International Studies Perspectives

International Studies Perspectives (2013) **14**, 199–220.

Disaggregating Torture Allegations: Introducing the Ill-Treatment and Torture (ITT) Country-Year Data*

Courtenay R. Conrad
University of North Carolina at Charlotte

Jillienne Haglund and Will H. Moore
Florida State University

The Ill-Treatment and Torture (ITT) Data Collection Project uses content analysis to measure allegations of government ill-treatment and torture made by Amnesty International (AI) from 1995 to 2005. ITT's country-year (CY) data quantify AI allegations of ill-treatment and torture at the country-year unit of observation and further across different responsible government agents and across different econo-socio-political groups of alleged victims. This paper introduces the Ill-Treatment and Torture country-year data, describes quantitative patterns likely to be of interest to researchers focused on the study of international non-governmental organizations (INGOs) and human rights, and suggests a number of theoretically motivated questions that can be explored using the ITT country-year data.

A primary mission of international non-governmental organizations (INGOs) like Amnesty International (AI) and Human Rights Watch (HRW) is to monitor government human rights records, calling attention to transgressions and pressuring states for reform. As part of their activities, INGOs periodically allege that a government is responsible for ill-treatment or torture, or they accuse a government of demonstrating a pattern of abusive behavior.[1] Several well-known data collection efforts have utilized INGO reports to create ordinal data about government human rights abuses (for example, Hathaway 2002; Gibney *et al.* 2009; Cingranelli and Richards 2010). These projects have been invaluable, supporting a wide variety of statistical analyses that have helped researchers better understand the covariates of

*A previous version of this paper was presented at the 2011 Annual Meeting of the International Studies Association, Montreal, Quebec. Thanks to Aysegul Aydin, Jessica Brandwein, Christian Davenport, Peter Haschke, Madhav Joshi, Emilia Powell, Chris Sullivan, Priyam Trivedi, Carrie Booth Walling, Eric Wiebelhaus-Brahm, and Reed Wood for helpful comments and suggestions on earlier drafts. The ITT project has received support from the US National Science Foundation (Grants #0921397 and #1123666), the Department of Political Science at Florida State University, Kroc Institute, University of Notre Dame, and the School of Social Sciences, Humanities, and Arts at the University of California, Merced. Additional information on the ITT Project is available at http://www.politicalscience.uncc.edu/cconra16/UNCC/ITT_Data_Collection.html.

[1]The defining characteristics of this behavior, often called information politics, include unbiased research, grassroots mobilization and fund raising, and the distribution of information to raise awareness of transgressions of human rights (for example, Keck and Sikkink 1998; Cmiel 1999; Clark 2001b; Ron *et al.* 2005).

Conrad, Courtenay R., Jillienne Haglund and Will H. Moore. (2012) Disaggregating Torture Allegations: Introducing the Ill-Treatment and Torture (ITT) Country-Year Data. *International Studies Perspectives*, doi: 10.1111/j.1528-3585.2012.00471.x.

human rights performance (for example, Poe and Tate 1994; Cingranelli and Richards 1999a; Apodaca 2001; Keith 2002; Davenport and Armstrong 2004; Bueno de Mesquita *et al.* 2005; Hafner-Burton 2005; Neumayer 2005; Abouharb and Cingranelli 2007; Hathaway 2007; Richards and Gelleny 2007; Vreeland 2008; Keith *et al.* 2009; Powell and Staton 2009; Simmons 2009; Cingranelli and Filippov 2010).

Although existing data have paved the way for large-N, cross-national research on human rights, the data produced by the Ill-Treatment and Torture (ITT) Data Collection Project are unique in at least two ways. First, rather than conceptualize INGO reports about human rights as evocative of the performance of states vis-à-vis their international obligations, ITT quantifies Amnesty International *allegations* of ill-treatment and torture. Allegations are distinguishable from "true" levels of state human rights violations, which are inherently unobservable (for example, Spirer 1990; Clark 2001a:57; Cingranelli and Richards 2001:230–1). This is of theoretical and empirical import, as the activity of INGOs is of considerable interest to a broad research community within international relations and international law. Furthermore, although differences between actual violations of human rights and allegations of that behavior result in validity and reliability challenges (Bollen 1986), human rights data collection projects to date have not grounded their efforts conceptually in allegations. The ITT Project does so.[2]

Second, in addition to producing an ordinal measure of AI allegations of torture in all countries with populations over one million, the Ill-Treatment and Torture data include information on a number of additional characteristics of the allegations advanced in AI publications. More specifically, the ITT project performed content analysis of all AI publications from 1995 to 2005 to measure allegations of torture at two units of observation: specific allegations (SA) and country-year (CY) allegations. The distinction between these two units of analysis involves the breadth of their spatial–temporal domain. Country-year allegations concern the general use of torture across a country throughout a year by a particular government agency (if specified). They are more general in nature than specific allegations, and they apply only to reports that describe torture occurring across an entire country over an entire year. The ITT project refers to allegations of torture occurring within a limited time (that is, less than a year) or space (for example, a region, a specific prison) as specific allegations (Conrad and Moore 2011:8–9). This paper reports information on country-year data.[3] The CY data code information on three characteristics not available in other cross-national data on state torture: the government agency alleged to have committed the abuse, the econo-socio-political group of which the alleged victim is a member, and whether or not AI claims the accused government obstructed NGO/INGO access to victims.

Because of their unique structure, the Ill-Treatment and Torture country-year data are the first cross-national human rights data that allow researchers to disaggregate allegations of torture by state agency and victim type. Consider an example. Cingranelli and Richards (2010), a commonly used quantitative measure of state torture,[4] codes both Liberia and Portugal as engaging in torture "frequently" for the majority of years from 1998 to 2001.[5] The ITT CY data tell a much more nuanced story. From 1998 to 2003, the ITT data report AI allegations of "system-

[2]Although ITT's distinction between the actual level of disrespect for a given right and allegations about violations is novel with respect to the collection of cross-national human rights data, it is a common distinction in the INGO community.

[3]The ITT SA data were released in early 2012 and will be described in a separate paper.

[4]The ordered Cingranelli and Richards (2010) measure of respect for the right not to be tortured ranges from zero to two. A score of zero indicates that torture was practiced "frequently," while a score of two indicates that torture did not occur in a given country-year.

[5]Liberia is coded zero in all years from 1998 to 2001; Portugal is coded zero in 1998, 1999, and 2001 and coded one in 2000 (Cingranelli and Richards 2010).

atic" torture in Portuguese prisons and no allegations of statewide torture against the Portuguese military. In Liberia, AI makes no allegations of statewide torture against either police or prison officials from 1998 to 2001, but instead alleges "systematic" and "widespread" torture against the Liberian military during the same period. Similarly, although the CIRI torture data code both Georgia and Zimbabwe as engaging in "frequent" torture, the ITT country-year data show that AI accuses the Georgian government of "widespread" torture against criminals and no torture against dissidents; Zimbabwe, on the other hand, goes unnoticed for criminal torture, but faces allegations of "systematic" or "widespread" dissident torture from 2000 to 2005. This type of variance is unobservable in other quantitative data on state torture. As such, we hope the ITT CY data encourage human rights scholars to refine theories about the mechanisms that influence domestic respect for rights across state agencies and across victim types.

In the following section, we provide an overview of the technical details of the ITT country-year data. After presenting the key variables, we discuss the reliability and validity of our measures and provide researchers with suggestions to deal with missingness in the data. We then look at the effect of commonly used covariates of state repression on AI allegations of government torture across state agencies and victim types, discussing in detail the steps researchers should take to draw inferences about violations of human rights more broadly rather than allegations specifically.

Nuts and Bolts of the ITT Country-Year (CY) Data

The Ill-Treatment and Torture country-year data code Amnesty International allegations of state torture and ill-treatment when the perpetrator is an agent of the state, the victim is a person detained under the state's control, and the alleged abuse meets the definition of torture in the United Nations Convention Against Torture (UN CAT):

> torture means any act by which severe pain or suffering, whether physical or mental, is intentionally inflicted on a person for such purposes as obtaining from him or a third person information or a confession, punishing him for an act he or a third person has committed or is suspected of having committed, or intimidating or coercing him or a third person, or for any reason based on discrimination of any kind, when such pain or suffering is inflicted by or at the instigation of or with the consent or acquiescence of a public official or other person acting in an official capacity. It does not include pain or suffering arising only from, inherent in or incidental to lawful sanctions.

ITT codes allegations of government torture only when a state is functioning: During periods of state collapse or foreign occupation, the project does not code allegations.[6] We do not produce data for country-years coded by the Polity project as not having a functioning state.

Key Variables of Interest

Level of Torture (LoT)

The main variable of interest included in the ITT country-year data is a measure of AI allegations of torture incidence: It is a modified version of the ordinal

[6]AI makes allegations about both state and nonstate actors and does so without regard to our definitions of state collapse. Coders were instructed to record all allegations against state actors, but we exclude from our data those allegations made during years when the Polity project codes the state as failed or occupied. In late 2012, all of the files needed to replicate the ITT data collection will be made available on the ITT project web site; interested researchers can use those files to obtain the data on failed or occupied states.

scale proposed by Hathaway (2002) to code the LoT alleged by AI to have occurred *throughout a country over the course of an entire year*, as described in Conrad and Moore (2011). Let us unpack that statement. As noted above, existing efforts to collect cross-national data on the extent to which governments engage in ill-treatment and torture endeavor to measure government behavior (for example, Hathaway 2002; Cingranelli and Richards 2010). The ITT project expressly does not code government behavior. Instead, it codes AI *allegations* about government ill-treatment and torture, taking into consideration that (i) it is not possible to know, much less report, the actual level of ill-treatment and torture occurring in any one country (Bollen 1986; Spirer 1990; Rejali 2007) and (ii) AI is a strategic actor with limited resources that issues a report only when it is highly confident about the accusation, and further where it believes it is most likely to influence governments (for example, Orentlicher 1990; Cmiel 1999; Clark 2001b; Hopgood 2006). Although ITT CY data explicitly contain *allegations* of state torture, we suggest methods for drawing inferences about *violations* of human rights below.

The ITT country-year data code only those allegations that make claims about abuse occurring throughout a country over the course of an entire year. To illustrate, if AI alleges that people held in a specific prison are frequently beaten, that allegation would not be recorded for the CY data because it is limited in its spatial domain. Similarly, allegations of torture during an election would not be recorded for the CY data because they are limited in their temporal domain. Instead, these allegations would be coded using ITT's specific allegation (SA) coding rules.[7] The Ill-treatment and Torture project adopts the Hathaway (2002) ordinal scale to measure the intensity of government ill-treatment and torture as reported by AI. Coders recorded country-wide allegations occurring throughout the year that used one of the following key words:[8]

- 1 = Infrequent
- 2 = Some(times)
- 3 = Frequent
- 4 = Widespread
- 5 = Systematic

Interested observers and scholars alike often wish to distinguish among types of torture, implicitly identifying some types as worse than others. For example, ill-treatment is often implicitly considered a less "intense" form of abuse than torture. Human rights activists, however, rarely make such distinctions. With respect to the CAT, a given act is either a violation under international law or it is not. AI publishes allegations about violations of the CAT, and the ITT project uses the LoT scale to code the frequency with which these allegations occur. As such, it is important to emphasize that the LoT variables in the CY data *do not* imply change in the "intensity" of violations and should not be used to measure such a concept. Instead, the LoT variables measure AI allegations of the frequency with which a state violates the CAT throughout a given country during a particular year.

[7] This difference suggests that the ITT data may not correlate strongly with the Hathaway (2002) data or the Cingranelli and Richards (1999a) data. Those projects code AI allegations about people held in a specific prison being frequently beaten, as well as allegations of abuse surrounding elections. We should note, however, that those projects code only Annual Reports; in its Annual Reports, AI by and large limits itself to broad allegations about abuses that occur throughout the country.

[8] We identify a number of synonyms for these terms. Please refer to Conrad and Moore (2011) for more information. In country-years for which AI did not make any allegations that met these criteria, the CY data record a value of 0: No Allegation.

Agency of Control (AoC)

The ITT country-year data are further distinct from previous data collection efforts on government torture and ill-treatment in that they code not only the incidence of torture alleged in AI reports, but also (i) the government agency AI alleges to be responsible for the abuse, and (ii) the type of victim that AI alleges was abused by the state. ITT distinguishes among five government agencies, which we label AoC: Police, Prison, Military, Intelligence, Immigration Detention, and Paramilitary (Conrad and Moore 2011:11–12). AI's allegations do not always identify a government agency, so the data include a sixth category: Unnamed/Not Stated. Because ITT codes AI's allegations, coders were instructed to code what AI alleged: They were not to use the AI report as a cue to divine what government agency was responsible for any given allegation of abuse.

Victim Type (VT)

Turning to VT, ITT coders distinguished between four groups of victims: Criminal, Political Dissident, Member of a Marginalized Group, and State Agent (Conrad and Moore 2011:13–14). As with the AoC variable, VT contains an Unnamed/Not Stated value. The Political Dissident category includes prisoners of conscience, human rights activists, and protesters. Members of marginalized religious and ethnic groups, the elderly and youths, and immigrants are all coded as Members of a Marginalized Group. If AI alleges that a victim is an illegal immigrant, the coders recorded both Criminal and Member of a Marginalized Group. Neither values on AoC nor values on VT are mutually exclusive.

Restricted Access (RA)

Aside from country-year data on LoT, AoC, and victim type, ITT codes data on RA, assigned a value of one in any year for which AI published a statement that it, or another INGO, had difficulty gaining access to detainees in that country (Conrad and Moore 2011:14). Because ITT LoT data report AI torture allegations, rather than actual torture incidence, we recommend using Restricted Access as a control in any statistical analyses that use LoT as a dependent variable.

Missing Data, Validity, and Reliability

ITT country-year data are available in four distinct structures, all of which report the alleged LoT for the relevant units:

- Country-Year (CY)
- Country-Year, Agency of Control (CYAoC)
- Country-Year, Victim Type (CYVT)
- Country-Year, Agency of Control, Victim Type (CYAVT)

In what follows, we discuss dealing with missing data in the CY data sets, as well as the validity and reliability of our key measures.

Incomplete/Missing Data

A number of AI allegations of torture and ill-treatment make explicit reference to continuation of, or change from, a status quo. Unfortunately, information about the status quo is not always provided in previous reports, or the referenced year does not fall into ITT's temporal domain. As such, the ITT data include negative values for phrases that AI uses with some frequency to identify a state's practice as improving, worsening, or staying the same.[9] When allegations

[9]Please see Conrad and Moore (2011:5–7).

that use these phrases do not include additional information that made it possible for coders to assign a value on the LoT scale, they assigned a value of “better,” “worse,” or “status quo.” In cases in which coders were able to assign a value in the preceding year, we wrote a computer batch file that added one to, subtracted one from, or simply used the preceding year’s value, accordingly. Nevertheless, after both human and computer assignment of change given the relevant previous year’s values, the CY data have a number of country-years for which no information is available. Thus, these cases have negative values in the data set.

Conrad and Moore (2011, 11) discuss a variety of options for addressing these incomplete/missing data, and they can be grouped into two categories: (i) drop the missing cases or (ii) use available information to estimate the missing values. Below, we report both descriptive and correlation findings produced with the data, and we adopt different approaches for each. For the correlational analyses, we adopt the multiple imputation approach pioneered by Little and Rubin (1987).[10] As useful as that approach is when one is studying relationships between variables, it cannot be implemented when one is studying univariate descriptives. For the descriptives presented below, we replace the incomplete/missing values with the modal value for that variable in the country for all years when the data were not incomplete/missing. This approach introduces measurement error relative to having complete/nonmissing data, but we believe it introduces less error than would be introduced if we dropped those cases. Other researchers are free to implement alternative solutions when they use the country-year data.

Validity

Although the Ill-Treatment and Torture country-year data on government torture allegations permit the disaggregation of rights violations by state agency and victim type, they are not the first cross-national data on state torture practices. We use two other cross-national measures of government torture, one from Cingranelli and Richards (2010) and one from Hathaway (2002), to establish the convergent validity of the CY LoT measures. Convergent validity describes the extent to which concepts or measures that should be related to one another in theory are actually related to one another in practice. Accordingly, Table 1 shows the correlation between ITT country-year LoT measures, the Cingranelli and Richards (2010) freedom from torture measure (in the second column), and the Hathaway (2002) torture measure (in the third column).[11]

The ITT LoT variables presented in Column 1 include measures of allegations across all state agencies (CY LoT), within the individual agencies described above (CYAoC), and against the individual victim types described above (CYVT). The Cingranelli and Richards (2010) measure of freedom from torture ranges from zero to two, with higher values indicating a greater respect for the right to freedom from torture. Accordingly, we expect each of the LoT measures to be *negatively* correlated with the Cingranelli and Richards (2010) freedom from torture measure. The Hathaway (2002) measure of torture, on the other hand, ranges from zero to five; higher values indicate higher levels of torture. We expect each of our LoT measures to be *positively* correlated with Hathaway (2002).[12]

[10]More specifically, we implemented multiple imputation on our LoT measures and performed econometric analyses on the imputed data using the Stata 11 suite of `mi` commands.

[11]Note that the sample size of the correlational analyses presented in Table 1 differs by measure. This is because the temporal domain of ITT data ranges from 1995 to 2005, whereas the temporal domain of Cingranelli and Richards (2010) data and Hathaway (2002) data ranges from 1981 to 2009 and 1985 to 1999, respectively.

[12]We do not assess convergent validity using another popular cross-national measure of human rights, the Political Terror Scale, because it is an indicator of physical integrity violations writ large rather than government torture specifically.

Table 1. Correlation of ITT CY LoT Measures with CIRI and Hathaway Measures

	CIRI Freedom from Torture	*Hathaway Torture*
CY LoT	−0.42***	0.46***
CYAoC LoT		
Unnamed	−0.38***	0.43***
Police	−0.24***	0.37***
Prison	−0.25***	0.41***
Military	−0.33***	0.43***
Intelligence	−0.04***	0.14***
Immigration	−0.07	0.08
Paramilitary	−0.14***	0.20***
CYVT LoT		
Unnamed	−0.39***	0.46***
Criminal	−0.28***	0.42***
Dissident	−0.30***	0.40***
Marginalized	−0.18***	0.34***
State Agent	−0.08***	0.20***
N	1576	698

CYAoC, Country-Year, Agency of Control; CYVT, Country-Year, Victim Type.)
(*Notes.* *$p < 0.10$; **$p < 0.05$; ***$p < 0.01$; (two-tailed).

As anticipated, each of the LoT measures shown in Column 1 is negatively correlated with the Cingranelli and Richards (2010) freedom from torture measure and positively correlated with data from Hathaway (2002), which measures violations.[13] Across the disaggregated ITT data, the correlations range in absolute value from 0.07 to 0.46, thus exhibiting almost no correlation through moderate levels of correlation. Not surprisingly, given the different scales used by Cingranelli and Richards (2010) (that is, three values) and Hathaway (2002) (that is, five values), the source for the ITT scale, the absolute value of the correlations is higher between ITT and Hathaway (2002) in all but one case. We take two points from this exercise: The absence of correlations with the wrong sign provides evidence of convergent validity, and the low to moderate size of the correlations demonstrates that the ITT data contain rather different information than that provided in other data.[14]

What best explains the low to moderate correlations? Our data were created to code AI allegations, whereas Cingranelli and Richards (2010) and the Hathaway (2002) project use those allegations to code state behavior. Yet despite this conceptual distinction, one might anticipate higher associations given that all three projects perform content analysis on documents that allege violations. But consider differences in the coding practices across the three projects. The Cingranelli and Richards (2010) project assigns its torture variable values based on the coder's judgment about all of the torture allegations against a particular state in a particular country-year (Cingranelli and Richards 2010). The Hathaway (2002) data were collected using the same approach. By comparison, ITT only codes allegations in our country-year data when AI claims that abuse occurs throughout the country over the course of the year.[15] Neither the Cingranelli and Richards (2010) nor the Hathaway (2002) data distinguish between allegations with a broad spatial–temporal domain matching the country-year and those

[13]The Cingranelli and Richards (2010) measure is correlated with the Hathaway (2002) measure at -0.6658.

[14]Differences in convergent validity may occur because the sources from which allegations are drawn differ across these data sets. Hathaway (2002) data are coded from US State Department reports; although Cingranelli and Richards (2010) data are sourced from AI when there is contention between AI and the US State Department, the data are primarily generated using US State Department reports.

[15]ITT codes AI allegations of ill-treatment and torture that are limited temporally or spatially, but does so as event data.

with a limited spatial and/or temporal domain. Which data are more useful, then, depends upon the research question and the goals of a particular project.

With that background, we briefly examine some of the specific correlations presented in Table 1. Consider first the country-year LoT measure, which reports the highest value of alleged country-wide torture over all agencies and victim types in a year. Since this is the ITT CY variable most similar to the Cingranelli and Richards (2010) and Hathaway (2002) variables, it is not surprising that it produces the largest (negative) correlation with the Cingranelli and Richards (2010) freedom from torture measure (−0.42), and the highest (positive) correlation with the Hathaway (2002) measure of torture (0.46). Among the various ITT CYAoC variables, Unnamed is more highly correlated with both the Cingranelli and Richards (2010) measure (−0.38) and the Hathaway (2002) measure (0.43). In contrast, the CYAoC measure least correlated with the Cingranelli and Richards (2010) measure is Intelligence (−0.04), while the CYAoC measure least correlated with the Hathaway (2002) measure is Immigration (0.08). These low values suggest that there is potentially interesting variation to be explored across government agencies.

Similar patterns exist when the data are divided into AI allegations of torture against different groups of victims. The CYVT Unnamed LoT measure exhibits the largest correlation with both the CIRI (−0.39) and Hathaway (0.46) variables. ITT LoT data on torture allegations against Criminals and Dissidents are correlated with the country-year measures generated by Cingranelli and Richards (2010) (−0.28, −0.30) and Hathaway (2002) (0.42, 0.40), respectively. The CYVT State-Agent is the least correlated disaggregated ITT measure of torture against specific types of victims across both the CIRI freedom from torture (−0.08) and Hathaway torture measures (0.20). These low values suggest to us that there is potentially interesting variation to be explored. We show below that when one moves from global to regional (and even country) comparisons, one continues to find differences that suggest the Ill-Treatment and Torture data may yield useful insights.

Reliability

To evaluate the reliability of the ITT coding rules, we conducted a series of inter-coder reliability checks during the year in which the content analysis was performed. Conrad and Moore (2011:15) provide a brief discussion of our coder recruitment and training process, as well as the analysis of inter-coder reliability. We plan to later release a detailed study of the results of our inter-coder reliability checks across all CY and SA measures. To assess the reliability of the measures contributing to the CY data (that is, LoT, AoC, Victim Type, and Restricted Access), we report both the overall proportion of agreement measure (Fleiss 1971, 1981):

$$P_{OA} = \frac{\sum_{j=1}^{C}\sum_{k=1}^{K} n_{jk}n_{jk-1}}{\sum_{k=1}^{K} n_{jk}n_{jk-1}}$$

and Krippendorff's alpha, αK (2004):[16]

$$\alpha_K = \frac{P_{OA} - P_e}{1 - P_e}$$

Table 2 shows that the four variables contributing to the CY data have inter-coder reliability scores ranging from 0.805 (Restricted Access) to 0.958 (LoT) for Krippendorff's α and from 0.902 to 0.979 for the proportion of agreement.

[16] P_e is the expected proportion of correct classification by all coders if the values were assigned randomly. In the ITT coding scheme, coders assigned binary measures for all of the variables that make up the variables in the data set (see Conrad and Moore 2010a for details), so $P_e = 0.5$.

TABLE 2. Intercoder Reliability Scores

	P_{OA}	α_K
Level of Torture (LoT)	0.979	0.958
Agency of Control (AoC)	0.971	0.942
Victim Type (VT)	0.939	0.888
Restricted Access	0.902	0.805

(*Notes.* P_{OA}: proportion of overall agreement; α_K: Krippendorff's α.)

Univariate Patterns of AI's CY Allegations

Frequencies

Table 3 displays frequencies of LoT in the CY, CYAoC, and CYVT data described above.[17] The first row of Table 3 records the distribution of LoT aggregated at the country-year: This is the highest level of country-wide annual torture alleged by AI for all AoCs and VTs in a particular country during a given year. In only 29% (486) of country-years did AI not issue at least one allegation of country-wide violation of the UN Convention Against Torture (CAT). Note that this is not a single alleged act of abuse: The ITT country-year data do not record AI's allegations of a single act of abuse.[18] These data only code allegations of abuse occurring throughout the country over the course of an entire year. In 37% of country-years, AI alleged either Widespread or Systematic abuse of the rights enshrined in the CAT (220 and 397, respectively). AI alleged either Several or Routine use of ill-treatment and torture in another 25% of country-years (212 and 206, respectively), and Infrequent abuse in only 9% (151) of country-years. This distribution is consistent with the global pattern described in Cingranelli and Richards (1999b:522), who report that rights outlined in the CAT are the most widely contravened in the world.[19]

What distribution emerges when one turns attention to the government agency AI claims is responsible for victimization at the global level? It is immediately apparent that Unnamed and Police are the government agencies that AI most commonly names and shames: There are 811 country-years for which AI issued an allegation without identifying the government agent (48%) and 746 country-years in which AI called out the Police for violations (46%). Prisons are named and shamed in 31% of allegations (503) and the Military in 28% of country-years (469). By comparison, Intelligence agencies, agencies responsible for detaining Immigrants, and Paramilitary organizations are named and shamed by AI in only 5–6% each of the country-years (85, 90, and 99 country-years, respectively). Finally, Police exhibits a fairly uniform spread across the levels: Several through Systematic all have between 128 and 172 country-years with allegations. For the other AoCs, the spread across values other than No Allegation is roughly similar to those seen above for CY LoT.[20]

In addition to observing the frequencies shown in Table 3, we consider graphical displays of the spread of LoT values for each of the four data sets described above. Box and whisker plots visually depict the central tendency and dispersion of the values of a variable. In the plots below, the median value is depicted as a

[17]Although there are 2,002 country-years covered by each of these data sets, the value in the Total column of Table 3 is 1,672. When states collapse or are occupied by foreign powers, those country-years are assigned a missing value code (Conrad and Moore 2011:8–9) and dropped from the data.

[18]As noted above, those allegations are coded in a distinct data set released in 2012.

[19]Over the years 1980–2008, the CIRI data record that 79% of the country-years exhibit either little or only some respect for the right to freedom from torture.

[20]Note that these patterns are true of the country-year allegations, which are not limited in space or time. If spatially or temporally limited violations were included, the frequencies may differ.

Table 3. AI Allegation Frequency of Level of Torture (LoT) by Measure

	None	*Infrequent*	*Several*	*Routinely*	*Widespread*	*Systematic*	*N*
CY LoT	486	151	212	206	220	397	1,672
CYAoC LoT							
Unnamed	861	167	133	131	159	221	1,672
Police	926	138	172	171	137	128	1,672
Prison	1,161	137	120	70	75	109	1,672
Military	1,203	103	52	79	92	143	1,672
Intelligence	1,587	49	6	7	1	22	1,672
Immigration	1,582	54	16	11	4	5	1,672
Paramilitary	1,573	52	6	11	11	19	1,672
CYAoC LoT							
Unnamed	701	205	152	153	211	250	1,672
Criminal	971	159	130	132	129	151	1,672
Dissident	1,227	105	104	78	48	110	1,672
Marginalized	1,114	120	113	123	75	127	1,672
State Agent	1,589	50	4	11	4	14	1,672

Notes. CY, Country-Year; CYAoC, Country-Year, Agency of Control; CYVT, Country-Year, Victim Type.

solid white horizontal line,[21] and a shaded rectangle covers the range of the 25th through the 75th percentiles of the variables' values. A thin vertical line stretches from the edge of the rectangle to encompass what are known as the upper and lower adjacent values. Technically, these values are the 75th (25th) percentile plus (minus) 1.5 times the mean value. Dots are used to depict any values that lie outside of the range of the lower and upper adjacent values. Figure 1 further amplifies the frequencies presented in Table 3: Both Unnamed and Prison have the largest spreads with the zero to 75th percentile ranging from zero to three. The mean of Unnamed is one, but the mean of each of the named AoCs is zero. Both Prison and Military reach the 75th percentile at one and have an upper adjacent value of two.

How should these data be interpreted? More specifically, can one claim that AI focuses its advocacy resources on abuses committed in prisons and by police and military? Or is it better to argue that these data are representative of the "true" distribution of state torture, effectively assuming that AI invests its monitoring, investigative, and reporting resources equally wherever abuse occurs? As noted above, we recommend the former approach and encourage scholars to research the extent to which AI (and other INGOs) invest their monitoring, investigative, and reporting resources in proportion to violations.[22] We have no reason to believe that AI does not invest its resources in proportion to actual state violations, but it strikes us as prudent to study that issue rather than assume it to be so.[23]

Returning to Table 3, VT LoT variables exhibit a similar emphasis on a few types in particular: The Unnamed and Criminal categories are more common than the Dissident and Marginalized victim groups. The State Agent group is especially rare. Figure 2 displays the same information graphically, showing that 51% of Unnamed country-years involved an allegation (691), 36% of the Criminal country-years contain an allegation (522), and 22, 27, and 2% of Dissidents, Marginalized Populations, and State Agents involved allegations, respectively (328, 401, and 30, respectively).

[21]If the horizontal line does not appear, the median value is zero.

[22]As noted above, this distribution across agencies is only evocative of country-wide, year-long AI allegations. Temporally and/or spatially limited violations are not included in the CY data.

[23]Please refer to the discussions in Gourevitch and Lake (2011), Hill *et al.* (2012).

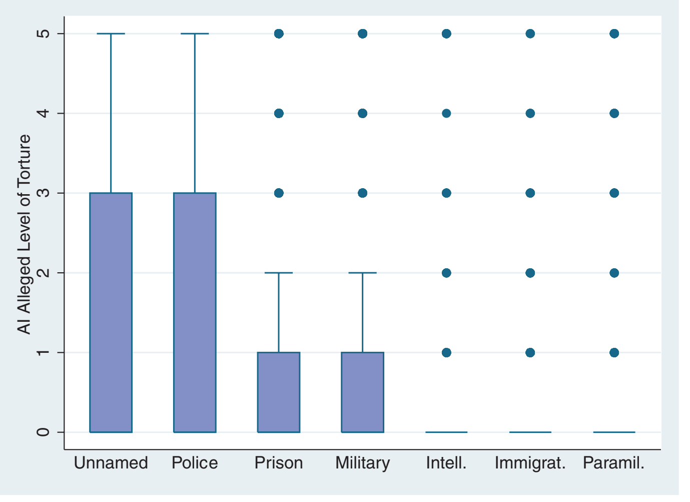

Fig. 1. AI Alleged Level of Torture by Agency of Control, 1995–2005 (*Notes*. Level of Torture: 0 = No Allegation; 1 = Infrequent; 2 = Some; 3 = Frequent; 4 = Widespread; 5 = Systematic. Agency of Control: Intell. = Intelligence; Immigrat. = Immigration; Paramil. = Paramilitary.)

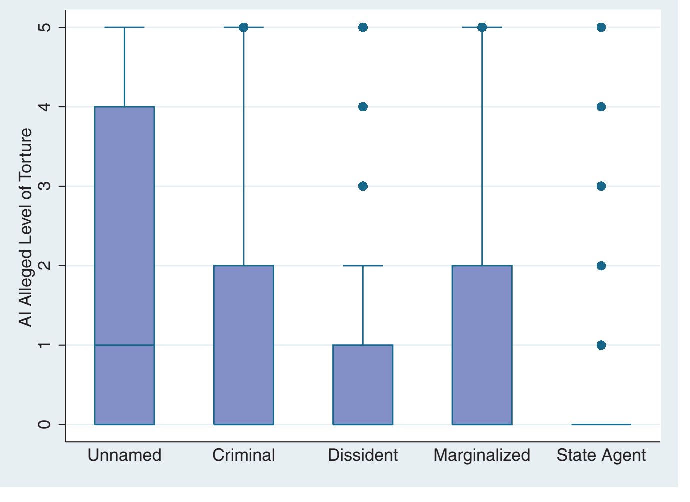

Fig. 2. AI Alleged Level of Torture by Victim Type, 1995–2005 (*Notes*. Level of Torture: 0 = No Allegation; 1 = Infrequent; 2 = Some; 3 = Frequent; 4 = Widespread; 5 = Systematic.)

Lastly, we examine the extent to which AI allegation patterns are associated with one another across AoC and VT. The Ill-Treatment and Torture CY data permit one to explore such questions using Country-Year, Agency of Control Victim Type (CYAVT) data. Figure 3 displays a heat map of the Goodman and Kruskal γ statistic across the AoC and LoT values for each country-year from 1995 to 2005. The Goodman and Kruskal γ is a measure of association for ordinal level variables that summarizes the frequencies one can observe in a contingency table. It ranges from -1 to 1, with $|1|$ indicating a perfect association and 0 indicating an absence of any association. Consequently, it allows us to examine the extent to which a given country-year's alleged LoT against a given AoC has the same value as the alleged LoT against a given VT.

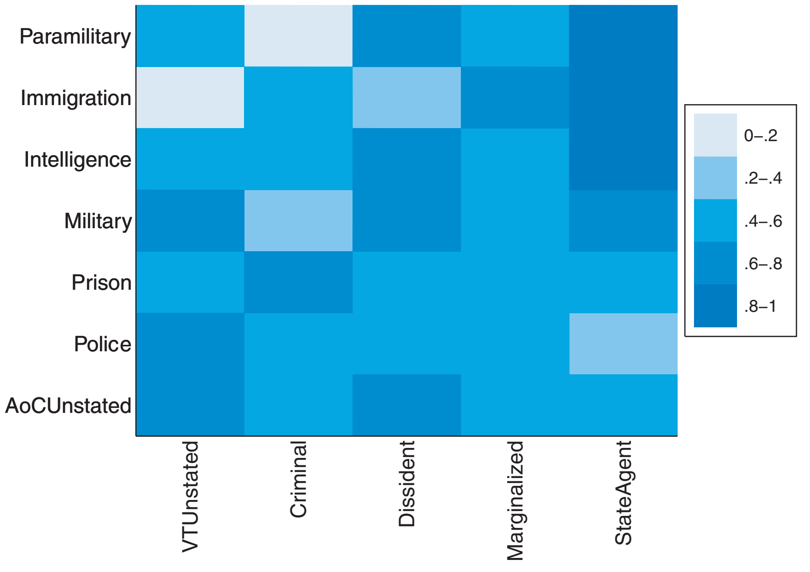

FIG. 3. AI Alleged Level of Torture, 1995–2005 (*Note.* Goodman and Kruskal's γ, which ranges from -1 to 1, is depicted in the cells.)

According to Table 3, State Agent is the VT least frequently identified in AI's allegations. Figure 3 indicates that the alleged LoT against State Agents is strongly associated with its alleged LoT against Intelligence, Immigration Detention, and Paramilitary AoCs, rather strongly associated with the alleged LoT against the Military, moderately associated with the alleged LoT against Prisons and Unstated AoCs, and only weakly associated with Police allegations. We can also observe that AI allegations about the abuse of Marginalized Populations are rather strongly associated with its LoT allegations about Immigration and Detention AoCs, and moderately associated with alleged LoT by other AoCs. Alleged LoT against Dissidents is most strongly associated with alleged LoT against Military, Intelligence, and Unstated AoCs, while the alleged LoT against Criminals is most strongly associated with the alleged LoT in Prisons, only weakly associated with the alleged LoT of the Military, and not at all associated with the alleged LoT of Paramilitary groups. These patterns demonstrate one way in which the ITT CY data can be displayed to reveal potentially interesting patterns heretofore unexplored.

Using ITT CY Data for Multivariate Analyses

In this section, we provide an illustration for how the ITT data can be used to study a state's (lack of) respect for the United Nations Convention Against Torture (CAT). Given ITT's explicit focus on AI allegations rather than state behavior, it may seem odd that we choose to conduct multivariate analyses to draw inferences about state torture rather than AI "naming and shaming" activity. But ITT data can be used to draw inferences about violations. To use the ITT data to study state behavior, researchers should develop a model of the process that links AI allegations to the unobservable (that is, latent) variable of interest: state abuse of (or respect for) the CAT. To illustrate how this might be done, we present illustrative statistical analyses in which we estimate the impact of commonly used covariates of state repression on government torture across state agencies and victim types. We discuss the steps researchers should take to use the Ill-Treatment and Torture country-year data as evocative of the "true," unobserved level of state abuse (c.f. Bollen 1986; Spirer 1990; Rejali 2007). Before discussing how one might control for the process that produces AI allegations, we first briefly introduce the covariates that we expect to affect state torture.

Although existing research finds that democratic institutions decrease the incidence and intensity of human rights violations (for example, Poe and Tate 1994; Poe *et al.* 1999; Davenport 2007), the effect of democracy and other common predictors of state repression may vary across the institutions responsible for—and the victims of—human rights abuse. In part because of the highly aggregated nature of existing cross-national data on state repression, there has been no research to date on the effect of democratic institutions on torture by heterogeneous agencies and against different types of victims. The ITT country-year data make it possible to explore whether patterns of ill-treatment and torture found at the aggregate level using Cingranelli and Richards (2010) or Hathaway (2002) data hold broadly across agencies and victim types. In what follows, we estimate coefficients for the "usual suspect" covariates of state repression using ordered probit regression models and the ITT country-year data—by agency and by victim type—as our dependent variables.

Our main independent variables are those found to be significant predictors of state repression in the seminal study conducted by Poe and Tate (1994): Democracy, Interstate War, Civil War, Country Wealth, and Country Population.[24] We use a minimalist, binary measure of democracy from the Democracy-Dictatorship (DD) data set (Cheibub *et al.* 2010).[25] To measure both international and civil war, we use the UCDP/PRIO Armed Conflict Dataset from Themnér and Wallensteen (2011). Because country wealth and population often have a statistically significant impact on state repression (for example, Poe and Tate 1994; Davenport 1995), we include measures of GDP per capita and country population from the World Development Indicators (WDI) in each of the models below. Following Poe and Tate (1994), we also included a lagged dependent variable in each of our models to account for temporal dependence.

As argued above, to use the country-year LoT variables in a statistical analysis of state behavior requires, at a minimum, the specification of control variables that influence the likelihood of an Amnesty International allegation. Researchers need to include measures that capture the likelihood that AI would observe a violation of human rights and then report it.[26] To that end, we include in our analyses measures that we expect to be related to the likelihood of AI "naming and shaming" states for violations of human rights. Because it is an information INGO, AI relies upon access to domestic sources for its reporting. In the extreme, AI would have no people with whom to interact to generate allegations against a particular country. Yet AI *does* generate allegations against most every

[24]We recognize that we have perhaps omitted from our analyses other measures commonly found to be associated with state repression. As such, our results are intended as illustrative: We leave for future work the analysis of fully specified models. This effort is solely intended to stimulate readers' thinking about how the CY data might be used.

[25]The measure classifies countries as democracies or dictatorships based on whether or not they hold free executive and legislative elections. In order for a country to be coded as a democracy, (i) the chief executive and the legislature must be selected through popular election, (ii) there must be ex-ante uncertainty about who will win the election, (iii) the electoral winner must take office following the election, and (iv) elections must occur at regular intervals. Because there is debate in the literature about the appropriate way to conceptualize and measure democracy (for example, Munck and Verkuilen 2002), particularly with respect to its relationship to respect for human rights (for example, Richards 1999; Richards and Gelleny 2007), we also measure democracy using a continuous indicator from Polity IV (Marshall and Jaggers 2009) that ranges from -10 to 10. Those results do not differ substantively from the results reported here.

[26]A more sophisticated approach would specify a model of the "naming and shaming" process, perhaps using selection models or latent measurement models (for example, Treier and Jackman 2008). We hope that the country-year data will attract attention within the political methodology community to more directly model these processes.

country in the world, thus demonstrating its effectiveness in preventing that extreme case from being realized. Between that extreme and the opposite situation of full and unimpeded access lies considerable ground. AI refers to countries with limited access as “closed countries” (Hopgood 2006:100) and comments in its reports when it lacks access in a particular country. Countries in which AI comments on lack of access may be more likely to draw AI attention because the INGO is likely to believe that countries that abuse their citizens’ rights have greater incentive to limit access. We thus include in the following analyses the ITT measure of restricted AI access. Further, given AI’s grass roots structure, its level of information about a country is also influenced by the size of its membership in that country (Hopgood 2006:65–71, 73–104, 204–223). As a proxy indicator for membership, we utilize a variable that codes whether AI has a National Office in a given country collected by Krüger (2008).

Table 4 reports coefficient estimates from ordered probit models where the independent variables are listed in Column 1, and the CYAoC LoT measures shown across the rows are the relevant dependent variables.[27] Democracy has a negative sign for all AoCs except Police, where it has no impact. This is particularly interesting in light of Rejali’s (2007) history of the rise of clean/stealth, or non-scarring, torture techniques; methods of interrogation leaving no permanent marks were pioneered by police agencies in Britain, France, and the United States after courts began to reject confessions from accused whose bodies were scarred (Rejali 2007:4, 13, 40, 70–79). He further argues that the success of Amnesty International’s 1973 *Report on Torture* stimulated a global monitoring regime that has stimulated all states to abandon scarring torture techniques in favor of clean/stealth techniques (Rejali 2007:8–15, 39–44, 105–117).

The results on Interstate and Civil War are striking. Interstate war has no effect on torture by any state agency except the Military, where it has a positive and highly significant effect. Civil war, on the other hand, is positively and significantly related to torture by Unnamed agencies, as well as the Military. Civil war is *negatively* and significantly related to Prison Torture. Country wealth has a negative sign for all AoCs except for Immigration Detention, where it is positive. This result is quite interesting as it suggests that the wealthier a country, the more likely it is to mistreat people held in its Immigration Detention centers. Since wealthier countries are both more likely to attract migrants *and* have state resources to arrest and incarcerate migrants who lack state approval to be there, this appears to be a reasonable finding. Country population has a positive impact on all AoCs except the Military and Immigration Detention, which are non-significant.

With regard to the covariates included to capture the process by which AI generates allegations, Table 4 shows that the presence of a National AI Office has a positive impact upon the level of alleged torture AI reports for all AoCs. AI Lack of Access also has a positive impact upon AI’s alleged level of AoC abuse, with the exception of Immigration Detention. With the exception of Immigration Detention, these results suggest that countries that restrict access attract greater criticism from AI.

Coefficient estimates from ordered probit models where the dependent variables are CYVT LoT measures are provided in Table 5. Democracy has a negative and highly significant effect on torture suffered by all groups except Marginalized Populations. This suggests that holding elections does not afford a measure

[27]Because listwise deletion due to missing values on our independent variables might bias the results, we also used the Stata 11 suite of `mi` commands to multiply impute the missing data. In general, there are no changes in sign from what we report below, though higher levels of significance are achieved across the board. Furthermore, the estimated coefficients for Democracy get a bit larger in size with the MI data, and the coefficient estimates for National AI Office and Restricted Access get a bit smaller. Those results will be made available in our replication data set upon publication.

Table 4. Determinants of Torture Allegations Across Government Agencies

	Unnamed	*Police*	*Prison*	*Military*	*Immigration*
Democracy$_t$	−0.355***	0.026	−0.250***	−0.165*	−0.338*
	(0.078)	(0.075)	(0.088)	(0.089)	(0.184)
Interstate War$_t$	0.213	−0.106	−0.012	0.719**	0.005
	(0.361)	(0.075)	(0.433)	(0.354)	(0.670)
Civil War$_t$	0.446***	0.066	−0.411**	0.338**	0.047
	(0.165)	(0.075)	(0.199)	(0.168)	(0.367)
Country Wealth$_t$	−1.52e-05**	−1.12e-05***	−1.01e-05***	−2.04e-05***	1.93e-06***
	(4.60e-06)	(4.13e-06)	(4.82e-06)	(5.77e-06)	(6.90e-06)
Country Population$_t$	1.44e-09***	7.65e-10***	3.91e-10*	9.81e-12	3.64e-10
	(3.25e-10)	(2.32e-10)	(2.41e-10)	(3.00e-10)	(4.00e-10)
Dependent Variable$_{t-1}$	0.406***	0.387***	0.480***	0.441***	0.644***
	(0.021)	(0.075)	(0.295)	(0.023)	(0.075)
National AI Office$_t$	0.177**	0.141*	0.299***	0.262***	0.353**
	(0.085)	(0.082)	(0.093)	(0.095)	(0.183)
AI Lack of Access$_t$	0.446***	0.243*	0.471***	0.268*	0.093
	(0.141)	(0.141)	(0.144)	(0.160)	(0.270)
N	1,239	1,239	1,239	1,239	1,239

(*Notes.* *$p < 0.10$ **$p < 0.05$; ***$p < 0.01$; (two-tailed). Estimates generated using Ordered Probit on data where negative values are set to the panel mean. Results on cutpoints are omitted above, but are available in our replication files. Intelligence and Paramilitary not shown due to fully determined observations and questionable SEs.)

TABLE 5. Determinants of Torture Allegations Across Victim Types

	Unnamed	*Criminal*	*Dissident*	*Marginalized*	*State Agent*
Democracy$_t$	−0.186***	−0.186**	−0.363***	−0.105	−0.234*
	(0.073)	(0.078)	(0.090)	(0.085)	(0.154)
Interstate War$_t$	0.199	0.036	−0.372	0.095	1.252***
	(0.333)	(0.382)	(0.538)	(0.073)	(0.434)
Civil War$_t$	0.468***	−0.067	0.302*	0.262*	−0.003
	(0.166)	(0.170)	(0.170)	(0.168)	(0.312)
Country Wealth$_t$	−1.67e-05***	−4.69e-06	−2.39e-06***	9.48e-07	−6.59e-06
	(4.13e-06)	(4.16e-06)	(6.16e-06)	(4.31e-06)	(3.88e-10)
Country Population$_t$	1.23e-09***	5.59e-10**	5.59e-10**	6.83e-10***	1.81e-10
	(3.01e-10)	(2.30e-10)	(2.39e-10)	(2.36e-10)	(3.34e-10)
Dependent Variable$_{t-1}$	0.396***	0.369***	0.369***	0.422***	0.612***
	(0.020)	(0.020)	(0.025)	(0.023)	(0.059)
National AI Office$_t$	0.173**	0.292***	0.273***	0.227**	−0.027
	(0.079)	(0.084)	(0.096)	(0.091)	(0.175)
AI Lack of Access$_t$	0.354***	0.321**	0.227*	0.424***	0.263
	(0.139)	(0.140)	(0.153)	(0.142)	(0.256)
N	1,239	1,239	1,239	1,239	1,239

(*Notes.* *$p < 0.10$ **$p < 0.05$; ***$p < 0.01$; (two-tailed). Estimates generated using Ordered Probit on data where negative values are set to the panel mean. Results on cutpoints omitted above, but are available in our replication files.)

of greater protection to marginalized minority groups, which is consistent with the concerns advocates of liberal democracy raise about majoritarian rule.[28] That democracy is negatively associated with Criminals as victims yet not Police as perpetrators is intriguing, though we hasten to point out that it is not an issue best explored using CY data; these data record the highest level of abuse alleged by AI to occur throughout the country in a given year. As such, one cannot use these data to directly link the AoC and VT LoT values. For example, consider a victim abused in a police station or a prison and identified as a criminal. Although values for AoC and VT cannot be linked across observations using the country-year data, researchers will be able to make direct connections using the specific allegation data. The CY data are certainly suggestive, but they do not permit valid inference of that kind.

Interstate War has little effect on torture across Victim Types. The exception is torture against State Agents; Interstate war is positively and significantly associated with State Agent torture. Civil war is positively and significantly related to torture against Unnamed victims, Dissidents, and Marginalized Individuals. Finally, the results on Country Wealth and Country Population are broadly consistent with those reported above across agencies of control, as well as the literature on state repression more generally. Country Wealth is only associated with statistically significant decreased torture against Unnamed victims and Dissidents, while Country Population is positively associated with torture against all Victim Types, with the exception of State Agents.

National AI Office and AI Lack of Access have a positive and statistically significant impact upon the LoT alleged by AI for all victim types except state agents (where we are unable to reject the null of no impact for either). These results suggest that both the presence of a national office and restrictions on AI access to prisoners *increase* criticism of a country's respect for the Convention Against Torture.

Although coefficient estimates from ordered probit models are suggestive, they only provide information on the effect of the independent variables on the upper and lower end of our LoT scale (that is, when LoT = 0 and LoT = 5). To estimate the effect of our independent variables across the full range of countries' use of torture, we used Clarify (Tomz *et al.* 2003) to calculate the change in the predicted probability of AI torture allegations in each category of LoT (that is, zero to five) based on a shift from nondemocracy to democracy (with all other independent variables set to their means).[29] Table 6 shows the point estimates of these results and their relevant confidence intervals across the agencies and victim types coded by ITT.

Although this is an illustrative exercise and the results should be interpreted with caution, some interesting patterns beyond those available in the coefficient estimates emerge. The values in each cell can be interpreted (with due caution) as indicating the probability that the average state is in compliance with (the None column) or at a given level of violation of (the other columns) the CAT given a change from Autocracy to a Democracy. Scanning the table shows that a change from Autocracy to Democracy has a small, though non-trivial, impact. With that context, the None column can be interpreted as reporting the change in the predicted probability that the state is in compliance with the CAT given that it is a Democracy. For Unnamed and Prison as AoC and Unnamed, Criminal, and Dissident as VTs, the probability increases from 0.054 to 0.110. Put plainly, states that hold free and fair elections are more likely than others to have Prisons that respect the CAT; respect for the CAT under these institutional

[28]See, for example, Federalist #10 (Hamilton *et al.* 2009).

[29]The exceptions are Restricted AI Access, Interstate War, and Civil War, which we hold constant at a value of zero.

Table 6. Change in Predicted Probability of Torture Allegations (Nondemocracy to Democracy)

	None	*Infrequent*	*Several*	*Routinely*	*Widespread*	*Systematic*
Unnamed AoC	0.110*	−0.029*	−0.026*	−0.024*	−0.021*	−0.010*
	[0.063, 0.154]	[−0.043, −0.016]	[−0.038, −0.014]	[−0.035, −0.013]	[−0.031, −0.012]	[−0.015, −0.005]
Police	−0.008	0.001	0.002	0.002	0.001	0.001
	[−0.057, 0.039]	[−0.008, 0.010]	[−0.010, 0.016]	[−0.011, 0.018]	[−0.006, 0.010]	[−0.003, 0.005]
Prison	0.056*	−0.021*	−0.019*	−0.007*	−0.006*	−0.003*
	[0.019, 0.096]	[−0.038, −0.006]	[−0.032, −0.006]	[−0.013, −0.012]	[−0.011, −0.012]	[−0.006, −0.001]
Military	0.034	−0.010	−0.006*	−0.001	−0.006	−0.004
	[−0.002, 0.070]	[−0.021, 0.001]	[−0.014, 0.001]	[0.016, −0.006]	[−0.012, 0.003]	[−0.009, −0.003]
Immigration	0.011	−0.008	−0.002	−0.001	−0.000	−0.000
	[−0.001, 0.023]	[−0.017, 0.001]	[−0.006, 0.000]	[−0.003, 0.000]	[−0.001, 0.000]	[−0.001, 0.000]
Unnamed VT	0.066*	−0.016*	−0.015*	−0.014*	−0.015*	−0.001*
	[0.016, 0.115]	[−0.028, −0.003]	[−0.028, −0.004]	[−0.026, −0.003]	[−0.026, −0.003]	[−0.011, −0.001]
Criminal	0.054*	−0.014*	−0.013*	−0.011*	−0.010*	−0.001*
	[0.009, 0.097]	[−0.026, −0.002]	[−0.024, −0.002]	[−0.021, −0.002]	[−0.018, −0.020]	[−0.010, −0.001]
Dissident	0.084*	−0.022*	−0.026*	−0.017*	−0.009*	−0.011*
	[0.044, 0.124]	[−0.034, −0.011]	[−0.040, −0.013]	[−0.026, −0.009]	[−0.013, −0.004]	[−0.018, −0.005]
Marginalized	0.023	−0.007	−0.007	−0.006	−0.002	−0.002
	[−0.015, 0.058]	[−0.018, 0.004]	[−0.018, 0.004]	[−0.014, 0.004]	[−0.005, 0.001]	[−0.004, 0.001]
State Agent	0.014	−0.008	−0.001	−0.003	−0.001	−0.002
	[−0.006, 0.035]	[−0.020, 0.003]	[−0.002, 0.000]	[−0.007, 0.001]	[−0.003, 0.000]	[−0.006, 0.001]

(*Notes.* * Significant at 95% confidence. Change in the predicted probability of torture allegations by agency and victim type. Restricted access, interstate war, and civil war set to 0; all other independent variables set to their in sample means. Intelligence and Paramilitary not shown due to fully determined observations and questionable SEs.)

arrangements is also expected to extend to Criminals and Dissidents. This tentative result represents an interesting stylized fact that we did not know prior to having access to the country-year data. Why might this be so? Similarly, why might democracies be no more likely than nondemocracies to respect the CAT with regard to Marginalized Groups or State Agents? Perhaps Publius' concern about Majoritarian rule is relevant (Hamilton *et al.* 2009); perhaps not. These are interesting examples of the types of questions that ITT data are able to illuminate.

Whether the findings for the impact of elections upon states' respect for the CAT shown in Table 6 have validity depends upon whether one believes that the model controls sufficiently for the process that leads AI to publish an allegation. The country-year data have immediate and apparent utility for studying the allegation behavior of AI. The extent to which researchers find the ITT data useful for studying states' respect for the CAT will vary considerably across (and perhaps even within) research communities, but will depend to a large extent on the validity of the statistical model. We hope that this illustrative exercise demonstrates one manner in which researchers can undertake such inquiry.

Conclusion

As with any behavior that actors prefer to hide from view, the quantification of human rights violations—including that of state torture—requires scholars to engage in careful data collection and thoughtful empirical analysis.[30] The Ill-Treatment and Torture country-year data provide a new resource for researchers interested in the behavior of information politics INGOs' and states' (lack of) respect for the United Nations Convention Against Torture, as well as other domestic laws prohibiting the practice. The data contain information about the level of ill-treatment and torture that Amnesty International alleges to have occurred throughout a given country from 1995 to 2005, as measured on a six-point ordinal scale. To recapitulate, ITT CY data differ from other cross-national, quantitative data on state torture in at least two important ways. First, the Ill-Treatment and Torture project explicitly codes AI allegations of government torture, rather than the "true" level of abuse occurring throughout a given country-year. As such, the project's conceptual focus does not concern the behavior of states insomuch as it concerns the behavior of AI. Researchers should take care to model the process by which AI generates allegations of government torture when they use the ITT country-year LoT variables as indicators of state abuse. Second, the CY data record the highest level of alleged abuse by different government Agencies of Control and across different Victim Types. This is a novel feature of the data that can be used to investigate the impact of domestic and international institutions on AI allegation behavior and/or state respect for the right to freedom from torture.

We hope the ITT CY data encourage human rights scholars to refine theories about the mechanisms that influence domestic respect for rights across state agencies and across victim types. We can imagine a variety of questions for which the ITT data are appropriate and conclude by highlighting two examples for future research. First, research on the effect of international human rights treaties has attracted considerable attention among international organization (IO) scholars in recent years (for example, Hathaway 2002; Hafner-Burton and Tsutsui 2005; Goodliffe and Hawkins 2006; Simmons 2009). These studies have largely found international law to be associated with little change in—and sometimes worse—human rights practices than would have been anticipated

[30]With regard to human rights data collection more generally, Bollen (1986) has an outstanding discussion of the issues that confront scholars interested in scientific inquiry and the study of human rights. See also Spirer (1990), Landman (2004), Landman and Carvalho (2009).

absent treaty commitment (Hathaway 2002; Hill 2010). The ITT data make it possible to explore whether patterns of ill-treatment and torture found at the aggregate level using Cingranelli and Richards (2010) or Hathaway (2002) data hold broadly across agencies and victim types. We also expect the effect of international institutions to vary across state agencies and victim types. For example, in the face of human rights treaty commitment, executives sometimes lack the ability to unilaterally limit human rights violations (Conrad and Moore 2010b). But executive (lack of) control may not be the same across agencies or across different types of victims. It is not difficult to imagine, for example, that executives may be better able to control torture within executive agencies than across prison systems. The ITT country-year data provide researchers with the ability to look at the effect of international human rights law across domestic agencies and across heterogeneous victim types.

Second, what types of victims are most protected from human rights violations by effective domestic courts? Because there are arguably few international costs associated with ratifying an international human rights treaty and failing to abide by its stipulations (Hathaway 2002; Von Stein 2005), recent literature focuses on the costs and benefits of signing human rights agreements that are associated with domestic political institutions (Vreeland 2008; Powell and Staton 2009). Domestic courts, in particular, have been found to have a notable effect on human rights outcomes. On average, effective domestic courts tend to limit human rights violations, including state torture (for example, Blasi and Cingranelli 1996; Cross 1999; Apodaca 2004; Howard and Carey 2004; Hathaway 2007). But does domestic judicial effectiveness have a greater depressing effect on violations by all government agencies, or only some of them? Do effective courts protect some victims more than others?

Finally, although the ITT country-year data increase the number of questions about allegations of state torture that can be investigated quantitatively, the unit of analysis of the data—the country-year—remains highly aggregated. Information on torture allegations is only included in the CY data when AI alleges that they occurred over the geographic span of an entire country over the course of an entire year. Clearly, many AI allegations are more limited either spatially or temporally. In early 2012, we released the ITT specific allegation (SA) data, in which the unit of analysis is the individual torture allegation rather than the country-year. At this disaggregated level of observation, the Ill-Treatment and Torture data provide information on AoC and VT for individual events, as well as information on the type of alleged torture and the state response (for example, investigation, adjudication) to each individual allegation. We hope that both data sets facilitate researchers' ability to test more precise theories about INGO behavior and state respect for human rights.

References

Abouharb, M. Rodwan. and David L. Cingranelli. (2007) *Human Rights and Structural Adjustment.* New York: Cambridge University Press.

Apodaca, Clair. (2001) Global Economic Patterns and Personal Integrity Rights after the Cold War. *International Studies Quarterly* 45: 587–602.

Apodaca, Clair. (2004) The Rule of Law and Human Rights. *Judicature* 87(6): 292–299.

Blasi, Gerald J. and David Cingranelli. (1996) Do Constitutions and Institutions Help Protect Human Rights? In *Human Rights and Developing Countries,* edited by David Cingranelli. Greenwich, CT: JAI Press pp. 223–37.

Bollen, Kenneth A. (1986) Political Rights and Political Liberties in Nations: An Evaluation of Human Rights Measures 1950 to 1984. *Human Rights Quarterly* 8(4): 567–591.

Bueno de Mesquita, Bruce, George W. Downs, Alastair Smith and Feryal M. Cherif. (2005) Thinking Inside the Box: A Closer Look at Democracy and Human Rights. *International Studies Quarterly* 49: 439–457.

Cheibub, Jose Antonio., Jennifer Gandhi and James Raymond Vreeland. (2010) Democracy and Dictatorship Revisited. *Public Choice* 143(1): 67–101.

Cingranelli, David L. and Mikhail Filippov. (2010) Electoral Rules and Incentives to Protect Human Rights. *Journal of Politics* 72(1): 243–257.

Cingranelli, David L. and David L. Richards. (2001) Measuring the Impact of Human Rights Organizations. In *NGOs and Human Rights: Promise and Performance*, edited by C.W. Welch. Philadelphia: University of Pennsylvania Press, pp. 225–237.

Cingranelli, David L. and David L. Richards. (2010) The Cingranelli-Richards (CIRI) Human Rights Dataset. URL: http://www.humanrightsdata.org

Cingranelli, David L. and David L. Richards. (1999a) Measuring the Level, Pattern and Sequence of Government Respect for Physical Integrity Rights. *International Studies Quarterly* 43(2): 407–418.

Cingranelli, David L. and David L. Richards. (1999b) Respect for Human Rights after the End of the Cold War. *Journal of Peace Research* 36(5): 511–534.

Clark, Ann Marie. (2001a) A Calendar of Abuses: Amnesty International's Campaign on Guatemala. In *NGOs and Human Rights: Promise and Performance*, edited by C.W. Welch. Philadelphia: University of Pennsylvania Press, pp. 55–71.

Clark, Ann Marie. (2001b) *Diplomacy of Conscience: Amnesty International and Changing Human Rights Norms.* Princeton: Princeton University Press.

Cmiel, Kenneth. (1999) The Emergence of Human Rights Politics in the United States. *The Journal of American History* 86(3): 1231–1250.

Conrad, Courtenay R. and Will H. Moore. (2010a) *The Ill-Treatment & Torture (ITT) Data Project Coding Rules & Norms.* Merced and Tallahassee: Ill Treatment and Torture Data Project. URL: http://www.politicalscience.uncc.edu/cconra16/UNCC/Data.html

Conrad, Courtenay Ryals and Will H. Moore. (2010b) What Stops the Torture? *American Journal of Political Science* 54(2): 459–476.

Conrad, Courtenay R. and Will H. Moore. (2011) *The Ill-Treatment & Torture (ITT) Data Project (Beta) Country-Year Data User's Guide.* Merced and Tallahassee: Ill Treatment and Torture Data Project. http://www.politicalscience.uncc.edu/cconra16/UNCC/Data.html

Cross, Frank B. (1999) The Relevance of Law in Human Rights Protection. *International Review of Law and Economics* 19(1): 87–98.

Davenport, Christian. (1995) Multi-Dimensional Threat Perception and State Repression: An Inquiry Into Why States Apply Negative Sanctions. *American Journal of Political Science* 39(3): 683–713.

Davenport, Christian. (2007) *State Repression and the Domestic Democratic Peace.* New York: Cambridge University Press.

Davenport, Christian and David Armstrong. (2004) Democracy and the Violation of Human Rights: A Statistical Analysis from 1976–1996. *American Journal of Political Science* 48(3): 538–554.

Fleiss, Joseph L. (1971) Measuring Nominal Scale Agreement Among Many Raters. *Psychological Bulletin* 76(5), 378–382.

Fleiss, Joseph L. (1981) The Measurement of Interrater Agreement. In *Statistical Methods for Rates and Proportions*, edited by J.L. Fleiss, B. Levin and M.C. Paik. New York: Wiley, pp. 212–236.

Gibney, Mark, Linda Cornett and Reed M. Wood. (2009) Political Terror Scale. Available online at: http://www.politicalterrorscale.org/. Accessed August 2009.

Goodliffe, Jay and Darren G. Hawkins. (2006) Explaining Commitment: States and the Convention against Torture. *Journal of Politics* 68(2): 358–371.

Gourevitch, Peter and David A. Lake. (2011) Credible Ethical Action. In *TBA*, edited by Peter Gourevitch and David A. Lake. Toronto: University of Toronto Press.

Hafner-Burton, Emilie M. (2005) Trading Human Rights: How Preferential Trade Agreements Influence Government Repression. *International Organization* 59(3): 593–629.

Hafner-Burton, Emilie M. and Kiyoteru Tsutsui. (2005) Human Rights in a Globalizing World: The Paradox of Empty Promises. *American Journal of Sociology* 110(5): 1373–1411.

Hamilton, Alexander, James Madison and John Jay. (2009) *The Federalist Papers.* New Haven: Yale University Press.

Hathaway, Oona A. (2002) Do Human Rights Treaties Make a Difference? *Yale Law Journal* 111: 1935–2042.

Hathaway, Oona A. (2007) Why Do Countries Commit to Human Rights Treaties? *Journal of Conflict Resolution* 51(4): 588–621.

Hill, Daniel W. (2010) Estimating the Effects of Human Rights Treaties on State Behavior. *Journal of Politics* 72(4): 1161–1174.

Hill, Daniel W., Will H. Moore and Bumba Mukherjee. (2012) Information Politics v Organizational Incentives: When Are INGO's "Naming and Shaming" Reports Biased? *International Studies Quarterly*. Forthcoming.

Hopgood, Stephen. (2006) *Keepers of the Flame: Understanding Amnesty International*. Ithaca, NY: Cornell University Press.

Howard, Robert M. and Henry F. Carey. (2004) Is an Independent Judiciary Necessary for Democracy? *Judicature* 87(6): 284–290.

Keck, Margaret E. and Katherine Sikkink. (1998) *Activists Beyond Borders: Advocacy Networks in International Politics*. Ithaca, NY: Cornell University Press.

Keith, Linda Camp. (2002) Constitutional Provisions for Individual Human Rights: Are They More Than Mere Window Dressing. *Political Research Quarterly* 55: 111–143.

Keith, Linda Camp, C. Neal Tate and Steve C. Poe. (2009) Is the Law a Mere Parchment Barrier to Human Rights Abuse? *Journal of Politics* 71(2), 644–660.

Krippendorff, Klaus. (2004) *Content Analysis: An Introduction to Its Methodology*. Thousand Oaks, CA: Sage Publications.

Krüger, Jule. (2008) Effects of International Human Rights Treaties: The Case of the UN Convention against Torture.

Landman, Todd. (2004) Measuring Human Rights: Principle, Practice and Policy. *Human Rights Quarterly* 26(4): 906–931.

Landman, Todd and Edzia Carvalho. (2009) *Measuring Human Rights*. New York: Taylor & Francis.

Little, Roderick J.A. and Donald B. Rubin. (1987) *Statistical Analysis with Missing Data*. New York: J. Wiley and Sons.

Marshall, Monty and Keith Jaggers. (2009) Polity IV Project: Political Regime Characteristics and Transitions, 1800–2007. Data Users' Manual. Center for Systemic Peace. Available online at: http://www.systemicpeace.org/inscr/p4manualv2007.pdf.

Munck, Gerardo and Jay Verkuilen. (2002) Conceptualizing and Measuring Democracy: Evaluating Alternative Indices. *Comparative Political Studies* 35(1): 5–35.

Neumayer, Eric. (2005) Do International Human Rights Treaties Improve Respect for Human Rights? *Journal of Conflict Resolution* 49(6): 925–953.

Orentlicher, Diane F. (1990) Bearing Witness: The Art and Science of Human Rights Factfinding. *Harvard Human Rights Journal* 3: 83–135.

Poe, Steven and C. Neal Tate. (1994) Repression of Personal Integrity Rights in the 1980's: A Global Analysis. *American Political Science Review* 88: 853–872.

Poe, Steven, C. Neal Tate and Linda C. Keith. (1999) Repression of the Human Right to Personal Integrity Revisited: A Global, Cross-National Study Covering the Years 1976–1993. *International Studies Quarterly* 43: 291–313.

Powell, Emilia J. and Jeffrey K. Staton. (2009) Domestic Judicial Institutions and Human Rights Treaty Violation. *International Studies Quarterly* 53(1): 149–174.

Rejali, Darius. (2007) *Torture and Democracy*. Princeton, NJ: Princeton University Press.

Richards, David L. (1999) Perilous Proxy: Human Rights and the Presence of National Elections. *Social Science Quarterly* 80(4): 648–668.

Richards, David L. and Ronald D. Gelleny. (2007) Good Things to Those Who Wait? Elections and Government Respect for Human Rights. *Journal of Peace Research* 44(4): 502–523.

Ron, James, Howard Ramos and Kathleen Rodgers. (2005) Transnational Information Politics: NGO Human Rights Reporting, 1986–2000. *International Studies Quarterly* 49(3): 557–588.

Simmons, Beth A. (2009) *Mobilizing for Human Rights: International Law in Domestic Politics*. New York, NY: Cambridge University Press.

Spirer, Herbert F. (1990) Violations of Human Rights–How Many? *American Journal of Economics and Sociology* 49(2): 199–210.

Themnér, Lotta and Peter Wallensteen. (2011) Armed Conflict, 1946–2010. *Journal of Peace Research* 48(4): 525–536.

Tomz, Michael, Jason Wittenberg and Gary King. (2003) CLARIFY: Software for Interpreting and Presenting Statistical Results. Available at: http://gking.harvard.edu/. Version 2.1.

Treier, Shawn and Simon Jackman. (2008) Democracy as a Latent Variable. *American Journal of Political Science* 52(1), 201–217.

Von Stein, Jana. (2005) Do Treaties Constrain or Screen? Selection Bias and Treaty Compliance. *American Political Science Review* 99(04), 611–622.

Vreeland, James R. (2008) Political Institutions and Human Rights: Why Dictatorships Enter into the United Nations Convention Against Torture. *International Organization* 62(1), 65–101.