



PROJECT MUSE®

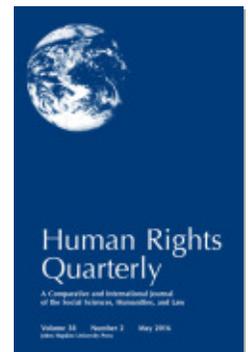
The Myth of Information Effects in Human Rights Data: Response to Ann Marie Clark and Kathryn Sikkink

David L. Richards

Human Rights Quarterly, Volume 38, Number 2, May 2016, pp. 477-492 (Article)

Published by Johns Hopkins University Press

DOI: [10.1353/hrq.2016.0033](https://doi.org/10.1353/hrq.2016.0033)



➔ For additional information about this article

<https://muse.jhu.edu/article/617749>

The Myth of Information Effects in Human Rights Data: Response to Ann Marie Clark and Kathryn Sikkink

David L. Richards*

I. INTRODUCTION

In August 2013, Ann Marie Clark and Kathryn Sikkink published “Information Effects and Human Rights Data: Is the Good News about Increased Human Rights Information Bad News for Human Rights Measures?” in this journal.¹ Their article examines important issues related to indicators of government respect for physical integrity rights from the Cingranelli-Richards (CIRI) Human Rights Data Project as well as the Political Terror Scale (PTS) index of physical integrity conditions. Using a data set about Latin America and the case examples of Brazil and Guatemala, they make an argument that an increased amount of information over time in the US State Department Country Reports on Human Rights Practices, as evidenced by growing word counts, is responsible for a downward bias in scores indicating government respect for human rights over time.²

* David L. Richards is Associate Professor of Human Rights and Political Science at The University of Connecticut. He is the author, with Jillienne Haglund, of *Violence Against Women and the Law* (Routledge, 2015) and the co-founder of the Cingranelli-Richards (CIRI) Human Rights Data Project. Email: david.richards@uconn.edu.

I wish to thank Jonathan Sykes for creating the word counts used herein. Thanks also to David Cingranelli, Michael Colaresi, and Amanda Murdie for their comments, all of which improved this work. Finally, I appreciate Ann Marie Clark meeting with me in 2011 to gain a better understanding of CIRI’s actual data-production processes before her article with Kathryn Sikkink was published.

1. Ann Marie Clark & Kathryn Sikkink, *Information Effects and Human Rights Data: Is the Good News about Increased Human Rights Information Bad News for Human Rights Measures?*, 35 *HUM. RTS. Q.* 539 (2013).
2. By “downwardly biased scores” I mean scores that show less government respect for human rights than they should because of measurement error.

While the article has already been cited in a number of prominent venues,³ its conclusions appear to have been accepted without scrutiny. However, given the extremely important impact that Clark and Sikkink's argument, if true, would have on the large number of published studies using CIRI and PTS in time-series form, it seems prudent to begin holding their claims to some inspection. Before settling into that task, I wish to say that I applaud without reservation the impetus for their article, which is that we should understand everything we can about the data we use to examine the world. With data, as with food, we are what we eat. Thus, caring as much or more about our data as we do our models is imperative for producing high-quality quantitative scholarship. Their article contributes to a highly important conversation.

In this article, I use a data set containing word counts of US State Department Country Reports on Human Rights Practices reports, as well as the CIRI Human Rights Data, to empirically address the viability of some of Clark and Sikkink's core arguments. I find the assumption that CIRI scores are declining in recent years, the conclusion that longer State Department reports have a substantive role in these lower scores, and the argument that changing standards affected the CIRI time series data, to be suspect.

II. DO THE DATA REALLY SHOW THINGS ARE GETTING WORSE?

Clark & Sikkink's key worry is that human rights scores are declining over time, while actual human rights practices may be getting better:

[S]cholars should keep in mind that the data are based on contemporaneous documents and understand the changing characteristics of the source materials over time. *Because of increased quality and quantity of information, the data may skew toward worse scores in later years.*⁴

They examine in detail the cases of Guatemala and Brazil as the basis for a generalized warning about the possibility of information effects downwardly biasing any country's data. However, if the information effects posed by Clark and Sikkink are in play at large in the CIRI and PTS data, we should see significant, downwards trends over time in the global data. Or, one could

3. Dara Kay Cohen, *Explaining Rape During Civil War: Cross-National Evidence (1980–2009)*, 107 AM. POL. SCI. REV. 461 (2013); Wade M. Cole & Francisco O. Ramirez, *Conditional Decoupling: Assessing the Impact of National Human Rights Institutions, 1981 to 2004*, 78 AM. SOCIOLOGICAL REV. 702 (2013); Christopher J. Fariss, *Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability*, 108 AM. POL. SCI. REV. 297 (2014); Jay Ulfelder, *It's Getting Better All The Time*, FOREIGN POL'Y, 24 Nov. 2015, available at <http://foreignpolicy.com/2015/11/24/its-getting-better-all-the-time-human-rights>.

4. Clark & Sikkink, *supra* note 1, at 567 (emphasis added).

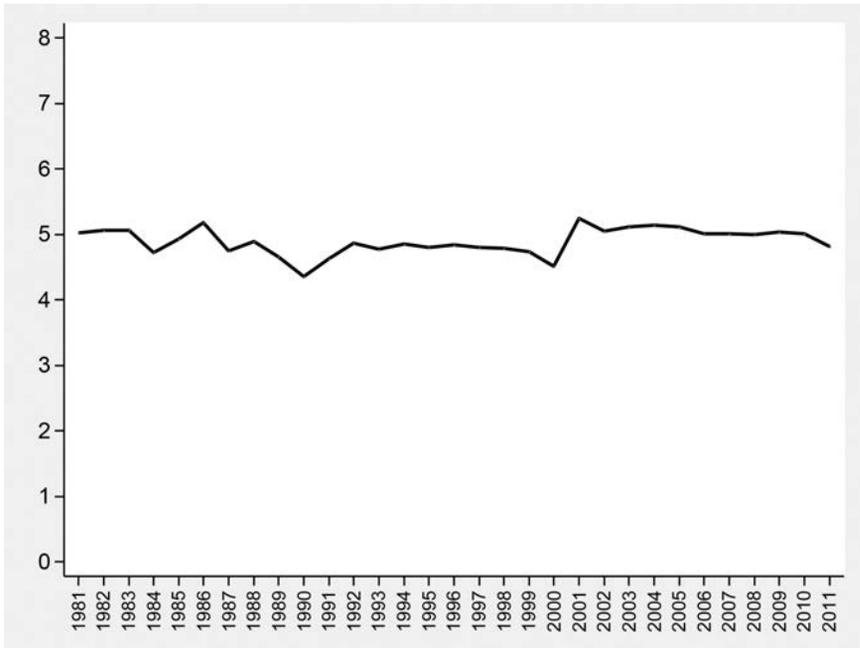


Figure 1. CIRI Physical Integrity Rights Index, Global Average, 1981–2011

also make the argument that stagnant trends in data -- in effect hiding real-world improvements-- might also be demonstrative of information effects. Both of these possibilities will be considered here.

A. Physical Integrity Rights

Let us look at a simple graph of the CIRI physical integrity index over time. This index combines measures of the level of government respect for extrajudicial killing, disappearance, political imprisonment, and torture into a nine-point additive ordinal index, ranging from 0 (no respect for any of these four rights) to 8 (full respect for all four rights).

What we see in Figure 1 is that the global average on this aggregate index of government respect for physical integrity rights is very stable over time. It is decidedly not decreasing over time, as one might expect were Clark and Sikkink's version of information effects at work. There is a small dip in average respect during the transition decade around the Cold War but, afterwards, there is a recovery in the average level of respect.

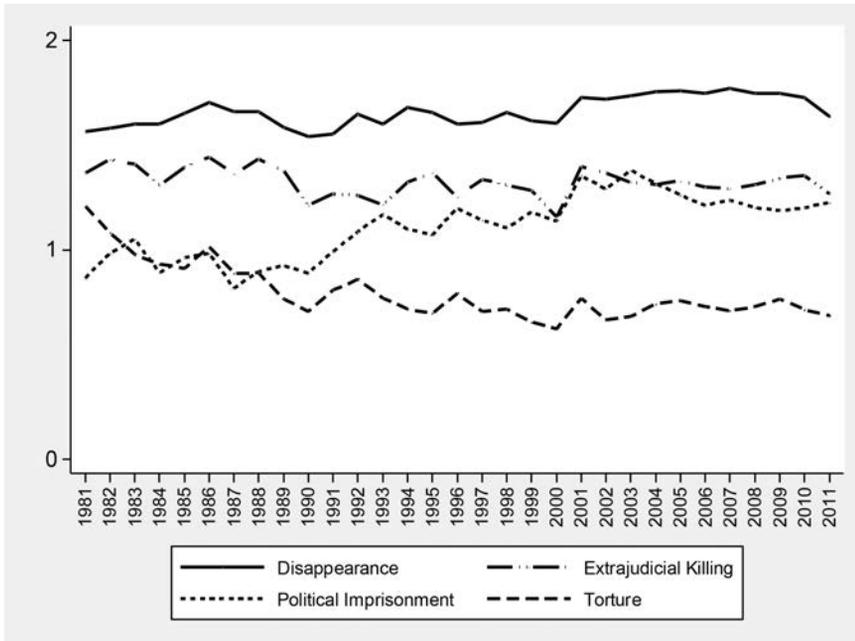


Figure 2. CIRI Physical Integrity Rights, Disaggregated Global Averages, 1981–2011

What about the possibility, however, that the stasis shown in Figure 1 is masking actual improvement in respect for these rights? Figure 2, which is a disaggregated view of Figure 1, shows this is the case; the culprit responsible for the flat trend in Figure 1 is not information effects. Instead, it is simply the effect of creating an additive index. As Figure 2 shows, there is much greater variation in average respect over time when looking at individual rights than when looking at the additive index.

What about the possibility, however, that information effects might be at play in these individual trends? Figure 2 shows only one of the four rights, torture, decreases in respect over time. Respect for the right against political imprisonment clearly improves over time. Respect against disappearance improves marginally, and respect against extrajudicial killings dips then recovers to its original level. There are no significant decreases in respect for these four rights over time in the CIRI data and, actually, there is noticeable improvement. This is counter to the expectations of the information effects argument.

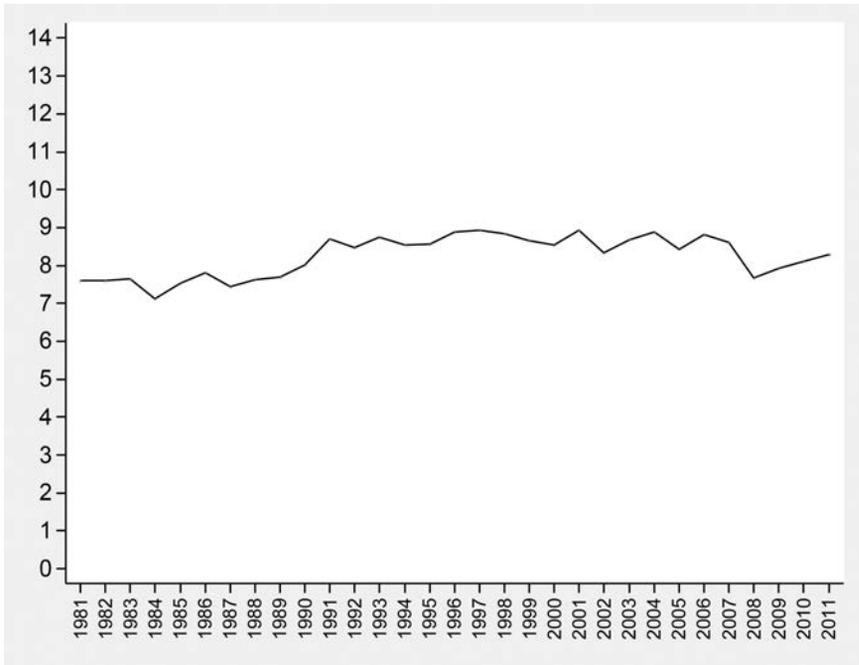


Figure 3. CIRI Empowerment Rights Index, Global Average, 1981–2011

B. Empowerment Rights

While the emphasis of Clark and Sikkink is on physical integrity rights, the framing of their arguments is “human rights measures.” This generalization of their argument from physical integrity rights to all human rights, therefore, requires taking a look at other human rights. The CIRI data do not cover all internationally-recognized human rights, but they do include what can be called “empowerment rights.” CIRI provided an additive index known as the “empowerment rights index” that aggregated individual measures of freedoms of domestic movement, foreign movement, religion, assembly and association, speech, workers’ rights, and electoral self-determination. This index ranged from 0 (no respect for any of these seven rights) to 14 (full respect for all seven of these rights).

Figure 3 illustrates the global average for the empowerment rights index from 1981–2011. As we saw with physical integrity rights in Figure 1, there is no pattern of decline in respect over time, and the overall trend is fairly flat. If forced to choose a direction of the trend, however, one would

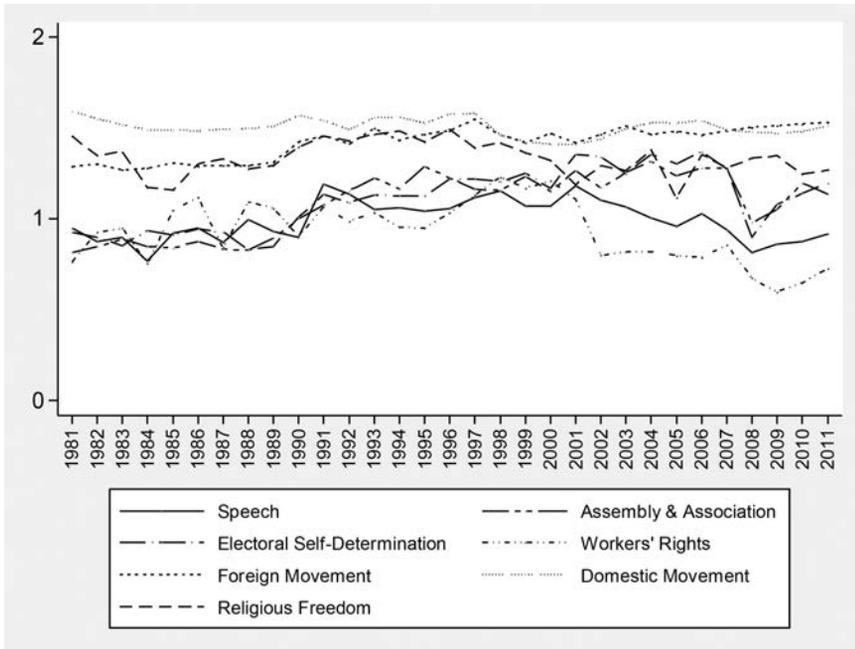


Figure 4. CIRI Empowerment Rights, Disaggregated Global Averages, 1981–2011

have to conclude it is ever-so-slightly upward over time. As with physical integrity rights, the Clark and Sikkink version of information effects does not appear supported.

Figure 4 is a disaggregated view of Figure 3.⁵ Here, we see considerable variation in changes in the average level of respect for rights over time, but we again do not see a general, significant decline in ratings over time. One of the most-noticeable features of Figure 4 is the post-2000 swoon in respect for the freedoms of speech and workers' rights. Where these rights end up in 2011, however, is close to their 1981 starting levels of respect. A look at summary statistics showed that while four of these seven indicators showed lesser averages of respect in 2011 than in 1981, there are some important qualifications to note before making too much of this. First, the differences ranged minimally, as follows: .03 (speech), .04 (worker's rights), .08 (domestic movement), and .19 (religion). Second, only one of these dif-

5. A color version of Figure 4 can be downloaded in PDF format at http://davidlrichards.com/images/fig_4_color.pdf.

ferences is statistically significant (freedom of religion, $t = 2.02$). Third, the number of states in the sample grew by 30% from 1981 to 2011. All these things considered, there seems to be no significant deflation of empowerment rights scores over time. And, as with physical integrity rights, the culprit in Figure 3's flat trend is simply the artifact of creating an additive index.

III. IS THERE REALLY MORE INFORMATION IN THE STATE DEPARTMENT REPORTS AS TIME GOES BY?

Clark & Sikkink assume there is more information in the State Department Country Reports on Human Rights Practices over time. They test this by looking at the number of words in each year's total report for Latin American Countries. It is my contention that examining total report lengths is an inappropriate method for this task. Rather, if one is examining torture scores, for example, one should be looking at the material in the section(s) of the report discussing torture, not the full country report. Likewise, if examining an aggregated physical integrity rights index like CIRI's, or the Political Terror Scale, one should look at the material in the sections of the report that contain information about these four physical integrity rights. The size of full country entries has grown, on average, over the years as a reflection of the incorporation of new human rights concerns such as sections on "Acts of Violence, Discrimination, and Other Abuses Based on Sexual Orientation and Gender Identity" and "HIV and AIDS Social Stigma." While it is true that the subsections of the State Department Reports are not hermetically sealed, the leakage of material about one type of human rights abuse into sections predominantly about other types of abuse is rather minor and, so far as the four physical integrity rights are concerned, mostly only relevant with regards to political imprisonment.

The analyses that follow use a dataset populated by a random draw of thirty countries for the years 1983, 1987, 1991, 1995, 1999, 2003, 2007, and 2011.⁶ For each of these country-years, the number of total words in their State Department report entries was counted, as were the number of words in the torture and extrajudicial killing subsections (separately) in these reports. Much more data would be necessary to authoritatively refute Clark and Sikkink's arguments, but thirty countries at eight points in time should give a good enough look to decide whether a closer, more-systematic examination of their reports with complete data is warranted. This is especially true since Clark and Sikkink base much of their critique of standards-based scores on the backs of only two case examples, Brazil and Guatemala.

6. To maximize generalizability and to deal with countries entering and leaving the CIRI data over the time series, a separate thirty-country drawing was done for each year.

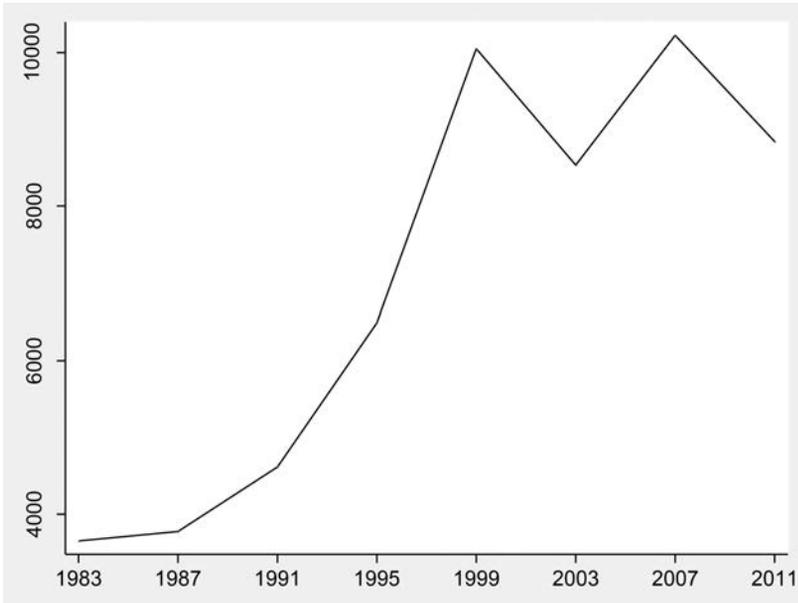


Figure 5. Median Total Word Count from Random 30-Country Draws, State Department Reports, 1983 – 2011

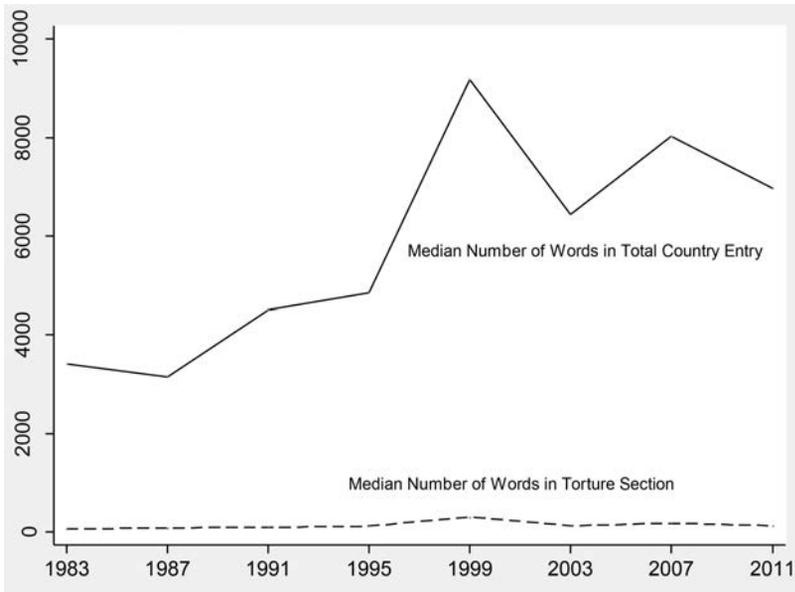


Figure 6. Median Total Country Entry Word Count and Torture Section Word Count, State Department Reports, 1983 – 2011

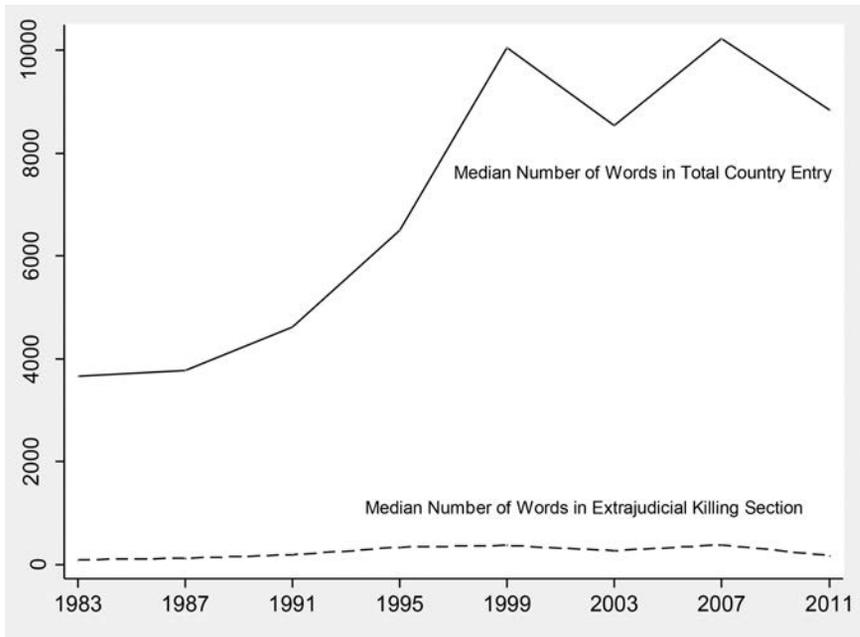


Figure 7. Median Total Country Entry Word Count & Extrajudicial Killing Section Word Count, State Department Reports, 1983 – 2011

Figure 5 shows the average total number of words in State Department Country Report entries from 1983 to 2011. Indeed, there is a huge jump in the average length of these reports after the end of the Cold War, although average report length oscillates greatly after 1999. However, what if we look at the word totals from just the section of the report used to create the physical integrity rights indicators? Figure 6 shows the median word totals for whole country report entries and the torture sections. Of course, complete country entries have more words than do subsections. However, the key here is that the median number of words is flat across the timeframe. There is no sharp jump in the number of words as time goes by.

The graph of extrajudicial killing word counts versus total counts shown in Figure 7 tells the same story.⁷

Lest one think the flat section-count trend lines in Figures 6 and 7 are the result of the Y axis' scale, Figures 8 and 9 take a closer look at these

7. The top lines in Figures 6 and 7 are slightly different because the random country draws for the torture and extrajudicial killing data were independent. Given that, their similarity is striking, however.

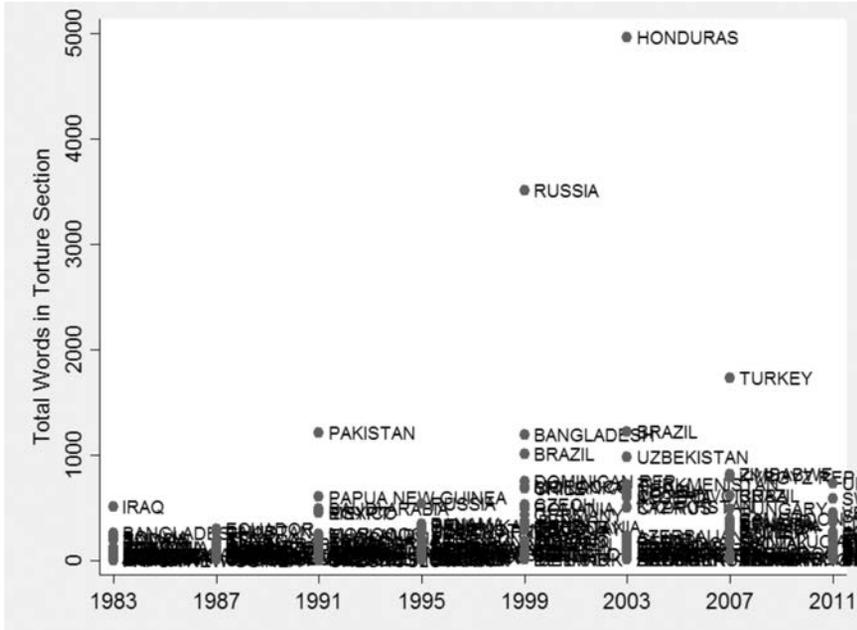


Figure 8. Torture Section Word Count Distribution by Country, State Department Reports, 1983 - 2011

counts over time. As Figure 8 shows below, there is very tight clustering among countries in terms of how many words the torture section of their report contains. Brazil, one of Clark and Sikkink’s case studies and a country about which they specifically discuss torture, is among the countries with the highest word counts. Also, we see that another Latin American country, Honduras, is the largest outlier here, by far.

Figure 9 shows there is also tight clustering among countries in terms of how many words the extrajudicial killing section of their report contains. Guatemala, one of Clark and Sikkink’s case studies, is the largest outlier in 1995 and 1999. Other Latin American countries are high in word count relative to their peers: Honduras, Mexico, Nicaragua, and Venezuela. The presence of so many Latin American countries among the top word counts and outliers raises a question of whether Clark and Sikkink’s findings might be biased towards confirmation because of their regional country sample.

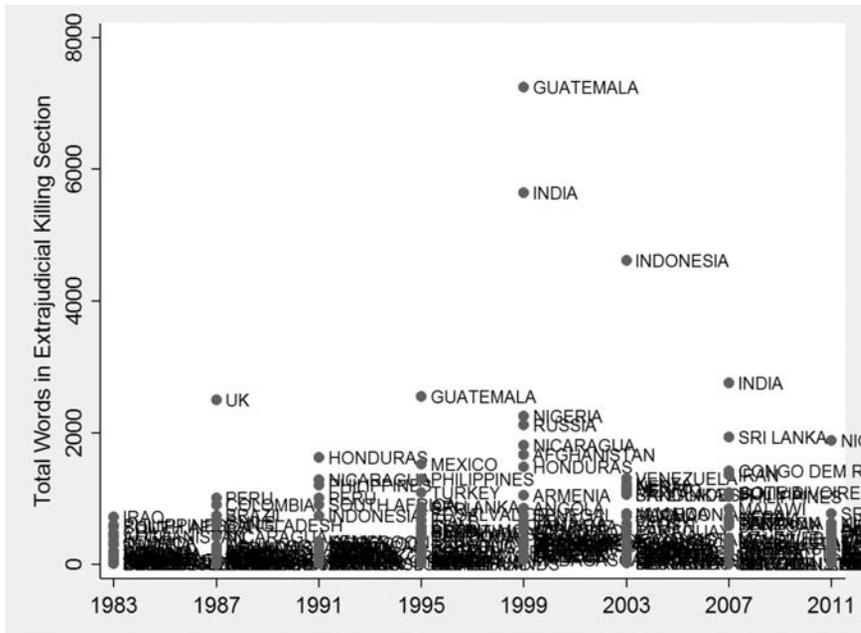


Figure 9. Extrajudicial Killing Section Word Count Distribution by Country, State Department Reports, 1983 – 2011

IV. EXAMINING THE “MORE WORDS = MORE INFORMATION = WORSE SCORES” PREMISE

Another premise underlying Clark and Sikkink’s argument is that more words in a report means more information in a report which leads to worse scores. Again, these scores are assumed to be downwardly biased because of information effects, not reflective of truly worse practices. While there is some intuitive appeal to the theory that more information should yield more-complete reporting of violations, this assumption that a greater number of words mechanically produces downwardly biased scores can also be challenged in several ways.

First, simple descriptive statistics certainly tell the story that countries with worse (lower) scores have higher median word counts in their sections than do countries with better scores. On its face, this would appear to at least partially support Clark and Sikkink’s overall worry about information effects. However, a few things must be taken into account. One such thing is the range of the number of words it takes to receive a particular score. For example, in these data, it could take anywhere from 22 to 7249 words

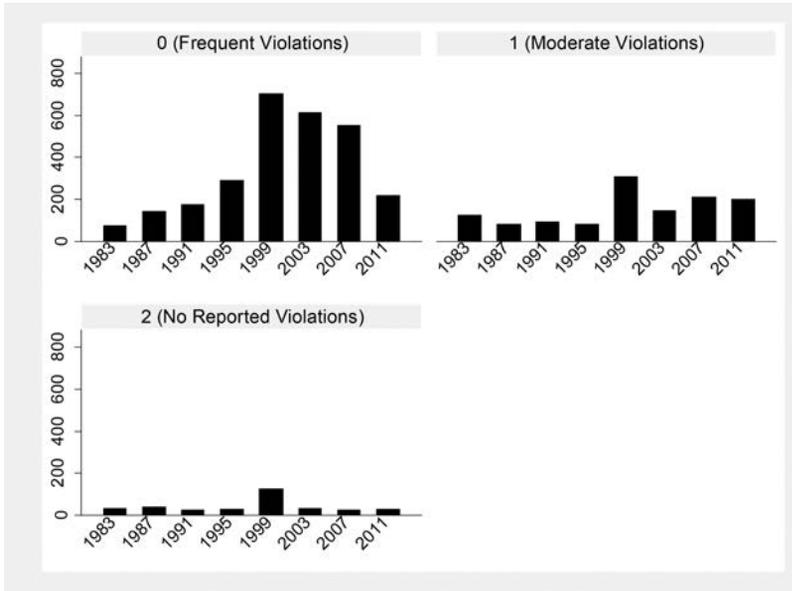


Figure 10. Median Word Count in Torture Section by CIRI Torture Score, By Year

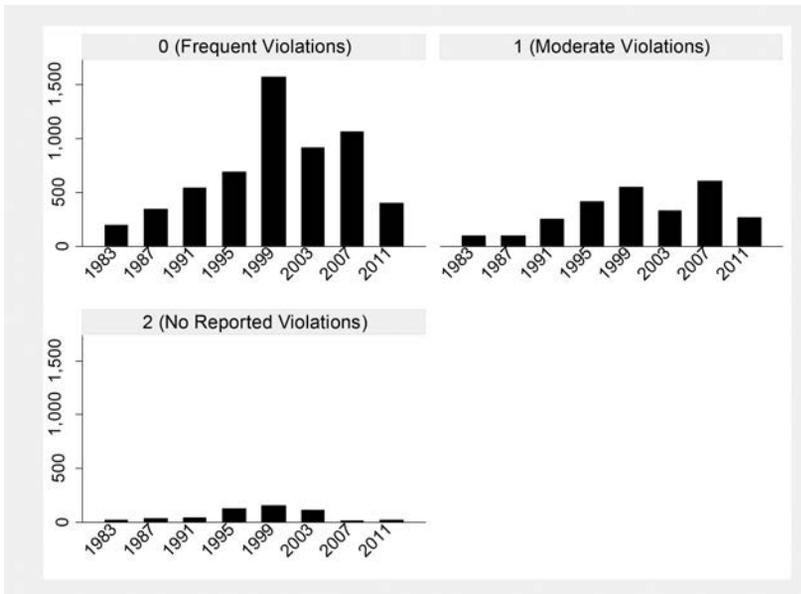


Figure 11. Median Word Count in Torture Section by CIRI Extrajudicial Killing Score, By Year

($md=586$) for a country to score a zero on extrajudicial killing and 7 to 1812 ($md = 271$) words to score a one.⁸ Likewise, it might take anywhere between 7 and 2507 ($md = 53$) words to score a perfect two. For torture, a country with a zero score might have a section anywhere between 17 and 3516 words ($md = 276$) whereas a country with a score of one could range from 13 to 4963 ($md = 126$) and those with a score of two could range anywhere from 7 to 534 words ($md = 35$).

Second, Clark and Sikkink's argument has a temporal element to it: that more words/information, over time, is leading to lower scores. It would indeed be troubling were greater word counts over time reliably associated with lower human rights ratings. However, the aforementioned wide ranges in word counts notwithstanding, it appears to have always been the case that more words are associated with lower scores. Figures 10 and 11 tell this story about torture and extrajudicial killing, respectively. These figures show the median word counts for these two indicators, over time, broken out by CIRI score received (0–2). It is clearly the case that throughout the time series, lower scores have longer substantive sections in the State Department reports. Further, the ratio in the number of words associated with scores of zero and two appears consistent over time.

Third, it does not appear to logically follow that more information necessarily creates scores that are biased downwards. It seems equally plausible that more information, if less biased than previous information, could produce lower scores that are more-precise estimates of human rights practices. This could happen, for example, by the State Department using three sources rather than one to describe what happened at a riot; clarifying that state agents were the first actors to use violence in a case where violence was used by both sides.

Finally, the qualities of standards-based data, generally, and how CIRI data were actually produced, in particular, do not fit with Clark and Sikkink's concerns. For example, figures two through six in their article seek to establish their argument by comparing standards-based scores from CIRI and PTS with numeric violations data from truth commissions. The underlying

8. A count of seven words for a CIRI score of one might appear low, but such cases typically indicate situations such as disagreement between the two sources used: State Department and Amnesty International. In these instances, the State Department would typically say "There were no reports of political killings." On the other hand, as in the case of Uruguay, Amnesty's Report might include language such as: "Demonstrations and violent disturbances broke out in April and November, when students and others protested against government policies at a time of increasing economic difficulties. On both occasions the protests were sparked off by the deaths of students, the first when a student was killed by a private citizen in March, and the second after another student died in police custody in November. The demonstrations resulted in violence and at least three people were shot dead by police." AMNESTY INTERNATIONAL 1988 ANNUAL REPORT 141 (1988).

ing logic of comparison is that the trend lines should assume similar shapes (albeit at different altitudes because of the different Y-scale axes). That is, it is assumed there should be a linear relationship between trend lines produced from events counts and those produced from standards-based data.⁹ First, this couldn't be expected to be the case, by definition, as ordinal data do not have equal intervals. While an events-count trend line will change with each change in count, the standards-based trend line won't change unless an ordinal score changes, which could be the result of any number of different magnitudes of change in the actual underlying phenomenon. Second, CIRI data were not produced using event counts as either the sole or dominant type of information on which scores were based. Below is a sample of language from the training packets used to train torture coders:

Coding will largely be done based on the language of the report. The language used in a report will always override a specific count of incidences, so make careful note of it. For example, if the report says violations were "widespread" or "systematic" (which would rate a score of "0") and then only mentions a few instances with a small number of cases (which would rate a score of "1"), the words "widespread" and "systematic" take precedence and the country scores a "0." A guideline follows.

- Instances where violations are described by adjectives such as "gross," "widespread," "systematic," "epidemic," "extensive," "wholesale," "routine," "regularly," or likewise, are to be coded as a ZERO (have occurred frequently).¹⁰

If a report explicitly stated that a country systematically or regularly tortured its opponents, for example, that country got a zero, no matter how short or long the illustrative materials included by the State Department or Amnesty International for context.¹¹ The total word count didn't matter to the score in these cases. What mattered was what words were used, and how.

Further, in coder training, CIRI explicitly trained coders to factor into their coding decisions whether a long list of examples of a type of physical

9. It is worth noting truth commission reports, as sources of authoritative documentation of human rights violations, may have their own issues of error. Alison Brysk, *The Politics of Measurement: The Contested Count of the Disappeared in Argentina*, 16 *HUM. RTS. Q.* 676, 685–86 (1994). Brysk documents in detail the politics surrounding the Argentine National Commission on the Disappeared's estimates of human rights violations by the military juntas of the 1970s and early 1980s. She shows, for example, that civil society groups and the truth commission were at odds about estimated violations, that the truth commission received "no responses to hundreds of requests of information from the military and police" and that accounting methods such as exhumations can be confounded by methods of victim disposal such as dumping at sea. *Id.* at 682.

10. DAVID L. CINGRANELLI, DAVID L. RICHARDS, & K. CHAD CLAY, *THE CINGRANELLI-RICHARDS (CIRI) HUMAN RIGHTS DATA PROJECT CODING MANUAL, VERSION 5.20.14*, at 17 (2014), available at <https://drive.google.com/file/d/0BxDpF6GQ-6fbWkpxTDZCQ01jYnc/edit>.

11. David L. Cingranelli & David L. Richards, *The Cingranelli and Richards (CIRI) Human Rights Data Project*, 32 *HUM. RTS. Q.* 401 (2010).

integrity rights violation was merely a long illustrative list, or representative of the entirety of a situation. For example, a country might have widespread torture with a dozen case examples listed and another country might have widespread torture with only a few case examples listed. They would both get a rating of zero, but the former's section would be much longer than that of the latter.

V. CHANGING STANDARDS

Human rights standards change over time. This happens for a variety of reasons. Most importantly, human rights law—which operationalizes what respect for human rights actually means—evolves via court cases, new international legal instruments, new municipal statutes, and by general comments by treaty committees. So, it is indeed the case that over time, information in reports such as that of the State Department could evolve with these changing legal understandings. I say “could” instead of “would” because the history of the State Department reports is a history of the President exerting influence over what content gets included in these reports. For example, the Reagan administration famously recast “economic and social rights” as “economic and social circumstances” in order to lessen their importance relative to the democratic components of political and civil rights the administration wished to promote as American ideals.¹²

Changing standards could indeed affect scores in a time series if, for example, states are all of a sudden held responsible for things in 2007 that they were not back in 1990. But does this possibility necessarily mean that CIRI scores were affected? I'm inclined to say any such effect, if at all extant, was minimal due to the way CIRI operated. Indeed, CIRI did change standards for certain variables over time as understandings of these human rights evolved in practice. However, when CIRI changed standards, the entire time series was recoded to include these new practices where reported. A good example is military hazing. At some point in the mid-2000s, CIRI started counting military hazing as torture. We then went back and checked all the previous torture data, recoding cases where such hazing was reported in the past. This was done to bring consistency to the time series. Admittedly, this was not perfect, because it is the case that before a common understanding of new standards emerged, the source materials were less likely to report related events. Such a threat to validity is most present with the State Department reports, as Amnesty reports have focused on a smaller set

12. US DEPARTMENT OF STATE, COUNTRY REPORTS ON HUMAN RIGHTS PRACTICES FOR 1981, at 6 (1982); David L. Richards, “*Rhetoric and Reality*” Revisited, 11 J. HUM. RTS. 337 (2012).

of well-addressed issues (e.g., political imprisonment) for which Amnesty has held fairly consistent standards over the years.

VI. CONCLUSION

Clark and Sikkink do everyone a favor by getting data users to think about potential threats to validity. However, it appears there is some reason to doubt some claims they make about information effects, at least in the CIRI data. First, it was found that scores are not necessarily declining over time, indeed many are improving. Moreover, the other information effects possibility—flat trends hiding improvement—is unsupported as the flat trends of the physical integrity rights and empowerment rights indices appear to be artifacts of aggregation. Second, it was shown that median word counts in the actual sections of the State Department reports used to create CIRI data have not changed much over time, as opposed to the total-country-entry counts used by Clark and Sikkink. Third, it was shown that the relationship between word counts from substantive State Department report sections and countries' human rights scores has remained steady over time. It is not new that countries with greater word counts in these sections tend to get lower scores; indeed, it has always been the case. Finally, while the issue of evolving human rights standards is definitely a threat to inferences drawn from time-series human rights ratings, it has not gone unheeded by metrologists; CIRI had practices in place to attempt to deal with this issue.

There are a few lessons from all of this for users of quantitative information about human rights, it appears. First, use disaggregated measures of human rights practices instead of aggregated indices such as those provided by CIRI and the Political Terror Scale. Any threat to the substantive inferences made from time-series analyses using aggregate indices of government respect for physical integrity and empowerment rights likely comes from the aggregation process used to create these indices, not from information effects. Second, when parsing raw sources of human rights information for quality, it is likely *which* words are being used, and *how*, rather than how many words are being used that guide what effect words have on human rights ratings. Finally, while no data are without some problems, there appear to be no information effects downwardly biasing CIRI physical integrity rights scores over time.