Conflict Event Data and Beyond

Ghassan Baliki, Ph.D

World Bank Training Workshop - ISDC,
Perugia, 20 March 2018
Outline

1. Introduction
2. Conflict Event Datasets
   - The Basics
   - Existing Conflict Event Datasets
3. Reliability of Event Data
   - Media Bias from Secondary Sources
   - Crowd-Seeding Event Data
4. Geographic Precision and Matching
   - The Significance of Geography
   - Matching Conflict Data with Household Surveys
5. Conclusion
Research requires more fine-grained data on conflict to better:

- analyze the exposure to violence at the household and individual level
- understand the dynamics behind violence and wars (prediction)
- provide meaningful measures for the quantification of conflicts
Why Do We Need Data on Conflict (Policy)

- Policy also requires better data on conflict to:
  - Inform development and humanitarian organizations (Mapping and targeting)
  - provide a better decision mechanism (peace processes)
  - facilitate interventions during wars and prescribe sound post-conflict policies
Aim

- Introduce the concept of geo-referenced conflict event data
- Introduce existing conflict event datasets
- Examine the reliability of conflict event data sources
- Provide an overview of various methods on the collection of conflict event data
  - Traditional (media sources) versus innovative (crowd-seeding)
- Test the usefulness of conflict event data (visualization and matching)
Outline

1. Introduction

2. Conflict Event Datasets
   - The Basics
   - Existing Conflict Event Datasets

3. Reliability of Event Data
   - Media Bias from Secondary Sources
   - Crowd-Seeding Event Data

4. Geographic Precision and Matching
   - The Significance of Geography
   - Matching Conflict Data with Household Surveys

5. Conclusion
Outline

1. Introduction

2. Conflict Event Datasets
   - The Basics
   - Existing Conflict Event Datasets

3. Reliability of Event Data
   - Media Bias from Secondary Sources
   - Crowd-Seeding Event Data

4. Geographic Precision and Matching
   - The Significance of Geography
   - Matching Conflict Data with Household Surveys

5. Conclusion
Temporal- and geo- coded data at the micro-level
Data are mainly collected from media sources
Definitions of “conflict events” vary (and it makes a difference)
Methodologies are simple (collection and quantification)
There exists a bank of information of conflict event data (ACLED, UCDP-GED, ...)
The number of studies using quantified conflict event data has risen sharply
Conceptualizing an “Event”

- Each Event constitutes an observation ⇒ each armed conflict can generate up to thousands of events
- But what constitutes an event?
  - political violence, civil conflicts, violence against civilians, state violence, …
  - protests and riots, militia presence and interactions, communal violence, …
  - ceasefires, truce, peace negotiations, conflict resolutions, …
- Breakdown of an event: action, actor, mean, outcome, time, location, …
Defining an “Event”

- e.g. “the incidence of the use of armed force by an organized actor against another actor, or against civilians, resulting in at least one direct death in either the best, low or high estimate categories at a specific location and for a specific time” (Sundberg, et al, 2010: definition of event - UCDP)
- So what about violent events that do not result in deaths, but displace hundreds of people
- or events that simply contain any type of violence?
Defining an “Event”

- e.g. “a rebel group attacked a livestock farm along Maramvya’s 15th Avenue, very near Bujumbura, the capital, stealing about 15 cows. They wounded four cows, including three calves. The other animals were not that lucky as the rebels shot dead 17 cows.” (Eck, 2012: on ACLED)

- Again: Back to the conceptual framework (what constitutes an armed conflict)
Operationalizing an “Event”

- Putting numbers to events (quantification)
  - e.g., Intensity of violence, number of deaths, dwellings destroyed, number of displaced
    - classification of variables (actors, outcomes, etc)
    - provision of weights (a sniper incident in Aleppo vs. a massacre in Duma)
    - precision of location and time
Geographic Precision

- How we handle geographic data impacts the results
- GPS coordinates at the lowest spatial level, but low matching with exposure from Household surveys.
  - Good and bad news: News reports don’t specify GPS coordinates
  - Need to use proxies (capital, centroid of districts and cities, etc)
- Advantage and disadvantages in choosing the right precision levels
This allows us to understand the dynamics of the violence.

Most datasets usually specify the exact date of an event, however, at other times we only know the week or month that an event took place in.

Lack of time precision (difficult to determine)

Both spatial and temporal precision levels allows us be able to identify the conflict

and enable matching with other types of datasets (such as household surveys)
Choice of Datasets

- Carefully assess the need/use for the dataset
  - violence versus other conflict actions
  - precision of data versus its widespread
  - local versus country level
  - policy versus research
Outline

1. Introduction

2. Conflict Event Datasets
   - The Basics
   - Existing Conflict Event Datasets

3. Reliability of Event Data
   - Media Bias from Secondary Sources
   - Crowd-Seeding Event Data

4. Geographic Precision and Matching
   - The Significance of Geography
   - Matching Conflict Data with Household Surveys

5. Conclusion
Existing Datasets

- Armed Conflict Location Event Database (ACLED)
- Uppsala Conflict Data Program (UCDP-GED)
- UCDP / PRIO
ACLED generates disaggregated conflict event data
- in Sub-Saharan Africa, South (-East) Asia, and recently in the Middle East and North Africa
- covers period from e.g., 1997 - current (updated on a monthly basis)

Uses mainly secondary sources
- international and local media / news reports
- publications from civil society
- security updates from international organizations

Applies human coding (coders read the reports and code them into the dataset)
ACLED - Codebook

- Types of events are based on a pre-designed codebook
- The codebook directs the coders on what to include and exclude
- e.g.: Battles (government, rebel groups); violence against civilians; riots/protests; remote violence
- Fatalities and other generic information of an event
Hence for an event to be coded in the final database:

- An event has to occur
- Secondary media sources should observe the event
- News reporting editors deem the event newsworthy (and report it)
- Coders access the specific news source
- Coders deem that the event matches the codebook specifications
- Coders quantify the event and add it into the database
ACLED - Data Access

From: 01/03/2016
To: 15/04/2017
Event Type: ~ALL~
Actor Type: All
Actor: All
Region: All
Country: All
Location: All
Keyword:
Export Type: □ Actor Based □ Compatibility Mode

Please note: Data export tool does not function well in Safari
ACLED - Example

Leaflet | Map tiles by Carto, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

- Battles
- Violence against civilians
- Remote violence
- Riots/protests
- Other

Ghassan Baliki, Ph.D

Conflict Event Data and Beyond
UCDP-GED - Dataset

- Similar to ACLED, however it has larger regional coverage
- This comes at the expense of the types of conflict and events coded
- Definition of an event: An incident of violence between two armed actor, or against civilians which results in at least 1 death at a specific date and specific time.
- Highest spatial precision is the “village” / and time is the “day”.
UCDP/PRIO Dataset

- The data is at the macro-level (country-level)
- Provides information if a country in a specific year is in conflict
- Onset versus intensity
- Battle-deaths dataset
Outline

1. Introduction
2. Conflict Event Datasets
   - The Basics
   - Existing Conflict Event Datasets
3. Reliability of Event Data
   - Media Bias from Secondary Sources
   - Crowd-Seeding Event Data
4. Geographic Precision and Matching
   - The Significance of Geography
   - Matching Conflict Data with Household Surveys
5. Conclusion
Introduction
Conflict Event Datasets
Reliability of Event Data
Geographic Precision and Matching
Conclusion

Outline

1 Introduction

2 Conflict Event Datasets
   - The Basics
   - Existing Conflict Event Datasets

3 Reliability of Event Data
   - Media Bias from Secondary Sources
   - Crowd-Seeding Event Data

4 Geographic Precision and Matching
   - The Significance of Geography
   - Matching Conflict Data with Household Surveys

5 Conclusion
Media put into test: Case of Syria

- We collect data from multiple secondary media sources
- Creation of online user friendly entry forms for human coders
- 60 days chosen randomly between March 2011 and June 2013
- All information on violent and political events are coded, per codebook
Coding of violent actions is divided into three consecutive segments:

- **Forms**: violence between armed actors; military offensives; violence against civilians.
- **Means**: mostly focusing on the use manufactured arms
- **Outcomes**: Death; destruction; and territorial gain (loss); displacement
Variation between Media Sources

- BBC Monitoring
- Al-Arabiya
- Agence France Press
- Associated Press
- Syrian Arab New Agency

Legend:
- under 10
- 10 - 50
- 50 - 100
- 100 - 250
- over 250
Excluding the Middleman

Two approaches have been previously utilized: Crowd-sourcing and Crowd-seeding

- Crowd-sourcing: Individuals freely participate in sharing information, usually to an online platform
- Crowd-seeding: A selection of participants (reporters) share specific information
Crowd-seeding Design

- Random Distribution of reporters into certain geographic areas (small areas of coverage)
- To combat biases and assure reliability
  - measure the geographical distance between the event reporter and the event
  - measure the degree of observation/information
- Reduce risk by tracking sensitivity of publishing events
Crowd-seeding Methodology

- Create platform and open-source survey tools
  - available and mostly free, yet fragile
- Use more expensive secure tools
- Conditional forms or more simplistically sms (quantity versus quality)
- Immediately stored, cleaned, and uploaded to a shared database
- Readily accessible
Generic Events - The Codebook

- Geo-location of an event
  - divided between cities and other geo-areas
  - constitutes of 6 precision levels (down to city sub-district level)
- Date and temporal precision levels
- Sources of information: trust and reliability for second degree observations
- Sensitivity to reduce risk
Peace Events - The Codebook

- Ceasefire and Peace resolution at the village level
  - Invitation and rejection to negotiate
  - Negotiation
  - Agreement, violation, and breakdown of a ceasefire or peace resolution
- Micro-political events are harder to capture versus macro events
# Main Advantages and Disadvantages

<table>
<thead>
<tr>
<th></th>
<th>C-seeding</th>
<th>C-sourcing</th>
<th>Traditional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-retrospective</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Elimination of media bias</td>
<td>✓</td>
<td>✓</td>
<td>X</td>
</tr>
<tr>
<td>Elimination of reporting fatigue</td>
<td>✓</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>More detailed and micro-level</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Codebook-oriented</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Validity and reliability</td>
<td>✓</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Less Risky</td>
<td>X</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Less Costly</td>
<td>X</td>
<td>✓</td>
<td>X</td>
</tr>
</tbody>
</table>
Outline

1 Introduction

2 Conflict Event Datasets
   - The Basics
   - Existing Conflict Event Datasets

3 Reliability of Event Data
   - Media Bias from Secondary Sources
   - Crowd-Seeding Event Data

4 Geographic Precision and Matching
   - The Significance of Geography
   - Matching Conflict Data with Household Surveys

5 Conclusion
Geographic Precision

- How we handle geographic data impacts the results
- Most geographic data is represented by spatial coordinates (lon and lat)
- It can be aggregated to various units
  - country; district; village; enumeration area;
- Arbitrary choice of the unit of analysis (moving beyond points)
- Advantage and disadvantages in choosing the right precision levels
Violence Spread - GPS Points
Aggregating - The Grid Level

The Significance of Geography
Matching Conflict Data with Household Surveys
Importance of Geography

- Using low spatial units helps us understand how conflict is correlated with many geographic factors
  - population densities
  - agricultural production and land use
  - road access and borders
  - weather variables
- Moreover, this helps in determining how households living in these spatial units are exposed to various types of violence
Deviation in Altitude
Population Density and Violence
Land Use (Farms) and Violence
Primary Roads and Violence
Outline

1. Introduction
2. Conflict Event Datasets
   - The Basics
   - Existing Conflict Event Datasets
3. Reliability of Event Data
   - Media Bias from Secondary Sources
   - Crowd-Seeding Event Data
4. Geographic Precision and Matching
   - The Significance of Geography
   - Matching Conflict Data with Household Surveys
5. Conclusion
Various matching techniques

- Number of fatalities per population; onset of violence; intensity of violence...
- However, it is important to determine the needs
  - For (over-)sampling
  - For mapping
  - For analysis
Traditionally, one takes the number of battle-related death per population density in a given district. This is then used as a proxy to exposure to households to conflict in this district. Advantages: easier to measure. Disadvantages: ignores the spatial dimension.
Example from Mexico on Crime

The rate of drug war deaths December 2006-December 2010 per 100,000 population

Ghassan Baliki, Ph.D
Another approach is to check all conflict events (not just fatality-related).

We can check the distributions of both the sample and the events.

Methodologically, there are many ways to approach this, e.g., classify conflict intensity (quartiles) and match with households.
North East Nigeria - spatial distribution
LSMS Nigeria and event data
LSMS Nigeria - EA Buffer

The Significance of Geography
Matching Conflict Data with Household Surveys

Introduction
Conflict Event Datasets
Reliability of Event Data
Geographic Precision and Matching
Conclusion

Ghassan Baliki, Ph.D
Conflict Event Data and Beyond
Outline

1. Introduction
2. Conflict Event Datasets
   - The Basics
   - Existing Conflict Event Datasets
3. Reliability of Event Data
   - Media Bias from Secondary Sources
   - Crowd-Seeding Event Data
4. Geographic Precision and Matching
   - The Significance of Geography
   - Matching Conflict Data with Household Surveys
5. Conclusion
How far can/should we go?

- micro-level conflict event data: there is still much to be done
- improved design, meaningful measures, and transparent methodologies
- How really is it needed? and for what purposes?
  ...Opinions
THANKS FOR YOUR ATTENTION