

**Appendix A**  
**to**  
**Mass Atrocities and Their Prevention**  
*Journal of Economic Literature*, 59 (4)

**Charles H. Anderton**  
*College of the Holy Cross*  
canderto@holycross.edu

**Jurgen Brauer**  
*Chulalongkorn University*  
*and Augusta University*  
brauer.jurgen@gmail.com

September 20, 2021

**Contents**

	<u>Page</u>
I. Selected Cross-Country Civilian Atrocity Datasets Available for Empirical Research	1
II. Civilian Atrocity Datasets and Protocols Used for Figure 2 and to Estimate Fatalities	1
III. State-Perpetrated Mass Atrocities Overlaps with Wars	4
IV. Comparison of Mass Atrocity Fatalities with Fatalities from War and Subwar Conflict	5
References	5

## **I. Selected Cross-Country Civilian Atrocity Datasets Available for Empirical Research<sup>1</sup>**

1. Political Instability Task Force Gen-Politicide Dataset (PITF-G) primarily state-perpetrated genocides and politicides for the period 1956–2018 (Marshall, Gurr, and Harff 2019).
2. Ulfelder and Valentino (UV) (2008) state-perpetrated mass killing cases for the period 1945–2006.
3. Easterly, Gatti, and Kurlat (EKG) (2006) state-perpetrated mass killing cases for the period 1820–1998.
4. Uppsala Conflict Data Program One-Sided Violence Against Civilians Dataset (UCDP-V) for state- and nonstate actor-perpetrated civilian killings for the period 1989–2020 (Eck and Hultman 2007; Pettersson 2021).
5. Political Instability Task Force Worldwide Atrocities Dataset (PITF-W) for state-, nonstate actor-, and unidentified actor-perpetrated civilian killings for the period 1995–2019 ([eventdata.parusanalytics.com/data.dir/atrocities.html](http://eventdata.parusanalytics.com/data.dir/atrocities.html)).
6. Armed Conflict and Location Event Dataset (ACLED) for state-, nonstate actor-, and unidentified actor-perpetrated civilian killings, which covers Africa for the period 1997–2021, the Middle East for the period 2016–21, and parts of Asia for the period 2010–21 (Armed Conflict Location & Event Data Project (ACLED) ([www.acleddata.com](http://www.acleddata.com)); Raleigh, *et al.* 2010).
7. Targeted Mass Killing (TMK) Dataset for state- and nonstate actor-perpetrated civilian killings at “low” and “high” levels for the period 1946–2018. Dataset also provides coded information on perpetrator intent (<https://politicsir.cass.anu.edu.au/about-targeted-mass-killing-dataset>; Butcher, *et al.* 2020).

## **II. Civilian Atrocity Datasets and Protocols Used for Figure 2 and to Estimate Fatalities**

### **1. State-Perpetrated Mass Atrocities, 1900–44**

We identified state-perpetrated mass atrocity cases for the 1900–44 period from EGK for cases in which at least 1,000 civilians were reported killed. If EGK reported a range of civilians killed, we estimated civilian deaths as the average of the two numbers. EGK cases in which fewer than

---

<sup>1</sup> Specialized country-specific datasets on civilian killings number in the many dozens and are available in the empirical case study literature (see online Appendix B).

1,000 civilians were killed or no civilian fatalities were reported were excluded from our data analysis. EGK cases that had onset years prior to 1900 were excluded from our analysis even if the atrocities continued into 1900 or later. One EGK case (China 1920s–49) did not have a point estimate for the onset year and was coded in the scatterplot with an onset year of 1925.

## 2. State-Perpetrated Mass Atrocities, 1945–54

We identified state-perpetrated mass atrocity cases for the 1945–54 period from UV. All cases in UV involve the intentional killing of at least 1,000 noncombatants from a discrete group (Ulfelder and Valentino 2008, p. 2). For each case, UV report low and a high fatality estimates. We took the average of the two estimates to arrive at our measure of estimated fatalities.

## 3. State-Perpetrated Mass Atrocities, 1955–2006

We identified state-perpetrated mass atrocity cases for the 1955–2006 period from UV and PITF-G. When the two datasets had overlapping cases, we recorded only one case to avoid double counting. The case that was chosen was the one that provided the higher fatality estimate. For UV, the fatality estimate was determined by the averaging protocol described in II.2. For PITF-G, death magnitude indexes were provided for each year of each case. We used the midpoint of PITF-G’s death magnitude index range to determine the fatality estimate for each year. For example, the death magnitude index for Sudan for 2002 was 2.5, indicating that estimated civilian fatalities that year ranged from 8,000–16,000, which we recorded as 12,000 in estimated fatalities. A complication arose in several PITF-G cases in which a year had a death magnitude index of 5, which signified estimated fatalities of 256,000 or more. Since a midpoint estimate could not be determined for index=5 years, we turned to Harff (2003, p. 60), a precursor of the PITF-G data, to estimate fatalities for cases in which at least one year contained an index of 5 for death magnitude.<sup>2</sup>

## 4. State-Perpetrated Mass Atrocities, 2007–18

We identified state-perpetrated mass atrocity cases for the 2007–18 period from the PITF-G dataset and by applying the UV coding protocol for mass killings to the UCDP-V data on civilian killings by governments. The UV coding protocol embodies the following:

A mass killing is “any event in which the actions of state agents result in the intentional death

---

<sup>2</sup> This led to the following cases and fatality estimates from Harff (2003) included in our set of mass atrocity cases: Cambodia 1975-79, 2,700,000; Pakistan 1971, 2,000,000; and Rwanda 1994, 750,000. Also included is one case – Sudan 1983-2002, 2,012,000 – that combined Harff (2003) data for 1983-2001 and PITF-G data for 2002.

of at least 1,000 noncombatants from a discrete group in a period of sustained violence” (Ulfelder and Valentino 2008, p. 2, emphasis removed);

and

“Mass killing events were considered to have begun in the first year in which at least 100 intentional noncombatant fatalities occurred. If fewer than 100 total fatalities are recorded annually for any three consecutive years during the event, the event was considered to have ended during the first year within that three-year period in which fatalities dropped below 100 per year (even if killing continues at lower levels in later years)” (Ulfelder and Valentino 2008, p. 7).

When the cases from PITF-G and the UV protocol UCDP-V overlapped, we recorded only one case to avoid double counting. The case that was chosen was the one that provided the higher fatality estimate. For UCDP-V, the fatality estimate was the dataset’s “best” estimate rather than an average of its low and high estimates.

#### 5. Nonstate Actor-Perpetrated Mass Atrocity Cases, 1989–2018

We applied the UV coding protocol in II.4 above to the UCDP-V data on civilian killings by nonstate actors (NSAs) to identify NSA mass atrocity cases and estimated fatalities per case. The fatality estimate for each NSA case was UCDP-V’s “best” estimate rather than an average of its low and high estimates. PITF-G also provides several cases of NSA-perpetrated genocides.<sup>3</sup> When PITF-G and UCDP-V-created NSA cases overlapped, we recorded only one case in order to avoid double counting. The case that was chosen was the one that provided the higher fatality estimate.

#### 6. Unidentified Actor-Perpetrated Mass Atrocity Cases, 1995–2019

Identifying mass atrocity cases perpetrated by “unidentified actors” is fraught with problems. Civilian killings by unidentified actors cannot be tied to a unified decisionmaking process or party responsible. As such, it is difficult to claim that civilian fatalities caused by unidentified actors cohere as a case of mass atrocity. Nevertheless, the growth of atrocity acts perpetrated by

---

<sup>3</sup> The nonstate cases from PITF-G include the Islamic State in Iraq and Syria from 2014–7 and Seleka and anti-Balaka militias in the Central African Republic from 2013–7. A complication arose in the PITF-G dataset in that two of the cases indicated that genocidal activity was perpetrated by the Angolan government and National Union for the Independence of Angola (UNITA) rebels (the two cases spanned the years 1975–94 and 1998–2002). In these two cases, we assigned half of the estimated fatalities to the state (Angola) and the other half to the NSA (UNITA). This procedure generated two additional NSA mass atrocity cases in the scatterplot (i.e., UNITA 1975 and 1998) based on the PITF-G data.

unknown actors has become a palpable empirical phenomenon, so we take preliminary steps to track such cases. Two datasets that track atrocities by unknown actors are PITF-W and ACLED.

We applied the following protocols to the PITF-W dataset to generate mass atrocity cases perpetrated by unidentified actors: (1) The case must meet the criteria in the UV mass atrocity protocol noted in section II.4 above. (2) Fatalities from unidentified attacks are counted on a country-year basis. (3) Suicide attacks are generally recorded as perpetrator unknown, but the PITF-W codebook indicates that the source of suicide attacks is often known. As such, we excluded data points in which the source of attack was a suicide attack. (4) When estimated fatalities spanned a range, we used the low number of the range; when estimated fatalities was “many,” we coded it as 5 fatalities (the low end of fatalities per attack tracked by PITF-W); when estimated fatalities was “dozens,” we coded it as 24. These procedures led to nine unidentified actor mass atrocity cases from the PITF-W dataset from 1995–2019.

We applied protocols 1, 2, and first part of 4 in the previous paragraph to ACLED data. The suicide attack issue and estimated fatalities of “many” and “dozens” were not applicable to the ACLED data. These procedures led to seven unidentified actor mass atrocity cases from the ACLED dataset from 1997–2019. There were no case overlaps between the PITF-W and ACLED unidentified actor cases. In total, 16 unidentified cases are in the scatterplot in Figure 2.

### **III. State-Perpetrated Mass Atrocities Overlaps with Wars**

A substantial portion of the empirical literature on mass atrocity risks includes a lagged measure of war (inter-, intra-, and/or extra-state) or other form of armed conflict as a right side explanatory or control variable. Our data analysis, however, finds that less than half of mass atrocities occur after war onset and that a nonnegligible number of mass atrocities begin before war onset or occur outside the context of war (see article Section 4). Our coding protocols to determine the overlap, or lack of overlap, between mass atrocity and war are as follows.

1. We use Correlates of War (COW) Project Interstate, Intrastate, and Extrastate Wars Dataset to code the onset and end years for inter-, intra-, and extra-state wars for 1900–2007 ([www.correlatesofwar.org/data-sets/COW-war](http://www.correlatesofwar.org/data-sets/COW-war)). All such wars involve at least one state.
2. We use our compiled dataset of state-perpetrated mass atrocities using the onset and end years available in the data sources we use.
3. We line up each mass atrocity in our dataset with any war, even if it is obvious that the war

has nothing to do with the mass atrocity. Note that this protocol is “conservative,” that is, it creates a mass atrocity/war overlap even if there is not an obvious mass atrocity/war overlap.

4. If a mass atrocity overlaps with many wars, we use the notes from the data sources for mass atrocities (EGK, UV, or PITF-G) to identify which war the mass atrocity lines up with and go with that one.
5. If the previous protocol does not give rise to an obvious war in which to make the alignment, we choose “conservatively,” that is, we go with the war that is least advantageous to our claim that mass atrocities do not generally occur contemporaneously with or in the later stages of war.

#### **IV. Comparison of Mass Atrocity Fatalities with Fatalities from War and Subwar Conflict**

Estimated battle-related fatalities of 1,408,882 for the 177 wars and sub-war conflicts of all types (interstate, extra-state, civil, and internationalized civil) fought from 1989–2018 are reported by the Uppsala Conflict Data Program ([ucdp.uu.se/downloads/index.html#battlerelated](http://ucdp.uu.se/downloads/index.html#battlerelated)) (Eck and Hultman 2007; Pettersson, Högladh, and Öberg 2019).

#### **References**

- Butcher, Charles R., Benjamin E. Goldsmith, Sascha Nanlohy, Arcot Sowmya, and David Muchlinski. 2020. “Introducing the Targeted Mass Killing Dataset for the Study and Forecasting of Mass Atrocities.” *Journal of Conflict Resolution* 64 (7-8): 1524–47.
- Easterly, William, Roberta Gatti, and Sergio Kurlat. 2006. “Development, Democracy, and Mass Killing.” *Journal of Economic Growth* 11 (2): 129–56.
- Eck, Kristine, and Lisa Hultman. 2007. “One-Sided Violence against Civilians in War: Insights from New Fatality Data.” *Journal of Peace Research* 44 (2): 233–46.
- Harff, Barbara. 2003. “No Lessons Learned from the Holocaust? Assessing Risks of Genocide and Political Mass Murder since 1955.” *American Political Science Review* 97 (1): 57–73.
- Marshall, Monty G., Ted Robert Gurr, and Barbara Harff. 2019. “PITF - State Failure Problem Set: Internal Wars and Failures of Governance, 1955-2018.” <http://www.systemicpeace.org/inscr/PITFProbSetCodebook2018.pdf>.
- Pettersson, Therese. 2021. *UCDP One-sided Violence Codebook v 21.1*. <https://ucdp.uu.se/downloads/>.
- Pettersson, Therese, Stina Högladh, and Magnus Öberg. 2019. “Organized Violence, 1989-2018 and Peace Agreements.” *Journal of Peace Research* 56 (4): 589–603.
- Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. “Introducing ACLED-Armed Conflict Location and Event Data.” *Journal of Peace Research* 47 (5): 651–60.
- Ulfelder, Jay, and Benjamin Valentino. 2008. “Assessing Risks of State- Sponsored Mass Killing.” Working Paper. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1703426](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1703426).