Civilian Impact of U.S. Drone vs. Non-Drone Strikes in Somalia and Yemen

Assessing Humanitarian Impact with Count Regression Models.

Shanmei Wanyan[1], Daniel Dai[1], Keivan Bolouri[1], Itaru Fukushima[1], Linxue Guo[1] and Evelyn Isaka[1]

1. University of California, Los Angeles, Department of Statistics and Data Science, 8125 Math Sciences Bldg, Los Angeles, CA 90095

NOAA Fisheries OpenSci
Tools for Open Science
https://keivanbolouri.github.io/finalProject140XP/

Table of contents

List of Figures

List of Tables

1 Abstract

Since 2002, the United States has conducted largely hidden counterterrorism campaigns in countries such as Somalia and Yemen, raising ongoing concerns about their humanitarian impact. This project asks how the characteristics and civilian costs of U.S. strikes differ between these two theaters of war. Using open-source strike records compiled by independent monitoring organizations, including the Bureau of Investigative Journalism, We construct a combined dataset of U.S. actions in Somalia and Yemen and analyze casualty patterns with negative binomial regression. The analysis tests three hypotheses: whether civilian casualty rates differ by country, whether drone strikes have different effects across countries, and whether reporting uncertainty varies between regions. The results show that, controlling for strike characteristics and total fatalities, strikes in Yemen are associated with nearly five times the civilian casualties of strikes in Somalia. By contrast, there is no evidence that the impact of drone strikes on civilian harm differs between the two countries, nor that overall reporting uncertainty is systematically higher in one region than the other. However, uncertainty is greater for drone and unconfirmed strikes and lower when U.S. involvement is confirmed. These findings underscore the unequal humanitarian burdens across theaters of U.S. counterterrorism and highlight the need for more transparent and consistent casualty reporting.

2 Introduction

2.1

Since 2002, the United States has waged a largely clandestine drone war in countries such as Yemen and Somalia, often far from public scrutiny [1]. Although these counterterrorism strikes aim to eliminate militant targets while minimizing risk to U.S. personnel, their humanitarian consequences remain a pressing concern. The cost to civilian life can be substantial. For example, an investigation found that roughly one-third of those killed by U.S. drone strikes in Yemen in 2018 were likely civilians or pro-government allies [2].

This research addresses a central question: **Do the characteristics and human costs of U.S. counterterrorism strikes differ between Somalia and Yemen—and if so, how?** Understanding such differences is important both theoretically and practically.
Theoretically, comparing two distinct drone theaters can reveal how local conditions—such as insurgent behavior, intelligence quality, and terrain—shape strike outcomes. Practically, identifying where drone operations are less effective in sparing civilians can guide improvements in targeting procedures, transparency, and accountability mechanisms.

These operations are often carried out "out of sight," yet their humanitarian consequences are very real [1]. Official reporting has historically underestimated civilian casualties, prompting independent organizations to investigate and publish alternative estimates [2]. For instance, the U.S. government once claimed only 64–116 civilian deaths from drone strikes outside declared warzones between 2009 and 2015, whereas independent monitors estimated several times more [2].

In response to these discrepancies, numerous efforts have emerged to document the drone war's toll. Pitch Interactive's *Out of Sight, Out of Mind* visualization cataloged CIA drone strikes and casualties in Pakistan [1]. The Economist released infographics demonstrating large gaps between official and independent casualty estimates. UCLA's **Drone Wars** project created a cross-country dataset covering Afghanistan, Pakistan, Somalia, and Yemen using records from the Bureau of Investigative Journalism (BIJ) [3,4].

These initiatives highlight the need for **rigorous, comparative analysis**. Yet no study has systematically compared Somalia and Yemen with respect to strike characteristics and humanitarian outcomes. This paper fills that gap by leveraging detailed open-source strike records from both countries to quantitatively assess differences in civilian harm.

We explicitly test three hypotheses:

1. **Hypothesis 1 – Civilian Harm Difference:**
   Somalia and Yemen differ in their civilian casualty rates.
2. **Hypothesis 2 – Drone Effectiveness Across Countries:**
   The effect of drone strikes on civilian casualties differs between Somalia and Yemen.
3. **Hypothesis 3 – Reporting Uncertainty:**
   Casualty reporting uncertainty differs between regions.

To evaluate these hypotheses, we construct a comprehensive dataset of U.S. counterterrorism strikes in Somalia and Yemen from independent monitoring organizations such as BIJ [1]. Because fatality reporting is often uncertain, we use minimum–maximum casualty ranges [1]. Civilian casualty counts exhibit strong overdispersion, so we employ negative binomial regression to estimate the effects of region and strike characteristics. This modeling framework allows us to determine whether "country" remains a significant predictor of civilian harm once contextual factors are controlled for.

3 Literature Review

Researchers and monitoring groups have spent many years examining how many people are killed in U.S. drone strikes, but most work focuses on one country at a time rather than comparing Somalia and Yemen directly.

Columbia Law School's Human Rights Clinic, in *Counting Drone Strike Deaths*, shows that official U.S. numbers often underestimate civilian deaths. They recommend using casualty ranges (minimum–maximum) because information from the ground is often unclear [**3**].

The Bureau of Investigative Journalism (BIJ) collected open-source reports for every known strike in Yemen, Somalia, Pakistan, and Afghanistan. Their database records both minimum and maximum death counts and distinguishes civilians from militants when possible, noting that reports are often uncertain or contradictory [**5**].

New America's *Counterterrorism Wars* project compiles strike data from Yemen and Somalia, listing total strikes and casualty ranges and explaining how they classify victims when reports are vague or disputed [**6**].

Together, these sources show that:

1. Independent groups usually find **more civilian deaths** than official U.S. reports.
2. Although detailed data exist for Yemen and Somalia, most previous analyses summarize each country separately rather than compare them statistically.

Our study fits into this work by using open-source strike records to conduct a direct, quantitative comparison between Somalia and Yemen. Using negative binomial regression, we test whether the countries differ in civilian casualty rates and the uncertainty of reported casualties, controlling for strike characteristics.

4 Methods

4.1 Statistical Methods

4.1.1 Hypothesis Tests

4.1.2 Hypothesis 1: Civilian Harm Difference

H_0 : Drone strikes have the same civilian impact in Somalia and Yemen.
H_1 : Drone strikes have different civilian impacts across the two regions.

To test whether Somalia and Yemen differ in civilian casualty outcomes, we estimate the model:

$$E(\text{civilian casualties}) = \beta_0 + \beta_1(\text{region}) + \beta_2(\text{drone}) + \beta_3(\text{US confirmed}) + \beta_4(\text{minimum strikes}) + \beta_5(\text{total killed})$$

Table 4.1: Variable Definitions

Variable	Description
Civilian casualties	Number of civilians reported killed in the strike (outcome variable).
Region	Country where the strike occurred (Somalia or Yemen).
Drone	Indicates whether the strike was carried out by a drone (1 = drone).
US confirmed	Whether the strike was officially confirmed by the U.S. government.
Minimum strikes	Minimum number of strike events associated with the record.
Total killed	Minimum number of total fatalities (civilians + militants).

4.1.3 Hypothesis 2: Drone Effectiveness Across Countries

H_0 : Drone use affects civilian casualties in the same way in both Somalia and Yemen.
H_1 : Drone use affects civilian casualties differently across Somalia and Yemen.

To evaluate whether the civilian impact of drone strikes varies by region, we include an interaction term between drone use and region:

$$\begin{aligned} E(\text{civilian casualties}) = \beta_0 + \beta_1(\text{drone}) + \beta_2(\text{region}) + \beta_3(\text{drone} \times \text{region}) \\ + \beta_4(\text{US confirmed}) + \beta_5(\text{minimum strikes}) + \beta_6(\text{total killed}) \end{aligned}$$

4.1.4 Hypothesis 3: Reporting Uncertainty Difference

H_0 : Reporting uncertainty does not differ between Somalia and Yemen.
H_1 : Reporting uncertainty differs between Somalia and Yemen.

To assess whether casualty reporting uncertainty differs between regions, we model the uncertainty metric. We modeled casualty reporting uncertainty (defined as `max_killed - min_killed`) region and strike characteristics as predictors.

$$E(\text{Uncertainty in casualties}) = \beta_0 + \beta_1 \text{ Region} + \beta_2 \text{ Drone} + \beta_3 \text{ US confirmed} + \beta_4 \text{ Minimum strikes}$$

where:

$$\text{Uncertainty in casualties} = \text{Maximum killed} - \text{Minimum killed}$$

5 Model Selection

Because our prediction variable is a count—specifically, the number of civilians killed in each strike—we use statistical models designed for count data. A natural starting point is the **Poisson regression**, which assumes that the mean and variance of the outcome are equal $E(x) = \mathrm{Var}(x)$. However, in our dataset the variance is much larger than the mean, a condition known as **overdispersion**. When overdispersion is present, Poisson regression underestimates the true variability and produces misleadingly small standard errors. To address this, we use a **negative binomial regression**, which adds a dispersion parameter that allows the variance to exceed the mean. This makes the negative binomial model much better suited for modeling drone-strike casualty counts and provides more reliable estimates of how factors such as region, drone use, and confirmation status relate to civilian harm.

```
mean_civ <- mean(combined_model$civilian_casualties)
var_civ  <- var(combined_model$civilian_casualties)

c(mean = mean_civ, variance = var_civ)
```

```
     mean  variance
0.4338521 8.0043879
```

6 Author Contributions

Author1 designed the research. Author2 carried out all simulations, analyzed the data. Author1 and Author2 wrote the article.

7 Acknowledgments

We thank G. Harrison, B. Harper, and J. Doe for their help.

8 References