the type of approach to use for the countries that have a different MV mechanism due to their status (protracting crisis, Small Island Developing States (SIDS), dictatorships).

6.1 Imputation methods

6.1.1 No imputation

In the case of missing data due to weak coverage, the approach is to introduce more than one high correlated indicator for the same component so that the indicators complement each other. Whenever certain values are missing, the aggregation process behaves as a tool to compensate a deficit in one dimension/category/component by creating a surplus in another.

This is the approach widely used so far in the INFORM GRI for most of the indicators.

6.1.2 Hot-deck imputations

6.1.2.1 Most recent

When available, the last observation is carried forward as imputation for the missing value. This represents the belief that if a measurement is missing, the best guess is that it hasn't changed from the last time it was measured.

This imputation method is broadly used in the current version of INFORM GRI, with the exception of indicators with high yearly variation (i.e. refugees, IDPs, people affected by disasters).

6.1.2.2 Country look-up

The available indicator's value of a country in similar conditions (geographical, economical, humanitarian) is used as a proxy for the missing data of the given country. Expert judgment is used for identifying those countries' relation.

This method is currently used in the INFORM GRI for the imputation of the missing values of the "Average Dietary Energy Supply Adequacy" and "Prevalence of Undernourishment" indicators. The selection of the look-up countries has been suggested by the FAO experts.

6.1.2.3 Regional average

Weighted averages for a region, taking into account the relative size of the relevant population of each country. The figures for countries with larger populations thus have a proportionately greater influence on the regional aggregates.[5]

This method, in combination with the country look-up method, is currently used in the INFORM GRI for the imputation of the missing values of the "Average Dietary Energy Supply Adequacy" and "Prevalence of Undernourishment" indicators.

6.1.3 Expert judgment: Zero filled or fixed value

Missing data are replace by a fix number according to an expert interpretation. This method is used when there is a clear knowledge of the reason why a value is missed.

The Zero-filled method is applied to the "Humanitarian Aid (FTS)" (donor's countries not receive humanitarian aid), "Malaria incidence" (malaria incidence is not reported if the disease is not present in the country), "Internally displaced persons (IDPs)" (we assume that if they are not reported, there are not IDPs in the country).

6.1.4 Linear regression

It is used when we know there is a correlation between the missing value and other variables. Missing values are substituted by the predicted values obtained from regression. The dependent variable of the regression is the individual indicator hosting the missing value, and the regressor(s) is (are) the individual indicator(s), showing a strong relationship with the dependent variable, i.e. usually a high degree of correlation (OECD, 2008).

Due to a strong relationship of HDI with the GDP (PPP) per capita, missing values were imposed with the predicted value of HDI based on the known GDP (PPP) per capita for specific countries obtained from regression analysis executed on the rest of the set (De Groeve, 2014).

6.1.5 Linear interpolation

Interpolation is a mathematical method that adjusts a function to your data and uses this function to extrapolate the missing data. The most simple type of interpolation is the linear interpolation, that makes a mean between the values before the missing data and the value after. This method works only when there are enough observation in the time-series and is particular efficient when the trend data pattern is linear (structural indicators).

6.1.6 Random Forest regression

See paragraph 5.2.

6.1.7 Others

Others imputation methods have been suggested in particular for composite indicators. These include k Nearest Neighbour (k-NN), Expectation-Maximization (EM). They will be tested as possible alternative for RFR.

6.2 Imputation strategy

The overview of the proposed imputation strategy for each indicator is finally summarised in **Table 12**. A more exhaustive description of the imputation strategy for each indicators is presented in the Annex 1 with the indicator fact-sheets.

[5] http://uis.unesco.org/en/glossary-term/regional-average

Table 12. Overview of MV imputation proposed for the future INFORM GRI

Indicator	Current Pre-imputation method	Proposed imputation method
Physical exposure to earthquakes of MMI VI	None	None
Physical exposure to earthquakes of MMI VIII	None	None
Physical exposure to flood	Fixed value	RFR or alternative
Physical exposure to tropical cyclone of Saffir-Simpson 1	None	None
Physical exposure to tropical cyclone of Saffir-Simpson 3	None	None
Physical exposure to surge from tropical cyclone	None	None
Physical exposure to tsunamis	None	None
Drought affected	None	None
Drought Frequency	None	None
Agriculture Stress Index Probability	None	None
Population density (people per sq. km of land area)	Most recent	RFR or alternative
Population living in slums (% of urban population)	Most recent	RFR or alternative
IHR capacity score: Food safety	None	None
Proportion of population with basic handwashing facilities on premises	Most recent	RFR or alternative
People practicing open defecation (% of population)	Most recent	RFR or alternative
Population at Risk to Aedes	None	None
Population exposed to Dengue	None	None
Populations at risk of PF malaria in 2010 - Stable transmission	None	None
Populations at risk of PF malaria in 2010 - Unstable transmission	None	None
Populations at risk of PV malaria in 2010 - Stable transmission	None	None
Populations at risk of PV malaria in 2010 - Unstable transmission	None	None
Population exposed to Zika	None	None
Population exposed to CCHF	None	None
Population exposed to EDV	None	None
Population exposed to Lassa Fever	None	None
Population exposed to MVD	None	None
Household size	Most recent	RFR or alternative
Children under 5 (% of population)	None	RFR or alternative
Urban population growth (annual %)	Most recent	RFR or alternative
Population living in urban areas (%)	Most recent	RFR or alternative
Number of vets	Most recent	RFR or alternative
Conflict HIIK National Power	None	None
Conflict HIIK Subnational	None	None

Indicator	Current Pre-imputation method	Proposed imputation method
GCRI Highly Violent Internal Conflict probability	None	None
GCRI Violent Internal Conflict probability	None	None
Human Development Index	Linear regression	Linear regression or RFR
Multidimensional Poverty Index	Most recent	RFR or alternative
Income Gini coefficient - Inequality in income or consumption	Most recent	RFR or alternative
Gender Inequality Index	Most recent	RFR or alternative
Volume of remittances (in USD) as a proportion of total GDP (%)	Most recent	RFR or alternative
Net ODA received (% of GNI)	Most recent	RFR or alternative
International humanitarian aid by the Financial Tracking Service	Zero-filled	Zero-filled
Official Development Assistance	Zero-filled	Zero-filled
Internally Displaced People	Zero-filled	Zero-filled
Refugees and asylum-seekers by country of asylum	Zero-filled	Zero-filled
Refugees (Palestine)	Zero-filled	Zero-filled
Returned refugees	Zero-filled	Zero-filled
Average dietary supply adequacy	Regional avg	Regional avg or RFR
Prevalence of undernourishment (% of population)	Regional avg	Regional avg or RFR
Estimated number of people living with HIV - Adult (>15) rate	Most recent	RFR or alternative
Number of new HIV infections per 1,000 uninfected population	Most recent	RFR or alternative
Malaria incidence per 1,000 population at risk	Most recent	RFR or alternative
Tuberculosis prevalence	Most recent	RFR or alternative
Number of people requiring interventions against neglected tropical diseases	Most recent	RFR or alternative
Children Under Weight	Most recent	RFR or alternative
Child Mortality	Most recent	RFR or alternative
People affected by Natural Disasters	None	None
Hyogo Framework for Action	Most recent	RFR or alternative
Corruption Perception Index	Most recent	RFR or alternative
Government Effectiveness	Most recent	RFR or alternative
Literacy rate, adult total (% of people ages 15 and above)	Most recent	RFR or alternative
Access to electricity	Most recent	RFR or alternative
Mobile cellular subscriptions (per 100 people)	Most recent	RFR or alternative
Internet users (per 100 people)	Most recent	RFR or alternative
People using at least basic drinking water services (% of population)	Most recent	RFR or alternative

Indicator	Current Pre-imputation method	Proposed imputation method
People using at least basic sanitation services (% of population)	Most recent	RFR or alternative
Roads, total network (km)	None	RFR or alternative
Proportion of the target population with access to 3 doses of diphtheria-tetanus-pertussis (DTP3) (%)	Most recent	RFR or alternative
Proportion of the target population with access to measles-containing-vaccine second-dose (MCV2) (%)	Most recent	RFR or alternative
Proportion of the target population with access to pneumococcal conjugate 3rd dose (PCV3) (%)	Most recent	RFR or alternative
Physicians Density	Most recent	RFR or alternative
Maternal Mortality	Most recent	RFR or alternative
Current health expenditure per capita	Most recent	RFR or alternative

6.3 Country outliers

Regardless of the type of indicator, a group of countries present a systematic lack of observed data (paragraph 4.1). This is due to the characteristics of the countries, being structural (small countries, geographical location), or temporary (protracting crisis, dictatorship).

It not only the large number of missing data, but also the time series patterns that makes those countries to be considered as outliers.

Countries in protracting crisis (e.g. Syria, Libya, Somalia, South Sudan) don't have the capacity to produce most of the vulnerability and coping capacity indicators, while only the strictly humanitarian indicators are available (refugees, IDP, humanitarian aid). In addition, the trend of the missing indicators are normally in contrast with the global trend as consequence of the crisis (e.g. the health system performance deteriorate, all the economic indicators present a negative trend, the malnutrition increase).

Autocratic regimes (e.g. Democratic Republic of Korea, Eritrea) are normally not willing to share the information about their country, therefore the availability of indicators is very weak. The real internal situation is large unknown, as well as the indicators that should help to describe it.

Small Island Developing States (SIDS) (i.e. Pacific islands, Caribbean), and in general small and unpopulated countries (e.g. Liechtenstein), share a lack of statistical capacity. Furthermore, earth observation derived indicators (e.g. Agricultural Stress Index) are often not available for such as small areas. A further challenge is that for a large set of indicators depending on the size of population or area, the patterns for those countries are substantially different compared to the rest of the countries.

Which data is normally available and what is missing for the countries outliers? In general data from Hazard & Exposure dimension, uprooted people (refugees, IDPs, returnees), and all the indicators not influenced by the crisis (malaria incident, recent shocks), are available. While are larger missing all the Vulnerability and Lack of Coping Capacity indicators impacted by the crisis and the ones reported by the national statistical offices (demographic, socio-economic, health, infrastructures).

For the country outliers the imputation methods don't perform well (paragraph 5.2). Traditional imputation methods based their predictions on the observed values in the time series of the same indicators (e.g. interpolation) or of the other indicators with high correlation (e.g. regression). But as we saw, the country outliers behave differently than the rest of the countries.

What we can do, then? Many possible options are available, but none of them is optimal:

- **Use proxy indicators**: indicators like GDP per capita or HDI have a good correlation with several of the indicators in the Vulnerability and Lack of Coping Capacity dimensions, and they could use in a regression for a broad estimation of the missing values. On the other hand, the same indicators are normally not available (or not reliable) for the countries in protracting crisis and for countries with autocratic regimes.
- **Use any other available information for the same country**: the idea is to use only observed data for the country to estimate what is missing. The cluster approach applied to the RFR (paragraph 5.2) goes to this direction.
- **Status quo**: use the pre-crisis data as proxies. This is what is currently done in INFORM GRI.
- **Remove the outliers**: as reliable data don't exist, we could decide to remove the country outliers from the INFORM GRI. In particular, assessing the risk for countries already in protracting crisis is not meaningful. INFORM has another tool for following the ongoing crisis (INFORM Severity), which should be used instead.

7 Conclusions

This report on imputation of missing values in the INFORM GRI model presents the follow up of the work started in 2017 with the experimental use of machine learning approach using the Random Forest algorithm for the prediction of missing data.

In the presented study we focused on better understanding the patterns and mechanisms of missing values in the INFORM GRI model, and on evaluating their impact on the model's outputs. The scope was to develop a missing data imputation strategy to be implemented in the INFORM GRI.

The main outcomes from the missing values analysis were that:

- The missing values are distributed very differently among the indicators, the countries and the time.
- A simple pre-imputation process is able to reduce the percentage of missing values from 50% to 11%.
- The main challenges are data updating, trend consistency and countries outliers.
- The influence of missing values is minimal on the final aggregated results (Risk score), while it increases in the underlying components and sub-components.
- A strategy based on a single imputation method seems not the ideal approach for dealing with the variety of the indicators used in INFORM GRI. Other similar indexes adopt a combination of different imputation methods according to the type of indicators. The criteria are normally based on expert knowledge of the characteristics of the single indicator.

Further improvements need to be done in the implementation of the Random Forest Regression, namely the clustering approach, as well as testing alternative methods.

As result of the analysis, we propose an imputation strategy based on multiple methods. The presented imputation strategy will be implemented by the next release of INFORM GRI on September 2020.

However, statistical imputation cannot solved all the missing data issues (country outliers, special cases). Often only the experts could give a plausible estimation of what is missing. The INFORM platform could be used for the collaborative collection of the expert opinions from the INFORM partners, and allow an integration with the INFORM data and results, although it might raise problems of confidentiality.

References

Damioli, G., 'Step 3: The identification and treatment of outliers', *05-09 November 2018 - 16th JRC Annual Training on Composite Indicators & Scoreboards*, 2018, JRC-COIN: https://composite-indicators.jrc.ec.europa.eu/sites/default/files/COIN%202018%20Step%203%20Outliers%20and%20Missing%20data.pdf

De Groeve, T., Poljansek, K., and Vernaccini, L., *Index for Risk Management - INFORM: Concept and Methodology*. Luxembourg: Publications Office of the European Union, 2014, doi:10.2788/78658.

Dondersa, et al., 'Review: A gentle introduction to imputation of missing values', *Journal of Clinical Epidemiology*, Vol. 59, 2006, pp. 1087-1091.

Kang H., 'The prevention and handling of the missing data', *Korean journal of anesthesiology*, Vol. *64,* No 5, 2013, pp. 402–406. doi:10.4097/kjae.2013.64.5.402

Marin-Ferrer, M., Doherty, B., Bejar Garcia, J., Luoni, S., and Vernaccini, L., *INFORM: Scientific and technical improvements in 2017: Missing values imputation and IT developments*, Publications Office of the European Union, Luxembourg, 2017a, doi:10.2760/076136.

Marin-Ferrer, M., Vernaccini, L., and Poljanšek, K., *Inform Index for Risk Management: Concept and methodology version 2017*, Publications Office of the European Union, Luxembourg, 2017b, doi:10.2760/094023.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., and Giovannini, E., *Handbook on constructing composite indicators: Methodology and user guide*, OECD Statistics Working Papers No 2005/03, OECD Publishing, Paris, 2005, doi:10.1787/533411815016.

OECD, *Handbook on constructing composite indicators*, OECD Publishing, Paris, 2008.

Paruolo, P., Saisana, M., and Saltelli A., 'Ratings and rankings: voodoo or science?', *Journal of the Royal Statistical Society A*, Vol. 176, No 3, 2013, pp. 609-634.

Pigott, M., Deshpande, A. Letourneau, I., Morozoff, C., Reiner, R., Kraemer, M., Brent, S., Bogoch, I, Khan, K., Biehl, M., Burstein, R., Earl, L., Fullman, N., Messina, J., Mylne, A., Moyes, C., Shearer, F., Bhatt, S., Brady, O., Gething, P., Weiss, D., Tatem, A., Caley, L., De Groeve, T., Vernaccini, L., Golding, N., Horby, P., Kuhn, J., Laney, S., Peter Piot, E., Sankoh, O., Murray, C., Hay, S., *Local, national, and regional viral haemorrhagic fever pandemic potential in Africa: a multistage analysis*, The Lancet, Vol. 390, Issue 10113, 2017, pp. 2662-2672, ISSN 0140-6736, doi:10.1016/S0140-6736(17)32092-5

Song, Q., and Shepperd, M., *Missing Data Imputation Techniques,* IJBIDM 2, pp. 261-291, 2017, doi:10.1504/IJBIDM.2007.015485.

Tang, F., and Ishwaran, H., 'Random Forest Missing Data Algorithms.' *Statistical analysis and data mining,* Vol. 10, No 6, 2017, pp. 363-377. doi:10.1002/sam.11348

Xie, Y., *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman, Hall/CRC, 2015, http://yihui.name/knitr/.

List of figures

List of tables

Annexes

Annex 1. Overview of the missing values imputation strategy by INFORM GRI indicators

Indicator	Missing values pattern					Indicator pattern	Imputation of MV	
Name	MV after pre-imputation	Reason for country's missingness	Time-series [2009-2018]	% MV [2009-2018]	Country coverage: No values in the time-series	Time series pattern [2009-2018]	Current pre-imputation method	Proposed imputation method
Hazards & Exposure								
Natural								
Earthquake								
Physical exposure to earthquakes of MMI VI	No		No	90%	0	Constant	None	None
Physical exposure to earthquakes of MMI VIII	No		No	90%	0	Constant	None	None
Flood								
Physical exposure to flood	Yes	Depends on the country size (small countries)	No	92%	32	Constant	Fixed value	RFR or alternative
Tropical cyclone								
Physical exposure to tropical cyclone of Saffir-Simpson 1	No		No	90%	0	Constant	None	None
Physical exposure to tropical cyclone of Saffir-Simpson 3	No		No	90%	0	Constant	None	None
Physical exposure to surge from tropical cyclone	No		No	90%	0	Constant	None	None
Tsunami								
Physical exposure to tsunamis	No		No	90%	0	Constant	None	None
Drought								
Drought affected	No		No	90%	0	Constant	None	None
Drought Frequency	No		No	90%	0	Constant	None	None

Indicator	Missing values pattern					Indicator pattern	Imputation of MV	
Name	**MV after pre-imputation**	**Reason for country's missingness**	**Time-series [2009-2018]**	**% MV [2009-2018]**	**Country coverage: No values in the time-series**	**Time series pattern [2009-2018]**	**Current pre-imputation method**	**Proposed imputation method**
Agriculture Stress Index Probability	Yes	Depends on the country size (small countries)	Yes	91%	13	Constant	None	None
Epidemic								
Population density (people per sq. km of land area)	No		Yes	1%	0	Trend	Most recent	RFR or alternative
Population living in slums (% of urban population)	Yes	Countries without slums + other countries	Yes	86%	68	Trend	Most recent	RFR or alternative
IHR capacity score: Food safety	Yes	Systematically missing for subset of countries	Yes	91%	15	Trend	None	None
Proportion of population with basic handwashing facilities on premises (% of population)	Yes		Yes	62%	96	Trend	Most recent	RFR or alternative
People practicing open defecation (% of population)	Yes	Random missing	Yes	12%	1	Trend	Most recent	RFR or alternative
Population at Risk to Aedes	No		No	90%	0	Constant	None	None
Population exposed to Dengue	No		No	90%	0	Constant	None	None
Populations at risk of Plasmodium falciparum malaria in 2010 - Stable transmission	No		No	90%	0	Constant	None	None

Indicator	Missing values pattern					Indicator pattern	Imputation of MV	
Name	MV after pre-imputation	Reason for country's missingness	Time-series [2009-2018]	% MV [2009-2018]	Country coverage: No values in the time-series	Time series pattern [2009-2018]	Current pre-imputation method	Proposed imputation method
Populations at risk of Plasmodium falciparum malaria in 2010 - Unstable transmission	No		No	90%	0	Constant	None	None
Populations at risk of Plasmodium vivax malaria in 2010 - Stable transmission	No		No	90%	0	Constant	None	None
Populations at risk of Plasmodium vivax malaria in 2010 - Unstable transmission	No		No	90%	0	Constant	None	None
Population exposed to Zika	No		No	90%	0	Constant	None	None
Population exposed to CCHF	Yes		No	92%	44	Constant	None	None
Population exposed to EDV	Yes	Data available only in Africa	No	97%	139	Constant	None	None
Population exposed to Lassa Fever	Yes	Data available only in Africa	No	97%	139	Constant	None	None
Population exposed to MVD	Yes	Data available only in Africa	No	97%	139	Constant	None	None
Household size	Yes	Random missing	Yes	89%	78	Trend	Most recent	RFR or alternative
Children under 5 (% of population)	Yes	Systematically missing for subset of countries	Yes	4%	7	Trend	None	RFR or alternative
Urban population growth (annual %)	Yes	Systematically missing for subset of countries	Yes	1%	1	Trend	Most recent	RFR or alternative

Indicator	Missing values pattern					Indicator pattern	Imputation of MV	
Name	**MV after pre-imputation**	**Reason for country's missingness**	**Time-series [2009-2018]**	**% MV [2009-2018]**	**Country coverage: No values in the time-series**	**Time series pattern [2009-2018]**	**Current pre-imputation method**	**Proposed imputation method**
Population living in urban areas (%)	Yes	Systematically missing for subset of countries	Yes	1%	1	Trend	Most recent	RFR or alternative
Number of vets	Yes	Random missing	Yes	66%	23	Trend	Most recent	RFR or alternative
Human								
Conflict risk								
Conflict HIIK National Power	No		Yes	0%	0	Random	None	None
Conflict HIIK Subnational	No		Yes	0%	0	Random	None	None
GCRI Highly Violent Internal Conflict probability	No		Yes	0%	0	Trend	None	None
GCRI Violent Internal Conflict probability	No		Yes	0%	0	Trend	None	None
Vulnerability								
Social-Economics Vulnerability								
Poverty & Development								
Human Development Index	Yes	Systematically missing for subset of countries	Yes	13%	5	Trend	Linear regression	Linear regression or RFR
Multidimensional Poverty Index	Yes	Developed countries + others	Yes	93%	89	Trend	Most recent	RFR or alternative
Inequality								

Indicator	Missing values pattern					Indicator pattern	Imputation of MV	
Name	**MV after pre-imputation**	**Reason for country's missingness**	**Time-series [2009-2018]**	**% MV [2009-2018]**	**Country coverage: No values in the time-series**	**Time series pattern [2009-2018]**	**Current pre-imputation method**	**Proposed imputation method**
Income Gini coefficient - Inequality in income or consumption	Yes	Random missing	Yes	69%	30	Trend	Most recent	RFR or alternative
Gender Inequality Index	Yes	Systematically missing for subset of countries	Yes	47%	31	Trend	Most recent	RFR or alternative
Economical Dependency								
Volume of remittances (in USD) as a proportion of total GDP (%)	Yes	Systematically missing for subset of countries	Yes	11%	17	Trend	Most recent	RFR or alternative
Net ODA received (% of GNI)	Yes	Systematically missing for subset of countries	Yes	37%	48	Random	Most recent	RFR or alternative
International humanitarian aid by the Financial Tracking Service	No		Yes	0%	0	Random	Zero-filled	Zero-filled
Official Development Assistance	No		Yes	35%	0	Random	Zero-filled	Zero-filled
Vulnerable Groups								
Uprooted people								
Internally Displaced People	Yes		Yes	73%	130	Random	Zero-filled	Zero-filled
Refugees and asylum-seekers by country of asylum	No		Yes	8%	0	Random	Zero-filled	Zero-filled
Refugees (Palestine)	No		Yes	98%	0	Random	Zero-filled	Zero-filled
Returned refugees	No		Yes	8%	0	Random	Zero-filled	Zero-filled

Indicator	Missing values pattern					Indicator pattern	Imputation of MV	
Name	**MV after pre-imputation**	**Reason for country's missingness**	**Time-series [2009-2018]**	**% MV [2009-2018]**	**Country coverage: No values in the time-series**	**Time series pattern [2009-2018]**	**Current pre-imputation method**	**Proposed imputation method**
Food Security								
Average dietary supply adequacy	Yes	Systematically missing for subset of countries	Yes	9%	17	Trend	Regional avg	Regional avg or RFR
Prevalence of undernourishment (% of population)	Yes	Systematically missing for subset of countries	Yes	16%	30	Trend	Regional avg	Regional avg or RFR
Health conditions								
Estimated number of people living with HIV - Adult (>15) rate	Yes	Random missing	Yes	45%	38	Trend	Most recent	RFR or alternative
Number of new HIV infections per 1,000 uninfected population	Yes	Random missing	Yes	41%	64	Trend	Most recent	RFR or alternative
Malaria incidence per 1,000 population at risk	Yes	Countries without malaria + other countries	Yes	55%	84	Trend	Most recent	RFR or alternative
Tuberculosis prevalence	Yes	Systematically missing for subset of countries	Yes	11%	1	Trend	Most recent	RFR or alternative
Number of people requiring interventions against neglected tropical diseases	Yes		Yes	21%	2	Trend	Most recent	RFR or alternative
Health of children under 5								
Children Under Weight	Yes	Random missing	Yes	90%	43	Trend	Most recent	RFR or alternative

Indicator	Missing values pattern					Indicator pattern	Imputation of MV	
Name	**MV after pre-imputation**	**Reason for country's missingness**	**Time-series [2009-2018]**	**% MV [2009-2018]**	**Country coverage: No values in the time-series**	**Time series pattern [2009-2018]**	**Current pre-imputation method**	**Proposed imputation method**
Child Mortality	Yes	Systematically missing for subset of countries	Yes	11%	1	Trend	Most recent	RFR or alternative
Recent shocks								
People affected by Natural Disasters	No		Yes	0%	0	Random	None	None
Lack of Coping Capacity								
Institutional								
DRR								
Hyogo Framework for Action	Yes	Systematically missing for subset of countries	Yes	69%	40	Trend	Most recent	RFR or alternative
Governance								
Corruption Perception Index	Yes	Systematically missing for subset of countries	Yes	9%	11	Trend	Most recent	RFR or alternative
Government Effectiveness	No		Yes	10%	0	Trend	Most recent	RFR or alternative
Infrastructure								
Communication								
Literacy rate, adult total (% of people ages 15 and above)	Yes	Random missing	Yes	77%	38	Trend	Most recent	RFR or alternative

Indicator	Missing values pattern					Indicator pattern	Imputation of MV	
Name	MV after pre-imputation	Reason for country's missingness	Time-series [2009-2018]	% MV [2009-2018]	Country coverage: No values in the time-series	Time series pattern [2009-2018]	Current pre-imputation method	Proposed imputation method
Access to electricity	No		Yes	10%	0	Trend	Most recent	RFR or alternative
Mobile cellular subscriptions (per 100 people)	No		Yes	11%	0	Trend	Most recent	RFR or alternative
Internet users (per 100 people)	Yes	Systematically missing for subset of countries	Yes	22%	2	Trend	Most recent	RFR or alternative
Physical Connectivity								
People using at least basic drinking water services (% of population)	No		Yes	11%	0	Trend	Most recent	RFR or alternative
People using at least basic sanitation services (% of population)	No		Yes	11%	0	Trend	Most recent	RFR or alternative
Roads, total network (km)	No		No	90%	0	Constant	None	RFR or alternative
Access to health care								
Proportion of the target population with access to 3 doses of diphtheria-tetanus-pertussis (DTP3) (%)	Yes	Systematically missing for subset of countries	Yes	11%	1	Trend	Most recent	RFR or alternative
Proportion of the target population with access to measles-containing-vaccine second-dose (MCV2) (%)	Yes	Systematically missing for subset of countries	Yes	35%	31	Trend	Most recent	RFR or alternative
Proportion of the target population with access to pneumococcal conjugate 3rd dose (PCV3) (%)	Yes	Systematically missing for subset of countries	Yes	59%	59	Trend	Most recent	RFR or alternative

Indicator	Missing values pattern					Indicator pattern	Imputation of MV	
Name	**MV after pre-imputation**	**Reason for country's missingness**	**Time-series [2009-2018]**	**% MV [2009-2018]**	**Country coverage: No values in the time-series**	**Time series pattern [2009-2018]**	**Current pre-imputation method**	**Proposed imputation method**
Physicians Density	Yes	Random missing	Yes	55%	6	Trend	Most recent	RFR or alternative
Maternal Mortality	Yes	Systematically missing for subset of countries	Yes	13%	9	Trend	Most recent	RFR or alternative
Current health expenditure per capita	Yes	Systematically missing for subset of countries	Yes	23%	4	Trend	Most recent	RFR or alternative

Annex 2. Example of other indicators

Social Progress Index – SPI [6]

The authors ensure that all indicators included in the Social Progress Index are missing as few observations as possible to avoid jeopardizing the statistical quality of the index. Missing values can stem from lack of coverage by the data source, incomplete reporting by the country to international organizations, or outdated data whose publication date is older than ten years. In cases where an indicator is missing a country data point, they assess their imputation methodology both before and during index calculation. They do not publish most imputations, other than those we generate for gaps between years in Access to Basic Knowledge, mentioned directly below.

Imputations prior to calculation

Missing data are imputed prior to calculation solely for Access to Basic Knowledge. These imputations are included in the published data, as they rely either on historical data from the same source or supplemental research. Data in this component are notoriously lacking, and we find this pre-calculation imputation step imperative to be able to include key countries in Social Progress Index rankings. For adult literacy rate and primary and secondary enrollment rates, the authors impute gaps between years to ensure smooth year-to-year estimates based on current and historical data and assuming linear change. In cases where there were data in recent years, but not for all years, they rely on data older than ten years (if available) to create **linear estimations** for the years in between. There are also several countries for which they impute adult literacy with 99%, based on **qualitative research**. These countries include: Australia, Austria, Belgium, Canada, Switzerland, Czech Republic, Germany, Denmark, Finland, France, Hungary, Ireland, Iceland, Japan, Luxembourg, Netherlands, Poland, Slovakia, Slovenia, Sweden, and United States.

Imputations during calculation

For countries for which there is no more than one missing data point per component in each of the twelve components (considered 'ranked countries') and for countries that have no more than one missing indicator data point in nine to eleven components (considered 'partial countries'), the authors use **regression imputation** to predict missing values during the calculation process. They use their country sample data of ranked and partial countries (including both current and historical Social Progress Index years, i.e. 2014-2018) to regress each indicator on the other indicators within a component. By constraining the regression to within component indicators, they can preserve the signal that the indicator provides to PCA. They review each imputation to ensure accuracy. In cases where imputations do not match expectations or qualitative research, they use **regional cohort estimates** or carry values consistently across time to minimize bias. For example, for many Middle Eastern countries where Gallup does not ask its survey question on gays and lesbians due to cultural sensitivities, they average the data values of countries within the regional groupings set by the Human Rights Campaign and based on LGBT criminalization laws. For indicators where imputations reflect substantial improvement or decline over time at the country level, they average the imputations across years. They applied this method to access to independent media, where imputations showed high gains for countries that were missing data for all years of the Social Progress Index. The estimation of missing values is necessary prior to undertaking PCA, which requires a complete dataset for the results to be sound. They do not impute values for countries that do not meet the criteria of ranked or partial countries noted above these countries are excluded from the main calculation process by which PCA weights are determined.

Global Peace Index – GPI [7]

Use a combination of **manual imputation by expert**, **most recent data** and **alternative sources**. Indicators are categorized in 5 classes.

Example: "Alternative Source: EIU. Where data is not provided, the EIU's analysts have filled them based on likely scores from the set bands of the actual data."

Ocean Health Index – OHI [8]

[6] https://www.socialprogress.org/assets/downloads/resources/2018/2018-Social-Progress-Index-Methodology.pdf
[7] http://visionofhumanity.org/app/uploads/2018/06/Global-Peace-Index-2018-2.pdf
[8] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4970671/bin/pone.0160377.s005.docx

First imputation with most recent data. Use ad hoc method for each indicator – ex. Zero-filled, regional average, regression.

The authors published a complete review of the impact of missing values and their gapfilling on the OHI results: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4970671/

States of Fragility – OECD [9]

Data availability is a key issue in calculating the OECD fragility framework. As the unit of analysis is the state or territory, it is important to select indicators that are comparable across those states and territories. While statistical imputation methods can be used to fill data gaps, such an approach is best used sparingly; preference should always be given to real-world data, even if it means dropping indicators or countries and territories that otherwise would have been included. The fragility framework methodology aims to strike a balance between the number of indicators, the contexts covered and the amount of imputation that would be required to build a complete data set. A criterion for inclusion in the OECD framework was at least 70% of the required data had to be available for a country or context. As a result, only 172 contexts could be included in the calculations.

The 43 indicators selected do not cover all contexts and imputation techniques have been used to fill in data gaps. Lack of data is the primary reason why a context may not be included. At least 70% of data for a context had to be available for it to be included in the OECD fragility framework. In 2018, this yields a list of 172 contexts. It is possible to assume that contexts missing from the dataset have a certain value for some indicators. For example, those missing from the datasets for battle deaths and deaths by non-state actors can be assumed to have a value of 0. Where no reasonable assumption could be made, data are imputed using **k-nearest neighbour (KNN)** imputation that uses statistical inference to fill in missing data from the k most similar contexts. In the OECD fragility framework, this has been done using the 15 most similar contexts for each missing data point.

Global Innovation Index GII – WIPO [10]

The GII developing team, for transparency and replicability, has always opted not to estimate missing data. The **'no imputation'** choice, which is common in similar contexts, might encourage economies not to report low data values. Yet this is not the case for the GII. After 11 editions of the GII, the index-developing team has not encountered any intentional noreporting strategy. The consequence of the 'no imputation' choice in an arithmetic average is that it is equivalent to replacing an indicator's missing value for a given country with the respective sub-pillar score. Hence the available data (indicators) in the incomplete pillar may dominate, sometimes biasing the ranks up or down.

9 https://www.oecd-ilibrary.org/docserver/9789264302075-en.pdf?expires=1561470472&id=id&accname=oid031827&checksum=DA4E777264842A28E2BF188D67B082DB

10 https://composite-indicators.jrc.ec.europa.eu/sites/default/files/JRC%20Statistical%20Audit%20of%20the%202018%20Global%20Innovation%20Index.pdf

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-30037-EN-C

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:10.2760/717316

ISBN 978-92-76-14725-1