CODING PERSONAL INTEGRITY RIGHTS: ASSESSING STANDARDS-BASED MEASURES AGAINST HUMAN RIGHTS LAW AND PRACTICE

MARGARET L. SATTERTHWAITE*

Clashing theories about the impact of broad social processes on human rights have long demanded indicators that allow for comparison of human rights conditions across countries and over time. Political science and international relations scholars have developed cross-national, time-series data sets to interrogate the relationships between human rights violations and things like civil unrest (do violations cause war?), terrorism (does government repression spur radicalization?), and economic development (do rising GDPs guarantee improved rights?) using sophisticated quantitative methods. In recent years, these debates have broadened beyond the academy. In development circles, there is concern that recent achievements in areas like health and education may have come at the expense of human rights—especially democratic governance and accountability, and conversely that deficits in human rights have led to failures in human development. At the same time as these changes in the development world have taken place, scholars of human rights have reached for metrics to answer some of the most difficult questions facing those who care about rights: does international human rights law—especially as codified in international treaties—actually influence state behavior over time? In other words, do human rights treaties and covenants work? This type of analysis requires rights metrics that can be used for traditional quantitative tests such as correlation and regression analysis, and which are amenable to integration into econometric models alongside indicators gathered from the development and policy analysis worlds. These synergistic discussions have led to a renewed focus in the human rights world on standards-based measures for human rights.

* Professor of Clinical Law, N.Y.U. School of Law. The author is grateful for feedback on earlier versions of this paper from David L. Richards, Mark Gibney, Deena Hurwitz, Mila Versteeg, participants at the “Indicators and the Ecology of Governance” conference (N.Y.U. School of Law, July 6–7, 2015) and the “Comparative Observation with Numbers: A Critical Assessment of Standards-Based Measures and Human Rights” workshop (University of Bielefeld, Germany, October 26–27, 2012). The author also thanks Tiffany Lin (N.Y.U. J.D. expected ‘16), Saif Ansari (N.Y.U. J.D. expected ‘16), Jessica Boulet (N.Y.U. J.D. expected ‘16), and Amy Zajac (N.Y.U. J.D. expected ‘16) for excellent editorial work, and Katherine Erickson (N.Y.U. J.D. ‘15) and Cristina Campo (U. Pa. B.A. expected ‘16) for research assistance. She gratefully acknowledges support from the Filomen D’Agostino Research Fund at New York University School of Law.
This Article examines several leading standards-based measures of human rights from the perspective of human rights law and practice. It is part of an emerging discussion in the literature assessing human rights data sets used in quantitative studies. Issues examined in this Article include: the fit between the concepts being coded into the data sets and the legal content of the relevant rights; the choice of rights to include and exclude from the coding process; the sources used by each data set; and the methods used by the organizations responsible for authoring those sources that may have unintended impacts on the coding process and potentially the conclusions reached by analysts using standards-based measures. The Article concludes that the drive to standardize, the push to quantify, and the desire to compare across time and space have produced data sets that are limited in specific respects, leading to possible distortions in the evidence base of studies relying on time-series data sets for human rights analysis. On the other hand, the measurement projects could provide important additional information for those seeking to integrate rights into development policy and practice, and they could be improved if they were brought closer into line with international human rights law.

I. INTRODUCTION .................................. 515
II. OVERVIEW OF POPULAR STANDARDS-BASED HUMAN RIGHTS MEASURES .................. 520
   A. Freedom House Civil and Political Liberties Scales .................................. 521
   B. Political Terror Scale ................................ 524
   C. Cingranelli and Richards Human Rights Data Project .................................. 526
III. THE LIMITATIONS OF STANDARDS-BASED MEASURES AS METRICS: NATIONAL AGGREGATION, VARIANCE TRUNCATION, AND LIMITED SOURCES ............ 529
   A. Limitations Related to Methodology: National Aggregation, Variance Truncation, and Limited Sources .................................. 529
   B. Limitations Related to Sources: The Development of Modern Human Rights Reports ........ 531
      1. Development of the Department of State Country Reports ............ 531
   C. The New Genre of the Human Rights Report ................................ 536
IV. LIFTING THE NUMERICAL VEIL: THE EXTENT TO WHICH STANDARDS-BASED MEASURES REFLECT INTERNATIONAL HUMAN RIGHTS LAW .............. 537
   A. Scope .................................. 540
      1. Scope of PTS and CIRI Coding .................. 540
2. **Scope of Department of State and Amnesty International Reports** .................................. 545

B. **Intensity** ................................................................. 551
   1. Intensity in PTS and CIRI Coding ............................. 551
   2. Intensity in Department of State and Amnesty International Reports .................................. 554

C. **Range** ................................................................. 559
   1. Range in PTS and CIRI Coding ................................. 559
   2. Range in Department of State and Amnesty International Reports .................................. 562

D. **Systematicity** .......................................................... 565
   1. Systematicity in PTS and CIRI Coding ...................... 565
   2. Systematicity in Department of State and Amnesty International Reports ...................... 569

E. **Interdependence** ..................................................... 571
   1. Interdependence in PTS and CIRI Coding ................ 571
   2. Interdependence in Department of State and Amnesty International Reports ................ 577

V. **Conclusion** .......................................................... 579

I. **Introduction**

Clashing theories about the impact of broad social processes on human rights have long demanded indicators that allow for comparison of human rights conditions across countries and over time. Political science and international relations scholars have developed cross-national, time-series data sets to interrogate the relationships between human rights violations and things like civil unrest (do violations cause war?), terrorism (does government repression spur radicalization?), and economic development (do rising GDPs guarantee improved rights?) using sophisticated quantitative methods. Until recently, such techniques have held little interest for human rights organizations, which have been dedicated to gathering action-oriented, qualitative, descriptive data that is close to the ground and almost always rooted in the testimony of victims and witnesses.¹ Questions about the systemic correlates of rights violations have seemed largely outside the focus of human rights organizations, which aim to document, de-

---

¹ For a discussion of human rights organizations’ working methods, see THE TRANSFORMATION OF HUMAN RIGHTS FACT-FINDING (Philip Alston & Sarah Knuckey eds., 2015).
nounce, and change policies in real time, largely through "naming and shaming."

Scholars in the fields of economics, political science, and international relations, on the other hand, have made wide use of these measures. One study found that ninety articles were published between 1999 and 2011 that used the leading standards-based measures data sets.

Standards-based measures of human rights have become a dominant feature of social scientific analysis, particularly in the fields of political science and international relations. Since the first cross-national statistical analysis of human rights in the 1980s (Mitchell and McCormick 1988), there has been a proliferation of studies using increasingly large and complex data sets and an expanding list of independent variables in which standards-based measures of human rights feature across most of the studies as a key dependent variable . . . . The comparable, standard and time-series nature of these measures has made them particularly popular in social science research projects that seek to make empirical generalizations across a large sample of countries and time.

In recent years, these debates have broadened beyond the academy. In development circles, there is concern that recent achievements in areas like health and education may have come at the expense of human rights—especially democratic governance and accountability, and conversely that deficits in human rights have led to failures in human development. As the international community conceptualized the Sustainable Development Goals, these concerns revealed a need for met-

---


4. Landman & Carvalho, supra note 2, at 86-88.

rics to measure human rights related to governance and democracy in a consistent, comparable manner. In the human rights advocacy community, it has become clear that strategies to integrate rights norms directly into development frameworks may be as important as monitoring the rights impacts of development processes from the outside through naming and shaming.

At the same time as these changes in the development world have taken place, scholars of human rights have reached for metrics to answer some of the most difficult questions facing those who care about rights: does international human rights law—especially as codified in international treaties—actually influence state behavior over time? In other words, do human rights treaties and covenants work? This question has come to the fore as the first major phase of the modern human rights movement, which focused on the codification and ratification of human rights law in universal treaties, draws to a close. In addition, scholars are increasingly developing sophisticated models to look at the impact of particular events on human rights conditions, effectively asking the same kinds


8. See Beth Simmons, Mobilizing for Human Rights: International Law in Domestic Politics 20–21 (2009) (finding, through an exhaustive quantitative empirical study, that "human rights treaties have positive effects" in that they "contribute to a political and social milieu in which these rights are more likely, on the whole, to be respected"); see also Oona A. Hathaway, Do Human Rights Treaties Make a Difference?, 111 Yale L.J. 1935, 2025 (2002) (finding that "ratification of the treaties by individual countries appears more likely to offset pressure for change in human rights practices than to augment it"). But see Ryan Goodman & Derek Jinks, Measuring the Effects of Human Rights Treaties, 14 Eur. J. Int’l L. 171, 171 (2003) (identifying "serious deficiencies in [Hathaway’s] empirical findings, theoretical model and policy prescriptions").
of questions that human rights advocates try to answer, but using different tools.  

Efforts to improve the evidence base of human rights advocacy, stemming in part from the influence of evidence-based development and policy-making discourses, have also made plain the need for data that enables analysts to test the impact of policy interventions that are the subject of recommendations by advocates—including, but not limited to recommendations to ratify treaties and pass laws.

This type of analysis requires metrics that can be used for traditional quantitative tests such as correlation and regression analysis. The metrics must also be amenable to integration into econometric models alongside indicators gathered from the development and policy analysis worlds. These synergistic discussions have led to a renewed focus in the human rights world on standards-based measures for human rights.

This Article examines several leading standards-based measures of human rights frequently used to study the interaction of human rights and large social, economic, and political phenomena like development, conflict, and terrorism. These measures are the same ones used by scholars examining the impact of human rights law. Standards-based measures such as the Freedom House Political Rights and Civil Liberties Scales, the Political Terror Scale, and the Cingranelli and Richards human rights data project rely on a process that turns country-specific information into indicators in relation to human rights standards. This indicator conversion process transforms heterogeneous qualitative data about rights violations into a homogeneous quantitative format, rendering it amenable to sophisticated statistical analysis.

This Article examines two of the most prominent measurement projects from the perspective of human rights law and practice. It is part of an emerging discussion in the literature assessing human rights metrics used in quantitative stud-

---

9. For example, Benedikt Goderis and Mila Versteeg ask whether 9/11 has had a significant impact on rights violations in the years following the attacks. Benedikt Goderis & Mila Versteeg, Human Rights Violations after 9/11 and the Role of Constitutional Constraints, 41 J. LEGAL STUD. 131 (2012).

10. LANDMAN & CARVALHO, supra note 2, at 68.
ies. Issues examined in this Article include: the fit between the concepts being coded through the scales and the legal content of the relevant rights; the choice of rights to include and exclude from the coding process; the sources used by each measurement project; and the methods used by the organizations responsible for authoring those sources that may have unintended impacts on the coding process and potentially the conclusions reached by analysts using standards-based measures.

As assessment of human development becomes more data-based, policy-makers insist on quantifying the impact of interventions, and the utility of human rights mechanisms is questioned, an assessment of the metrics that may be used in these contexts is called for. The Article concludes that the drive to standardize, the push to quantify, and the desire to compare across time and space have produced measurement projects that are limited in specific respects, leading to possible distortions in the evidence base of studies relying on the resulting time-series data sets for human rights analysis. On the other hand, the measurement projects could provide important additional information for those seeking to integrate rights into development policy and practice, and the metrics could be made more valid and reliable if they were brought closer into line with international human rights law.

Three prominent standards-based measures focus on measuring practices directly related to international human rights norms: the Freedom House Civil Rights and Political Liberties Scales, the Political Terror Scale (PTS), and the Cingranelli and Richards Human Rights Data Project (CIRI). Part II of this Article will provide a brief overview of each measure, explaining the qualitative sources they use and their approach to coding. Part III examines some commonly acknowledged

limitations to standards-based measures, focusing on the PTS and CIRI measures. These two measures are chosen because of their prominence, and because their commitment to methodological transparency makes them more attractive than the Freedom House measures. Limitations related to methodology are discussed, including: limited source material, relative lack of precision, and national aggregation. Limitations related to the development of the source materials used are also discussed: the U.S. Department of State (DOS)'s annual Country Reports on Human Rights Practices and Amnesty International (AI)'s Annual Report. Part IV of the Article focuses on assessing the PTS and CIRI Physical Integrity Rights Index from the perspective of international human rights law and practice. By comparing the coding conventions behind the CIRI and PTS measurement projects to human rights norms, the Article considers the validity of these measures as metrics for human rights. The Article thus allows readers to assess the fit between what the metrics capture, what is set out in international law, and what the source materials can be understood to encompass. Part V concludes by noting that measures like these might be improved through a stronger grounding in human rights law and practice, and that human rights organizations might learn lessons from the measurement projects that could improve their work.

II. OVERVIEW OF POPULAR STANDARDS-BASED HUMAN RIGHTS MEASURES

Standards-based measures are created by experts who score units of information in relation to a specific standard—here, human rights. As the UN Office of the High Commissioner for Human Rights explains, these measures are a form of data based on expert coding:

12. In this way, this Article is in conversation with concerns raised by other scholars of human rights, including Ryan Goodman. See, e.g., Ryan Goodman, The Difference Law Makes: Research Design, Institutional Design, and Human Rights, 98 AM. SOC'Y INT'L. L. PROC. 198, 199 (2004) (expressing concern that indicators of human rights violations are often not calibrated to capture violations that fit the legal definitions of particular violations).
13. Many thanks to David L. Richards for helpful clarification on this point.
[Standards-based coding] usually involves a more systematic use of diverse sources of information, including the media, government reports and reports from NGOs, by a limited number of experts (e.g., advocacy groups, academic researchers, social scientists, managers) who are asked to evaluate and score the performance of States. . . Data based on experts' judgements have been frequently used for cross-country ranking and comparisons over time.¹⁴

The use of such ratings to compare countries is in stark contrast to the traditional reluctance—and often, outright refusal—by human rights organizations to rank or even compare countries, despite their firm commitment to a universal standard of rights.¹⁵ For this reason, it is not surprising that a content analysis of all field research reports published by AI and Human Rights Watch (HRW) in 2000 and 2010 revealed that not a single report referred to the leading standards-based measures of human rights.¹⁶

A. Freedom House Civil and Political Liberties Scales

Freedom House is a U.S.-based NGO that was founded in 1941 by a bipartisan group of prominent Americans.¹⁷ It was originally established to call Americans' attention to the threats to freedom presented by Hitler's ascendancy.¹⁸ In 1955, the organization began publishing an annual Balance


Sheet of Freedom, which presented the “credits” and “debits” of freedom in the world. The Balance Sheet was replaced with the Comparative Survey of Freedom in the World, which was published between 1972 and 1978, and applied a standards-based rating system developed by scholar Raymond Gastil in 1972. The report expanded into an annual book-form publication in 1978; since then, Freedom House has published Freedom in the World. Gastil remained at the helm of the project until 1989 when Freedom House expanded to include a professional staff of analysts. Until this time, the methodology for the production of Freedom in the World was opaque; Gastil himself admitted in an article published in 1990 that he used a “loose, intuitive rating system.” This team was supplemented in the mid-1990s with external analysts based in the academy, media outlets, and human rights organizations.

Freedom House measures “global freedom as experienced by individuals” through two scales: the Political Rights Scale and the Civil Liberties Scale. Ratings for countries on each scale are derived from a total score drawn from a checklist comprising ten political rights questions and fifteen civil liberties questions:

Each country is assigned a numerical rating from 1 to 7 for both political rights and civil liberties, with 1 representing the most free and 7 the least free. The ratings are determined by the total number of points (up to 100) each country receives on 10 political rights questions and 15 civil liberties questions; countries receive 0 to 4 points on each question, with 0 representing the smallest degree and 4 the greatest degree of freedom. The average of the political rights

19. Id. at 33–34.
20. Id. at 33–35. See generally Our History, FREEDOM HOUSE, supra note 17.
22. Id.
25. Id.
and civil liberties ratings, known as the freedom rating, determines the overall status: Free (1.0 to 2.5), Partly Free (3.0 to 5.0), or Not Free (5.5 to 7.0). Freedom House also assigns upward or downward trend arrows to countries which saw general positive or negative trends during the year that were not significant enough to result in a ratings change.\(^{26}\)

The ratings are based on a description of "general characteristics" associated with the different ratings.\(^{27}\) Freedom House staff consult with regional experts and scholars individually and in regional meetings to compare actual country conditions to these general characteristics.\(^{28}\) When formulating rankings, the staff relies upon sources of information such as media accounts, scholarly materials, reports published by NGOs and think tanks, and regional visits.\(^{29}\) Although the questions included in the checklists that are used to guide the scoring process are made public, the specific steps and decisions involved in transforming information about each country from these various sources into particular scores are not reported.\(^{30}\)

Over the years, much of Freedom House's funding has come from grants by the U.S. government.\(^{31}\) In part because of the close (funding) relationship between Freedom House and the U.S. government, its scales have sometimes been perceived to be biased,\(^{32}\) although these critiques have been called "anecdotal" by other scholars.\(^{33}\) An extensive historical account of *Freedom in the World* concludes that while the record

\(^{26}\) Id.
\(^{27}\) Id.
\(^{28}\) Id.
\(^{29}\) Id.
\(^{32}\) See, for example, Edward S. Herman & Noam Chomsky, *Manufacturing Consent* (1994). For a broader discussion of allegations of bias, see Bradley, supra note 18, at 43-46 (outlining the history of Freedom House's dependence on U.S. government funding).
does not support allegations of intentional bias, there does appear to have been "a self-reinforcing process: ideological convergence" between U.S. government positions on the connection between rights protection and regime types and Freedom House ratings "but without any conspiracy." 34

Apart from concerns about bias and intellectual capture, Freedom House's relative methodological opacity 35 and reliance on expert opinion has led to a reluctance by some international actors to use its scales. 36 This Article will therefore focus the remaining discussion on the PTS and CIRI Physical Integrity Rights Index, scales that have made their sources and coding schemes more transparent and thus hold more promise for human rights advocates and policy-makers. 37

B. Political Terror Scale

The Political Terror Scale (PTS) is another well-established measurement project that assesses the human rights situation on a yearly basis for a large number of countries. Focusing on "physical integrity rights," the PTS extends back to 1976 and includes data on more than 180 countries. 38 PTS reports that it is the "most commonly used indicator of state violations of citizens' physical integrity rights." 39 The coding scheme uses a scale developed by the same scholar as Freedom House's scales—Raymond Gastil—who created the scale used by PTS

34. Bradley, supra note 18, at 45.
35. Landman & Carvalho, supra note 2, at 88–89; see also Jorgen Moller & Svend-Erik Skaninng, Evaluating Extant Rule of Law Measures, in THE RULE OF LAW: DEFINITIONS, MEASURES, PATTERNS AND CAUSES 41, 54 (2014) (noting that Freedom House "provides little more than checklists defining the parameters of the indicators and virtually no guidance about how to interpret the numerical scores"); David A. Armstrong II, Stability and Change in the Freedom House Political Rights and Civil Liberties Measures, 48 J. PEACE RES. 653, 661–62 (2011) (calling on Freedom House to release data generated in the computation of its scales to allow researchers to investigate "potentially serious problems").
36. See, e.g., UNDP, supra note 6, at 2 (rejecting "existing empirical indicators of democracy and democratic governance," including Freedom House, as "flawed").
37. For an exhaustive and insightful discussion of the history and context of Freedom in the World, see generally Bradley, supra note 15.
39. Id. at 368.
in 1979. This scale codes information about extrajudicial executions, torture, political imprisonment, and exile. Dividing human rights violations into five levels, the scheme attends to three different “dimensions” of violence: “scope, intensity, and range,” where the levels are understood to describe a “continuum of human rights practices,” with each country receiving scores for each year based on its place on that continuum. Higher scores indicate worse practices.

Box A: Political Terror Scale

Level 5: Terror has expanded to the whole population. The leaders of these societies place no limits on the means or thoroughness with which they pursue personal or ideological goals.

Level 4: Civil and political rights violations have expanded to large numbers of the population. Murders, disappearances, and torture are a common part of life. In spite of its generality, on this level terror affects those who interest themselves in politics or ideas.

Level 3: There is extensive political imprisonment, or a recent history of such imprisonment. Execution or other political murders and brutality may be common. Unlimited detention, with or without a trial, for political views is accepted.

Level 2: There is a limited amount of imprisonment for non-violent political activity. However, few persons are affected, torture and beatings are exceptional. Political murder is rare.

Level 1: Countries under a secure rule of law, people are not imprisoned for their views, and torture is rare or exceptional. Political murders are extremely rare.

Source: http://www.politicalterrorscale.org/about.php

Each country is given two scores for each year: one based on the information provided in AI’s Annual Report, and the other based on the information provided in the U.S. DOS’s annual

40. Landman & Carvalho, supra note 2, at 73.
41. Id. at 75.
42. Wood & Gibney, supra note 38, at 373–74.
Country Reports on Human Rights Practices, the scores are reported in two separate indexes. Senior coders Mark Gibney and Reed Wood, assisted by student researchers, assign scores. PTS has not published a coding manual, though it does provide coding examples on its website. The coding process is self-consciously "subjective," with the coding process embracing "contextual factors found in the reports [that] effectively prohibit purely objective coding criteria." Further, the scores are not meant to reflect at the closest possible approximation of the number of violations in a year; instead, the scores reflect "patterns of abuse" that occur across countries. Despite the inclusion of subjective, contextual factors in the coding process, the PTS reports very high levels of inter-coder reliability:

After each person codes (separately), the scores are then compared. In approximately 80% of the cases, the scores that each coder comes up with are exactly the same. However, when there are differences, there is invariably an informal discussion between several coders to determine how a particular score was assigned. This clears up any discrepancies between the scores.

C. Cingranelli and Richards Human Rights Data Project

The Cingranelli and Richards Human Rights Data Project (CIRI) takes a more disaggregated approach to coding countries' human rights practices. Scholars David Cingranelli and David Richards have developed fifteen different measures.

43. About the Political Terror Scale, http://www.politicalterrorscale.org/About/FAQ/. Amnesty International did not publish an Annual Report for a few isolated years. PTS coded Human Rights Watch's World Reports for those years. See id.
44. Wood & Gibney, supra note 38, at 372.
46. Wood & Gibney, supra note 38, at 374.
47. Id.
(though one was discontinued) designed to assess government practices concerning specific earlier human rights or clusters of related rights (see Box B). The approximately 202 countries in the dataset are given separate scores for each year from 1981-2011 for each of the measures. Some of the rights are measured on a 0-2 scale, and others are measured on a 0-3 scale; in both schemes, higher numbers suggest better human rights practices. In addition to these separate measures, CIRI also offers two indexes: a Physical Integrity Rights Index (PIR Index), and an Empowerment Rights Index. The former will be the focus of analysis in this Article.

<table>
<thead>
<tr>
<th>Box B: Variables in the CIRI Human Rights Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Political and Other Extrajudicial Killings/Arbitrary or Unlawful Deprivation of Life**</td>
</tr>
<tr>
<td>Disappearance**</td>
</tr>
<tr>
<td>Torture and Other Cruel, Inhumane or Degrading Treatment or Punishment**</td>
</tr>
<tr>
<td>Political Imprisonment**</td>
</tr>
<tr>
<td>Freedom of Speech and Press</td>
</tr>
<tr>
<td>Freedom of Religion</td>
</tr>
<tr>
<td>Freedom of Domestic Movement</td>
</tr>
<tr>
<td>Freedom of Foreign Movement and Travel</td>
</tr>
<tr>
<td>Freedom of Assembly and Association</td>
</tr>
<tr>
<td>Electoral Self-Determination</td>
</tr>
<tr>
<td>Worker Rights</td>
</tr>
<tr>
<td>Women’s Political Rights</td>
</tr>
<tr>
<td>Women’s Economic Rights</td>
</tr>
<tr>
<td>Women’s Social Rights (Discontinued in 2005/2007)</td>
</tr>
<tr>
<td>Independent Judiciary</td>
</tr>
</tbody>
</table>

**These rights are included in the CIRI PIR Index.


50. CIRI Human Rights Data Project: FAQ, http://www.humanrightsdata.com/p/faq.html (last visited Jan. 28, 2016) (noting that several dozen states were added later than others, so the data for these countries is available for a shorter period of time than the others).

51. CIRI Coding Manual, supra note 49.
CIRI codes information from the U.S. DOS's annual *Country Reports on Human Rights Practices* for all of its measures, and also uses AI's *Annual Report* as a source for the rights included in the PIR Index. David Cingranelli and David Richards explain this choice of sources as follows:

Measuring government respect for human rights for nearly every country in the world requires systematic qualitative information—meaning standardized information about the same rights for each country, annually. The annual *US State Department Country Reports on Human Rights Practices* (*Country Reports*) is the only such existing source, and our coders use it to code all variables. When coding physical integrity rights, coders also use Amnesty International’s *Annual Report*. When there is a difference between the two sources, our coders treat the Amnesty International assessment as authoritative.

AI’s reports are used only for physical integrity rights since the organization has not historically and systematically reported on other types of violations; indeed, the organization’s mandate was largely limited to such violations until the early 2000s (see discussion below).

Scores for each variable are determined by trained coders under the supervision of David Cingranelli and David Richards. At least two people code each data point, following instructions set out in a 100-page manual. The project has reported an impressive inter-coder reliability statistic.

---

53. Cingranelli & Richards, supra note 52, at 400.
54. *Id.*
55. *Id.* at 403. *See generally* CIRI Coding Manual, supra note 49.
56. Cingranelli & Richards, supra note 52, at 403–04.
III. THE LIMITATIONS OF STANDARDS-BASED MEASURES AS METRICS: NATIONAL AGGREGATION, VARIANCE TRUNCATION, AND LIMITED SOURCES

A. Limitations Related to Methodology: National Aggregation, Variance Truncation, and Limited Sources

As discussed previously, standards-based measures of human rights have been widely used by political science and international relations scholars, and are being used more and more by human rights scholars.57 The measures are seen as especially useful for comparing countries over time, identifying major differences among countries, as well as offering the capacity to examine the causes, correlates, determinants, and consequences of government violation or respect for core human rights.58 A literature has developed comparing the main standards-based human rights measures in this context. This discussion has appeared in the political science, international relations, and interdisciplinary human rights journals and attends to the issues of greatest concern to social science scholars.59 Among the more technical concerns are those well


59. See generally, e.g., David Cingranelli & David L. Richards, Measuring the Level, Pattern, and Sequence of Government Respect for Physical Integrity Rights, 43
summarized by Landman and Carvalho as concerns about national aggregation and variance truncation.\textsuperscript{60}

Concerns about national aggregation arise because CIRI and PTS both code rights on a national basis only and are thus insensitive to differences within countries, which are often especially significant for human rights advocates.\textsuperscript{61} For example, specific ethnic or religious groups may suffer human rights violations at a higher rate than the general population, but when the national data is examined, the abuses appear less severe due to the aggregation of information on a national basis.\textsuperscript{62}

Variance truncation is a problem that is inherent in the design of the standards-based measures. Since the scales all use an ordinal, interval approach to rating country performance, coding countries on a limited-value scale allows for relative comparison but not for ranking.\textsuperscript{63} The measures' designers have explained that the limited-value scales are necessary because the source materials do not provide sufficiently precise or systematic information to code using more values. As David Cingranelli and David Richards explain:

\begin{quote}
We use the ordinal level of measurement because measurement schemes should complement the nature of the source information. Human rights information is far from perfect \ldots Thus, accurate numbers of violations are rare, and all counts contain an inherent amount of measurement error \ldots We use the latitude provided by ordinal score categories to allow for measurement error resulting from the qualitative information on which our scores are based.\textsuperscript{64}
\end{quote}

The designers of the major standards-based measures agree on the propriety of using limited-value scores, though they disagree about how many ordinal categories or values are

\begin{itemize}
\item \textsuperscript{60} Landman & Carvalho, \textit{supra} note 2, at 88-90.
\item \textsuperscript{61} \textit{Id.}
\item \textsuperscript{62} See Carey, Gibney & Poe, \textit{supra} note 58.
\item \textsuperscript{63} Landman & Carvalho, \textit{supra} note 2, at 88-90.
\item \textsuperscript{64} Cingranelli & Richards, \textit{supra} note 52, at 400.
\end{itemize}
optimal for measuring human rights practices given the limitations of existing data sources.\footnote{See, e.g., id. at 404 (asserting that more factors in scoring within a category lead to lower reliability); Wood & Gibney, supra note 38, at 378 (critiquing CIRI for having too few categories for meaningful differentiation among levels of violation); see also Armstrong, supra note 30, at 654 (discussing the distortion of data that results from Freedom House’s method of converting scores to a coded categories); Jørgen Møller & Svend-Erik Skaaning, supra note 35, at 56 (noting that Freedom House’s limited-value scale produces a “ceiling-effect, meaning that variation among the best performing countries . . . is not captured”).}

Another major set of issues addressed in the scholarly literature are the inherent limits of measurement projects that are based on very limited sources. The single two sources used by both CIRI and PTS are the U.S. DOS’s annual \textit{Country Reports on Human Rights Practices} (DOS Country Reports) and AI’s \textit{Annual Reports}. A brief overview of the two sources and their development is therefore relevant here, and provides important context for the discussion of the validity of CIRI and PTS as human rights metrics in Part IV of this Article.

\section*{B. Limitations Related to Sources: The Development of Modern Human Rights Reports}

\subsection*{1. Development of the Department of State Country Reports}

As mentioned above, DOS \textit{Country Reports} were chosen because they were the only source that provided “standardized information about the same rights for each country, annually,” for a significant period of time.\footnote{Cingranelli & Richards, supra note 52, at 400.} The \textit{Country Reports} were mandated by Congress through a law passed in 1976 that aimed at addressing the need for an objective assessment of whether countries had “engage[d] in a consistent pattern of gross violations of internationally recognized human rights.”\footnote{Judith Innes de Neufville, \textit{Human Rights Reporting as a Policy Tool: An Examination of the State Department Country Reports}, 8 \textit{Hum. Rts. Q.} 681, 683–84 (1986).} This assessment was made necessary by a 1975 amendment to the Foreign Assistance Act prohibiting U.S. development assistance to any country that met this standard.\footnote{Id. at 683.} At the time, no NGO was yet producing high-quality, relatively uniform data about human rights practices in relation to all—or even...
most—counties in the world. Indeed, the NGO-produced human rights report as we know it today did not exist in the mid-1970s.

The DOS therefore had to develop an approach to researching, vetting, and presenting data that until that time had not been regularly gathered cross-nationally. In many ways, the DOS was in both the best and the worst position to conduct this kind of research. It was privileged to have a large foreign service staff in its employ who could gather facts on the ground, the soft power of the government of the United States—useful in persuading governments to allow access to sensitive information—as well as access to classified information. No NGO could dream of having this combination of forces. However, the DOS also had the burden of being an arm of the U.S. government, a status that placed it in a delicate situation in relation to human rights information, and a status that has left its reports open to accusations of bias that have never disappeared, even as its reports have become more uniform and professional.

While early reports were uneven and not subject to uniform formats or reporting guidelines, this quickly changed and the reports took on a more standard format, tone, and content by the mid-1980s. By 1985, one scholar found a consensus among report users that the reports were a "basically fair effort to display an accurate picture" of the human rights situation in each country. Most of the reports were seen as accurate and free of falsified data, though a recent study by David Richards and Jillienne Haglund found that in a "near majority" of the reports examined for a study of violence against women, misleading or factually inaccurate legal information was presented. In addition, a consensus developed that a handful of the country reports were routinely subject to

69. Classified information may be used to verify facts, or otherwise drawn upon in an unclassified manner. See U.S. Gov't Accountability Office, GAO-12-561R, Human Rights: State Department Followed An Extensive Process To Prepare Annual Country Reports 7 (2012) [hereinafter US-GAO, Human Rights].

70. de Neufville, supra note 67, at 686–89.

71. Id. at 691.

political bias, making the reports for those countries less accurate.\textsuperscript{73}

The well-respected human rights NGO, the Lawyers Committee for Human Rights (LCHR, now Human Rights First), along with HRW and its earlier antecedents, the Watch Committees, dedicated significant resources to providing in-depth critiques of the \textit{Country Reports}, beginning in 1980.\textsuperscript{74} Specific issues raised by LCHR and others will be addressed in the analysis in Part IV; relevant here, the issue of political bias was a regular feature of the LCHR critiques, though the organization concluded that these biases were confined to a small subset of countries in each report. LCHR's careful critique and others like it have been credited with many of the improvements in the DOS reporting practices.\textsuperscript{75} Now, the DOS issues annually updated reporting instructions, which include guidelines for preparing the reports, as well as a mandated outline with headings and sub-headings to be used for each country report.\textsuperscript{76} While those guidelines were not made available to NGOs in the early years, they have been subject of careful analysis and commentary in later years.\textsuperscript{77}

In 2012, the federal Government Accountability Office conducted an audit of the \textit{Country Reports} and concluded that the DOS "has an extensive process designed to make the country reports on human rights as comprehensive, objective, and uniform as possible."\textsuperscript{78} Scholars who have examined the issue of bias empirically have found, on the whole, fairly minimal differences between the DOS \textit{Country Reports} and the AI An-

\textsuperscript{73} Id.

\textsuperscript{74} See \textit{Americas Watch Committee et al.}, \textit{supra} note 73, at i.

\textsuperscript{75} de Neufville, \textit{supra} note 67, at 688–89.

\textsuperscript{76} USGAO, \textit{Human Rights}, \textit{supra} note 69, at 6–7.


\textsuperscript{78} See USGAO, \textit{Human Rights}, \textit{supra} note 69, at 3.
In an article examining the impact of information effects on PTS and CIRI, Clark and Sikkink find some “pronounced” cases of political bias in State Department reports; they conclude that such bias has fluctuated over time. Poe, Carey and Vasquez report that:

Drawing on earlier, qualitative research and existing theories, we tested hypotheses consistent with arguments that the US State Department’s human rights [reports] are biased. While such arguments and allegations abound, to our knowledge there had never been any kind of systematic quantitative tests conducted to find if the historical record is consistent with those arguments. For the most part, hypotheses concerning those biases found limited support in our general analyses. The results indicate that the State Department’s reports, in comparison to those of Amnesty International, have at times favored US friends and trading partners while discriminating against its (perceived) leftist foes . . . these analyses gave us no reason to believe that State Department biases affected their assessments of the vast majority of cases during the twenty year period our data covered. This is because hypotheses consistent with the critics’ allegations of State Department bias explained only a very small percentage of the variance in the differences between the two reports. Further ‘good news’ . . . is that the two reports have clearly converged in their assessment of human rights violations over time. It seems likely this is because the US State Department has instituted improvements in the reports in response to its critics.

Even if the two sources tend to be fairly consistent with each other, each is produced by a Western-based institution using roughly similar methods. As one scholar has concluded,

79. Steven C. Poe, Sabine C. Carey & Tanya C. Vasquez, How are These Pictures Difference? A Quantitative Comparison of the U.S. State Department and Amnesty International Human Rights Reports, 1976-1995, 23 Hum. RTS. Q. 650 (2001). For a further discussion concerning the potential sources of bias in these sources, see infra Part IV(b)(ii).

80. The authors also report that U.S. officials have themselves admitted bias in specific country reporting. Clark & Sikkink, supra note 3, at 547, 553.

81. Poe, Carey & Vasquez, supra note 67, at 677.
"a more or less exclusive use" of "reference materials provided by organizations, media, etc. situated in western countries" will "undoubtedly introduce a systemic bias." 82

Surely the most pervasive systemic errors contained in the PTS and CIRI measurement projects, however, are those based on the inherent problem of reporting on physical integrity rights. The problem,

[A]s many human rights statisticians have explained, is that standard variables in this field only measure recorded and reported human rights violations, not actual violations. The problem is that improving human rights conditions increases access to information on the extent of violations . . . This limitation in the data can produce perverse measurement results: the more rights protective a state becomes the worse the state's record may appear in terms of detected human rights violations. 83

The question is whether the PTS and CIRI data sets examined here are so replete with systematic measurement error of this kind as to impact their reliability.

2. Development of Amnesty International's Annual Report

AI published its first Annual Report in 1962, 84 but it was not until the mid-1970s that this report took on the features that later made it an attractive source for systematic coding. Before that time, the organization was small, its work was heavily focused on political prisoners ("prisoners of conscience"), and the tools of human rights research and fact-finding were fledgling at best. 85 Between 1962 and the mid-1970s, the organization's Annual Reports featured information about the activities of the various AI membership groups, the finances of the organization, and cases "adopted" by AI. During the mid-to late-1960s, the Annual Reports began to feature short country summaries discussing the activities AI had undertaken in rela-

82. Skaaning, Measuring Civil Liberty, supra note 45, at 730.
83. Goodman & Jinks, supra note 8, at 175 (internal citations omitted).
84. ANN MARIE CLARK, DIPLOMACY OF CONSCIENCE: AMNESTY INTERNATIONAL AND CHANGING HUMAN RIGHTS NORMS 5-6 (2001).
85. See generally Clark, supra note 84.
tion to prisoners in specific countries. The number of countries covered increased steadily, from fifteen in 1966 to 107 in 1975. In 1975, the section once entitled "Countries in Which Amnesty has been Particularly Active" had become the more general and evaluative "Prisoners and Human Rights: Country by Country." In 1977, the year that the first DOS Country Reports were published, AI's Annual Report featured summaries concerning abuses against prisoners in 116 countries. In the 1970s and 1980s, these summaries grew into descriptions of human rights concerns related to political imprisonment, torture, disappearances, and extrajudicial executions.

C. The New Genre of the Human Rights Report

The PTS extends back to 1976, and the CIRI data set begins in 1981. While the early reports used by the PTS were quite uneven, by the time CIRI began coding them, the DOS Country Reports and AI's Annual Report were fairly standardized and uniform, and they have grown even more so over time. These qualities were the result of the professionalization of


90. By 2012, Amnesty International had a detailed style guide used by all researchers and editors of AI publications, with the Annual Report representing the most assiduously standardized report published by the organization each year. Telephone interview with Senior Amnesty International Employee A (Oct. 1, 2012).
human rights fact-finding both within government and in the NGO sector. During the first decade of their existence, the language used in the reports shifted, from occasionally "emotional" to detached, clinical, stripped of adjectives. Terms characterizing the quantum or quality of evidence began unevenly but became more regular. A new genre was born: the field research-based human rights report.

Relevant to this discussion are both fact this new genre’s existence and its relative youth at the time the PTS and CIRI begin coding reports. Indeed, human rights investigative and fact-finding methods were still being perfected in these early years, and were far from systematic. This and other realities behind the production of the two primary sources used by the PTS and CIRI are relevant to assessing their validity as measures of human rights because they result in reports that have varying levels of reliability. The next section explores some of the more significant background conditions relevant to this assessment.

IV. LIFTING THE NUMERICAL VEIL: THE EXTENT TO WHICH STANDARDS-BASED MEASURES REFLECT INTERNATIONAL HUMAN RIGHTS LAW

As empirical scholarship theorizes about what makes human rights law effective or ineffective using standards-based measures to assess the dependent variable of rights violation or protection, and as policy-makers and human rights advocates consider embracing and deploying standards-based measures of human rights in their work, it is important to assess the fit

91. See de Neufville, supra note 67, at 687; Clark, supra note 84, at 84–92 (describing AI’s sophisticated efforts to clarify terminology and categorize rights violations in reports and campaigns).

92. See de Neufville, supra note 67, at 687; Clark, supra note 84, at 45 (describing “AI’s characteristically cautious, understated, objective tone”). In 1996, AI made it an official policy to “avoid emotive or abusive expressions.” STEPHEN HOPGOOD, KEEPERS OF THE FLAME: UNDERSTANDING AMNESTY INTERNATIONAL 74 (2006) (internal citations omitted).

93. AMERICAS WATCH COMMITTEE ET AL., supra note 73, at 3 (finding that terms such as “alleged,” “claims,” and “documented” were used unevenly).

between what the measures purport to assess—violation of human rights, or subsets of such rights—and the coding schemes used to generate the indicators supporting that measurement. Ensuring such a fit is crucial in advancing efforts to make human rights work more empirically-grounded. This Part will examine the validity of the CIRI PIR Index and the PTS, assessing how their coding definitions and conventions match with or diverge from international human rights law.

This Part thus examines the conceptualization of specific human rights norms as defined in the coding conventions of each measurement project. It will also identify those realities behind the fact-finding efforts that produce human rights reports that are relevant to assessing the validity of CIRI and PTS data. Examining both law and practice are crucial to this assessment.95

As discussed previously, the creators of the PTS have explained that the measurement project captures three dimensions of personal integrity rights: “scope, intensity, and range,” suggesting that the different levels of the PTS scale reflect a “continuum of human rights practices.”96 This Part uses the three PTS dimensions as a conceptual framework for analysis, adding two more: systematicity and interdependence. Box C presents the major elements of the discussion. As explained in this Part, the two additional dimensions are added to reflect international human rights law and reporting practice, and because they are implicitly addressed, though not in these terms, by both the PTS and CIRI.

95. As Kenneth A. Bollen explains, “[s]ystematic error can arise from the subjective nature of ratings and the information on which ratings are based.” Political Rights and Political Liberties in Nations: An Evaluation of Human Rights Measures, 1950 to 1984, 8 Hum. RTS. Q. 567, 568 (1986).

96. Wood & Gibney, supra note 38, at 573–74.
<table>
<thead>
<tr>
<th>HR Dimensions</th>
<th>PTS</th>
<th>CIRI's PIR Index</th>
<th>How Approached</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scope:</strong> type of human rights violations</td>
<td>Physical integrity rights, including: murder, torture, forced disappearances, political imprisonment</td>
<td>Physical integrity rights, comprising: disappearances, extrajudicial executions, torture, and political imprisonment</td>
<td>PTS: list of specific abuses; aggregated CIRI: very specific definitions of violations; disaggregated</td>
</tr>
<tr>
<td></td>
<td>Committed by the government or with its acquiescence</td>
<td>Strict attribution only</td>
<td></td>
</tr>
<tr>
<td><strong>Intensity:</strong> number and frequency of human rights violations</td>
<td>Relative numbers/ frequency taken into account in coding, but no numerical thresholds are provided</td>
<td>Uses number thresholds, where such data are available, and key terms</td>
<td>PTS: frequencies are plotted on an overall scale and terms are part of coding; Abuses scaled to population size CIRI: number thresholds and specific terms are correlated with specific levels of abuse: numerous, epidemic, wholesale, extensive, many</td>
</tr>
<tr>
<td>HR Dimensions</td>
<td>PTS</td>
<td>CIRI’s PIR Index</td>
<td>How Approached</td>
</tr>
<tr>
<td>---------------</td>
<td>-----</td>
<td>-----------------</td>
<td>----------------</td>
</tr>
</tbody>
</table>
| **Range** | Abuses against certain victim groups are coded more strictly (“labor leaders” v. “apolitical peasants” as victims of EJEs). Extraterritorial abuses taken into account | Not explicitly measured, though abuses against non-citizens are explicitly excluded | PTS: identity or status of victims assessed  
CIRI: Not explicitly measured, but certain terms used in coding process are in fact capturing range: widespread, extensive, wholesale |
| Systematicity | Not measured | Not measured explicitly | PTS: not measured  
CIRI: Not explicitly measured, but certain terms used in coding process are in fact capturing systematicity: systemic, systematic, routine |
| Interdependence | Addressed through scaling and composite scoring | Addressed through summing of separate scores | PTS: Violations are scored together  
CIRI: Separate scores are summed for Index value |

A. Scope

1. Scope of PTS and CIRI Coding

Both the PTS and the CIRI PIR Index focus on rights considered to be “personal integrity” rights, encompassing disappearances, extrajudicial executions (“state-sanctioned killings”\(^\text{97}\)), torture, and political imprisonment. These four rights are clearly articulated and defined in international

---

\(^{97}\) About the Political Terror Scale, supra note 43.
human rights law and this specific grouping, developed and used by political scientists, has a clear logic as an analytical tool from a legal perspective.

The PTS does not use a manual to define the rights it codes. Instead, it sets out the basic contours of the violations encompassed and identifies included and excluded categories, types, or practices. In contrast, CIRI cites to international human rights law. For example, CIRI defines "political and other extrajudicial killings/arbitrary or unlawful deprivation of life" very specifically in its coding manual, citing to the right to life as set out in Article 6 of the International Covenant on Civil and Political Rights. Despite these differences, the conc-

98. Cingranelli & Richards, Measuring, supra note 59, at 407 (1999) (noting that "almost all empirical human rights research has focused on government respect for one category of human rights, physical integrity rights, as the main concept of theoretical interest").

99. The basis for the combined scoring, or scaling, of these rights, will be addressed in a later section of this Article.

100. For example, the designers of the PTS have described the category of government-sponsored killings, or extrajudicial executions as follows:

The PTS focuses on state-sponsored killings that take place outside of the normal judicial setting. These "extrajudicial" executions or killings include death squad killings of political enemies, unlawful use of lethal force by police forces (e.g., shooting unarmed suspects), intentional killing of civilians by security forces during combat, and other arbitrary deprivation of life by state actors. The PTS does not include state-sanctioned executions that occur after trials that conform to international standards. However, what constitutes a legitimate "legal" execution and what constitutes an "extrajudicial" killing is difficult to determine. As a general rule, the PTS will code summary executions or those that take place outside the context of a legal proceeding as illegitimate executions and exclude those killings that take place after legal proceedings.

Wood & Gibney, supra note 38, at 370–71.

101. The coding manual provides this definition:

Extrajudicial killings are killings by government officials without due process of law. They include murders by private groups if instigated by government. These killings may result from the deliberate, illegal, and excessive use of lethal force by the police, security forces, or other agents of the state whether against criminal suspects, detainees, prisoners, or others. Deaths resulting from torture should be counted, as these deaths occurred while the prisoners were in the custody of government or its agents. Deaths from military hazing also count. In most cases, the US State Department [USSD] reports indicate cases of political killings by explicitly referring to these killings as "political." A victim of politically-motivated
tours of the abuses that count, for example, as extrajudicial killings for coding purposes appear to be essentially the same in the two datasets, comprising the following elements: (a) state sponsored or state-instigated (b) killings or executions (c) carried out without due process of law. These elements largely mirror international human rights law, though some illustrative differences should be noted, as well as the fact that the source cited by CIRI is only one of many interlocking sources of relevant protections under treaty law. Whether a particular lethal use of force is lawful or not under international law depends on many factors, including the context in which it occurs.

For example, PTS includes "intentional killing of civilians by security forces during combat" while international human rights and humanitarian law would approach the elements differently, and depending on the context, would either exclude as lawful some such deaths or include as unlawful others not described by the measurement project's definitions. Additionally, while PTS apparently would count all civilians intentionally killed by military actors in combat as abuses, international law would require a more searching inquiry and would not count as unlawful certain of these killings. As a general matter, there are two different frameworks of law relevant to such killings: international human rights law (IHRL), which is always applicable, and international humanitarian law (IHL), which is only applicable in the context of armed conflict. IHL rules vary depending on whether the armed conflict is international or non-international, and it applies concurrently with IHRL, unless there are conflicts between the two regimes, when IHL will apply as lex specialis. As the UN Special Rapporteur on Extra-judicial Executions has explained:

Under the rules of IHL: [Killing] is only lawful when the target is a "combatant" or "fighter" or, in the case of a civilian, only for such time as the person "directly participates in hostilities." In addition, the killing

---

killing is someone who was killed by a government or its agents as a result of his or her involvement in political activities or for supporting (implicitly or explicitly) the political actions of opposition movements against the existing government. While they may be the result of different motives, both extrajudicial killings and political killings are to be treated identically for purposes of coding.

must be militarily necessary, the use of force must be proportionate so that any anticipated military advantage is considered in light of the expected harm to civilians in the vicinity, and everything feasible must be done to prevent mistakes and minimize harm to civilians. These standards apply regardless of whether the armed conflict is between States (an international armed conflict) or between a State and a non-state armed group (non-international armed conflict), including alleged terrorists. Reprisal or punitive attacks on civilians are prohibited.

[...]

*Under human rights law:* A State killing is legal only if it is required to protect life (making lethal force proportionate) and there is no other means, such as capture or nonlethal incapacitation, of preventing that threat to life (making lethal force necessary). The proportionality requirement limits the permissible level of force based on the threat posed by the suspect to others. The necessity requirement imposes an obligation to minimize the level of force used, regardless of the amount that would be proportionate, through, for example, the use of warnings, restraint and capture.\(^{102}\)

Because PTS appears to code all killings of civilians in combat as violations, PTS is likely to code as violations some killings that would in fact be acceptable under international law.

On the other side of the equation, the CIRI coding manual specifies that to count for the coding process, killings must be either committed by state actors or "instigated" by them,\(^{103}\) and PTS explains that it also focuses on acts by governments.\(^{104}\) However, IHRL would often qualify killings as unlawful.

---


103. CIRI Coding Manual, *supra* note 49, at 7 (extrajudicial killings “include murders by private groups if instigated by government”).

104. Wood & Gibney, *supra* note 38, explain at 371 that “[o]n some occasions, it is not clear if the state is the actor directly responsible for a given
ful killings at a lower threshold of state involvement—especially situations in which ostensibly private groups operate with the tacit support of the government (such as with death squads or paramilitaries). As the Special Rapporteur on Extrajudicial Executions has explained, killings by non-state actors will be the state’s responsibility under international law in at least the following circumstances:

(a) The State has direct responsibility for the actions of non-State actors that operate at the behest of the Government or with its knowledge or acquiescence. Examples include private militias controlled by the Government (which may, for example, be ordered to kill political opponents) as well as paramilitary groups and death squads;

(b) Governments are also responsible for the actions of private contractors (including military or security contractors), corporations and consultants who engage in core State activities (such as prison management, law enforcement or interrogation).

In addition, governments become responsible under international law for “fully ‘private’ killings, such as murders by gangs, vigilante justice, ‘honor killings’ or domestic violence killings” when:

[T]here is a pattern of killings and the Government’s response (in terms either of prevention or of accountability) is inadequate, the responsibility of the State is engaged. Under human rights law, the State is not only prohibited from directly violating the right to life, but is also required to ensure the right to life, and must meet its due diligence obligations to take appropriate measures to deter, prevent, investigate, prosecute and punish perpetrators.\textsuperscript{106}


\textsuperscript{106} Id., ¶ 46(d).
CIRI and PTS appear to include most of the killings encompassed in categories (a) and (b) above, though CIRI may be more conservative than international law by requiring a higher threshold of collusion with private groups. Neither measurement project appears to encompass killings in the final category, however, even where the failure to protect is established. This issue will be discussed again below in relation to the issue of range.

After assessing the validity of the metrics in relation to the scope of human rights abuses, it becomes apparent that the narrow construction of state attribution rules and the failure to examine due diligence obligations by PTS and CIRI lead the resulting data sets to be under-inclusive of unlawful killings. The same is true for torture and disappearances, which are often carried out by private actors. While the citation of human rights law is a welcome feature of the CIRI project, the narrow approach reflected in the manual can lead to under-inclusive categories. This failure to fully encompass the full scope of violations of IHL and IHRL is problematic because it may have the effect of distorting the evidence base upon which users of the data sets rely.

2. Scope of Department of State and Amnesty International Reports

The DOS Country Reports and the AI Annual Reports encompass the human rights coded by CIRI and the PTS, and they do so for a broad range of countries, but the geographical and substantive range of the reports evolved over time. Therefore, attention to the development of the reports’ range is relevant to assessing the validity of the PTS and CIRI PIR Index.

When Congress asked the DOS to report on whether potential recipients of U.S. assistance had “engage[d] in a consistent pattern of gross violations of internationally recognized human rights,” there was disagreement about what rights should be included in the reports.107 This is not surprising in light of the broad wording of the directive, and since this was one of the first instances of the institutionalization of human rights in modern U.S. policy.108

107. de Neufville, supra note 67, at 685.
Historian Samuel Moyn has placed the birth of the modern human rights era in the mid-1970s, calling 1977 the "year of human rights" since it was 1977 when President Carter invoked human rights as a central tenet in his inauguration speech and AI was awarded the Nobel Peace Prize. Although 1948's Universal Declaration of Human Rights had been translated into concrete binding norms in the 1960s through the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social, and Cultural Rights (ICESCR), the real work of implementing and interpreting the treaties was only beginning at the time of the Congressional mandate. More importantly, the idea of human rights as a movement, the object of public advocacy, came to the fore only in the 1970s. Human rights NGOs were a new phenomenon. AI had been established in 1961 but did not gain its worldwide status or reach until the 1970s. While several "Watch Committees" had been set up in the 1970s, beginning with Helsinki Watch in 1978, HRW was not created as a single NGO until 1988. By the late 1970s, this new human rights movement was "working to make the world a slightly less wicked place." As its leading institution, AI was shining a light on the most egregious violations by reporting on physical integrity rights. As historian Samuel Moyn explains, the human rights movement was born of a minimalist view of political utopia, one in which personal integrity rights were the focus and uniting locus, and economic, social, and cultural rights were very much an afterthought.

In this context, the U.S. DOS focused its reporting efforts on this same core of rights, though the reports did also include some discussion of "vital needs" such as food, shelter, health care, and education. The DOS looked to the Univer-

109. Id. at 120–75.
110. Id. at 155–58.
111. Id. at 3–4.
113. MOYN, supra note 108, at 147 (quoting Eckel).
114. Id.
115. See U.S. Dep't of State, 1977 Country Reports on Human Rights Practices (1978), http://babel.hathitrust.org/cgi/pt?id=mdp.39015078705632;view=1up;seq=3. The Department of State ceased reporting on even these "vital needs" during a period in the 1980s. See THE WATCH COMMITTEES AND LAWYERS COMMITTEE FOR HUMAN RIGHTS, CRITIQUE: REVIEW OF THE U.S.
sal Declaration of Human Rights for normative guidance.\textsuperscript{116} The \textit{Country Reports} are produced for all countries receiving U.S. foreign assistance and all member states of the United Nations each year.\textsuperscript{117} The \textit{Annual Reports} now encompass 194 country reports, which in 2010 ranged from nine pages (the Republic of San Marino) to 145 pages (China). They cover "internationally recognized civil, political, and worker rights" including "freedom from torture or other cruel, inhuman, or degrading treatment or punishment; from prolonged detention without charges; from disappearance or clandestine detention; and from other flagrant violations of the right to life, liberty, and the security of the person."\textsuperscript{118}

AI's early reports had a very narrow focus, spending most of their attention on the right to be free from political imprisonment. The organization had been founded to campaign impartially for the release of those the organization defined as "prisoners of conscience," individuals jailed for their beliefs, provided that they had not used or advocated violence.\textsuperscript{119} Ann Marie Clark describes AI's early reporting this way:

The origins of Amnesty International's independent reporting initiatives were modest and driven by the qualities of loyalty to principle and objectivity. Reaching for objectivity while remaining faithful to principles of truth, the nonviolent expression of political opinion, and most particularly to the well-being and release of individuals unjustly imprisoned, was Amnesty International's mainstay. As AI gathered knowledge of specific cases, it acquired a range of information on human rights conditions that few others could claim.\textsuperscript{120}

This unique evolution, from a focus on specific individual cases to an embrace of both cases and broader human rights conditions, reflects AI's organizational development away

\textsuperscript{116} See USGAO, \textit{Human Rights}, supra note 69, at 1.

\textsuperscript{117} \textit{Human Rights Reports}, U.S. \textsc{Dep't of State}, http://www.state.gov/j/drl/rls/hrrpt/.

\textsuperscript{118} \textit{Overview and Acknowledgments}, U.S. \textsc{Dep't of State}, http://www.state.gov/j/drl/rls/hrrpt/humanrightsreport/index.htm#wrapper.

\textsuperscript{119} See Clark, supra note 84, at 14–15.

\textsuperscript{120} \textit{Id.} at 18.
from the narrow prisoner-based focus at its founding. The narrow focus is reflected in AI's early Annual Reports, which focus on prisoner cases in a small number of countries. As the organization grew, efforts focused on increasing the range of AI's research while retaining the quality of its information. Limits of resources and access meant that it was impossible for AI to ever cover the entire globe in the annual Report.

As AI grew, and as the uses for the Report expanded, debates arose about the optimal approach to determining which countries would get the organization's attention. The organization's commitment to universal human rights and to a particular kind of impartiality led to the embrace for some time of the principle of universal coverage, which meant that AI strove to have a basic grasp of the human rights situation in all countries of the globe. This principle was later operationalized in the concept of "minimum adequate coverage," which "entailed some ongoing capacity to speak with authority on any country and a commitment in principle to cover all countries in the Annual Report." At some point in the 1990s, the principle of minimum adequate coverage was modified by the idea of "strategic coverage," through which countries were covered "for three main reasons: gravity of abuses, possibility of making a difference, or as test cases." While minimum adequate coverage remains a goal, with AI's expanded focus it has become more difficult to cover the whole world, and strategic coverage is a way of prioritizing the work in a considered way. Further, it is AI policy that it will not include an entry on a country in its annual Report unless it has undertaken some specific research in relation to that country in the relevant year; this policy limits the number of countries covered in each report.

The focus of AI's reporting has greatly expanded: from political imprisonment, beginning in 1961 to torture, beginning in 1968, and—as governments adapted their abusive methods—to disappearances, beginning in 1974 and extraju-

121. HoAcooD, supra note 92, at 100–01.
122. Id. at 182.
123. Id. at 182–83.
dicial executions, beginning in 1980.\textsuperscript{126} Then, AI's mission expanded beyond the original core of physical integrity violations by state actors in the 1990s and 2000s; specific non-state actor abuses were taken on in 1991,\textsuperscript{127} and economic and social rights were taken on in the 2000s.\textsuperscript{128} As AI's work grew, the topics covered in the annual Reports also expanded. With each expansion, the reporting was less than comprehensive, growing more systematic over time. This pattern has almost definitely led to unevenness in reporting year by year as the topics changed and grew over time.

Indeed, from the start, the annual Report has been "a record of Amnesty International's work and concerns"\textsuperscript{129} during the year under review, presenting information about human rights concerns as filtered through the organization's programmatic priorities and research limitations.\textsuperscript{130} In order to keep the reports a manageable size, AI has had to adjust its approach to scope of topical coverage. For example, it is now standard practice to include an entry pertaining to the specific issues that are the subject of a global membership campaign during the relevant years, in as many country entries as possible.\textsuperscript{131} Thus, during the violence against women campaign, the annual Report included information about such violence in

\begin{itemize}
  \item \textsuperscript{126} AI was founded to work for the release of "prisoners of conscience," defined as individuals imprisoned for their beliefs, provided that they had not used or advocated violence. As AI's work expanded, the organization took on the fight against torture by government officials, no matter the identity of the victim. \textit{See Clark, supra} note 84, at 37 (torture), 70 (disappearances), and 101 (extrajudicial executions). This work expanded into a global campaign in 1972. \textit{See Jayne Huckerby & Sir Nigel Rodley, Outlawing Torture: The Story of Amnesty International's Efforts to Shape the U.N. Convention against Torture, in Human Rights Advocacy Stories} 17 (Huwitz & Satterthwaite with Ford eds., 2009).
  \item \textsuperscript{127} \textit{Amnesty Int'l, Amnesty International Report} 1992 13–14 (1992).
  \item \textsuperscript{129} \textit{Amnesty Int'l, Introduction, in Report} 1988 I (1988).
  \item \textsuperscript{130} Telephone interview with Senior Amnesty International Employee A (Oct. 1, 2012).
  \item \textsuperscript{131} Interview with Senior Amnesty International Employee B, in New York City, N.Y. (Sept. 27, 2012).
\end{itemize}
relation to numerous countries.\textsuperscript{132} This inclusion in some years and not in others might appear to signal that such violence fluctuated; instead, it is often a reflection of the programmatic priorities adopted within the organization toward specific human rights issues. Further, conscious of its role as one among many human rights organizations, AI sometimes focuses less heavily on a specific issue in a particular country in favor of another priority issue if it knows that another credible human rights organization is doing work on that topic.\textsuperscript{133} This can lead to less comprehensive reporting on some issues in the annual Report that might seem to reflect changing patterns of human rights abuse (where fewer cases are reported, for example, than might have been uncovered if that issue had been a focus of research and reporting), but which in fact signal a change in emphasis or even an informal division of labor among organizations.\textsuperscript{134}

Another dimension that impacts scope of coverage are the different practices of the DOS and AI concerning reporting on human rights abuses by non-state actors in the context of collusion or collaboration. As discussed previously, while CIRI operates under a strict view and will not count abuses unless they were “instigated” by the government, the PTS appears to take into account a broader range of non-state actor abuses, such as killings committed by death squads. In the reporting context, this issue translates into whether there is sufficient evidence of government involvement to make an allegation of such a connection, and the language used to characterize such connections. This question has been debated over the years in relation to both DOS and AI reports. The Lawyers Committee for Human Rights has critiqued the Country Reports for framing evidence such that it is “presented in a fashion that appears to diminish the government’s responsibility for violations.”\textsuperscript{135} One example given by the Lawyers Committee involved the Country Report on Georgia for 1994, which sug-

\textsuperscript{132} See generally, e.g., AMNESTY INT’L, AMNESTY INTERNATIONAL REPORT 2005 (2005). The “Stop Violence Against Women” campaign was launched in 2004. See Hopgood, supra note 92, at 33.

\textsuperscript{133} Telephone interview with Senior Amnesty International Employee A (Oct. 1, 2012).

\textsuperscript{134} Id.

\textsuperscript{135} LAWYERS COMM. FOR HUMAN RIGHTS, CRITIQUE: REVIEW OF THE DEPARTMENT OF STATE’S COUNTRY REPORTS ON HUMAN RIGHTS PRACTICES 1994
gested that specific violations were committed by an ostensibly non-governmental paramilitary group; in fact, according to the Lawyers Committee for Human Rights, this "paramilitary" was an official government department.136

AI has long been committed to reporting on the activities of groups that purport to be separate from the government, but which in fact operate in collusion with the state. Tracing such connections was a feature of AI's work on disappearances in the 1970s and 1980s, in part because the organization would then only work on cases in which such a link could be identified.137 This issue became more straightforward when AI decided to devote attention to abuses by non-state actors in 1991.138 The new policy meant that abuses committed by paramilitaries or other armed groups would be more broadly investigated and documented by the organization than in the past, when such abuses would need to be clearly attributable to the government to make it into AI reports. This change could be assumed to produce an increase in the number of reported cases within the CIRI and PTS coding rules.

While the precise timing of these changes in policy and practice are not important here, these different approaches are relevant in assessing the data drawn from the AI Annual Report, since its comprehensiveness and quality should be expected to fluctuate—perhaps significantly—over time, from country to country, and across rights topics. If coding decisions do not take these fluctuations into account, the validity of the data could be impacted.

B. Intensity

1. Intensity in PTS and CIRI Coding

PTS and CIRI both measure intensity—the number or frequency of human rights violations—but each does so differently. CIRI uses numeric thresholds as guidelines for its coders. For example, the coding guide specifies that a country where fifty or more extrajudicial killings have been reported in one year is a country where such acts are "practiced fre-
quently," meriting the country a score of 0 for this right. A country where between one and forty-nine such killings are reported is a country where killings are "practiced occasionally," and should be coded a 1. Countries where no extrajudicial killings have been reported should receive a 2.

The PTS designers have critiqued this approach, pointing out that most reports do not specify numbers of cases at all. This makes sense since such numbers—of cases of extrajudicial executions, torture, political imprisonment, or disappearances—are highly unreliable in the rare instances that they are available. As indicated above, this is true for a variety of interlocking reasons. First, governments and related forces usually do not admit to using torture, enforced disappearance, or extrajudicial execution. When governments do admit to being responsible for the underlying acts—such as in the case of political imprisonment—they usually reject the label, characterizing such detentions as security measures, responses to "hooliganism," or the like. This makes counting cases very complex.

The CIRI designers respond to the PTS critique by explaining that while the numeric thresholds are intended to guide coding decisions and enhance inter-coder reliability, they are in fact rarely used. CIRI instructs coders to also examine the terms being used to describe the frequency of the abuses:

In practice, these numeric thresholds are rarely used to produce scores because usable numeric estimates

139. Cingranelli & Richards, supra note 52, at 8.
140. Id.
141. Id.
142. Wood & Gibney, supra note 38, at 379.
143. In some rare cases, such as in relation to the U.S. practices of extraordinary rendition, secret detention, and "enhanced interrogation techniques" during the presidency of George W. Bush, governments have aggressively redefined the relevant legal standards, rejecting the labels otherwise applicable to these acts under international law. UN bodies have made clear repeatedly that extraordinary rendition and secret detention as carried out by the United States amount to enforced disappearance, and that "enhanced interrogation techniques" like waterboarding are torture. See Margaret L. Satterthwaite, Extraordinary Rendition and the Role of Law, 75 GEO. WASH. L.R. 1333 (2007).
144. Cingranelli & Richards, Measuring, supra note 59, at 407.
145. Cingranelli & Richards, supra note 52, at 401–02.
are seldom provided in our qualitative source material. Thus, most of our physical integrity scores are based on qualitative descriptions, just like the indicators that are not based on countable phenomena. For example, CIRI coders are instructed that 'instances of 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).' We take great care in putting together these lists of key terms, going so far as to regularly use concordance software to analyze the language of the Country Reports so that simple changes in language do not affect scores. When doing content analysis on an annual multi-authored source over time, one needs to ensure that scores are based on practices, not on changes in the way similar things are described from year to year.146

On the other hand, the PTS designers assert that their scale captures intensity better than the CIRI PIR Index does.147 Not only do they critique CIRI for using numeric thresholds,148 they also argue that CIRI suffers from more significant variance truncation than PTS since CIRI uses scales of 0-2 or 0-3, where PTS uses a scale with more gradations (1-5).149 PTS also takes into account the size of the population of

146. Cingranelli & Richards, supra note 52, at 401-02.
147. Wood & Gibney, supra note 38, at 373, 376.
148. Wood and Gibney assert:
It is incorrect to make the assumption that the event counts necessarily convert into qualitative categories. For example, we are somewhat puzzled by how '50 or more' instances of abuse translate into 'widespread' abuse. 'Widespread, systematic, and extensive' are qualitative terms that reflect, to a certain degree, the selectivity of abuse and its relative frequency, whereas counts of abuse are objective values that are constant across population size and do not reflect any inherent pattern of violence. Both approaches are valid (though we prefer the former), but we question the logic of mixing the two. We are likewise concerned about the potential for creating inaccurate comparisons between those cases in which the score is based on a count and those in which the score is based on narrative descriptors.

Id. at 380.
149. Id. at 378-79.
a country in relation to the frequency of violations; CIRI’s designers have explicitly rejected such an approach, pointing out that this approach treats the same instance of abuse as having more weight in a small country in comparison to its coded value in a populous country. However, they do concede that “China and India almost always receive our lowest scores,” in part because they are so populous.

Finally, while the CIRI coding guidance, by providing both numeric thresholds and key terms, is aimed at ensuring reliability in the face of qualitative data that is often widely divergent and far from uniform, the analysis in the later sections of this article suggests that some of the terms identified in the CIRI coding manual as useful for assessing intensity/frequency are in fact capturing other dimensions that are also relevant under human rights law, albeit with varying significance.

2. Intensity in Department of State and Amnesty International Reports

The use of numbers, especially statistics, in DOS and AI reports, has been contested, and merits close attention in understanding the limits of the PTS and CIRI coding schemes. As they worked to standardize and professionalize their annual reports, AI and the U.S. DOS stood in very different positions in relation to the quantitative measurement of human rights violations. As noted above, the DOS had a comparatively large staff which often had access to foreign governments and institutions on a level that an NGO could not expect to equal. This meant that the DOS could, theoretically, obtain statistics on issues within its reporting ambit. On the other hand, numbers related to political imprisonment, torture, disappearances, and extrajudicial executions was not the everyday work of the official statistics offices or criminal justice institutions to which the U.S. government had comparably better access. Indeed, any official statistics about the types of abuses reported in the DOS Country Reports were likely to be erroneous and often intentionally so. Thus, while the U.S. government could be expected to gather information about rights violations through its Embassy staff, in many ways AI had a comparative advantage: close relationships with victims, their families, and those

150. Cingranelli & Richards, supra note 52, at 415.
151. Id.
working on their behalf. In the 1970s and 1980s, data about specific human rights violations were being collected largely through the events-based reporting of such close-to-the-ground advocates. Interviews with rights professionals who have followed the DOS *Country Reports* carefully have revealed that in many countries, the Embassy staff responsible for the reports used the same sources as AI and other human rights organizations, though in the early days, many DOS human rights officers would instead seek information through the large international human rights organizations themselves.\(^{152}\)

In the late 1970s, the DOS's Bureau of Human Rights and Humanitarian Affairs issued guidelines to Embassy staff responsible for human rights reporting that requested that, where possible, quantitative estimates of violations should be provided with their human rights reports.\(^{153}\) The numbers were to be coupled with assessments about their reliability.\(^{154}\)

One scholar identified the following internal effects of this drive to quantify:

> First, the [reporting] requirement has altered standard operating procedures, organizational norms and goals, and the relative power of interests within the Department of State. The fact that reporting includes an effort to quantify violations has reinforced these effects. In general, the reporting has directed organizational attention to human rights. The effort to quantify has required the building of expertise and has meant that over time many F.S.O.s [foreign service officers] have learned in direct, empirical ways about human rights. The data have helped create internal advocates for human rights by changing the perceptions of the data gatherers and, in turn, the data lend credibility to these advocates in internal policy debates and help mobilize public support for their efforts.\(^{155}\)

*Country Reports* often include numerical information drawn from secondary sources. Sometimes these numbers con-

---

154. *Id.* at 687–88.
155. *Id.* at 695–96.
lict, and the DOS will often provide several numbers along with their sources; for example, the 2010 report on Colombia includes two different estimates about the number of trade union activists killed, specifying that the lower number is drawn from a government source, which defined trade union membership more narrowly than the NGO that produced the higher number. A recent change in the Country Reports may impact the ability of coders to draw inferences about intensity; in 2011, the DOS:

[D]eveloped a streamlined format for each country report. As a result, we do not attempt to catalog every incidence, however egregious, of a particular type of human rights abuse in a country. Rather, we spotlight examples that typify and illuminate the types of problems frequently reported in 2011 in that country. The mention of fewer cases in a particular report should not be interpreted as a lessening of concern for the overall human rights situation in any particular country. Rather, our goal is to shed light on the nature, scope, and severity of the reported human rights abuses.

Counting cases will now be an even less effective method for assessing intensity in relation to the Country Reports.

AI has a complex history with numerical information about human rights abuses. The 1973 Annual Report includes a feature page titled “How many prisoners in . . . ?” that reads as follows:

The statistical question most frequently asked of Amnesty International is how many political prisoners there are in a certain country or in the entire world. It is a question almost impossible to answer. Few governments acknowledge that they have political prisoners as such, much less publish complete lists of them. Even when brought to trial, political prisoners are usually charged under the common criminal law of the country and subsequently absorbed into statistics covering criminal imprisonment. Amnesty Inter-

156. See USGao, Human Rights, supra note 69, at 7.
national's information about political prisoners is gleaned from a variety of other sources—newspapers, released prisoners, prisoners' families, exiles, reputable international organizations, etc. Sometimes the information, coupled with government announcements and detailed knowledge of a country's political situation, enables Amnesty to estimate total numbers fairly accurately. Most often, however, the information precise though it is about individuals and groups of prisoners—is at best only a rough pointer at the overall numbers detained. No accurate national figures can be given. Nor do the numbers of prisoners of conscience adopted by Amnesty International in a country reflect the total picture. Limited resources may make it physically impossible to adopt more than a percentage of the prisoners. Furthermore, these percentages vary widely according to individual circumstances. Thus a large total of adoptees in one country does not necessarily mean that it has more political prisoners than a country in which a smaller number are adopted.\textsuperscript{158}

A similar statement was made more recently by AI in response to a 2007 study by researchers associated with the Colombia-based Centro de Recursos para el Análisis de Conflicto critiquing AI and HRW reporting on Colombia. The study criticized AI and HRW for their use of numerical estimates in relation to abuses in Colombia, pointing to the use of these estimates in the CIRI and PTS measurement projects; the authors conclude: "we are very skeptical of human rights panel data sets, such as the PTS and CIRI data, that are built" in part on AI data.\textsuperscript{159} Rejecting the criticism as inaccurate and ill-informed, AI explained that its objectives are to end human rights violations, not to create statistics about such abuses, and

that it relies on qualitative methods almost exclusively to reach this goal.\footnote{Amnesty Int'l, \textit{Amnesty International Response to Andrés Ballesteros, Jorge A. Restrepo, Michael Spagat, Juan F. Vargas, The Work of Amnesty International and Human Rights Watch: Evidence from Colombia, CERAC, Colombia, February 2007, AI Index AMR 23/006/2007 (Feb. 21, 2007).}

Senior AI staff interviewed for this Article echoed this general analysis, and one explicitly rejected the idea of transforming certain words used by the organization into estimates about intensity. The staff member suggested that the word "widespread," for example, could not be reliably coded for systematic analysis, stating "it is going to be different in different countries or different in differing contexts. While there is attention [within AI] to these things, it is not enough to support quantitative analysis."\footnote{Telephone interview with Senior Amnesty International Employee A (Oct. 1, 2012).} As another expert who works for a different organization explained, estimating frequency in human rights reporting is "more of an art than a science" and such artfully conceived estimates should not be the basis for quantitative analysis.\footnote{Interview with Senior Human Rights Watch Employee A, New York City, N.Y. (Oct. 5, 2012).}

As Cingranelli and Richards have explained, "usable numeric estimates are seldom provided" in the DOS or AI reports.\footnote{Cingranelli & Richards, \textit{supra} note 52, at 401–02.} Wood and Gibney agree, and eschew attempts to use numeric values in the coding process.\footnote{Wood & Gibney, \textit{supra} note 38, at 378.} Despite these protes-
tations, both measurement projects attempt to capture intensity of abuse.

C. Range

1. Range in PTS and CIRI Coding

The designers of the PTS assert that one of the advantages of its methodology is that range—understood as the identity or status of the populations targeted for violation—is taken into account in the coding process:

[R]ange is intended to differentiate among abused groups based on the groups' observable actions, behaviors, or associations. That is, the PTS would assign a lower score to a state that was responsible for killing hundreds of political activists, labor leaders, and protesting students compared to a state that was responsible for executing the same number of apolitical peasants.165

This appears to be potentially problematic from a human rights perspective. It may be an interesting political science inquiry to theorize the difference between a government that tortures or kills a group of labor leaders as compared to another government that tortures or kills the same number of peasants. This victim-based distinction would not be one human rights law would find significant, however. Indeed, quite the opposite: under human rights law, all persons have the same right to be free of unlawful killing, and under the right to equality and equal protection under the law, a government must not draw distinctions like those PTS finds valuable here. Further, the right to freedom of assembly and freedom of expression would provide additional protections relevant to the analysis here, since "labor leaders" may need added protection related to their activism in some circumstances.

CIRI does not draw this type of distinction. CIRI's coding scheme does, however, exclude from coding violations against two groups of people that human rights law does in fact protect: non-citizens and victims of the extraterritorial actions of a state.166 These are both areas in which there has been signifi-

165. Id. at 374.
166. CIRI Coding Manual, supra note 49, at 4-5. Note that CIRI does include extraterritorial violations in "certain cases of occupation." Id. at 5.
cant evolution in human rights law in recent years. With respect to non-citizens within the territory of the state, it does not make sense under human rights law to exclude violations of their physical integrity rights from the CIRI measurement project. While there are some rights that governments can limit or deny in relation to non-citizens (such as the right to vote), this is not the case for three of the rights included in the PTS and CIRI PIR Index. The rights to be free from torture, unlawful killing, and disappearance must be upheld regardless of the citizenship status of the person suffering the violation, and human rights bodies have affirmed the right of all people, including non-citizens, to enjoy freedom from these egregious violations. The right to be free of political imprisonment is more complex, since there are situations when governments can limit the rights of freedom of expression and assembly of non-citizens. However, since such restrictions are allowed only insofar as they are aimed at protecting national security, public order or the rights and freedoms of others, it is very unlikely that cases of political imprisonment of non-citizens would fall into these exceptions. In sum, the CIRI PIR Index is out of line with international human rights law when it comes to the rights of non-citizens.

CIRI's exclusion of extraterritorial abuses is more understandable but still problematic. Until recently, there has been intense debate about whether and how human rights law applies to the activities of states when they act abroad. In recent years, court decisions and the authoritative interpretations of human rights bodies of relevant instruments have made clear that norms such as the rights to be free from torture, disappearances, political killing, and political imprisonment do extend to cover any person or territory under the jurisdiction of a state. Thus, torture, disappearances, extrajudicial execution, and political imprisonment carried out abroad would plainly be covered by human rights law. By excluding such violations, the CIRI PIR Index is out of step with human rights law.

It is not entirely clear how the PTS codes extraterritorial abuses. In the PTS's "Frequently Asked Questions," its designers have explained that they take a "common sense" approach to coding such violations:

167. See Satterthwaite, supra note 135.
What about political violence that a state engages in outside of its own territorial borders? This is a difficult question. In coding for the United States, for example, should this score also compute American activities in Iraq? For the most part, we are only measuring political violence that a state carries out within in its own territorial borders. However, a situation such as American activities at Guantanamo simply cannot be ignored. Thus, we seek to use common sense on such matters.168

This explanation raises more questions than it answers. First, the only example given is the United States, which is arguably a sui generis case, especially in relation to practices in Iraq and Guantánamo. Second, what is common sense in this case? Why not refer to human rights law for the relevant standard, which would allow for meaningful distinctions between those activities abroad that should be coded into the data set (here, violations that occurred under the jurisdiction of the United States) and those that should not (violations where the U.S. link is not so tight)?

Range is also impacted by the approach of CIRI and PTS concerning non-state actors, since some groups will be targeted more often by non-state actors instead of state agents. Although the state would become responsible for such abuses under the standard of due diligence, such abuses would be excluded by PTS and CIRI without proof of government collusion. Here we have an example of how range is impacted directly by scope—overly strict rules of attribution and a failure to take into account states' affirmative obligations under human rights law distort the PTS and CIRI data sets in this respect. Further, the impact can be expected to be disproportionate, falling more heavily on reports about groups that often face private actor abuse, such as when "undesirable" members of a society are routinely targeted for death.169 Even more pervasive is private actor violence against women, which appears not to be taken into account by either PTS or CIRI

168. About the Political Terror Scale, supra note 43.

unless it is instigated directly by the state. Under human rights law, states must protect individuals against private violence by taking a wide range of specific steps. If a state has not done so with due diligence, it becomes responsible for human rights violations. Such acts may amount to torture in some circumstances, as the UN Committee Against Torture has explained:

[W]here State authorities or others acting in official capacity or under colour of law, know or have reasonable grounds to believe that acts of torture or ill-treatment are being committed by non-State officials or private actors and they fail to exercise due diligence to prevent, investigate, prosecute and punish such non-State officials or private actors consistently with the Convention, the State bears responsibility and its officials should be considered as authors, complicit or otherwise responsible under the Convention for consenting to or acquiescing in such impermissible acts. Since the failure of the State to exercise due diligence to intervene to stop, sanction and provide remedies to victims of torture facilitates and enables non-State actors to commit acts impermissible under the Convention with impunity, the State's indifference or inaction provides a form of encouragement and/or de facto permission. The Committee has applied this principle to States parties' failure to prevent and protect victims from gender-based violence, such as rape, domestic violence, female genital mutilation, and trafficking.170

The PTS and CIRI databases thus appear to be systematically excluding a large swath of violations that amount to torture under human rights law.

2. Range in Department of State and Amnesty International Reports

Putting to one side abuses against non-citizens and the issue of extraterritoriality, reporting by both AI and the DOS has evolved over time in relation to range in ways relevant to

CIRI and PTS coding practices. An important example here are changes that impact the extent to which reporting by the DOS or AI will have accurately reflected the populations targeted.

The DOS guidelines have evolved over time, encompassing specific reporting requirements related to abuses based on race, gender, religion, disability, and language and social status. In a series of revisions to the reporting guidelines beginning with a comprehensive overhaul in 1993, the DOS set out instructions about such reporting.

AI has frequently highlighted specific populations in campaigns in ways that impact reporting on those groups in the Annual Report. The organization has launched campaigns on specific populations, such as its groundbreaking first report on violations against women in 1991, Women in the Front Line, and a six-year global campaign on violence against women between 2004 and 2010. AI also often campaigns on particular groups as a part of larger campaigns, such as its current work for the rights of indigenous peoples in connection with its broader economic and social rights campaign, Demand Dignity. In the time leading up to these campaigns and while they are under way, AI devotes additional research and reporting resources to the populations under examination. This can be expected to have the effect of increasing the depth and frequency of reporting on such violations. Such changes may impact the validity of any conclusions drawn about the range of abuses from AI's Annual Reports.

In addition to the changes in reporting that result from specific policy changes and campaign choices concerning specific populations, there have also been important changes that


172. Id. For example, now the Department of State routinely reports on the pervasiveness of different forms of violence against women and state efforts to eradicate it.


stem from the evolution of human rights practice that may be more subtle. An illustrative example is the change in reporting on grave human rights abuses committed against lesbian, gay, bisexual, and transgendered people (LGBTs). While personal integrity abuses such as torture, extrajudicial executions, and disappearances have been within the reporting ambit of both the DOS Country Reports and the AI Annual Report no matter who the victims were, in reality abuses against LGBT persons were largely excluded until particular policies and practices were adopted relating to abuses of LGBT persons. Within AI, a long debate about whether to work for the release of individuals imprisoned for their LGBT identity led to the embrace of such campaigning in 1991.175 While technically the organization would have called for the cessation of torture, killings and disappearances of LGBT persons in the same circumstances as it would have done for others before 1991 (and the organization did report on such abuses sporadically), it was not until 1991 that the organization began to lay the groundwork needed to report in a systematic manner on such abuses.176 This groundwork included steps such as making contact with local and national LGBT organizations, beginning intentional work to identify relevant cases, and delineating guidelines for which cases would be taken up by AI and which would not.

Similarly, the DOS began reporting on human rights abuses against LGBT people in 1990.177 However, the reporting has evolved slowly, with LGBT organizations issuing a call to action in 2007 to improve the quality of the reporting through better training of human rights officers.178 Abuses against LGBT populations pose additional challenges that are

relevant here: in many countries, LGBT people are still unable to operate openly, meaning that organizing efforts, documentation work, and networking with human rights organizations in such places remains nascent or non-existent. In other places, LGBT organizations are strong and capable, allowing them to make known the abuses they suffer. Reports on countries in the former category may include fewer cases of abuse in comparison with countries in the latter, reflecting the on-the-ground reality in reverse. This example demonstrates that reporting on the range of abuses in any given country will depend not only on the reality on the ground, but also on how resources are directed, whether needed networks have been established, and how assiduously the relevant fact-finders have been in chasing down leads. This reality means that even in cases where there is a good match between human rights law and the PTS and CIRI coding conventions, the reports being coded may vary unpredictably, and this variability could have undetected impacts on the resulting data sets.

D. Systematicity

1. Systematicity in PTS and CIRI Coding

Another dimension that is relevant for human rights measurement, but which neither the PTS nor CIRI explicitly code is the dimension of systematicity—the degree to which human rights violations reflect a pattern of abuse resulting from government policy or amounting to a de facto policy. Although the CIRI designers have not addressed the issue on these terms, the data project attends to words that implicitly capture systematicity: “systematic,” “routine,” and “regularly” are set out in the coding manual as key words to look for in relation to assessing intensity/frequency.\textsuperscript{79} The coding manual instructs coders to understand these terms as indicating that the specific violations “have occurred frequently,” meriting the country a score of zero for that violation.\textsuperscript{80} A more accurate reading of these terms might be to understand them not as being about frequency per se, but instead as designating that there is evidence that a particular type of abuse was used as a

\begin{itemize}
\item include a recommended program of training and awareness-raising for state actors regarding non-discrimination and equality for LGBT people).
\end{itemize}

\textsuperscript{80} \textit{Id.}
matter of explicit or de facto policy—of either action or inaction.

Under human rights law, it is relevant whether a pattern of abuse can be established, and whether it stems from a formal or informal policy, since government responsibility—including individual criminal responsibility for government officials—will attach differently in the context of a pattern of violations. While systematicity and frequency are related, they are separate issues under human rights law and practice. For example, the Convention Against Torture and Other Forms of Cruel, Inhuman or Degrading Treatment or Punishment empowers the Committee Against Torture to initiate an inquiry into situations when it "receives reliable information which appears to it to contain well-founded indications that torture is being systematically practiced."181 This phrasing notably does not include mention of frequency. Indeed, violations could be systematic but not frequent, as in the case of violations against a specific, but small, group of persons such as high-value terrorism suspects. Thus, U.S. waterboarding, which was allegedly used against only a handful of persons, was perhaps not frequent, though it was systematic, having been carefully designed and implemented according to specific authorizations by high-level officials.

The phrase "grave or systematic" was used by the drafters of the Optional Protocol to the Convention on the Elimination of All Forms of Discrimination Against Women (CEDAW) as a way of encompassing abuses that are particularly severe ("grave") as well as those that are being practiced as a matter of policy ("systematic"). Article 8 of the Optional Protocol authorizes the CEDAW Committee to undertake an investigative inquiry into violations of the Convention when the "Committee receives reliable information indicating grave or systematic violations by a State party of rights set forth in the Convention."182 The conjunction "or" here signals the different meanings of the two terms.


The dimension of systematicity is especially well-developed in international criminal law. Indeed, the element of systematicity, in some circumstances, transforms the treatment of a set of abuses as not only rights violations, but also as crimes against humanity, which are directly punishable at the international level. The Rome Statute, for example, sets out the chapeau of the definition of crimes against humanity as follows:

For the purpose of this Statute, “crime against humanity” means any of the following acts when committed as part of a widespread or systematic attack directed against any civilian population, with knowledge of the attack. . . .

In this definition, the term for frequency, “widespread” is set out as an alternative to the term capturing systematicity, “systematic.” Attention to the existence of a policy is also included in the Rome Statute’s description of war crimes, where the existence of “a plan or policy” is a relevant factor in determining whether the International Criminal Court will have jurisdiction over specific crimes. In addition, the crime of genocide includes an important element that is closely related to systematicity: genocide only exists if the perpetrator acted with specific genocidal intent, committed particular acts against specified groups, and “[t]he conduct took place in the context of a manifest pattern of similar conduct directed against that group or was conduct that could itself effect such destruction.”

It is important to note, however, that while the dimension of systematicity most often signifies the existence of a policy, it


185. Id. at art. 8, “War Crimes” (“The Court shall have jurisdiction in respect of war crimes in particular when committed as part of a plan or policy or as part of a large-scale commission of such crimes.”).

also encompasses situations where a government fails to act in response to a pattern of systematic abuses carried out by non-governmental actors. Thus, as one human rights expert explained in relation to NGO reporting, the quantum of evidence required to establish systematicity will in part depend on who the perpetrator is: less is required for government actors, and more for private actors. The latter form of systematicity is especially important in relation to abuses against groups that are targeted for discrimination and marginalization in the private sphere. A commentary to the CEDAW Optional Protocol explains:

The term "systematic" refers to the scale or prevalence of violations or to existence of a scheme or policy directing the violations. Violations that do not rise to the level of severity implied by the term "grave" may nonetheless be the subject of an inquiry if they evince a consistent pattern or are committed pursuant to a scheme or policy. Violations may be systematic in character without resulting from the direct intention of the State Party. For example, discrimination may be the result of social and cultural factors that the State has difficulty in affecting or a gap between policy and law as determined by the government and implementation at the local level. . . . The existence of a scheme or policy is therefore not a necessary element of systematic violations.

A similar point is included in the comments accompanying the ICC's Elements of Crimes in relation to crimes against humanity. The commentary clarifies that a failure to act could amount to a policy in exceptional circumstances where the context demonstrates that the inaction was "consciously aimed

---

at encouraging" a systematic attack against a civilian population.  

While international crimes are not captured as such by the PTS and CIRI data sets, some of the constitutive acts that make up these crimes are included (torture, murder, disappearance, certain forms of imprisonment), meaning that in practice many such crimes are coded into the data set as human rights violations. Further, since the CIRI coders are specifically asked to code terms that capture systematicity as though they were terms related to frequency, this dimension may not be accurately reflected in the resulting data set.

2. Systematicity in Department of State and Amnesty International Reports

Evidence suggests that there has been increasing attention to systematicity in reports by the DOS and AI. In the 1980s, HRW and the Lawyers Committee for Human Rights stressed the need for the DOS to identify and report on patterns of violations "when necessary to shed light on responsibility for, or the significance of, a violation."  

The DOS’s 1993 instructions to drafters of country entries specified that the reports should “render an opinion on whether a pattern of abuse is present.”  

The Lawyers Committee for Human Rights wrote in 1995 that the Country Reports had a “tendency to conceal patterns of state responsibility” by virtue of its “atomized” approach to laying out allegations of abuse: in relation to Colombia, LCHR concluded that the “complex and pervasive nature of state involvement in human rights abuse is obscured by the report’s excessive focus on isolated events.”  

In 2012, under the helm of the former Executive Director of LCHR, then-Assistant Secretary of State Michael Posner, the


192. Id. at xvii.
DOS’s *Country Reports* changed its approach to identifying and reporting on patterns of abuse, using “a streamlined format for each country report,” spotlighting “examples that typify and illuminate the types of problems frequently reported in 2011 in that country.”

AI’s original focus on individual cases of “prisoners of conscience” meant that the organization’s efforts were devoted, in significant part, to researching and reporting on specific cases. As the organization’s work expanded—especially when political killings were taken on board—attention to systems of abuse became more important. This change in emphasis grew over time, with AI concerned to determine whether patterns and policies existed in relation to all abuses within its ambit—from torture and disappearances to violence against women and economic and social rights.

Further, AI became an important actor in the fight against impunity around the world, especially in campaigns to establish the ad hoc criminal tribunals for Rwanda and the former Yugoslavia, and later for the International Criminal Court. The global fight against impunity involved establishing facts sufficient to demonstrate the systematicity of human rights violations, identifying their authors, and pushing for individual punishment. AI’s work to establish institutions at the international level to hold leaders to account for their criminal actions was born of the organization’s experience in seeking redress for individual victims, which often foundered in the face of systematic failures to hold such leaders to account: impunity. Such impunity characterized the transitional phase in Latin America once the military dictatorships responsible for disappearances and political killings were dismantled.

AI’s reports often focus on impunity and on facts needed to establish crimes that are by definition crimes of systematicity such as genocide and crimes against humanity, as well as crimes that do not require such evidence including torture,

---


extrajudicial executions, and disappearances. AI is now engaged in a global campaign for international justice: from 2010 to 2016, the organization is working to ensure "justice, truth and full reparations for victims of serious human rights violations." The existence of this campaign means that AI's Annual Reports for these years will likely highlight the systematicity of violations where possible.

E. **Interdependence**

1. **Interdependence in PTS and CIRI Coding**

   It is an axiom of human rights law that rights are "indivisible, interdependent, and interrelated," which is taken to mean that rights form a single net of protection; that "the realization of one right often depends, wholly or in part, upon the realization of others," and that some rights cannot be prioritized over others. Whelan has argued that this axiom has been inadequately explored, and carries more political force than conceptual power. Despite the lack of precision in these terms, it is important to note that the relationship among rights is recognized and advanced by human rights advocates. At this stage in the human rights movement, interrelatedness is in many ways a corrective to the traditional tendency to prioritize civil and political rights over—often to the exclusion of—economic and social rights. But in relation to measurement, the issue of interrelatedness is closely linked to the issue of aggregation and disaggregation: the extent to which rights are considered so linked that they can be coherently aggregated in a coding system or must be disaggregated, with individual scores developed for separate rights.

---


199. *Id.*

Both the PTS and the CIRI measurement projects capture this type of interrelatedness through their grouping together of specific rights as "physical integrity rights." PTS and CIRI group together, and score together, violations of rights that have been seen as especially severe: unlawful killings, torture, disappearances, and political imprisonment are all considered violations of fundamental human rights, and even \textit{jus cogens} violations—peremptory norms binding on all states.\footnote{See \textit{Restatement (Third) of Foreign Relations Law} § 702 (1987) (cataloguing these and other \textit{jus cogens} violations); M. Cherif Bassiouni, \textit{International Crimes: Jus Cogens and Obligatio Erga Omnes}, 59 \textit{Law \& Contemp. Probs.} 63, 67 (1996) (discussing \textit{jus cogens} status and sources).} Further, as noted above, many of the international crimes that have reached \textit{jus cogens} status, such as genocide, crimes against humanity, and war crimes, encompass constitutive acts such as torture, disappearances, extrajudicial executions, and arbitrary imprisonment. As a result, coding for these specific acts will automatically also include some instances of these crimes. The PTS designers have characterized the violations included in the data set as the "most egregious and severe crimes against humanity."\footnote{Steven C. Poe \& C. Neal Tate, \textit{Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis}, 88 Am. Pol. Sci. Rev. 853, 854 (1994).} While this is a misstatement from a legal perspective (since not all instances of these violations are in fact crimes against humanity), it does indicate the intentions of the designers—to capture the most severe violations of fundamental personal integrity rights.

The CIRI designers, on the other hand, have stressed both the importance of capturing information about physical integrity rights and the need to expand analyses of the type that utilize standards-based human rights measures beyond this subset of human rights.\footnote{Cingranelli \& Richards, \textit{supra} note 52, at 396–97.} This is why the PIR Index is only one part of the CIRI measurement project, which extends to other civil and political rights and even a limited set of economic and social rights.

CIRI has disaggregated its variables into specific rights categories: the four rights included in its PIR Index are each coded separately (disappearances, political and extrajudicial killings, torture, and political imprisonment), whereas the PTS codes them all together. The PTS designers argue that the CIRI approach generates a "pretense of precision and accu-
racy” that is misleading based on the source materials, while the CIRI designers retort that the lack of disaggregation in the PTS “can hide variation in government respect for human rights” that is significant for analysts.

However, the CIRI designers did feel that it was meaningful to create an index that considers physical integrity rights together through a composite score. The decision to do so was based in part on the fact that “[a]lmost all empirical human rights research has focused on government respect for one category of human rights, physical integrity rights, as the main concept of theoretical interest.” The justification for creating a single score, however, was based on an empirically provable relationship among the specific rights violations that justified their treatment under the larger concept of “physical integrity rights.” This relationship, unidimensionality, is an indicator of “the extent to which the items in a scale measure one and only one construct,” here “physical integrity rights.” If items in a scale are unidimensional, a scale or index pooling the individual scores will be justified. The CIRI designers found “strong scalability” in their four physical integrity rights and concluded that “because of this strong scalability, we can confidently sum the scores of our four indicators of government respect for particular physical integrity rights into an additive scale showing the overall level of government respect for physical integrity in a country.”

With this empirical grounding, the CIRI designers present both disaggregated scores for each of the four rights, and a composite score for all of them summed together for each country/year—a PIR Index score. Each country/year will receive a score of 0, 1 or 2 for each right, and the PIR Index scores will therefore range from 0 (worst) to 8 (best).

The PTS presents only a composite score, and does not provide disaggregated scores for each right. While the PTS designers do not prove the scalability of their index in the way
the CIRI designers do, they argue that their data set is more accurate because its coding scheme captures the dimensions of scope, intensity, and range in an integrated scale.\(^{212}\) It does so by plotting different types of violations on the scale alongside others and presenting descriptions of patterns of abuse that are used as models for coders.\(^{213}\) To understand the function of this scaling technique, the language used by the PTS to describe the intensity or frequency of different violations at each level is distilled in Box D below.\(^{214}\)

<table>
<thead>
<tr>
<th>PTS Level</th>
<th>Torture</th>
<th>EJE</th>
<th>Political Imprisonment</th>
<th>Disappearances</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>committed against &quot;the whole population&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>committed against &quot;large numbers of the population&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>may be common</td>
<td>may be common</td>
<td>extensive</td>
<td>[none]</td>
</tr>
<tr>
<td>2</td>
<td>exceptional</td>
<td>rare</td>
<td>limited amount</td>
<td>[none]</td>
</tr>
<tr>
<td>1</td>
<td>rare or exceptional</td>
<td>extremely rare</td>
<td>none</td>
<td>[none]</td>
</tr>
</tbody>
</table>

In sum, differing frequencies are associated with each level of the PTS scale for different human rights violations.\(^{215}\) This is in sharp distinction to CIRI, where the numeric guidelines are constant for each of the rights in the PIR Index.\(^{216}\) On this point, Cingranelli and Richards assert that the PTS assumes "an a priori ordering among types of physical integrity viola-

\(^{212}\) Wood & Gibney, supra note 38, at 379–80.
\(^{213}\) See supra Box A, Section II(b).
\(^{214}\) The Political Terror Scale, Documentation: Coding Rules (2015), http://www.politicalterroryscale.org/Data/Documentation.html. The brackets around ‘none’ in relation to disappearances is because these violations are not mentioned at Levels 1, 2, and 3.
\(^{215}\) Cingranelli and Richards construct a table of “a priori expected patterns of government violations of particular physical integrity rights as specified by the Political Terror Scale” that is slightly different than Box C since they have translated the PTS language into categories of none/rare/some/many/common/very common. See Cingranelli & Richards, supra note 52, at 415.
\(^{216}\) CIRI Coding Manual, supra note 49, at 8 (EJEs), 13 (disappearances), 18 (torture and other cruel, inhuman and degrading treatment or punishment), 22 (political imprisonment).
tions” that has not been “empirically justified” in the way their unidimensional scale has been.\textsuperscript{217} Here, the critique centers on the structure of the PTS scale itself, which presents a specific pattern of worsening human rights violations—where some abuses increase at one rate and others at another rate—without proof that the pattern is in fact reflective of facts on the ground.

Both the CIRI PIR Index and the PTS treat the category “personal integrity rights” as so interrelated that a single score is seen as valid and helpful for quantitative analysis. As noted above, there has been some criticism of such composite scores from within the political science literature, but the concerns are different in nature than the concerns arising from human rights law and practice. James McCormick and Neil Mitchell argue that:

[H]uman rights violations differ in type not just amount, such that they cannot be clearly represented on a single scale. That is, there is a substantive difference between the use of imprisonment on the one hand and the use of torture and killing on the other. Substantively, these are quite different types of government activity, with differing consequences for the victims, differing use of governmental resources and capabilities, and differing costs for the government, both domestically and internationally. Normatively, too, there is a considerable distance between a regime that relies on imprisonment as a method of political control and one that relies on torture and killing.\textsuperscript{218}

Svend-Erik Skaaning suggests that aggregation rules about how the different components of a composite score will be combined should be based in “an explicit theory concerning the relationship between the attributes.”\textsuperscript{219} Instead, standards-based measures, including the CIRI PIR Index, mostly use simple addition without testing alternatives.\textsuperscript{220} This approach to

\textsuperscript{217} Cingranelli & Richards, \textit{supra} note 52, at 412.
\textsuperscript{219} Skaaning, \textit{Measuring Civil Liberty, supra} note 45, at 730 (internal citations omitted).
\textsuperscript{220} Id. at 729 (internal citations omitted).
aggregation is common to many such metrics, but the lack of theoretical grounding is problematic when dealing with complex concepts like human rights.221

These concerns—that the reality of governmental abuse may not be accurately reflected in standards-based measures, and how and to what extent individual types of abuses co-occur—are relevant to human rights practitioners. Certainly, human rights law would agree that "political imprisonment and torture are qualitatively different activities," but not because of the comparative costs each imposes on the government committing the violations, and not because the items within the indices may not be "unidimensional." Instead, human rights law is concerned with setting out norms that proscribe violative behavior in clear and mandatory terms. Measurement of the reality on the ground against such norms cannot be aggregated to form a score that would be meaningful under law. It matters a great deal as a legal matter whether a country has tortured people, imprisoned individuals for political reasons, disappeared people, and unlawfully killed individuals. Each abuse matters in and on its own terms, and so aggregated measures, which confound analysis of the differing prevalence levels of each of the component abuses, and which re-aggregate rights concepts that have been carefully disentangled and articulated are unlikely to be particularly helpful for the human rights practitioner.

While using composite scores may be important for comparisons over time or space or for predictive analysis, most human rights practitioners are not focused on such things. Instead, they seek actionable information that they can use for naming and shaming or other advocacy in the specific context of their action. The focus is not, in other words, on identifying a typical pattern of rights violations across a large set of countries. Instead, the specific pattern of abuses in this country at this time is what matters. Human rights advocates certainly do attend to patterns, but they do so at different levels, and for different reasons. Like human rights law, such practitioners are intensely focused on identifying the specific patterns of abuses in particular countries during particular periods.

221. Moller and Skaaning find that, with respect to measures of the rule of law, "[i]n no case was the aggregation procedure grounded in theory." Moller & Skaaning, supra note 35, at 60.
2. Interdependence in Department of State and Amnesty International Reports

The type of interdependence embodied in the PTS and CIRI PIR Index is not represented in DOS and AI reports. Indeed, the recommendation that measures be disaggregated would face few practical hurdles in the organizations' manner of reporting for the reasons identified above. Indeed, the CIRI data project has used DOS and AI reports for their disaggregated measures from the start, apparently without incident.

From the beginning, the Country Reports and the AI Annual Report have organized their factual material using the categories set out in international human rights law, adjusting the categories as needed over the years to take into account evolving law and changing patterns of violations. DOS reports have even been faulted for doing this to an extreme; the Lawyers Committee for Human Rights complained that the separate, "atomized" accounts of violations of different rights had the effect of "obscure[ing] the connections between different categories of human rights abuse."222

However, this type of reporting makes is especially amenable to disaggregated coding. The CIRI coding manual, for instance, identifies specific headings in the Country Reports that will contain information relevant to coding each of the individual rights in the PIR Index.223 Earlier AI reports are not organized this way, but more recent Annual Reports include useful headings within country entries concerning the specific rights violations under examination. For these reasons, the reports underlying the coding process by the PTS and CIRI are well designed for disaggregated coding of the type called for by analysts who have examined this issue carefully.

Finally, there is a perverse form of interrelatedness among human rights violations and the reporting that exposes them, which was first observed during the period when the AI and DOS reports became regularized. The theory is that the human rights movement's success (and AI's in particular) in bringing government violations to light by shaming those who held and tortured political prisoners can be measured through the pivot made by many governments from imprison-

222. LAWYERS COMMITTEE FOR HUMAN RIGHTS, supra note 171, at vii.
223. See CIRI Coding Manual, supra note 49.
ing their foes to disappearing and killing them.\textsuperscript{224} A similar theory has been advanced in recent years as the Obama Administration prohibited the notorious practice of rendition to torture of terrorism suspects while ramping up their covert killing by drone (targeted execution). Relatedly, rights advocates in Colombia have documented a shift in paramilitary abuses in relation to reporting practices: in a context where the killing of four or more persons at once was labeled a massacre, paramilitaries shifted to the tactic of “multiple homicides,” in which people are “killed over a period of several days and their bodies scattered, evidently in order to avoid the designation of a massacre.”\textsuperscript{225} Darius Rejali has advanced the related, but more subtle argument that human rights pressure on governments has changed the type of torture now used: while torturers in the past would destroy bodies and create scars, current-day torture is “clean” and seeks to leave no marks.\textsuperscript{226}

To observe this kind of interrelatedness requires disaggregated measures that allow for assessment of governments’ “choice strategies” concerning different types of human rights violation.\textsuperscript{227} Scales like the PTS are not granular enough to allow for such assessments. Using disaggregated measures, studies must also be carefully designed to measure governments’ “strategic behaviour and substitutability” related to types of abuse:

Measuring one area of human rights without concurrently measuring the others would have misconstrued the patterns and prevalence of human rights conditions on the ground . . . . greater compliance with one obligation (e.g., reduction in disappearances)

\textsuperscript{224} See Clark, supra note 84, at 104 (“[T]here appeared to be a haunting possibility that the apparent effectiveness of prisoner adoption techniques had caused governments to impose more final, and secret, ways of dealing with opponents. Political prisoners now attracted sustained international attention; outright killing and disappearances ending in death were more deniable— and irreversible.”); see also Goodman & Jinks, supra note 8, at 174.

\textsuperscript{225} Winifred Tate, Accounting for Absence: The Colombian Paramilitaries in U.S. Policy Debates, in SEX, DRUGS, AND BODY COUNTS: THE POLITICS OF NUMBERS IN GLOBAL CRIME AND CONFLICT 215–46, 241–42 (Peter Andreas & Kelly M. Greenhill eds., 2010).

\textsuperscript{226} See Darius Rejali, TORTURE AND DEMOCRACY 1–33 (2009).

\textsuperscript{227} McCormick & Mitchell, supra note 218, at 516.
can show up as lower compliance with another (e.g., increase in unfair trials). Therefore, a model studying only unfair trials would show human rights conditions worsening, even though the overall country conditions may be improving . . . 228.

This form of interrelatedness is of crucial importance to understanding and advancing human rights.

V. CONCLUSION

This Article has drawn back the curtain on several popular standards-based measures, examining the concepts, sources, and methods involved in the coding process from the perspective of human rights law and practice. The drive to standardize, the push to quantify, and the desire to compare across time and space have produced data sets that suffer from some specific and seemingly meaningful validity problems. While determining whether these issues are significant enough to introduce distortions in the evidence base of studies relying on these data sets for human rights analysis is beyond the scope of this Article, recent scholarship suggests that they may be. 229

The analysis has also uncovered idiosyncrasies attributable to practices within the institutions responsible for creating the qualitative reports upon which standards-based measures are based. While the identification of some of these problems may appear irrelevant or unfair to human rights advocates who do not prepare their reports in order to have them subjected to coding by political scientists, other issues may well be the type of concern that advocates would in fact be interested in addressing. Specifically, greater methodological transparency by human rights organizations is called for. If it is true that, as Beth Simmons has found, quantitative analysis of human rights organizations uncovers "some patterns that, to date, more detailed case studies have not brought squarely to our attention," then it is in everyone’s interest to improve the data underlying this work.

228. Goodman & Jinks, supra note 8, at 174.
229. See Dai, supra note 11, at 91 (concluding that "the improvement in human rights practice concerning political rights is more visible in some areas using some indicators, than in other areas with different indicators").
230. Simmons, supra note 8, at 11.