PHOTO: USAID

ZIMBABWE GOVERNANCE INDICATORS ANALYSIS

METHODS DOCUMENT

February 5, 2024

Jeremy Springman | Ryan Hatano

This publication was produced for review by the United States Agency for International Development. It was prepared by the Cloudburst Group. The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government

Prepared By:
The Cloudburst Group
8400 Corporate Drive, Suite 550
Landover, MD 20785-2238
Tel: 301-918-4400

TABLE OF CONTENTS

ACRONYMS

AFIDEP&GI	African Institute for Development Policy & Global Integrity
ARM	Average Running Minimum
GoZ	Government of Zimbabwe
GPP	General population poll
IIAG	Ibrahim Index of African Governance
MLP	Machine Learning for Peace
QRQ	Qualified Respondents' Questionnaire
PTS	Political terror scale
V-Dem	Varieties of Democracy
WJP	World Justice Project
ZDERA	Zimbabwe Democracy and Economic Recovery Act of 2001

EXECUTIVE SUMMARY

PURPOSE AND RESEARCH QUESTIONS

Zimbabwe's current government has verbally committed to the implementation of key reforms under the Zimbabwe Democracy and Economic Recovery Act of 2001 (ZDERA), including crucial governance reforms to improve accountability and the protection of constitutionally guaranteed rights. USAID has commissioned researchers at Cloudburst Group and the University of Pennsylvania to 1) assess indicators proposed to measure the Government of Zimbabwe's (GoZ's) progress toward its governance reform commitments and 2) advise on targets for those indicators. The results of the study will be used by USAID to inform engagement with the GoZ and with other stakeholder groups.

The core research questions that motivate this project are:

1. What level of improvement constitutes meaningful progress on the agreed-upon indicators?
2. What real-world events and policies lead to changes in these indicators?
3. What are the potential limitations/weaknesses of these indicators and what are the possible solutions to those limitations/weaknesses?

METHODS

This technical report analyzes cross-national variation on key governance indicators included in ZDERA criteria to establish empirically grounded targets for GoZ's governance reforms. To provide a substantive understanding of these targets, this report complements the comparative analysis with visualizations of the data, identification of countries that made lasting improvements on the ZDERA indicators and an analysis of their change over time, and a comparison of changes in indicators with event data from the Machine Learning for Peace (MLP) dataset.

FINDINGS

The report yields several notable findings that are summarized below for each research question

1. What level of improvement constitutes meaningful progress on the agreed-upon indicators?
    - The team recommends USAID target improvements in the 50th or 75th percentile of observed changes in the absolute value of Mo Ibrahim Foundation's Ibrahim Index of African Governance (IIAG) sub-indicators (civil society space, impartiality of the judicial system, and democratic elections). For the fourth IIAG sub-indicator (absence of violence against civilians), Zimbabwe's 2019 value is at a high level, making a large improvement difficult.
    - The team finds that the publicly available version of the World Justice Project (WJP) indicators has methodological problems that limit its ability to reliably detect changes in governance over time. To ameliorate these concerns, the team recommends that USAID obtain versions of WJP measures that exclude data from the general population poll (GPP).
2. What real-world events and policies lead to changes in these indicators?
    - The team finds that event data has a limited ability to predict changes in IIAG scores. However, an exploratory analysis suggests several interesting patterns, including increased reporting on corruption in periods of opening civic space and major surges in civic activism in response to election-related government repression.

3. What are the potential limitations/weaknesses of these indicators and what are the possible solutions to those limitations/weaknesses?
    - The WJP'S index combines a Qualified Respondents' Questionnaire (QRQ), updated annually, a GPP, updated sporadically, and a political terror scale (PTS) index, updated annually. Due to its irregular updates and changing sample, the GPP is vulnerable to major composition effects and renders the aggregate WJP indicators an unreliable measure of changes in governance over time. WJP measures should be reconstructed without GPP indicators.
    - Both the WJP and IIAG indicators rely on subjective measures of governance based on expert opinion. Little and Meng (2023) raise concerns that expert opinions may overestimate changes as a reaction to major political events. To mitigate this, the team recommends pairing analyses of IIAG and WJP data with event data to identify major events that may influence subjective scores.

BACKGROUND

Zimbabwe's current government has committed to the implementation of key reforms, including crucial governance reforms to improve accountability and protection of constitutionally guaranteed rights. USAID seeks to assess selected governance indicators to measure the GoZ's progress toward its governance reform commitments. The results of the study will be used by USAID to inform engagement with the GoZ and assess indicator targets for an evaluation of Zimbabwe's political reforms for discussion with interagency and multi-stakeholder groups.

This study will assess the selected indicators measuring the GoZ's progress toward its governance reform commitments and provide clarity about the degree of progress expected by the GoZ and inform future deliberations about dialogue with GoZ. It will also assess whether the selected indicators can be gamed by the GoZ. In order to do this, the research team addresses the following research questions:

1. What level of improvement constitutes meaningful progress on the agreed-upon indicators?
2. What real-world events and policies lead to changes in these indicators?
3. What are the potential limitations/weaknesses of these indicators? What are possible solutions to those limitations/weaknesses?

APPROACH

This technical report analyzes cross-national variation on key governance indicators included in ZDERA criteria. The comparative analyses herein are designed to help establish reasonable targets that indicate meaningful improvements in ZDERA indicators. The report begins by describing the data covered by ZDERA indicators as well as the strengths and limitations of these sources. It then describes each of the three driving research questions and presents the findings from the analyses.

This identifies a range of potential ZDERA indicator targets that would constitute a low, medium, and high bar for defining meaningful improvements. The report also briefly describes the progress GoZ has made in recent years and notes the substantively large discrepancies between WJP and IIAG measures, concluding that the IIAG data is a much more accurate measure of countries' changes in governance over time.

The team then drew on event data capturing major changes in civic space to assess the correlation between real-world events from the MLP dataset and durable improvements in IIAG democracy indicators. Ultimately, the team found that the event data cannot reliably predict changes in IIAG scores; this is likely due to the relatively small sample size and the different levels of aggregation of these data (monthly for event data and annual for IIAG) rather than a failure of either source to capture valuable information about changes in governance.

However, the team did observe a strong correlation between the frequency of several major political events and durable improvements in IIAG scores. To facilitate a deeper understanding of these events, the team used a GPT-4 model to summarize several thousand news articles scraped by MLP from domestic news outlets in countries that experienced a large number of these political events during non-improvement years. However, USAID should be cautious when using this to benchmark whether or not meaningful improvements are occurring during periods for which IIAG data is unavailable. The report then briefly describes recent trends in event data for Zimbabwe and reflects on the potential use of the MLP data to continue tracking these trends.

DATA, INDICATORS, AND LIMITATIONS

This section describes the indicators selected by USAID and other stakeholders to evaluate GoZ's progress toward its governance reform commitments. The team drew these indicators from the WJP Rule of Law Index and the IIAG. The team explored the construction of these index measures, the underlying data used, and the sources' limitations. The limitations of WJP create significant challenges to confident inferences about governance improvements over time. For this reason, any improvements in GoZ's WJP governance scores should be corroborated by similar improvements captured by other data sources.

WORLD JUSTICE PROJECT RULE OF LAW INDEX

WJP relies primarily on two sources of original data: the QRQ and the GPP. The QRQ is a survey of in-country practitioners and academics with expertise in governance and law soliciting their perceptions of their country's rule of law. These surveys include questions on the efficacy of courts, the strength of regulatory enforcement, and the reliability of accountability mechanisms. WJP collects this data annually for every country beginning in 2012. However, WJP significantly revised its surveys in 2015, preventing comparisons between data collected before 2015 with data collected from 2015 forward (Ponce, 2015).

The GPP solicits ordinary citizens' perceptions of the rule of law in their country. These surveys include questions on the ease of interacting with state bureaucracy, the extent of bribery and corruption, the availability of dispute resolution systems, and the prevalence of common crimes. Depending on the country and year of the survey, this questionnaire is administered to either a representative sample of respondents in the three largest cities or to a nationally representative sample. On average, the GPP is conducted every four to five years for each country. Importantly, the timing of GPP surveys is highly variable across countries. Each year's index measures are calculated using data from the country's most recent GPP. According to the 2023 WJP report, three countries have GPP data from 2023, 42 have data collected between 2019 and 2022, 103 countries have data collected between 2016 and 2018, and six countries had their most recent GPP data collected before 2016.

Table 1: The WJP

DATA SOURCES	METHODS	RESPONDENTS	YEARS
GPP	Survey of the public's firsthand perceptions of rule of law	1,000 randomly selected citizens	2016, 2018*a*
QRQ	Survey of lawyers, academics, and civic actors	24 in-country experts	2015–2017, 2019–2023
PTS	Index of levels of political violence and terror	Reports from international sources (Amnesty International, U.S. State Department, Human Rights Watch)	2015–2021

[a] GPP years vary by country. Years for Zimbabwe are reported in this table.

Table 1 summarizes the WJP data sources and coverage. In addition to the GPP and QRQ, WJP also incorporates third-party data sources into several indicators. To aggregate individual measures drawn from the QRQ, GPP, or third-party sources to create index measures, WJP normalizes all measures to a 0–100 score and takes a simple average. According to WJP, indicators draw most heavily on data from the QRQ, with 65.32 percent of the total measures factored into their indices coming from the QRQ, while 25.83 percent come from the GPP and 8.85 percent come from third-party sources.

Table 2 lists the WJP indicators selected by USAID and other stakeholders to evaluate GoZ's progress toward its governance reform commitments. For the United States Government, these indicators relate to congressionally identified conditions under the ZDERA. Looking specifically at the indicators used for this analysis, the data from third-party sources included in these WJP indicators includes only the PTS.

The WJP data have important characteristics that limit their suitability for measuring sustained improvements in governance with confidence. First, while the QRQ is conducted annually, the GPP is conducted intermittently. This means that large changes in public perception will only be factored into WJP scores if there was a survey conducted after these changes in perceptions occurred. For Zimbabwe, the WJP conducted the most recent GPP in 2018 and first included it in the 2019 data (Ponce, 2018, 163). All scores for Zimbabwe beginning with 2019 utilize the 2018 GPP data, while all Zimbabwe scores between 2018 and 2016 utilize the 2016 GPP data. This creates several challenges.

Table 2: WJP Indicators

INDICATOR	INDICATOR DEFINITION	ZDERA CRITERIA	SOURCES
1.4: Government officials are sanctioned for misconduct	Measures whether government officials are investigated, prosecuted, and punished for official misconduct and other violations.	4. Military and Police Subordinate to Civilian Government	WJP-QRQ WJP-GPP
1.6 Transition of power is subject to the law	Measures whether government officials are elected or appointed in accordance with the rule of law, whether elections take place, and the integrity of the electoral process.	2. Free and Fair Elections OR Improved Pre-Election Conditions	WJP-QRQ WJP-GPP
4.2 Right to life and security of person effectively guaranteed	Measures whether the police inflict physical harm upon criminal suspects and whether political dissidents or members of the media are subjected to intimidation or violence.	1. Restoration of Rule of Law	WJP-QRQ WJP-GPP PTS
4.4 Freedom of opinion and expression is effectively guaranteed	Measures whether independent media, civil society organizations, or political parties, are free to report on government policies without fear of retaliation.	1. Restoration of Rule of Law	WJP-QRQ WJP-GPP
4.7 Freedom of assembly and association is effectively guaranteed	Measures whether people can freely attend community meetings, join political organizations, or hold peaceful public demonstrations.	1. Restoration of Rule of Law	WJP-QRQ WJP-GPP
6.5 Government does not expropriate without lawful process and adequate compensation	Measures whether the government respects property rights, refrains from illegal seizure, and provides adequate compensation when property is legally expropriated.	1. Restoration of Rule of Law 3. Equitable, Legal, and Transparent Land Reform	WJP-QRQ WJP-GPP
8.7: Due process of the law and rights of the accused	Measures whether the basic rights of criminal suspects are respected, including the presumption of innocence, access to evidence, non-abusive treatment, and provisions of legal assistance. Also includes the basic rights of prisoners after they have been convicted.	1. Restoration of Rule of Law	WJP-QRQ WJP-GPP

Most importantly, changes detected in 2016 or 2019 may be driven by the introduction of new GPP data. In 2016, the introduction of GPP data for the first time may cause an abrupt change in governance scores, even if these citizen perceptions were no different in 2016 than they were in 2015. While the addition of

data measuring citizen perceptions adds an important perspective, it may create a false appearance of change even as the actual quality of governance remains constant. Similarly, the 2016 GPP relied on a sample from the country's three largest cities (Harare, Bulawayo, Chitungwiza) while the 2018 GPP surveyed a nationally representative sample of citizens. For this reason, changes in 2019 may be driven by differences in the composition of the sample (for example, the inclusion of rural citizens) rather than true changes in citizen perceptions of rule of law over time.

Second, because the 2018 GPP is used for 2019–2023, any changes after 2019 are not captured, creating a false impression of stability. This is especially troubling given the 2017 change in government. If citizens were overly optimistic about a new government during the 2018 survey, these elevated GPP scores are included in the aggregate scores for 2019–2023. Not only are any changes in citizen perceptions after 2018 not captured in the data for those years, but also the scores for 2018 are re-used every year, muting the extent to which changes in the QRQ measures (which are conducted every year) are able to drive changes in the aggregate score.

Finally, the QRQ and GPP may be vulnerable to manipulation. While many expert surveys rely on opinions from individuals residing both inside and outside countries of interest, both sources rely entirely on responses from residents of Zimbabwe. QRQ respondents include heads of leading law firms, universities, and non-governmental organizations, which may be subject to pressure from from GoZ. Similarly, GPP respondents or the survey firm implementing the GPP may be vulnerable. For this reason, additional information about the firms involved in data collection and the experts consulted for the QRQ may be necessary to establish an appropriate level of confidence for these measures.

IBRAHIM INDEX OF AFRICAN GOVERNANCE

The Mo Ibrahim Foundation's IIAG measures the quality of governance across 54 African countries. IIAG relies primarily on publicly available third-party data. Table 3 summarizes the IIAG data sources and coverage. Looking specifically at the IIAG indicators selected to evaluate GoZ's progress toward its governance reform commitments, IIAG uses the African Institute for Development Policy & Global Integrity (AFIDEP&GI), Varieties of Democracy (V-Dem), Freedom House, PTS, and the Armed Conflict Location & Event Data Project. IIAG's indices cover four dimensions of governance: security & rule of law; participation, rights, & inclusion; foundations for economic opportunity; and human development. These categories are made up of 16 sub-categories, consisting of 81 total indicators. Within the security & rule of law theme, this report focuses on the sub-themes of civil society space, impartiality of the judicial system, and participation in democratic elections. Table 4 lists the IIAG indicators selected by USAID and others to evaluate GoZ's progress toward its governance reform commitments.

IIAG data is released biannually, and the available IIAG data cover 2012 to 2021. The next update is scheduled for release in 2024 and will extend the data's coverage through 2023. As with WJP, IIAG data has several important limitations. Fortunately, these limitations do not appear as problematic for measuring sustained improvements in governance with confidence.

First, although the external data sources that compose IIAG's measures are global in nature, the IIAG measures are only available for the 54 African countries. This limits the number of countries that the team can compare over time. Second, IIAG's reliance on external indicators makes it vulnerable to methodological changes that could drive sudden changes in the aggregate scores. Although the initial

review did not uncover any such changes, a more thorough review is still necessary. Third, the biannual release cycle of IIAG makes it impossible to detect the most recent changes in GoZ governance.

It is important to note that several of these limitations can be mitigated by consulting the underlying third-party data that constitute the IIAG index measures. Specifically, several of these underlying data sources, V-Dem in particular, cover longer time periods for a much larger sample of countries and are released annually (rather than bi-annually).

Table 3: The IIAG

SOURCES	METHODS	CODERS	YEARS
African Institute for Development Policy	Rule of law score based on legal and scholarly reviews, expert interviews, and media review.	Country experts	2012–2021
V-Dem	Governance characteristics scored based on expert opinion.	25 country experts (2/3 local and 1/3 foreign)	2012–2021
Armed Conflict Location and Event Data Project	Dataset of reported political violence and protest events from local media.	Research assistants	2012–2021
PTS	Index of political violence and terror based on reports from Amnesty International, the U.S. State Department, and Human Rights Watch.	Country experts	2015–2021
Freedom House	Ratings from field research, local consultations, and desk research.	International consultants	2012–2021

Table 4: IIAG Indicators

INDICATOR	INDICATOR DEFINITION	ZDERA CRITERIA	SOURCES
Civil society space	Assesses the extent to which civil society and non-governmental organizations are free to establish and operate and are free from repression and persecution.	1. Restoration of Rule of Law	AFIDEP&GI V-Dem
Rule of law and justice—impartiality of the judicial system	Assesses the extent to which the judicial system is impartial based on independence of the courts, autonomy of judges, and appointment of judges.	1. Restoration of Rule of Law	AFIDEP&GI V-Dem
Participation—democratic elections	Assesses the extent to which elections are free and fair and the extent to which election monitoring bodies and agencies	2. Free and Fair Elections OR	AFIDEP&GI V-Dem

INDICATOR	INDICATOR DEFINITION	ZDERA CRITERIA	SOURCES
	are independent and have operating capacity, including for reporting.	Improved Pre-Election Conditions	Freedom House
Absence of violence against civilians	Measures the number of violent events against civilians committed by government forces and non-state actors, as well as the levels of political violence in a country.	2. Free and Fair Elections OR Improved Pre-Election Conditions 4. Military and Police Subordinate to Civilian Government	The Armed Conflict Location & Event Data Project PTS

Finally, the broader range of sources that IIAG data are drawn from and the mix of experts residing within and outside the relevant country reduce concerns about manipulation. Although manipulation cannot be completely ruled out, it would be extremely difficult to influence the responses of such a broad range of experts residing in multiple countries and reporting to multiple organizations.

RESEARCH QUESTIONS AND FINDINGS

In this section, the team presents analyses of the three research questions motivating this report in the hopes of informing USAID's determination of targets that would satisfy GoZ's ZDERA obligations on key democracy indicators, identify political events that might signal forthcoming improvements, and assess the level of confidence admitted by the underlying data sources used in this analysis. The report includes targets across both WJP and IIAG indicators. However, the discussion focuses on IIAG because the team believes that WJP indicators are likely misleading. If more detailed WJP data is made available, future drafts may include a similar discussion of WJP targets.

RQ1: WHAT LEVEL OF IMPROVEMENT CONSTITUTES MEANINGFUL PROGRESS ON THE AGREED-UPON INDICATORS?

KEY TAKEAWAYS

The team recommends USAID target improvements in the 50th or 75th percentile of observed changes in the absolute value of Mo Ibrahim Foundation's IIAG sub indicators (civil society space, impartiality of the judicial system, and democratic elections). For the fourth IIAG sub-indicator (absence of violence against civilians), Zimbabwe's 2019 value is at a high level, making a large improvement difficult

The team finds that the publicly available version of the WJP indicators has methodological problems that limit its ability to reliably detect changes in governance over time. To ameliorate these concerns, the team recommends that USAID obtain versions of WJP measures that exclude data from the GPP.

To construct evidence-informed targets for each indicator, the team analyzed cross-national variation on key governance indicators included in ZDERA criteria. The selected indicators cover a relatively short period of history for each country. Because the data does not cover multiple historical episodes of governance reforms for Zimbabwe, it is impossible to generate an evidence-informed target from a within-country analysis of Zimbabwe. For this reason, the team defined targets based on an analysis of the size of durable improvements made by other countries. This allowed the team to identify a range of potential ZDERA indicator targets that would constitute low, medium, and high bars for meaningful improvements on these indicators.

IDENTIFYING DURABLE IMPROVEMENT EPISODES ACROSS COUNTRIES

To identify improvement targets for each governance indicator, the team began by calculating the size and duration of changes in indicator scores. For each country indicator, the team identified the minimum value that had been observed so far at each year in the data (a "running minimum"). After identifying the running minimum in each country indicator time series, the team averaged across that minimum value and the

values for the two previous years, creating an average running minimum (ARM).[1] The team then compared this ARM to the value of each proceeding year to calculate the percentage change in the value for each year relative to the ARM. The team then identified countries that experienced a 10 percent or greater increase in their governance score over the proceeding years, taking 10 percent as the minimum size for an increase to be considered substantively meaningful. This minimum ensures that the team describes the size of durable improvements only among countries with some meaningful improvement.[2] Furthermore, to limit the sample to countries with sustained improvement, the team also restricted the sample to increases that endured consistently for at least five years (meaning that the indicator score never fell below the 10 percent improvement from the ARM value). This was defined as every year within the five-year period being at least 10 percent above the ARM value.[3]

This approach will allow USAID to compare the size of GoZ's observed increases after 2019 with those of other countries across the full sample period. It is important to note two characteristics of this approach. First, focusing on percentage increases in indicators benefits countries with very low scores because smaller increases are required to reach the threshold. This seems reasonable, given GoZ's relatively low scores and the importance of improvements from the worst-performing countries. Second, the ARM's inclusion of values from two years before the running minimum risks overstating increases for countries where the minimum is observed in 2012 or 2013 (because two years of earlier values are not available to be incorporated into the ARM).

Looking at the IIAG sample, the team observed 55 episodes of durable improvements over the 10 years and 54 countries in the sample. Thirteen countries (including Zimbabwe) experienced durable improvements on at least two indicators and five countries experienced durable improvements on three indicators.[4] Turning to the WJP data, only 14 countries experience durable improvements across any of these indicators, and while Zimbabwe increases on four of the seven indicators, no other country improves on more than one. Figure 1 and Figure 2 visualize the number and size of durable increases across indicators. While the primary objective of this research question is to establish targets for future improvements, these figures provide a sense of GoZ's improvements during the sample years. While GoZ is among the best-performing countries in the WJP data (an effect the team believes may be an artifact of composition changes in the data), its performance on IIAG indicators is good but not anomalously so.

Although the team uses a 10 percent improvement over a five-year period as the main specification, Appendix 1 explores less strict thresholds, including a 5 percent increase threshold and a three-year

[1] Taking the average of the minimum value and the two preceding values mitigates the influence of sudden, brief decreases in indicator values. This may happen in moments of political crisis in which some rights are temporarily suspended or as a result of measurement error. See Zimbabwe's scores in 2016 on the 'Right to Life' and 'Private Property' panels in Figure 12, which correspond with waves of anti-government protests that were met with heavy-handed repression.

[2] This approach resembles that of the Episodes of Regime Transformation dataset, which uses V-Dem data to identify periods of time during which countries are undergoing democratization or autocratization (Maerz et al., 2021). See Maerz et al. (2021) for a full description of the Episodes of Regime Transformation methods, which define a meaningful improvement as being an overall increase of at least 10 percent over the duration of a transition episode. See Boese et al. (2021) for an application of these data to an empirical analysis of the impact of democratization episodes on economic growth.

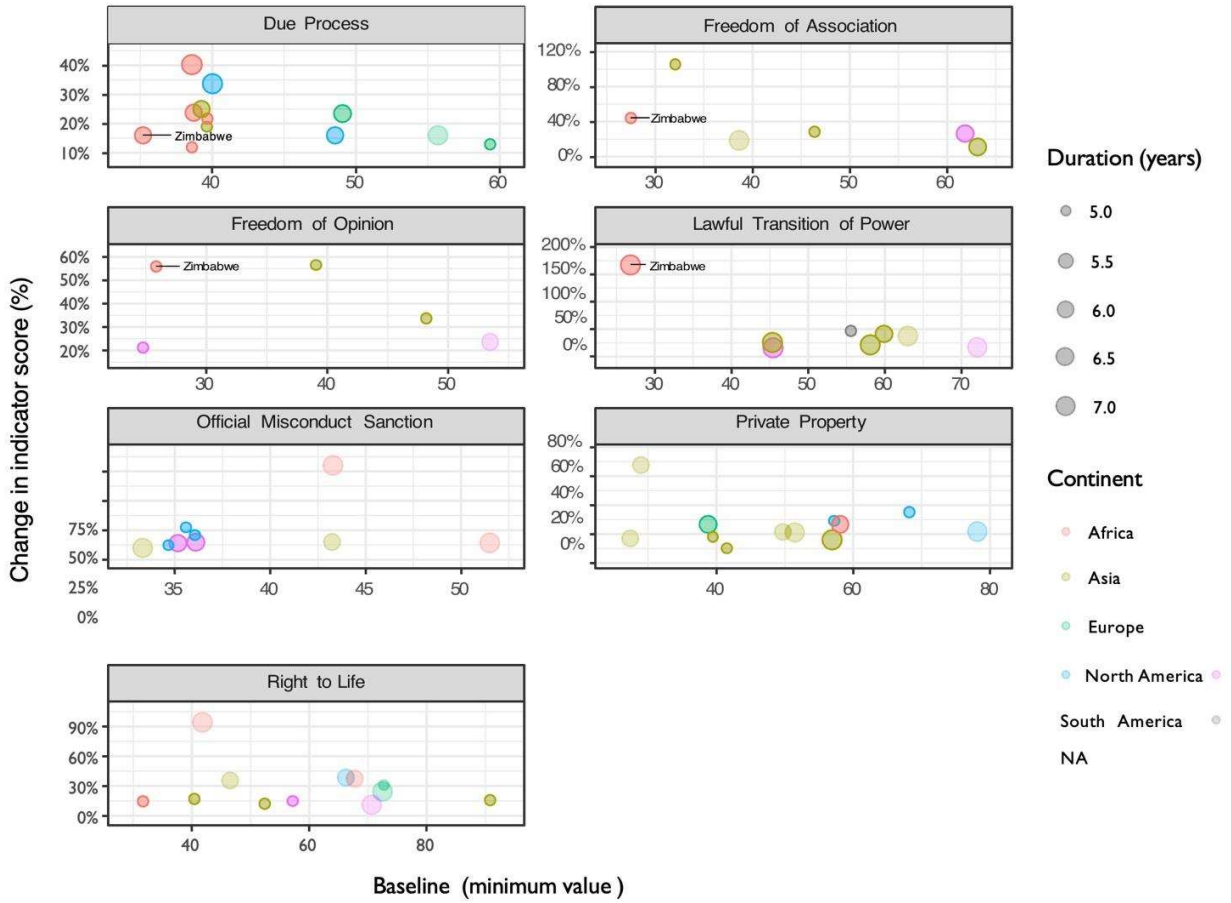
[3] The team argues that requiring improvements to be sustained for a minimum of five years helps to isolate periods of concentrated progress and reform. Conceptually, years in which scores fall back to within 10 percent of the ARM value.

[4] The limited episodes of sustained improvement observed in the Absence of Violence Against Civilians indicator compared to Democratic Elections or Impartiality of the Justice System reflect the higher baseline scores in this area. Starting from a higher initial value presents two challenges for achieving durable improvement. First, reaching a specific percentage gain, like 10%, requires a larger absolute improvement compared to cases with lower starting points. Secondly, improvements often exhibit diminishing returns, meaning that the same effort invested in a system with a higher starting point will yield a smaller relative improvement compared to a system starting from a lower point. Conversely, the scores for Civil Society are comparable to those of Democratic Elections or Impartiality of the Justice System, indicating a genuine lack of improvement events in this sector. For a visualization of the change in scores from 2012 to 2019, please refer to Figure 27.

improvement threshold. Table 10 demonstrates how these thresholds affect the commonness of durable increases in the sample.

To identify targets for GoZ's progress, the team identifies the size of increases for each indicator that would constitute small, medium, and large increases relative to durable improvements experienced in other countries. Table 5 describes the distribution of the size of improvement for a given country-year for each indicator. Specifically, the team recorded the size of the durable improvement (defined as the percent increase from the ARM of a country-year) at the 25th, 50th, and 75th percentiles of the distribution. For example, Zimbabwe's due process score for 2023 is 35.2, representing a 16 percent increase from Zimbabwe's average minimum score of 30.3 in 2015. This method of calculation enables the research team to identify specific country-years in which durable improvements begin or large changes in scores occur.[5]

Figure 1: Size and Duration of Increases on WJP Indicators



[5] The team considered alternative methods of calculating durable improvement by country-episode, rather than country-year. These included taking the average of the annual improvements by a country over an improvement episode or taking the difference between the ARM and the improvement episode endpoint. These aggregated methods, however, limit the ability to identify specific events or policies that may have attributed to changes in the scores.

Each point represents specific countries that sustained a minimum 10 percent increase for at least five years in a given indicator. The number of dots on each panel represents the total number of countries for each indicator that experienced such an improvement. The y-axis represents the magnitude of each country's increase, with larger increases represented by higher values on the graph. The x-axis is each country's baseline score, with countries further to the left starting from a lower (worse) score. Countries in the top left corner experienced large magnitude increases from low baseline conditions, while countries in the bottom right experienced smaller increases starting from higher baseline values. For indicators where Zimbabwe experiences such an increase, the corresponding point is labeled.

Figure 1: Size and Duration of Increases on WJP Indicators

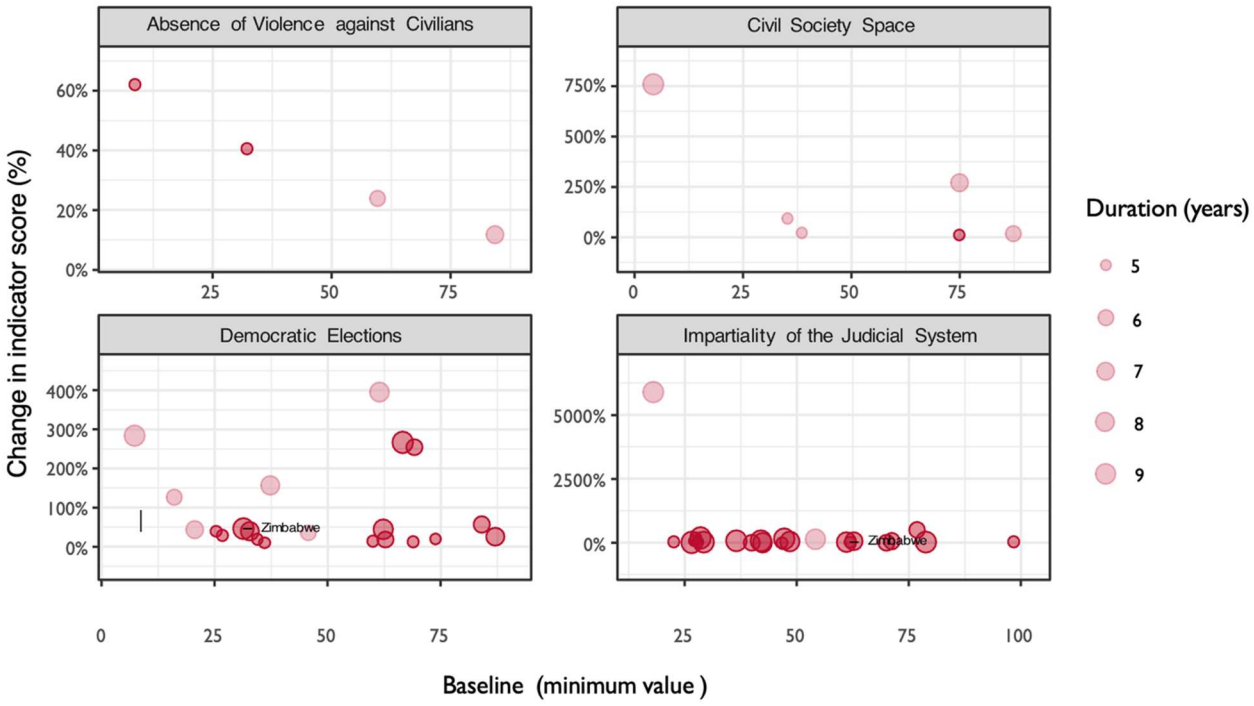


Figure 2: Size and Duration of Increases on IIAG Indicators

Each point represents specific countries that sustained a minimum 10 percent increase for at least five years in a given indicator. The number of dots on each panel represents the total number of countries for each indicator that experienced such an improvement. The y-axis represents the magnitude of each country's increase, with larger increases represented by higher values on the graph. The x-axis is each country's baseline score, with countries further to the left starting from a lower (worse) score. Countries in the top left corner experienced large magnitude increases from low baseline conditions, while countries in the bottom right experienced smaller increases starting from higher baseline values. For indicators where Zimbabwe experiences such an increase, the corresponding point is labeled.

ESTIMATING THE SIZE OF DURABLE IMPROVEMENTS

To measure uncertainty, the team used bootstrapping to generate confidence intervals around these numbers. These intervals provide USAID with a range of values around each quartile score that can be used to confirm whether or not a country falls above or below a given quartile score after accounting for

the amount of variation in the data. For example, if Zimbabwe exhibits a 50 percent increase in the judicial system impartiality score in a given year, this would indicate that Zimbabwe's improvement in that year is greater than the median improvement for the data on that indicator.[6]

The WJP indicator of due process has seen the smallest improvement percentages, at 14 percent, 18 percent, and 23 percent, for the 25th, 50th, and 75th percentiles, respectively. The team observed larger improvement percentages in the IIAG indicators, such as judicial system impartiality (26 percent, 51 percent, and 95 percent), civil society space (20 percent, 43 percent, and 270 percent), and democratic elections (28 percent, 45 percent, and 112 percent). The IIAG indicators are skewed at the highest end of their distribution by countries that scored extremely low during certain years, such as Somalia and Eritrea, and thus could show large percentage improvements with relatively small actual gains. For this reason, the team removed the bottom 10 percent of countries from this calculation. Appendix 2 shows the quartile scores and confidence intervals with these outlier countries retained.

Table 5: Quartiles of Improvement

SOURCE	INDICATOR	25TH QUARTILE	50TH QUARTILE	75TH QUARTILE
IIAG	Civil Society Space	0.19 (0.16 0.21)	0.21 (0.19 0.28)	0.3 (0.23 2.4)
IIAG	Judicial System Impartiality	0.25 (0.2 0.28)	0.42 (0.35 0.49)	0.8 (0.65 0.92)
IIAG	Democratic Elections	0.26 (0.21 0.29)	0.39 (0.32 0.45)	0.62 (0.5 0.81)
IIAG	Absence of Civilian Violence	0.14 (0.12 0.18)	0.18 (0.14 0.28)	0.28 (0.18 0.4)
WJP	Official Misconduct Sanction	0.17 (0.15 0.19)	0.21 (0.19 0.23)	0.25 (0.23 0.28)
WJP	Lawful Transition of Power	0.18 (0.15 0.23)	0.25 (0.22 0.32)	0.4 (0.3 0.46)
WJP	Right to Life	0.16 (0.14 0.18)	0.2 (0.18 0.23)	0.3 (0.24 0.34)
WJP	Due Process	0.14 (0.12 0.16)	0.17 (0.16 0.2)	0.22 (0.2 0.25)
WJP	Freedom of Opinion	0.22 (0.14 0.27)	0.26 (0.21 0.34)	0.32 (0.25 0.39)
WJP	Freedom of Association	0.15 (0.12 0.22)	0.22 (0.16 0.26)	0.26 (0.22 0.31)
WJP	Private Property	0.18 (0.16 0.2)	0.22 (0.2 0.26)	0.29 (0.25 0.31)

Bootstrap generated 95 percent confidence intervals for the 25th, 50th, and 75th percentile, with the lowest 10 percent of countries removed

ESTABLISHING REPRESENTATIVE DURABLE IMPROVEMENTS

To provide additional context for these scores, in Table 6, the team matched the median score (50th percentile) for each IIAG indicator to the improvement episode with the closest value. Countries whose increases are close to each quartile value can provide a qualitative reference point illustrating the substantive importance of increases of various sizes. For example, Morocco's improvement episode on the democratic elections indicator was the median improvement episode in the sample, with a 54 percent

[6] The improvement episode was still ongoing in 2021, which was the last year of the data.

increase in their 2018 score from their ARM value in 2014.[7] This qualitative reference point can be used to determine, for each indicator, whether a target larger or smaller than the median improvement episode should be used to judge GoZ's improvement.

Table 6: Median Country-Year Episodes of Improvement

COUNTRY	PERIOD	INDICATOR	%
Angola	2012–2019	Civil Society Space	0.49
Morocco	2014–2018	Judicial Impartiality	0.54
Ethiopia	2012–2017	Democratic Elections	0.46
Central African Republic	2014–2017	Absence of Civilian Violence	0.28

Country-episodes with the median percentage of improvement in selected IIAG indicators. If Zimbabwe's targets are set at the median levels of improvement, then the team can expect an improvement comparable to the improvement seen in the following four countries over the amount of time indicated in the period.

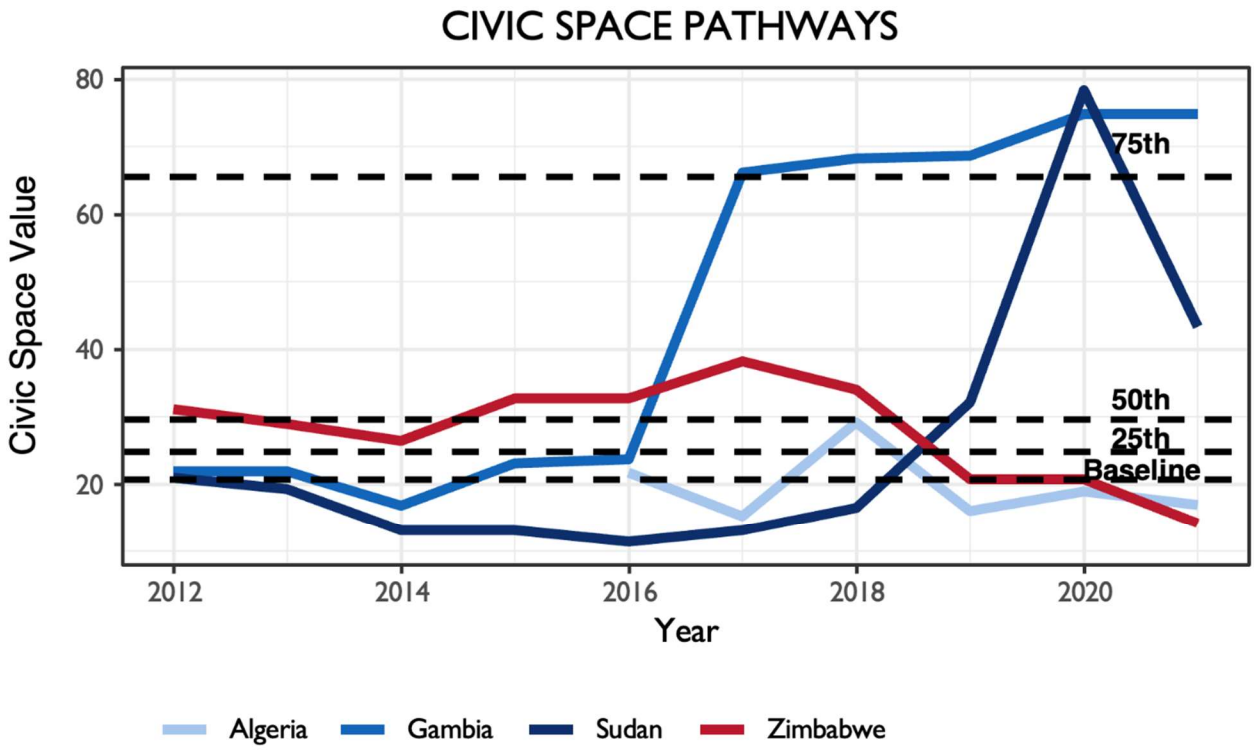
COMPARING THE TRAJECTORY OF COUNTRIES COMPARABLE TO ZIMBABWE

To provide additional qualitative reference points for each IIAG indicator, the team plotted the trajectory of countries that had a score similar to Zimbabwe's 2019 score (the baseline year for ZDERA reforms) at some point in the data. These figures show how these countries with comparable scores changed over time, beginning from the year when their score matched that of Zimbabwe's 2019 score. In 2019, Zimbabwe's civic space score was its worst-performing indicator. With a score of 20.7, Zimbabwe was in the bottom 13th percentile of countries. The environment for civic space in Zimbabwe in 2019 was comparable to that of Algeria in 2016, Gambia in 2012, and Sudan in 2012. While Algeria engaged in a brief period of improvement, reaching the 50th percentile of countries in 2018, the country ultimately declined to a level lower than its 2016 starting value. Sudan and Gambia experienced both longer and more dramatic improvements; by 2020, both countries reached the top quartile of improvements. However, Sudan began to decline after 2020. While Sudan offers a warning about the sustainability of brief improvements, Gambia represents a success case, remaining stable at an improved level.

[7] The team considered alternative methods of calculating durable improvement by country-episode, rather than country-year. These included taking the average of the annual improvements by a country over an improvement episode or taking the difference between the ARM and the improvement episode endpoint. These aggregated methods, however, limit the ability to identify specific events or policies that may have attributed to changes in the scores.

CIVIC SPACE PATHWAYS

Figure 3: Pathways of Countries from the Year in Which They Matched Zimbabwe's 2019 Score

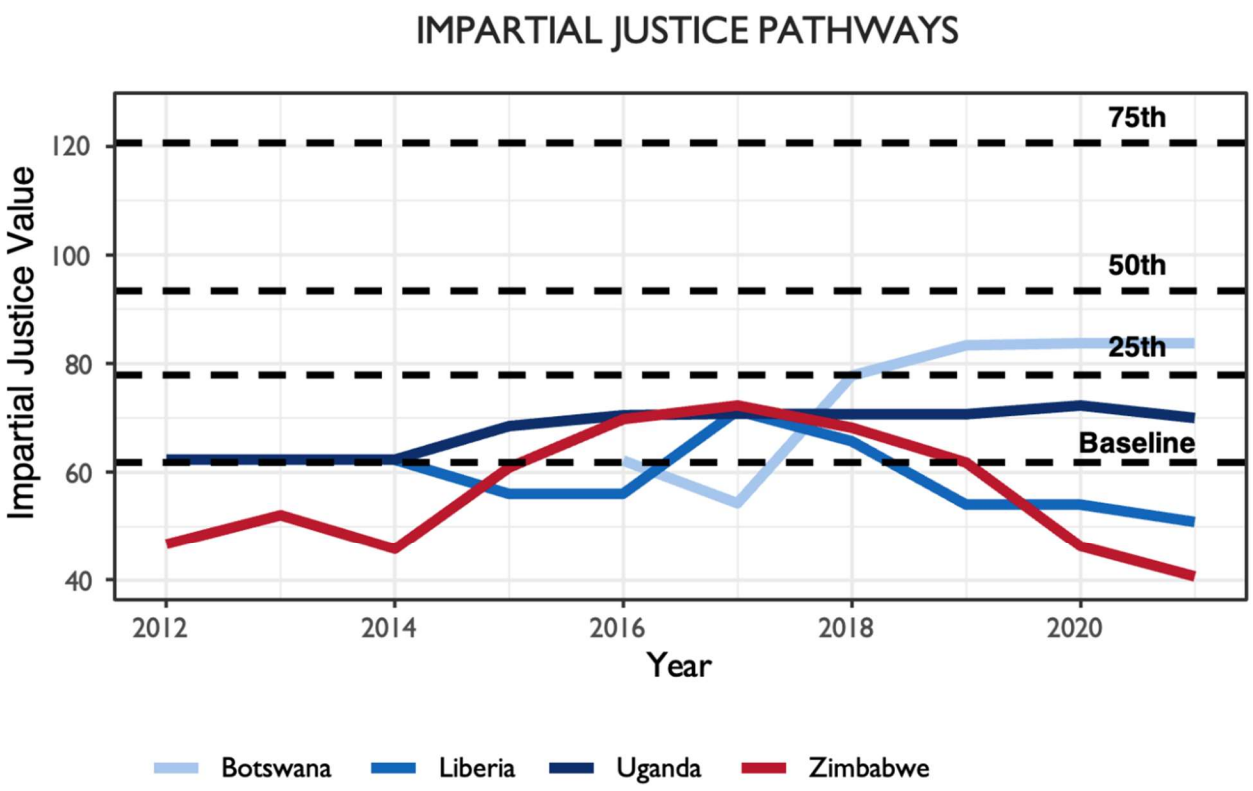


Zimbabwe's history since 2012 is shown in red, while three selected similar countries are shown in blue. These example countries begin their paths at the point in time in which they were equivalent to Zimbabwe in 2019. Horizontal dotted lines represent levels that would represent the 25th, 50th, and 75th percentile level of improvement from Zimbabwe's 2019 baseline, as calculated in Table 5.

Zimbabwe's impartial justice value in 2019 was 61.8, representing a score slightly higher than average. As seen in Figure 4, Zimbabwe's justice system was comparable to Botswana in 2016, Liberia in 2014, or Uganda in 2012. While Botswana and Uganda achieved mild improvements to their impartial justice scores, the Liberian score has fallen below the baseline value. Zimbabwe risks following an even more severe decline, with the 2021 score reaching the lowest level in the country's history.

IMPARTIAL JUSTICE PATHWAYS

Figure 4: Pathways of Countries from the Year in Which They Matched Zimbabwe's 2019 Score.



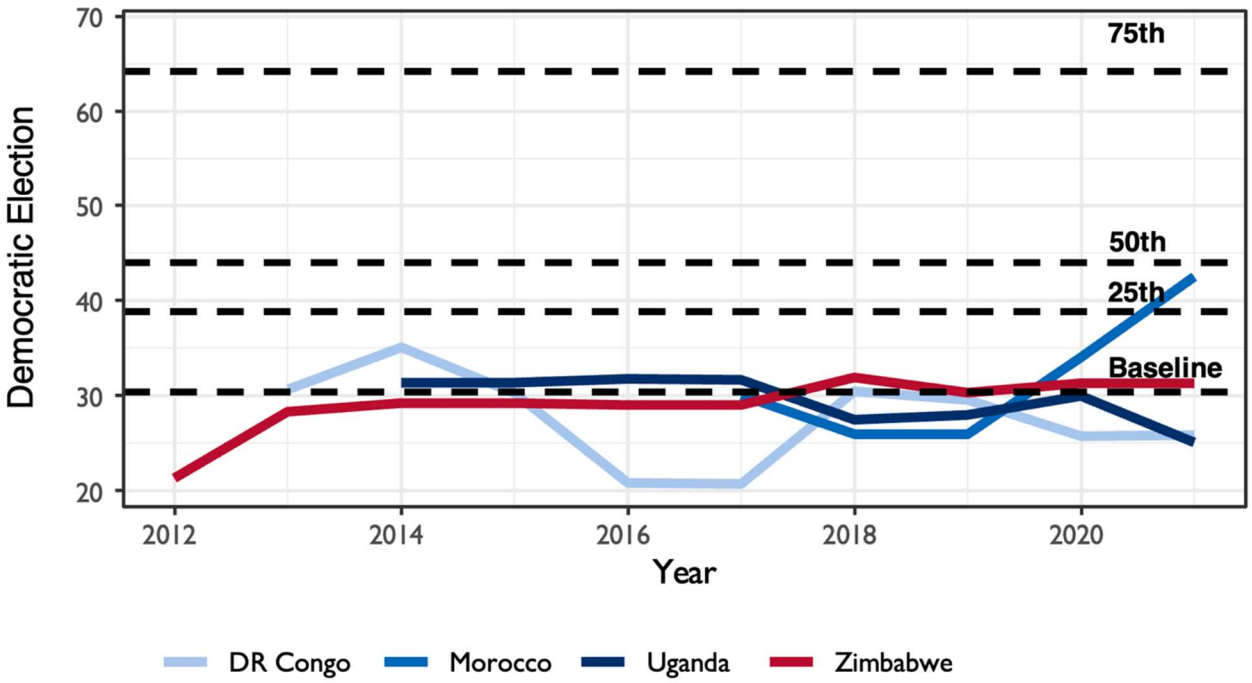
Zimbabwe's history since 2012 is shown in red, while three selected similar countries are shown in blue. These example countries begin their paths at the point in time in which they were equivalent to Zimbabwe in 2019. Horizontal dotted lines represent levels that would represent the 25th, 50th, and 75th percentile level of improvement from Zimbabwe's 2019 baseline, as calculated in Table 5.

Zimbabwe's second-worst-performing indicator was democratic elections, exhibiting a 24th percentile[8] value of 30.3 in 2019. Zimbabwe's 2019 electoral environment was comparable to the Democratic Republic of the Congo in 2013, Morocco in 2017, or Uganda in 2014, as seen in Figure 5. From Zimbabwe's position, Morocco successfully reformed, while the Democratic Republic of the Congo and Uganda stagnated or declined.

[8] This quartile score is different from the quartiles of durable improvements; they are measures of the full distribution, including non-improvers.

DEMOCRATIC ELECTIONS PATHWAYS

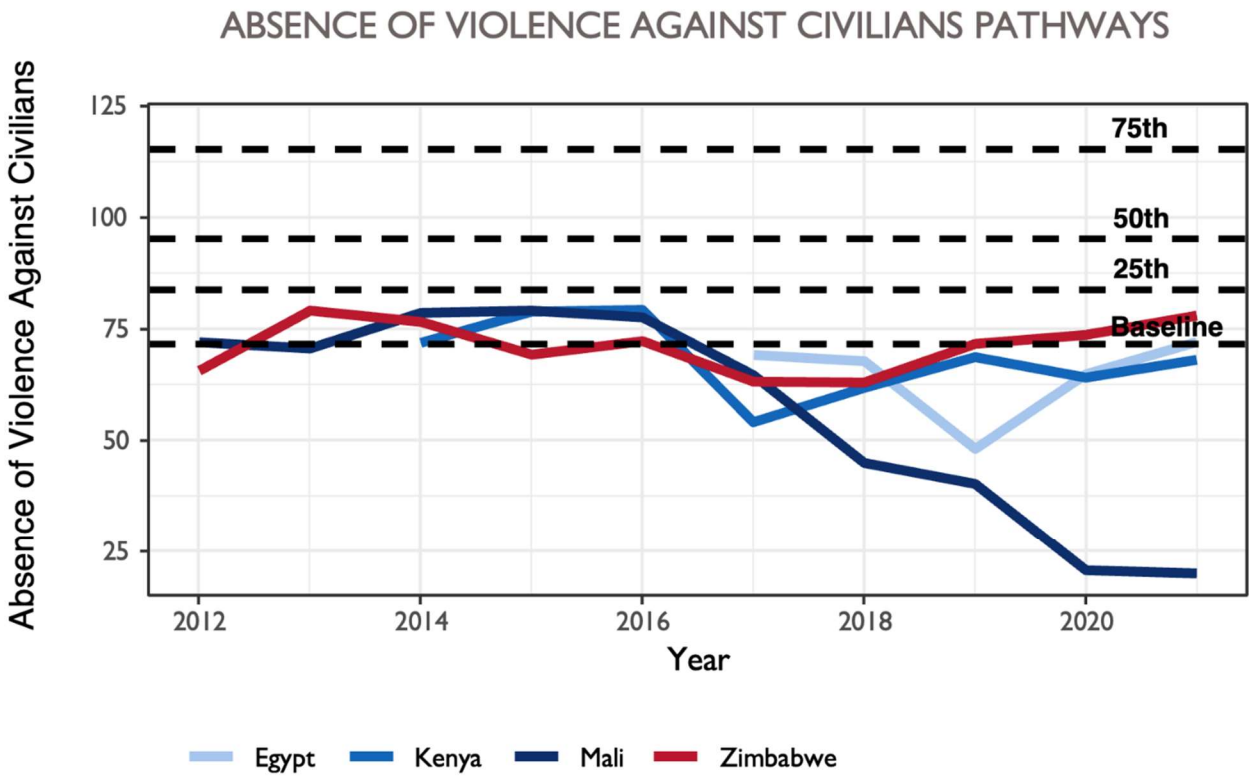
Figure 5: Pathways of Countries from the Year in Which They Matched Zimbabwe's 2019 Score.



Zimbabwe's history since 2012 is shown in red, while three selected similar countries are shown in blue. These example countries begin their paths at the point in time in which they were equivalent to Zimbabwe in 2019. Horizontal dotted lines represent levels that would represent the 25th, 50th, and 75th percentile level of improvement from Zimbabwe's 2019 baseline, as calculated in Table 5.

Zimbabwe's best-performing indicator was the absence of violence against civilians, with a 2019 score of 71.7 (top 67th percentile of countries). Zimbabwe's 2019 situation resembled Egypt in 2017, Kenya in 2014, or Mali in 2012. As seen in Figure 6, Kenya and Mali represent two possible, diverging pathways for the future of Zimbabwe. While Kenya has maintained its low levels of violence, the security situation in Mali deteriorated dramatically over the past six years. While the team recommends using the 50th or 75th percentile as the improvement targets for the other three indicators, for the absence of violence against civilians, the team recommends setting a target of either maintaining its 2019 baseline value or using the 25th percentile of improvement. Comparable countries with this high of a level of development show that they have little room to improve, although they remain at risk of decline.

Figure 6: Pathways of Countries from the Year in Which They Matched Zimbabwe's 2019 Score.



Zimbabwe's history since 2012 is shown in red, while three selected similar countries are shown in blue. These example countries begin their paths at the point in time in which they were equivalent to Zimbabwe in 2019. Horizontal dotted lines represent levels that would represent the 25th, 50th, and 75th percentile level of improvement from Zimbabwe's 2019 baseline, as calculated in Table 5.

RQ2: WHAT REAL-WORLD EVENTS AND POLICIES LEAD TO CHANGES IN THESE INDICATORS?

KEY FINDINGS

The team finds that event data has a limited ability to predict changes in IIAG scores. However, an exploratory analysis suggests several interesting patterns, including increased reporting on corruption in periods of opening civic space and major surges in civic activism in response to election-related government repression.

To understand what real-world events and policies led to changes in the identified indicators, the team identified patterns between major shocks detected in the MLP event data and the presence (or absence)

of durable improvement episodes on IIAG indicators. The team then used GPT-4 models to summarize thousands of articles reporting on events in country-years that resembled the patterns detected in the cross-national analysis. Synthesis of this qualitative information gives a sense of the type of events that tend to happen more and less frequently when countries are experiencing a durable improvement episode. In doing so, this builds confidence in the IIAG by showing that the major events accompanying these changes appear to be meaningful in their implications for democracy. For example, although the positive correlation between durable improvements in civil society space and spikes in reporting on corruption is not immediately intuitive, the AI-assisted qualitative analysis suggests that opening civic space can clear the way for increased awareness of and accountability for corrupt practices. Similarly, the team sees that large spikes in activism are negatively associated with durable improvements on democratic elections. Again, the AI-assisted synthesis suggests that the biggest instances of civic mobilizations detected in the MLP data are those focused on resisting the types of repression that often precede elections and contesting results in the aftermath of elections. While these results are compelling, **the weak predictive performance of the linear probability models should encourage caution in the use of these patterns to benchmark whether or not meaningful improvements are occurring during periods for which IIAG data is unavailable.**

One major advantage of the MLP data is that it is updated every ninety days and is accompanied by quantitative forecasts that predict when major political events are likely to occur over the next six months. These data can be used to monitor major political events in Zimbabwe on an ongoing basis as USAID waits for updated IIAG data. For example, in the four years after 2019, Zimbabwe has seen a decline in both the maximum number and the average number of shocks per year for MLP's measures of political cooperation, government purges, political threats, government raids, protests, and non-lethal violence; alternatively, the team has seen increases in the prevalence of activism and lethal violence. Relating these trends back to the patterns found in the cross-national analysis, the team found a decrease in one event (cooperation) that is negatively associated with durable improvements in democratic elections and an increase in one event (activism) that is negatively associated with durable improvements in democratic elections. The next update of MLP data and forecasts for Zimbabwe is expected in late February 2024.

To analyze how changes in governance indicators correspond with real-world events, the research team utilized data from the MLP dataset. MLP produces monthly data for 59 countries from 2012–2024, tracking 20 types of domestic political events bearing on civic space. MLP's unique database builds continuously on a repository of nearly 100 million news articles capturing daily coverage from more than 350 online newspapers (including five newspapers based in Zimbabwe and two regional African sources) published in more than 35 languages. MLP uses large language models to process huge volumes of news into a rich array of event types along with scraping and parsing tools customized to ensure accurate, comprehensive capture of news published by a highly curated list of high-quality sources.

For this project, the team relied on MLP's measures tracking monthly levels of reporting by local media outlets on civic space events. To identify months in which major events took place, referred to as shocks, the team trained an ensemble of algorithms to detect major spikes in reporting on each event. These algorithms were trained to define shocks as spikes in reporting that happen no more than 12 times over

a 12-year period.[9] To match this monthly data with IIAG's annual indicators, the team counted the number of months during which each country experienced a shock across all 20 indicators. Figure 9 in Appendix 3 visualizes the number of months with shocks across each event category.

Taking this measure of the number of months per year with major shocks, the team used simple linear probability models to assess the correlation between the number of months in each year with shocks and whether or not each year was a year of no improvement or a year within a durable improvement episode on the relevant IIAG indicator.[10] Figure 7 reports the results. **Ultimately, measures of model fit indicate that movements in the MLP event data cannot reliably predict changes in IIAG scores. Given the relatively small sample size and the conceptual difficulties inherent in comparing data collected at different frequencies, this is disappointing but not surprising. However, the team did observe a statistically significant and substantively meaningful negative relationship between the frequency of several major political events and durable improvements in IIAG scores. While it is difficult to determine whether these events could serve as a reliable predictor of changes in governance scores outside of this sample, the large magnitude of the coefficients and the qualitative investigation that follows provide some reason for optimism.**

[9] A frequency of once per year was used to guide training of the shock detection algorithms for several reasons. Most importantly, traditional statistical approaches to anomaly detection yielded an average of 1.5 –2 shocks per year across event types. Qualitative validation efforts suggested that the type of events detected were often less significant than was desired. Additional details are available in MLP's Pipeline Report at mlpeace.org.

[10] The team excluded years during which there is an improvement but that are not contained within a durable improvement episode. This is intended to ensure that the team is comparing years of stagnation or deterioration against years that are part of a concerted reform effort.

Figure 7: Linear Probability Model Regressing IIAG Indicators on the Count of Months with Shock across MLP Event Measures

	Absence of Violence against Civilians	Civil Society Space	Democratic Elections	Impartiality of the Judicial Sys
Censorship	0.004 (0.657)	-0.011 (0.542)	-0.025 (0.471)	0.016 (0.643)
Defamation Case	0.004 (0.652)	-0.001 (0.981)	0.038 (0.354)	-0.035 (0.332)
Election Activity	-0.013+ (0.083)	0.003 (0.843)	0.022 (0.564)	0.045 (0.179)
Election Irregularities	0.005 (0.458)	0.001 (0.974)	-0.014 (0.744)	-0.024 (0.536)
Legal Action	-0.023* (0.042)	0.008 (0.701)	-0.011 (0.771)	0.043 (0.221)
Legal Change	0.000 (0.958)	0.010 (0.614)	0.038 (0.260)	-0.009 (0.790)
Lethal Violence	-0.008 (0.385)	0.004 (0.862)	0.027 (0.457)	0.049 (0.158)
Nonlethal Violence	0.002 (0.912)	-0.011 (0.386)	-0.038 (0.271)	-0.029 (0.397)
Disaster	-0.002 (0.647)	-0.011 (0.207)	-0.022 (0.362)	-0.003 (0.887)
Corruption	0.013 (0.275)	0.031+ (0.096)	0.026 (0.513)	-0.041 (0.239)
Arrest	0.011 (0.340)	-0.011 (0.491)	-0.030 (0.355)	0.002 (0.944)
Security Mobilization	-0.009 (0.182)	-0.011 (0.413)	0.008 (0.807)	-0.023 (0.487)
State of Emergency	-0.001 (0.887)	0.018 (0.241)	0.006 (0.845)	0.012 (0.675)
Raid	0.014 (0.412)	0.000 (0.983)	-0.034 (0.377)	0.002 (0.965)
Protest	0.000 (0.990)	0.000 (0.994)	0.044 (0.207)	-0.034 (0.237)
Activism	0.002 (0.789)	-0.009 (0.507)	-0.051+ (0.077)	-0.007 (0.830)
Cooperation	0.014 (0.263)	-0.004 (0.827)	-0.056+ (0.083)	0.027 (0.426)
Purge	-0.013 (0.176)	0.003 (0.904)	0.022 (0.549)	-0.008 (0.831)
Threats	0.008 (0.667)	0.012 (0.580)	0.047 (0.202)	-0.021 (0.563)
Irregular Transition	0.011 (0.389)	-0.024 (0.202)	-0.076+ (0.068)	0.026 (0.496)
Num.Obs.	198	180	167	182
R2	0.072	0.059	0.102	0.064

Models use HC1 robust standard errors.

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

For example, one additional month with a spike in election activity or legal actions decreases the probability that a country is experiencing a durable improvement in the absence of violence against civilians indicator by 1.3 percent and 2.3 percent, respectively. For countries with particularly high levels of activity on these indicators, which experienced up to 7 and 6 months of shocks on these event categories respectively, this results in a 9.1 percent and a 13.8 percent reduction in the probability that the country is experiencing a durable improvement in the absence of violence against civilians indicator. Of the six MLP measures with a statistically significant correlation with an IIAG indicator, these two have the smallest coefficients. Because the correlation between these six indicators and the IIAG indicators is relatively

large in magnitude, the team used them to indicate the type of events that are associated with changes in the IIAG indicators in the data and explore their potential to serve as early indicators of a likely change in IIAG scores.

Figure 8: Significant Coefficients from Linear Probability Model Regressing IIAG Indicators on the Count of Months with Shock across MLP Event Measures

Correlation between IIAG durable-improvement years and count of months with major MLP events

Direction indicates whether the relationship is positive or negative

IIAG	MLP	Direction
Absence of Violence against Civilians	Election Activity	-
Absence of Violence against Civilians	Legal Action	-
Civil Society Space	Corruption	+
Democratic Elections	Activism	-
Democratic Elections	Cooperation	-
Democratic Elections	Irregular Transition	-

Figure 8 highlights the specific MLP measures that have a significant correlation with one of the IIAG indicators. To further illustrate the real-world context behind these correlations, Figure 10 in Appendix 4 presents the two country-years (more country-years are presented when there is a tie across country-years) with the highest level of activity on the MLP measures while having IIAG indicator scores that correspond with the direction of the coefficient (non-improvement years for negative coefficients and durable improvement years for positive coefficients). For each country-year in Figure 10, the team identified the specific months when shocks occurred and used OpenAI's GPT-4-0125-preview model to summarize the thousands of articles from domestic news outlets in the MLP database that were reporting on the underlying relevant events. Although these AI-generated summaries are too long to include in this

document due to the large number of distinct events that occurred, the team synthesized the information contained in these summaries in the paragraphs below.[11]

ACTIVISM AND DEMOCRATIC ELECTIONS IN ANGOLA, GHANA, AND RWANDA

Spikes in activism are more prevalent in countries that *are not* experiencing a durable improvement in IIAG's Democratic Elections indicator. Consistent with the pattern detected in the data, of the three country-years with the highest levels of activism in the MLP data, the activism activities in two of these country-years were focused squarely on issues relevant to elections.

Angola experienced extremely high levels of activism throughout 2015. This activism often took the form of conflict between political movements and resistance against government repression of activists. In particular, the majority of activism reported during this period centered around mobilization against the government's legal persecution of a high-profile group of political activists, which eventually resulted in their release to house arrest. Although these events were not directly associated with a change in elections, they marked a repressive turn by the incumbent government that shaped the 2017 elections.

Ghana experienced extremely high levels of activism throughout 2013. Much of this activism was centered around a legal battle challenging the results of the 2012 presidential election, including demands on the Electoral Commission to produce the collation for several constituencies and an eventual Supreme Court decision dismissing the case. Rwanda experienced high levels of activism throughout 2018. In contrast to Ghana and Angola, this activism did not appear to concentrate on contentious political issues and was largely focused on service delivery.

POLITICAL COOPERATION AND DEMOCRATIC ELECTIONS IN TUNISIA AND MOZAMBIQUE

Surprisingly, spikes in political cooperation are also more prevalent in countries that are not experiencing a durable improvement in IIAG's Democratic Elections indicator. However, dynamics in the two country-years with the highest levels of political cooperation in the data show that the cooperation being detected is often motivated by a broader political crisis.

Tunisia experienced extremely high levels of political cooperation in 2013 after a major political crisis. The major events being reported on included attempts to create new political parties and coalitions, contentious efforts to form a new government, and months of negotiation over the willingness of different political groups to participate in a national dialogue. These events transpired as Tunisia was experiencing multiple delays in elections that were initially promised by the ruling parties in 2011. Mozambique experienced extremely high levels of political cooperation in 2016. Much of this cooperation took the form of attempts at negotiating a peace agreement between the country's two major political parties after

[11] This report does not include a discussion of the reporting on election activities and irregular transitions because their association with the IIAG indicators is straightforward. Specifically, the election activities measure captures reporting on things like election campaigns and administration. The negative association between election activities and the absence of violence against civilians indicator suggests that governments are more likely to engage in violence during the periods surrounding elections. Similarly, the irregular transitions measure captures reporting on changes in government authority that are not in the rules of succession or transition, such as military coups or plots to remain in a government office after losing an election or through changes to the rules or constitution. In the data, this includes the 2013–2014 Tunisian political crisis and Algerian President Bouteflika's 2019 resignation after publicly losing the support of the military. The negative association between irregular transitions and the democratic elections indicator is not surprising.

contested election results led to conflict between government forces and militias associated with the opposition.

CORRUPTION AND CIVIL SOCIETY SPACE IN ANGOLA AND SOUTH AFRICA

In contrast to the last two categories, spikes in reporting on corruption are more prevalent in countries that are experiencing a durable improvement in IIAG's civil society space indicator. South Africa and Angola both experienced extremely high levels of reporting on corruption in 2018. Both countries had major corruption scandals involving top government officials and business leaders. For Angola, these major scandals occurred the year after a new government took office and reduced restrictions on non-governmental organizations, paving the way for more freedom in the ability of civil society to expose corruption.

LEGAL ACTIONS AND ABSENCE OF VIOLENCE AGAINST CIVILIANS IN ALGERIA, MOZAMBIQUE, AND RWANDA

Spikes in legal actions are more prevalent in countries that are not experiencing a durable improvement in IIAG's absence of violence against civilians indicator. Algeria and Mozambique experienced extremely high levels of legal action in 2019. In Algeria, the vast majority of this activity focused on actions related to corruption charges. In Mozambique, much of this activity was focused on a major corruption scandal involving the country's international debt, but there were also important actions in response to partisan violence and the murder of an activist, legal challenges of voter registration, and criminal charges brought against several journalists. Rwanda experienced extremely high levels of legal action in 2020. Legal actions in Rwanda were more focused on the actions of former militia members and military leaders.

RQ 3: WHAT ARE THE POTENTIAL LIMITATIONS/WEAKNESSES OF THESE INDICATORS? WHAT ARE THE POSSIBLE SOLUTIONS TO THOSE LIMITATIONS/WEAKNESSES?

KEY FINDINGS

The WJP's index combines a QRQ, updated annually, a GPP, updated irregularly, and a PTS index, updated annually. Due to its irregular updates and changing sample, the GPP is vulnerable to major composition effects and renders the aggregate WJP indicators an unreliable measure of changes in governance over time. WJP measures should be reconstructed without GPP indicators.

Both the WJP and IIAG indicators rely on subjective measures of governance based on expert opinion. Little and Meng (2023) raise concerns that expert opinions may overestimate changes as a reaction to major political events. To mitigate this, the team recommends pairing analyses of IIAG and WJP data with event data to identify major events that may influence subjective scores.

As noted above, the WJP indicators are vulnerable to composition changes due to the introduction of new GPP data in 2019 and major changes in GPP sampling between periods. WJP indicators also appear

more susceptible to manipulation due to a heavy reliance on in-country samples and a lower level of transparency about data collection. For this reason, the team recommends that USAID focus on IIAG data in establishing targets for ZDERA criteria.

In the future, the team hopes that WJP will provide access to disaggregated data that would allow researchers to separately analyze the QRQ and GPP data. Analyzing data from these two sources separately would allow researchers to understand how changes in the composition of the data affect the changes in WJP indicator scores over time. For Zimbabwe, this would enable the assessment of the extent to which the massive improvements in GoZ scores after 2019, which are not present in the IIAG data, are driven by the changes in the sampling between the 2016 and 2018 GPP survey and the unfortunate timing of the 2018 GPP, which coincided with a coup that may have affected public opinion in ways that overstated the magnitude of the country's improvement. By including results from the 2018 GPP in the indicator scores for every year after 2018, the WJP methodology exacerbates these problems.

Importantly, both the WJP and IIAG indicators rely on subjective measures of governance based on expert opinion. Little and Meng (2023) raise concerns that subjective indicators are vulnerable to changes in coder bias over time and may overestimate changes to national indicators as a result. For example, major political events may change coders' beliefs about the governance trajectory of their country and encourage "motivated beliefs" that influence the interpretation of future events. While subjective measures describe a worldwide democratic decline, Little and Meng (2023) show that "objective" measures based on directly observable criteria, such as the alteration of political power between parties, show no such widespread decline.

While these results should encourage caution when relying exclusively on subjective data, "objective" measures of many characteristics of governance are not available, making subjective measures indispensable. Recent work by Baron et al. (2023) encourages researchers to use both subjective data and event data in tandem to assess changes in governance over time. Using event data from the Democratic Erosion Event Dataset for Turkey and Brazil, their results demonstrate that the main dataset used to create IIAG—V-Dem—produces estimates of democratic erosion that reflect (though slightly overstates) trends in event data that are missed by the objective indicators used by Little and Meng (2023).

These findings increase the team's confidence that IIAG indicators provide a useful metric for ZDERA criteria. To further bolster the analysis, the team used event data from the MLP dataset to explore patterns between durable improvements captured in IIAG and major political events. The team used MLP data because the Democratic Erosion Event Dataset is only available through 2020 and relies on student-coded data collected from a limited number of predominantly English news sources while MLP utilizes machine coding to process millions of articles from domestic news outlets publishing in dozens of languages. To provide a deeper understanding of these dynamics, the team then used a GPT-4 model to summarize thousands of articles reporting on relevant political events in countries that fit the patterns found in the cross-national analysis. Through this exercise, the patterns between these data sources reinforced the team's confidence in their suitability for this analysis.

RECOMMENDATIONS

The team concludes that WJP indicators *do not reliably* assess changes in GoZ's governance. However, a review of the methodology as well as a validation exercise using event data from the MLP dataset suggests that IIAG's indicators can be relied on to assess GoZ's performance. Specifically, the team recommends targeting improvements in the 50th or 75th percentile of observed changes across the IIAG indicators, civil society space, the impartiality of the judicial system, and democratic elections. For the fourth IIAG indicator (absence of violence against civilians), Zimbabwe's 2019 value is at a high enough level that one would not expect to see as large of an improvement. Zimbabwe's score, however, has decreased over the past two years. For this indicator, the research team recommends targeting a return to Zimbabwe's 2019 value.

REFERENCES

Baron, Hannah, Robert A Blair, Jessica Gottlieb, and Laura Paler. 2023. "An Events-Based Approach to Understanding Democratic Erosion." *PS: Political Science & Politics*, 1–8.

Boese, Vanessa A, and Markus Eberhardt. 2021. "Democracy Doesn't Always Happen over Night: Regime Change in Stages and Economic Growth."

Little, Andrew T, and Anne Meng. 2023. "Measuring Democratic Backsliding." *PS: Political Science & Politics*, 1–13.

Maerz, Seraphine F, Amanda Edgell, Matthew C Wilson, Sebastian Hellmeier, and Staffan I Lind- berg. 2021. "A Framework for Understanding Regime Transformation: Introducing the ERT Dataset." *V-Dem Working Paper* 113.

Ponce, Alejandro. 2015. "World Justice Project Rule of Law Index® 2015."

"World Justice Project Rule of Law Index® 2017-2018."

APPENDIX

BOOTSTRAP FOR MEDIAN VALUES

Table 7: 10 Percent Improvement and Three-Year Duration

SOURCE	INDICATOR	25TH QUARTILE	50TH QUARTILE	75TH QUARTILE
IIAG	Civil Society Space	0.19 (0.17 0.21)	0.28 (0.22 0.37)	1.14 (0.49 2.4)
IIAG	Judicial System Impartiality	0.25 (0.22 0.3)	0.48 (0.41 0.55)	0.9 (0.78 1.05)
IIAG	Democratic Elections	0.21 (0.19 0.26)	0.35 (0.29 0.41)	0.63 (0.54 0.81)
IIAG	Absence of Civilian Violence	0.16 (0.15 0.18)	0.23 (0.17 0.34)	0.57 (0.34 1.02)
WJP	Official Misconduct Sanction	0.16 (0.14 0.18)	0.21 (0.18 0.23)	0.26 (0.24 0.32)
WJP	Lawful Transition of Power	0.16 (0.15 0.18)	0.22 (0.2 0.24)	0.31 (0.28 0.4)
WJP	Right to Life	0.17 (0.15 0.18)	0.21 (0.2 0.26)	0.35 (0.31 0.4)
WJP	Due Process	0.13 (0.12 0.14)	0.17 (0.16 0.18)	0.21 (0.19 0.25)
WJP	Freedom of Opinion	0.15 (0.12 0.18)	0.21 (0.17 0.28)	0.37 (0.27 0.52)
WJP	Freedom of Association	0.14 (0.12 0.18)	0.2 (0.16 0.26)	0.3 (0.23 0.51)
WJP	Private Property	0.18 (0.16 0.2)	0.23 (0.22 0.25)	0.29 (0.27 0.32)

This table includes point estimates and confidence intervals for the 25th, 50th, and 75th percentiles using bootstrapping. The values represent the point estimates for each indicator's percentiles of improvement. In parentheses are the corresponding indicator's 95 percent confidence intervals.

Table 8: 5 Percent Improvement and Three-Year Duration: Quartiles of Improvement

SOURCE	INDICATOR	25TH QUARTILE	50TH QUARTILE	75TH QUARTILE
IIAG	Civil Society Space	0.14 (0.1 0.17)	0.21 (0.18 0.27)	0.5 (0.32 1.49)
IIAG	Judicial System Impartiality	0.17 (0.14 0.21)	0.36 (0.31 0.46)	0.82 (0.66 0.93)
IIAG	Democratic Elections	0.17 (0.14 0.21)	0.29 (0.27 0.36)	0.58 (0.46 0.7)
IIAG	Absence of Civilian Violence	0.08 (0.07 0.09)	0.12 (0.11 0.15)	0.24 (0.17 0.37)
WJP	Official Misconduct Sanction	0.09 (0.07 0.1)	0.14 (0.11 0.16)	0.22 (0.19 0.25)
WJP	Lawful Transition of Power	0.1 (0.08 0.11)	0.16 (0.14 0.19)	0.24 (0.22 0.3)
WJP	Right to Life	0.11 (0.09 0.13)	0.18 (0.16 0.2)	0.29 (0.23 0.34)

SOURCE	INDICATOR	25TH QUARTILE	50TH QUARTILE	75TH QUARTILE
WJP	Due Process	0.1 (0.09 0.11)	0.13 (0.12 0.16)	0.19 (0.17 0.21)
WJP	Freedom of Opinion	0.1 (0.08 0.12)	0.17 (0.12 0.2)	0.28 (0.21 0.38)
WJP	Freedom of Association	0.1 (0.09 0.12)	0.14 (0.12 0.17)	0.22 (0.19 0.29)
WJP	Private Property	0.1 (0.08 0.11)	0.16 (0.14 0.19)	0.25 (0.22 0.27)

This table includes point estimates and confidence intervals for the 25th, 50th, and 75th percentiles using bootstrapping. The values represent the point estimates for each indicator's percentiles of improvement. In parentheses are the corresponding indicator's 95 percent confidence intervals.

Table 9: 5 Percent Improvement and Five-Year Duration: Quartiles of Improvement

SOURCE	INDICATOR	25TH QUARTILE	50TH QUARTILE	75TH QUARTILE
IIAG	Civil Society Space	0.17 (0.1 0.21)	0.25 (0.2 0.49)	2.4 (0.49 8)
IIAG	Judicial System Impartiality	0.19 (0.16 0.25)	0.43 (0.34 0.51)	0.88 (0.71 1.05)
IIAG	Democratic Elections	0.22 (0.2 0.27)	0.37 (0.31 0.44)	0.71 (0.58 1.12)
IIAG	Absence of Civilian Violence	0.08 (0.07 0.1)	0.14 (0.11 0.16)	0.28 (0.16 0.57)
WJP	Official Misconduct Sanction	0.1 (0.1 0.13)	0.16 (0.14 0.19)	0.25 (0.22 0.27)
WJP	Lawful Transition of Power	0.1 (0.09 0.12)	0.17 (0.15 0.2)	0.29 (0.23 0.37)
WJP	Right to Life	0.12 (0.1 0.15)	0.18 (0.16 0.21)	0.3 (0.24 0.36)
WJP	Due Process	0.11 (0.1 0.12)	0.16 (0.13 0.17)	0.21 (0.18 0.23)
WJP	Freedom of Opinion	0.1 (0.08 0.12)	0.15 (0.12 0.21)	0.28 (0.21 0.41)
WJP	Freedom of Association	0.11 (0.11 0.14)	0.16 (0.13 0.22)	0.27 (0.19 0.51)
WJP	Private Property	0.12 (0.1 0.16)	0.2 (0.17 0.22)	0.27 (0.23 0.3)

This table includes point estimates and confidence intervals for the 25th, 50th, and 75th percentiles using bootstrapping. The values represent the point estimates for each indicator's percentiles of improvement. In parentheses are the corresponding indicator's 95 percent confidence intervals.

PERCENTAGE AND NUMBER OF COUNTRIES THAT EXPERIENCED DURABLE IMPROVEMENT

Table 10: Indicators

INDICATOR	10% FIVE YEARS	10% THREE YEARS	5% FIVE YEARS	5% THREE YEARS
Absence of Civilian Violence	7.4% (4)	22.2% (12)	20.4% (11)	38.9% (21)
Civil Society Space	11.1% (6)	33.3% (18)	13% (7)	44.4% (24)
Democratic Elections	38.9% (21)	66.7% (36)	48.1% (26)	74.1% (40)
Due Process	13.32% (12)	21.09% (19)	21.09% (19)	34.41% (31)
Freedom of Association	6.66% (6)	9.99% (9)	8.88% (8)	18.87% (17)
Freedom of Opinion	5.55% (5)	13.32% (12)	11.1% (10)	17.76% (16)
Judicial System Impartiality	44.4% (24)	59.3% (32)	50% (27)	70.4% (38)
Lawful Transition of Power	8.88% (8)	25.53% (23)	21.09% (19)	36.63% (33)
Official Misconduct Sanction	9.99% (9)	14.43% (13)	16.65% (15)	29.97% (27)
Right to Life	13.32% (12)	24.42% (22)	19.98% (18)	34.41% (31)
Private Property	13.32% (12)	19.98% (18)	18.87% (17)	32.19% (29)

The percentage of countries who achieved durable improvement for each indicator under the conditions of 10 percent for five years, 10 percent for three years, 5 percent for five years, and 5 percent for three years.

BOOTSTRAPPED INDICATOR VALUES WITHOUT OUTLIERS REMOVED

Table 11: Quartiles of Improvement

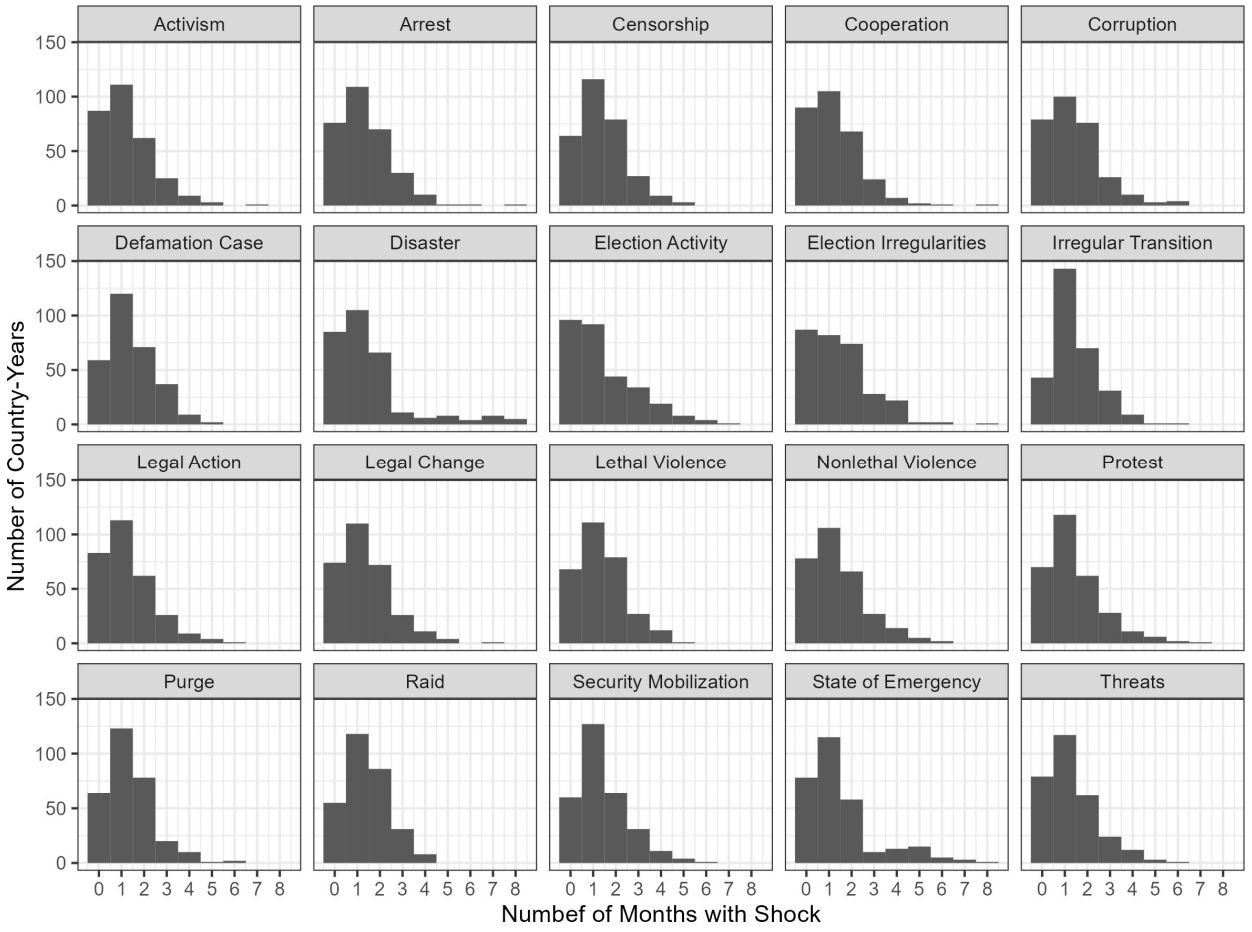
SOURCE	INDICATOR	25TH QUARTILE	50TH QUARTILE	75TH QUARTILE
IIAG	Civil Society Space	0.2 (0.17 0.28)	0.43 (0.23 2.38)	2.71 (0.93 8.4)
IIAG	Judicial System Impartiality	0.26 (0.24 0.31)	0.51 (0.43 0.58)	0.95 (0.83 1.08)
IIAG	Democratic Elections	0.28 (0.23 0.31)	0.45 (0.4 0.55)	1.12 (0.68 1.57)
IIAG	Absence of Civilian Violence	0.17 (0.13 0.29)	0.33 (0.19 0.59)	0.61 (0.37 0.96)
WJP	Official Misconduct Sanction	0.17 (0.16 0.19)	0.23 (0.19 0.25)	0.28 (0.25 0.33)
WJP	Lawful Transition of Power	0.21 (0.16 0.24)	0.29 (0.23 0.38)	0.45 (0.37 0.51)

SOURCE	INDICATOR	25TH QUARTILE	50TH QUARTILE	75TH QUARTILE
WJP	Right to Life	0.16 (0.15 0.18)	0.21 (0.19 0.26)	0.33 (0.26 0.42)
WJP	Due Process	0.14 (0.13 0.16)	0.18 (0.16 0.2)	0.24 (0.21 0.27)
WJP	Freedom of Opinion	0.21 (0.18 0.28)	0.3 (0.22 0.43)	0.56 (0.31 0.58)
WJP	Freedom of Association	0.15 (0.13 0.19)	0.21 (0.16 0.27)	0.31 (0.23 0.58)
WJP	Private Property	0.18 (0.17 0.21)	0.23 (0.22 0.27)	0.29 (0.27 0.31)

Point estimates and confidence intervals for the 25th, 50th, and 75th percentiles using bootstrapping. The values represent the point estimates for each indicator's percentiles of improvement. In parentheses are the corresponding indicator's 95 percent confidence intervals.

MLP SHOCK DETECTION

Figure 9: Distribution of MLP Shocks by Event Category



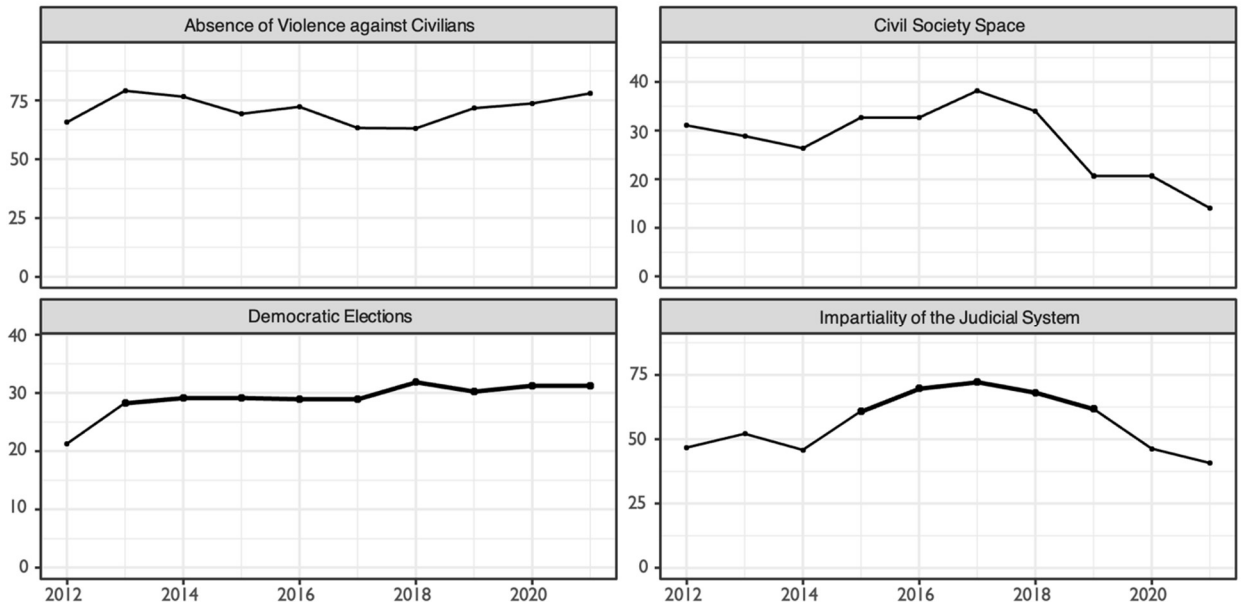
MLP HIGH ACTIVITY COUNTRY-YEARS

Figure 10: Country-Years with High MLP Event Activity and Corresponding IIAG Target Value

Highest Activity MLP Country-Years across IIAG Targets				
Country	Year	Months	MLP	IIAG
Ghana	2013	7	Activism	Democratic Elections
Angola	2015	5	Activism	Democratic Elections
Rwanda	2018	5	Activism	Democratic Elections
Tunisia	2013	8	Political Cooperation	Democratic Elections
Mozambique	2016	6	Political Cooperation	Democratic Elections
Tunisia	2013	6	Irregular Transition	Democratic Elections
Algeria	2019	4	Irregular Transition	Democratic Elections
Angola	2018	4	Corruption	Civil Society Space
South Africa	2018	4	Corruption	Civil Society Space
Mozambique	2019	6	Legal Actions	Absence of Violence against Civilians
Algeria	2019	5	Legal Actions	Absence of Violence against Civilians
Rwanda	2020	5	Legal Actions	Absence of Violence against Civilians
DR Congo	2018	7	Election Activity	Absence of Violence against Civilians
Cameroon	2018	6	Election Activity	Absence of Violence against Civilians
Tanzania	2015	6	Election Activity	Absence of Violence against Civilians
Tunisia	2014	6	Election Activity	Absence of Violence against Civilians

IIAG INDICATORS FOR ZIMBABWE

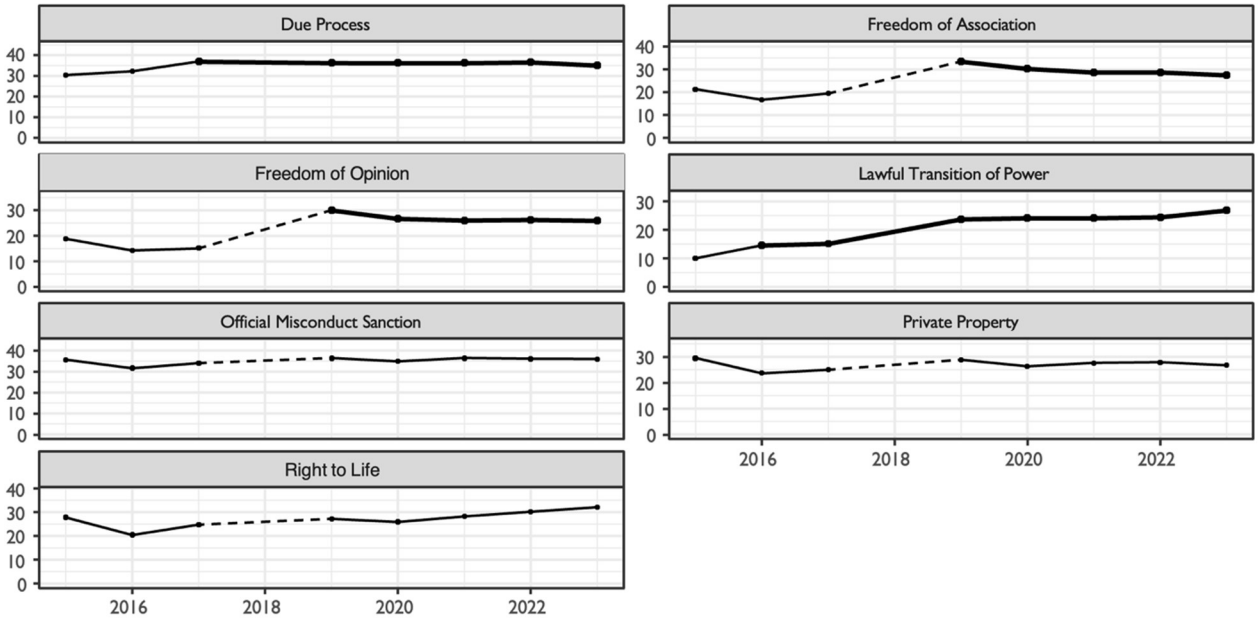
Figure 11: GoZ's Scores on Four IIAG Indicators Selected by USAID



Bold lines and larger points capture 10 percent or larger increases in an indicator value that last for at least five years. Bolding disappears when a sustained increase is reversed.

WJP INDICATORS FOR ZIMBABWE

Figure 12: GoZ's Scores on Seven WJP Indicators Selected by USAID

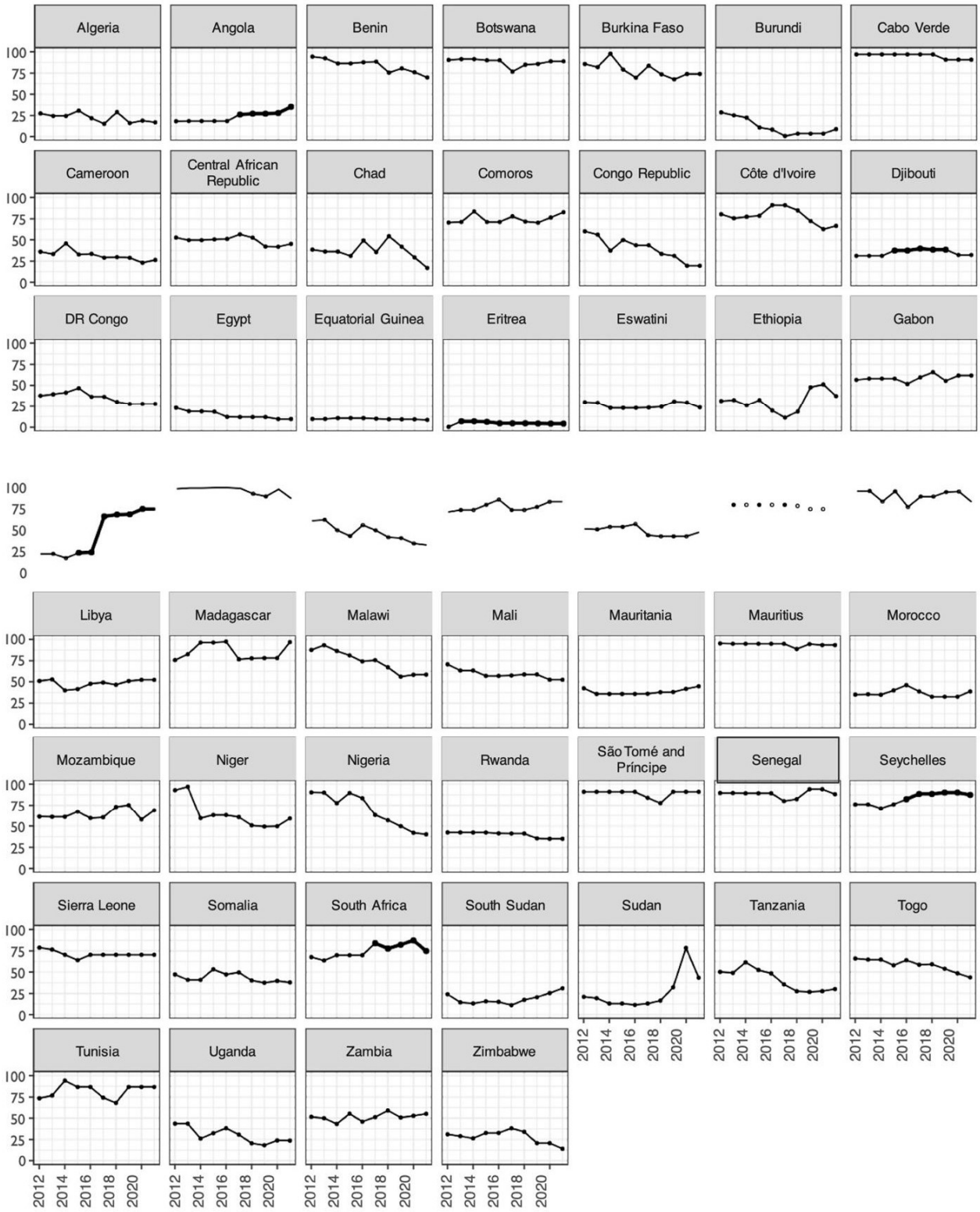


Dashed lines indicate linear interpolation of data for missing years. Bold lines and larger points capture 10 percent or larger increases in an indicator value that last for at least five years. Bolding disappears when a sustained increase is reversed.

IIAG INDICATORS FOR ALL COUNTRIES

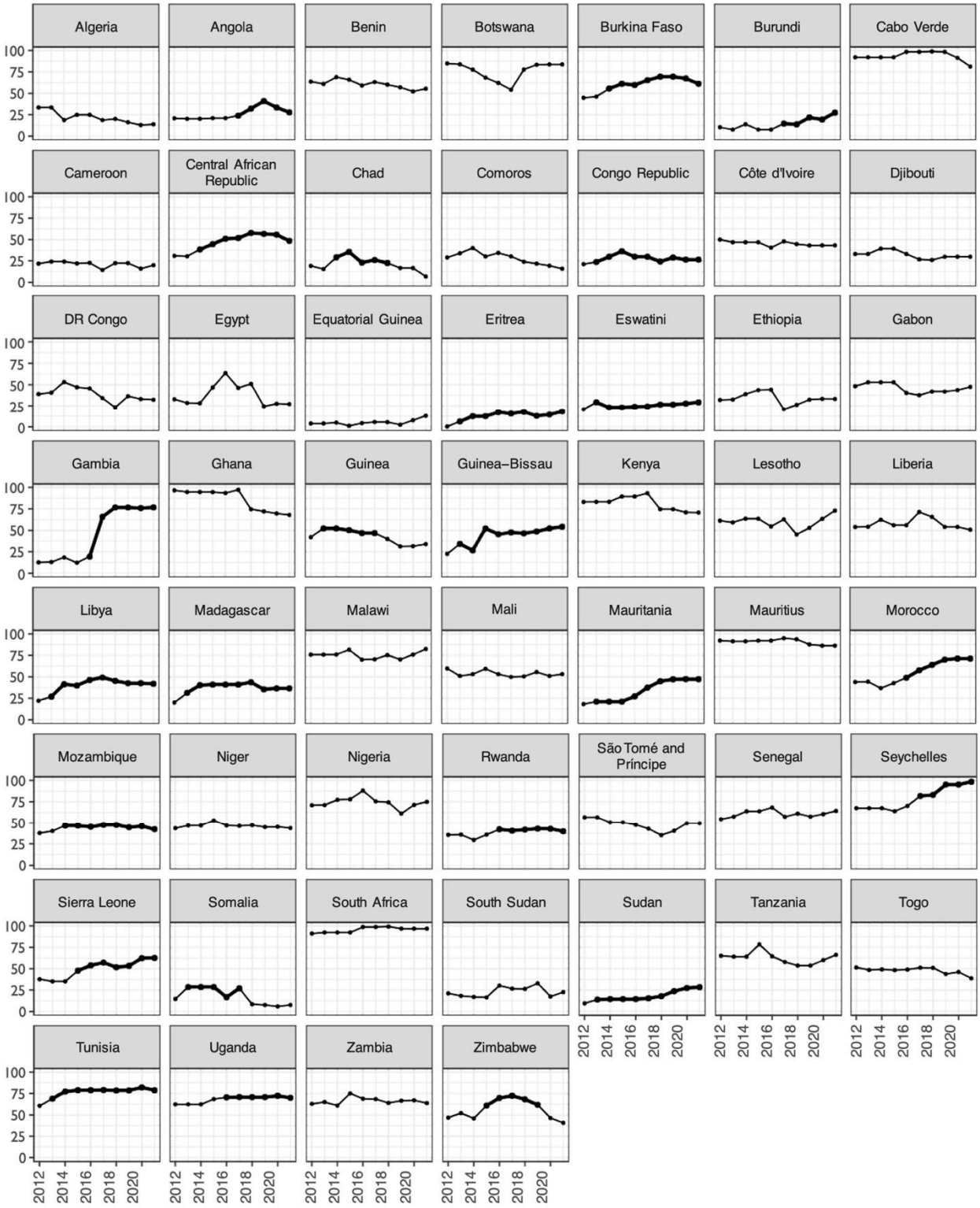
CIVIL SOCIETY SPACE

Figure 13: Country Scores on Four IIAG Indicators Selected by USAID



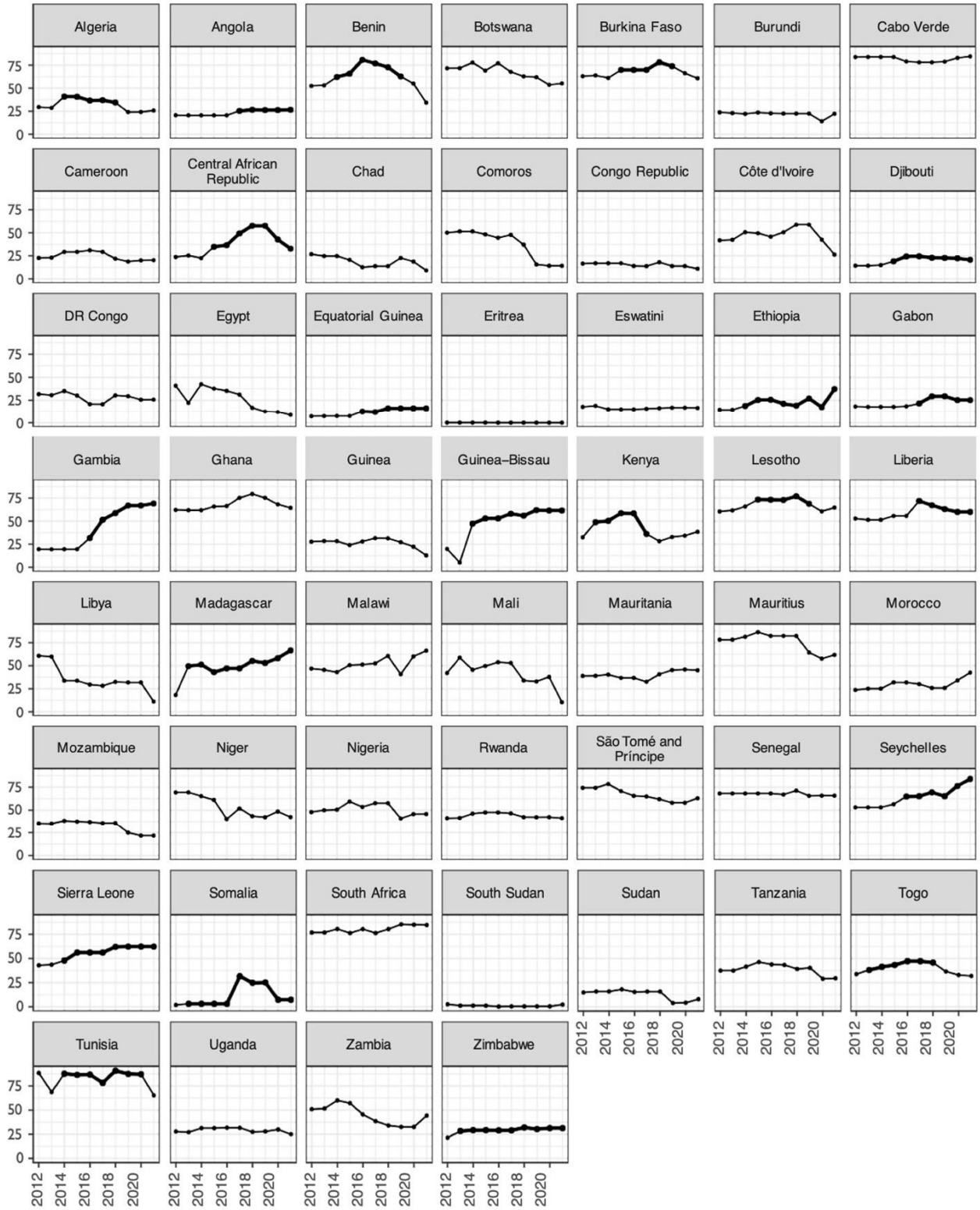
IMPARTIALITY OF THE JUDICIAL SYSTEM

Figure 14: Country Scores on Four IIAG Indicators Selected by USAID



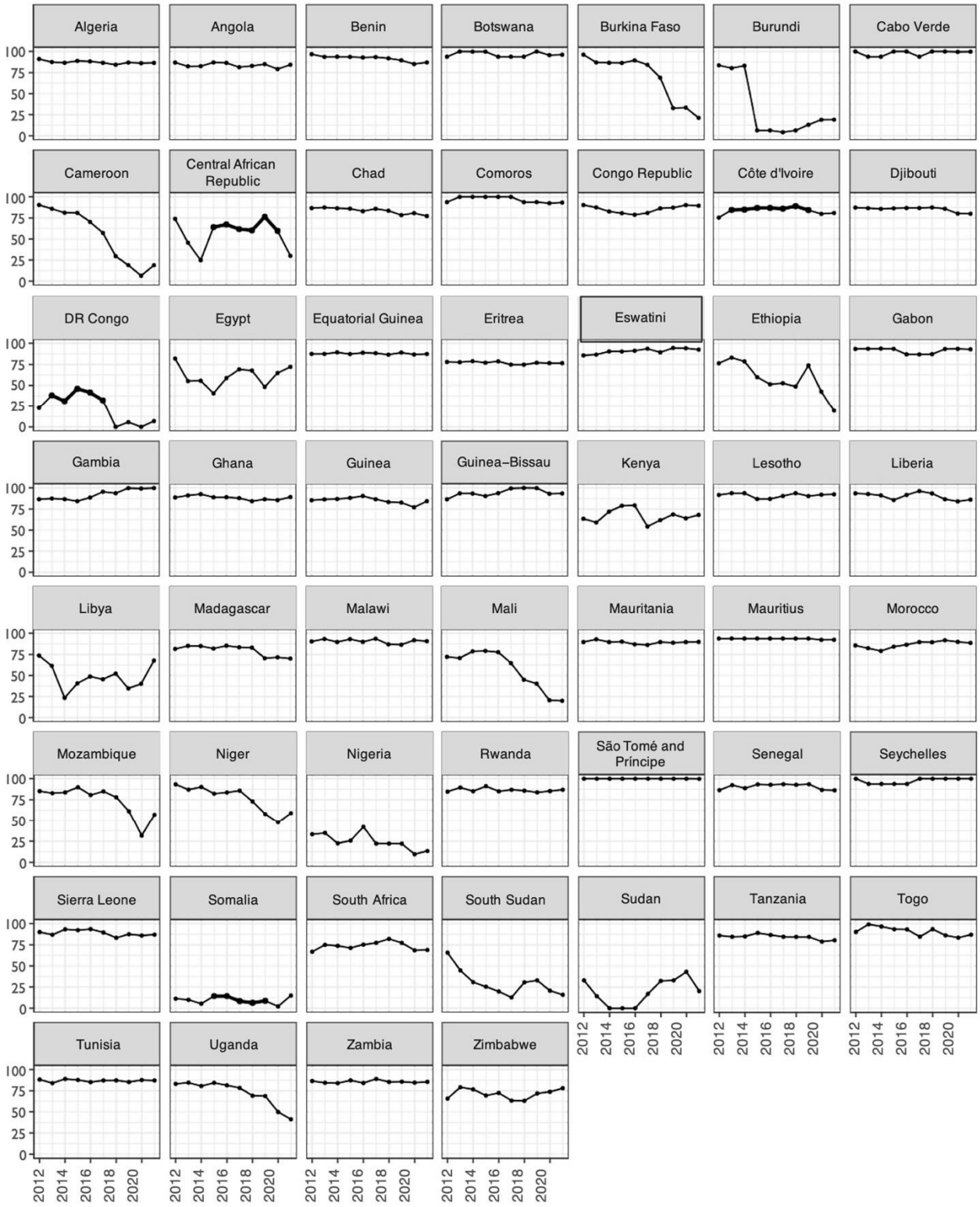
DEMOCRATIC ELECTIONS

Figure 15: Country Scores on Four IIAG Indicators Selected by USAID



ABSENCE OF VIOLENCE AGAINST CIVILIANS

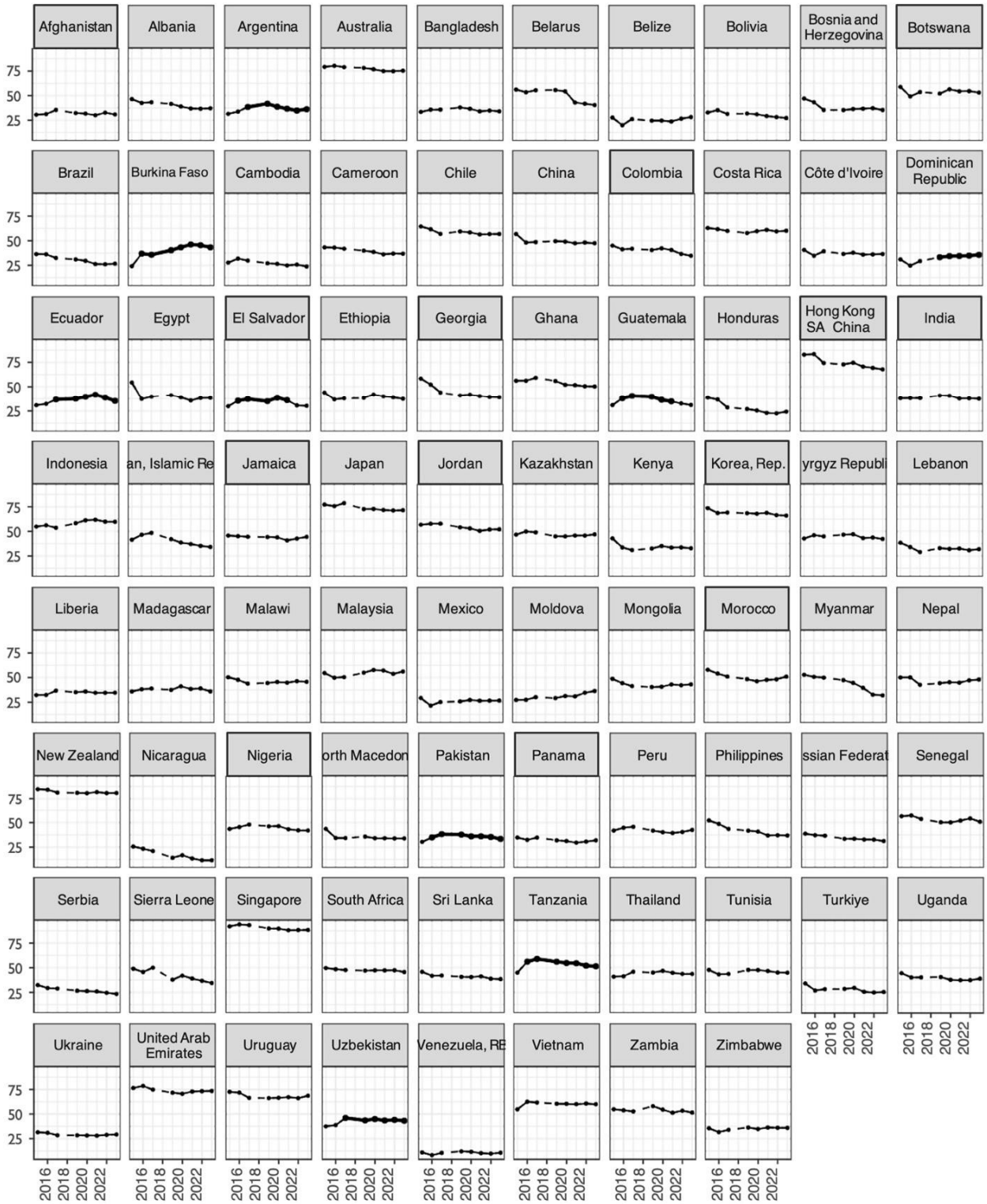
Figure 16: Country Scores on Four IIAG Indicators Selected by USAID



WJP INDICATORS FOR ALL COUNTRIES

OFFICIAL MISCONDUCT SANCTION

Figure 17: Country Scores on Seven WJP Indicators Selected by USAID



LAWFUL TRANSITION OF POWER

Figure 18: Country Scores on Seven WJP Indicators Selected by USAID

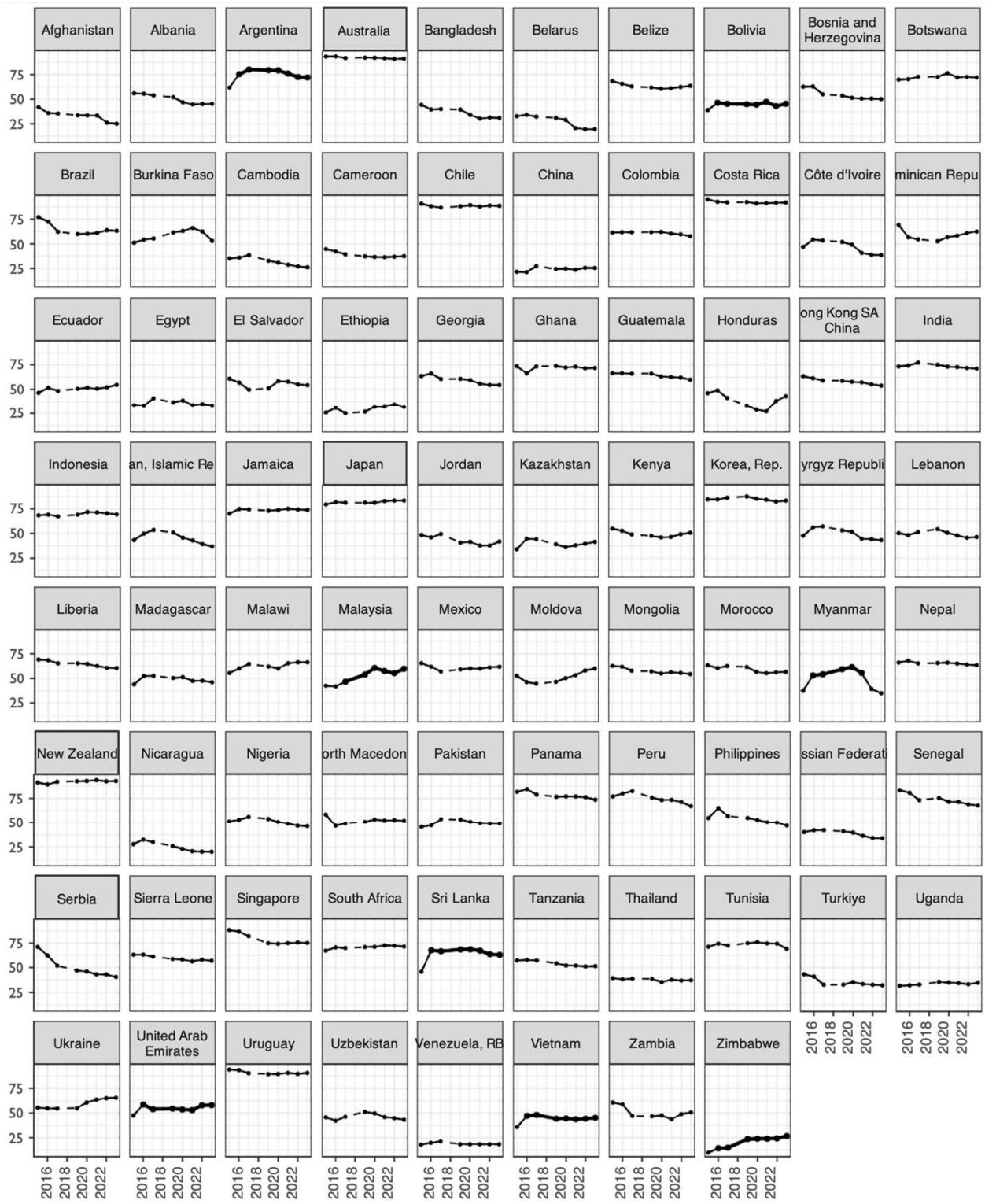
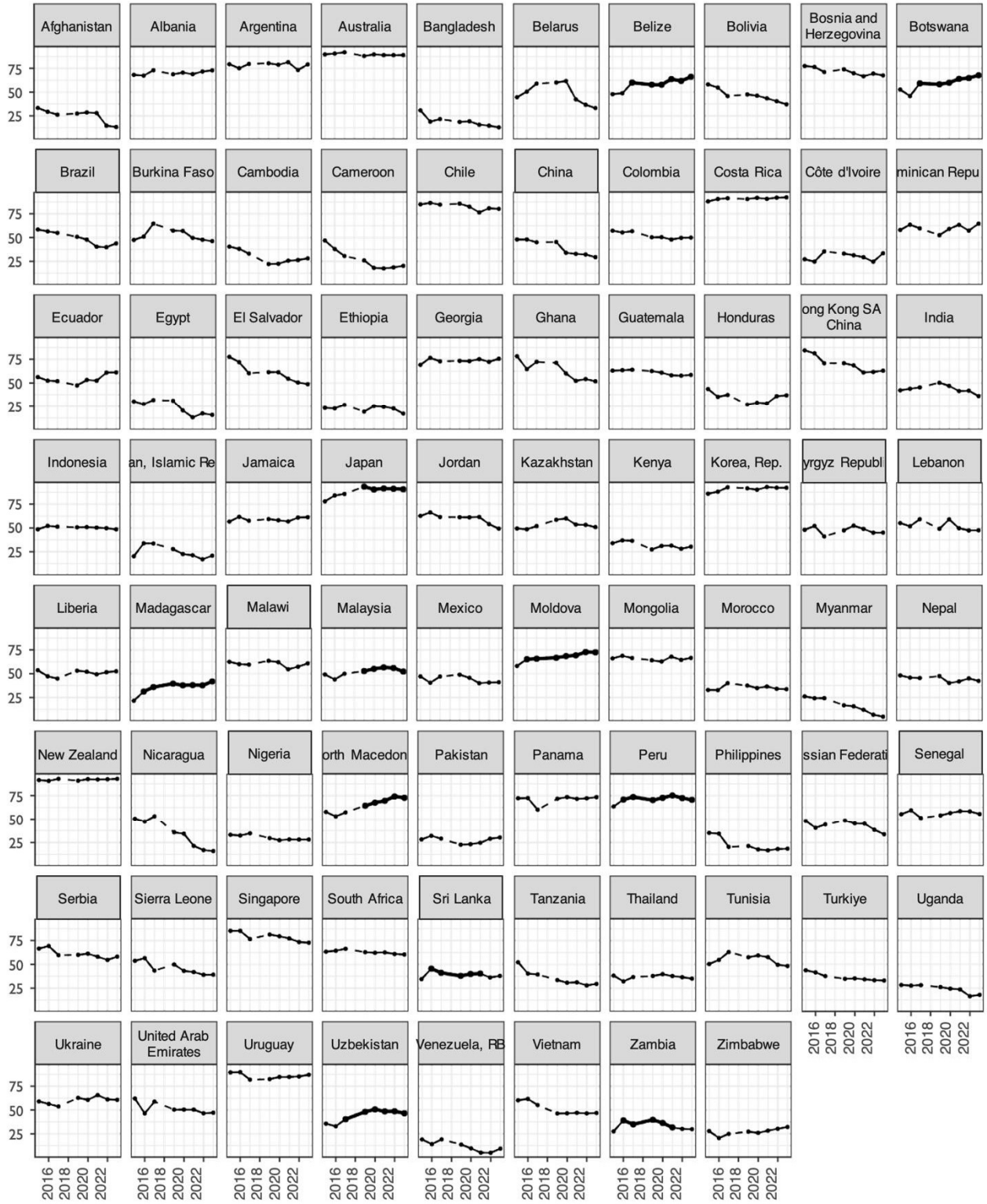
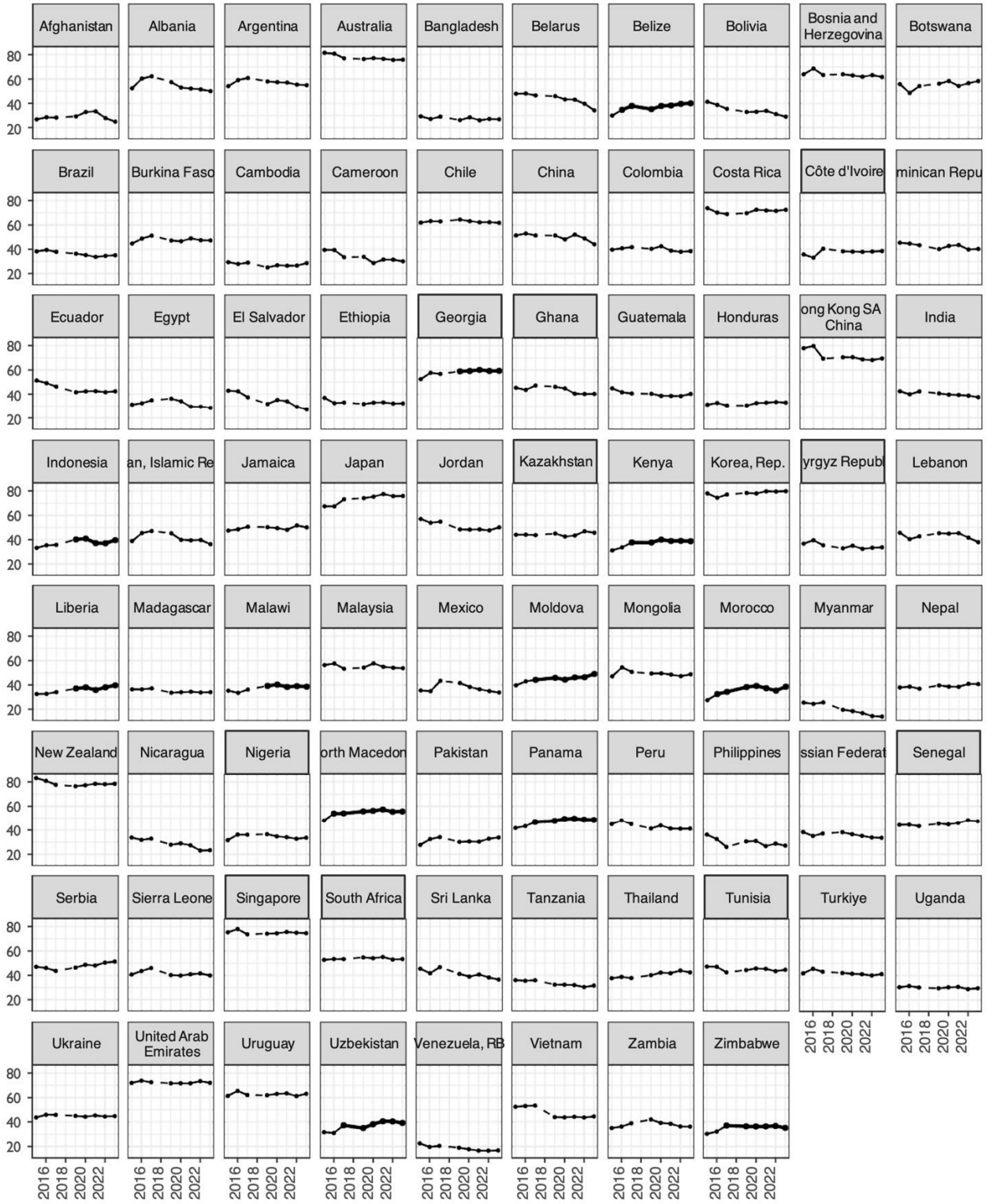


Figure 19: Country Scores on Seven WJP Indicators Selected by USAID



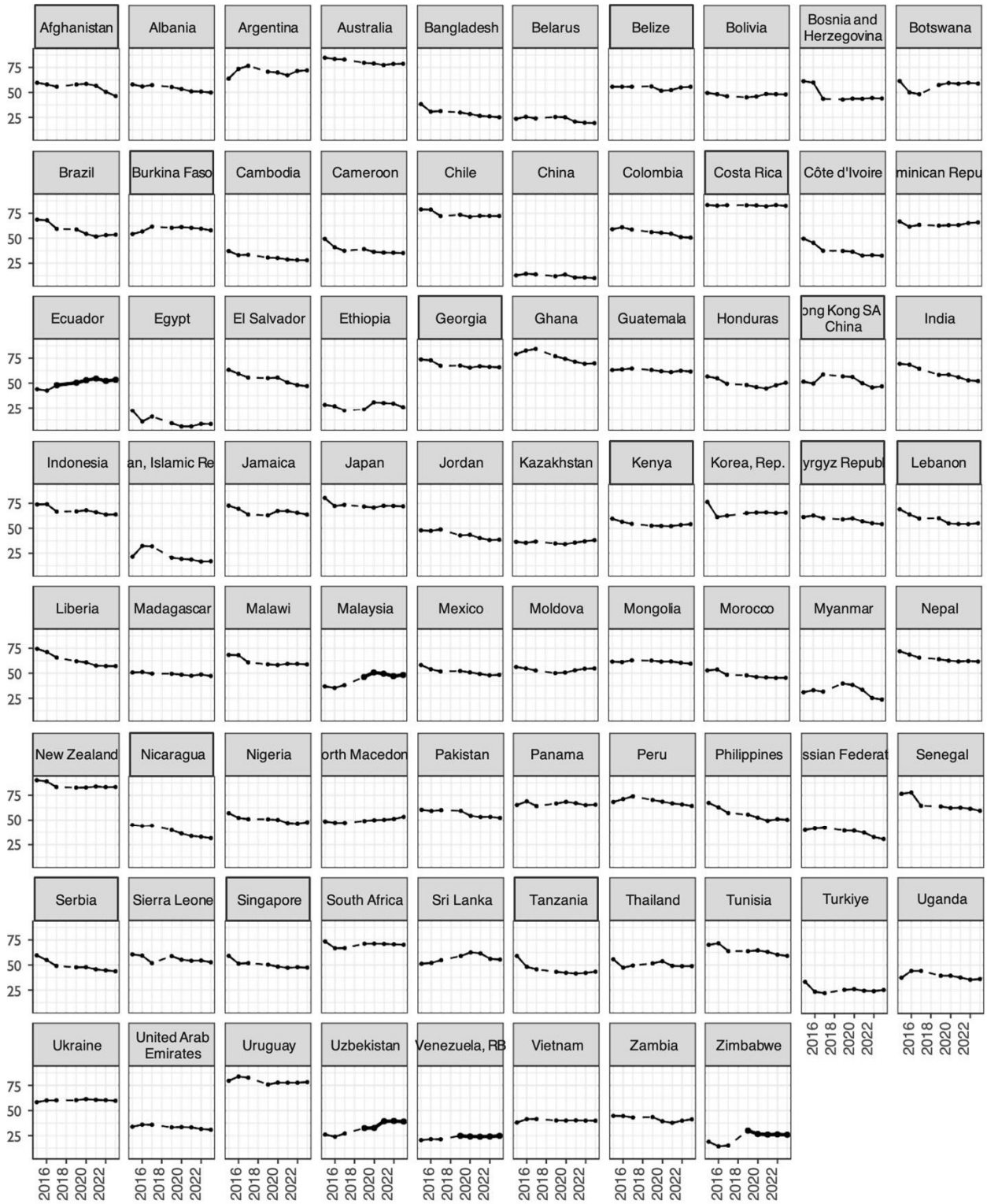
DUE PROCESS

Figure 20: Country Scores on Seven WJP Indicators Selected by USAID



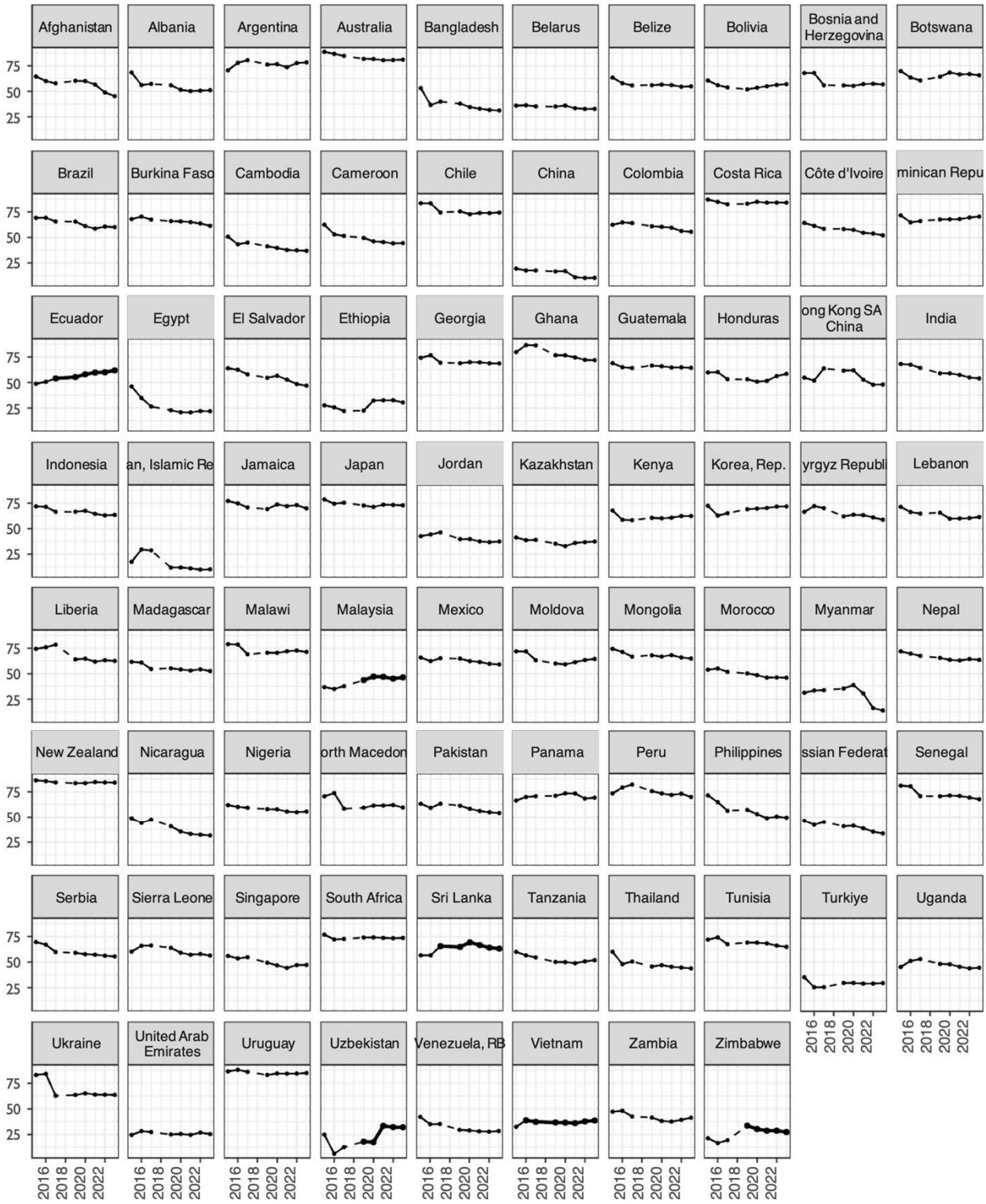
FREEDOM OF OPINION

Figure 21: Country Scores on Seven WJP Indicators Selected by USAID



FREEDOM OF ASSOCIATION

Figure 22: Country Scores on Seven WJP Indicators Selected by USAID



PRIVATE PROPERTY

Figure 23: Country Scores on Seven WJP Indicators Selected by USAID

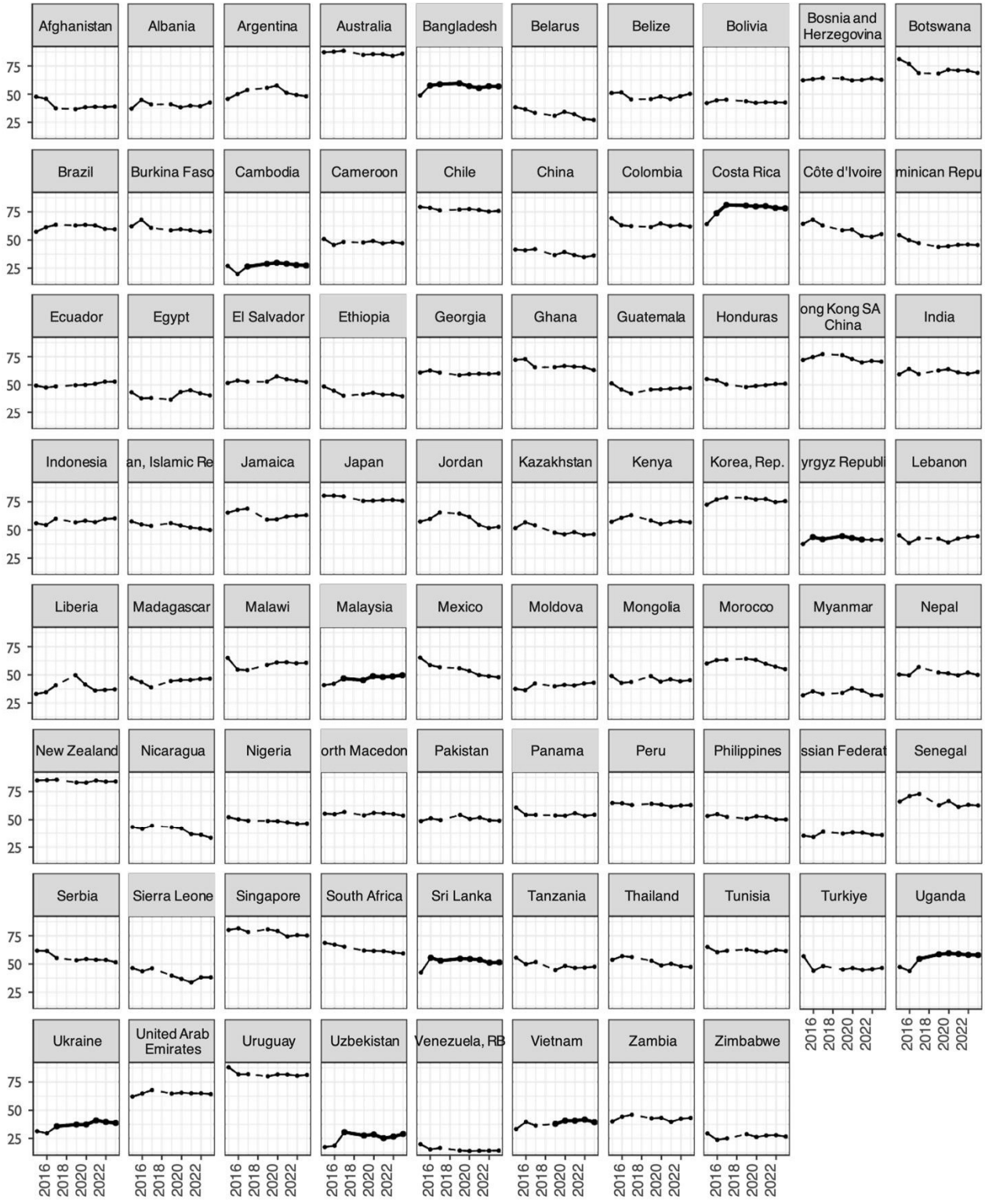
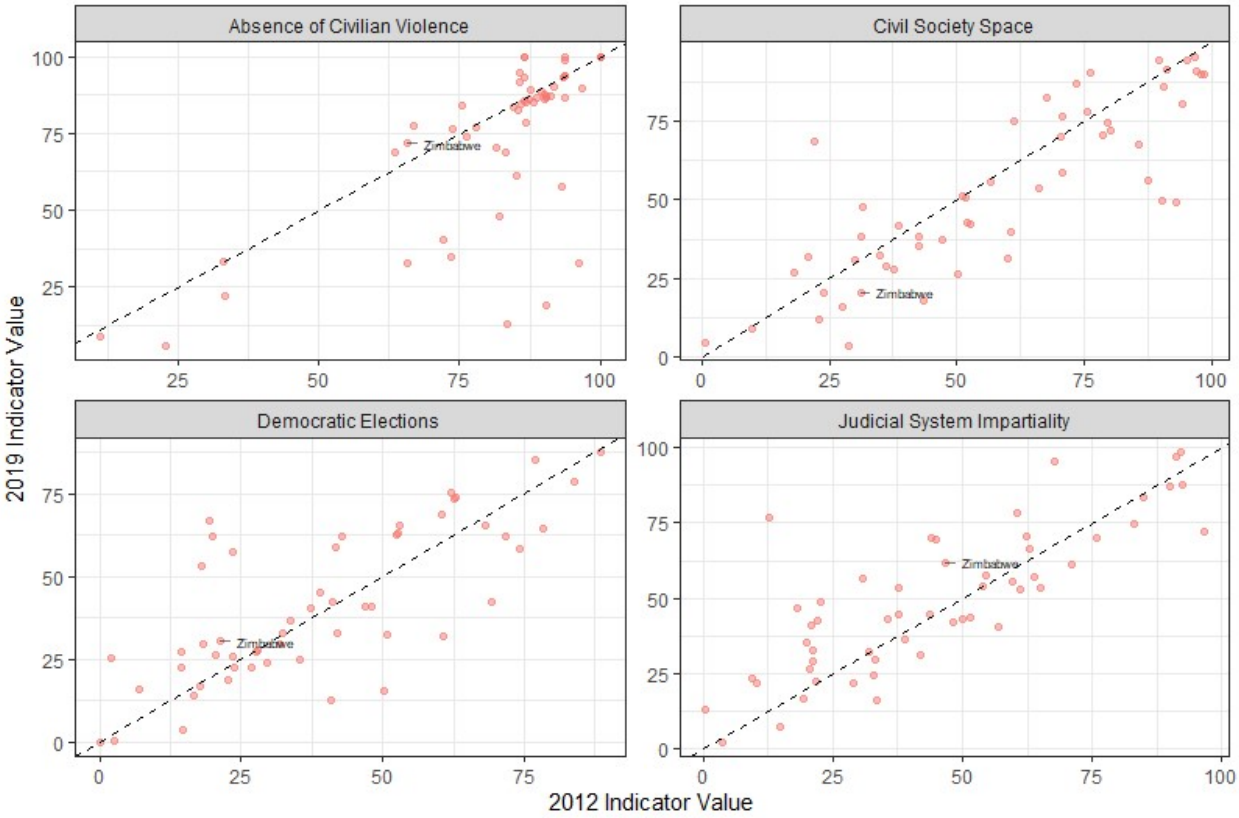


Figure 24: Change in Country Scores in Selected IIAG Indicators from 2012 to 2019



Countries above the diagonal improved, while those below declined. Absence of Civilian Violence exhibits the highest scores, as shown by the concentration of points in the top right quadrant. Within these scores, however, some notable declines exist, as evidenced by points in the bottom right quadrant. While both Democratic Elections and Judicial System Impartiality show cases of substantial improvement (points significantly above the dotted line), Civil Society Space exhibits fewer instances of progress.