Investigation of US Strikes in Yemen

Jason Scruggs, Aditya Bharath, Dave Le, Angelo Desiderio, Anthony Ko

1 Abstract

This report examines the civilian impact of U.S. strikes in Yemen, focusing on data from 324 instances and 528 total strikes conducted primarily between 2009 and 2019. The study leverages extensive data analysis, including hypothesis testing and exploratory data analysis, to explore trends in civilian casualties, regional disparities in impact, and the lethality of confirmed versus possible strikes. Key findings highlight significant civilian harm, with approximately 15,000 deaths reported during the period, concentrated in contested regions like Abyan, a critical hub for resources and supply routes [1]. The analysis also identifies temporal patterns, revealing higher lethality in early strikes and regional variance in casualty numbers. This report highlights the negative externalities of the war in Yemen.

2 Introduction

The conflict in Yemen has not only devastated the economy and infrastructure of the country, but has also resulted in one of the most severe humanitarian crises the world has seen. With more than 15,000 civilian deaths and millions in need of aid, the human toll of this conflict demands a critical examination of the factors that contribute to the devastation. One such factor has been the US led strikes as part of the counter-terrorism efforts against Al-Qaeda, ISIS, and Houthi's. Although they have helped curb terrorism, strikes have also caused massive damage to the country's citizens, leaving countless dead, injured, and in dire need of help.

Figure 1: Map of Yemen

The strikes, primarily conducted between 2009 and 2019, have dispropor- tionately affected areas such as Abyan and Zinjibar, regions critical for Yemen's oil and gas supply. While military actions aim to eliminate terrorist threats, the high civilian death toll underscores the devastating human cost of these operations. The data collected on over 500 U.S. strikes reveals a troubling pattern, with early strikes being particularly lethal. This report will analyze the civilian impact of these strikes, focusing on the death toll, the geographic spread of casualties, and the temporal trends in the conflict. By exploring these factors, we aim to offer a clearer understanding of the long-term consequences of such military interventions on Yemen's population.

3 Data

The dimensions of the original dataset are 324 rows x 18 columns. The 324 rows count each unique strike ID, with a total of 528 strikes. Below is a table with all listed variables:

StrikeID	Date	Location	Province
TypeOfAttack	ConfirmedOrPossible	AirOperation	DroneStrike
MinimumStrikes	MaximumStrikes	MinimumPeopleKilled	MaximumPeopleKilled
MinimumCiviliansKilled	MaximumCiviliansKilled	MinimumChildrenKilled	MaximumChildrenKilled
MinimumPeopleInjured	MaximumPeopleInjured		

DroneStrike and AirOperation are dummy variables with either 0/1 values. The variables listed as minimum and maximum provide an overarching range of possible kills/injuries since wartime toll counting is difficult. Below is a screenshot of part of the dataset.

	Strike.ID	Date	Location	Province	Type.of.attack	Confirmed..possible.US.attack.	Air.operation.	Drone.strike	Minimum.number.of.strikes	Maximum.number.
1	YEM001	03/11/2002	Unknown	Marib	Drone strike	Confirmed	1	1	1	
2	YEM002	17/12/2009	Al Majala	Abyan	Cruise missile strike	Confirmed	0	0	1	
3	YEM003	17/12/2009	Arhab	Sanaa	US-Yemen ground operation Possible drone strike	Confirmed	0	0	1	
4	YEM004	24/12/2009	Rafd	Shabwa	Airstrike Possible cruise missile strike	Confirmed	1	0	1	
5	YEM005	12/01/2010	Unknown	Shabwa	Yemen ground operation Possible US assistance	Possible	0	0	1	
6	YEM006	15/01/2010	Al Ajashir	Saada	Airstrike	Possible	1	0	1	
7	YEM007	20/01/2010	Erq al Shawan	Marib	Airstrike	Possible	1	0	4	
8	YEM008	31/01/2010	N/A	Across Yemen	US ground operations	Possible	0	0	24	
9	YEM009	14/03/2010	Jeezat al Qotn Moudia district	Abyan	Airstrike	Confirmed	1	0	1	
10	YEM010	15/03/2010	Moudia	Abyan	Airstrike	Possible	1	0	1	
11	YEM011	24/05/2010	Unknown	Marib	Airstrike	Confirmed	1	0	1	
12	YEM012	05/05/2011	Unknown	Shabwa	Drone strike Airstrike with US conventional aircraft	Confirmed	1	1	1	
13	YEM013	03/06/2011	Zinjibar	Abyan	Drone strike	Confirmed	1	1	1	
14	YEM014	10/06/2011	Zinjibar	Abyan	Airstrike Possible drone strike	Possible	1	1	1	
15	YEM015	10/06/2011	Raia	Abyan	Airstrike Possible drone strike	Possible	1	1	1	
16	YEM016	18/06/2011	Jaar	Abyan	Airstrike Possible drone strike	Possible	1	1	1	
17	YEM017	14/07/2011	Wadeea district Mudiya	Abyan	Drone strike	Confirmed	1	1	1	
18	YEM018	14/07/2011	Unknown	Abyan	Drone strike	Possible	1	1	1	
19	YEM019	27/07/2011	Karadeef Zinjibar	Abyan	Airstrike Possible drone strike	Possible	1	1	1	
20	YEM020	01/08/2011	Al Khamila	Abyan	Drone strike Possible Yemen Air Force airstrike	Confirmed	1	1	2	
21	YEM021	24/08/2011	Zinjibar	Abyan	Airstrike	Possible	1	0	1	
22	YEM022	24/08/2011	Arkoub	Abyan	Airstrike Possible drone strike	Possible	1	1	1	
23	YEM023	25/08/2011	Wadi Hassan	Abyan	Airstrike	Possible	1	0	1	
24	YEM024	31/08/2011	Unknown	Abyan	Airstrike	Confirmed	1	0	2	
25	YEM025	05/09/2011	Jaar	Abyan	Airstrike	Possible	1	0	1	
26	YEM026	05/09/2011	Jaar	Abyan	Airstrike	Possible	1	0	2	
27	YEM027	07/09/2011	Mahfed	Abyan	Drone strike	Confirmed	1	1	2	
28	YEM028	21/09/2011	Mahfed	Abyan	Drone strike	Confirmed	1	1	1	
29	YEM029	21/09/2011	Shaqra	Abyan	Airstrike	Possible	1	0	1	
30	YEM030	30/09/2011	Unknown	Jawf	Drone strike	Confirmed	1	1	1	
31	YEM031	05/10/2011	Al Arqoub	Abyan	Airstrike Possible drone strike	Possible	1	1	1	
32	YEM032	14/10/2011	Azzan	Shabwa	Drone strike	Confirmed	1	1	1	
33	YEM033	14/10/2011	Azzan	Shabwa	Drone strike	Confirmed	1	1	1	
34	YEM034	14/10/2011	Unknown	Southern Yemen	Airstrike Possible drone strike	Possible	1	1	1	
35	YEM035	14/10/2011	Azzan	Shabwa	Airstrike Possible drone strike	Possible	1	1	1	
36	YEM036	22/12/2011	Zinjibar	Abyan	Airstrike Possible drone strike	Possible	1	1	1	

Figure 2: Snapshot of the Data

4 Methodology

1. There were no missing values in our dataset, however there was one year mislabeled as 2002 which required deletion.
2. For variables like deaths and injuries which had a minimum and a maximum, we took the arithmetic mean to get a singular value for analysis.
3. Our focus for the research questions was civilian deaths.
4. For research questions, we used a combination of R, Tableau, and Python for geographical data analysis.

5 Research Questions and Findings

5.1 Where did the attacks occur?

We conducted an analysis of the attack locations for the major provinces in Yemen. Our findings were concurrent with historical context about the war.

1. The province of Abyan and it's capital city of Zinjibar has been a highly contested territory over the years. Abyan serves as a key supply route for oil and gas, which makes it an appealing region for either party to control. Additionally, it is also very natural resource rich. In this region, there is also a constant power vacuum created by the ever-changing local government.

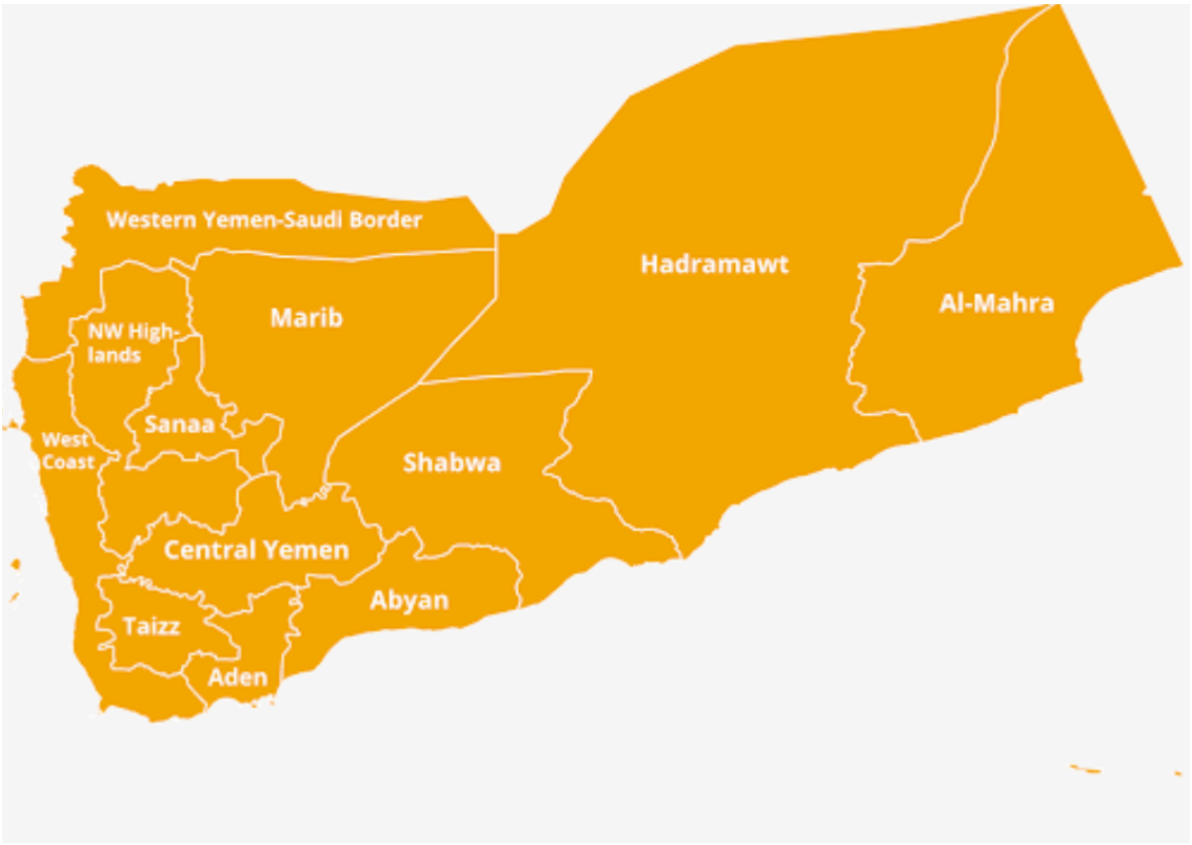

Figure 3: Map of Yemen

2. Sanaa is the largest city (and capital) in Yemen with almost 2 million residents is Sanna but has the lowest amount of civilian casualties. The Houthis have fortified the city, making it a strategic stronghold. Although Sanaa has been targeted by airstrikes, its status as a key base for the Houthis may have resulted in more calculated attacks to

avoid excessive civilian casualties. Additionally, there is the possibiility that being the capital city, it received better aid and attention compared to other regions.

Average number of civilians killed by year.

	Province								
Year of Date	Abyan	Bayda	Damar	Hadhra..	Jawf	Marib	Multiple provinces	Sanaa	Shabwah
2009	44.00								
2010	10.00					3.50			
2011	42.50								3.50
2012	28.00	15.00		2.00					4.50
2013		12.00	2.00	4.50	1.00	5.50		2.00	
2014	1.00	16.50		1.00		0.50			9.00
2015				2.50					1.50
2016				3.00					
2017		25.00				5.00	1.00		5.50
2018		2.50		4.50					7.50

Figure 4: Civilian deaths by region and year

5.2 When did the damage occur

By 2012, the majority of injuries occurred (69%) and a similar trend was observed in civilian deaths (54%). However, attacks took the longest to accumulate, reaching only 35%. Our data suggest that early strikes were significantly more lethal. This is potentially due to the ebb and flow of the war, along actions taken by the international community to slow down the war.

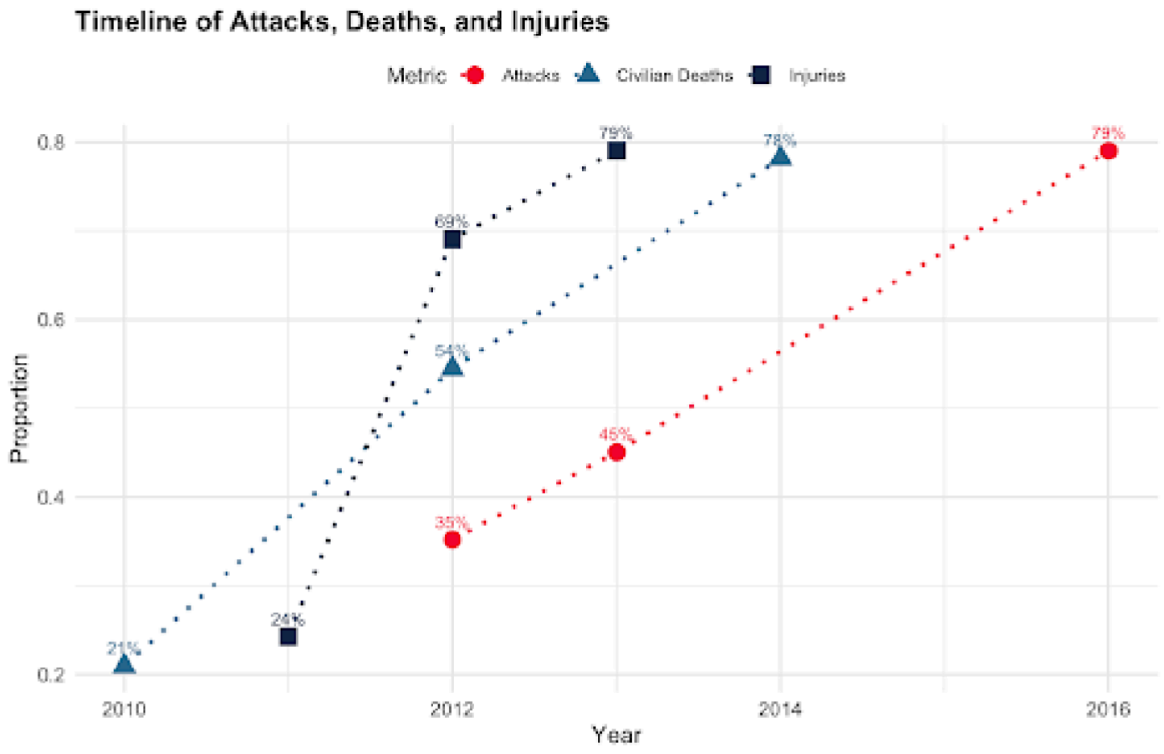

Figure 5: Cumulative density timeline of war casualties

5.3 Does the type of attack make a difference in injuries and deaths?

Non-airstrike attacks averaged more civilian deaths, but were heavily rightskewed. We conducted a two-sample t-test to determine whether the mean injuries and deaths of civilians were impacted based on the type of attack. The test yielded a p-value > 0.05 and showed no significant difference in deaths between airstrikes and non-airstrikes. It is good to note that cruise missile attacks caused the most injuries and deaths per incident as well as drone strikes being a common feature across various attack types, showing a possible dominance/preference in air-based operations.

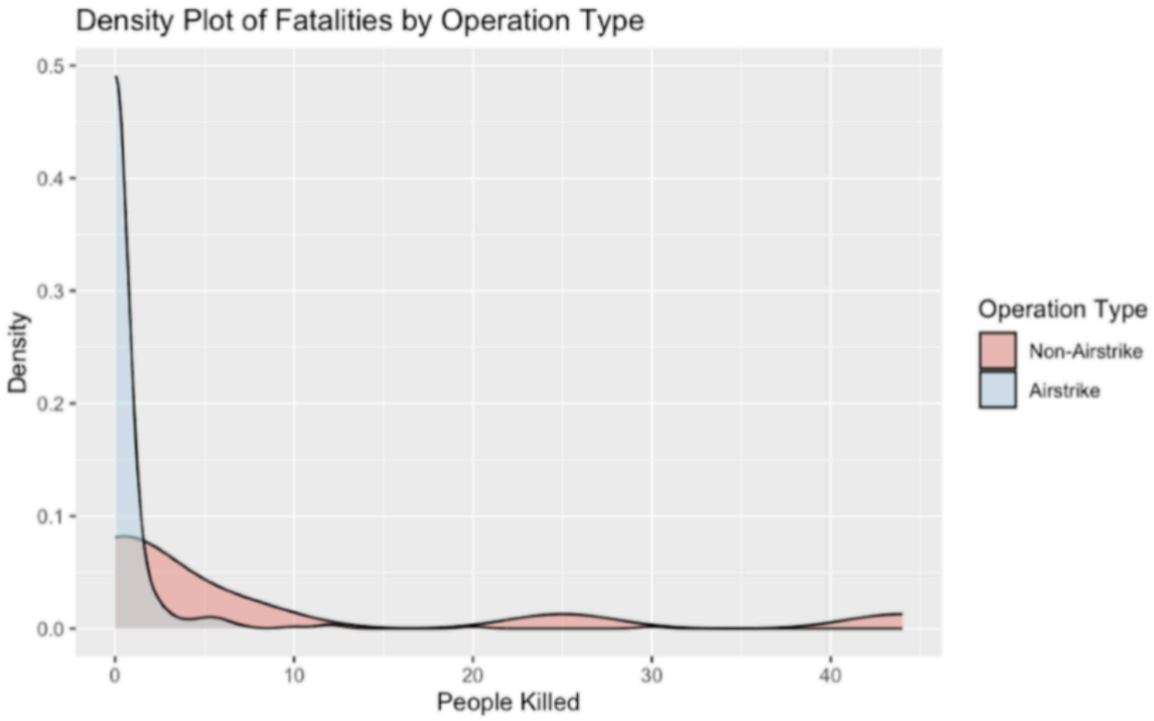

Figure 6: Airstrike vs Non-Airstrike Death Tally

5.4 Does "Confirmed" or "Possible" Classification increase variance of the kill range?

Interested on whether the "Confirmed" and "Possible" classifications had a difference in the variance of the kill range, we performed an F-test to test the variances between the two groups. The null hypothesis for this test assumes that the variances in the range of the death counts in these two groups are equal, while the alternative hypothesis posits that the variances differ. The range of deaths was calculated using the maximum minus the minimum deaths for each recorded attack. The test yielded an F-statistic of 0.1348 and a corresponding p-value of 2.2e^16, which indicated that there is sufficient statistical evidence to reject the null hypothesis. The observed difference in variances between the "confirmed" and "possible" classifications should be attributed to an inherent difference in the variability of death counts between the two groups instead of random variation.

6 Conclusion

Our analysis revealed that U.S. strikes that were carried out earlier in the decade were statistically more deadly for civilians, despite not being the primary targets of conflict. Furthermore, the focus of the strikes was on capturing land for the interested parties and gaining power, which in this case was Al-Qaeda. Deeper analysis and additional data would be required to uncover potential motives.