A METHODOLOGICAL TOOLKIT

BY COLLECTIVE AID

EVIDENCE-BASED ACTION FOR HUMAN RIGHTS AT BORDERS

Methodological Toolkit and Pilot Project Aegean Incidents (2024/2025)

THIS PROJECT WAS DELIVERED BY COLLECTIVE AID INTERNATIONAL PROGRAMME AND MADE POSSIBLE THROUGH THE GENEROUS SUPPORT OF HEINRICH-BÖLL-STIFTUNG THESSALONIKI TO COLLECTIVE AID CHARITABLE FOUNDATION.

I. INTRODUCTION AND PURPOSE

1.1 Why This Toolkit Matters

The Aegean Sea remains one of Europe's most contested maritime borders, characterized by complex interactions between state authorities, maritime accidents, rescue operations, and human rights concerns. Official statistics from Greek and Turkish Coast Guards often conflict, media reporting varies in accuracy, and humanitarian organizations document gaps between what authorities report and what communities experience.

This Unified Monitoring Toolkit addresses a critical need: standardized, transparent methodology for collecting, coding, verifying, and ethically using incident-level data on border violence, maritime deaths, apprehensions, rescues, and related incidents. By establishing clear operational definitions, consistent coding rules, and verification procedures, this framework enables multiple stakeholders, researchers, advocates, policymakers, legal teams, to work with compatible datasets that maintain scientific rigor while honoring the human reality these numbers represent.

For the pilot of the project the geographic focus remains the Aegean Sea. The design will allow for transferability to other EU external border regions.

1.1 Why This Toolkit Matters

This toolkit documents the incident-level Aegean Sea dataset, which captures discrete maritime border events (pushbacks, rescues, shipwrecks, apprehensions, boat chases) from January 2024 to December 2025, with data collected from official Coast Guard statistics (Greek and Turkish), IOM Missing Migrant project, media monitoring, NGO reports (Aegean Boat Report, Alarm Phone). The dataset currently tracks 2554 incidents in 2024 and 1062 incidents in 2025, across Greek islands (Lesbos, Chios, Samos, Kos, Rhodes, Leros, and others) and Turkish coastal regions.

What this toolkit covers:

Unified data collection methodology aligned with international documentation practices

Multi-source verification procedures with four-tier reliability assessment

Quality assurance and missing data management with transparent reporting

Analysis guidance for researchers, advocates, and policymakers

Worked examples showing real incident coding from multiple sources

Consistent incident-level coding framework with clear operational definitions

Ethical and security protocols with do no harm implementation checklist

Intersectionality and vulnerable groups considerations

Analysis guidance for researchers, advocates, and policymakers

Transferability principles for adaptation to other border contexts

What this toolkit does NOT cover:

The technical data science memorandum (maintained separately by data specialists), see Annex 2

Specific policy recommendations (reserved for research reports and advocacy materials)

1.3 How to Use This Document

For data collectors

Sections III-VII provide step-by-step guidance on what to collect and how to code it, with decision trees and practical examples.

For data analysts and data scientists

Section XII offers guidance on appropriate analytical approaches, including descriptive statistics, and inferential tests. The accompanying Technical Memorandum (Annex 2) details

For policymakers and advocates

Sections I, II, VIII, and XIII explain what the data can and cannot tell you, what ethical considerations apply, and how to responsibly communicate findings.

For independent verifiers and external partners

The entire framework is designed for replicability, sufficient detail that another qualified analyst could produce comparable results using the same sources and methodology.

For organizations planning adaptation to other regions

Section XVIII provides detailed transferability guidance for geographic, sectoral, and operational adaptation.

II. TARGET USERS AND STAKEHOLDER GROUPS

This toolkit is designed for multiple audiences working across different contexts:

	Research teams	Human rights monitors	NGOs and civil society	Data scientists and analysts	Journalists and media
Primary need	Standardized methodology for academic or advocacy research	Incident documentation for evidence gathering and accountability	Data for policy advocacy, parliamentary questions, and public communication	Clean, standardized incident data for advanced analysis	Credible incident-level data for investigative reporting
How This Toolkit Helps	Provides reproducible framework, statistical foundations, documentation protocols, and shared definitions	Verification tiers, ethical protocols, do no harm guidance, and secure data handling procedures	Quality assurance procedures, limitations clearly stated, appropriate disclaimers, and confidence intervals	Codebook specifications, missing value protocols, data quality indicators, and analytical guidance	Source documentation, multi-source corroboration procedures, and ethical safeguards for survivor protection

Key principle: Anyone following this methodology should be able to produce datasets compatible with others using the same framework, enabling collaboration and comparative analysis while maintaining data quality standards.

III. CORE CONCEPTS AND CLEAR DEFINITIONS

3.1 What is an Incident?

An incident is a single, discrete border-related event that:

Occurs at a specific point in time (even if the date range is uncertain)

Involves identifiable location(s) (Turkish and/or Greek coastal waters/ territory, or both)

Can be described using a consistent set of core characteristics

Is documented by at least one credible source

A single pushback event (e.g., vessel intercepted by Greek Coast Guard, pushed back to Turkey)

One rescue operation (e.g., Turkish Coast Guard saves 35 people from disabled boat)

One shipwreck event (e.g., boat capsizes near Lesbos, people rescued by Greek Coast Guard)

One combined event sequence (e.g., attempted crossing → pushback by Greek Coast Guard → rescue by Turkish Coast Guard)

One apprehension event (e.g., Turkish Coast Guard stops vessel with 42 people)

One body recovery (e.g., deceased body of foreign national found on shore, recovered by authorities)

How to decide: If you can clearly distinguish the event by time and location, and core incident characteristics differ, code as separate incidents. When in doubt, consult the decision tree for incident boundaries in Section V.

3.2 Border Violence Typology

Pushback Forced return or expulsion by authorities preventing access to asylum procedures.	Includes: Any form of coercive return; situations involving violence, intimidation, confiscation preventing landing	Excludes voluntary returns; redirections initiated by people themselves; administrative processing
Rescue Direct intervention/ emergency evacuation to save individuals from immediate danger	Life-threatening distress; emergency intervention; all rescue modalities (Coast Guard, NGO, merchant vessel, independent)	Excludes Routine disembarkation; administrative handling; apprehension without distress component

Apprehension State interception/ stoppage/control before/during/ after attempted crossing	Authorities board/seize vessel; stop onward movement; coercive control	Excludes Voluntary redirects; routine port processing
Shipwreck Vessel capsizes, sinks, breaks apart, or is destroyed creating life-threatening conditions	Full or partial sinking; capsizing; severe structural damage	Excludes Functional boats; controlled grounding; minor damage

Boat Chase Vessel pursued at sea by authorities using high-speed pursuit, intimidation, or coercive tactics	High-speed pursuit; use of sirens/lights/ warnings/shots; tactical interception	Excludes Ordinary escorting; shadowing without coercion
Criminalization Individuals arrested/ charged/ suspected in connection with incident, typically as alleged smugglers or boat drivers	Formal or suspected charges under smuggling, trafficking, boat driving, or immigration provisions	Excludes Administrative registration without criminal charges; document discovery without charges
Body Recovery Deceased person discovered and recovered, linked to border context	Confirmed or probable border-related death; maritime context; recovery by authorities/ individuals	Excludes Deaths unrelated to crossing; disappearances without body recovery

IV. DATA COLLECTION ARCHITECTURE

4.1 Multi-Source Collection Strategy

The Aegean dataset employs triangulation across three distinct data streams, each maintained separately until verification procedures are applied:

STREAM 1: OFFICIAL STATISTICS (PRIMARY STREAM)

Turkish Coast Guard (TCG) records:

- COA (Coast Guard Apprehended): Turkish Coast Guard apprehension records
- COR (Coast Guard Rescued): Turkish rescue operations
- COR+PB (Turkish rescue AFTER Greek pushback): Combined incidents documenting Greek enforcement followed by Turkish response

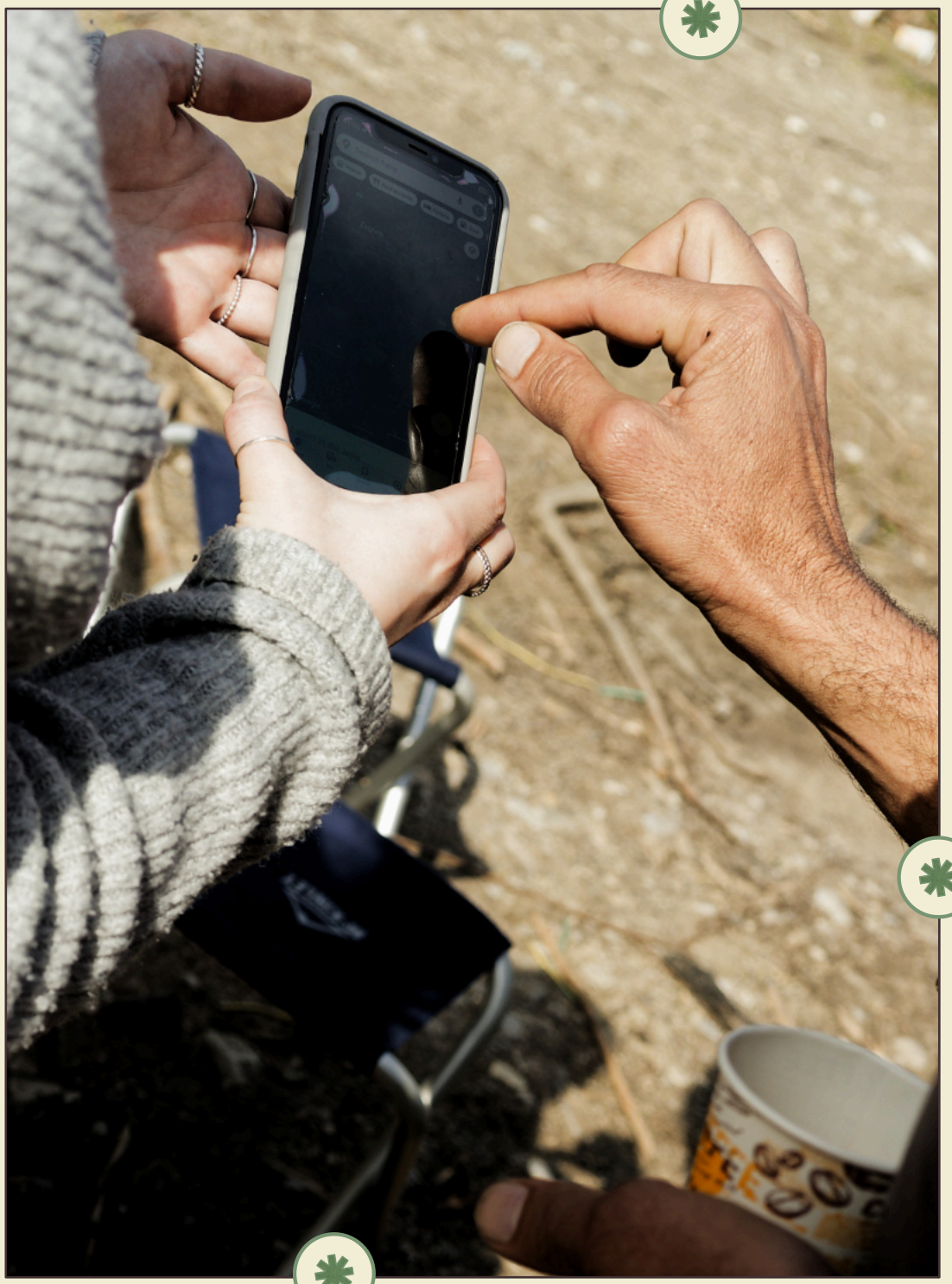
Hellenic Coast Guard (HCG) records:

- Official rescue, people found and boat chase reports
- Arrest and criminalization records (when applicable)
- Body recovery notifications

Processing: Official data are coded directly into the dataset with source verification coded as Single Official. Authority figures (dead, missing, people counts) are prioritized in conflict resolution procedures (see Section V). All official sources are listed with hyperlinks for where the official data is available.

Caution: We acknowledge that official sources do not provide a complete or fully transparent account of all incidents and may contain limitations, omissions, or institutional biases. Nevertheless, they represent the most authoritative and consistently available official data currently accessible, and are therefore included as the primary statistical stream.

The secondary stream of information is therefore important to verify and cross check the official information provided.



STREAM 2: MEDIA AND NGO MONITORING (SECONDARY STREAM)

Daily media and NGO monitoring of:

- Greek media outlets: Sto Nisi (local), Lesvos News, 24/7 news (news247.gr), Kathimerini, Political Lesvos and island-based news portals.
- International outlets: InfoMigrants, Al Jazeera, international press agencies (AP, Reuters), BBC, The Guardian, Politico are also followed.
- NGOs reporting: Focuses on Aegean Boat Report and Alarm Phone, two organizations, which provide emergency hotlines for people.

Processing: Media and NGO reports are reviewed for incident verification, new incident identification, additional context, and conflicting information. Reports go into the Multisource official verification category when linked to official data. When only reported by media and/or NGO the source reliability is classified as Single/Multiple without the official tag.

STREAM 3: CROSS-VERIFICATION DATABASE (INTEGRATION POINT)

Monthly cross-check with:

- IOM Missing Migrant Project database entries for maritime deaths and missing persons
- Comparative analysis of official figures (Turkish vs. Greek Coast Guard reports for same incidents)
- Identification of apparent duplicates or conflicting accounts across sources
- Reconciliation of discrepancies using conflict resolution procedures

Processing: Processing: Conflicting data undergo systematic verification procedure (Section V) to establish final verified figures. All conflicts documented in the Additional Information field with full audit trail.

V. CODING FRAMEWORK AND DATA MODEL

5.1 Incident-Level Structure

The dataset maintains one row per incident structure because this is the standard format in quantitative social science research, allowing for:

- Statistical testing: Descriptive statistics andinferential statistics
- Filtering and aggregation: All pushbacks in January; All incidents involving children
- Missing value handling: Transparent tracking of unknown vs. confirmed values
- Temporal analysis: Trends by week, month, location, incident type
- Cross-tabulation: Relationships between variables (e.g., pushback × shipwreck)

Date
(YYYY-MM-DD)

Clearly stated incident date or explicit day reference

Publication dates; later discovery dates; date ranges without start date

Time
(HH:MM, 24-hr)

Hour of incident from any reliable source

Durations; reporting times; vague references (early morning)

Area Turkey
Text

Named Turkish coastal region/province/maritime area

Vague references; non-geographic locations; unmarked locations

Shipwreck
Categorical (0/1)

Vessel sank, capsized, or was destroyed

Functional boats; minor damage; controlled grounding

Area Greece
Text

Named Greek island or coastal area

Vague references; administrative centers; unmarked locations

Land or Sea
Categorical

Incident location domain (0=Land; 1=Sea, 2=both)

Insufficient information; ambiguous cases

Means of Travelling
Text

Named Turkish coastal region/province/maritime area

Specific boat type or transport mode used

Gun Shot Fired
Categorical (0/1)

Shots fired during pursuit/enforcement action

General tactics not proven in incident; unrelated shooting

Found by GA
Categorical (0/1)

Group located by Greek authorities on land

Sea locations; other finders; apprehension without custody

Apprehension by TA
Categorical (0/1)

Turkish authorities intercepted/stopped vessel/people

Voluntary redirects; other agencies; routine port processing

Pushback
Categorical (0/1)

Forced return by authorities (primarily Greek CG)

Voluntary returns; administrative processing; routine redirects

Boat Chase
Categorical (0/1)

Pursuit by authorities (primarily Greek)

Ordinary escorting; non-coercive movement; routine patrol

Rescue by GA
Categorical (0/1)

Greek authorities emergency evacuation from distress

Apprehension without rescue; disembarkation; custody transfer

Rescue by TA
Categorical (0/1)

Turkish authorities emergency evacuation from distress

Apprehension; routine recovery; routine disembarkation

Rescue by Other
Categorical (0/1)

NGO, merchant vessel, or independent rescue

State actor rescue; non-distress interventions; administrative assist

Source Verification
Categorical

Evidence strength (Single Official / Single / Multisource Official / Multisource)

No sources; unverifiable reports

Number of People
Numerical

Total individuals involved in incident

Counts unrelated to incident; estimates without numbers

Number of Children
Numerical

Minors (under 18) explicitly identified, if no information labeled as NA

Age groups without specific child identification; 'several' without count

Nationalities Present
Text

All nationalities reported with the number of them

Ethnicities without nationality; mixed without breakdown

Additional Information
Text

Sequence of events, conflicts, context not captured elsewhere, criminalization information.

Information already in other variables; redundant notation

Number Dead
Numerical

Confirmed fatalities caused by incident

Speculative deaths; unrelated deaths; missing persons (separate variable)

Number Missing
Numerical

Individuals unaccounted for post-incident

Reporting gaps; unclear absences; speculative counts

Criminalization
Categorical (0/1)

Individuals arrested/ charged/suspected as part of incident

Administrative registration without charges; document discovery

Charges
Text

Specific legal charges applied (if any)

Vague allegations without formal/ suspected charges

5.1 Incident-Level Structure

RULE 1: ALWAYS CODE UNKNOWN VALUES AS NA (NOT 0)

– **CORRECT**: Number of children = NA (source doesn't mention)

– **WRONG**: Number of children = 0 (implies none were present)

Principle: 0 = known to be zero; NA = unknown or not reported

RULE 2: WHEN TWO REPORTS DESCRIBE SAME INCIDENT, MATCH ON TWO OF

– Location (same area or clearly overlapping)

– Time (within 2 hours on same date)

– Number of people (within 10% of reported figure)

– Incident type (same core category)

RULE 3: CONFLICTING INFORMATION RESOLUTION HIERARCHY

When official and non-official sources conflict:

– Official source priority: Use official figure first (Coast Guard figures take precedence over media estimates)

– If multiple official sources conflict: Use median/most common figure; document discrepancy in Additional Information

– If all non-official sources: Use most reliable source (news outlet with editorial standards > social media > unverified reports)

– When no authority exists: Record lower bound.

Example:

Turkish Coast G
uard says 42 people rescuedGreek media says 40 people rescued
Code in dataset: 42 people

Additional Information: Turkish COR official data: 42 people. Greek news reports: 40 people. Discrepancy documented; official source figure used.

RULE 4: EVENTS OCCURRING ACROSS DATES

If an incident spans multiple days (e.g., vessel reported missing on Jan 5, rescue on Jan 6):

– Code earliest date in Date variable

– Document full sequence in Additional Information field

– Flag ambiguity for analyst awareness

Example:

Additional Information: Shipwreck occurred Jan 5 evening; rescue completed Jan 6 early morning; bodies recovered Jan 7. Coded with incident start date Jan 5. Multiple dates reflect the sequence of events.

VI. VERIFICATION TIERS SYSTEM

6.1 Why Verification Tiers Matter

Incident data quality varies enormously. A count reported by the Turkish Coast Guard differs in reliability from a count reported on social media. This toolkit addresses this variation through transparent verification tiers that allow analysts to:

- Filter data by confidence level (e.g., Multisource official)
- Report uncertainty honestly
- Justify what conclusions are defensible
- Identify which findings are robust vs. sensitive to data quality
- Support responsibility in advocacy and policy use

6.2 Four Verification Tiers

TIER 1: SINGLE OFFICIAL SOURCE

Incident reported by only ONE official authority (Coast Guard, Frontex, Hellenic Police, port authority).

Reliability: Moderate-High for event occurrence; Numbers subject to authority-specific biases and selective reporting incentives.

Example: Turkish Coast Guard reports 35 people apprehended on Jan 5 at Bodrum. No other sources mention this incident.

What can be published: Event occurrence (apprehension happened), numbers, descriptive trends (e.g., apprehensions increased in January).

What requires caution: Individual numbers should be reported as approximately X rather than exact; Confidence intervals recommended in analysis; Note authority interests (apprehensions = enforcement success).

TIER 2: SINGLE

Incident reported by news media, NGO, or independent monitor, but NO official authority corroboration.

Reliability: Lower for numbers (media often estimates); Moderate for event description; Depends heavily on source reputation.

Example: InfoMigrants reports boat chase incident on Jan 8; neither Turkish nor Greek Coast Guard publishes data on this incident.

What can be published: Event description; incident category; caveats clearly stated.

What requires extreme caution: Numbers are speculative; only report with explicit qualification (e.g., reportedly 30+ people); Avoid citing as fact in policy documents; Consider whether the report may reflect bias or incomplete information.

TIER 3: MULTISOURCE OFFICIAL

Incident corroborated across TWO OR MORE official sources (e.g., Turkish CG + Greek CG + Frontex) OR official source + independent non-official corroboration.

Reliability: High for event occurrence; High for numbers IF all sources agree; Moderate if sources conflict (apply resolution procedure; document discrepancy).

Example: Turkish COR reports 47 people rescued near Samos; Greek news cites Hellenic Coast Guard confirming rescue of ~47 people; IOM Missing Migrants Project includes incident in database.

What can be published: Event with confidence; numbers as documented; trends are robust.

TIER 4: MULTISOURCE

Incident documented across non-official sources such as multiple media outlets and NGO sources.

Reliability: Middle- multiple independent sources confirm the same facts from different angles, but no official data.

Example: + Greek news reports NGO monitor all confirm the same incident with consistent numbers (±5%) and circumstances.

What can be published: Incident with middle confidence; exact numbers defensible; incident details; named individuals (when appropriate and consented).

6.3 Practical Verification Decision Tree

Start with incident from any source

Is it reported by official authority (HGC/Frontex/TGC/IOM)?

- NO → Go to Non-official sources only path
- YES, only one official source → Check for corroboration
  - Find news/NGO confirmation? → Check strength and convergence
    - Single media outlet or NGO reporting? → Tier 3 (Multisource Official)
    - No other sources? → Tier 1 (Single Official)

Non-official sources only path:

Is it reported by news outlets and/or NGOs?

- YES, single reputable outlet → Tier 2 (Single)
- YES, 2+ independent reputable outlets → Check convergence
  - Convergent accounts? → Tier 4 (Multisource)

Assign verification tier in dataset
Document all sources used with hyperlinks in different columns
Note conflicts in Additional Information

VII. DATA QUALITY AND MISSING VALUES

7.1 Missing Value Protocol

Missing data are inevitable in monitoring work. This toolkit is explicit about missingness rather than concealing it.

Coding missing data:

Use NA for any unknown value (never 0, never blank)
Never interpret absence of information as presence of zero

Why this matters:

Scenario 1: Report says 30 adults, no mention of children
✓ Number of children = NA (data not available)
✗ Number of children = 0 (implies confirmed absence)

Scenario 2: Official data shows 42 people apprehended; death toll not mentioned
✓ Number dead = NA
✗ Number dead = 0

Software implementation: In Excel, use the text NA; in statistical software (R, Python, Stata), use native missing value codes.

7.2 Reporting Missingness: Required % Missing Column

Every summary table or dataset should include % Missing column to indicate data reliability and limitations:

Example table:

Variable	N Valid	N Missing	% Missing
Number of People	95	5	5%
Number of Children	72	28	28%
Number Dead	68	32	32%
Number Missing	45	55	55%

Interpretation: The Number Missing variable is most problematic (55% missing), suggesting survivor reports are often unavailable. Analysts should restrict claims about missing persons to the subsample where data exist (N=45) rather than treating it as complete incident-level fact. Report as: Of incidents where data on missing persons were available (N=45), X% documented missing individuals.

7.3 Handling Common Missing Data Scenarios

Scenario A

Number of people given as range (e.g., 25-30 people)

Code: Lower bound (25) with notation in Additional Information

Add % Missing notation to highlight range uncertainty

Note in Additional Information: Source provides range 25-30; coded as lower bound (25)

Caveat in analysis: Incident counts conservative; actual range likely higher

Scenario B

Time not specified (only date given)

Code Time: NA

Incident remains in dataset for date-based analysis

Incidents excluded from time-of-day analysis (e.g., Are night incidents more dangerous?)

Scenario C

Children mentioned but count not specified (several children aboard)

Code: NA (not a count; several is not a number)

Add to Additional Information: Multiple children reported; exact number unknown

Analysts can still note presence of minors without claiming specific numbers

Possible analytical approach: Binary variable Children present: Yes/No alongside count variable

Scenario D

Deaths reported as range or uncertain (e.g., 2-5 dead or at least 2 dead)

Code: If official statistics use that source, otherwise lower confirmed bound (2) with notation

Add Additional Information: Source reports 25 deaths; 2 confirmed; others suspected

Report as: Minimum deaths documented: 2; suspected total: 2-5

Caveat: Mortality figures conservative; actual deaths likely higher

VIII. DATA ACCESS AND SECURE STORAGE

TIER 1: RAW DATA COLLECTED THROUGHOUT THE MONTH/SUBMITTED BY PARTNERS (MOST SENSITIVE)

– Location: Encrypted storage on Kobo with log in

– Usage: Case analysis, legal documentation, internal verification only

TIER 2: WORKING DATASET (PUBLIC USAGE)

– Location: Openly accessible repository with stable DOI (Digital Object Identifier), downloadable together with the codebook and methodological toolkit from Collective Aid Website

– Access: Global public; free download

– Contains: Incident-level data with complete codebook, all identifiers removed, source types documented

– Variables: All variables

IX. INTERSECTIONALITY AND GENDERED-DATA

Note on this section: This Aegean dataset does not currently capture gender/sex and sexuality information consistently. However, the framework is developed for potential future expansion. Current guidelines are adapted for contexts where such data ARE collected, and recommendations for other border contexts where gender data ARE available.

9.1 Handling Unknown/Unspecified Gender Data

When gender is not reported (common in official statistics), the toolkit:

– Does NOT assume: Absence of gender specification is not interpreted as all male

– Codes transparently: Gender field marks Unknown separately from confirmed male/female (if/when gender data added)

– Analyzes gender gaps: Acknowledges that under-reporting of women creates analytical blind spot and may reflect both data limitations and actual patterns of enforcement/risk

9.2 Other Intersectionality Indicators (Children, Unaccompanied Minors, Current Dataset)

Current Aegean dataset tracks:

– Children presence/count: Available in ~x% of incidents; encoded as Number of children variable

– Language: Use neutral, dignity-preserving language in narratives; avoid reinforcing stereotypes

X. TECHNICAL SPECIFICATIONS FOR ANALYSTS

10.1 Variable Types and Appropriate Analytics

Categorical variables (Incident Type, Area Greece, Area Turkey, Means of Travelling):

Use: Frequency tables, proportions, chi-square tests, logistic regression

Format: Text; never force into arbitrary numeric codes (don't code Lesbos = 1, Chios = 2)

Example: Frequency of vessel types: IB (45%), LB (32%), Speedboat (15%), Other (8%)

String/Text Variables (Additional information, Charges, Nationality):

Use: For narrative reports, frequency analysis

Format: Text

Example: Suspected facilitators from Turkey were mentioned x times.

Numerical variables (Number of People, Number Dead, Number Missing, Number of Children):

Use: Mean, median, percentiles, standard deviation, range; regression, ANOVA, t-tests

Do NOT use: These on dummy variables (Pushback: Yes/No)

Caution: Report missingness carefully; avoid imputation without justification; use robust statistics if outliers present

Example: Mean group size = 38 (SD = 15, Mdn = 36, range = 1-77); 5% missing data (N=95 valid)

Categorical variables (Incident Type, Area Greece, Area Turkey, Means of Travelling):

Use: Frequency tables, proportions, chi-square tests, logistic regression

Format: Text; never force into arbitrary numeric codes (don't code Lesbos = 1, Chios = 2)

Example: Frequency of vessel types: IB (45%), LB (32%), Speedboat (15%), Other (8%)

Dummy variables (Pushback Y/N, Rescue Y/N, Shipwreck Y/N, Criminalization Y/N):

Use: Logistic regression, cross-tabulation, difference-in-proportions tests, correlation

Format: 0/1; analyze as binary outcomes

Example: 64% of incidents (N=87) involved pushback; 45% involved rescue; 18% involved both

Options for handling NA values:

Listwise deletion (safest for small samples)

– Remove any observation with missing data on variables used

– Transparent; reduces sample size but avoids imputation bias

– Use when: N is large enough to sustain reductions; missing data clearly random

Example: Analysis restricted to incidents with complete data on number of people and death outcomes (N=68; 13 incidents excluded due to missingness)

Available case analysis (variable-specific)

– Use all available data for each variable

– Report N for each analysis; acknowledge potential bias

– Use when: Missingness mechanism differs by variable

Example: Number of people analyzed for N=95; number of dead for N=68

XI. WORKED EXAMPLE: INCIDENT CODING IN PRACTICE

11.1 Example Incident: Multi-Source, Conflicting Data (Real Case from Aegean Dataset) Raw Sources:

Source 1 - Turkish Coast Guard (COR+PB Records):

"On 09 July 2024 at 10.47 a.m., it was reported that there was a group of irregular migrants on Karaada in Çeşme, that 1 irregular migrant was taken from the sea by a fishing vessel present in the area and that there may be irregular migrants in the sea. 4 TUR CG Boats (TCSG-30, TCSG-912, KB-38, KB-4309), 1 TUR CG Helicopter (TCSG-507) and 1 TUR CG Diving Team (DEGAK-06) were immediately dispatched to the scene.

As a result of the search and rescue operations carried out by TUR CG assets, a total of 19 irregular migrants, 18 from the island and 1 from the fishing vessel, were rescued in good health and the lifeless bodies of 8 irregular migrants were found.

It was learned upon the initial statements received from the irregular migrants that they were left adrift on a life boat by Greek assets in an area close to Turkish territorial waters, that they got on Karaada by their own means after the life boat hit the rocks and sank, and that there were 27 irregular migrants on the life boat. The search and rescue operations initiated have been finalized.

Legal action has been taken by the Çeşme Public Prosecutor's Office regarding the incident."

Source 2 - Aegean Boat Report

"This morning disaster struck again, this time outside the Greek island of Chios, where 8 people drowned, as a direct result of a pushback performed by the Greek Coast Guard.

Turkish authorities announced this morning that a total of 19 people were found alive, 18 on the Turkish island of Karaada, and 1 picked up in the sea by a local fisherman, while 8 bodies had been recovered from the sea.

Survivors said that they had been placed in a life raft by the Greek Coast Guard, and that when the raft hit the rocky shore of the island, those who couldn't swim drowned."

Source 3 - Alarm Phone

"A group of 27 people arrived on the Greek island of Inousses. They were then subject to a deadly pushback. They were forced into life rafts and left adrift at sea by Greek authorities. After their life rafts crashed onto rocks, 8 people died.

In the morning hours of the 8th of July, Alarm Phone received a distress case concerning 27 people, who had landed on the Greek island of Inousses. The people had arrived safely on land but needed support as they had run out of water and were far from any camp or other facility. They wanted to apply for asylum in Greece.

We managed to establish direct contact to the people, who confirmed on the phone that they were on the Greek island of Inousses. They spoke of their urgent need to be rescued as soon as possible by the competent authorities. After confirming these details, we alerted Greek authorities at 06:45 CEST by mail, and sent a second mail shortly after with all the names of the people in the group. We also mentioned their intention to apply for asylum in Greece. The group later informed us they were moving to ta little church nearby, to wait in a safe place for their rescue.

Just after they started moving towards the small church, they sent us a message: "We are with the local police, do we have to be afraid?". After this moment, they went offline and could not be reached anymore. Over the next hours and days, we tried many times to reach them but to no avail.

We called various authorities to try and get information about the group's whereabouts. Despite having been in receipt of the email alert, local authorities on Chios claimed not to be aware of the situation. They only noted down details of the case over the phone and then asked us to call local police of Inousses. The local police station of Inousses also claimed that they were not aware of this situation and told us it was outside of their

outside of their responsibility. When calling the local port authority of Inousses, they claimed that it was not their responsibility and hung up the phone on us.

We continued to call all the different authorities in Greece throughout the day, but neither the Hellenic Coast Guard nor the Greek police picked up the phone or shared any information with us.

We were worried: the people were offline for too long, and we began to fear another brutal push back. We called Turkish authorities during the afternoon, however, at the time they had not rescued anybody in this area.

It was a long night and we kept trying to get information from the Greek authorities, without success.

In the early morning of the 9th of July, 24 hours after our last contact with the people – who had at that time been on land on a Greek island – we learned with horror from Turkish newspapers about 2 life rafts that had crashed on a Turkish island. Greek media reported on the incident but failed to mention the life rafts that the people had been in, hiding the pushback and cause of the crash. Following this discovery, we managed to confirm that it was the same group we had been in contact with the day before. We also learned that 7 people from this group had died in the sea and one was missing.

And so, it became clear: the group had been pushed back by Greek authorities. After having arrived on Inousses, they had been forced into life rafts, left adrift at sea and then suffered a deadly crash against the rocks of the Turkish Island of Kara Ada. Following the collision, the people had tried to rescue themselves by swimming to the island. However, not everybody made it.

We spoke with Turkish authorities, who confirmed the terrible news. Turkish authorities also confirmed that after a search and rescue operation has been conducted operation has been conducted, they had found the dead body of the missing person. This brought the confirmed death toll of the pushback up to 8 people.

These 8 people had arrived on European soil with their families 24 hours before. They were looking for safety, trying to escape the circumstances they were living in. When speaking with the people the day before, they had sent us photos of themselves, seeming happy and relieved to have made it to land on the island of Inousses. However, they did not find safety, but only torturous violence and death.

Europe: You are responsible for these deaths. These 27 people were looking for safety, instead 8 of them have been murdered by Greek authorities and the European border regime."

11.2 Coding Decision Process (Step-by-Step)

STEP 1: Is this one incident or multiple?

Analysis: All sources report the same date (1/5/24), overlapping time window with official mentioning 10:47 AM and NGO reporting saying morning, same location (Samos), same sequence (pushback, shipwreck then rescue), same approximate people count.

Decision: ONE INCIDENT - The event sequence (pushback → shipwreck → Turkish rescue) comprises one continuous border incident. The sequence of events is part of a single incident narrative, not separate incidents.

STEP 2: Which variables to code and what values?

Date
1/5/24
All sources agree; use official date format

Time
10:47
Use time of Turkish rescue (final intervention); note sequence in Additional Info

Area Turkey
İZMİR/Çeşme
Incident occurred between Turkish and Greek waters

Area Greece
Chios
Incident occurred between Turkish and Greek waters

Number of Children
8
Specified in TGC source

Nationalities Present
14 Afghanistan, 4 Iraq, 1 Morocco + 8 dead
TGC data most detailed; totals 27 persons

Number Dead
8
Explicitly confirmed by all sources

Number Missing
0
No mention in any source

Land or Sea
2
Both land and sea.

Means of Travelling
NA
Nothing explicitly mentioned

Found by GA
0
Greek authorities did not find/rescue; they intercepted

Apprehension by TA
0
Turkish authorities did not apprehend; they rescued

Criminalization
0
No arrests or charges mentioned

Source Verification
Multisource Official
Turkish CG official + Greek media + IOM corroboration

Source 1
Turkish COR official database
[Hyperlink]

Source 2
ABR
[Hyperlink]

Pushback
1 (Yes)
Greek authorities clearly redirected vessel coercively

Rescue by GA
0
Greek authorities did not rescue

Rescue by TA
1 (Yes)
Turkish Coast Guard conducted emergency rescue

Rescue by Other
0
No NGO or merchant vessel involvement documented

Source 3
AP
[Hyperlink]

Source 4
IOM Missing Migrant Project
[Hyperlink]

Additional Information
19 rescued 8 dead as a result of a push back from Inusses. People were left drifting in 2 life rafts that crashed into the rocks. All sources verify the number of deaths and people.
Document sequence, data sources, and verification

Boat Chase
0
No pursuit documented; Greek interception was enforcement, not chase

Gun Shot Fired
0
No gunshots mentioned in any source

Shipwreck
1
Vessel nonfunctional, people drowning

Number of People
27
Official Turkish CG figure; verified by NGO reporting

STEP 3: Conflict Resolution Applied and Documented

Conflict Identified

– Turkish Coast Guard (COR official records) report 27 people on the life boat, 19 rescued and 8 dead.

– Aegean Boat Report confirms 19 survivors and 8 recovered bodies, referencing survivor testimony.

– Alarm Phone initially reports 7 dead and 1 missing, later confirming that the missing person was recovered, bringing the death toll to 8.

Decision Process

1. Primary authority assessment: Turkish Coast Guard is the rescuing authority and conducted the final search and recovery operation. Its figures are considered authoritative for final outcomes.

2. Temporal discrepancy evaluation: Alarm Phone's initial report of 7 dead and 1 missing reflects an intermediate stage of the incident, before the search and rescue operation concluded.

3. Outcome convergence:

Alarm Phone later confirms, based on communication with Turkish authorities, that the missing person was found deceased, aligning with the Turkish CG total of 8 deaths.

4. Cross-source validation:

– All sources converge on:

– 27 total people

– 19 survivors

– 8 fatalities

– 0 missing (final status)

Resolution

Code the incident with 27 people total, 19 survivors, 8 dead, and 0 missing.

Temporary discrepancies are resolved by using final post-SAR figures confirmed by the rescuing authority and corroborated by NGOs.

Documentation in Additional Information

Turkish Coast Guard COR official data report 27 people on board, with 19 rescued and 8 bodies recovered. Alarm Phone initially reported 7 dead and 1 missing; subsequent confirmation from Turkish authorities verified recovery of the missing person, bringing the death toll to 8. The Aegean Boat Report corroborates survivor and fatality counts. Final figures reflect completed search and rescue operations.

STEP 4: Completed Additional Information Field

Sequence of Events:

– 08 July 2024 (morning hours) – A group of 27 people arrives on the Greek island of Inousses and requests assistance and asylum.

– The group is later forcibly transferred into life rafts and pushed back toward Turkish waters by Greek authorities.

– During the night, the life raft(s) crash into rocks near Karaada, Çeşme, İzmir, after which individuals attempt to swim to shore.

– 09 July 2024, 10:47 hrs – Turkish Coast Guard launches search and rescue operations.

– 19 people are rescued alive; 8 lifeless bodies are recovered from the sea.

Source Consistency:

– Turkish Coast Guard COR records: 27 total / 19 rescued / 8 dead.

– Aegean Boat Report: confirms 19 survivors and 8 deaths.

– Alarm Phone: confirms pushback sequence and final death toll of 8 after recovery of initially missing person.

Verification Status:

– Multisource corroboration achieved.

– Official rescue data supported by independent NGO documentation and media reporting.

– Minor discrepancies resolved through temporal sequencing and final outcome verification.

XII. ANALYSIS GUIDANCE FOR RESEARCHERS AND ADVOCATES

12.1 Recommended Analytical Approaches

Incident-level datasets documenting migration enforcement, rescue, and harm are well suited to descriptive and inferential statistical analysis aimed at identifying patterns rather than causal proof. Descriptive statistics are typically used as a first step to quantify the prevalence of key event types (e.g., pushbacks, rescues, fatalities) through proportions, rates, and confidence intervals. Stratified summaries by time period, location, verification tier, or incident type allow comparison across contexts while maintaining transparency about data coverage and missingness. These analyses help establish the empirical contours of documented practices and outcomes, while acknowledging that such datasets often reflect only a subset of all events due to reporting and detection biases.

Where sample size and data completeness permit, inferential methods can be applied to test associations and explore predictors of outcomes. Chi-square tests or Fisher's exact tests are appropriate for examining relationships between categorical variables, such as enforcement action and rescue occurrence, while logistic regression models can be used to assess how multiple factors jointly relate to the likelihood of severe outcomes (e.g., deaths or missing persons). More on the usable test in Annex II.

12.2 Statistical Power and Sample Size Considerations

Current dataset documenting 3616 incidents on the Aegean Sea across 2024 and 2025 is sufficient for: descriptive statistics and inferential statistics.

12.3 Visualization Guidance

Appropriate charts:

– Bar charts: Incident type frequency, pushback vs. rescue by location, criminalization prevalence

– Time series: Incident count by month, cumulative incidents over time, seasonal patterns

– Heatmaps: Pushback vs. rescue by island × month, incident types by season, nationality composition shifts

– Proportional plots: % of incidents with rescue by incident type, % involving children, % by verification tier

– Confidence interval plots: Pushback prevalence by island with error bars; comparison of rescue rates across time periods

XIII. LIMITATIONS, BIASES, AND TRANSPARENCY STATEMENTS

13.1 Known Limitations

Temporal: Dataset current through January 2024 to December, 2025. It does not show a comprehensive picture of the Aegean sea for a longer time. Patterns may have changed; results not predictive of future months without caution about structural changes.

Geographic: Coverage strongest for major islands (Lesbos, Chios, Samos, Kos, Rhodes) due to monitoring infrastructure and media presence. Leros, Kalymnos, smaller islands, and remote coastal areas may be underrepresented. Turkish mainland coastline coverage is more fragmented than Greek islands.

Source availability: Turkish media reporting in English limited compared to Greek; international media coverage episodic (increased around major incidents). Smaller NGO monitoring activities may be underdetected if not well-publicized.

Official reporting: Coast Guard data reflect only detected incidents. Dark figure of undetected pushbacks, deaths, rescues likely significant. Authorities may have incentives to under-report certain incident types (deaths) and over-report others (apprehensions = enforcement success).

13.2 Potential Biases

Selection bias: Incidents resulting in rescue more likely reported than apprehensions. Incidents with deaths more likely to trigger official response and media coverage. Quiet apprehensions or silent deaths at sea may go entirely undocumented. Effect: Dataset likely underrepresent dramatic/deadly incidents relative to total incident universe.

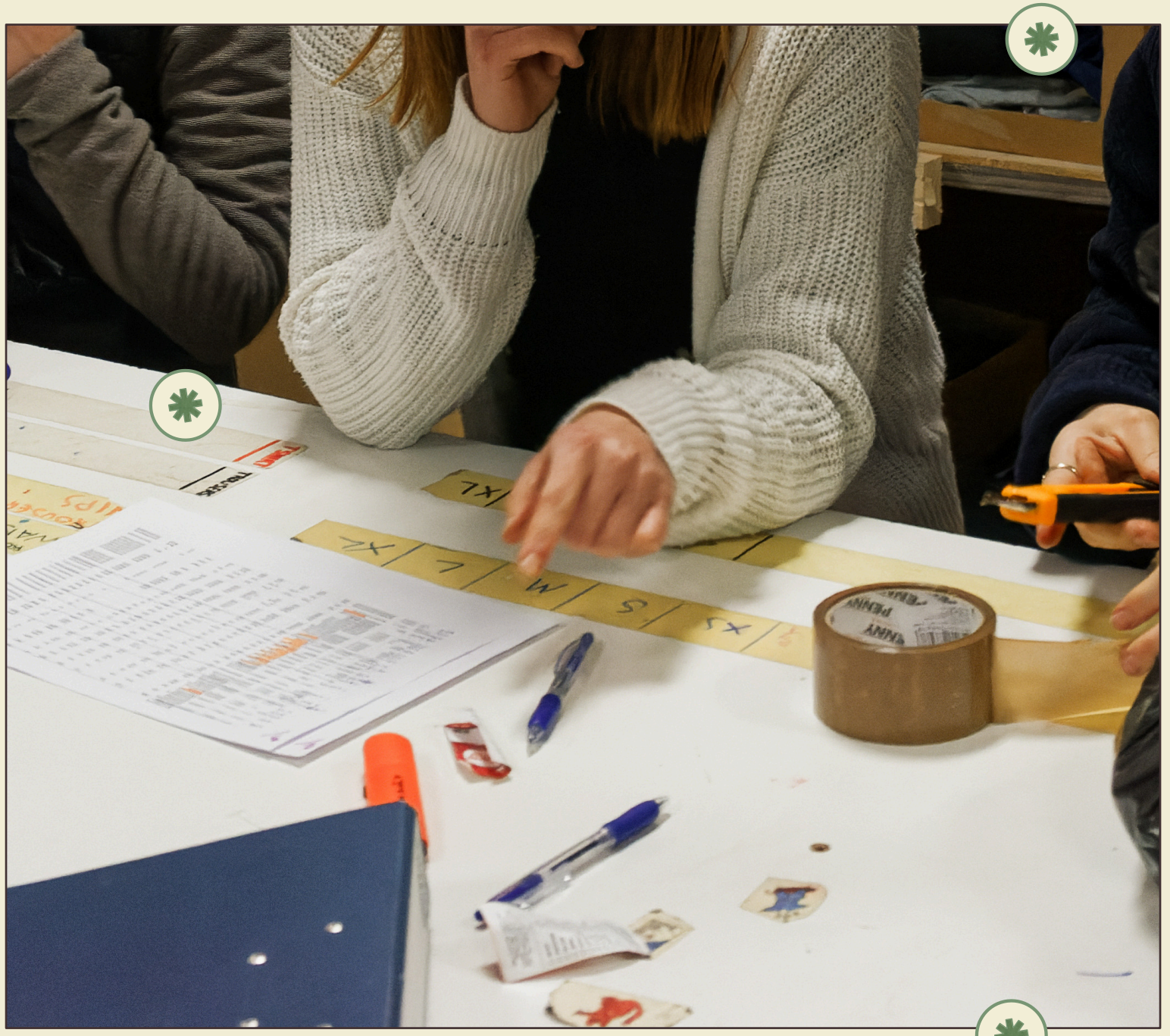
Authority bias: Official statistics may overcount apprehensions (demonstrates enforcement effectiveness) and undercount deaths (liability concerns; may not report unconfirmed deaths). Turkish authorities may report different figures than Greek authorities for the same incident, reflecting different operational perspectives. Effect: Numbers depend heavily on which authority reports; aggregate figures uncertain.

13.3 Required Transparency Statement for All Publications

Any study or report using this data MUST include the following methodological transparency statement:

Data Limitations: This analysis uses the Aegean Sea Border Incident Dataset, which documents incidents from official authorities (Turkish and Greek Coast Guards, Frontex) and complementary monitoring (media, NGOs, IOM). The dataset is subject to:

Interpretation: Findings should be interpreted as conservative estimates. True incident frequency, death toll, and harm are likely higher than documented.



XIV. TRANSFERABILITY AND ADAPTATION TO OTHER BORDER CONTEXTS

14.1 Transferability Principles

This toolkit was explicitly designed for replication and adaptation to other EU external border regions and, with modifications, to non-EU border contexts. The framework's transferability rests on several core principles:

Principle 1: Clear Operational Definitions Over Context-Specific Assumptions

The toolkit defines concepts (pushback, rescue, apprehension, shipwreck) using operational criteria rather than context-specific actions. This allows the definitions to apply across:

– Different state actors (Greek CG → Italian CG)
– Different routes (Aegean Sea → Central Mediterranean)
– Different vehicles (boats → desert vehicles → rail transport)
– Different enforcement types (maritime pushback → pushback on land)

Example: Pushback is defined as forced return or expulsion by authorities preventing access to asylum procedures - this applies equally to other border contexts.

Principle 2: Variables and Variables Structure Separate from Geographic Context

The dataset structure (one row per incident with consistent variable types) is universal. What changes across contexts is:

– Geographic variables (Area Turkey → Area Morocco OR Border Crossing Point)
– Authority variables (Turkish Coast Guard → Italian Coast Guard OR Spanish GC)
– Transport variables (Inflatable Boat → Fibre Boat OR Desert vehicle OR Train carriage)
– Specific incident types (rare in maritime context may be common in desert/land contexts)

The underlying structure and verification procedures remain unchanged.

Principle 3: Source Architecture is Modular

The three-stream data collection model (Official + Media + NGO monitoring) works across contexts:

– Stream 1 adapts: Which authorities produce public statistics varies (EU countries publish regularly; some non-EU partners less transparent)

– Stream 2 adapts: Media landscape differs (English-language outlets available in most countries; local media language varies)

– Stream 3 adapts: Available monitors vary (IOM present in some contexts; local NGOs in others)

The logic remains: triangulate across available sources → verify → resolve conflicts → document.

14.2 Step-by-Step Adaptation Guide

STEP 1: Geographic Adaptation

Current (Aegean) Adaptation Needed

Example: Central Mediterranean (Italy-Libya)

Area Greece Area of destination/ arrival

Area Italy (Sicily, Lampedusa)

Area Turkey Area of origin/ departure

Area Libya, Area Egypt

Action steps:

Map geographic boundaries for your context
Identify all named locations (islands → cities/towns; maritime zones → regions)
Adapt Area Turkey and Area Greece variable names to your context
Create codebook with local geographic terms (spell out clearly; avoid abbreviations)

STEP 2: Authority Adaptation

Current (Aegean) Adaptation Needed

Central Med Example

Area Greece Area of destination/ arrival

Area Italy (Sicily, Lampedusa)

Turkish Coast Guard Authority of origin

Libyan Coast Guard

Action steps:

Identify ALL state and non-state authorities operating at border in your context
Document their roles (enforcement, rescue, coordination)
Note their transparency/public reporting levels (some publish regularly; others don't)

Adapt variable names: Apprehension by TA → Apprehension by [Name of Authority]

Create codebook listing all authority abbreviations and full names

STEP 3: Source Adaptation

Current (Aegean) Adaptation Process

Central Med

Greek news outlets (Sto Nisi, etc.)

Which media cover border issues?

Al Jazeera, Politico, Mediterranean media

Local NGO monitoring networks

Which organizations document incidents?

Alarm Phone, maritime NGOs

Turkish Coast Guard official data

Which authorities publish data?

Libyan CG, Italian CG (both limited transparency)

IOM Missing Migrants Project

Which international monitors are active?

IOM (yes), UNHCR, MSF

Action steps:
Conduct desk research: Which authorities publish incidents? In what formats? How frequently?
Identify media landscape: Which outlets cover border issues? In what language(s)? What is their editorial approach?
Map NGO/international monitoring: Which organizations operate in your context? What data do they collect? How accessible are they?
Assess data accessibility: Some data freely public; some requires FOIA/requests; some restricted
Build sustainable source architecture: Use regularly-updated sources; avoid relying on single outlet

14.3 Sectoral Adaptation (Non-Maritime Contexts)

This toolkit was developed for maritime incidents in the Aegean. Adapting to land borders, desert crossings, or other transport modalities requires specific adjustments in the incident types. However, the methodological toolkit serves as a guide on how to achieve this.

ANNEX I - CODEBOOK

Date
(YYYY-MM-DD)

Clearly stated incident date or explicit day reference

Time
(HH:MM, 24-hr)

Hour of incident from any reliable source

Area Turkey
Text

Named region, island, or maritime area in Turkey

means_of_travelling
Text

Type of boat, mean of travelling

Area Greece
Text

Named region, island, or maritime area in Greece

land_sea
Categorical (dummy)

0 = Land; 1 = Sea

found_by_GA
Categorical (dummy)

0 = No; 1 = Yes

apprehension_by_TA
Categorical (dummy)

0 = No; 1 = Yes

pushback
Categorical (dummy)

0 = No; 1 = Yes

Boat chase
Categorical (dummy)

0 = No; 1 = Yes

Gun shot fired
Categorical (dummy)

0 = No; 1 = Yes

shipwreck
Categorical (dummy)

0 = No; 1 = Yes

number_dead
Numerical

Integer ≥ 0

number_missing
Numerical

Integer ≥ 0

criminalization_occurred
Categorical (dummy)

0 = No; 1 = Yes

Charges
Text

Text listing formal and suspected charges

Additional_information
Text

Free-text narrative

source_verification
Categorical

singlesource_official; singlesource; multisource_official; multisource

Source_links
Text

URLs separated in columns hyperlinked

ANNEX II. TECHNICAL MEMORANDUM OF UNIFIED MONITORING TOOLKIT

1. Scope

The aim of this Technical Memorandum is to inform readers of the Unified Monitoring Toolkit on the method used to transform conflicting data into the Toolkit's final findings. It acts as a guarantee that the methodology used to merge data is transparent, auditable and reproducible by other data scientists, lawyers and organizations who are seriously interested in the Toolkit's methodology. This Technical Memorandum has been written by Collective Aid's data science partner, Erica Grauso.

2. Purpose of the Unified Monitoring Toolkit

There are considerable hurdles to documenting and estimating migration-related incidents in the Aegean Sea, one of Europe's most contested maritime borders, where official statistics and media reporting contain significant gaps. This pilot project focuses on the Aegean Sea and introduces a standardized, transparent methodology and toolkit for collecting, coding, verifying, and ethically using incident-level data on border violence, maritime deaths, apprehensions, rescues, and related incidents.

The scope covers estimates of migration-related deaths between January 2021 and December 2025, drawing on multiple sources to produce an incident-based dataset tracking incidents across Greek islands and Turkish coastal regions.

3. Data Sources

The dataset results from the analysis of incident-related data from various cross-checked sources, encompassing monitoring NGO's, news reporting media, and Coast Guards official figures. To achieve the final findings of the dataset, and to achieve as full of a picture as possible of the events on the ground, triangulation across three different data streams was used, each maintained separately until verification procedures were applied.

The three streams are:

– Official statistics (the Turkish Coast Guard (TCG), and the Greek Hellenic Coast Guard (HCG). This acted as the primary stream.

– Media and NGO monitoring, including Greek media, international media outlets, and NGO reporting.

– Cross-verification with data from the IOM Missing Migrant Project, comparative analysis, identification of duplicates and reconciliation.

Due to the sensitivity of the incident data, gaining access to these sources was not unrestricted nor without challenges. For instance, incident reporting by the Turkish and Greek Coast Guards and news reporting were not present for every incident. Some of the gaps in reporting by these sources may be linked to an institutional bias, which will be explored in a later section.

4. Data Preparation and Verification

A thorough coding system was set up to standardalize records of each incident, by type, location, dates and other identifying factors. Each value was clearly defined with both inclusion and exclusion criteria.

In addition, coding rules were laid out to approach gaps and unknowns in a standard manner.

Finally, a conflicting information resolution hierarchy was set up to deal with inconsistencies, such as conflicting data probably referring to the same incident, or incidents spanning more than one day. Through this systematic verification procedure, the final verified figures were achieved despite the differences in data reporting in each source. All conflicts documented in the Additional Information field with full audit trail.

To address variation in data quality across sources, this project applies a four-tier Verification System that classifies migration-related incidents according to source type and level of corroboration. The system distinguishes between single official reports, single non-official reports, multisource official corroboration, and multisource non-official corroboration, enabling analysts to filter incidents by confidence level and assess the robustness of findings. Each tier specifies acceptable analytical uses, required caveats, and limitations, ensuring that uncertainty is documented transparently and that conclusions drawn from the data are defensible. A standardized decision tree guides tier assignment, while all sources and discrepancies are systematically documented within the dataset to support reproducibility and responsible reuse.

This project applies a standardized protocol for handling missing and uncertain data to ensure transparency and analytical integrity. All unknown values are explicitly coded as missing (NA), and absence of information is never interpreted as zero. To make data limitations visible, all summary outputs include a required percentage-missing indicator for each variable, guiding appropriate interpretation and restricting claims to observed subsamples where necessary. The protocol also specifies consistent treatment of common uncertainty scenarios—including ranges, unspecified times, and indeterminate counts - by applying conservative coding rules, documenting uncertainty in accompanying fields, and requiring explicit analytical caveats. This approach supports reproducibility, prevents false precision, and enables responsible use of incomplete migration incident data.

5. Statistical Method

This section outlines the statistical approaches that could be applied to the dataset to summarize patterns, test associations, and explore predictive relationships. Variables and some of the data tests referenced are documented in the Data tab of the accompanying Google Sheet file.

The dataset lends itself to descriptive statistical analysis to provide an overall characterization of recorded incidents. Such analyses can include the calculation of frequencies and percentages for different incident types, such as pushbacks, rescues, boat chases, and shipwrecks, as well as the frequency of involvement by different actors, including the Greek Coast Guard, Turkish Coast Guard, and Frontex. For numerical variables, descriptive measures such as the mean, median, minimum, maximum, standard deviation, and selected percentiles can be used to summarize the number of people involved per incident, including children, deaths, and missing persons.These descriptive statistics may also be stratified by key categories, such as land versus sea incidents or incident type, allowing for comparisons across contexts. Together, these summaries can illustrate how common different types of incidents are, how frequently particular actors appear, the typical scale of incidents, and how outcomes vary across settings.

Beyond descriptive summaries, the data allow for a range of inferential statistical analyses to assess relationships between variables and test specific hypotheses. Associations between categorical variables can be examined using chi-square tests of independence, for example to assess whether pushbacks are more likely to occur in sea incidents than land incidents, whether fatalities are more likely when certain actors are involved, or whether shipwrecks occur more frequently during particular seasons. To model binary outcomes such as the occurrence of a pushback, rescue, or death, logistic regression models can be applied, using predictors such as incident location, actor involvement, group size, and vessel type. These models make it possible to estimate the

likelihood of specific outcomes while accounting for multiple explanatory variables simultaneously.

The dataset may also support comparisons of numerical variables across groups. Depending on the distribution of the data, t-tests or Mann–Whitney U tests can be used to compare outcomes between two groups, such as the number of people involved in land versus sea incidents, while analysis of variance (ANOVA) can be used to compare outcomes across multiple categories, such as average deaths across different incident types. In addition, correlation and regression analyses can be applied to examine relationships between numerical variables, for example to assess whether the number of children present in an incident is associated with a higher likelihood of rescue or whether group size is related to fatal outcomes.

Any inferential analyses conducted using these methods would rely on standard statistical assumptions appropriate to each technique, including independence of observations and relevant distributional assumptions. Uncertainty can be quantified through confidence intervals and sensitivity analyses, particularly where results may be affected by missing data or alternative model specifications.

6. Biases and Limitations

This dataset is subject to important temporal, geographic, and source-related limitations that affect interpretation. Coverage is strongest for major Greek islands and weaker for smaller islands, remote coastal areas, and parts of the Turkish mainland, reflecting uneven monitoring and media presence. Official data capture only detected incidents and may reflect institutional reporting incentives, while media and NGO coverage is uneven across languages and time periods. As a result, selection and authority biases are likely, and documented figures should be understood as conservative estimates rather than complete counts. All research using this dataset must include a transparency statement acknowledging these limitations and noting that true incident frequency, mortality, and harm are likely higher than reported.

7. Reproducibility details

For the readers who may wish to produce incident-based datasets compatible with other ones using the same Toolkit framework, enabling comparative analysis and collaboration while maintaining data quality standards, the following are some helpful details.

This database requires a two-tier software and storage infrastructure that balances data security with public accessibility. Raw data are collected and stored using encrypted, login-protected data collection platforms (e.g., Kobo) to support internal verification, case analysis, and legal documentation. A separate working dataset is produced for public use and published in an openly accessible repository with a stable DOI, alongside a complete codebook and methodological toolkit. This public dataset contains fully anonymized, incident-level data with documented source types and all variables, enabling global access, reproducibility, and reuse for

independent research.

ANNEX III - DATASET

Complementary data were collected alongside the development of the methodological toolkit to support its implementation and practical use. All documented incidents are publicly accessible and can be explored via a Google Sheet with embedded hyperlinks, as well as downloaded in HTML and CSV formats from the Google Sheet for further review and data analysis.

Further, some of the suggested statistics has been tested to showcase how the data could be used.

Incident datasets:

– Incidents 2024
Google Sheet – 2024 incidents

– Incidents 2025
Google Sheet – 2025 incidents

– Data analysis
Google Sheet - Graphs and Charts