

Article

Crystal Clear: Investigating Databases for Research, the Case of Drone Strikes

Giampiero Giacomello ^{1,*}  and Damiano Martinelli ²¹ Department of Political and Social Sciences, University of Bologna, 40125 Bologna, Italy² ELT Elettronica Group, 00131 Rome, Italy; damiano.martinelli@gmail.com

* Correspondence: giampiero.giacomello@unibo.it; Tel.: +39-051-209-2668

Abstract: The availability of numerous online databases offers new and tremendous opportunities for social science research. Furthermore, databases based on news reports often allow scholars to investigate issues otherwise hard to tackle, such as, for example, the impact and consequences of drone strikes. Crucial to the campaign against terrorism, official data on drone strikes are classified, but news reports permit a certain degree of independent scrutiny. The quality of such research may be improved if scholars can rely on two (or more) databases independently reporting on the same issue (a solution akin to ‘data triangulation’). Given these conditions, such databases should be as reliable and valid as possible. This paper aimed to discuss the ‘validity and reliability’ of two such databases, as well as open up a debate on the evaluation of the quality, reliability and validity of research data on ‘problematic’ topics that have recently become more accessible thanks to online sources.

Keywords: database analysis; triangulation; media reports; drones strikes; terrorism



Citation: Giacomello, G.; Martinelli, D. Crystal Clear: Investigating Databases for Research, the Case of Drone Strikes. *Data* **2021**, *6*, 124. <https://doi.org/10.3390/data6120124>

Received: 8 July 2021

Accepted: 15 November 2021

Published: 25 November 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last 20 years, scholarly research has benefited tremendously from the deluge of information available on the World Wide Web and other digital outlets. Nonetheless, the risk of unreliable or, worse, deliberately false sources has considerably increased, especially with the emergence of social networks as major avenues of communication to/from users and as news feeds, too, as the “Russiagate” and related problems for the Trump Administration demonstrated, or the “infodemia” of (more or less trustworthy) news about COVID-19. Scholars are not immune from such perils because they themselves rely increasingly on publicly available information and online databases. Hence, greater care and scrutiny are necessary to ensure that data and sources used to investigate social phenomena are valid, reliable and trustworthy. Data *triangulation* may be a viable solution. Much as the orienteering method, the idea is to discover the “location” of a third point, by using two known relative positions, or, in this case, by comparing two different, independent sources of information on the same issues and seeing if their findings coincide, namely that they “say the same thing”. If possible, correlation coefficients, with all the necessary due attention, may give the researcher an actual measure of “how much” the two independent sources overlap, hence increasing confidence about their reliability. Incidentally, using different sources of data collected for diverse purposes to find creative solutions to problems is one of the linchpins for Big Data (BD) analytical methods [1,2].

Although triangulation has been hailed as a highly valuable methodology for the social sciences, it may be too plagued by scarcity of data or bad quality of the same, and indeed, the debate on whether or not triangulation is a sustainable alternative for the social sciences has been lengthy [3,4]. Denzin gives four examples of ‘triangulation’: (1) data triangulation, (2) investigator triangulation, (3) theory triangulation and (4) methodological triangulation [1,5]. In this paper, we would like to focus on ‘data triangulation’, which has three subtypes: (a) time, where observations are collected at different moments, (b) space,

which requires different contexts, and finally, (c) persons. We thus examine data triangulation for different times and locations specifically for two databases reporting on drone strikes, namely one created by the New America Foundation (NAF) and the other by The Bureau of Investigative Journalism (TBIJ). We used the most updated datasets available as of the end of 2021. Unfortunately, after 2017, NAF stopped distinguishing between “air” and “drone” strikes for Yemen and Somalia, but kept the division for Pakistan, thus making it quite problematic for a comparative analysis of all the cases after the period 2017–2018. We nonetheless believe that the methods suggested and explored in this paper are quite useful examples for social researchers who have to work with different datasets on the same problem.

Nowadays, unmanned combat aerial vehicles (UCAVs) such as the Predator have become a linchpin in the struggle against terrorist organizations, such that the number of studies focusing on drone strikes and warfare is particularly notable. Several such studies rely on databases generated by news reports [6–8]. Furthermore, the ethical and political [9] consequences of drone warfare for counterterrorism imply that scholarly research on such topics would have essential implications for public policy. Likewise, research questioning the dependability of sources and databases is also on the rise [10,11]. For these reasons, we decided to focus specifically on the two databases to see how ‘compatible’ they are with one another and if they indeed ‘say the same thing’. If these conditions were to be satisfied, then the results would be consistent with Denzin’s data triangulation, which, in turn, would further strengthen the validity of papers relying on those databases.

As is well known, validity and reliability are the two most important issues when it comes to datasets. While reliability concerns ‘the extent to which (...) any measuring procedure yields the same results in repeated trials’, validity pertains to ‘the crucial relationship between concept and indicator’. Both are, nonetheless, a ‘matter of degree’ [12]. These issues are critical in the drone debate where think tanks, NGOs and the US government provide significantly different data about the casualties. We are fully aware that when stretching the boundaries of probability samplings, authors need to be very careful with the generalizations they reach based on these databases. On the other hand, these datasets are the only viable sources to investigate counterterrorism drone campaigns.

In this context, reliability is problematic, as common methods to measure it (e.g., test/retest, split-half) can hardly be applied. This is, nonetheless, a frequent obstacle in social science fields such as security studies or international relations, much less, for example, in clinical psychology. For the two databases considered here, the ultimate sources of reliability would be the quality, ethics and professionalism of the witnesses, observers and journalists who collected information and data on the strikes and their casualties and damage. Measuring reliability at such levels, however, is far beyond the scope of this paper.

Validity of the two databases (that is, if they indeed measure what they are intended to measure) can be managed with a greater degree of confidence. Of the three basic types of validity—content validity, criterion-related validity and construct validity—the first two have ‘limited usefulness in assessing the quality of social science measures’. Construct validation, on the other hand, has ‘generalized applicability in the social sciences’. More specifically, ‘if the performance of the measure is consistent with theoretically derived expectations, then it is concluded that the measure is construct valid’ [12]. In the case of the two databases, studies in the relevant literature are those using news coverage for scholarly analysis that allows for a relatively positive reply to the expectation of whether the databases are valid sources to assess the impact of drone strikes.

As observed by Trochim, there are four possible options when assessing reliability and validity [13]. The data are (a) both reliable and valid, (b) neither reliable nor valid, (c) valid and not reliable, or (d) reliable and not valid. Of these, the first is clearly the most preferable, and the second is the worst case. Between the third and the fourth (albeit both not without problems), it is preferable that data are valid and (relatively) reliable rather than fully reliable but not valid, thus not measuring what the researcher is interested in. In our case, the datasets were valid, as they measured the casualties in the countries

concerned, but were not entirely reliable due to the inherent difficulties of building a dataset on news reports and media outlets [11,14]. As Franzosi noted, newspapers have long been relied upon by historians and social scientists in general as informal sources of historical data. The author defends the use of news reports as data sources because they often constitute the only available source of information, not all events or items of information are equally liable to misrepresentation in the press, and the type of bias likely to occur in mass media consists more of silence and emphasis rather than outright false information [2].

The ultimate example of dataset building on collective news media is the GDELT Project, the largest, most comprehensive, and highest resolution open database of human society ever created [15]. Another great example is the Global Terrorism Database, an open-source database including information on terrorist events around the world from 1970 through 2016 [16].

2. Data Description

This section of the paper is devoted to the analysis of the methodologies and aggregated data of the various NAF and TBIJ data collections. The next part will then focus on the data trends, and finally the third will analyse the data in depth, strike by strike in order to verify the validity and the degree of agreement between the datasets.

2.1. The NAF Datasets

NAF has developed three datasets for Pakistan, Yemen and Somalia to provide information on UCAV operations through data collected from credible news reports [17]. NAF research is based on media outlets from three international news agencies (AFP, AP and Reuters), leading newspapers in the areas considered, major South Asian and Middle Eastern TV networks and Western media networks with extensive reporting capabilities in the countries concerned. Regarding the data collection methodology, NAF registers every strike only in the presence of two credible sources, and multiple strikes are listed as a single event only if they take place in the same area and in a 2-hour range. NAF researchers report the casualty figures as numerical intervals representing the minimum and maximum number of people killed as reported by different news sources. They describe every strike with a total casualties range and subtotals referred to as civilians, militants and unknowns killed. The differentiation among militants, civilians and unknowns follows these criteria:

- If two or more sources identify the casualties as militants while others refer to them as “people” or with other neutral terms, they are considered militants.
- Casualties referred to as civilians, children or women are listed as civilians.
- If only one source refers to the casualties as civilians and the majority of the news reports do not, they are listed as unknowns.
- The casualties are categorized as unknowns when different sources are too contradictory to come to a definitive conclusion.
- If a media outlet reports some civilians among the casualties without defining an exact number and no other source indicates a precise amount, NAF researchers classify a third of the total casualties as civilians or unknowns.
- The term ‘foreigners’ is considered indicative of militants unless the term ‘civilians’ is specified; the term “tribesman” is considered neutral [18].

In NAF datasets, the sum of the minimum or maximum subtotal values corresponds to the minimum or maximum total value. This regularity, shown in Equations (1) and (2), is a particular feature of these datasets facilitated by the addition of the unknowns’ columns. In these equations, T refers to totals, M to militants, C to civilians and U to unknowns. The subscripts “min” and “max” indicate the minimum and maximum values.

$$T_{\min} = M_{\min} + C_{\min} + U_{\min} \quad (1)$$

$$T_{\max} = M_{\max} + C_{\max} + U_{\max} \quad (2)$$

NAF Aggregate Data

The multi-annual aggregate data on the various countries are presented in Appendix A following the original datasets model, which comprises eight columns for the minimum and maximum values of the total casualties ($T_{\min} - T_{\max}$), militants ($M_{\min} - M_{\max}$), civilians ($C_{\min} - C_{\max}$) and unknowns ($U_{\min} - U_{\max}$).

In this subsection, we analysed the aggregate data of the three datasets unified in a comprehensive spreadsheet that provided an overview of UCAV effects outside areas of active hostilities. Table 1 presents the data relative to the three contexts and a complete overview of the UCAV campaign. There have been 655 UCAV strikes. The gap between the minimum and maximum values expanded with the increase in strike numbers (N_{att}), reaching 1600 units for the total values visible in Table 1's last row. The considered period extended from 3 November 2002 to 31 December 2017, presenting only 1 year without registered strikes. The second part of Table 1 shows the data evolution over time. The Pakistan data had a great influence on the results, as proven by the highest N_{att} value in 2010. On the other hand, the decrease in this indicator after 2010 was slower for the contribution by Yemen and Somalia data; in fact, these two countries experienced growth in the number of operations after 2010. The campaign intensity was significant from 2008 to 2017, as highlighted by the total and militants' data analysis. In these 9 years, the total and militants' data were constantly above 177, moving close to a thousand in 2010. The 2006 data were peculiar because only two strikes took place in Pakistan in that year with a large impact on civilians. For this reason, some of the graphs that follow do not present the 2006 data.

Table 1. NAF comprehensive dataset data in aggregated form and distribution over time.

Country	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax	Umin	Umax
Pakistan	414	2366	3702	1910	3071	245	303	211	328
Yemen	210	920	1208	841	1100	47	53	36	60
Somalia	36	169	205	148	180	11	11	10	10
Year	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax	Umin	Umax
2002	1	6	6	6	6	0	0	0	0
2004	1	7	7	1	1	1	1	5	5
2005	3	15	15	5	5	4	4	6	6
2006	2	88	100	1	1	87	99	0	0
2007	4	48	77	37	65	0	0	11	12
2008	36	236	334	169	247	24	33	43	54
2009	54	372	712	241	500	56	67	75	145
2010	123	642	1033	592	969	19	24	31	40
2011	81	456	695	371	571	49	62	36	62
2012	97	522	740	478	678	20	24	24	38
2013	51	225	292	201	262	20	26	4	4
2014	42	237	323	232	318	5	5	0	0
2015	38	178	200	175	197	2	2	1	1
2016	54	187	212	171	196	3	3	13	13
2017	68	229	358	212	320	13	17	8	18
2018	5	7	15	7	15	0	0	0	0
Total	660	3455	5119	2899	4351	303	367	257	398

2.2. TBIJ: Methodology

Over the past few years, TBIJ has conducted a research project called Drone Warfare that includes three datasets regarding UCAV strikes in Pakistan, Yemen and Somalia [19]. These datasets are based on news reports, statements, documents and press releases from national and international media, similarly to the NAF datasets. The TBIJ methodology differs in some respects from that used by NAF. For example, strikes launched more than an hour apart or more than 2 miles apart are registered as separate strikes. However, the data presentations are similar: numerical ranges to reconcile different sources. TBIJ datasets present, besides total casualties data, the number of civilians, children and injured involved in every strike. Regarding the counting of civilian casualties, TBIJ researchers use the formula $0-X$ where X refers to the maximum reported number of civilians killed when there is a source with precise information on the number of civilians killed but no other confirming sources. This formula is used to include possible civilian casualties in the count even when the media reports mentioned “people”, “local tribesmen” or “family members” being killed.

The TBIJ research team did not register the number of militants’ casualties for every strike because the term ‘militant’ is considered politically and emotively charged and it does not yet have an accepted legal definition (The Bureau of Investigative Journalism). In these datasets, the subtotals sum does not coincide with the total casualties figures because of the specific counting method ($0-X$) and because of the lack of the militants’ data. This is an operational limitation of these datasets. Indeed, the militants’ figures are essential for a complete comparison with NAF datasets. For this reason, we have elaborated the original datasets to add the militant casualties calculated as follows (Negative values were reported as zero):

$$M_{\min} = T_{\min} - C_{\max} \quad (3)$$

$$M_{\max} = T_{\max} - C_{\min} \quad (4)$$

TBIJ: Aggregate Data

In the TBIJ datasets analysis, we ignored the data referring to the children and injured involved in the strikes, while we added the militants’ data. Furthermore, these datasets do not present the columns related to the unknowns. The data related to each dataset are reported in Appendix B. Table 2 shows the result of the elaboration of a comprehensive database including all the Pakistan strikes and the UCAV “Confirmed” operations conducted in Yemen and Somalia (see Appendix B). The registered strikes totalled 616 over 16 years, and the Pakistan data were higher for every indicator, followed by Yemen and Somalia. The gaps between minimums and maximums reached significant values; in particular, C_{\max} exceeded 1000 units for the first time, more than twice C_{\min} . The highest values of all the indicators were in 2010 with the exception of the civilians’ data. In fact, C_{\min} and C_{\max} reached the top in 2009 and maintained relevant values in the next 2 years. The N_{att} decrease after 2010 was more gradual than the Pakistan case because of the inclusion of data on Yemen and Somalia. The 2017 N_{att} was less than half the NAF one due to differences in the accounting methodology. The civilians’ figures were peculiar because they presented small values in the first and last 4 years, including 2003.

Table 2. Data from the TBIJ comprehensive database in aggregated form and distribution over time.

Country	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax
Pakistan	430	2515	4026	1704	3602	424	969
Yemen	153	640	925	568	857	68	109
Somalia	34	273	387	260	384	3	15
Year	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax
2002	1	6	6	6	6	0	0

Table 2. *Cont.*

2004	1	6	8	4	6	2	2
2005	3	16	16	5	11	5	11
2006	2	94	105	0	15	90	100
2007	5	36	56	5	45	11	46
2008	38	252	401	115	342	59	173
2009	54	471	753	292	653	100	210
2010	128	755	1108	573	1019	89	197
2011	85	410	770	289	682	88	188
2012	84	421	685	357	665	20	76
2013	50	190	327	156	310	17	41
2014	43	215	331	202	327	4	13
2015	41	155	221	142	218	3	16
2016	48	304	426	298	422	4	6
2017	33	96	122	87	119	3	14
2018	1	1	3	1	3	0	0
Total	617	3428	5338	2532	4843	495	1093

3. Methods: Data Comparative Analysis

An in-depth datasets analysis and the study of two relevant indexes will describe, in the next section, the trends and impact of the targeted killing campaigns outside areas of active hostilities. The analysed indexes were the number of strikes per year and the ratio (C/M) between the number of civilians and militants killed annually: C_{\min}/M_{\min} and C_{\max}/M_{\max} . The study of these indicators provided a first comparison of the datasets that helped us verify and consolidate the validity of the data.

3.1. Pakistan

Pakistan is the main case study regarding UCAV use, considering the great number of reported strikes. Figure 1 represents the evolution of N_{att} over time: NAF data, in orange, and TBIJ data, in blue, can be compared in the graph. There were small differences from the two data series as confirmed by the graph trend. Both data series reached the top in 2010 reporting: 122 and 128 strikes. This six-unit gap was the highest in the years considered. Moreover, both trends showed a rapid increase from 2007 to 2010 and a fast decrease after 2010. Considering the methodological differences related to the registering of multiple strikes, NAF and TBIJ datasets presented significantly similar data referring to N_{att} in the Pakistani area.

In conclusion, the N_{att} index confirms the reliability of the data from the two organizations since the trends and the values are very similar and consistent.

The C/M Ratio

The annual C/M ratio is the more relevant index for the analysis. Figures 2 and 3 show the trends of the C/M respectively for the NAF and TBIJ datasets. For a readability issue, the graphs do not present the 2006 ratios that amounted to 87 and 99 for the NAF dataset and to 90/0 and 6.67 in the case of the TBIJ data. Figure 2 provides a clear image of the progressive decrease in the ratio. After 2006, the values never exceeded 0.23, which means that for every four militants hit, a maximum of one civilian was killed. The situation improved further after 2011, when the values did not exceed 0.09. The total C/M values describe a campaign in which from 10 to 13 civilians were killed for every 100 militants.

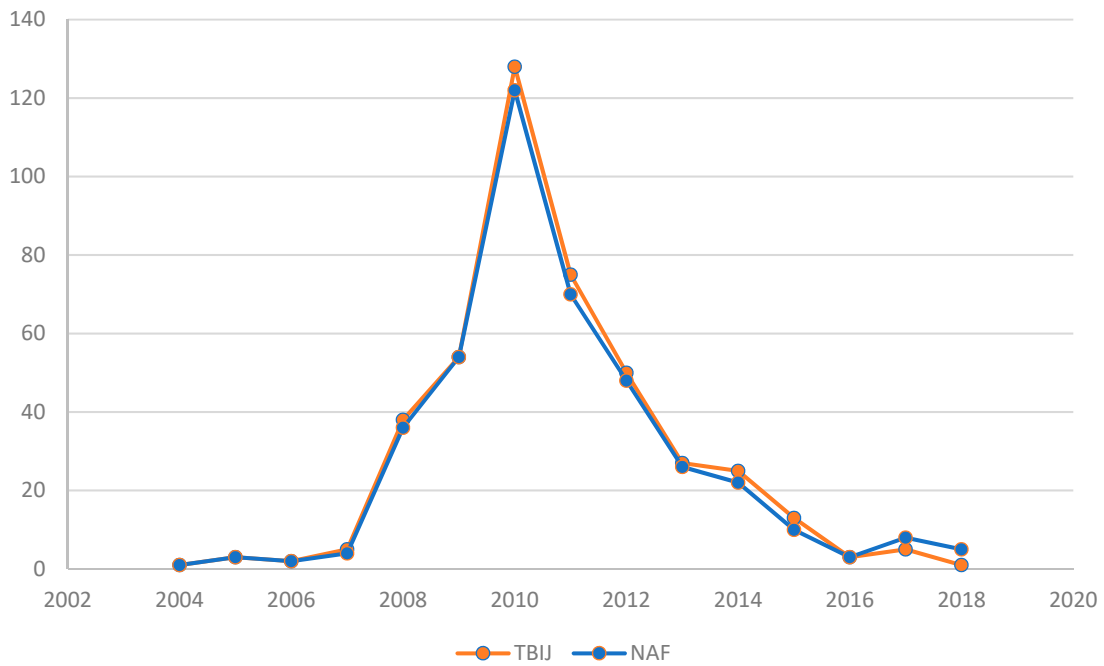


Figure 1. N_{att} in Pakistan from 2004 to 2018.

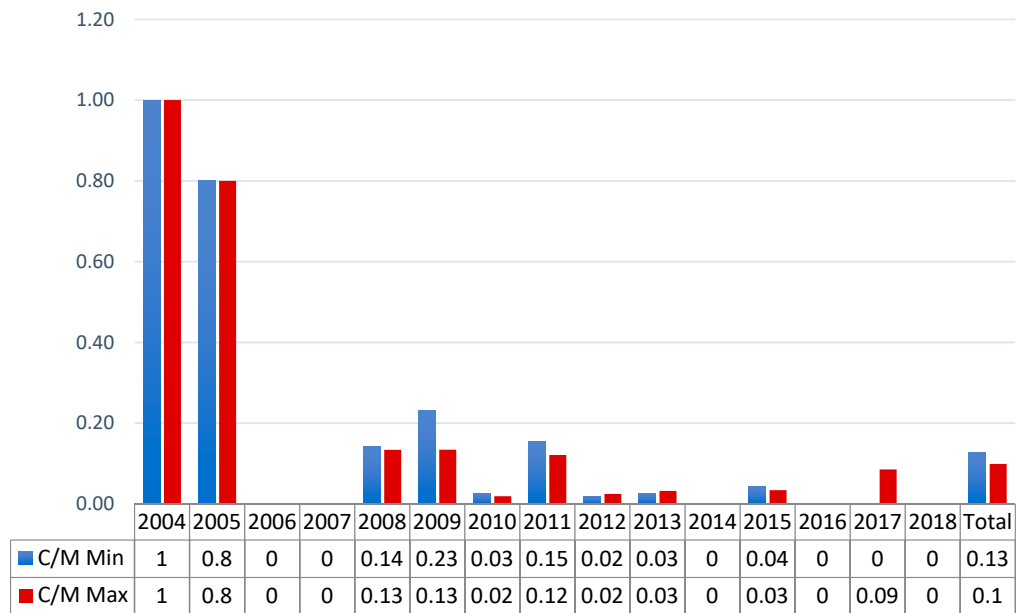


Figure 2. NAF data C/M ratio for Pakistan from 2004 to 2018.

For TBIJ data, the C/M index was higher over many years and reached a peak in 2006–2007. The total results (0.25–0.27) highlight the differences from the NAF dataset. The main discordance was in the 2007 data, where there were null values in Figure 2 and ratios from 1.02 to 2.20 in Figure 3. Nonetheless, Figure 3 shows a decrease over time in the analysed ratio; the index never exceeded 0.25 after 2009, and the highest value in the final 5 years was 0.14 in 2017. In conclusion, despite some relevant differences, both datasets indicate a decrease in the ratio and hence an improvement over time in the targeted killing operations’ accuracy. In terms of reliability, we can conclude that the C/M indicator presents relevant shifts in the values, but the trends were sufficiently consistent considering the methodological dissimilarity and the use of media outlets as the only source.

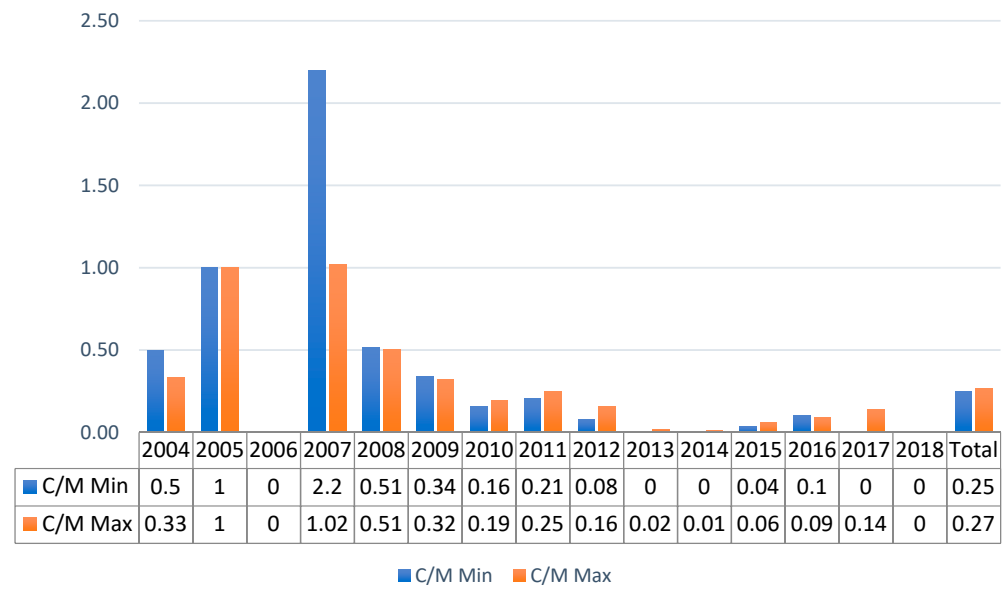


Figure 3. TBIJ data C/M ratio for Pakistan from 2004 to 2018.

3.2. Yemen

The Yemeni case study had a secondary role in the analysis but was significant to verify the trends described in the Pakistan datasets. The first aspect analysed was the evolution of N_{att} over time, shown in Figure 4.

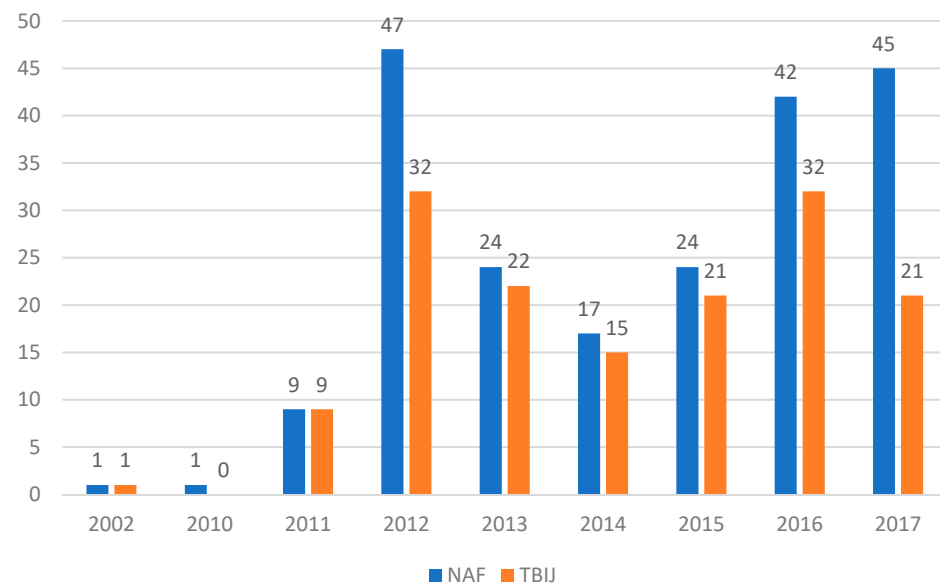


Figure 4. N_{att} in Yemen from 2002 to 2017.

Figure 4 shows that the TBIJ reported a N_{att} always smaller or equal to the NAF, but the trend was similar. Both data series reached a peak in 2012, significantly decreasing in the next 2 years and presenting a renewed increase in 2015–2016. The UCAV campaign intensity was significant in the period 2011–2017, with a minimum of nine strikes per year in both datasets, and there was a long gap between the first and second strikes. To summarise, the two datasets agreed on the N_{att} trend over the years, despite TBIJ reporting 57 strikes less than NAF.

In the validity analysis, we must consider the methodological differences and the confusion related to the inclusion in one dataset of strikes executed not only with UCAV but also with special forces raids, cruise missiles and conventional aircraft as well (see

Appendices A and B). Considering these limits, the N_{att} index provides us with a mixed picture in terms of a reliability assessment; while the trend is sufficiently consistent, the gaps between the annual values are in some cases wide and significant.

The C/M Ratio

The C/M ratio is crucial for an understanding of the UCAV impact in Yemen. Figures 5 and 6 show the evolution of this index over time. Figure 5, elaborated from NAF data, shows values lower than 0.22 for all years except 2010. In that year, the C/M reached a peak, but only one strike was registered. Four years reported null values and the 2012, 2014 and 2017 ratios were lower than 0.06. The C/M total values present an improvement in the UCAV campaign from the Pakistani situation, with a comprehensive ratio of 5–6 civilian casualties for every 100 militants killed.

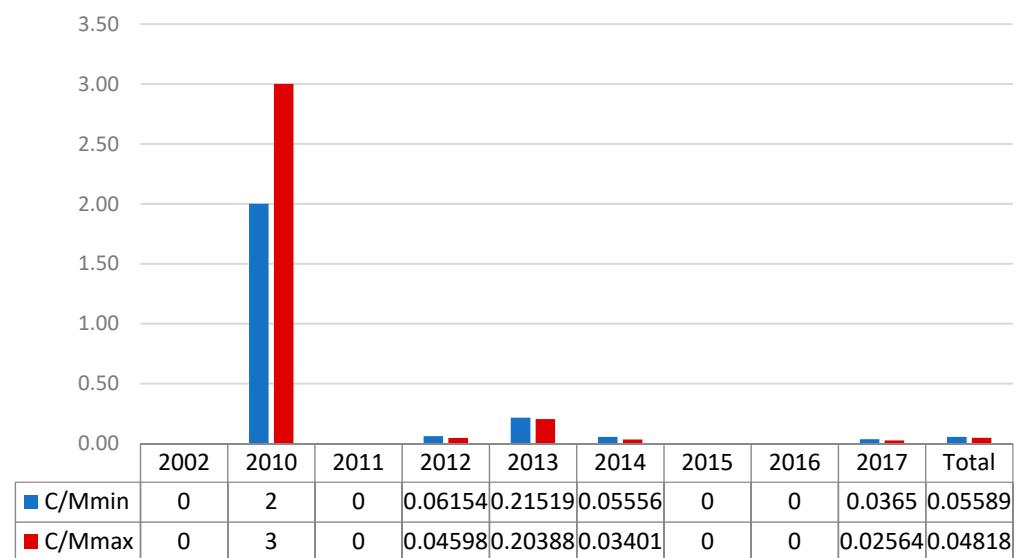


Figure 5. NAT data C/M ratio for Yemen from 2004 to 2017.

Figure 6 presents TBIJ data that reached a peak in 2011 with relatively small values, 1.06 and 0.55. Only in 3 years did the data exceed 0.07, and there were null values in 2002 and 2016. The total C/M values were twice the NAF ratios but approximately half the values in the Pakistan TBIJ dataset. Hence, even the TBIJ data presented a relevant drop in this index for Yemen when compared to Pakistan. In conclusion, both datasets confirm a significant decrease over time in the C/M index and an improvement from that for Pakistan. Examining these indexes from a reliability point of view, we can see general agreement in the overall trend, but relevant differences in the annual data due to the limits highlighted in the N_{att} indicator.

3.3. Somalia

The Somalia case had a minor role in the analysis due to the small number of strikes registered. The N_{att} values were in fact 36 and 34, respectively, for NAF and TBIJ. Figure 7 presents a graph that shows the N_{att} evolution. The indicator peaked in 2016–2017, highlighting a UCAV campaign at the peak of its intensity. The total C_{min}/M_{min} ratio was 0.01 for the TBIJ and 0.07 for the NAF, while the C_{max}/M_{max} values were respectively 0.04 and 0.06. These ratios were smaller than the results for Pakistan and Yemen. In conclusion, the congruity between the NAF and TBIJ and the exiguity of the values must be noted. In terms of reliability, the Somalia datasets share the limits of the Yemen datasets (see Appendices A and B), although we can observe a similarity in the N_{att} trend until 2016 and a general agreement in the total C/M values.

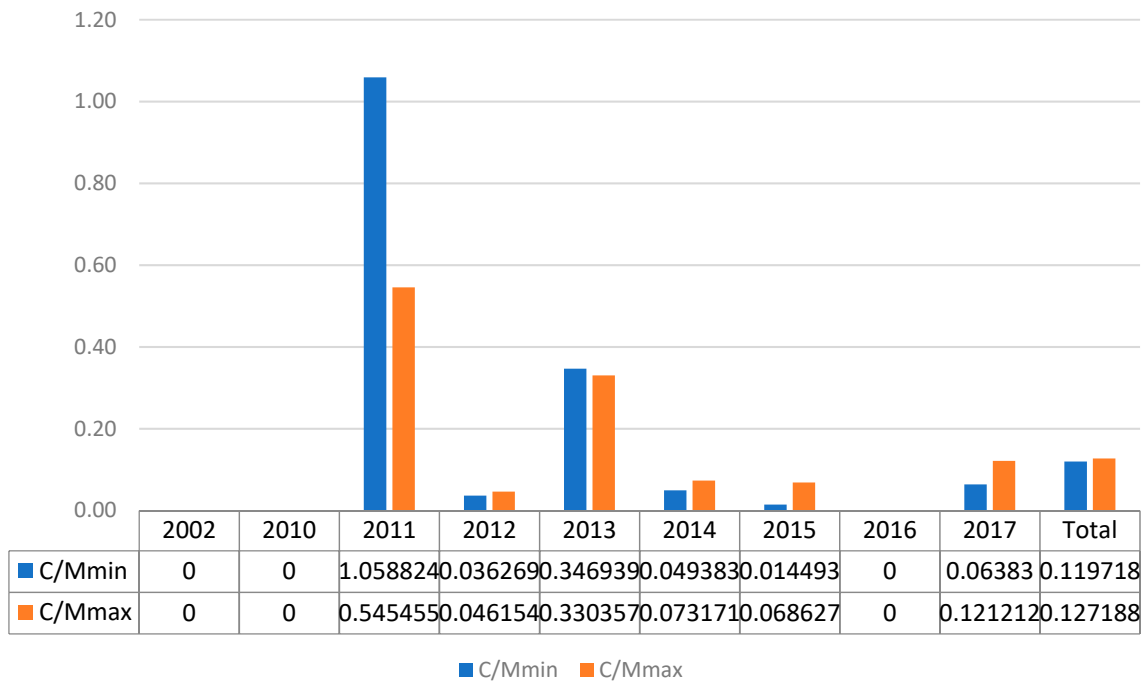


Figure 6. TBIJ data C/M ratio for Yemen from 2004 to 2017.

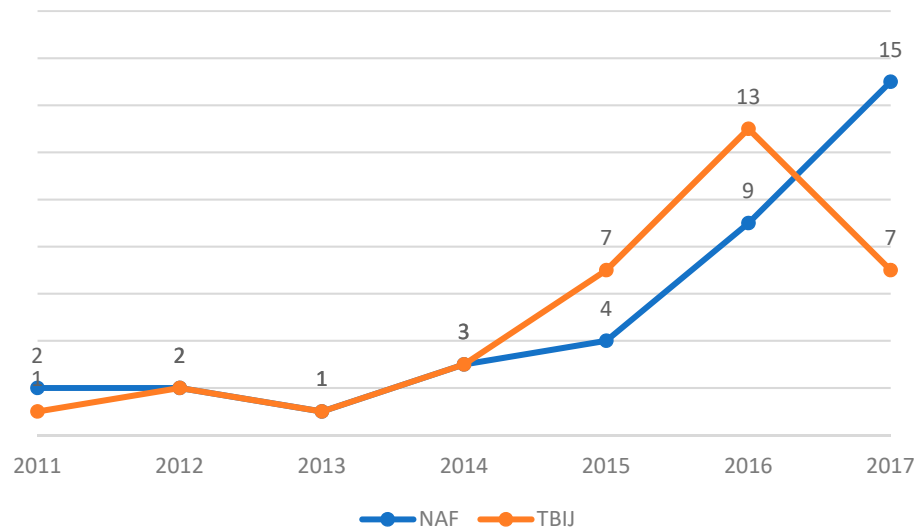


Figure 7. N_{att} in Somalia from 2011 to 2017.

4. Comparison and Discussion of Results

In this second part of the paper, we analyse and compare the data strike by strike to further assess the validity of the two organizations’ datasets. This process will lead to the elaboration of a ‘strike by strike’ matrix for every country, which includes all the information contained in the original datasets. The matrix methodology and results are presented in the following sections.

4.1. Methodology and Results

The comparative matrix building process started with the elaboration of the NAF and TBIJ datasets to a unique format. To achieve this goal, we calculated the militants’ data for the TBIJ dataset. The chosen format contained ten columns for the strike date and location, the total results, the militants’ data, the civilians’ data, the unknowns’ data and the link to the strike web page. In the case of the TBIJ Pakistan and Yemen datasets, we kept the columns related to the ‘Area’ and ‘Province’ of the strike. Furthermore, we

kept the columns referring to the organizations targeted in every strike included in NAF datasets. The comparison process took into account, first of all, the date and location columns and classified the strikes into three categories: identical, similar and different. The identical strikes were those pairs of events where date, location and numerical data coincided in both the NAF and TBIJ datasets. Similar strikes presented the same date and location but different casualty figures. Strikes registered as different were those that did not have a corresponding event in the dataset of the other organization. The date margin of tolerance was 1 day, and in the location comparison, a flexible approach was necessary due to the difficulties in locating different places and to linguistic issues with the geographical designation of various areas, villages and regions.

We considered places geographically close to the same location, taking into account the researcher's difficulties in precisely locating the strikes. In some cases, strikes registered by the same organization and occurring on the same date and location were merged into a single strike. This operation was necessary only 13 times and in particular conditions to facilitate the comparison process. Merging the strikes was justified by the different methodologies used by the two organizations to catalogue strikes occurring on the same day at different times. The final stage of the process consisted of building a matrix for every country including the identical strikes registered singularly for different strikes by the NAF and TBIJ and the non-numerical information on the similar strikes. This matrix was the basis for the construction of many possible comparative databases depending on the criteria chosen for the elaboration of numerical data on similar strikes. Therefore, we could elaborate these data in different ways according to the purpose of the analysis.

4.2. Pakistan

The Pakistan comparative matrix (PAKO) presents the results shown in Table 3. Regarding the methodology, for the strike reported on the date of 19 June 2004 by the NAF and on the date of 17 June 2004 by the TBIJ, we made an exception to the date margin of tolerance (The two strikes were the only ones reported by the two organizations in 2004). In 15 cases, two strikes with different dates were considered similar or identical, and the merger of multiple strikes was necessary eight times (see Appendix C).

Table 3. Number of strikes categorized by type and country from the comparative databases.

Country	Identical	Similar	Different	Total
Pakistan	72	329	32	433
Yemen	63	72	91	226
Somalia	8	13	25	46
Total	143	414	148	705

The comparison results report 72 identical strike pairs, 329 similar and 33 different events. Eleven different events derived from the NAF dataset and 22 from the TBIJ dataset. This fact reflects the higher N_{att} in the TBIJ dataset. Therefore, PAKO included 434 strikes and classified about 7.5% as different strikes. The data show how the two original datasets agree on date and location in 92% of the cases but only in 17% of their numerical data. The aggregate data for only the identical strikes are interesting for two reasons: first, the unknowns' data were null because the TBIJ did not include it, and second, the civilians' data amount to three casualties. This aspect demonstrates the difficulties of finding agreement between the two organizations' datasets in terms of the number of civilian casualties, and was related to methodological and terminological differences. It must be stressed that the criteria for the "Identical" category were very strict. For example, if we considered only the total values, the number of identical strikes would increase by 28, and more than half of the pairs of similar strikes would present a difference between the sum of the respective minimum and maximum total values that did not exceed 2.

4.3. Yemen

The Yemen comparative matrix (YEMO) did not differ from PAKO in methodology or format. We considered similar or identical pairs of strikes with different dates and merged multiple strikes into a single event only two times (see Appendix C). In the TBIJ Yemen dataset, the location was missing in some cases while the province was always specified; in these cases, we considered only the province indicator for the comparison. The third line in Table 3 shows the YEMO results. The main difference from the PAKO data was the number of different strikes that represented more than a third of the total. Fifteen of these strikes derived from the TBIJ dataset and 76 from the NAF dataset.

These data reflected differences in N_{att} only in part and reflected variations in data collection methodology. In fact, both the original datasets registered the UCAV strikes along with strikes conducted with other military tools. Moreover, the TBIJ Yemen dataset presented strikes where US responsibility was not confirmed. The inclusion of different types of strikes in the datasets caused a methodological issue highlighted by the high number of different strikes observed. Nonetheless, it is remarkable that almost 60% of the strikes were identical or similar, and more than a fourth were identical. The aggregate data related only to the identical strikes confirmed the issues seen in the PAKO case. The unknown data were null, and the two datasets agreed on a strike with civilian casualties only once, reaffirming the difficulties of finding agreement between the two organizations' datasets in terms of civilian casualties.

4.4. Somalia

In the Somalia comparative matrix (SOMO), a date margin of tolerance was necessary in two cases, and we merged multiple strikes only three times (see Appendix C). The comparison results, presented in Table 3 (fourth line), diverged from those of YEMO and PAKO because more than half of the strikes were different, and only 17% found a perfect match. The strikes registered as different derived from the NAF in 13 cases and from the TBIJ in the other 12. The identical strikes data confirmed the issues for civilian casualty data that were null. In conclusion, SOMO was different from the other comparative matrices and reaffirmed the methodological issues previously highlighted.

5. Conclusions

Given the continuing attention to the drone debate [20] and, in particular, the casualties data and the number of articles published on the subject, we consider it an important contribution to the field that two of the most relevant databases were thoroughly tested for coherence and concordance. In this respect, our analysis showed that the two databases were generally in agreement on the overall picture. In particular, the distribution over time of the number of strikes was highly consistent in the two databases and their civilian/militant ratios also showed similar trends. Furthermore, the comparison results presented strong concordance between the datasets. The comparison total results, presented in Table 3, show the number of strikes by type and country. The most relevant aspect was the exiguity, in percentage, of the number of different strikes in PAKO. This indicates that PAKO was the matrix in which the NAF and TBIJ agreed more often on date and location. On the other hand, the number of identical strikes was higher in relative terms in YEMO, at 25%, against percentages of around 17% for SOMO and PAKO. The total results show that 20% of strikes were identical, 59% were similar and 20% were different. In conclusion, the two think-tanks agreed in 79% of the cases on date and location while presenting one in five identical strikes. The inclusion in a single matrix of all these strikes provides a more complete picture of the targeted killing campaign because strikes registered by one organization and not from the other have been included in the 'Different' category.

In the literature about triangulation between the four cases highlighted by Denzin [1,5], the one referring to the use of different databases has been given less attention, perhaps due to a longstanding issue of data scarcity at the time of writing (late 1970s). Today with plenty of digital data available and a growing number of news sources, the possibility of

creating very large databases such as GDELT and GTD is becoming more common and user-friendly, as demonstrated by the field of Big Data analysis. It is indeed likely that triangulation by different data sources will become the most common approach. Avenues for future research include merging the two databases into a single one built using the comparative databases as a starting point. Furthermore, the development of new studies correlating strike data with other variables such as daily reactions in the media as captured by GDELT or the impact on the terrorist organizations is also likely. Clearly, the obstacles are still quite substantial. For example, in the two databases, even if they present similar trends, the possibility of eliminating systematic error is quite challenging, as we can see with the differences in total results comparing civilian casualties in Tables 1 and 2.

As is already known, most empirical evidence of validity is correlational [21]. Since the two databases show a significant number of identical and similar strikes, we can safely affirm that a satisfactory level of validity has been achieved. As for reliability, the final assessment is more problematic. While the literature on the reliance of using news coverage for research is quite established, we have to conclude that reliability is only partial, as highlighted by the 143 different strikes reported in Table 3. Thus, more effort should be invested in this area by finding ways to (a) improve the level of detailed information in the databases and (b) comparing news reports with other data sources that may support (or reject) the conclusions reached. Using Trochim's four alternatives for reliability and validity, we can assert that, for the two databases considered here, we have a 'data are valid but relatively reliable' case, that is, a 'second best' (preferable to 'reliable but not valid') scenario. In more general terms, this is an achievement to strive for. Naturally, it would be much more preferable to have the 'valid and reliable' assessment, but, for the social sciences, the second best is *the most* a researcher often can hope for.

In our specific case considered here, the databases are likely to become even more central to any future research on drone strikes. In fact, as it seems, moving away from the greater transparency that had begun to appear during the last phase of the Obama administration, the Trump administration, as shown by a recent report [22], moved toward more secrecy and thus more opacity with the data made available for this type of research. As the alternative would be to do 'less' research on drone strikes, a second-best outcome of 'valid even if relatively reliable' does not seem too 'unpalatable'.

A final, general, lesson from the cases examined in this paper is that more data, of the most diverse types, available online (e.g., GDELT) may present a valued opportunity for social scientists to improve the overall quality of their research. Perhaps these are excessive expectations, as has already been the case in the past, or the social sciences are inexorably doomed to remain 'imprecise' by their own nature, but, as we mentioned in our introduction, (big) data could definitely be a game-changer by harnessing, for example, unrelated data to provide researchers with undiscovered insights into the problems at hand. Such a path will certainly increase the relevance of the validity/reliability issue, but it will also offer tremendous opportunities for new research avenues that would be too tempting to leave untried.

Author Contributions: Conceptualization, G.G. and D.M.; Data curation, D.M.; Formal analysis, D.M.; Methodology, D.M.; Supervision, G.G.; Validation, G.G. and D.M.; Writing—original draft, D.M.; Writing—review & editing, G.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the datasets used in this paper are available (in .zip format) at Centre for Computational Social Science of the University of Bologna <https://centri.unibo.it/computational-social-science/en/publications> (accessed on 11 November 2021), under the CC BY 4.0 licence.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Concerning the Pakistan campaign, Table A1 shows the strike data in an aggregated form and in their distribution over time. The period considered extends from 19 June 2004 to 31 December 2017. The Yemen dataset includes data about air strikes conducted by UCAV and by other military instruments such as cruise missiles and raids. Table A2 shows the total results and the data for the different types of strikes. The first line is the most significant for the analysis. The period considered in this case extended from 3 November 2002 to 31 December 2017, and Table A2 displays the data evolution over time. The Somalia dataset also reports strikes conducted with either UCAV or other military means such as Special Forces raids and cruise missiles. The first section of Table A6 shows the total results and the data for the different types of strike. Table A3 presents the data distribution from 23 June 2011 to 31 December 2017.

Table A1. NAF data for Pakistan.

Year	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax	Umin	Umax
2004	1	7	7	1	1	1	1	5	5
2005	3	15	15	5	5	4	4	6	6
2006	2	88	100	1	1	87	99	0	0
2007	4	48	77	37	65	0	0	11	12
2008	36	236	334	169	247	24	33	43	54
2009	54	372	712	241	500	56	67	75	145
2010	122	636	1025	590	967	15	18	31	40
2011	70	393	614	317	514	49	62	27	38
2012	48	230	359	213	325	4	8	13	26
2013	26	123	162	120	157	3	5	0	0
2014	22	128	157	128	157	0	0	0	0
2015	10	49	61	47	59	2	2	0	0
2016	3	8	11	8	11	0	0	0	0
2017	8	26	53	26	47	0	4	0	2
2018	5	7	15	7	15	0	0	0	0
Total	414	2366	3702	1910	3071	245	303	211	328

Table A2. NAF data for Yemen.

Type	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax	Umin	Umax
UCAV	210	920	1208	841	1100	47	53	36	60
Others	20	350	455	289	366	56	81	5	8
Total	230	1270	1663	1130	1466	103	134	41	68

Year	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax	Umin	Umax
2002	1	6	6	6	6	0	0	0	0
2010	1	6	8	2	2	4	6	0	0
2011	9	62	79	53	55	0	0	9	24

Table A2. *Cont.*

2012	47	287	376	260	348	16	16	11	12
2013	24	100	128	79	103	17	21	4	4
2014	17	95	152	90	147	5	5	0	0
2015	24	91	101	90	100	0	0	1	1
2016	42	127	147	124	144	0	0	3	3
2017	45	146	211	137	195	5	5	8	16
Total	210	920	1208	841	1100	47	53	36	60

Table A3. NAF data for Somalia.

Type	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax	Umin	Umax
UCAV	36	169	205	148	180	11	11	10	10
Others	36	794	966	758	894	62	100	0	38
Total	72	963	1171	906	1074	73	111	10	48

Year	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax	Umin	Umax
2011	2	1	2	1	2	0	0	0	0
2012	2	5	5	5	5	0	0	0	0
2013	1	2	2	2	2	0	0	0	0
2014	3	14	14	14	14	0	0	0	0
2015	4	38	38	38	38	0	0	0	0
2016	9	52	54	39	41	3	3	10	10
2017	15	57	90	49	78	8	8	0	0
Total	36	169	205	148	180	11	11	10	10

Appendix B

Table A4 presents the TBIJ Pakistan dataset data in an aggregate form and in their distribution over time. The TBIJ Yemen dataset includes UCAV strikes, military operations of the Yemeni and US Special Forces and air strikes conducted by cruise missiles and conventional aircraft. A peculiarity of this dataset is the presence of two columns reporting the minimum and maximum number of operations for every event. In the analysis, we treated every dataset line as a single event ignoring this information. Furthermore, the TBIJ dataset includes strikes in which US responsibility was not confirmed. Indeed, the Yemeni Air Force and the Royal Saudi Air Force also conducted air strikes in Yemen.

As reported in the dataset notes, the strikes were registered as “Confirmed” only if recognized as US operations by a named or unnamed US official, by a named senior Yemeni official, or by three or more unnamed official Yemeni officials in different published sources. The first part of Table A5 presents the data divided by strike type and by Confirmed or Possible US responsibility. The second part of this table shows the distribution over time of the data only for UCAV ‘Confirmed’ strikes. The TBIJ Somalia dataset shares the peculiarities of the Yemen dataset. For our study, we treated every line of the dataset as a single event and analysed only the ‘Confirmed’ strike data. Table A6 shows data in an aggregate form, for UCAV strikes and other military instruments and the division between ‘Possible’ and ‘Confirmed’ strikes. The last section of the table presents the evolution over time of the data for the ‘Confirmed’ strikes conducted by UCAV. The described period extended from 23 June 2011 to 31 December 2017.

Table A4. TBIJ data for Pakistan.

Year	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax
2004	1	6	8	4	6	2	2
2005	3	16	16	5	11	5	11
2006	2	94	105	0	15	90	100
2007	5	36	56	5	45	11	46
2008	38	252	401	115	342	59	173
2009	54	471	753	292	653	100	210
2010	128	755	1108	573	1019	89	197
2011	75	362	666	253	614	52	152
2012	50	212	410	161	397	13	63
2013	27	109	195	105	195	0	4
2014	25	115	186	113	186	0	2
2015	13	60	85	55	83	2	5
2016	3	11	12	10	11	1	1
2017	5	15	22	12	22	0	3
2018	1	1	3	1	3	0	0
Total	430	2515	4026	1704	3602	424	969

Table A5. TBIJ data for Yemen.

Type	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax
UCAV	237	1001	1440	885	1346	94	170
Others	57	467	607	351	502	105	134
Total	294	1468	2047	1236	1848	199	304
Possible	84	361	515	322	489	26	61
Confirmed	153	640	925	563	857	68	109
Year	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax
2002	1	6	6	6	6	0	0
2011	9	46	102	34	66	36	36
2012	32	205	267	193	260	7	12
2013	22	79	129	49	112	17	37
2014	15	90	127	81	123	4	9
2015	21	75	103	69	102	1	7
2016	32	89	122	89	122	0	0
2017	21	50	69	47	66	3	8
Total	153	640	925	568	857	68	109

Table A6. TBIJ data for Somalia.

Type	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax
UCAV	38	313	449	300	446	3	15
Others	42	250	382	236	344	11	57
Total	80	563	831	536	790	14	72
Possible	4	40	62	40	62	0	0
Confirmed	34	273	387	260	384	3	15
Year	Natt	Tmin	Tmax	Mmin	Mmax	Cmin	Cmax
2011	1	2	2	2	2	0	0
2012	2	4	8	3	8	0	1
2013	1	2	3	2	3	0	0
2014	3	10	18	8	18	0	2
2015	7	20	33	18	33	0	4
2016	13	204	292	199	289	3	5
2017	7	31	31	28	31	0	3
Total	34	273	387	260	384	3	15

Appendix C

In the PAKO elaboration process, we merged the strikes reported by the TBIJ for certain dates:

- 6 January 2010 labelled Ob55 e Ob56.
- 10 March 2010 labelled Ob75 e Ob76.
- 6 December 2010 labelled Ob167 e Ob168.
- 6 June 2011 labelled Ob214 e Ob215.
- 11–12 July 2011 labelled Ob227 e Ob228.
- 26 December 2014 labelled Ob356 e Ob357.

Concerning the NAF dataset, we merged the strikes registered on the following dates:

- 18 June 2009.
- 12 July 2011 in the location 'Bray Nishtar, Shawal area in North Waziristan'.

In the YEMO building process, we merged the strikes reported by the TBIJ on the date of 14 October 2011 labelled YEM032 e YEM033 and on date of 19 January 2013 labelled YEM117, YEM118 e YEM119. In the elaboration of SOMO we merged the operations reported by the NAF on the dates of 23 June 2011 and 11 April 2016 and the strikes registered by the TBIJ on the date of 15 July 2015 labelled SOM25 and SOM26.

References

1. Denzin, N.K. *The Research Act: A Theoretical Introduction to Sociological Research Methods*, 3rd ed.; Prentice Hall: Hoboken, NJ, USA, 1989.
2. Franzosi, R. The press as a source of socio-historical data: Issues in the methodology of data collection from newspapers. *Hist. Methods J. Quant. Interdiscip. Hist.* **2013**, *20*, 5–16. [CrossRef]
3. Jack, E.P.; Raturi, A.S. Lessons learned from methodological triangulation in management research. *Manag. Res. News* **2006**, *29*, 345–357. [CrossRef]
4. Trochim, W.M. *The Research Methods Knowledge Base*, 2nd ed.; Version Current as of 20 October 2006. Available online: <http://www.socialresearchmethods.net/kb/> (accessed on 4 July 2021).
5. Denzin, N.K. Triangulation: A case for methodological evaluation and combination. In *Sociological Methods*; Denzin, N.K., Ed.; MacGraw-Hill: New York, NY, USA, 1978; pp. 339–357.

6. Avery, P.; Fricker, M.S. Tracking the predators: Evaluating the US drone campaign in Pakistan. *Int. Stud. Perspect.* **2012**, *13*, 344–365.
7. Johnston, P.B.; Sarbahi, A.K. The impact of US drone strikes on terrorism in Pakistan. *Int. Stud. Q.* **2016**, *60*, 203–219. [CrossRef]
8. Jordan, J. The effectiveness of the drone campaign against Al Qaeda central: A case study. *J. Strateg. Stud.* **2014**, *37*, 4–29. [CrossRef]
9. START—National Consortium for the Study of Terrorism and Responses to Terrorism. Global Terrorism Database 2018. Available online: <http://www.start.umd.edu/gtd/> (accessed on 4 July 2021).
10. Bentley, M. Fetishised data: Counterterrorism, drone warfare and pilot testimony. *Crit. Stud. Terr.* **2018**, *11*, 88–110. [CrossRef]
11. Woolley, J.T. Using media-based data in studies of politics. *Am. J. Political Sci.* **2000**, *44*, 156–173. [CrossRef]
12. Carmines, E.G.; Zeller, R.A. *Quantitative Applications in the Social Sciences: Reliability and Validity Assessment*; SAGE Publications Ltd.: Thousand Oaks, CA, USA, 1979.
13. Weber, J. Keep adding: On kill lists, drone warfare and the politics of databases. *Environ. Plan. D Soc. Space* **2016**, *34*, 107–125. [CrossRef]
14. Barranco, J.; Wisler, D. Validity and systematicity of newspaper data in event analysis. *Eur. Soc. Rev.* **1999**, *15*, 301–322. [CrossRef]
15. GDELT Project. Available online: <https://www.gdeltproject.org/> (accessed on 4 July 2021).
16. Stimson Center. *An Action Plan on U.S. Drone Policy: Recommendations for the Trump Administration*; Stimson Center: Washington, DC, USA, 2018; Available online: <https://www.stimson.org/wp-content/files/file-attachments/Stimson%20Action%20Plan%20on%20US%20Drone%20Policy.pdf> (accessed on 4 July 2021).
17. Penney, J.; Schmitt, E.; Callimachi, R.; Koetti, C. Drone mission, curtailed by Obama expand in Africa. *The New York Times*, 10 September 2018; p. 1. Section A.
18. New America Foundation. America’s Counterterrorism Wars. NAF International Security Program. 2018. Available online: <http://www.newamerica.org/in-depth/americas-counterterrorism-wars/> (accessed on 4 July 2021).
19. Bureau of Investigative Journalism. Drone Warfare 2018. Available online: <https://www.thebureauinvestigates.com/projects/drone-war> (accessed on 4 July 2021).
20. Shane, S. Drone strikes reveal uncomfortable truth: U.S. is often unsure about who will die. *The New York Times*, 24 April 2015; p. 1. Section A.
21. Guion, R.M. Validity and reliability. In *Blackwell Handbooks of Research Methods in Psychology: Handbook of Research Methods in Industrial and Organizational Psychology*; Rogelberg, S.G., Ed.; Blackwell: Oxford, UK, 2002; pp. 57–76.
22. Thurmond, V.A. The point of triangulation. *J. Nurs. Scholar.* **2001**, *33*, 253–258. [CrossRef] [PubMed]