

Using practitioner surveys to measure human rights: The Human Rights Measurement Initiative's civil & political rights metrics

K Chad Clay, University of Georgia

Ryan Bakker, University of Essex

Anne-Marie Brook, Motu Economic and Public Policy Research

Daniel W Hill, Jr, University of Georgia

Amanda Murdie, University of Georgia

Abstract

Most cross-national data sets of civil and political rights practices have relied on internationally-distributed English language secondary sources as the core source of information for their metrics. This approach has yielded data that are highly reliable, but also suffer from the fact that their information sources under-represent the overall level of abuse internationally and do so in a way that is biased across countries. The combined knowledge of the individual human rights practitioners working to directly monitor the abuses occurring within a country would likely serve to overcome much of this biased under-reporting, but it is difficult to compare that knowledge across country and cultural contexts. In this article, we discuss how we overcome these problems in the Human Rights Measurement Initiative (HRMI) civil and political rights data. Using an expert survey that contains anchoring vignettes in concert with Bayesian scaling techniques, we present a new methodology for collecting and aggregating data on the intensity and distribution of respect for eight separate civil and political rights.

Keywords: human rights, measurement, expert surveys, data

Corresponding author: kcclay@uga.edu

Introduction

Civil and political rights ensure the ability to live and engage in religious, political, intellectual, or other activities free from coercion, abuse, or discrimination. They have been enshrined, along with other human rights, in a vast body of international law that has proliferated since the end of World War II. While international organizations, advocacy groups, scholars, and members of the public are rightly concerned about violations of these rights around the world, obtaining reliable information about civil and political rights violations has proven to be an elusive goal, for several reasons.

First, governments often frame and contest reports of abuse, arguing that such acts were necessary to maintain security. For example, in 2015, there were reports that Egyptian police were targeting and killing people they suspected of ‘terrorism and other criminal activity.’ Human rights advocates argued that many of these killings were extrajudicial executions accompanied by evidence of torture. However, the Egyptian ‘Ministry of Interior claimed the suspects had been killed after opening fire on police officers’ (USDS, 2016). Second, government agents often attempt to commit violations in secret and hide evidence after the fact, as occurred in Bangladesh in 2015, when ‘members of security forces in plain clothes arrested dozens of people and later denied knowledge of their whereabouts’ (Amnesty International, 2016, 83). Third, many cases of abuse are never publicly reported at all. All of these factors make it difficult to accurately assess the level of government respect for civil and political rights around the world.

Previous human rights data projects have attempted to mitigate these problems by relying on the public documentation produced by governments and human rights organizations (HROs). Using highly replicable, standards-based coding procedures, these projects have successfully produced data and addressed some of the measurement problems noted above. Still, crucial weaknesses remain. While the information in public documentation is credible and unlikely to exaggerate abuse (Hill, Moore, and Mukherjee, 2013), it is also subject to political, legal, and resource constraints. The allegations of abuse in such reports biasedly understate the level of abuse in countries worldwide (Conrad & Moore, 2011; Conrad, Haglund & Moore, 2014).

More recent data collection projects, especially the Varieties of Democracy Project (Coppedge et al., 2019), have eschewed the information in public reports in favor of surveys of academics and other experts. While this approach has promise, it is not clear that it obviates the problems with the annual reports, since many respondents are likely to base their rankings on publicly available information. Further, since V-Dem is focused on measuring attributes of democracy rather than human rights, no previous project has applied this expert-survey approach to concepts defined by international human rights law.

We have had numerous conversations with government leaders, bureaucrats, and human rights advocates who express skepticism of existing human rights data due to the problems mentioned above. These conversations often ended with some variant of the question: ‘We are on the ground in Country X, so why not just ask us for the information directly?’ The Human Rights Measurement Initiative’s (HRMI) approach to measuring civil and political rights takes this question seriously, basing its data on information supplied directly by human rights practitioners who are responsible for monitoring human rights practices in particular countries or regions. The results of our data collection give us good reason to believe that our method appropriately captures information on civil and political rights, while providing better and more detailed information on what is occurring in countries than has been provided by any previous cross-national human rights data collection effort.

What do existing measures of civil and political rights miss?

Several existing projects have attempted to measure civil and political rights in different ways, e.g. Cingranelli, Richards & Clay (2014a), Conrad and Moore (2010), and Gibney et al. (2015). If there are so many cross-national measures, what could they possibly be missing?

According to Goldstein (1986), attempts to generate quantitative data on human rights face challenges associated with definitions, data reliability, and data interpretation. With regard to definitions, most projects have hewed closely to the definitions found in international human rights treaties, often aided by the treaty bodies that interpret those documents; HRMI also adopts this approach. However, concerning problems of data reliability and interpretation, HRMI avoids many of the shortcomings of existing approaches and provides more detailed, contextualised information on the distribution of abuse and the groups who are most affected.

Problems of information

The contentious nature of human rights is perhaps the most serious impediment to the collection of information about civil and political rights practices. When violations are reported, states often attempt to frame abuses as either committed out of necessity or carried out by rogue agents without the state’s permission (McCoy, 2012: 52). Likewise, by their very nature, many violations of civil and political rights are clandestine, with violators seeking to conceal their actions (e.g. Conrad, Hill & Moore, 2018; Rejali, 2007). Success in concealing abuses exacerbates the problems surrounding any attempt to collect comparable information across countries.

Most previous attempts to collect cross-national data on civil and political rights have relied on public documentation from the US State Department and HROs, primarily Amnesty International and Human Rights Watch (e.g. Cingranelli, Richards & Clay, 2014a; Conrad & Moore, 2010; Gibney et al., 2015). These projects have produced data that are highly reliable (Fariss, 2014). However, either explicitly

(Conrad & Moore, 2010) or implicitly (Cingranelli, Richards & Clay, 2014a; Gibney et al. (2015)), these projects also acknowledge limitations in the information sources on which their quantitative data are based. As Bollen (1986) discusses, human rights violations often go unreported, even when journalists or HRO members know about them. Human rights organizations have limited resources with which to promote better human rights practices. Understandably, they strategically focus their effort on places and issues where they are most likely to have an impact (Barry, et al., 2015; Hendrix & Wong, 2014). Concerns about expending valuable time and resources with little or no resulting improvement inhibits organizations from publicizing all of the information they collect. Also, these organizations must maintain a credible international image. While HROs are unlikely to produce false allegations of abuse, they also do not report every actual instance of abuse of which they are aware (Hill, Moore & Mukherjee, 2013). Allegations are always contentious and politically sensitive, which sometimes makes organizations hesitant to publicize them. The effect of these organizational incentives is that all publicized allegations are credible, but not all credible allegations are made public.

The distance between what is reported about human rights abuses and what an organization knows about them is almost certainly larger for some countries than others. Some countries have more journalists and HROs; some receive a greater share of international attention. As such, estimates of human rights abuse based on the information sources used by previous measurement projects are likely biased downwards. That is, they overestimate the degree to which human rights are enjoyed everywhere, but more in some places than others (Conrad & Moore, 2011; Conrad, Haglund & Moore, 2014).

Many have tried to address this problem. The ordered categorical scales used by the Political Terror Scale (PTS) (Gibney et al., 2015) and the CIRI Human Rights Data Project (CIRI) (Cingranelli, Richards & Clay, 2014a) acknowledge the lack of precision in data created from public human rights reports. However, even these limited scales may be subject to the problem of underestimation, and it is possible that, in earlier periods, more limited documentation of human rights abuse may have produced a more severe undercount than does current documentation (e.g. Clark & Sikkink, 2013; Fariss, 2014). Since the bias in existing human rights data may extend across both space and time, some have suggested using statistical methods to account for it (e.g. Bagozzi, et al., 2015; Conrad, Hill & Moore, 2018). While this strategy may help with the validity of inferences drawn from secondary analyses, it does not provide easy-to-understand measurements for a wide audience. Fariss (2014) produces data that account for changing standards of accountability over time by utilising a measurement model drawn from multiple data sets of various types of abuse. While this approach is promising, it would be better to have higher quality data for each type of abuse in the first place. It would also be advantageous to forgo the extreme data reduction process necessary to obtain these estimates, which combines several distinct human rights practices into a single scale.

V-Dem has attempted to sidestep these problems by turning to another source of information: individuals with expertise on the politics of countries or regions (Coppedge et al., 2019). In most cases, a V-Dem Country Expert holds a PhD degree, suggesting that most respondents are likely to be academics. This approach is appealing since it effectively pools many sources of information (respondents' answers) together instead of relying on two or three public reports. While this approach is a welcome step forward, it also suffers from significant limitations. Academics undoubtedly know more about the relevant country than the average person, but there is good reason to believe they rely heavily on secondary sources.¹ The more they rely on similar sources, particularly if those sources are the public reports discussed above, the more that estimates of abuse taken from their responses are likely to suffer from the same bias that has afflicted previous measures. Additionally, the precision of these estimates will likely be overstated. While there is nothing inherently wrong with using information from secondary sources, the logic of 'crowdsourcing' does not apply when members of the crowd do not have access to distinct sources of information. Additionally, academic expert surveys can differ significantly from the measures created using public reports. There is less agreement between the measures than most would expect, and no clear indication that the disagreement decreases in more recent years. Importantly, the two are *negatively* correlated across time for many countries, suggesting a potentially large and consequential amount of error in one or both measures (Cope, Crabtree & Fariss 2020).

Problems of interpretation

Another challenge facing all human rights measurement projects involves the interpretation of the limited information to which they have access. We focus on two interpretive issues that we hope to improve upon: (1) the accurate representation of uncertainty and (2) the dimensionality of civil and political rights abuse.

Both the PTS and the CIRI projects handled the problem of uncertainty in the information contained in human rights reports by using ordered scales with a few broad categories. For instance, the CIRI measure for torture and ill-treatment allowed for grouping states into only three categories (Cingranelli, Richards & Clay, 2014b). While this is a reasonable approach given the level of precision in the reports, a country with 500 documented instances of torture and another with 5,000 would fall in the same category of frequent abuse, each receiving a score of 0. Both countries certainly engage in high levels of abuse, they

¹ This is especially true of V-Dem estimates of conditions in the distant past (as far back as 1900). As questions reach further into the past, respondents are increasingly likely to rely on similar source materials. HRMI's expert survey methodology means that we will only be able to collect information moving forward and will not be able to collect historical information. However, for the reasons stated throughout this article, there are many reasons to be sceptical of data on historical abuse.

are not equivalent. Even though most academic researchers understand the limitations of categorical scales, they can create confusion among wider audiences.²

Further, the most commonly used measures have been constructed in a way that makes it impossible to know the degree of certainty around a country's categorical placement. Returning to CIRI's torture measure, did a state receive a 1 because, based on the report, researchers were confident that the state engaged in only a few instances of torture? Or was it because there was not enough information to tell whether it belonged in the worst category? Was it close to the borderline between categories or quite far away? Just as two countries in the same category may have different human rights practices, two countries in different categories may have relatively similar practices. With a single score and no indication of the quality of information, it is impossible to know.

Further, most previous attempts to collect cross-nationally comparable civil and political rights data have ignored the dimensionality of abuse. Stohl et al. (1986, 600-603) note that there are three dimensions to the violation of civil and political rights: (1) scope, (2) intensity, and (3) range. 'Scope' refers to the number of different types of abuse the violator has engaged in, i.e. the rights being violated. 'Intensity' refers to the frequency of each type of abuse. 'Range' refers to the proportion of the population that has been targeted.

While researchers have long recognized these distinct characteristics of abuse, previous projects have failed to create measures that capture these as distinct dimensions. For instance, while PTS captures aspects of scope, intensity, and range, it collapses all of them into a single index, treating three separate dimensions as if they can be captured on a single scale (Gibney et al., 2015). Like PTS, Fariss (2014) produces a single score for all physical integrity rights. While CIRI uses disaggregated measures of different types of abuse, its individual scores only measure the intensity of those particular types of abuse, with no comparable measure of range. V-Dem's measures also convey no information about range (Coppedge et al., 2019) and are also limited in scope as they exclude measures of political imprisonment and disappearances.

HRMI's approach to civil and political rights measurement

We aim to produce a suite of metrics that covers the full range of human rights listed in the core United Nations human rights treaties (HRMI, 2018).³ We seek to create measures in a way that ensures cross-national comparability, while remaining transparent about how those measures are created.

² See Ophir Bar-Zohar's article from 14 December 2011, 'Israel Earns Another Failing Score on Freedom of Religion Index,' in which Israel received the same score as China, Iran, Saudi Arabia, and Afghanistan: <https://www.haaretz.com/1.5219143> (Last Accessed: 18 March 2018).

³ The HRMI data also include measures of five Economic and Social Rights, based on the measurement strategy employed by Fukuda-Parr, Lawson-Remer & Randolph (2015).

Ultimately, we want to create data that are useful for human rights advocates, researchers, journalists, and anyone else seeking information on human rights worldwide.

Compared to existing approaches, we expect information collected directly from human rights practitioners to have several advantages. One is that the information will reflect a more complete set of human rights abuses than those included in public reports. As discussed above, it is not feasible for any human rights organization to include in their public reports all of the information at their disposal. Asking practitioners directly about the countries they are responsible for monitoring will help circumvent this problem, as individual practitioners that work for or with these organizations will be less constrained by the incentives that lead to omissions in public reports. While practitioners certainly have limited time, their responses in an anonymous survey will be less prone - compared with public reports - to pragmatic, organization-driven considerations about which country or abuse demands the most attention. Practitioners also need not worry about the political consequences of their answers, since they will not be publicized in an official report bearing their organization's name.

We expect that data collected from surveys of human rights practitioners will be an improvement over data collected from academic surveys. Relative to academics, human rights practitioners are more likely to have access to primary sources and direct experiences that help them make their assessment and will have been consuming human rights information as part of their daily routine.

The HRMI civil and political rights practitioner survey

We developed a survey that has been used to collect annual data on eight civil and political rights in 2017 and 2018. Each of the eight civil and political rights are founded in the International Covenant on Civil and Political Rights (ICCPR) and other relevant international law. These are the rights to political participation (Article 25), opinion and expression (Article 19), assembly and association (Articles 21 and 22), freedom from torture and ill-treatment (Article 7 and the Convention against Torture), freedom from extrajudicial execution (Article 6), freedom from the death penalty (Second Optional Protocol to the ICCPR), freedom from arbitrary or political arrest and detention (Articles 2, 9, 11, 18, 19, 21, 22, and 26), and freedom from disappearance (Articles 9 and 10, and the Convention on Enforced Disappearances).. Each section of the survey contains (1) a definition of the right under consideration, (2) questions about the intensity of respect for that right, and (3) questions regarding the range of respect for that right, i.e. who was targeted for abuse. The definition of each right was determined on the basis of international law and its interpretation by the appropriate treaty bodies at the United Nations. Questions about the intensity of violations of rights generally follow the pattern of this question on torture and ill-treatment:

In 2018, how often did government agents, such as soldiers, police officers, and others acting on behalf of the state, commit acts of torture or ill-treatment?

Respondents answered this question on an 11-point scale, ranging from 0, which represented an answer of ‘Never’, up to 10, which represented an answer of ‘Constantly.’⁴

After that, we asked our respondents to provide us with more specific information about range, e.g. asking who was especially likely to experience torture:

In 2018, which groups of people, if any, were especially vulnerable to torture and ill-treatment by government agents, such as soldiers, police officers, and other state-sanctioned actors? (Select all that apply.)

Respondents could select from 31 pre-selected identifiers or provide us with other potential identifiers that we did not have the foresight to include. We followed this with an open-ended question asking for more information on why the respondent chose their responses.

We also include sections asking our respondents to score the intensity of three hypothetical countries on their respect for the rights under consideration. These hypothetical cases help account for differences in the interpretation of our 11-point intensity scale and contribute meaningfully to the final intensity scores produced for each country. The entire survey is in the online appendix.

Selection of pilot countries and survey respondents

Our respondents are human rights practitioners who are actively monitoring the civil and political rights situation in each country. They typically work for an international or domestic non-governmental organization. We have also allowed for participation by human rights lawyers, journalists covering human rights issues, and staff working for National Human Rights Institutions if that institution has been rated as fully compliant with the Paris Principles (United Nations, 2010; GANHRI, 2019).⁵

We have mainly relied on respondents located within the country in question. However, for some more closed and repressive countries, we had to rely on a higher proportion of respondents who are based outside of the country in question. Because our goal has been to collect information from respondents who are first points of contact for human rights information in the country and have access to primary sources, we do not rely on academics as respondents. Likewise, in order to ensure that our measures are independent from government-backed sources, we excluded staff members at government-organized NGOs.

⁴ The survey question is inverted from the final score presented in our results below, in which higher scores represent better respect for the right in question. The scores are inverted prior to the BAM estimation procedure described below.

⁵ The countries in our 2018 sample with an NHRI that met this criterion were: Australia, Liberia, Mexico, Nepal, New Zealand, South Korea, and the United Kingdom.

The sample of potential respondents was determined by a two-step process. First, we sought nominations from human rights advocates worldwide for countries to include in the study. Thirteen countries were nominated in 2017, and we selected all for inclusion in the pilot, as together they provided significant diversity in government type, country size, level of development, geographic location, and many other factors. Following nominations given at the HRMI co-design workshop in Johannesburg, South Africa in 2018, an additional six countries were added to the sample, resulting in the 2019 sample of nineteen countries: Angola, Australia, Brazil, Democratic Republic of Congo (DRC), Fiji, Jordan, Kazakhstan, Kyrgyzstan, Liberia, Mexico, Mozambique, Nepal, New Zealand, Saudi Arabia, South Korea, United Kingdom, United States, Venezuela, and Vietnam. Second, relying on volunteer HRMI ambassadors and other trusted partners in non-governmental human rights organizations, we engaged in a snowball sampling technique. As potential respondents were added to the list, those respondents were also asked if they could recommend potential respondents. By the end of the 2019 process, we had identified between 12 and 74 potential survey respondents per country, each of whom was sent a single-use survey link to ensure that the survey link was not shared with unintended respondents. In total, 771 people were invited to take the survey in 2019, and we received 215 fully completed surveys. The number of fully completed surveys per country ranged from 6 to 19, with an average of 11 completed surveys per country. Responses from partially-completed surveys were also used, to the extent possible, and some responses were dropped for some rights if the respondent failed to correctly order the anchoring vignettes or if their responses were extreme outliers.

Producing intensity scores: Model description

The most common approach to aggregate survey responses into a single score for each country is to take the average of the responses for a given country. While this approach is simple and intuitive, it makes some assumptions that are likely to be violated when collecting survey responses from different people in different contexts. The two most problematic assumptions are (1) that each item relates equally as well to the underlying latent dimension and (2) that our respondents view the underlying scales in the same way. The latter assumption relates to the potential for what is called differential item functioning, or DIF.

To aggregate the survey responses to the country level, we employ the Bayesian Aldrich-McKelvey (BAM) scaling algorithm developed by Hare et al (2015). This model estimates a latent concept using observable indicators of that concept and allows us to correct for the potential of DIF in the data. The unobserved concept of interest is the intensity of human rights respect in a given country and the observed outcomes are survey responses from practitioners in that country.

The BAM routine assumes there is a ‘true’ score for each country/right and that the survey responses are linear transformations of this true score. Formally, the model is:

$$Y_{ij} = \alpha_i + \beta_i \theta_j + \tau_{ij}$$

where Y_{ij} is respondent i 's placement of a given right in country j . The α and β parameters, which are indexed by respondent, are the linear transformation parameters that map each respondent's placement onto the 'true' score for a given country/right, θ_j , and τ_{ij} is a compound error term consisting of both respondent and country/item level random variation.

Because we are estimating a Bayesian version of the model, we must supply distributional information about the unknown quantities in the model. As the survey responses have 11 categories we treat them as normally distributed. We can write our model:

$$\begin{aligned} Y_{ij} &\sim N(\mu_{ij}, \tau_{ij}) \\ \mu_{ij} &= \alpha_i + \beta_i \theta_j \\ \tau_{ij} &= \tau_i \tau_j \end{aligned}$$

Each α_i represents respondent i 's tendency to place countries lower/higher on the scale. A respondent with a negative α tends to rank countries on the low end of the scale, while one with a positive α tends to push their rankings towards the high end. Each β_i represents how well a respondent distinguishes between poor and good human rights conditions. Respondents with β s closer to 0 place countries with different human rights performances relatively close together on the scale, while those with more positive β s place countries with different performances relatively far apart on the scale. A negative value of β would indicate that the respondent ranks countries with worse performance higher than those with better performance. We allow the variation in survey responses, τ_{ij} , to be a function of both respondent and item level variation.

One of the advantages of our approach versus taking the simple mean of the responses is that our approach can handle differences in how respondents may view the underlying response across different countries. That is, what one person may view as a 6 another may view as a 4. To account for this, we include a set of hypothetical countries, described in the survey, that all respondents place regardless of their country of expertise. These 'anchoring vignettes,' combined with the Bayesian Aldrich-McKelvey model described above, allow us to correct for any potential differences in how practitioners view the underlying scales in our survey. That is, we use questions about hypothetical countries as 'bridging observations' in order to estimate the model and to create a scale that is cross-nationally comparable. An example data matrix for our model, with 6 respondents from 3 countries, is shown in Table I.⁶

⁶ For a more detailed discussion, see Bakker et al. (2014).

Table I. Example data for Bayesian Aldrich-McKelvey model

Respondent	Country 1	Country 2	Country 3	Vignette 1	Vignette 2	Vignette 3
1	5	—	—	1	5	7
2	3	—	—	5	7	8
3	—	1	—	3	5	10
4	—	4	—	2	3	5
5	—	—	6	6	8	9
6	—	—	9	3	6	8

We estimate our model via Markov chain Monte Carlo simulation. We adopt the following non-informative prior distributions for the parameters in our model:

$$\begin{aligned}
\alpha_i &\sim U(-100, 100) \\
\beta_i &\sim U(-100, 100) \\
\Theta_{ij} &\sim N(0, 1) \\
\tau_j &\sim \text{Gamma}(0.1, 0.1) \\
\tau_i &\sim \text{Gamma}(v, \omega) \\
v &\sim \text{Gamma}(0.1, 0.1) \\
\sigma &\sim \text{Gamma}(0.1, 0.1)
\end{aligned}$$

We let three chains run for 22,000 iterations and stored the last 2,000 draws from the posterior distributions to summarise the model parameters. We assessed convergence via visual inspection of density plots and the Gelman-Rubin statistic, and no parameters showed evidence of non-convergence. The posterior means range from approximately -0.9881 to 1.57, and the standard deviations range from approximately 0.01 to 0.5.

Presentation of selected physical integrity rights data

Our models, combined with the additional information collected in our survey, produce measures of both intensity and range of abuse for 8 different civil and political rights. In the interest of space, we present results from our measurement models using the 2018 data on the most commonly measured physical integrity rights: the rights to be free from torture and ill-treatment, extrajudicial killing, disappearance, and political or arbitrary arrest and imprisonment.

Figure 1 presents the estimates for the intensity of respect for each physical integrity right. Because the models that produce these estimates are Bayesian, they can be used to calculate a mean score for each country along with an estimate of uncertainty around each score, based on the estimated standard deviations of the posterior distributions. In Figure 1, mean scores for each country are presented as dots. The horizontal lines around each dot show the 90% credible interval for that country.

Figure 1. Estimates of selected physical integrity rights performance, 2018

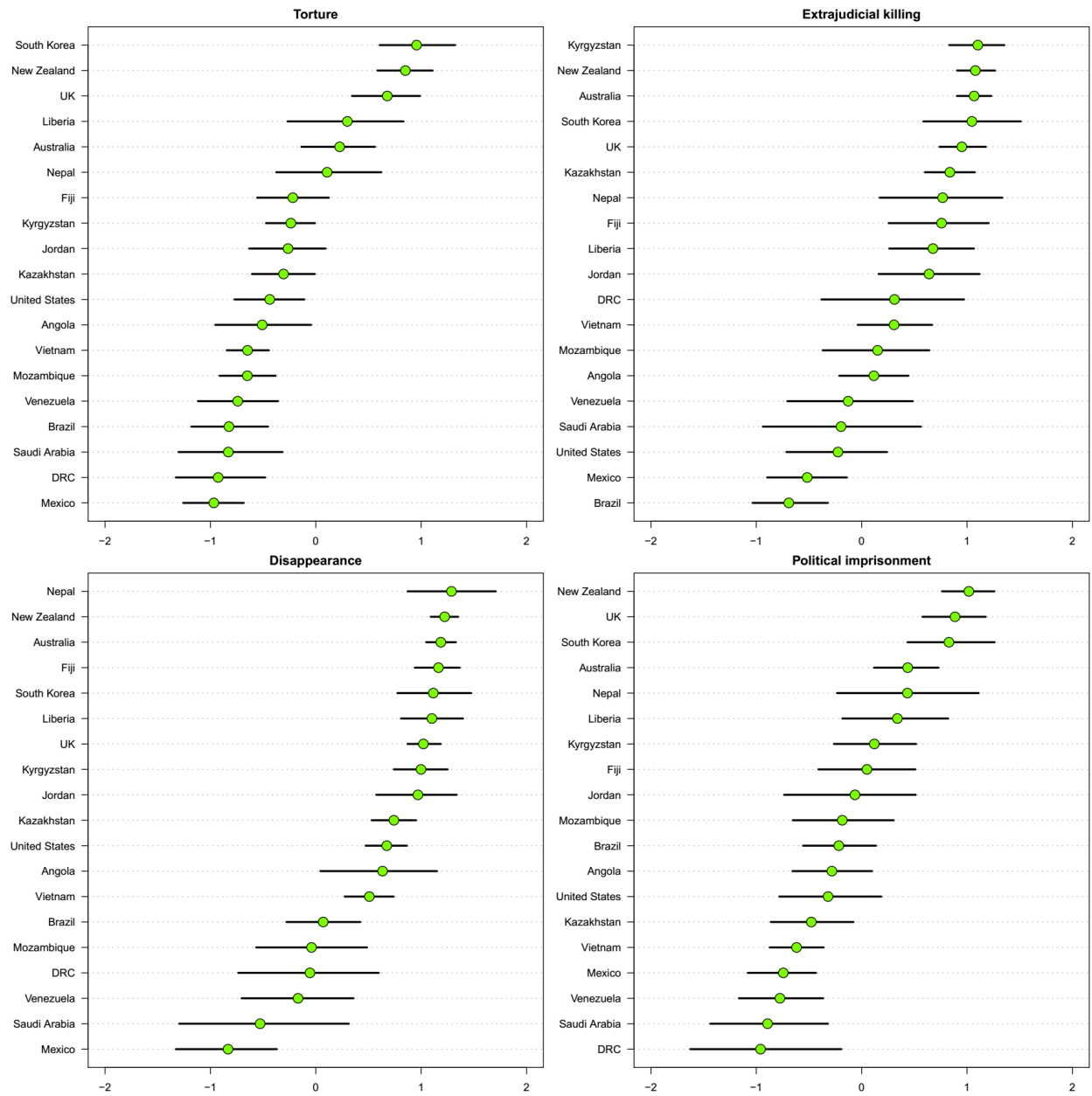


Figure 2. Alpha parameters for respondents in torture model

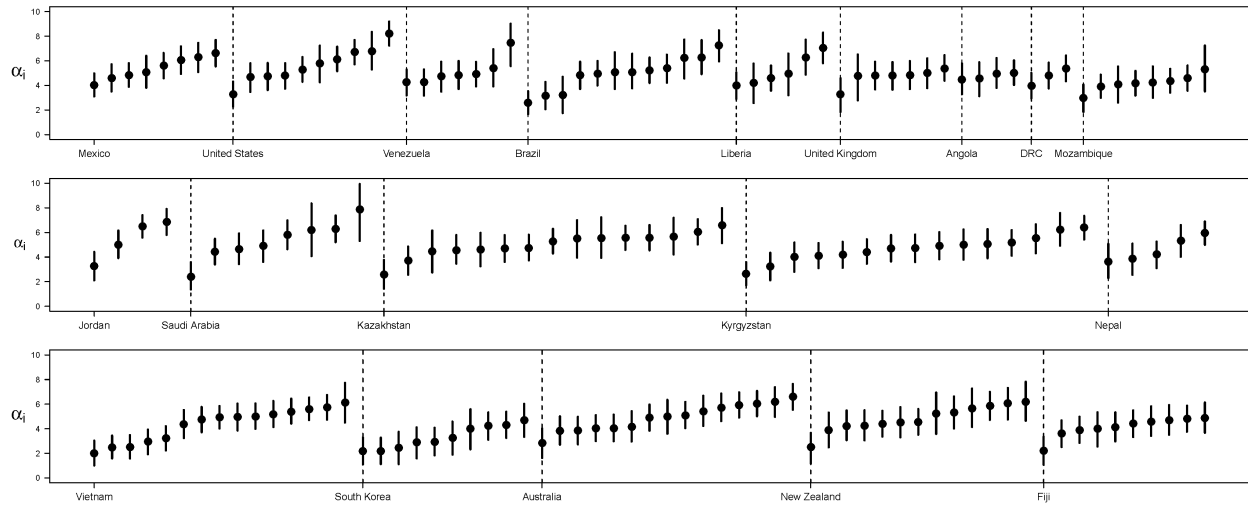
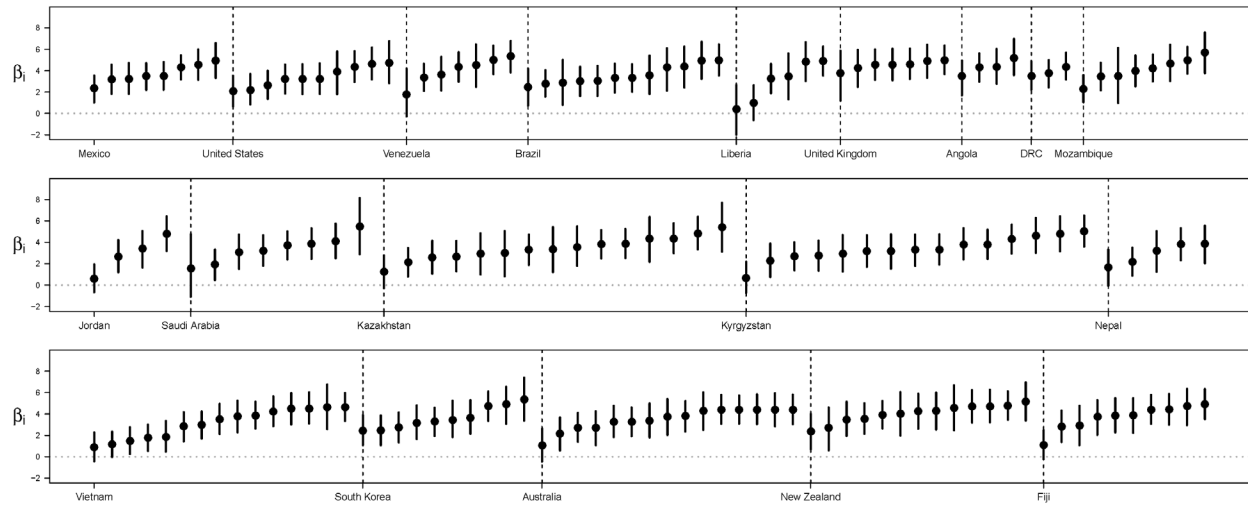


Figure 3. Beta parameters for respondents in torture model



For each value in Figure 1, we have respondent-specific alpha and beta parameters; Figures 2 and 3 display these parameters for torture and ill-treatment. The alpha parameters indicate whether a respondent tends to place countries toward the low or high end of the scale. These estimates are all positive, suggesting that most of our respondents tend to give countries relatively favourable ratings. Estimates for the beta parameters are also positive (though some are close to 0), which means most respondents included assigned substantially better scores to countries with better practices. Interestingly, for both parameters there is generally more variation within countries than across, suggesting that differences in respondent parameters are more likely the result of characteristics of the respondents themselves rather

than characteristics of the countries they monitor. Note that a model with a single parameter for each item would force these parameters to be the same across respondents. There is enough variation in our estimates to suggest that such a model would lead to erroneous inferences.

We also gather rich information on the range of abuse. Figure 4 shows the number of identifiers that at least two respondents claimed placed a person at risk for four different types of physical integrity rights violations. As shown, there is substantial variation across countries in the number of identifiers chosen across both countries and rights. Further, the number of identifiers chosen per country does not perfectly correlate with that country's intensity score shown above, with Spearman rank correlation coefficients ranging from approximately -0.6 for disappearance to -0.4 for torture.⁷

Figure 5 digs deeper into the specific identifiers that respondents claimed were targeted for torture. There is substantial variation across countries in some of the other particular identifiers that made one highly likely to experience torture at the hands of government agents. For instance, *indigenous people* were selected by more than 50% of respondents in Australia, Brazil, Mexico, and New Zealand, while *people of particular races* were selected by more than 50% of respondents in Brazil, the United Kingdom, and the United States. Respondents claimed that *people with particular religious beliefs or practices* were especially likely to be targeted in Saudi Arabia, the United States, and Vietnam, and *human rights advocates* were likely to be subject to torture in Angola, Brazil, the DRC, Mexico, Mozambique, Saudi Arabia, and Vietnam.

⁷ Higher intensity scores indicate fewer rights violations, while higher numbers of selected identifiers indicate that more identities were at risk for violations. As such, the negative correlation is to be expected.

Figure 4. Number of identifiers selected for individuals at risk for four types of physical integrity violations
(Excluding 'None' and 'All')

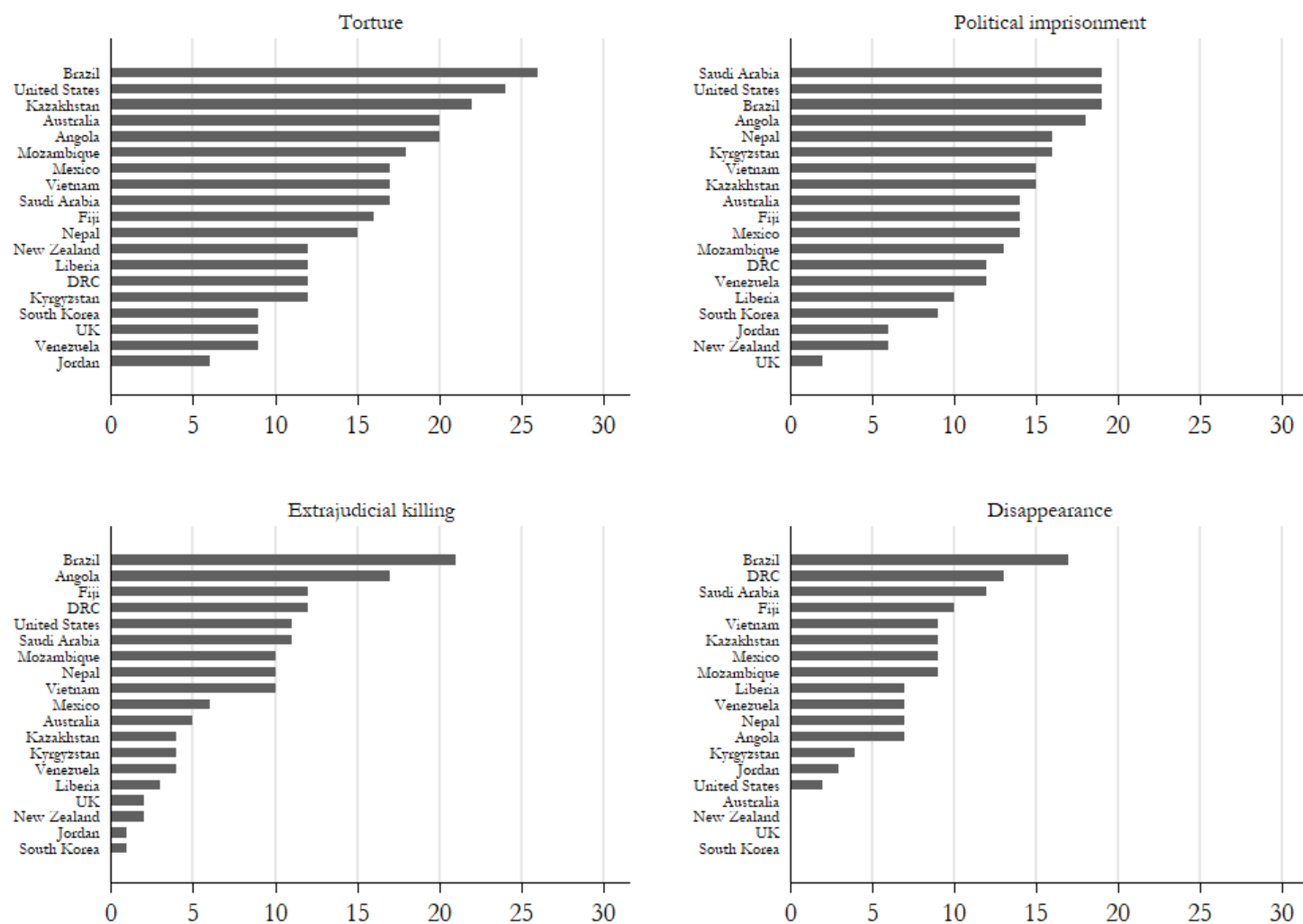


Figure 5. Identifiers selected for individuals at risk for torture, 2018
(As a proportion of total respondents)

	total	none	all	indigenous	ethnicity	race	culture	religion	nationality	location	foreign	women	LGBTQIA+	children	older	disability	medical
Angola	13	0.00	0.00	0.00	0.00	0.00	0.08	0.46	0.46	0.31	0.08	0.46	0.08	0.15	0.08	0.15	0.08
Australia	19	0.05	0.00	0.89	0.42	0.21	0.26	0.32	0.11	0.16	0.16	0.00	0.05	0.26	0.11	0.42	0.16
Brazil	15	0.00	0.07	0.60	0.60	0.80	0.33	0.47	0.27	0.67	0.07	0.53	0.73	0.13	0.07	0.07	0.13
DRC	6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.17	0.33	0.00	0.17	0.00
Fiji	14	0.00	0.14	0.14	0.07	0.07	0.07	0.07	0.07	0.07	0.00	0.14	0.29	0.07	0.07	0.14	0.07
Jordan	6	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Kazakhstan	18	0.06	0.22	0.11	0.06	0.00	0.06	0.39	0.06	0.11	0.11	0.11	0.22	0.11	0.06	0.06	0.06
Kyrgyzstan	20	0.00	0.45	0.00	0.35	0.05	0.05	0.40	0.30	0.00	0.00	0.05	0.45	0.05	0.05	0.00	0.05
Liberia	18	0.11	0.06	0.06	0.00	0.00	0.06	0.00	0.00	0.06	0.06	0.33	0.39	0.17	0.06	0.06	0.06
Mexico	10	0.00	0.00	0.60	0.10	0.10	0.10	0.00	0.20	0.20	0.00	0.30	0.20	0.00	0.00	0.00	0.00
Mozambique	11	0.00	0.00	0.09	0.09	0.00	0.09	0.18	0.18	0.18	0.00	0.09	0.18	0.09	0.09	0.18	0.00
Nepal	15	0.07	0.20	0.27	0.33	0.07	0.13	0.13	0.00	0.20	0.00	0.13	0.27	0.07	0.00	0.00	0.07
New Zealand	17	0.24	0.00	0.65	0.18	0.06	0.00	0.00	0.06	0.06	0.00	0.12	0.12	0.12	0.00	0.35	0.18
Saudi Arabia	10	0.00	0.10	0.10	0.20	0.10	0.30	0.50	0.10	0.30	0.20	0.30	0.00	0.20	0.00	0.00	0.10
South Korea	12	0.25	0.00	0.00	0.00	0.08	0.00	0.08	0.00	0.00	0.08	0.08	0.25	0.00	0.00	0.08	0.00
UK	7	0.00	0.00	0.00	0.71	0.71	0.14	0.14	0.00	0.00	0.14	0.00	0.00	0.14	0.00	0.43	0.14
United States	14	0.00	0.00	0.29	0.57	0.93	0.36	0.50	0.43	0.36	0.21	0.36	0.50	0.29	0.00	0.21	0.29
Venezuela	10	0.00	0.00	0.10	0.00	0.00	0.00	0.10	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00
Vietnam	20	0.05	0.40	0.15	0.35	0.05	0.00	0.60	0.00	0.20	0.00	0.05	0.05	0.00	0.00	0.10	0.00

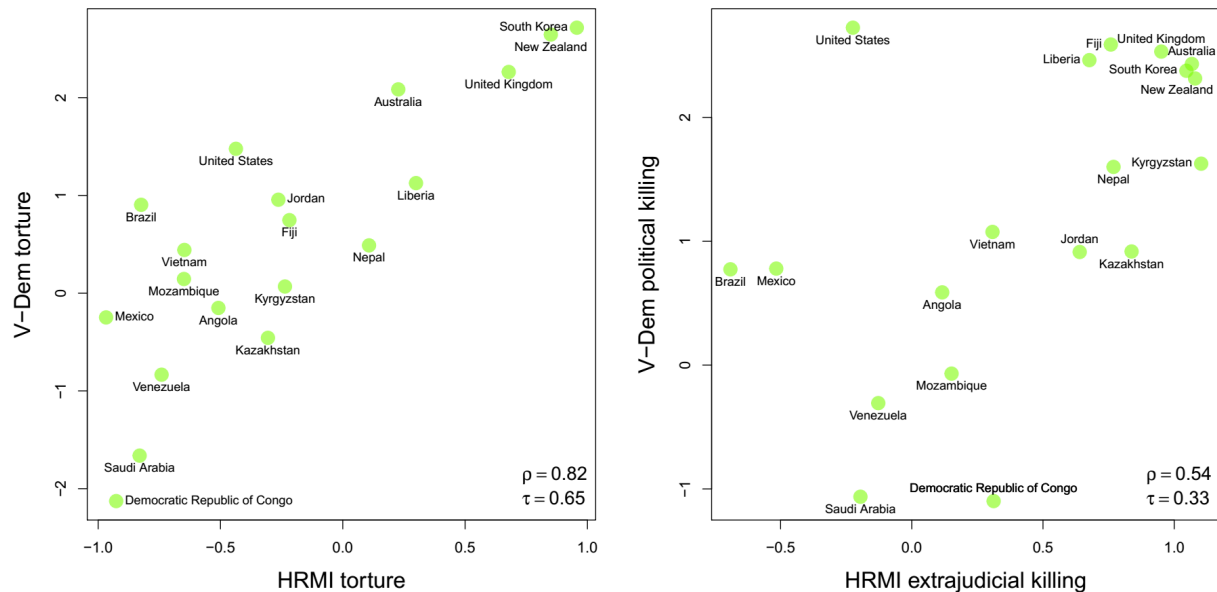
	total	politics	journalists	advocates	unions	protest	pol viol	terrorism	detainees	immigrants	refugees	socioeconomic	homeless	education	academics	profession	other
Angola	13	0.46	0.54	0.54	0.31	0.54	0.08	0.31	0.69	0.62	0.62	0.69	0.31	0.38	0.38	0.23	0.00
Australia	19	0.00	0.00	0.00	0.00	0.11	0.05	0.21	0.53	0.21	0.68	0.26	0.42	0.11	0.00	0.00	0.05
Brazil	15	0.40	0.20	0.73	0.40	0.87	0.13	0.13	0.60	0.27	0.27	0.80	0.60	0.53	0.20	0.13	0.07
DRC	6	0.83	0.83	1.00	0.00	1.00	0.50	0.83	0.67	0.00	0.00	0.33	0.00	0.50	0.50	0.17	0.33
Fiji	14	0.50	0.36	0.29	0.43	0.36	0.00	0.00	0.71	0.07	0.14	0.36	0.07	0.21	0.14	0.14	0.07
Jordan	6	0.50	0.17	0.17	0.33	0.67	0.50	0.67	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00
Kazakhstan	18	0.67	0.28	0.28	0.22	0.67	0.33	0.61	0.72	0.11	0.11	0.44	0.28	0.17	0.06	0.06	0.17
Kyrgyzstan	20	0.05	0.00	0.10	0.00	0.10	0.00	0.60	0.60	0.00	0.00	0.35	0.15	0.35	0.00	0.00	0.00
Liberia	18	0.28	0.50	0.44	0.06	0.33	0.22	0.06	0.50	0.11	0.06	0.17	0.06	0.11	0.00	0.00	0.06
Mexico	10	0.20	0.30	0.60	0.10	0.60	0.20	0.30	1.00	0.60	0.20	0.60	0.10	0.30	0.00	0.00	0.20
Mozambique	11	0.91	0.73	0.64	0.18	0.55	0.27	0.64	0.55	0.27	0.27	0.36	0.00	0.18	0.45	0.09	0.09
Nepal	15	0.27	0.00	0.00	0.07	0.27	0.33	0.20	0.67	0.07	0.00	0.27	0.07	0.20	0.00	0.00	0.07
New Zealand	17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.00	0.06	0.18	0.12	0.12	0.00	0.00	0.18
Saudi Arabia	10	0.70	0.70	0.80	0.00	0.70	0.50	0.30	0.50	0.20	0.10	0.00	0.10	0.00	0.30	0.10	0.40
South Korea	12	0.33	0.08	0.08	0.42	0.17	0.08	0.00	0.25	0.25	0.33	0.33	0.25	0.08	0.00	0.00	0.00
UK	7	0.00	0.14	0.00	0.00	0.00	0.14	0.29	0.43	0.29	0.43	0.29	0.43	0.00	0.00	0.00	0.00
United States	14	0.14	0.21	0.14	0.07	0.36	0.07	0.36	0.71	0.64	0.79	0.50	0.50	0.29	0.00	0.00	0.07
Venezuela	10	0.90	0.60	0.40	0.40	0.80	0.30	0.30	0.70	0.10	0.00	0.20	0.00	0.10	0.00	0.10	0.10
Vietnam	20	0.75	0.40	0.95	0.25	0.90	0.65	0.50	0.85	0.05	0.00	0.30	0.20	0.45	0.05	0.05	0.05

Comparison with V-Dem

The Varieties of Democracy (V-Dem) Project serves as a useful comparison due to its similar methodology. While V-Dem does not measure all the rights on which HRMI focuses, it does include indicators of the rights to freedom from torture and political killing. V-Dem does not explicitly tie the definitions of these rights to international law. The definition of torture used by V-Dem is the closest to its international legal definition, particularly as contained in the CAT (Coppedge et al., 2019). Even in that case, V-Dem uses a more constrained definition than we do and limits ‘torture’ to acts committed with the intent to ‘extract information or intimidate victims, who are in a state of incarceration’ (Coppedge et al., 2019: 161). Consistent with the CAT, we allow for torture to be for ‘any specific purpose.’ While HRMI’s practitioner survey defines extrajudicial killing as including ‘any death caused by government actions outside of armed combat, immediate self-defense, or a legally-sanctioned death penalty,’ V-Dem’s measure of political killing only includes killings committed by government actors ‘for the purpose of eliminating political opponents’ (Coppedge et al., 2019: 161).

Even with these differences in definitions, we would expect our measures to correlate positively with theirs. Figure 6 shows the 2018 V-Dem indicators plotted against the relevant HRMI indicator. The figure shows a Pearson correlation (ρ) and a rank correlation (Kendall's tau) for each pair of indicators. The torture scales are relatively highly correlated, though notably Brazil, Mexico, and the United States do relatively worse on the HRMI scale than on V-Dem. The extrajudicial killing scales are only moderately correlated, with most of the difference due to differences in scores for Brazil, Mexico, and the United States. The positive correlations provide indirect evidence for the validity of our measures, but also show we are not duplicating existing scales. Differences can presumably be explained by our broader definition of extra-legal killing (which, for example, captures excessive use of lethal force by the police against suspected criminals) and the fact that our respondents rely on information sources that are distinct from those used by V-Dem respondents.

Figure 6. Comparison with V-Dem indicators



Conclusion

We believe the HRMI approach produces data on civil and political rights that are valid, reliable, and accessible to academics and practitioners alike. First, the HRMI data collection approach has been co-designed in consultation with the practitioner community and relies heavily on engagement and interaction with that community. Second, a practitioner-survey methodology is less likely to produce the biases demonstrated by international media or NGO reporting and provides us with a previously untapped source of human rights information. Third, the methods we use to convert intensity information into quantitative metrics allow us to be honest about uncertainty and permit sensible cross-country comparisons. Fourth, HRMI collects data not only on the intensity of abuse, but also on its distribution, opening new avenues for exploring and understanding the nature of human rights violations.

HRMI publishes data on a growing subset of rights and countries every year. We want to expand the sample of countries to include the global population, while at the same time expanding our coverage of rights to include all of those included in core international human rights treaties. Still, HRMI's data already allow for analyses that have not previously been conducted in the quantitative human rights literature. While there is a long history of analyses of the intensity of physical integrity rights abuse (e.g. Poe, Tate & Keith, 1999; Hill & Jones, 2014), cross-national analyses of the distribution of such abuses in the population, including abuses of civil and political rights beyond the traditional four physical integrity rights, have been almost non-existent. Why do states target some people for repression and not others?

What explains the violation of civil and political rights for those that are not seen as political rivals? HRMI data can begin to answer questions such as these immediately. Further, if HRMI achieves its goal of comprehensive rights and country coverage, it will open entire areas of cross-national human rights research beyond the physical integrity rights that have received the bulk of attention in the literature thus far.

Replication data

The dataset, codebook, and script files for the empirical analysis in this article can be found at <http://www.prio.org/jpr/datasets>. Analyses were conducted using R and Stata. For up-to-date information on HRMI's latest data and analysis, please visit <https://humanrightsmeasurement.org/>.

Acknowledgments

We are indebted to hundreds of people who have helped make the HRMI civil and political rights data a reality. In particular, we would like to thank all of the survey respondents for their time and knowledge, and the HRMI ambassadors, who helped connect us with potential survey respondents. Further, we are grateful to the participants of the HRMI co-design workshops, and the many others who have given their time to help us design and translate our work. Finally, we would like to thank all of the other members of the HRMI team, past and present.

Funding

This research was supported in part by a grant from the Open Society Foundations.

References

- Amnesty International (2016) Amnesty International Report 2015/16: The state of the world's human rights (<https://www.amnesty.org/download/Documents/POL1025522016ENGLISH.PDF>).
- Bagozzi, Benjamin E; Daniel W Hill, Jr, Will H Moore, & Bumba Mukherjee (2015) Modeling two types of peace: The zero-inflated ordered probit (ZiOP) model in conflict research. *Journal of Conflict Resolution* 59(4): 728-752.
- Bakker, Ryan; Seth Jolly, Jonathan Polk, & Keith Poole (2014) The European common space: Extending the use of anchoring vignettes. *Journal of Politics* 76(4): 1089-1101.
- Barry, Colin M; Sam R Bell, K Chad Clay, Michael E Flynn, & Amanda Murdie (2015) Choosing the best house in a bad neighborhood: Location strategies of human rights INGOs in the Non-Western world. *International Studies Quarterly* 59: 86-98.
- Bollen, Kenneth A (1986) Political rights and political liberties in nations: An evaluation of human rights measures, 1950 to 1984. *Human Rights Quarterly* 8(4): 567-591.
- Cingranelli, David L; David L Richards & K Chad Clay (2014a) The CIRI human rights dataset. Version 2014.04.14 (<http://www.humanrightsdata.com>).
- Cingranelli, David L; David L Richards & K Chad Clay (2014b) The Cingranelli-Richards (CIRI) human rights data project coding manual. Manual version 5.20.14 (<http://www.humanrightsdata.com/p/data-documentation.html>).
- Clark, Anne Marie; & Kathryn Sikkink. (2013) Information effects and human rights data: is the good news about increased human rights information bad news for human rights measures? *Human Rights Quarterly* 35(3):539–568.
- Conrad, Courtenay R; & Will H Moore (2011) The Ill-Treatment and Torture (ITT) data collection project country-year data user's guide. 19 July (http://faculty.ucmerced.edu/cconrad2/Academic/Data_files/ITT_CY_UsersGuide19July11.pdf).
- Conrad, Courtenay R; Jillienne Haglund, & Will H Moore (2014) Torture allegations as events data: Introducing the Ill-Treatment and Torture (ITT) specific allegation data. *Journal of Peace Research* 51(3):429–438.
- Conrad, Courtenay R; Daniel W Hill, Jr, & Will H Moore (2018) Torture and the limits of democratic institutions. *Journal of Peace Research* 55(1): 3–17
- Conrad, Courtenay R; & Will H Moore (2010) The Ill-Treatment and Torture (ITT) data project coding rules and norms. 5 May (http://faculty.ucmerced.edu/cconrad2/Academic/Under_the_Hood_files/ITTCodebookNorms6May10.pdf).
- Conrad, Courtenay R; & Will H Moore (2011) The Ill-Treatment and Torture (ITT) data collection project country-year data users' guide. July 19. Version 2.1 (http://faculty.ucmerced.edu/cconrad2/Academic/Data_files/ITT_CY_UsersGuide19July11.pdf).
- Cope, Kevin L; Charles Crabtree, & Christopher J. Fariss (2020) Patterns of disagreement in indicators of state repression. *Political Science Research and Methods* 8(1): 178-187.
- Coppedge, Michael; John Gerring, Carl Henrik Knutsen, Staffan I Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Anna Lührmann, Kyle L Marquardt, Kelly McMann, Pamela Paxton, Daniel Pemstein, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Agnes Cornell, Lisa Gastaldi, Haakon Gjerløw, Valeriya Mechkova, Johannes von Römer, Aksel Sundtröm, Eitan Tzelgov, Luca Uberti, Yi-ting Wang, Tore Wig, &

- Daniel Ziblatt (2019) V-Dem Codebook v9. Varieties of Democracy (V-Dem) Project (<https://www.v-dem.net/en/data/archive/previous-data/data-version-9/>).
- Fariss, Christopher J (2014) Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review* 108 (2): 297–318.
- Fukuda-Parr, Sakiko; Terra Lawson-Remer, & Susan Randolph (2015) *Fulfilling Social and Economic Rights*. New York: Oxford University Press.
- Gibney, Mark; Linda Cornett, Reed Wood, Peter Haschke, & Daniel Arnon (2015) Political Terror Scale 1976-2015 (<http://www.politicalterroryscale.org>). Last Accessed: 18 March 2018.
- Global Alliance of National Human Rights Institutions (GANHRI) (2019) Chart of the status of national institutions. 4 March ([https://nhri.ohchr.org/EN/Documents/Status%20Accreditation%20Chart%20\(04%20March%202019.pdf\)](https://nhri.ohchr.org/EN/Documents/Status%20Accreditation%20Chart%20(04%20March%202019.pdf))).
- Goldstein, Robert Justin (1986) The limitations of using quantitative data in studying human rights abuses. *Human Rights Quarterly* 8(4):607–627.
- Hare, Christopher; David A Armstrong, II, Ryan Bakker, Royce Carroll, & Keith T Poole (2015) Using Bayesian Aldrich-McKelvey scaling to study citizens' ideological preferences and perceptions. *American Journal of Political Science* 59 (3): 759-774.
- Hendrix, Cullen S; & Wendy H Wong (2014) Knowing your audience: How the structure of international relations and organizational choices affect amnesty international's advocacy. *The Review of International Organizations* 9 (1): 29-58.
- Hill, Daniel W, Jr.; Will H Moore, & Bumba Mukherjee (2013) Information politics versus organizational incentives: When are Amnesty International's "naming and shaming" reports biased? *International Studies Quarterly* 57: 219-232.
- Hill, Daniel W, Jr.; & Zachary M Jones (2014) An empirical evaluation of explanations for state repression. *American Political Science Review* 108 (3): 661-687.
- Human Rights Measurement Initiative (HRMI) (2018) Human Rights Measurement Initiative website (<https://humanrightsmmeasurement.org/>).
- McCoy, Alfred W (2012) In the minotaur's labyrinth: Psychological torture, public forgetting, and contested history. In: Julie A Carson & Elisabeth Weber (eds.) *Speaking About Torture*. Fordham University Press, 37-58.
- Poe, Steven C; C Neal Tate, & Linda Camp Keith (1999) Repression of the human right to personal integrity revisited: A global cross-national study covering the years 1976-1993. *International Studies Quarterly* 43 (2): 291-313.
- Rejali, Darius (2007) *Torture and Democracy*. Princeton: Princeton University Press.
- Stohl, Michael; David Carleton, George Lopez, & Stephen Samuels (1986) State violation of human rights: Issues and problems of measurement. *Human Rights Quarterly* 8(4):592–606.
- United Nations (2010) National human rights institutions: History, principles, roles and responsibilities. Professional Training Series, Number 4 (Rev. 1) (http://www.ohchr.org/Documents/Publications/PTS-4Rev1-NHRI_en.pdf).
- United States Department of State, Bureau of Democracy, Human Rights, and Labor (USDS) (2016) 2015 country reports on human rights practices: Egypt. 13 April (<https://www.state.gov/j/drl/rls/hrrpt/2015/nea/252921.htm>).

Bio sketches

K CHAD CLAY, b. 1980, PhD in Political Science (Binghamton University, 2012); Associate Professor, University of Georgia (2012-); Co-Founder and CPR Metrics Lead, HRMI.

RYAN BAKKER, b. 1971, PhD in Political Science (University of North Carolina, 2007); Reader, University of Essex (2019-).

ANNE-MARIE BROOK, b. 1970, MPA in Economics (Princeton University, 1997); Policy Fellow, Motu Economic and Public Policy Research (2016-); Co-Founder, HRMI.

DANIEL W HILL, JR, b. 1982, PhD in Political Science (Florida State University, 2012); Associate Professor, University of Georgia (2012-).

AMANDA MURDIE, b. 1980, PhD in Political Science (Emory University, 2009); Professor, University of Georgia (2016-).