Database on Suicide Attacks Codebook

October 2020

I. The Database on Suicide Attacks

A. Project Overview

The Chicago Project on Security and Threats Database on Suicide Attacks (DSAT), the foundation upon which the institute was built, comprises the most complete list of suicide attacks since 1974 currently available. We are proud to make our data freely available to the public. This allows users to review and analyze the complete set of suicide attacks independent of CPOST's analysis and findings.

The database includes information about the location of attacks, the target type, the weapon used, and systematic information on the demographic and general biographical characteristics of suicide attackers. The database expands the breadth of the data available in English by using native language sources (e.g., Arabic, Hebrew, Russian, Tamil) that are likely to cover information that secondary sources do not. DSAT data is collected from multiple sources including newswires, tweets, and video sources, and is coded for over 70 variables. This codebook will expound upon the uses and coding behind each of the individual variables accounted for in the DSAT.

B. Citing the DSAT

The use of CPOST's DSAT data should be acknowledged using the following citation:

Chicago Project on Security and Threats (CPOST). 2021. Database on Suicide Attacks (October 2, 2020 Release). [Data File]. Retrieved from http://cpost.uchicago.edu/

II. Resources and Collection

A. Collection Sources

To identify and verify suicide attacks, CPOST uses a comprehensive three-step collection process followed by a two-step verification process. Attack data is first collected using optimized search strings through the database ProQuest. In the second round of data collection, researchers utilize customized "Google Alerts" which scrape the web daily for any news of suicide attacks. Finally, researchers conduct comprehensive regional searches using multiple search engines and platforms including social media accounts (e.g. Twitter), blogs, live maps, militant group websites, martyr videos, and more to locate claims for attacks and attacker biographical data not reported in the media. In previous years, archives such as LexisNexis and OpenSource.gov were included in the collection process. CPOST regularly expands and edits its collection based on newly discovered information.

After the three-step collection process, data is hand-verified by two senior researchers. These researchers scrutinize the data of each individual attack, checking for mistakes or updates independently before resolving discrepancies between their findings. Through this process,

CPOST consistently provides the most reliable suicide attack data available. Every attack is looked at by at least three individual researchers, and often by more than three. Attacks are frequently reverified to ensure the most up-to-date and reliable coding.

B. Confirmed Attacks vs. Possible Attacks

To be included in the database as a confirmed suicide attack, an attack must meet two criteria:

1. At least one attacker must kill him or herself to kill others.
2. The attack must be verified by at least two independent sources.

CPOST separates all potential suicide attacks into two categories: confirmed attacks and possible attacks. To be counted as a confirmed suicide attack, the attack must be reported by two independent sources. These do not include two sources that gain information from the same newswires. For instance, if two sources both use a Reuters newswire as the basis of their report, the sources will not be considered independent. Group claims also serve as an independent source for attacks.

CPOST also collects data on possible suicide attacks. Possible attacks fall into three categories: attacks with only one source, attacks that only appear in group claims, and attacks where news sources conflict as to whether the attack was a suicide or not. CPOST collects possible suicide attacks because information may become available in the future to confirm the attack. CPOST periodically reviews possible attacks. If, upon review, new information confirms a possible attack as suicide, the attack is updated in the database as confirmed. The opposite is also true: if new evidence reveals that an attack no longer meets CPOST's criteria for inclusion, it is removed.

Failed suicide attacks – the attacker does not kill him or herself – and suicide missions – attacks where the attacker dies, but not by his or her own hand – are not collected.

Double verification of the attack is crucial: it greatly reduces bias from any single source to ensure the most accurate record of suicide attacks. Accordingly, users of CPOST data can have high confidence that the attacks in the Database on Suicide Attacks did occur.

C. Sources of Potential Bias

Data on suicide attacks comes from a variety of sources, including governments, media, and militant groups. Each has its own potential for bias. Governments may have an incentive to over-report attacks as suicide to mobilize support for costly counterinsurgency measures, or to under-report such attacks to create a false impression of competence and security. Media reports might be based on preliminary evidence or rumors. Militant groups routinely over-report suicide attacks and the casualties they inflict to mobilize support or create a perception of strength. Requiring a

minimum of two sources increases our confidence that the suicide attacks in the DSAT are accurate and complete.

III. What is Considered a Suicide Attack?

A. Definition of a Suicide Attack

CPOST defines a suicide attack as an attack in which an attacker kills himself or herself in a deliberate attempt to kill others. CPOST includes only suicide attacks perpetrated by non-state actors; attacks authorized by national governments are not included. The classic example is a suicide bomber detonating an explosive vest (a "belt bomb") or explosives in a vehicle the bomber is driving (a "suicide car bomb"). The critical criteria is suicide: the attacker must kill him or herself, even if no one but the attacker dies in the attack. The CPOST-DSAT does not include (1) failed suicide attacks where explosives do not detonate or are detonated by someone other than the attacker (e.g. the explosives were set off by a gunshot from police); or (2) "suicide missions," where the attacker expects to be killed while killing others, but does not directly kill himself or herself.

B. A Note on "Terrorism"

Suicide attacks are often associated with "terrorism" in the media, policy, and academia. The boundaries of what counts as "terrorism," however, is notoriously difficult to define. CPOST is agnostic on whether a given suicide attack qualifies as an act of terrorism or might better be described as some other form of violence (e.g., an attack with no obvious political motivation). Beyond the qualification that the attack must be carried out by a non-state actor, any attack meeting the definition of suicide attack and our two-source requirement is included.

IV. DSAT Variables and Coding Guidelines

A. DSAT Attacks (dsat_attacks)

Category	Field Name	Data Type	Description
Dataset	data_release	DATE	Date when data was exported from the DSAT
Attack Information	event_id	INT	Unique attack identifier
	summary	LONGTEXT	Attack's summary produced by CPOST
	status	CHAR	Status of the attack: Confirmed Suicide or Possible
	ct_sources	INT	Number of sources found for each attack
	event_date	DATE	Date of the attack
	date_year	INT	Year of attack
	date_month	INT	Month of attack
	date_day	INT	Day of attack
	wounded_low	INT	Low estimate of wounded people
	wounded_high	INT	High estimate of wounded people
	killed_low	INT	Low estimate of fatalities
	killed_high	INT	High estimate of fatalities
	killed_low_civilian	INT	Low estimate of civilian fatalities
	killed_high_civilian	INT	High estimate of civilian fatalities
	killed_low_political	INT	Low estimate of political fatalities
	killed_high_political	INT	High estimate of political fatalities
	killed_low_security	INT	Low estimate of security fatalities
	killed_high_security	INT	High estimate of security fatalities
	ct_belt_bomb	INT	Number of belt bombs used
	ct_truck_bomb	INT	Number of truck bombs used
	ct_car_bomb	INT	Number of car bombs used
	ct_weapon_oth	INT	Number of other weapons used
	ct_weapon_unk	INT	Number of unknown weapons used
	weapon_id	INT	ID code for weapon used in the attack
	weapon_txt	CHAR	Weapon (car bomb, belt bomb etc.) used
	weapon_type_txt	CHAR	Type of weapon used
	cbrn	CHAR	Chemical, biological, radiological or nuclear warfare used
UCDP Connectors	ucdp_conflict_sgv_1_id	INT	Conflict ID as used in UCDP/PRIO Armed Conflict Dataset
	ucdp_conflict_sgv_2_id	INT	Conflict ID as used in UCDP/PRIO Armed Conflict Dataset
	ucdp_conflict_osv_1_id	INT	Conflict ID as used in UCDP One-sided violence Dataset
	ucdp_conflict_osv_2_id	INT	Conflict ID as used in UCDP One-sided violence Dataset
	ucdp_dyad_sgv_1_id	INT	Dyad ID as used in UCDP/PRIO Armed Conflict Dataset
	ucdp_dyad_sgv_2_id	INT	Dyad ID as used in UCDP/PRIO Armed Conflict Dataset
	ucdp_dyad_osv_1_id	INT	Dyad ID as used in UCDP One-sided violence Dataset
	ucdp_dyad_osv_2_id	INT	Dyad ID as used in UCDP One-sided violence Dataset

Category (cont.)	Field Name	Data Type	Description
Aimpoint	region_id	INT	ID code for region
	subregion_id	INT	ID code for subregion
	admin0_id	INT	ID code for country
	admin0_cow_id	INT	ID code for country (used by UCDP/Correlates of War datasets)
	admin1_id	INT	ID code for province/state
	city_id	INT	ID code for city, village or district
	region_txt	CHAR	Region where attack took place
	subregion_txt	CHAR	Subregion where attack took place
	admin0_txt	CHAR	Country where attack took place
	admin1_txt	CHAR	Province/State where attack took place
	city_txt	CHAR	City, village or district where attack took place
	aimpoint_desc	CHAR	Specific location targeted
	latitude	DECIMAL	City's Latitude
	longitude	DECIMAL	City's Longitude
Target	target_desc	CHAR	The intended target of the attack
	tt1_id	INT	ID code for Target Type 1
	tt2_id	INT	ID code for Target Type 2
	tt3_id	INT	ID code for Target Type 3
	tt1_txt	CHAR	Target Type 1: Type of target attacked (Civilian, Political, Security)
	tt2_txt	CHAR	Target Type 2: Sub-category for Target Type 1
	tt3_txt	CHAR	Target Type 3: Sub-category for Target Type 2
	tt_org_id	INT	ID code for target's organizational affiliation
	tt_natl_id	INT	ID code for target's nationality
	tt_ethnic_id	INT	ID code for target's ethnicity
	tt_org_txt	CHAR	Target's organizational affiliation
	tt_natl_txt	CHAR	Target's nationality
	tt_ethnic_txt	CHAR	Target's ethnicity
	tt_relig_id	INT	ID code for target's religious affiliation
	tt_denom_id	INT	ID code for target's religious denomination affiliation
	tt_sect_id	INT	ID code for target's religious sect affiliation
	tt_relig_txt	CHAR	Target's religious affiliation
	tt_denom_txt	CHAR	Target's religious denomination affiliation: Sub-category for religious affiliation
	tt_sect_txt	CHAR	Target's religious sect affiliation: Sub-category for religious denomination affiliation

Category (cont.)	Field Name	Data Type	Description
Claim(s)	ct_claimed	INT	Number of groups that **claimed** responsibility for the attack
	ct_denied	INT	Number of groups that **denied** responsibility for the attack
	ct_suspect	INT	Number of groups **suspected** for conducting the attack
	group_clm_1_id	INT	ID code for group 1 who **claimed** responsibility for the attack
	group_clm_2_id	INT	ID code for group 2 who **claimed** responsibility for the attack
	group_clm_3_id	INT	ID code for group 3 who **claimed** responsibility for the attack
	group_clm_1_txt	CHAR	Group 1's name who **claimed** responsibility for the attack
	group_clm_2_txt	CHAR	Group 2's name who **claimed** responsibility for the attack
	group_clm_3_txt	CHAR	Group 3's name who **claimed** responsibility for the attack
	group_den_1_id	INT	ID code for group 1 who **denied** responsibility for the attack
	group_den_2_id	INT	ID code for group 2 who **denied** responsibility for the attack
	group_den_3_id	INT	ID code for group 3 who **denied** responsibility for the attack
	group_den_1_txt	CHAR	Group 1's name who **denied** responsibility for the attack
	group_den_2_txt	CHAR	Group 2's name who **denied** responsibility for the attack
	group_den_3_txt	CHAR	Group 3's name who **denied** responsibility for the attack
	group_susp_1_id	INT	ID code for group 1 **suspected** by government official for conducting the attack
	group_susp_2_id	INT	ID code for group 2 **suspected** by government official for conducting the attack
	group_susp_3_id	INT	ID code for group 3 **suspected** by government official for conducting the attack
Attacker(s)	ct_attackers	INT	Number of suicide attackers involved
	ct_attackers_fem	INT	Number of female attackers
	ct_attackers_mal	INT	Number of male attackers
	ct_attackers_unk	INT	Number of attackers where gender is unknown

B. DSAT Claims (dsat_claims)

Category	Field Name	Data Type	Description
Dataset	data_release	DATE	Date when data was exported from the DSAT
Claim Information	event_id	INT	Unique attack identifier
	claim_id	INT	Unique claim identifier
	group_id	INT	CPOST Militant Group ID
	ucdp_group_id	INT	UCDP Militant Group ID
	group_name	CHAR	Group's name
	claim	CHAR	Claim status: Claimed, Denied, Suspected, Unclaimed
	ucdp_conflict_sgv_1_id	INT	Conflict ID as used in UCDP/PRIO Armed Conflict Dataset
	ucdp_conflict_sgv_2_id	INT	Conflict ID as used in UCDP/PRIO Armed Conflict Dataset
	ucdp_conflict_osv_1_id	INT	Conflict ID as used in UCDP One-sided violence Dataset
	ucdp_conflict_osv_2_id	INT	Conflict ID as used in UCDP One-sided violence Dataset
	ucdp_dyad_sgv_1_id	INT	Dyad ID as used in UCDP/PRIO Armed Conflict Dataset
	ucdp_dyad_sgv_2_id	INT	Dyad ID as used in UCDP/PRIO Armed Conflict Dataset
	ucdp_dyad_osv_1_id	INT	Dyad ID as used in UCDP One-sided violence Dataset
	ucdp_dyad_osv_2_id	INT	Dyad ID as used in UCDP One-sided violence Dataset

C. DSAT Attackers (dsat_attackers)

Category	Field Name	Data Type	Description
Dataset	data_release	DATE	Date when data was exported from the DSAT
Attacker Information	event_id	INT	Unique attack identifier
	attacker_id	INT	Unique attacker identifier
	gender	CHAR	Attacker's gender
	birth_year	INT	Attacker's birth year
	age_time_death	INT	Attacker's age at time of death
	birth_admin0_id	INT	ID code for attacker's birth country
	birth_admin0_cow_id	INT	ID code for attacker's birth country (used by UCDP/Correlates of War datasets)
	birth_admin1_id	INT	ID code for attacker's birth province/state
	birth_city_id	INT	ID code for attacker's birth city
	birth_admin0_txt	CHAR	Country where the attacker was born
	birth_admin1_txt	CHAR	Province/State where the attacker was born
	birth_city_txt	CHAR	City, village or district where the attacker was born

V. DSAT Variables Clarification

A. Attack Information

- **Killed Low:** The lowest number of killed reported across sources.
- **Killed High:** The highest number of killed reported across sources.
- **Wounded Low:** The lowest number of wounded reported across sources
- **Wounded High:** The highest number of wounded reported across sources
- **Civilian/Political/Security Fatalities:** This category is coded by the disaggregation of victims in an attack into the categories of 'civilian,' 'political,' or 'security.' In optimal circumstances, explicit terms are used to identify victims as such. In other cases, heuristics are used to identify victims (i.e. people at an embassy are political, people at a market are civilian, people with guns are security). When multiple victims of different affiliations are targeted, the coder disaggregates based on the context, and may have to divide the casualty counts by the number of different targets being hit. Note that *only deaths* are included in these variables, not wounded.

B. Additional Notes on Causalities

Often sources will give the casualty count for multiple attacks combined, e.g. "two suicide bombings rocked Maiduguri, resulting in 30 deaths." If the two attacks have different targets, two entries are created in the DSAT. If no other information to help split the casualties is given in the sources, we split the fatalities/wounded evenly between the attacks. If there is other relevant information, we use our best judgement to estimate the split of the casualties.

There will often be news reports that differ on the number of people killed or wounded in a suicide blast. In Deaths/Wounded high, we input the highest number found in our sources, in Deaths/Wounded low, we input the lowest number found.

Categorical Casualty Counts: Sources will not always use exact numbers for casualty counts. Instead, they will use vague language such as "several deaths" or "many wounded." In these situations, we use the following conversion table:

A couple	2	Dozens	24
A few/others	3	Scores	40
Multiple	4	At least #	#
A number/several	5	More than #	# + 1
Many	7	Nearly #	# - 1
Tens	20		

C. Target Location

- **Attack place:** *(Short Text)* Attack place refers to a more specific location within the city. For example, an attack place would be "Route 71 Bus Station" or "Food Stalls at Central Market." If no additional information is available, the attack place variable will be coded as "Unknown."
- **Target:** *(Short Text)* Indicates a descriptive target of the attack. The "Target" variable contains information conveyed in the target types. "Target" is not to be confused with location.
- **Target Type 1:** Suicide attacks against security and political targets often kill many civilians. In such cases CPOST codes the intended target. CPOST also uses primary sources such as claim texts issued by groups to determine the target of attacks. If no information on a likely target is present, the target is coded as unknown.

  Rules for Ambiguous Cases

  - **Target Change:** When a bomber detonates while attempting to proceed to some other target (e.g. outside of a mosque), the target of the attack is coded as the destination the bomber intended to reach. This does not apply if the bomber changes targets (e.g. targets security forces guarding a sporting event instead of the event itself).
  - **Joint Targeting:** When there are multiple targets of one security attack (e.g. a joint convoy of U.S. and Afghan troops), the order of precedence is foreign actor, domestic actor, non-state actor. When prioritizing one actor over another, the remaining TTs are coded according to the chosen actor.
  - **Coalitions:** In cases where there is a dominant state in the coalition that is targeted, we code their troops' Organizational Affiliation as Foreign Government and the nationality as the dominant state. If a non-dominant partner in a coalition is targeted, we code the nationality as that country but code the Organizational Affiliation as the appropriate coalition. If no specific information about which partner in a coalition is targeted is known, we code Nationality as Multinational and Organizational Affiliation as the coalition.

- **Nationality:** We do not assume that the nationality of the target is that of the state that the attack occurred within unless there is some reason to think the target was indeed a citizen of that nation or a foreigner. Any distinction between nation and state is ignored.
- **Religion, Denomination, Sect:** For targets for which these variables are applicable, religion is coded as “"Unknown" unless the sources indicate that the target was of a certain religion.
  - Non-state groups with a strong religious or ethnic affiliation (e.g. the Anbar Awakening) may be coded with the group's religion.
  - Civilian targets may have religions, denominations, and sects if the sources give evidence for attributing them.
  - State actors do not have ethnicities or religions, and as such their religions are coded as “N/A”.

  Exception: Cases where it is required to assume religion without any supporting evidence from the source or claim (e.g. Sadr City, Baghdad).

- **Ethnicity:** Ethnicity is coded as “"Unknown" unless the sources indicate that the target was of a certain ethnic group. Ethnicity is not generally assumed, especially when coding civilian targets. For non-state groups with strong ethnic affiliations (e.g. the Popular Mobilization Front), attacks on them may be coded with the group's ethnicity. State actors do not have ethnicities or religions, and as such are coded as “NA.” There are certain cases where CPOST coders assume ethnicity without any supporting evidence from the source or claim (e.g. Bashir: a predominantly Iraqi Turkmen village south of Kirkuk).

  Exception: The Peshmerga and other actors loyal to Iraqi Kurdistan are coded with the ethnicity “Kurd” even though we code them as Security Forces rather than Rival Militia.

  In cases of assassination, we add all relevant information about the target indicated in the sources, including religion and ethnicity, even if the target is a state actor.

D. Attack Attribution (Claims by Groups)

- **Claim:** There are four ways to code attribution in the DSAT:

  1. **Claimed:** there is explicit mention of a group claiming responsibility for a suicide attack in at least one source used to code the attack.
  2. **Denied:** there is explicit mention of a group denying responsibility for a suicide attack in at least one source used to code the attack.
  3. **Suspected:** there is explicit mention of an official government source stating that a group carried out an attack. For our coding rules, media suspicion is not sufficient.
  4. **Unclaimed:** there is not explicit mention of a group claiming, denying, or suspected of having responsibility for a suicide attack in any source used to code the attack.

E. Attacker(s)

- **#Attackers:** Includes only the number of people who detonated, not the total number of people who were involved in the attack. CPOST does not count non-suicide attackers.