2024 Ibrahim Index of African Governance

—

Methodology

OCTOBER 2024
MO IBRAHIM FOUNDATION

CONTENTS

INTRODUCTION TO THE IIAG

The Ibrahim Index of African Governance (IIAG) provides a biannual statistical assessment of governance performance in 54 African countries.

Data from diverse sources are combined into composite governance scores – an *Overall Governance* score, categories, sub-categories, clustered indicators, as well as clustered sub-indicators. The 2024 IIAG provides comparable data for the ten-year time period 2014-2023 enabling analysis of national and regional trends over time.

This document contains a detailed explanation of how the IIAG is calculated. The methodology is simple and transparent. All of the underlying data and information used in the construction of the IIAG are freely available and transparently published by the Mo Ibrahim Foundation (MIF), alongside the scores and uncertainty estimates. Given the impossibility to observe directly the nature of governance, all measures are made with error and users of the Index are encouraged to use the confidence intervals when interpreting the scores. This document will outline how MIF conceptualises governance, assumptions made, how variables are selected for inclusion, how the IIAG is calculated and how we estimate uncertainty around the scores.

MEASURING GOVERNANCE

MIF defines governance as the provision of the political, social, economic and environmental goods and services that every citizen has the right to expect from their state, and that a state has the responsibility to deliver to its citizens. The IIAG is concerned with operationalising these public goods mostly through outputs and outcomes of policy, as opposed to inputs or de jure indicators. Although the latter are selected in occasions when suitable output or outcome variables are not available. The framework of the IIAG has been constructed to reflect this definition and consists of four main pillars of governance: *Security & Rule of Law*, *Participation, Rights & Inclusion*, *Foundations for Economic Opportunity* and *Human Development*. These categories are comprised of 16 sub-categories. The structure of the 2024 IIAG (to sub-category level) can be seen in Table 1 below.

Table 1. 2024 IIAG framework (to sub-category level)

OVERALL GOVERNANCE			
SECURITY & RULE OF LAW	PARTICIPATION, RIGHTS & INCLUSION	FOUNDATIONS FOR ECONOMIC OPPORTUNITY	HUMAN DEVELOPMENT
Security & Safety	Participation	Public Administration	Health
Rule of Law & Justice	Rights	Business & Labour Environment	Education
Accountability & Transparency	Inclusion & Equality	Infrastructure	Social Protection & Welfare
Anti-Corruption	Women's Equality	Rural Economy	Sustainable Environment

Each construct is operationalised as a composite indicator of its sub-components. The *Overall Governance* score is constructed by calculating an unweighted average of its underlying four categories; these categories are constructed by calculating an unweighted average of their respective sub-categories. Likewise, sub-category scores are the result of aggregating the scores of their underlying indicators. There are 96 indicators in the 2024 IIAG, 90 of which are clustered and 6 are standalone indicators. Indicators are formed by variables. A variable is anything that can be constituted as a raw data from source. There are 322 variables collected to calculate the 2024 IIAG. In total, the 2024 Index contains 492 measures of governance (taking into account the variables collected from source as well as every composite score provided in the IIAG dataset). The data comes from 49 data sources, a mix of quantitative and qualitative assessments.

Following an internal review of the IIAG framework as well as advice from our IIAG Expert Panel, each sub-category now includes one indicator composed of public perception measures sourced from Afrobarometer. These were previously available within a standalone public perception dataset named *Citizens' Voices*. Including these into the IIAG framework allows for an additional level of interpretation and reflects elements that may not be captured in the expert assessment and official data included in the IIAG.

TECHNICAL ASSUMPTIONS

The central assumption is that each indicator score is the true value of its respective sub-category score plus or minus some error. Each constituent indicator measures something specific – for example *Education Enrolment*- but when included in a sub-category it is expected to measure a broader concept; in this case whether a country has a comprehensive education system. In some countries which have high education enrolment rates, completion rates may be a lot lower because of the lack of adequate support systems for students at risk of dropping out, and therefore enrolment rates may overestimate the adequacy of education provided in a country. It is this misestimation which is captured in the error term, as discussed further in the section on uncertainty estimates.

The IIAG makes the assumption of independence across indicator errors. This is illustrated in the *Civil Registration* and *Budgetary & Financial Management* indicators from the *Public Administration* sub-category. Both indicators misestimate *Public Administration* by some amount, but because the indicators come from very different sources – in the case of *Civil Registration*: the Africa Integrity Indicators from the African Institute for Development Policy, in the case of *Budgetary & Financial Management*: the Country Policy and Institutional Assessments from the African Development Bank and the World Bank- there is no reason to suppose that when one indicator misestimates *Public Administration* the other indicator will do the same. It is assumed that the errors are not related and will cancel each other out when the indicators are aggregated. The more indicators are added the more likely it is that the errors sum to zero. Any correlation between indicators arises as a result of a latent governance dimension. This misestimation is elaborated in greater detail in the section on uncertainty estimates.

VARIABLE SELECTION & CRITERIA FOR INCLUSION

To be included in the IIAG, a variable has to be linked to MIF's definition of governance and should preferably measure outputs and outcomes of governance, not inputs or de jure measures. For example, to measure the level of education in a country, we use the indicator *Education Quality*, as opposed to expenditures made by a government on education. Further considerations around inclusion of a variable include its methodological soundness, timeliness and accessibility, as well as the credibility of the data provider.

For each dataset under consideration, missing data is a large issue. In particular, the number of African countries covered by each dataset in a certain year; the number of years covered; the periodicity of the data and the most recent year of published data are all parameters which impact the precision of the composite scores.

Furthermore, comparability across countries and over time is fundamental. Data should be based on common concepts, definitions, classifications, and methodology. For this reason, data collected nationally (e.g. by a government agency) which are not internationally comparable cannot be used.

Given these considerations, **to be included in the IIAG, a variable must have at least two years' worth of data since the beginning of the time series (2014 in the case of the 2024 IIAG dataset) for at least 33 countries and the latest data point for these 33 countries must exist within the last three years (2021-2023 in the case of the 2024 IIAG dataset)**. Furthermore, in order to differentiate between scores, numerical granularity is taken into consideration, with all the variables selected being on a four-point scale or more.

When a variable is deemed to be suitable for inclusion, it is assigned to the sub-category in which it sits best conceptually. As dimensions of governance are not independent and variables may be deemed to be suitable for multiple sub-categories, this process goes through an additional stage of consultation with the IIAG Expert Panel, a group of expert advisors.

Following advice from the European Commission's Competence Centre on Composite Indicators and Scoreboards (CC-COIN), some additional guiding principles when selecting variables for the 2024 IIAG framework were the following:

- Missingness: The percentage of values missing from a variable should not overcome 20% in a year. Minimising the amount of missing values leads to a lower proportion of imputed data points and helps ensure that country sub-category scores are not dependent on a single variable.
- Correlation structure: The level of correlation for every variable/indicator should be higher than 0.3 with every level of the index (variables, sub-indicators, indicators, sub-categories, categories, Index).
- Correlation structure: There should be no cases of low correlation between variable/indicator and higher levels of the Index.
- Correlation structure: There should be no cases of redundant variables/indicators and/or highly correlated with every level of the Index that are of no added value for the overall Index.
- Principal Component Analysis (PCA): The percentage of variance explained by the first principal component should be ideally higher or equal to 70%. In cases failing this, alternate frameworks were trialed to maximise the result while maintaining conceptual coherence.

In Annex 1, a full variable selection tree is provided.

OUTLIERS & TREATMENT

In some instances, a variable includes observations which lie far away from the mass of the rest of the distribution. Including these extreme observations in the IIAG would bias the measure scores as, after normalisation, the outliers would make the range skewed and differentiation between most of the countries' scores would be difficult. To prevent this, the raw data is analysed to determine whether any of the variables require treatment to address outliers.

Outlier diagnostics identify outliers in the raw data, for each variable independently. All of these diagnostics are based on Tukey's method, which measures the distance of extreme observations from an upper and lower threshold. These thresholds are calculated using the interquartile range (IQR), the 1st and 3rd quartile and a factor. If a point lies above the upper threshold or below the lower threshold, it is considered an outlier.

$$\text{Interquartile range (IQR)} = 3^{rd}\text{ quartile} - 1^{st}\text{ quartile}$$
$$\text{Upper threshold} = 3^{rd}\text{ quartile} + IQR * \text{factor}$$
$$\text{Lower threshold} = 1^{st}\text{ quartile} - IQR * \text{factor}$$

In the 2024 IIAG, 12 variables were treated for outliers: *Absence of Violent Events in State-Based Conflict, Absence of Violent Events in Non-State Conflict, Absence of Government Violence against Civilians, Absence of Non-State Actor Violence against Civilians, Absence of Refugees, Cost of Mobile Internet, Internet Bandwidth Usage, Cost of Broadband Internet, Internet Security, Sustainable Agriculture, Control of Primary Forest Loss* and *Marine Protected Areas*.

These variables were filtered according to the following scheme: the trimmed mean and trimmed standard deviation of the measure were computed on the central 95% of the distribution; i.e. the bottom and top 2.5% were not used to compute the mean and standard deviation. All observations further than 3 trimmed standard deviations away from the trimmed mean were replaced by the trimmed mean plus 3.1 times the trimmed standard deviation, if they were in the right-hand tail, and replaced by the trimmed mean minus 3.1 times the trimmed standard deviation, if they were in the left-hand tail.

DEALING WITH MISSING DATA

The majority of the variables included in the IIAG have a degree of missing data over the time series (2014-2023). To ensure continuity between composite scores over time, values for these years are imputed.

In order to determine the most suitable method of imputation for the IIAG, simulation experiments were conducted, in which a proportion of the data were deleted using various missingness mechanisms. The deleted data were then imputed using a range of methods.

From the point of view of accuracy, precision and the amount of missingness remaining after imputation, it was determined that the best method of imputation was linear interpolation for missing data which are in the interior of the time series. The interior missing values are replaced with numbers incrementally higher or lower than the neighbouring data points.

For missing data points that are located in the exterior of the available time series, data for a country in a previous or following year was deemed to be the best proxy available to measure governance in the given year. Hence, the exterior missing values are replaced using the closest data point from source (last value carried forward - LVCF - or first value carried backward - FVCB -).

As an example, Country A has data missing for years 2014, 2017 and 2018. The second row of Table 2 contains raw data. For the interior missing values in years 2017 and 2018, linear interpolation is used to obtain the value, as shown in the third row of Table 2. For the exterior missing value in year 2014, the missing data is imputed using FVCB to obtain the data as shown in the fourth row.

Table 2. Imputation example for Country A

Year	2014	2015	2016	2017	2018	2019
Raw data from source		85.1	84.8			82.8
Between data points		85.1	84.8	*84.1*	*83.5*	82.8
Outside data points	*85.1*	85.1	84.8	*84.1*	*83.5*	82.8

Let tx be the raw data value of a variable for a country in year t. If there is no data for year t_2, and there are data for the years t_1 and t_3, whereby t_1 and t_3 are the closest years to t_2 such that $t_1<t_2<t_3$, linear interpolation can be illustrated as follows:

$$x_2 = x_1 + (t_2 - t_1) * \frac{x_3 - x_1}{t_3 - t_1}$$

Using the above formula in the previous example, a missing value in 2017 would be imputed as follows using linear interpolation:

$$84.1 = 84.8 + (2017 - 2016) * \frac{(82.8 - 84.8)}{2019 - 2016}$$

Exceptions

There are exceptions for when imputation is not applied. For countries which have missing data for the entire time series for a variable, no imputation is carried out. When a country only has one data point from source across the time series, this has been excluded from the IIAG dataset because imputing this point across the whole time series using LVCF/FVCB would be inappropriate due to the inaccuracies incurred.

NORMALISATION

Data used in the construction of the IIAG are diverse and include both quantitative and qualitative assessments provided by a range of data providers. At source, the variables collected are produced on different scales and can also have different polarities (e.g. in raw data terms, while for *Private Property Rights* higher score is better, for *Absence of Government Violence against Civilians* higher score is worse). In order for them to be meaningfully combined and compared, raw data are standardised before being included in the IIAG.

The data points for each variable are transformed using the min-max normalisation method. This method performs an order preserving linear transformation of the data, in which the maximum value in the raw data becomes the highest score and the minimum value in the raw data becomes the lowest score for each variable.

The min-max normalisation method subtracts from the actual value the minimum value of the entire raw dataset (2014-2023) in the specific variable and divides by the range of the variable values. The scores for each country's value in a particular year, where the polarity of the raw data is 'higher score is better', are transformed as follows:

$$x = 100 * \frac{\text{actual value} - \text{minimum raw value}}{\text{maximum raw value} - \text{minimum raw value}}$$

If the polarity of the raw data is 'higher score is worse', the scores must be inverted. Therefore, the previous formula must be altered as follows:

$$x = 100\text{-}100 * \frac{\text{actual value} - \text{minimum raw value}}{\text{maximum raw value} - \text{minimum raw value}}$$

This method of standardisation allows all scores to be published in common units and within the same bounds of 0.0-100.0, in which 100.0 reflects the best performance in Africa across the ten years. 100.0 does not imply perfect governance but indicates the best score in Africa according to the raw data. Thus, meaningful comparisons between variables and countries can be made. Other advantages of this method include preserving the skewness and kurtosis and not making any assumptions about the distribution of the data.

AGGREGATION & WEIGHTING

The IIAG uses linear, additive aggregation and weights each sub-component equally within its dimension. While there are a number of different types of aggregation methods with respective pros and cons, there is no set standard for aggregation in composite indices. The linear aggregation method has advantages in its simplicity, transparency and accessibility. The decision to weight the four overarching categories equally in the IIAG was taken based on the judgment that the four governance dimensions of the IIAG – *Security & Rule of Law*, *Participation, Rights & Inclusion*, *Foundations for Economic Opportunity* and *Human Development*- are of equal importance in measuring governance.

While the weight of the categories and sub-categories is equal in the *Overall Governance* composite score, indicators have different implicit weighting as a result of the structure of the IIAG - with the number of indicators per sub-category going from four to six, with the majority having five. This effectively means that each of the indicators of sub-categories such as *Sustainable Environment*, which only has five underlying indicators, contribute more to the *Overall Governance* score than those of sub-categories such as *Health*, which has seven underlying indicators.

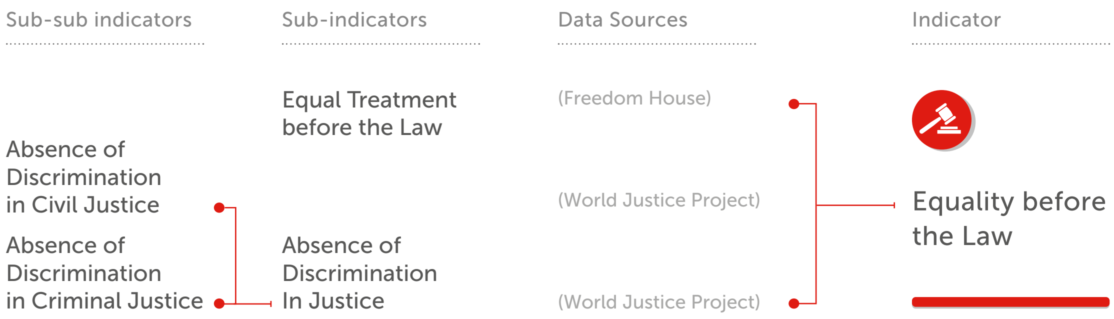
CLUSTERING

The underlying 322 variables included in the IIAG are constituents of 96 indicators, 90 of which are clustered indicators. A clustered indicator or a sub-indicator is composed of a number of underlying variables which capture the same dimension.

The data included in the IIAG comes from 49 different institutions and are a mix of quantitative and qualitative assessments. The diversity of sources and types of data means the measurement error in the composite values is minimised.

A measure can be clustered as a result of similar variables being available from multiple sources or when different variables can be considered a proxy for the same wider concept.

For example, variables measuring narrow concepts such as *Absence of Discrimination in Civil Justice* and *Absence of Discrimination in Criminal Justice*, collected from the World Justice Project, are clustered together to create the *Absence of Discrimination in Justice* sub-indicator. In turn this sub-indicator is clustered together with the sub-indicator *Equal Treatment before the Law*, collected from Freedom House, to create the indicator *Equality before the Law*.



QUANTIFYING UNCERTAINTY

The Foundation publishes standard errors and confidence intervals alongside the *Overall Governance* and category scores to reflect degrees of uncertainty. The standard errors and confidence intervals allow users of the IIAG to discriminate, to a certain degree, between changes in the values of the IIAG that can be confidently treated as actual changes in the state of governance and changes that might be due to "noise", or are at least insufficiently sizeable to be able to ascribe a high likelihood to such change being statistically significant.

Users of the IIAG are encouraged to use these to interpret changes in country scores over time and differences between country scores. For example, score or rank comparisons between countries with overlapping confidence intervals should generally be avoided as they represent a statistical tie. Similarly, users are encouraged not to over interpret marginal differences in a country's score change and refer to the confidence intervals for statistically significant movements.

Standard errors

For previous IIAG iterations, the method that has been used to calculate the standard errors was based on a process called *Bootstrapping*. It relies on creating multiple new versions of the dataset, by replacing values of measures with other values from the dataset (allowing for duplicates). It is a very common method and has the advantage of not making assumptions on the distribution of the data.

Bootstrapping, however, has been criticised. The bootstrap process is done at the sub-category level. Consequently, it will only have up to seven indicator values (or fewer for different sub-categories or when data is missing) to bootstrap from, which is likely to produce standard errors that will underestimate the true standard deviation of the scores.

Another issue with bootstrapping is its lack of replicability because of the random nature of the selection of values for the creation of resampled datasets. As such if an external party were to calculate the standard errors themselves, even if they followed the methodology to write their own bootstrapping algorithm, they would not be able to arrive at the same results.

Additionally, since bootstrapping is done with replacement, in the presence of missing values, the resampled datasets may not have the same amount and/or the same positions of missingness as the original dataset, which could also lead to incorrect estimating of the true standard errors.

The MIF Research Team has worked with Prof. Ali Hadi, a longstanding member of the IIAG Expert Panel, to implement a new standard error calculation method, which solves some of the issues related to bootstrapping. The new method is theoretically-based and replicable (that is, it does not depend on the re-sampling methodologies as does the Bootstrap). Since it is given in a closed-form equation, it drastically cuts down the time needed for the programme to run. Additionally, the resultant standard errors are more accurate than those provided by the Bootstrap.

Confidence intervals

The standard errors are used to construct confidence intervals for the country scores. These confidence intervals allow users of the IIAG to discriminate between country scores and country trends over time. While MIF chooses to use 90% confidence intervals, other degrees of confidence (80%; 85%; 90% and 95%) are available on our website.

READING THE RESULTS

Historical revision

The IIAG is refined and revised on a biannual basis to continually improve its measurement of governance. Improvements are a result of either methodological changes, or based on the inclusion of new data. Equally, if previously included measures undergo fundamental methodological changes or do not meet the criteria for inclusion it may be necessary to exclude them from future iterations. It is also necessary to update previously published data if retrospective revisions are made to data at source.

As a result of these changes, the IIAG is re-calculated every two years. The retrospective revision means that **no previous publications should be compared to the 2024 IIAG scores** as differences may be a result of a change in framework or an update in data from source rather than due to a change in score. **Score and rank comparisons between years should be made entirely within the 2024 IIAG.**

Relativity

The IIAG country scores and ranks are all relative, taking into account a country's performance in relation to the other 54 African countries. This is a result of the normalisation procedure, which transforms the raw data into a scale of 0.0-100.0, whereby 100.0 is the best score. This means that a country's change in score and rank may be reflective of other countries performing better or worse. Users of the IIAG are encouraged to treat marginal differences in scores and ranks with caution and refer to the standard errors for statistically significant changes.

Annex 1: Variable selection tree

A. Conceptual framework

1. Is the variable linked to MIF's definition of governance?

- Yes, the concept is defined in a way that captures the governance aspect. When "proxy" series are used, there needs to be evidence of their relationship with "target" series.
- No, the concept is not defined in a way that captures the governance aspect.

2. Does this variable measure an input or outcome of governance?

- It preferably focuses on outcomes of governance (e.g. level of quality in education).
- It focuses on inputs of governance (e.g. laws or expenditure levels).

B. Data availability and methods

3. What are the data provider/source and its methods?

- Widely credible data source and the methods it uses are sound. Data are perceived to be objective, produced professionally in accordance with appropriate statistical standards and policies and practices are transparent.
- Not credible data source, using unknown, biased methods, or lacks rigorous data collection.

4. Is the data coherent/comparable?

Yes, coherence/comparability over time and across countries can be ensured. Data are based on common concepts, definitions, classifications, and methodology, and any differences are explained and can be accounted for.

No, coherence/comparability over time and across countries cannot be ensured. Incoherence over time refers to breaks in a series resulting from changes in concepts, definitions, or methodology. Data collected nationally (e.g. by a government agency) which are not internationally comparable cannot be used.

5. Can we expect the variable to be discontinued any time soon?

No, the source plans to keep producing the data, maintaining its operationalisation/methodology.

Yes, the source plans to stop producing the data, or change its operationaliation/methodology.

6. What is the country coverage?

It covers at least 33 African countries.

It covers fewer than 33 African countries.

7. How many data points are there at the country level?

There are at least 33 African countries with two data points from source since the beginning of the time series (2014 in the case of the 2024 IIAG dataset).

There are fewer than 33 African countries with two data points from source since the beginning of the time series (2014 in the case of the 2024 IIAG dataset).

8. How old is the latest data point at the country level?

There are at least 33 African countries whose latest data point is no more than three years old (2021-2023 in the case of the 2024 IIAG dataset).

There are fewer than 33 African countries whose latest data point is no more than three years old (2021-2023 in the case of the 2024 IIAG dataset).

9. What is the level of missingness? (Data coverage)

The percentage of values missing from a variable does not overcome 20% in a year.

The percentage of values missing from a variable overcomes 20% in a year.

10. What is the numerical granularity of the variable's raw data?

Raw data from source comes on a four-point scale or more, which is essential to meaningfully distinguish country performance.

Raw data from source comes on less than a four-point scale.

C. Statistical/internal structure

11. How does the variable correlate with other components/levels?

The variable has a correlation coefficient higher than 0.30 with every level of the index (variables, sub-indicators, indicators, sub-categories, categories, index).

The variable has a correlation lower than 0.30 or is negatively correlated with higher levels of the index (the indicator, sub-category or category in which the variable sits).
The variable is redundant and/or highly correlated with every level of the index (other components at the same level, as well as higher levels of the index).

mo.ibrahim.foundation

FB→ /MoIbrahimFoundation

X→ @Mo_IbrahimFdn #IIAG

IG→ moibrahimfoundation