Political Science Research and Methods (2020), **8**, 178–187
doi:10.1017/psrm.2018.62

RESEARCH NOTE

Patterns of disagreement in indicators of state repression

Kevin L. Cope[1], Charles Crabtree[2*] and Christopher J. Fariss[2]

[1]School of Law, University of Virginia, Charlottesville, VA, USA and [2]Department of Political Science, University of Michigan, Ann Arbor, MI, USA
*Corresponding author. Email: ccrabtr@umich.edu

Abstract
Until recently, researchers who wanted to examine the determinants of state respect for most specific negative rights (i.e., physical integrity and empowerment rights) needed to rely on data from the CIRI or the Political Terror Scale (PTS). The new Varieties of Democracy (V-Dem) dataset offers scholars a potential alternative to the individual human rights variables from CIRI. We analyze a set of key Cingranelli–Richards (CIRI) Human Rights Data Project and V-Dem negative rights indicators, finding unusual and unexpectedly large patterns of disagreement between the two sets. First, we discuss the new V-Dem dataset by comparing it to the disaggregated CIRI indicators, discussing the history of each project, and describing its empirical domain. Second, we identify a set of disaggregated human rights measures that are similar across the two datasets and discuss each project's measurement approach. Third, we examine how these measures compare to each other empirically, showing that they diverge considerably across both time and space. These findings point to several important directions for future work, such as how conceptual approaches and measurement strategies affect rights scores. For the time being, our findings suggest that researchers should think carefully about using the measures as substitutes.

Recently, one of the most vibrant lines of international law, comparative politics, and international relations research has been the quantitative study of government repression. This literature typically models the determinants of states' rights performance, exploring how institutions like economic and military foreign policy, international conventions,[1] civil society, and domestic-government structure, among other factors, impact negative rights outcomes (Hill and Jones 2014). These studies have resulted in the creation of the "standard model": a common model specification that regresses state respect for human rights on a series of variables capturing differences in "domestic and external threats (civil and/or international war), regime type (democracy, military, and leftist), and socio-economic conditions (economic development, population size, and colonial legacy)" as well as international legal commitments (Keith 2012, 79).

The outcome measure in most of these models is some index or indicator that measures state respect for physical integrity rights[2] or empowerment rights.[3] Among the most commonly used outcomes are the Political Terror Scale (PTS), the CIRI Physical Integrity Rights Index (Cingranelli,

[1]See Cope and Creamer (2016) for a summary and analysis of this literature.

[2]These rights include protection from torture, extrajudicial murder, forced disappearance, and political imprisonment (Goldstein 1978).

[3]These rights include electoral self-determination, the right to domestic movement, the right to foreign movement, the right to religious freedom, the right to freedom of speech, and the right to assembly and association (Cingranelli, Richards and Clay 2015).

© The European Political Science Association, 2019.

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press

Richards and Clay 2015), the CIRI Empowerment Rights Index, the latent Human Rights Protection Scores (Fariss 2014), and the DO-IRT Empowerment Rights Index (Schnakenberg and Fariss 2014).[4] Each measure assesses states based on their *aggregate* level of performance on physical integrity rights, empowerment rights, or both. In other words, they measure how well states perform across a bundle of rights.

In many cases, however, researchers are interested in testing theories that relate to specific negative rights, such as extrajudicial killing. Until recently, researchers who wanted to examine the determinants of state respect for most individual rights needed to rely on CIRI's data.[5] In 2016, the Varieties of Democracy Project (V-Dem) dataset was fully released, offering researchers an alternative. V-Dem offers researchers a rich new dataset, covering 173 countries and colonies from 1900 to 2015.[6] It contains 350 unique indicators related to democracy, including several measures of negative rights. These include, among others, freedom from torture and political killings, and freedom of speech and religion. Researchers are increasingly likely to use these V-Dem scores as an alternative or complement to their CIRI counterparts; the project has already been cited over 300 times since 2013.[7]

Though the CIRI and V-Dem measures ostensibly capture similar underlying constructs, we show that the CIRI–V-Dem measurement pairs diverge significantly across time periods. Moreover, the measures are negatively correlated for a large set of countries. These findings stand in contrast to the relatively high correlations between other sets of popular negative rights measures and suggest that researchers should use caution before using the two measures as substitutes for one other. They also raise important broader questions about how rights measures are designed.

The CIRI and V-Dem rights measures

Cingranelli and Richards' CIRI measures have been used more than any other disaggregated human rights data, appearing in at least 1136 studies published in comparative politics, political economy, international relations, international law, and other law and social science journals. The dataset quantifies information contained in the U.S. State Department's *Country Reports on Human Rights Practices*. For physical integrity rights, it also considers AI's *World's Human Rights Reports*.[8] From these reports, CIRI coders produce ordinal-level country ratings on 15 state negative human rights practices. The dataset currently covers 202 countries from 1981 to 2011.

Released nearly two decades later, V-Dem is an international data collaboration project that currently covers 173 countries and colonies, spanning the period 1900–2015 (Coppedge et al. 2011, 2015; Pemstein et al. 2015). Of the more than 350 unique indicators related to democracy, about half of them are based on "factual information obtainable from official documents such as constitutions and government records."[9] The others are based on subjective decisions by approximately 2500 country experts around the world, many of whom have lived or worked in the countries they are rating (Pemstein et al. 2015). In most cases, the expert coders rely on their own impressions of countries or refer to secondary sources about country conditions.[10] The

[4]PTS measures states' protection of physical integrity was introduced in 1982 and has since been used in over 500 published articles. The CIRI scores, which include both summative and disaggregated measures of physical integrity rights and empowerment rights, were developed in 1994. Fariss (2014) published a latent variable measure of state respect for physical integrity rights two decades later.

[5]Scholars who want to test theories related to torture can use a wider set of indicators that includes the Ill-Treatment and Torture (ITT) dataset (Conrad, Haglund and Moore 2013) and the Hathaway Torture Data (Hathaway 2002).

[6]Some indicators extend only to 2012.

[7]Google Scholar search for "'Varieties of Democracy' and 'Coppedge,"' performed October 2, 2017.

[8]Where the two disagree, CIRI considers the AI reports as more authoritative (http://www.humanrightsdata.com/p/faq.html).

[9]https://www.v-dem.net/media/filer_public/f9/08/f908eb53-c0e2-40f0-9294-e067537d8f0b/v-dem_policybrief_5_2016.pdf

[10]We thank several anonymous V-Dem coders for this information.

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press

project uses multiple coders per country-year unit in a process that was coordinated by over 50 academics across multiple universities and institutes. From these data, V-Dem generates latent scores based on categorical questions answered by each coder for each country-year item.[11] The model accounts for disagreement between coders and it generates measurements of uncertainty conditional on the number of and agreement between coders, as well as coder reliability over time (Coppedge et al. 2015). The V-Dem data are increasingly used in comparative politics, political economy, international relations, and sociology.[12]

Both datasets include a range of measures that capture how well states respect negative human rights. Specifically, they provide measures of the extent to which states respect individual physical integrity rights, empowerment rights, or bundles of these rights. Our focus here is on the disaggregated measures that capture state performance in relation to a *particular* right. CIRI provides 15 such measures. Depending on how negative human rights are defined, V-Dem provides somewhere between two and several dozen of these measures. Many of them examine specific subcomponents of disaggregated rights. For instance, V-Dem does not provide a measure of free speech generally; it disaggregates free speech into several measures that capture specific aspects of free speech, such as *freedom of discussion* and *freedom of discussion for women*.

The codebooks for each data project conceptualize these six measures in ways that share many similarities but also include key differences. For example, both sets measure torture, but they define it differently. The CIRI standard defines torture in part as, "the purposeful inflicting of extreme pain, whether mental or physical, by government officials or by private individuals at the instigation of government officials." The V-Dem torture indicator defines torture more narrowly, as "the purposeful inflicting of extreme pain, whether mental or physical, with an aim to extract information or intimidate victims, who are in a state of incarceration." It therefore requires that victims be incarcerated, and it limits the torturer's motive to "extract[ing] information or intimidat[ing] victims," thus excluding, among other things, the motive of "punishment" and "discrimination." Torture as defined by various international law sources often occurs outside of the context of incarceration. So even without observing the data, we would expect the CIRI measure to count many more torture incidents than its V-Dem counterpart.

Just as the CIRI and V-Dem measures capture related but not identical underlying constructs, CIRI and V-Dem measure the six variables listed in Table 2 below in ways that are similar in some respects but different in others. Methods for measuring repression levels can be divided into three broad approaches based on how the source material is quantified: (1) events-based; (2) standards-based, and (3) survey-based (Landman 2004).[13] CIRI claims to use a "standards-based" approach, meaning that it attempts to assess the frequency and level of violations in some time and place without documenting or counting specific instances. It claims to do this by comparing actual human rights practices in any given year to contemporary international law standards (Cingranelli and Richards 2010, 405–06). Yet some commentators argue that CIRI's method is actually events-based. For instance, Wood and Gibney (2010) remind us that the original CIRI approach was based on events-based thresholds such that the categories 2, 1, and 0 correspond to zero violations ("full respect"), 1-to-49 violations ("moderate respect"), and

[11]There are some rare exceptions to this rule. Thanks to the V-Dem team for alerting us to this.

[12]https://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse journal_id = 2602515SSRN's V-Dem Institute Working Paper Series already included more than 70 papers as of December 28, 2017. Many papers not in this series use V-Dem indicators in supplementary analyses or in robustness checks.

[13]Events-based methods attempt to identify every instance of rights violation, including information on things like context, type of right violated, severity, victim, perpetrator, and motive. These instances are aggregated to arrive at a tally or level of violation for a given time and place, which can take the form of an interval or ratio variable. Standards-based methods instead assess the frequency and level of violations in some time and place without documenting or counting specific instances. Those violation levels are assigned to an ordinal scale. Survey-based methods sample relevant populations, usually to derive estimates of public perceptions of violations. This approach therefore measures directly only subjective impressions of violations, but they are often treated as a proxy for actual, objective levels of violations.

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press

50-or-more violations ("no respect"), respectively, and therefore describe a count of discrete rights violations. Though this was originally one of the coding guidelines developed by the CIRI team, they discovered it was not practical for most of the country-year reports that they coded.

V-Dem also claims to use a standards-based approach. Unlike other standards-based indicators, the V-Dem project controls the standards used to assess each of its variables (i.e., the questions' wording and format, which is displayed in Online Appendix A). Moreover, because the coders complete the questions over the relatively short time span of four years, it is unlikely that the V-Dem human rights scores are temporally biased in the same way that standards-based human rights data might be (Fariss 2014). That is, unlike the human rights reports, the V-Dem data are based on question responses that are *produced* consistently with respect to time. However, like events-based data, the V-Dem expert coders rely on their knowledge of evidence from the historical record, and the historical record provides different levels of information for certain cases.

Table 1 summarizes the several characteristics of the CIRI and V-Dem measures. Given their nominally similar concerns, researchers might assume that the two measures correlate strongly, even if they note that descriptions in the respective indicators' codebooks differ. This intuition is understandable given that previous comparisons of leading aggregated measures have found fairly high correlations between the measures. For instance, the Kendall's τ-b coefficient for CIRI and the PTS is approximately 0.73 (for those PTS scores derived from State Department reports) and 0.65 (for those derived from AI reports) (Wood and Gibney 2010, 375).

So it would be reasonable to assume that CIRI and V-Dem are good substitutes. That assumption might lead researchers to choose one over the other based on traits like past or recent popularity, scope of time, or countries covered. It could also lead researchers to replace one measure for another as a robustness check. In fact, our conversations with researchers suggest that this is already widely done, even if not always reported in papers.

But treating two measures as interchangeable could be a pitfall for researchers, particularly if the measures are in fact concerned with different constructs. For instance, an indicator can be *institution-oriented* or *individual-oriented*. Institution-oriented indicators are concerned with compliance, that is, how well government institutions perform concerning individual rights or bundles of rights. In contrast, individual-oriented indicators are concerned with enjoyment—to what extent the country's rights performance affects a typical national resident. Using an individual-based measure for a question relating to government rulemaking (or an institution-based

Table 1. Overview of Datasets

Dataset	Source Material	Years	Countries	Scale	Approach
CIRI	AI reports, State Department reports	1981–2011	202	0 (worst)–2 (best)	(Contested)
V-Dem	Primary source documents–Original data provided by expert coders	1900–2012/14	173	0 (worst)–4 (best)	Standards-based

Note: Table 1 describes the source material for each dataset, its empirical coverage, scale, and measurement approach.

Table 2. Comparable Disaggregated Human Rights Measures

Right	CIRI	V-Dem
Physical integrity		
Extrajudicial killing	KILL	v2clkill
Torture	TORT	v2cltort
Empowerment		
Association	ASSN	v2x_frassoc_thick
Domestic movement	DOMMOV	v2xcl_dmove
Foreign movement	FORMOV	v2clfmove
Religious freedom	NEW_RELFRE	v2clrelig

Note: The columns contain the underlying construct they are trying to capture, the CIRI variable name, and the V-Dem variable name.

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press

measure for a question related to how citizens respond to repression) would answer a question different from the researcher's question of interest, producing misleading results.

One way to shed light on different constructs is to study the respective databases' codebooks, examining the assumptions underlying the measurement approaches. Another option is to do what we do here: empirically compare pairs of theoretically similar rights. This method may not tell the whole story, because significant divergence may stem from measurement error. But where empirical correlation is low, it constitutes strong evidence that the two scores are not in fact measuring the same phenomenon.

An empirical comparison of the disaggregated measures

To begin, we selected a set of six broadly comparable rights that often interest empirical researchers. They include the physical integrity rights of (1) protection from extrajudicial killing and (2) protection from torture, and the empowerment rights of (3) freedom of domestic movement, (4) freedom of foreign movement, (5) freedom of religion, and (6) freedom of association. For each of those rights, we reviewed the codebooks for each dataset and identified one measure from CIRI and one from V-Dem that appeared to most closely capture that right. We took a conservative approach to identifying comparable indicators in CIRI and V-Dem, requiring high congruence between indicator definitions. Table 2 lists these CIRI–V-Dem measurement pairs. The first column provides a general description of the broadly conceived underlying construct they purport to capture.

We examine the extent to which these measures agree by constructing a panel dataset that contains the six CIRI and V-Dem variables listed in Table 2 for all 5647 country-year observations shared across both datasets. We then evaluate the degree to which the measures agree (Lee Rodgers and Nicewander 1988) by calculating the Spearman's rank correlation coefficient, or Spearman's ρ, for each of the six CIRI–V-Dem measurement pairs.[14] The correlation coefficient is calculated using the CIRI value and the mean V-Dem posterior value for each observation.[15] The statistic indicates the extent to which the measures agree with each other, with values closer to 1 representing increased agreement and values closer to −1 representing increased disagreement.

We first calculate ρ for each measure-measurement pair using all available country-year observations. Since we want to compare the coded values for country-years, we drop all observations that are not coded by both data projects.[16] Table 3 summarizes these calculations. The first column contains the name of the right measured by the CIRI–V-Dem pair, the second contains ρ for that pair, the third contains the number of observations coded by both measures, and the fourth column contains the number of country years dropped because of missingness in CIRI or V-Dem measures.

This analysis reveals two important patterns. One is that the correlations for empowerment rights measures are higher than the correlations for physical integrity rights. An explanation for this could be that it is easier to observe violations of empowerment rights, since they often occur in more public settings than violations of physical integrity rights. In addition, governments rarely take the same measures to conceal violations of civil rights. The other pattern that we observe is a moderate relationship between nearly all of the measurement pairs. The range of correlation coefficients is 0.471–0.741, and the average is 0.562. In comparison, most measures of democracy correlate with each other at 0.8 or higher (Pemstein et al. 2010). We might expect that some human rights measures would correlate at even higher levels than democracy measures do. Though concepts like "torture" and "extrajudicial killing" are disputed in the rights, psychology, and medical literatures (e.g., McDonnell et al. 2011), and in law (e.g., Waldron 2005), they are still perhaps clearer than the long-contested concept of democracy. And as mentioned above, the Kendall's τ-b coefficient for

[14]Spearman's ρ is the appropriate measure of association to use since we are comparing ordinal (CIRI) and continuous (V-Dem) indicators (Carroll 1961).

[15]This means that we do not incorporate information about the uncertainty of V-Dem estimates into the calculation.

[16]In many cases, it would be appropriate to impute missing values. Since we want to compare the *coded* values across datasets, we do not consider imputation appropriate here.

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press

Table 3. Correlations Between CIRI–V-Dem Measurement Pairs

Right	ρ	N	Dropped
Physical integrity rights			
Torture	0.471	4436	1211
Extrajudicial killing	0.526	4471	1217
Empowerment rights			
Association	0.741	4441	1206
Domestic movement	0.476	4682	965
Foreign movement	0.585	4684	963
Religious freedom	0.575	4677	970

Note: Table 3 shows the correlation coefficient for each CIRI–V-Dem measurement pair. The first column contains the name of the right measured by the CIRI–V-Dem pair, the second contains the calculated ρ for that pair, the third contains the number of observations coded by both measures, and the fourth contains the number of country years dropped because of missingness in CIRI or V-Dem measurements or in any country-year information.

CIRI and the PTS is approximately 0.73 (State Department-derived PTS reports) and 0.65 (AI-derived reports) (Wood and Gibney 2010, 375). Our initial comparison of the CIRI and V-Dem indicators suggests that the measures disagree far more than most probably would have expected.

By looking only at the correlation coefficients for the entire sample, we might miss important divergences within subsets of the sample. For instance, it might be that these measures converge more for some years. If so, we would probably expect more agreement for more recent years; recent years' records are better, and the V-Dem coders' memories about them are fresher. We calculate the pairwise correlations for individual years. Figure 1 plots ρ for each measurement pair across time. Larger values indicate increased agreement over human rights practices within years. Between the CIRI and V-Dem data projects on extrajudicial killing and religious freedom, the plots are consistent with our expectations: agreement is higher for more recent years. The ρs for both sets of measurement pairs increase by about 50 percent from the first year in our data to the last year. But surprisingly, as to association and foreign movement rights, agreement *declines* as time progresses. The correlation between both sets of measures decreases by approximately 30 percent from 1981 to 2011. Curiously, agreement initially decreases over time for the domestic movement and torture measurement pairs and then increases.

These plots present several new puzzles for human rights researchers. Future work should examine why different sources agree more in some years than others. Scholars might also investigate why agreement about rights practices in recent years varies across datasets.

Next, we investigate whether the measures correlate better within some places than others. To do this, we calculate ρ for all countries in our sample.[17] The panels in Figure 2 plot the correlation coefficients for each pairwise measurement comparison by country name.[18] Larger values indicate increased agreement about human rights practices within countries. The horizontal gray line denotes $\rho = 0$.

In this case, there are no clear patterns. Rather, the plotted correlation coefficients for each human right appear as if they were randomly drawn from a uniform distribution bound between $[-1, 1]$. For each right, the range of correlation coefficients covers nearly this entire interval. The correlation coefficients for extrajudicial killing cover the smallest range (−0.789 to 0.866) while the correlation coefficients for domestic movement cover the largest range (−1.000 to 1.000). This means that the measures substantially diverge when it comes to measuring trends within countries over time.

[17]In calculating the correlation for countries, we have to drop (a) countries where one or both measures do not vary over time and (b) countries with fewer than two observations per measure. This is because a correlation coefficient cannot be estimated for those countries. Online Appendix A contains additional details on the number of countries dropped per measurement pair.

[18]For scholars who are interested in identifying ρ for particular countries and measurement pairs, we provide in Online Appendices B-G a series of dot plots with country names on the vertical axis and correlation coefficients on the horizontal axis.

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press

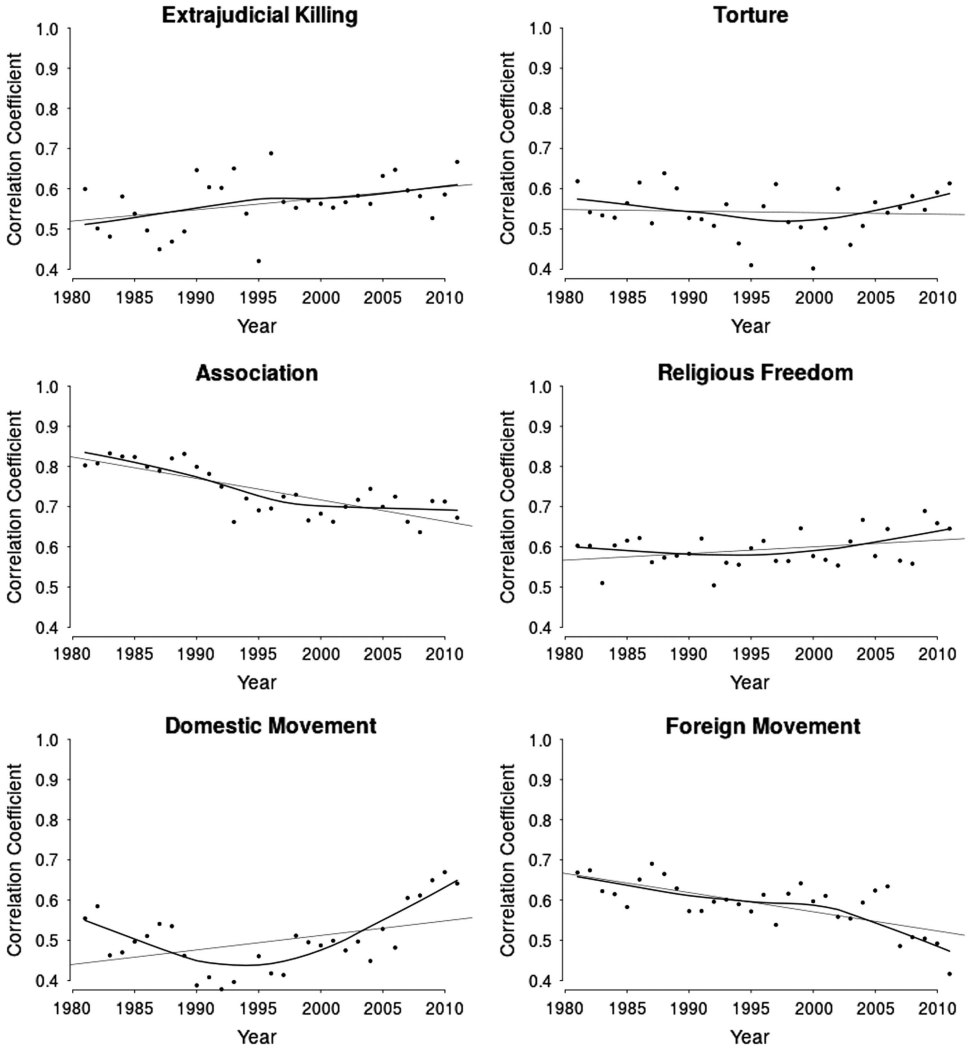


Figure 1. Within-year correlations between CIRI–V-Dem measurement pairs
Note: Figure 1 plots the correlation coefficient for each measurement pair across time.

Surprisingly, the measures negatively correlate for approximately half the countries in the sample. In other words, while one measure records relatively higher respect for human rights, the other records relatively lower respect for human rights. Table 4 shows the number and percentage of countries in which rights measures are negatively correlated. The first column contains the name of the right measured by the CIRI–V-Dem pair. The second column contains the number of countries for which the relationship between the paired measures is negative along with the total number of countries included in the measurement comparison. The third column presents the percentage of countries for which the relationship between the CIRI and V-Dem measures is negative. In line with the patterns we observe in Table 3, the data projects disagree more about state respect for physical integrity rights within countries than state respect for empowerment rights. The CIRI and V-Dem physical integrity rights measures are, on average, negatively correlated for about 37 percent of the countries in our data. In contrast, the CIRI and V-Dem empowerment rights measures are negatively correlated for around 25 percent of countries on average.

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press

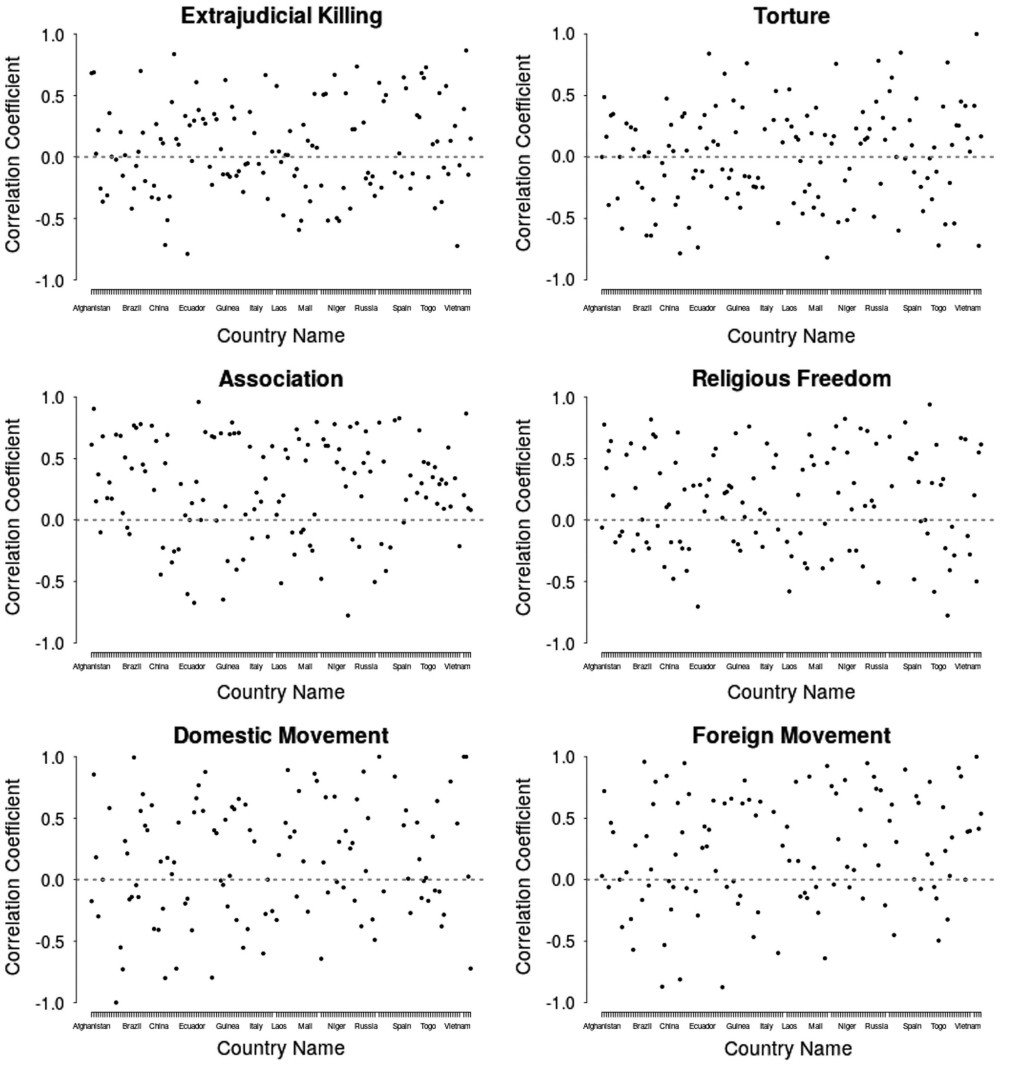


Figure 2. Within-country correlations between CIRI–V-Dem measurement pairs
Note: Figure 2 plots the correlation coefficients from countries for each pairwise measurement comparison.

Finally, we examine the extent to which the measures agree within cases. The idea here is to see if some measures are consistently positively or negatively correlated for a set of countries that often interest both scholars and the public: China, Germany, Iran, Russia, South Africa, and the USA.[19] As we show in Table 5, the measures do not exhibit stability across these countries. The first column contains the name of the right measured by the CIRI–V-Dem pair, and the remaining columns contain the correlation coefficient for individual countries. We see that agreement varies considerably across measures within countries. In China, for example, the CIRI–V-Dem torture measures correlate at 0.090, while the extrajudicial killing measures correlate at −0.342. That the CIRI and V-Dem data projects disagree about the over-time human

[19]We focus on these countries here because each is a major geopolitical player that has recently received significant attention for its rights repression, its response to other countries' repression, or both. Others might be interested in the pairwise correlations for different countries. We present a series of dot plots in Online Appendices B–G that can be used to find those correlations.

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press

Table 4. Negative Within-Country Correlations Between CIRI–V-Dem Measurement Pairs

Right	Negative *N* (Total *N*)	% Negative
Physical integrity rights		
Extrajudicial killing	62 (167)	37
Torture	64 (167)	38
Empowerment rights		
Association	34 (167)	20
Domestic movement	48 (167)	29
Foreign movement	39 (167)	23
Religious freedom	48 (167)	29

Note: Table 4 shows the percentage of countries in which rights measures are negatively correlated. The first column contains the name of the right measured by the CIRI–V-Dem pair, and the second contains the number of countries for which the relationship is negative.

Table 5. Within-Country Correlations Between CIRI–V-Dem Measurement Pairs for Important Cases

Right	USA	China	Germany	Russia	South Africa	Iran
Physical integrity rights						
Extrajudicial killing	0.577	−0.342	NA	−0.175	−0.128	−0.052
Torture	0.257	0.090	−0.171	0.450	−0.013	−0.171
Empowerment rights						
Association	0.296	NA	NA	0.721	0.811	NA
Domestic Movement	NA	−0.408	NA	0.070	0.839	−0.402
Foreign movement	NA	−0.011	NA	0.740	0.896	−0.266
Religious freedom	NA	0.125	0.280	0.621	0.794	NA

Note: Table 5 shows the correlation coefficient between CIRI–V-Dem measurement pairs across several cases. The first column contains the name of the right measured by the CIRI–V-Dem pair and the remaining columns contain the ρ for individual countries.

rights trends within countries, and even within similar types of rights, raises significant questions about how they code state practices. Future work might examine this puzzle.

Taken together, the results of our empirical comparison of the CIRI–V-Dem measurement pairs suggest that researchers should be careful when considering them as substitutes. Indeed, the data projects disagree about human rights performance over time and within countries. Scholars might theorize about the causes of measurement disagreement in future work.

Conclusion

The CIRI and V-Dem data projects largely disagree about the extent of rights practices within years and within countries. Although the two measures appear on their face to measure the same underlying concept, looking below the surface reveals important potential conceptual differences. Indeed, they correlate only moderately within years. Even more surprisingly, the variables are only weakly or moderately correlated within countries, and, for a non-trivial number of countries, the measures are negatively correlated. In sum, the CIRI and V-Dem indicators disagree far more than most researchers would probably expect.

The source of the observed divergence probably stems from some combination of two distinct phenomena: (1) different conceptualizations and (2) different measurement strategies. When we observe empirical divergence between CIRI and V-Dem scores, we cannot know the exact cause of the divergence, though patterns of correlation may point us in the direction of certain explanations.

These findings indicate that researchers should think carefully about using the measures as substitutes. These findings also suggest several important directions for future work, such as how conceptual approaches (e.g., based in international law versus elsewhere, or institution- versus individual-oriented) and measurement strategies (e.g., based on independent research versus government/NGO documents, or interval versus ordinal scoring) affect rights scores. For the time being, researchers considering whether to use one of these measures might consider first studying closely the assumptions and measurement techniques that underlie them.

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press

The next step is to determine whether certain types of variables, periods, regions, or states differ more than others. If so, what explains this cross-measure variation? And perhaps most important, how do studies' choice of measures affect their findings? If findings are robust across different measures, it would of course boost our confidence in the results. If not, we would need to probe the possible determinants of measurement disagreement. If certain factors are systematically related to measurement agreement, then scholars should take these factors into account while conducting statistical work.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/psrm.2018.62

Acknowledgments. We thank Holger L. Kern, Ronald F. Inglehart, Matt Golder, Sona Golder, Jule Krüger, Dom Nardi, Kathryn Sikkink, and Yuri Zhukov for their comments. We also thank attendees of the 2017 American Political Science Association annual meeting. All data and code necessary to replicate the analysis presented in the article will be publicly available upon publication at dataverse repositories maintained by the authors. This work is dedicated to the memory of Will Moore.

References

Carroll JB (1961) The nature of the data, or how to choose a correlation coefficient. *Psychometrika* **26**(4), 347–72.

Cingranelli DL and Richards DL (2010) The Cingranelli and Richards (CIRI) Human Rights Data Project. *Human Rights Quarterly* **32**(2), 401–24.

Cingranelli DL, Richards DL and Clay KC (2015) 'The Cingranelli-Richards (CIRI) Human Rights Data Project Coding Manual Version 2014.04.14'. Available at http://www.humanrightsdata.com/p/data-documentation.html, accessed 12 March 2017.

Conrad CR, Haglund J and Moore WH (2013) Disaggregating torture allegations: introducing the ill-treatment and torture (ITT) country-year data. *International Studies Perspectives* **14**(2), 199–220.

Cope KL and Creamer C (2016) Disaggregating the human rights treaty regime. *Virginia Journal of International Law* **55**, 459–80.

Coppedge M, Gerring J, Altman D, Bernhard M, Fish S, Hicken A, Kroenig M, Lindberg SI, McMann K, Paxton P and Semetko HA (2011) Conceptualizing and measuring democracy: a new approach. *Perspectives on Politics* **9**(02), 247–67.

Coppedge M, Gerring J, Altman D, Bernhard M, Fish MS, Hicken A, Kroenig M, Lindberg SI, McMann K, Paxton P, Semetko HA, Skaaning S-E, Staton J and Teorell J (2015) "V-Dem Dataset v5." *Varieties of Democracy (V-Dem) Project*. Available at https://www.v-dem.net/en/data/data-version-5/

Fariss CJ (2014) Respect for human rights has improved over time: modeling the changing standard of accountability in human rights documents. *American Political Science Review* **108**(2), 297–318.

Goldstein RJ (1978) *Political Repression in Modern America, From 1870 to Present*. Cambridge, MA: G. K. Hall.

Hathaway OA (2002) Do human rights treaties make a difference? *Yale Law Journal* **111**(8), 1935–2042.

Hill Jr. DW and Jones ZM (2014) An empirical evaluation of explanations for state repression. *American Political Science Review* **108**(3), 661–87.

Keith LC (2012) *Political Repression Courts and the Law*. Philadelphia: University of Pennsylvania Press.

Landman T (2004) Measuring human rights: principle, practice, and policy. *Human Rights Quarterly* **26**(4), 906–31.

Lee Rodgers J and Nicewander WA (1988) Thirteen ways to look at the correlation coefficient. *The American Statistician* **42**(1), 59–66.

McDonnell M-HM, Nordgren LF and Loewenstein G (2011) Torture in the eyes of the beholder: the psychological difficulty of defining torture in law and policy. *Vanderbilt Journal of Transnational Law* **44**(2011), 87–122.

Pemstein D, Marquardt KL, Tzelgov E, Wang Y-T and Miri F (2015) 'The V-Dem measurement model: latent variable analysis for cross-national and cross-temporal expert-coded data'. Working Paper. Available at http://www.https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3167764.

Pemstein D, Meserve SA and Melton J (2010) Democratic compromise: a latent variable analysis to ten measure of regime type. *Political Analysis* **18**(4), 426–49.

Schnakenberg KE and Fariss CJ (2014) Dynamic patterns of human rights practices. *Political Science Research and Methods* **2**(1), 1–31.

Waldron J (2005) Torture and positive law: jurisprudence for the White House. *Columbia Law Review* **105**(6), 1681–750.

Wood RM and Gibney M (2010) The Political Terror Scale (PTS): a re-introduction and comparison. *Human Rights Quarterly* **32**(2), 367–400.

Cite this article: Cope KL, Crabtree C, Fariss CJ. (2020). Patterns of disagreement in indicators of state repression. *Political Science Research and Methods* **8**, 178–187. https://doi: 10.1017/psrm.2018.62

https://doi.org/10.1017/psrm.2018.62 Published online by Cambridge University Press