People have been dying while trying to cross the external borders of the European Union (EU) for three decades. However, exactly how many have died and how dangerous irregular border crossing routes are is unknown. Quantitative data on border deaths is notoriously poor (Weber and Pickering 2011; Brian 2014; Last and Spijkerboer 2014; Heller 2015; Williams and Mountz 2016), in part due to the circumstances in which irregularised border-crossers die or disappear (Grant 2011; Tazzioli 2015), in part due to the lack of official reporting of these deaths (Chapter 2). Moreover, there are significant differences in trends between existing datasets of EU border deaths (Chapter 2: Figure 2.4).

This chapter presents a two-staged research into EU border death data. The first study investigates: What are the main sources of border death data in academic literature? The second investigates: How reliable is this data for assessing mortality among irregularised border crossers along the southern external borders of the EU? This chapter is organised as follows: First, the methodology and findings of each study are presented in turn. Then, in the subsequent section, the findings are discussed and the question whether more reliable estimates are attainable is also explored. These questions are particularly poignant when death data is used to assess the impact of existing policies.

---

99 Co-authored with Joke Harte.
Study 1: What are the main sources of border death data in academic literature?

Method

To discover the sources of quantitative data used in academic literature concerned with EU border deaths, relevant academic works were searched for border death data. These works were selected from an exhaustive search (no temporal filters) of major publishers of English-language, peer-reviewed journals and academic volumes for literature concerned with EU border deaths, as well as a number of unpublished works attained by the author via email and ad hoc online searches between 2013-2016. A detailed description of the selection is provided in Chapter 5). In short, works were selected for inclusion in the study if they sought to explain the phenomenon, in particular, the relation between deaths and policy. The search and selection process elicited 39 works relevant to the study. The results of the search for border death data in these 39 works were examined on the original source and use of data.

Findings

There are important empirical contributions among the literature studied, based on qualitative field work among local, national and European authorities and other actors in border communities, as well as desk-based analytical research. The intensity of field work varies but tends to be single-sited, incident-focused, and/or short term (Albahari 2006; Carling 2007; Klepp 2011; Spijkerboer 2013; Topak 2014; Heller 2015; Kovras and Robins 2016; Squire 2016; Oliveri 2016). Desk-based research includes gathering information for a comprehensive overview (Kiza 2008; Weber and Pickering 2011; Weinzierl and Lisson 2007; Spijkerboer 2007, 2013; Grant 2011; Basaran 2014; Albahari 2006; Cuttitta 2004) and complex, technical analyses (Pickering and Cochrane 2012; Williams and Mountz 2016; Blanchard, Clochard and Rodier 2012; Heller and Pezzani 2016; Heller 2015). The remaining works are theoretical in the sense that they do not question or explore the data that they reference. Overall, death data is chiefly used to illustrate arguments that are based on theoretical constructions of the situation.

Figure 4.1 illustrates the referenced and original sources of border death data in academic literature. Sources were included in Figure 4.1 if they were cited by 3 or more reviewed works. Where there were several citations for essentially the same source, these were combined. For example, Brian (2014) and Brian and Lazcko (2014) are publications linked to the International Organisation for Migration’s (IOM) Missing Migrants Project, with which UNHCR corroborates its aggregated data (Al Tamimi et al 2017). Scholars marked in boxes are also frequently cited as sources of quantitative data.
Data on border deaths along southern EU external borders

Study 1: What are the main sources of border death data in academic literature?

Method
To discover the sources of quantitative data used in academic literature concerned with EU border deaths, relevant academic works were searched for border death data. These works were selected from an exhaustive search (no temporal filters) of major publishers of English-language, peer-reviewed journals and academic volumes for literature concerned with EU border deaths, as well as a number of unpublished works attained by the author via email and ad hoc online searches between 2013-2016. A detailed description of the selection is provided in Chapter 5. In short, works were selected for inclusion in the study if they sought to explain the phenomenon, in particular, the relation between deaths and policy. The search and selection process elicited 39 works relevant to the study. The results of the search for border death data in these 39 works were examined on the original source and use of data.

Findings
There are important empirical contributions among the literature studied, based on qualitative field work among local, national and European authorities and other actors in border communities, as well as desk-based analytical research. The intensity of field work varies but tends to be single-sited, incident-focused, and/or short term (Albahari 2006; Carling 2007; Klepp 2011; Spijkerboer 2013; Topak 2014; Heller 2015; Kovras and Robins 2016; Squire 2016; Oliveri 2016). Desk-based research includes gathering information for a comprehensive overview (Kiza 2008; Weber and Pickering 2011; Weinzierl and Lisson 2007; Spijkerboer 2007, 2013; Grant 2011; Basaran 2014, 2015; Last & Spijkerboer 2014). The remaining works are theoretical in the sense that they do not question or explore the data that they reference. Overall, death data is chiefly used to illustrate arguments that are based on theoretical constructions of the situation.

Figure 4.1 illustrates the referenced and original sources of border death data in academic literature. Sources were included in Figure 4.1 if they were cited by 3 or more reviewed works. Where there were several citations for essentially the same source, these were combined. For example, Brian (2014) and Brian and Lazcko (2014) are publications linked to the International Organisation for Migration’s (IOM) Missing Migrants Project, with which UNHCR corroborates its aggregated data (Al Tamimi et al 2017). Scholars marked in boxes are also frequently cited as sources of quantitative data.

Figure 4.1 Sources of EU border death data in academic literature

As Figure 4.1 shows, most death data in the literature comes from one original source – news reports. There is a small group of academics who have used official aggregated statistics – which are published by national authorities on an ad hoc basis (Last & Spijkerboer 2014) – or death registries (Spijkerboer 2013). The Deaths at the Borders Database (DatBD) is the first and only dataset sourced from state authorities, providing a vital alternative to news media-sourced data (Chapter 2); however, it was only published open source in 2015 and is not kept up-to-date. The IOM’s Missing Migrants Project also corroborates their primary source (news reports) with local officials whenever possible. Some academics use news reports directly for data on particular incidents, while others use news reports to build datasets themselves (Cuttitta 2004; Kiza 2008; Williams and Mountz 2016). Most works, however, reference available datasets that compile news reports on deaths: UNITED, the Fortress Europe blog by Italian journalist Gabriele del Grande, the collections of Institute of Race Relations (IRR) or Asociación Pro Derechos Humanos de Andalucía (ADPHA), and IOM’s Missing Migrants Project. Fargues and his colleagues at the Migration Policy Center in Florence are also frequently referenced, not as a source of data but for the mortality trends they have published in policy briefs, which are derived from the Fortress Europe blog data. In addition, as shown in Table 4.1, for the period 2000-2016, a team of journalists merged UNITED and the Fortress Europe blog into a sixth dataset, the Migrant Files, which was not included in Figure 4.1 because it was only cited by two reviewed works. As there are a number of datasets and empirical academic contributions, it appears as though there is a wide array of sources of data on EU border deaths. However, ultimately, almost all descriptive statistics, quantitative analyses and narrative details of incidents are sourced primarily – if not exclusively – from news reports.
Table 4.1 Publically available datasets on EU border deaths

<table>
<thead>
<tr>
<th></th>
<th>UNITED100</th>
<th>FEblog101</th>
<th>Migrants Files102</th>
<th>IOM MMP103</th>
<th>DatBD104</th>
</tr>
</thead>
<tbody>
<tr>
<td>year data collection commenced</td>
<td>1993</td>
<td>2006 (TBC)</td>
<td>2013</td>
<td>2013</td>
<td>2014</td>
</tr>
<tr>
<td>geographical scope</td>
<td>EU and neighbouring countries</td>
<td>EU and neighbouring countries</td>
<td>EU and neighbouring countries</td>
<td>global</td>
<td>southern EU Member States</td>
</tr>
<tr>
<td>frequency updated</td>
<td>annually</td>
<td>irregularly, last updated February 2016</td>
<td>N/A</td>
<td>weekly</td>
<td>N/A</td>
</tr>
<tr>
<td>primary source</td>
<td>news reports</td>
<td>news reports</td>
<td>UNITED &amp; Feblog</td>
<td>news reports</td>
<td>death certificates</td>
</tr>
<tr>
<td>other sources</td>
<td>Feblog, organisations in the network, researchers</td>
<td>NGOs</td>
<td>Puls (University of Helsinki)</td>
<td>local officials, IOM field staff, IOs, NGOs, survivors</td>
<td>cadaver inspection/ autopsy reports, burial permits, cemetery registers, coast guard operation reports</td>
</tr>
<tr>
<td>working definition of border death</td>
<td>deaths attributable to the policies of Fortress Europe, on the way, after arrival, during deportation or after repatriation</td>
<td>irregular migrants who do not survive the journey to Europe</td>
<td>refugees and migrants who died in their attempt to reach or stay in Europe</td>
<td>deaths that occur at physical borders and while en route to an international destination</td>
<td>people who died attempting to cross the southern EU external borders</td>
</tr>
</tbody>
</table>

Although the literature recognises the lack of (official) data (Fekete 2004; Weinzierl and Lisson 2006; Kiza 2008; Weber and Pickering 2011; Pickering and Cochrane 2012; Saucier and Woods 2014; Shields 2015; Kovras and Robins 2016) and the unreliability of news-sourced data (Webber 2004; Albahari 2006; Carling 2007; Grant 2011; Weber and Pickering 2011; Pickering and Cochrane 2012; Oliveri 2016; Topak 2014), the specific limitations and weaknesses of death data are rarely discussed in any detail (cf Last and Spijkerboer 2014; Tazzioli 2015; Williams and Mountz 2016). There is little done to scrutinize the effects of

---

100 [http://unitedagainstrefugeedeaths.eu/about-the-campaign/about-the-united-list-of-deaths/](http://unitedagainstrefugeedeaths.eu/about-the-campaign/about-the-united-list-of-deaths/)
101 [http://fortresseurope.blogspot.nl/p/la-strage.html](http://fortresseurope.blogspot.nl/p/la-strage.html)
103 [https://missingmigrants.iom.int/](https://missingmigrants.iom.int/)
104 [http://www.borderdeaths.org](http://www.borderdeaths.org)
unreliable media reporting of facts, aside from citing corroborating references, which often rely on the same data. Even Carling (2007), who models different scenarios reflecting the possible biases of reported numbers, assumes that if bodies are found or people reported missing by survivors this information is accurately and consistently reported in the news. Academics who are aware of and acknowledge the problems of the available data, nonetheless draw policy-relevant conclusions from that data.

On a related note, studies and statistics from other border regions marking fault lines between the Global North and Global South – especially the US-Mexico border – are frequently cited as evidence supporting claims about deaths along the EU external borders. This occurs both in literature that takes a comparative perspective, looking at more than one border region in which deaths-during-border-crossing occur (Mountz and Loyd 2013; Weber and Pickering 2011; Pickering and Cochrane 2012; Squire 2016; Jones and Johnson 2016), and also in literature that aims to explore only the European region or even a particular national context within that region (Fekete 2004; Webber 2004; Lutterbeck 2006; Albahari 2006; Kiza 2008).

**Study 2: How reliable is existing border death data?**

**Method**

As shown in Figure 4.1, academic literature relies heavily on news reports as the original source of data for EU border deaths. This second study aims to investigate the reliability of this main source by comparing news-sourced data with another source. The only database of border deaths not sourced primarily from news reports is DatBD. Sourced from death management systems, DatBD records deceased irregularised border-crossers whose bodies were found or brought to municipalities of Member States along the southern EU external borders between 1990-2013. The data collection process, compilation of the Database and its limitations are published in detail elsewhere (Chapter 2). For the purposes of this study, due to the nature of its source, DatBD does not cover all EU border death cases but it represents a baseline for the minimum number of deaths because there is no chance that a death certificate is issued without evidence of a death. UNITED’s list of deaths was selected to represent news-sourced data in the comparison for a number of reasons. It is the longest-standing dataset of border deaths, covering much of the same period as DatBD (see Table 4.1). It is also the most commonly cited dataset in academic literature. Finally, UNITED compiles its data in an accessible format, coded according to the date of the incident, the number of deceased and missing persons, the personal information of deceased and missing persons, the circumstances and place of death and the source of data. Both datasets are ordered chronologically. This study consists of

---

105 Although the earliest death recorded by FEblog is dated in 1988, this was the result of a historical search conducted by the author of the blog, Gabriele del Grande, when he first began compiling news reports on deaths online in 2006.

106 There has been a shift in sources of data in academic literature. Since 2015, IOM’s MMP appears to be succeeding as the primary reference for border death data. However, IOM’s MMP only began recording deaths at the end of 2013 so it is not useful for analysis of long-term trends over the three decades that border deaths have been occurring at the EU’s southern borders.
matching records between UNITED and DatBD and analysing the overlap between them and gaps in each.

The two datasets cover different deaths related to their primary sources, temporal coverage and their working definitions of ‘border death’ (see Table 4.1). To enable comparison, records were selected from UNITED in accordance with the narrower working definition of DatBD: people who died attempting to cross the southern EU external borders. In addition, UNITED records were coded according to whether bodies had been found or not and whether they had been found in/near or brought to an EU member state or not. These variables were used to interpret the results. DatBD and UNITED cover slightly different periods (1990-2013 versus 1993-2017), so only records from 1993-2013 were selected from each dataset for comparison. In sum, 3,030 death records from DatBD were compared (manually, following a Protocol) with 13,397 death records from UNITED.

Perfect matching between UNITED and DatBD is impossible primarily because the majority of deceased persons remain unidentified (see Chapters 2 and 3). However, for the purposes of assessing the reliability of news-sourced data for analysing EU border mortality, perfect record matching is unnecessary. It is not important whether the news reports the same individual whose body was found; it is important whether the news reports the story of a body being found (as recorded in DatBD), including the right number of bodies and the right information about those bodies. In other words, the aim was not to test whether UNITED has ‘recaptured’ the same deceased individuals ‘captured’ by DatBD, but whether UNITED has ‘recaptured’ the news of individuals ‘captured’ by DatBD. To this end, the system dependence between UNITED and DatBD (a body being found increases the chances of that death being reported in the news) was exploited in order to match records. For example, the news may report a body being found of North African origin on a given date, while the DatBD records an identified man from Morocco dying on the same date. It can be assumed that the journalist reported the story before the identity of the man was known and guessed his region of origin based on his appearance or the origin of his survived fellow passengers. For another example, the news may report bodies of Sub-Saharan Africans washing up on the beach on a given date, believed to be 3 of 6 missing from an earlier shipwreck, while the DatBD records 4 decomposed bodies of black men who died between a week and 2 weeks before. In this case, it can be assumed that the death management system records physical description while the journalist uses these details to guess the region of origin, and that the journalist reported the dates relevant to the story of the body being found while the death management system records the date of death estimated by the pathologist conducting the cadaver examination. It can also be assumed that the journalist reported the story before the fourth body was found. Thus, system dependence between UNITED and DatBD was integrated into the Protocol used to match records between the datasets.

### Table 4.2 Classifications for comparison of records between UNITED and DatBD records

<table>
<thead>
<tr>
<th>Category</th>
<th>Description of circumstances/cause of death</th>
<th>Number of bodies found</th>
<th>Place (town/small island) / region / location (e.g. hospital/beach/boat)</th>
<th>Date of death / date found / date registered</th>
<th>Personal details (sex/age/origin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitive match</td>
<td>All available data is the same or very similar and the difference is explainable (e.g. 1/2 days difference in date, North African and Moroccan, etc), no conflicting data, same number of bodies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial match</td>
<td>Definitive match except different number of bodies reported by UNITED</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possible match</td>
<td>Some conflicting data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No match</td>
<td>No corresponding record or conflicting data too significant or better fit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The datasets were divided into three sub-sets based on country for ease of comparison: (1) Greece, (2) Italy and Malta, and (3) Spain and Gibraltar. The Protocol set the criteria and procedure for matching. Records were linked on the basis of the ‘best fit’ of the following information:

- Date of death / date found / date registered
- Place (town/small island) / region / location (e.g. hospital/beach/boat)
- Number of bodies found
- Personal details (sex/age/origin)
- Description of circumstances/cause of death

This information is not available for all UNITED or DatBD records and differences are common between details of the two datasets as a result of their different sources. Four classifications were available as shown in Table 4.2. For the purposes of analysis, definitive matches, partial matches and possible matches were all treated as matched records. The difference in the number of bodies found among partial matches were taken into account in all calculations.

**Table 4.2 Classifications for comparison of records between UNITED and DatBD records**

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definitive match</td>
<td>All available data is the same or very similar and the difference easily explainable (e.g. 1/2 days difference in date, North African and Moroccan, etc), no conflicting data, same number of bodies</td>
</tr>
<tr>
<td></td>
<td>Exception can be made for records with conflicting data if an unusual detail is the same in both records (e.g. name, particular circumstance of death or discovery of the body)</td>
</tr>
<tr>
<td>Partial match</td>
<td>Definitive match except different number of bodies reported by UNITED and recorded by DatBD</td>
</tr>
<tr>
<td>Possible match</td>
<td>Some conflicting data</td>
</tr>
<tr>
<td></td>
<td>Insufficient data to determine match definitively (e.g. if DatBD does not record date died or date found, or if UNITED does not report any personal details with a vague description of cause of death)</td>
</tr>
<tr>
<td></td>
<td>When a DatBD case could match with more than one UNITED record (e.g. because of possible double-counting in UNITED). However, a UNITED record should never be matched with more than one DatBD case unless they are part of the same incident.</td>
</tr>
<tr>
<td>No match</td>
<td>No corresponding record or conflicting data too significant or better fit with another case</td>
</tr>
</tbody>
</table>
The procedure established by the Protocol begins by filtering records in a sub-set by year. This reduces the number of cases being compared at any given time to a manageable number and was intended to reduce the likelihood of human error. In a first round of comparison, UNITED records were searched for potential matches with a particular DatBD record based on the information listed above; then in a second round, all DatBD records for that year were searched for potential matches with remaining unmatched UNITED records. In this second round, UNITED records with dates close to the beginning or end of the year were also compared with DatBD records from the previous or next year, respectively. In this way, it was possible to determine whether each DatBD was matched according to the principle of ‘best fit’. After the second round, remaining unmatched DatBD records were recorded as no match. Results were recorded in a table. Once a potential match was identified, the UNITED record number was entered into the column next to the corresponding DatBD case number(s), all similarities and differences in the information recorded in the matched records from each dataset were listed in a third column, and the category of match was coded in a fourth column. The corresponding records were also colour-coded according to the category of match in each dataset.

Record matching was done manually due to the considerable nuance in the data, the fact that UNITED records incidents while DatBD records individuals, and because the task was not overwhelming given the relatively small size of the datasets. The disadvantage is that no sensitivity analysis was possible. Instead interrater reliability of the matching Protocol was established by a second matcher (Korhonen 2017) who independently scored a sample of the cases (n=303, Kappa=0.848, percentage agreement=89.4%). The sample was drawn from each sub-set by selecting every 10th case, starting from a row number between 1-9 randomly generated at www.graphpad.com/quickcalcs/randomN2/. The sample contains 10% of DatBD cases: 79 cases from Greece, 123 cases from Italy/Malta, and 101 cases from Spain/Gibraltar. Most of the disagreement in results between the original and second matchers concerned the assignment of categories 1-3 (see Table 4.2). For the purposes of further analysis, however, categories 1-3 are grouped as ‘match’. In terms of the classification of match/no match, only 3 cases (<1%) were coded differently between the test sample and the original comparison. This means that even the low level of subjectivity in implementing the Protocol has no impact on the findings presented below.

Findings

Table 4.3 and Figure 4.2 show the results of the comparison between DatBD and UNITED. Each dataset has gaps reflecting their limitations (see Table 4.3). DatBD only records deceased bodies found or brought to southern EU Member States; it does not record missing persons or bodies found or brought to non-EU Member States. UNITED only records those deaths and disappearances reported in the news. The 11,568 missing from DatBD are largely a result of the transparent systematic bias of its original source (for details, see Chapter 2). The 1,201 deaths missing from UNITED provide evidence, for the first time, of the limitations resulting from the unpredictable bias of news-sourced data on EU border deaths. The total number of deaths missing from both datasets remains unknown.
Table 4.3. Overlap between and gaps in UNITED and DatBD, 1993-2013

<table>
<thead>
<tr>
<th>Recorded in UNITED</th>
<th>Recorded in DatBD</th>
<th>Missing from DatBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,829</td>
<td></td>
<td>11,568</td>
</tr>
<tr>
<td>1,201</td>
<td></td>
<td>unknown</td>
</tr>
</tbody>
</table>

The shaded area in Figure 4.2 shows the matched records that appear in both datasets per year, while the lines show the remaining unmatched records from each dataset, illustrating the under-counting summarised in Table 4.3. The average proportion of bodies found that were reported by the news per year is 50.6% (σ=18.8%, min 0.0%, max 86.2%, n=21). UNITED and DatBD show different trends over the 21 years in which they are comparable. The overall increase in deaths is significantly more gradual in DatBD than in UNITED and, in general, the number of deaths per year captured by DatBD are more stable than those captured by UNITED. Although many of the peaks and dips occur in the same years, demonstrating some consistency between the two datasets, there are two periods in which opposite trends can be observed: 1996-1998 and 2004-2007. This demonstrates the significance of better understanding the bias inherent to news-sourced data.

While Table 4.3 and Figure 4.2 reveal that both datasets under-count border deaths, Figure 4.3 shows the proportion of DatBD cases matched in UNITED over time. There has been a general upward trend, an improvement in UNITED’s coverage of EU border deaths. It is possible that the persons responsible for collecting news reports and compiling UNITED improved their methodology with experience, or that discovery of news reports has become easier with the development of online search engines. However, the general upward trend could also reflect an increase in coverage of border deaths by news media, perhaps related to an increase in public/political attention to the phenomenon. Either way, despite the general upward trend, Figure 4.3 shows that the proportion of bodies found that are reported in the news varies considerably from year to year and can drop (e.g. between 2004-2008 and 2010-2012). The drop in proportion of matches from 66.4% in 2009 to 42.0% in 2010 is particularly interesting as it partially explains the dramatic drop in the number of deaths in UNITED over the same years, as shown in Figure 4.2.
Figure 4.2 The overlap between UNITED and DatBD

Figure 4.3 Scatterplot of the percentage of bodies found (DatBD) that were reported in the news (UNITED)

The under-counting of bodies managed by European authorities varies not only per year but also notably from route to route. Figure 4.4 maps the proportion of bodies managed by local authorities in southern EU member states that were reported in the news according to route. The Atlantic route (between West Africa and the Canary Islands) has by far the highest proportion of matches (81.7%). This was even higher during the peak years of the Atlantic route; between 1999-2009, 85.4% of the 308 bodies managed by local authorities in the Canary Islands were reported by the news. The Central Mediterranean route has the second highest proportion of matches (72.2% of 710 bodies). However, there is more variation in the proportion of bodies found and reported in the news over the years on the Central Mediterranean route than on the Atlantic route.107

In contrast, as Figure 4.4 shows, comparison of the Adriatic land route resulted in no matches. But there were relatively few deaths recorded on this route in both datasets. The proportion of bodies found reported in the news on the remaining routes varies from 38.6% along the Greek-Turkish land border and 54.9% in the Aegean Sea, to 47.3% in the Adriatic Sea, to 45.0% on the Western Mediterranean sea route between Morocco and mainland Spain and 44.4% in Ceuta and Melilla, the Spanish enclaves in Morocco.

107 For Atlantic route, 1993-2013, $\sigma = 39.8\%$ (min 0.0%, max 100%, n=21) but for the peak years, 1999-2009, $\sigma = 17.0\%$ (min 44.4%, max 100%, n=11). The difference between the overall standard deviation and that of the peak years is explainable by the low number of bodies found outside of the peak years and thus the sensitivity of the percentage to whether or not these few records were matched. For Central Mediterranean route, 1993-2013, $\sigma = 29.2\%$ (min 0.0%, max 92.2%, n=21). The Central Mediterranean route does not have one clear peak period like the Atlantic route but if you take the same years, 1999-2009, $\sigma = 22.9\%$ (min 12.5%, max 87.0%, n=11).
The under-counting of bodies managed by European authorities varies not only per year but also notably from route to route. Figure 4.4 maps the proportion of bodies managed by local authorities in southern EU member states that were reported in the news according to route. The Atlantic route (between West Africa and the Canary Islands) has by far the highest proportion of matches (81.7%). This was even higher during the peak years of the Atlantic route; between 1999-2009, 85.4% of the 308 bodies managed by local authorities in the Canary Islands were reported by the news. The Central Mediterranean route has the second highest proportion of matches (72.2% of 710 bodies). However, there is more variation in the proportion of bodies found and reported in the news over the years on the Central Mediterranean route than on the Atlantic route.107

In contrast, as Figure 4.4 shows, comparison of the Adriatic land route resulted in no matches. But there were relatively few deaths recorded on this route in both datasets. The proportion of bodies found reported in the news on the remaining routes varies from 38.6% along the Greek-Turkish land border and 54.9% in the Aegean Sea, to 47.3% in the Adriatic Sea, to 45.0% on the Western Mediterranean sea route between Morocco and mainland Spain and 44.4% in Ceuta and Melilla, the Spanish enclaves in Morocco.

107 For Atlantic route, 1993-2013, $\sigma=39.8\%$ (min 0.0%, max 100%, n=21) but for the peak years, 1999-2009, $\sigma=17.0\%$ (min 44.4%, max 100%, n=11). The difference between the overall standard deviation and that of the peak years is explainable by the low number of bodies found on this route outside of the peak years and thus the sensitivity of the percentage to whether or not these few records were matched. For Central Mediterranean route, 1993-2013, $\sigma=29.2\%$ (min 0.0%, max 92.2%, n=21). The Central Mediterranean route does not have one clear peak period like the Atlantic route but if you take the same years, 1999-2009, $\sigma=22.9\%$ (min 12.5%, max 87.0%, n=11).
The proportion of bodies found reported by the news does not appear to be related to the country in which the body was found or brought to shore, indicating that it is unrelated to national language, press, or death management systems. Rather, the chance that bodies found are reported by the news appear to be related to characteristics of the route. The three routes with the lowest proportion of DatBD records ‘captured’ by UNITED are the three land routes across the southern EU external borders. The average for land routes (27.7%) is well below that of sea routes (60.2%). There are two irregular migration routes between Turkey and Greece: the land route across the Evros/Meric river and the sea route across the Aegean to the Greek islands. On the land route, 55.0% of UNITED records were matched with DatBD records, while 61.0% of DatBD records were not reported in the news. In contrast, on the sea route, 19.7% of UNITED records were matched with DatBD records, and 38.7% of DatBD-recorded deaths were unreported in the news. These differences between land and sea routes mirrors the tendency in policy documents concerning border deaths to focus on the ‘loss of life at sea’ (see Chapter 5). The results of the comparison indicate that the dangers of sea routes claim more public attention than those associated with land routes.

The two routes with the highest proportion of DatBD records matched in UNITED are the longest sea routes, famous for shipwrecks of boats carrying dozens (Atlantic), hundreds (Central Mediterranean) of people. On these longer routes traversed by bigger boats carrying many passengers, a report of one fatal incident in UNITED often matches with many individual death records in DatBD. For example, the high proportion of bodies found reported by the news in 2013 (see Figure 4.2) is to a large extent attributable to a shipwreck that occurred on 3rd October within the territorial waters of Lampedusa. Due to the proximity of the shipwreck to the island and the provision of national support in the form of personnel, equipment and expertise, 364 dead bodies of victims were recovered from this shipwreck. These factors, among other things, also contributed to the 3rd of October 2013 becoming arguably the most

Figure 4.4 Map showing municipalities where official data was collected for DatBD and the proportion of bodies found reported in UNITED according to irregular border-crossing route.
famous shipwreck of a migrant boat in the Mediterranean. As a result, these deaths are found in both DatBD and UNITED.

The proportion of deceased bodies reported in the news could be related to peaks in the use of a particular route. Sharp increases in irregular arrivals by land or sea tend to draw public and political attention and attract journalists, in the same way that shipwrecks with hundreds of fatalities do. However, it is not possible to determine a relation between the proportion of bodies reported by the news and peaks in the use of an irregular migration route across the southern EU external borders because there is insufficient data available on arrivals per route over the same period as investigated in this study. Frontex provides interception data on various routes but only for the period 2008-2013 (or 2006-2013 in the case of the Atlantic route to the Canary Islands). This period excludes the peaks in use of Western Mediterranean and Adriatic routes, undermining any conclusions drawn. Data for earlier periods is available from some national authorities but this data is aggregated, so routes cannot be compared. Also, lack of standardisation undermines comparison on the basis of this data (Takle 2017).

Among the matched records, there are differences between the two datasets in details such as the number of bodies found per incident, time and place of death and personal information about the deceased (origin, sex, age). These differences not only made the matching process more challenging, they also demonstrate the unreliability of the details reported by the news. The differences in time of death were usually minor, especially for analysis of trends in aggregated numbers of border deaths (e.g. per year or per route). However, the differences in the number of bodies found per incident demonstrate further potential for under-counting deaths, as well as potential for over-counting deaths following a particular incident. For instance, among matched incidents (where at least one body was recorded in DatBD that matched an incident recorded in UNITED), UNITED recorded 13 more dead bodies found in Greece than were recorded in DatBD, 11 more in Italy and Malta and 36 more in Spain. Meanwhile, the differences in personal information – already limited to a minority of records in UNITED – demonstrate news-sourced data’s unreliability for comparative analysis of different groups of people based on sex, age or origin.

A substantial difference in Figure 4.2 between the shaded area (representing matched records) and the dotted line (representing total UNITED records) was expected because records could only be matched of deceased persons’ whose bodies were found in or brought within the jurisdiction of local authorities of southern EU member states (DatBD’s limitation). Due to the parameters of their primary sources, UNITED can capture disappearances and bodies that wash up on beaches in Morocco, Libya and Turkey, while DatBD cannot. The comparison enables the calculation of the proportion of bodies found in the EU among reported border deaths. The average proportion of UNITED records ‘captured’ in DatBD over the period 1993-2013 is 15.9% with a standard deviation of 11.5% (min 0.0%, max 52.3%, n=21); in other words, quite consistently low. This finding is relevant to arguments that Search and Rescue (SAR) policies (Heller and Pezzani 2016) or externalisation of border controls (Albahari 2006; Zagaria 2011; Williams and Mountz 2016) have contributed to keeping deaths out of sight of the European

---

108 If no body is found, there is no paper trail in the death management system (the source of data in DatBD).
public (Weber 2010; Grant 2011; Weber and Pickering 2011). Of course, one has to keep in mind that is not possible to test the reliability of reporting in UNITED of missing persons or bodies found in non-EU countries.

Interestingly, missing persons and bodies managed by non-EU authorities do not account for all records in UNITED that are missing in DatBD. The comparison revealed that UNITED records bodies found in EU territory which do not match any DatBD record. There are a few geographical gaps in DatBD relating to the few archives researchers did not gain access to, or to places that fell outside the data collection catchment area (e.g. stowaways discovered in vehicles inland or in major ports such as Patras in Greece, Genova in Italy and A Coruña in Spain). These gaps are laid out clearly in Chapter 2; the total number of deaths missing from the DatBD as a result of these gaps is estimated to be less than 100. Another explanation for the records in UNITED missing from DatBD is over-counting: matching records between the two datasets revealed double-counting in UNITED (represented in grey in Figure 4.5).

Figure 4.5 Double-counts in UNITED per year

Figure 4.5 shows the proportion of double counts discovered during the comparison of DatBD and UNITED. The columns show the number of deaths per year reported in UNITED; the grey sections represent double-counts. 4.9% of UNITED records (1993-2013) were found to be double counts. The potential for over-counting in news-sourced data was recognised by Last and Spijkerboer (2014). Records of incidents in UNITED are often vague, lacking details about the precise place and circumstances of the incident or who the victims were. If a shipwreck happens during the night – which many do – one journalist may report the shipwreck as occurring on one date and another journalist reports the shipwreck as occurring the following day. Given the difference in date provided in the two news reports, it is understandable that UNITED might record them as two separate incidents, especially if the two news stories focus
on different details (e.g. one mentions the number of women and children on board while the other shares the passengers’ nationalities, or one describes the events leading up to the shipwreck while the other reports on the rescue operation). At what point the journalist reports the news story can also lead to double counting in UNITED. For instance, one journalist reports the breaking news of a distress call while another journalist reports on the whole incident a few days later. Within those few days, many of the details in the initial news report may have been corrected, leading the two reports to be recorded in UNITED as separate incidents. Finally, and most predictably, UNITED double counts when it records both a new report of a shipwreck and a news report of bodies washing up on the beach days or weeks or months later. The 663 double counts shown in Figure 4.5 were revealed by the comparison because they were related to the same DatBD records. Therefore, the grey sections in Figure 4.5 represent the minimum over-count, as it is possible that there are more double counts in UNITED concerning incidents not recorded in DatBD (i.e. incidents in which no bodies were found and managed by local authorities in southern EU member states).

To sum up the findings of Study 2, the comparison of UNITED and DatBD revealed that each dataset has gaps (Table 4.3) determined by their particular methodological limitations and that the proportion of matched records between the datasets varies over time (Figures 4.2 and 4.3) and between routes (Figure 4.4). News-sourced data both under- and over-counts deaths (Table 4.3 and Figure 4.5). The comparison of the two datasets also enables a better understanding of the trends in each (Figure 4.2): there has been a general increase in the proportion of bodies found that are reported in the news (Figure 4.3), while the proportion of reported border deaths where a body is found in the EU has been quite consistently low.

Discussion

The findings of Study 1 reveal a dependence on news reports as the main source of data on EU border deaths. The availability of collections of news reports for advocacy purposes by NGOs and journalists has somehow dissuaded academic researchers from addressing the absence of official mortality statistics despite 30 years of EU border deaths. There is awareness in the literature of the problems associated with news-sourced data but this does not deter academics from drawing conclusions about trends in deaths and their relation with policy. Moreover, chain citations (citing another academic work rather than the dataset they used) lend a false sense of credibility and obscure the original source of data. The lack of reliable data is also implicated in the use of analyses from other border regions to support arguments concerning the EU external borders. It is reasonable to assume that this dependence on news-sourced death data extends to NGOs, journalists and government and EU offices.

The dependence on news reports as the primary source of border death data is especially worrying given the findings of the comparison of news-sourced data with data sourced from death management systems (Study 2). News-sourced data is not reliable for studying trends in absolute deaths or mortality because news-sourced data over-counts and under-counts deaths
Chapter 4

in unpredictable proportions over space and time.\textsuperscript{109} The few details that UNITED reports concerning age, sex and origin of deceased persons are also unreliable for comparative analysis between groups.

Border mortality is a valuable indicator of the suffering of irregularised border crossers and the negative impacts of border enforcement. If existing sources of data are unreliable for studying EU border deaths, what can be done to advance knowledge in this field? During the course of this research, available options for calculating a more reliable estimate of deaths were explored. For instance, Patrick Ball and his colleagues at the Human Rights Data Analysis Group (HRDAG)\textsuperscript{110} have developed capture-recapture techniques to provide reliable estimates from existing, problematic datasets in the context of human rights violations. In particular, they use Multiple Systems Estimation (MSE), which involves matching records between datasets, merging the datasets by excluding double records, and modelling an estimate of the number of records missing from the merged dataset based on the particular context and nature of each contributing dataset. Heller (2015: 204) has suggested that these techniques could also be employed to produce the reliable estimates of EU border deaths necessary for calculation of mortality trends. Unfortunately, MSE is not applicable to EU border deaths for two reasons:

1. The high proportion of unidentified among the dead (Chapter 3; Grant 2016; Kovras and Robins 2016) makes it impossible to achieve perfect matching of records necessary for employing capture-recapture techniques. Perfect matching is one of the four assumptions on which MSE depends, and the only one that cannot be compensated through modelling.

2. A second important assumption underpinning MSE is system independence: that the possibility of being ‘captured’ in one dataset is independent of the possibility of being ‘captured’ in another. As Figure 4.1 shows, there are only 2 primary sources of death data – news reports and local authorities – and they are not independent. The discovery of dead bodies increases the newsworthiness of shipwrecks or arrivals, making it more likely that such cases will be reported in the news. It requires 3 or more sources of data to adjust the model to overcome such system dependence.\textsuperscript{111}

It was initially hoped that linking records between UNITED and DatBD could inform a correction of the UNITED estimates per year and per route. However, the irregularity of under- and over-counting in news-sourced data exclude the possibility of using the DatBD and the analysis presented in this chapter to produce more accurate estimates of EU border deaths for the period 1993-2013. Moreover, due to the geographical limitations of DatBD, it fell outside

\textsuperscript{109} Conclusions about trends in deaths are based on whether there has been an increase or decrease between time A and time B. If UNITED records more deaths in B than in A, the conclusion would be that there has been an increase in deaths, that route has become more dangerous. However, because UNITED is based on news reports, there is an unknown probability of both under- and over-counting in any given period of time. Count A may be a net under-count, count B may be a net over-count, and thus there has in fact been a decrease in deaths over time.

\textsuperscript{110} https://hrdag.org/

\textsuperscript{111} Some might argue that IOM’s MMP could count as a third Mediterranean-wide source of data on deaths, but (1) their data collection only began towards the end of 2013, and (2) they also rely heavily on news reports. It might be possible to ‘zoom in’ on one particular area where a third source of data on deaths might be available (e.g. Coast Guards in the North Aegean have recorded the dead and missing from boat incidents they responded to for many years).
the scope of Study 2 to ascertain the reliability of reporting of missing persons and bodies found in non-EU countries recorded in UNITED. Therefore, the overall bias of news reports as a source of data for EU border deaths remains unknown.

Thus, the options for producing more accurate estimates of EU border deaths from existing sources have been exhausted. The search for more reliable sources has also been exhausted (Chapter 2). The reality that it is the only EU border death data available does not change the fact that news-sourced data is unreliable for the purposes of analysing trends in mortality over time. Figure 4.1 reveals a growing trend in academic literature on the subject in recent years to rely on IOM’s Missing Migrants Project (MMP) rather than UNITED or the Fortress Europe blog. But as Table 4.1 shows, IOM’s MMP is based on the same primary source of data – news reports – and therefore is inflicted by the same biases revealed in Study 2. Data generated by a project of an intergovernmental organisation and UN agency may appear official and, therefore, reliable, but IOM’s Missing Migrants Project is not proven to be more reliable than any other news-sourced dataset (Al Tamimi et al 2017). Thus, scientific claims and policy impact assessments should not depend on quantitative border death data.

Finally, for the quantitatively-inclined, a potential source of death data exists that has not been explored: survivors of border crossings and families of missing persons know who has been lost. However, survivors have precarious legal status in the EU or transit countries, are usually restricted to detention or reception facilities, and receive little – if any – mental health assistance to process the traumas they experienced during their journey. Survivors are interviewed for immigration processing and for smuggling intelligence, but not for the purpose of identifying the dead or naming the missing (Chapter 3; Zagaria 2011). Family members often face the same constraints on their international mobility that led their missing relatives to cross the southern EU external borders irregularly, and there is no independent platform or agency to whom they can turn to aide them in their search (Chapter 3; Kovras and Robins 2016). These issues must be dealt with if survivors and families are to be provided the opportunity, support and security necessary to exchange information (and it must be a genuine exchange, see Grant 2016) on their fellow travellers and missing relatives.

**Conclusions**

Nothing in this chapter should be construed as undermining the integrity of the List of Deaths maintained by UNITED for Intercultural Action or any of the datasets shown in Figure 4.1 and Table 4.1. News-sourced border death data has considerable value (1) in advocacy and raising awareness of border deaths, and (2) as an indication of fatal incidents among irregularised travellers. Where states have failed to document the lives lost crossing their borders, journalists, international organisations and NGOs have not.

Nonetheless, news-sourced death data is not reliable when used to calculate mortality rates or assess trends in deaths over time or space (e.g. for the purposes of determining whether a given policy has made irregular migration more or less dangerous in general or on a given route). Yet, academic research has presented this data as though it were reliable, giving the misguiding
impression that collecting more and better data is unnecessary and that findings of analyses of such data are conclusive. Repeated use and chain citations of news-sourced data lends these sources a false sense of credibility and reinforces assumptions as to the quantitative value of available data.

Given the absence of reliable data on deaths, trends in absolute numbers of EU border deaths or border mortality cannot be analysed quantitatively. Instead, debates surrounding irregularised travel and policy interventions should derive their empirical elements from qualitative data gathered by academic and civil society researchers, including interviews with survivors, families and rescuers, and ethnographic studies of border control. In short, as many others have argued (e.g. Tazzioli 2015), we need to stop relying on numbers to discuss the impact of current migration and border policy.