Evaluation of the Database of the Kosovo Memory Book

Jule Krüger and Patrick Ball

December 10, 2014

Human Rights Data Analysis Group
everybody counts.

Contents

Chapter 1

Executive summary

In March, 2012, the Humanitarian Law Centre (HLC) in Belgrade, Serbia, asked the Human Rights Data Analysis Group to evaluate their database of human losses during armed conflict in the territory of the Former Yugoslav Republic (FYR) between 1998 and 2000. In this report, we present our findings with regard to the completeness and quality of the Kosovo Memory Book database (KMBD). The database is a joint project between the HLC in Belgrade and the Humanitarian Law Centre Kosovo (HLC-Kosovo).

Our evaluation consisted of four visits to HLC-Belgrade, three by Jule Krüger (July 2012, March 2014, and November 2014) and one by Patrick Ball (November 2012).[1] The visits consisted of technical demonstrations of the database by KMB staff; discussions with the KMB team about their research and field processes; the methodology by which they classified deaths by type of death; and extensive reviews of how KMB staff determine whether multiple reports refer to the same victim or to different victims (called *matching*). In addition to the discussions with the KMB team in Belgrade, we conducted a descriptive analysis of the KMB database itself. A companion report to this one has been prepared by Professor Michael Spagat, and it comes to substantially the same conclusions as those we present here.

From our discussions with the KMB team and our analysis of the database, we conclude that the Kosovo Memory Book documents all or nearly all the human losses during conflicts in Kosovo during the period 1998–2000. By "nearly all," we mean that in our opinion, it is very unlikely that there are more than a few tens of undocumented deaths.

Our analysis is based on a comparison of the KMB database to ten other databases of victims in Kosovo during this period. We matched records from the other databases

[1] Further information was obtained from the KMB team in Belgrade via email, video calls, and their provision of supplementary information.

to the KMB to determine whether any of the deaths documented by earlier projects had been missed by the HLC and HLC-Kosovo research teams. In a three-stage process, HRDAG and the KMB team compared all ten databases to the KMB, and we determined that it is very unlikely that there are any additional killings in those databases not already documented in the KMB. This finding, combined with our finding that the KMB documents more deaths for every period and every region than any of the other databases,[2] leads us to conclude that the KMB register of war victims is very close to a complete enumeration.

The Kosovo Memory Book database covers human losses in the conflict during 1998–2000, and it documents 13,517 war victims. In addition, there are 1,603 potential victims whose war victim status has not been confirmed. Based on the documentation patterns, we believe most of these records will not be confirmed, and therefore they will be added to the list of 2,881 deaths classified as 'not war' victims (see Sections 4.1, 5.1, and 6.4).

The descriptive analysis in Chapter 4 shows that the great majority of war victims are civilians, most were killed (as opposed to disappeared), most were male, most were Albanian, and most were killed in 1999; in these examples, "great majority" and "most" mean more than 80%. One curious finding is that demographically, the group most at risk of being killed were older people, especially older men. This is partly because Kosovo had a relatively young population, so there are proportionally fewer older people, but it is nonetheless noteworthy; see Section 4.7 for more discussion.

The report begins with a discussion of HLC's and HLC-Kosovo's project, including the legal definitions that were used to define "victims" of the conflict in Chapter 2. Chapter 3 briefly describes the technical aspects of the database design. Chapter 4 opens a substantive analysis of the database by providing statistical descriptions of the war victims, and Chapter 5 examines the source materials that HLC used to register and document all victims. Chapter 6 is perhaps the most important chapter: it considers how the KMB can be understood as a complete enumeration of all the victims of the conflict in Kosovo from 1998–2000. Chapter 7 reviews our conclusions and recommendations.

[2]With two puzzling exceptions: see Section 6.2.1 for a discussion of weekly counts of victims documented by the Democratic League of Kosovo (LDK) that may be due to date encoding errors in this particular data source. See Section 6.2.2 for a discussion of possibly inaccurate victims records reported for the municipality of Kaçanik/Kačanik by the OSCE.

Chapter 2

Project description: The Kosovo Memory Book

The Kosovo Memory Book database (KMBD) of the HLC and HLC-Kosovo documents human losses in connection with the war in Kosovo between January 1, 1998, and December 31, 2000.[1] The KMB's focus on human losses is geared towards lethal human rights violations in connection with the armed conflict.[2] The objective is to document all individuals who were killed or disappeared in connection with the war in Kosovo.

The HLC and HLC-Kosovo pursue various goals in establishing a register of war victims, beginning with a clarification about what happened in this past episode of armed conflict. The most important goal is to restore the dignity of those affected by naming the victims and providing them a monument. Further objectives are to support processes of transitional justice and criminal prosecution, as well as to assist other initiatives of

[1]The definition of "armed conflict" (i.e., war) is based on a decision of the International Criminal Tribunal for the Former Yugoslavia (ICTY) Appeals Chamber in the *Tadic* case, Paragraph 70 (October 1995). "Armed conflict" is defined as a period that involves the use of armed force or the occurence of protracted armed violence between armed groups across states or within a state. Hereby, armed groups are understood as either governmental authorities or non-governmentally organized armed groups. The temporal focus was established in orientation to the official conflict dates adopted by the ICTY. The ICTY's mandate period ranges from February 28, 1998, to mid-June 1999. Because killings and abductions of ethnic Serbs and Roma, as well as so-called 'collaborators' and proponents of the Kosovo Liberation Army continued after the official conflict end date, the HLC and HLC-Kosovo extended their temporal focus to the end of 2000 when post-conflict violence finally subsided. The territorial choice covers murders and disappearances in connection with the war in Kosovo in the territories of Kosovo, Montenegro, and Serbia, as well as on Kosovo's border areas with Albania and Macedonia. For details on definitions, as well as the temporal and spatial focus, please refer to the accompanying methodology document by Nataša Kandić of the HLC.

[2]With the current focus on lethal violations, the HLC and HLC-Kosovo are evaluating approximately 20% of their physical archive. It is the HLC's and HLC-Kosovo's explicit goal to expand the evaluation of their archive to non-lethal violence in the future.

post-conflict resolution.

These goals are being achieved by publishing the KMB in four volumes of victim narratives, by providing a digital list of victims online,[3] and in advocating with regional partners for the establishment of a regional truth commission that is to investigate war crimes.

HLC started to collect data on human rights violations during field missions in Kosovo in 1998 and maintained a field office in Pristina until September 2010.[4] In addition to field research trips, information on human rights violations was obtained from family members and witness statements, local and foreign newspapers, as well as other media sources, reports of local and international governmental and non-governmental organizations, community leaders, and many other types of material.[5]

The idea that a Kosovo Memory Book database could be created by systematizing comprehensive information on human losses emerged in 2006. A digital database of human losses was started in 2005. The HLC and HLC-Kosovo continue with independent field research and the collection of witness statements until today. In January 2012, the HLC and HLC-Kosovo expanded data collection to obtain other types of documents, to verify personal facts of victims, to compare the war victim register to those of other organizations, and to assess their completeness. The goal is to create an enumeration of all human losses in connection with the war in Kosovo between 1998 and 2000.

[3]The version of the Kosovo Memory Book database as of November 7, 2014, which was analyzed for the purposes of this evaluation, can be found at the following websites: http://www.kosovomemorybook.org/ (English), http://www.liberkujtimiikosoves.org/ (Albanian), and http://www.kosovskaknjigapamcenja.org/ (Serbian).

[4]Since HLC-Kosovo became an independent partner organization, the KMB database is now being shared between both groups.

[5]In evaluating information, a very wide definition of what material may constitute a 'source' is adopted. Please see the accompanying methodology document by Nataša Kandić (HLC) for details on the KMB team's data collection strategy and terminology.

Chapter 3

Database structure and quality

The database is written in PHP for an html interface on top of a MySQL (5.1) database.[1] The database structure is extremely complicated, with over 150 tables organized in the following groups: shared codebooks, users, employment, courses, tags, documents, persons, events, trials, armed formations, financial, and analysis log.

One of the most valuable aspects of the database is the separation into "source" and "judgment" layers. Among the most important parts of a database of this kind is managing multiple streams of potentially contradictory information. The HLC's and HLC-Kosovo's database preserves all the original source information, and offers a second layer in which their judgments can be recorded. This approach means that rates of agreement and disagreement in the original sources can be calculated, and information is never lost, even when contraditory sources are merged.

The database is designed so that nearly every string field can be recorded in English, Croatian, Serbian, and Albanian. This provides tremendous detail for reports, and is an astonishing level of work that should be acknowledged.

In the KMB database, one person could be linked to many events by playing different roles in each event (as witnesses, victims, or perpetrators); similarly, each event has the possibility to include many people. The event structure is an extremely useful grouping structure which in theory enables the database to describe a kind of historical narrative, colligating the individual victims' experience into events. This will enable a wide range of possible analytic approaches, both quantitative and historical.

[1] The database software is being developed and maintained by Abacus in Zagreb, Croatia.

Chapter 4

Descriptive analysis of database fields

In this chapter, we review the various fields in the Kosovo Memory Book database that provide information on individuals who were killed or disappeared in connection with the war in Kosovo in the period of 1998–2000. The goal is to evaluate the quality of the information presented.

4.1 Victim categories and types of violations

The database presented for evaluation contained a total of 18,001 individuals the HLC and HLC-Kosovo reviewed in preparation of the Kosovo Memory Book for the three-year period 1998–2000. These individuals were classified into different categories of victims based on whether they can be confirmed as victims of the conflict, and what they endured during the war.

Individuals in the database were divided into three categories: 'war victims,' 'potential victims,' and 'not war victims.' A 'war victim' is a person who lost her/his life or went missing during or in connection with the armed conflict, irrespective of whether she/he was a civilian or a member of armed forces under the control of one of the parties to the conflict at the time of her/his death or disappearance.

'Potential victims' are individuals whom the KMB team has yet been unable to identify as war victims. This is due to missing information on the circumstances of the relevant deaths or disappearances that would allow for confirming a connection with the war in Kosovo. Potential victims are currently still subject to field research for purposes

of determining whether they are war victims or not. 'Not war victims,' in turn, are individuals for whom it was established that a death/disappearance in connection with the war in Kosovo did not occur.[1]

The KMB differentiates among the following violation categories for war victims: 'killings,' 'disappearances,' and 'death caused by war.' 'Killings' designate the deaths of individuals that occurred in connection with the armed conflict. A 'disappearance' is defined as the deprivation of freedom, detention, abduction or any other form of captivity during hostilities followed by denial or refusal of the responsible actors to reveal the fate or location of the missing person. Based on this definition, individuals are registered as 'disappeared' when they or their bodies have not been found to date. A 'death caused by war' denotes the death of civilians due to wartime food shortages, harsh conditions while on the run or in hiding, helplessness in the wake of forced separation from family members, lack of medicine or medical assistance, stress related to deportation, etc.[2]

For potential victims, the type of violation remains 'unknown' until sufficient information on the case is available that helps to determine whether the case is to be added to the war victims or the 'not war' victims register.

Among victims whose deaths are not attributed to the war ('not war victims'), the following categories were used to denote why an individual was not considered a victim in relation to armed conflict in Kosovo: 'found alive,' 'natural death,' 'death by accident,' and 'unknown.' For example, some officially missing persons were later on located in Serbian prisons. For other victims, family members stated that they died of natural causes, for instance, because of age, or as a result of ordinal criminal offense, or in a traffic accident that was unrelated to the war. 'Unknown' victims, in turn, are individuals for whom HLC/HLC-Kosovo were unable to establish their identities within the local communities, i.e., individuals with these names don't seem to exist. Furthermore, some missing persons requests were canceled at relevant institutions for unknown reasons.

Table 4.1 gives a summary of victims in terms of their classification into victim categories and violations. As can be seen, there is no missing information in the victim category and violation fields. Killings constitute the largest share of violations observed (11,532 total killings).[3]

[1] Exact definitions of terms used in the KMB are provided in the methodology document by Nataša Kandić (HLC) that accompanies this report.

[2] See previous footnote.

[3] The database provided for evaluation contained an additional 377 'not war' victims. These individuals were, however, excluded from Table 4.1 because the date of their death occurred outside of the 1998–2000 time period.

Table 4.1: Reported victims, by violation and victim category.

Violation/Category	War	(%)	Potential	(%)	Not war	(%)	Total	(%)
Killing	11532	85.3	0	0.0	0	0.0	11532	64.1
Unknown	0	0.0	1603	100.0	1194	41.4	2797	15.5
Disappearance	1704	12.6	0	0.0	0	0.0	1704	9.5
Found alive	0	0.0	0	0.0	1308	45.4	1308	7.3
Death caused by war	281	2.1	0	0.0	0	0.0	281	1.6
Death by accident	0	0.0	0	0.0	215	7.5	215	1.2
Natural death	0	0.0	0	0.0	164	5.7	164	0.9
Total	13517	100.0	1603	100.0	2881	100.0	18001	100.0

4.2 Victim code

Each victim has a unique identifier number assigned in the KMB database which helps to differentiate between victim dossiers.

4.3 Victim names

Individual victims are identified by their name information which consists of a last name, a father's name, and a first name.

In general, when a piece of name information was unknown HLC staff left the relevant name field blank. However, after a closer inspection of name values, we found that in some cases the codes 'NN', 'N', or 'N.' – likely for Serbian 'nepoznat,' i.e., unknown, – may have also been used to denote missing name information. Furthermore, we noticed that in some cases name field information only consisted of a single or two characters, or one character followed by a punctuation mark.[4]

Because name field information of only one or two characters is not meaningful to sufficiently identify individuals, we count such field values as missing. Here especially the use of 'N' or 'N.' is misleading as to whether a name value is missing or a name started with the letter 'N'. In Table 4.2, the resulting amount of name missingness, after name values of fewer than three characters were accounted for, is summarized by victim categories and name fields. Please note that the name missingness categories in Table 4.2 are mutually exclusive. For example, records with first and father's name missing are not also counted among records with first, father's, and last name missing.

An anonymous victim in this study is defined as an individual who does not have

[4]Serbian or Albanian names of only three-character length exist.

Table 4.2: Name missingness, by name field and victim category.

Field(s) NA/Category	War	(%)	Potential	(%)	Not war	(%)	Total	(%)
Father's name	169	0.9	1095	6.1	1440	8.0	2704	15.0
First name	3	0.0	33	0.2	40	0.2	76	0.4
First, father's name	0	0.0	18	0.1	31	0.2	49	0.3
Last name	0	0.0	22	0.1	14	0.1	36	0.2
Father's, last name	0	0.0	15	0.1	12	0.1	27	0.1
First, father's, last name	0	0.0	5	0.0	1	0.0	6	0.0
First, last name	0	0.0	2	0.0	1	0.0	3	0.0

information on the first and last name combined, i.e., the last two rows in the above table. As can be seen in Table 4.2, there are no anonymous records among war victims, as the number of cases with missing information on last and first name combined is 0.[5] Among potential victims, 7 are anonymous. Among 'not war' victims, 2 are anonymous.

The fact that no anonymous victims are contained in the war victims register attests to the quality and intended accuracy of this database. The fact that there are anonymous victims within the potential and 'not war' victims registers reflects the uncertainty associated with these records. The collection of name information of conflict victims is in contrast to data projects that provide anonymous victim counts to establish the number of individuals affected by some violence. The name information provided by HLC allows for cross-validation of whether stated human losses can be traced back to real people.

In the remainder of the descriptive analysis in Chapter 4, we are only concerned with the total of 13,517 war victims that HLC/HLC-Kosovo established to be victims directly related to armed conflict in Kosovo, or deaths caused by the war.

4.4 Victim status and membership in armed formations

In the Kosovo Memory Book, war victims are classified according to individual status. The possible status types are 'civilian,' 'soldier,' and 'police.' A civilian is anyone who is not a member of an armed force or a member of an organized armed group that is party to the conflict. Status information is summarized in Table 4.3. There is no missing information on war victims' individual status at the time of death/disappearance.

[5]Two of the 3 war victims with missing first name were babies who died before their parents gave them a name. The third war victim is a KLA soldier who died in battle with Serbian forces. In the future, the HLC and HLC-Kosovo may move this record into the potential victims register given the missing first name information.

Table 4.3: Reported war victims, by status.

Victim status	Victims	Percent of victims
Civilian	10305	76.2
Soldier	2848	21.1
Police	364	2.7
Total	13517	100

War victims of 'police' or 'soldier' status can be further differentiated with regard to their membership in the various armed formations involved in the war over Kosovo. This information is summarized in Table 4.4.

Table 4.4: Reported war victims, by membership in armed formation and status.

Formation/Status	Civilian	%	Soldier	(%)	Police	(%)	Total	(%)
None	10305	100.0	0	0.0	0	0.0	10305	76.2
KLA	0	0.0	2123	74.5	0	0.0	2123	15.7
VJ	0	0.0	719	25.2	0	0.0	719	5.3
MUP	0	0.0	0	0.0	364	100.0	364	2.7
FARK	0	0.0	3	0.1	0	0.0	3	0.0
KFOR	0	0.0	3	0.1	0	0.0	3	0.0
Total	10305	100.0	2848	100.0	364	100.0	13517	100.0

As can be seen in Table 4.4, civilians are not associated with any armed formation. The largest number of war victims was associated with the Kosovo Liberation Army (KLA), followed by soldiers of the Yugoslav Army (VJ), policemen of the Serbian Ministry of the Interior (MUP), soldiers of the Armed Forces of the Republic of Kosovo (FARK), and soldiers of the Kosovo Forces led by NATO (KFOR).

In the remainder of the descriptive analysis, victims with 'police' or 'soldier' status were summarized into one category of 'members of armed formations' (MOAF). This led to a total of 3,212 members of armed formations (23.8%).

In Table 4.5, a summary is given of all war victims by their status as well as the type of violation they endured. As can be seen, the majority of victims are civilian victims who were killed. The number of disappearances among members of armed formations is small (136 individuals, or 1% of all war victims) compared to 1,568 civilian victims (11.6% of all war victims) who disappeared.

Table 4.5: Reported war victims, by violation and status.

Violation/Status	Civilian	(%)	MOAF	(%)	Total	(%)
Killing	8526	82.7	3006	93.6	11532	85.3
Disappearance	1568	15.2	136	4.2	1704	12.6
Death caused by war	211	2.0	70	2.2	281	2.1
Total	10305	100.0	3212	100.0	13517	100.0

4.5 Victim sex

The database reports the sex of each war victim. There is no missing information on victim sex. Table 4.6 summarizes all war victims by their sex and status information. As can be seen, males constitute the majority of victims. Among civilian victims, 82.2% are male. The victims who were members of armed formations are almost entirely male (99.3%).

Table 4.6: Reported war victims, by sex and status.

Status/Sex	Civilian	(%)	MOAF	(%)	Total	(%)
Male	8472	82.2	3188	99.3	11660	86.3
Female	1833	17.8	24	0.7	1857	13.7
Total	10305	100.0	3212	100.0	13517	100.0

4.6 Victim ethnicity

The KMB database reports the ethnicity of war victims. As can be seen in Table 4.7, the majority of war victims are Albanian (regardless of victim status), followed by Serbs. There is no missing information on war victims' ethnicity (0 cases).

We note that the 'ethnicity' field conflates two concepts, citizenship and ethnicity. For individuals within Serbia, and therefore including the territory of Kosovo, this field is being used to designate membership in different ethnic groups as this was the main dynamic driving the conflict. For example, individuals resident within Kosovo are of Serbian citizenship during the time of the war, and in this field being characterized with regard to their ethnic group membership, i.e., Serb, Albanian, Roma, etc. For war victims resident outside of the national territory of Serbia, however, this field seems to denote their citizenship in a specific state. However, an individual of German citizenship could still belong to an ethnic group such as Albanian or Serb.

To separate these ideas, we recommend that two fields be established in the future to

Table 4.7: Reported war victims, by ethnicity and status.

Ethnicity/Status	Civilian	(%)	MOAF	(%)	Total	(%)
Albanian	8661	84.0	2131	66.3	10792	79.8
Serb	1197	11.6	1000	31.1	2197	16.3
Roma	151	1.5	9	0.3	160	1.2
Bosnian	86	0.8	9	0.3	95	0.7
Hashkali	78	0.8	0	0.0	78	0.6
Montenegrin	54	0.5	21	0.7	75	0.6
Egyptian	37	0.4	1	0.0	38	0.3
Gorani	11	0.1	3	0.1	14	0.1
Hungarian	0	0.0	14	0.4	14	0.1
Turk	9	0.1	1	0.0	10	0.1
Croatian	3	0.0	3	0.1	6	0.0
Macedonian	5	0.0	0	0.0	5	0.0
Bulgarian	3	0.0	2	0.1	5	0.0
Slovenian	3	0.0	2	0.1	5	0.0
Yugoslavian	1	0.0	3	0.1	4	0.0
Chinese	3	0.0	0	0.0	3	0.0
Russian	0	0.0	3	0.1	3	0.0
German	2	0.0	0	0.0	2	0.0
English	0	0.0	2	0.1	2	0.0
Italian	0	0.0	2	0.1	2	0.0
Czech	1	0.0	0	0.0	1	0.0
Algerian	0	0.0	1	0.0	1	0.0
French	0	0.0	1	0.0	1	0.0
Romanian	0	0.0	1	0.0	1	0.0
Rusyn	0	0.0	1	0.0	1	0.0
Slovakian	0	0.0	1	0.0	1	0.0
Ukrainian	0	0.0	1	0.0	1	0.0
Total	10305	100.0	3212	100.0	13517	100.0

clearly differentiate between these two concepts of ethnicity and citizenship. When ethnic group membership is unknown for citizens of countries other than Serbia, the ethnicity field can be left blank.

4.7 Victim date of birth

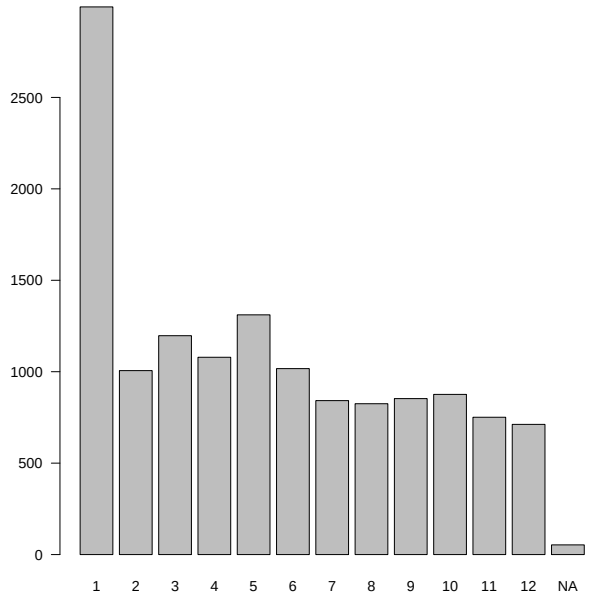
The KMB database reports the date of birth of war victims in the format 'month/day/4-digit year.' Figure 4.1 reports the distribution of the birth date values with regard to month, day, and year information, respectively.

As can be seen in Figure 4.1(a), the distribution of reported birth month information shows a comparatively higher frequency of the month of 'January' compared to any other month. For 3,048 war victims the birth month is 'January,' while for 53 war victims' birth month information is missing. Similarly, in Figure 4.1(b) it can be seen that for 2,770 war victims, the birth day is reportedly the first day of a given month resulting in a comparatively higher share of war victims with a birth date on the first day of a given month. For 53 war victims' birth day information is missing. The spread of birth year values ranges between 1899 and 1999 (see Figure 4.1(c)), while for 53 war victims birth year information is missing.

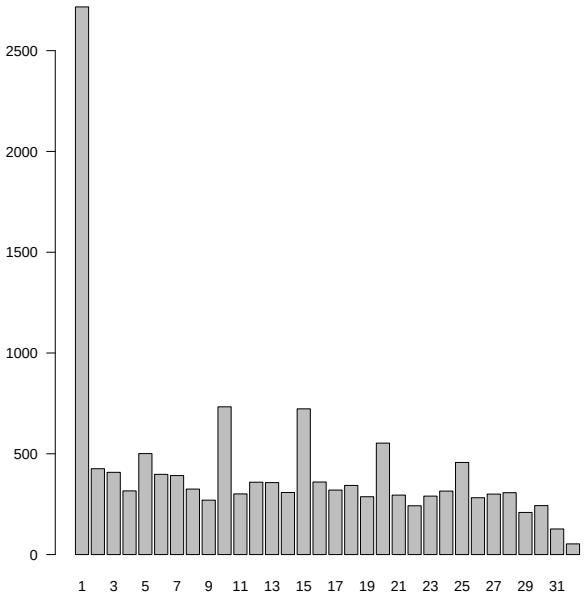
The observation of a high count of 'January,' as well as first day of the month birthday in the date of birth field is due to the KMB coding practice. Whenever date of birth information was entirely unknown (53 individual war victims), the DOB field was left blank. However, when the birth year was known but month and day information was unclear, the known year was entered and both month and day information were assigned '1.' Or, when the birth year and month were known, those were entered and the missing day of birth was assigned to the first day of the given month.

The missing date coding practice results in a comparatively high share of war victims with a birth date of 'January 1' across birth years (2,115 war victims) or a day of birth on the first of a given month. Furthermore, this rule hides victims with a true birth date of 'January 1' or on the first day of a given month who can no longer be distinguished from war victims for whom only the birth year or birth year and birth month are known. We recommend that the HLC and HLC-Kosovo update victim date of birth information by introducing a special 'Unknown' code (e.g., 99 or NA) for unknown birth day and/or month information.

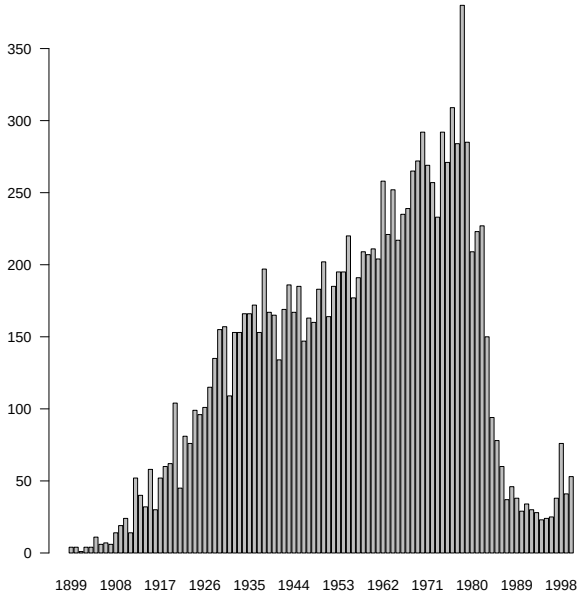
Victim date of birth information can be used to study the age distribution of war victims in the KMB database. We obtain age information at the time of a violation (i.e.,



(a) Month counts

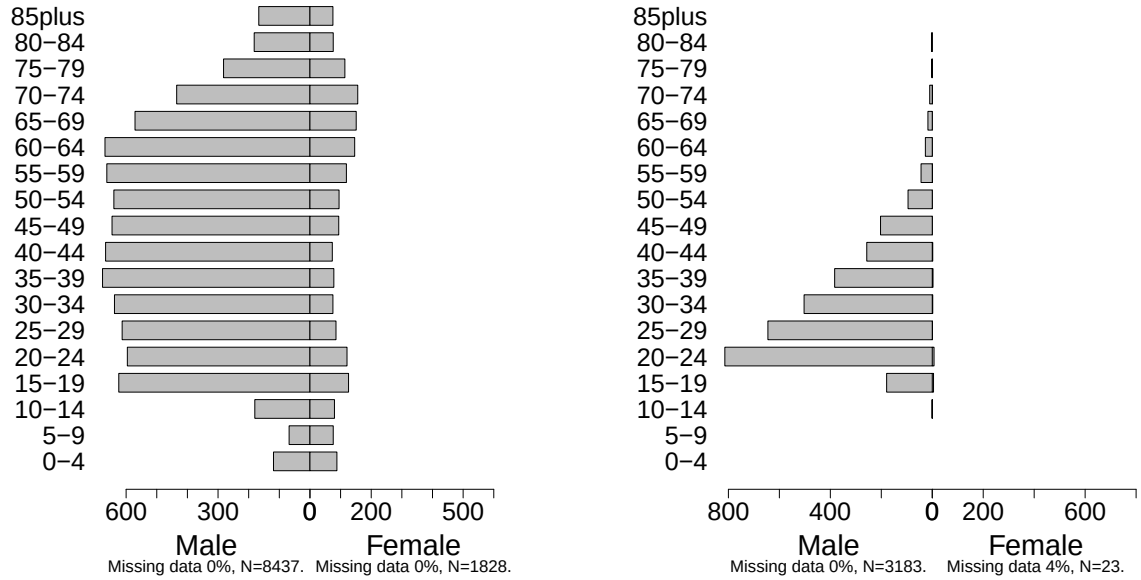


(b) Day counts



(c) Year counts

Figure 4.1: Distribution of war victims' date of birth values.



(a) Civilians

(b) Members of armed formations

Figure 4.2: Reported age-sex distribution of war victims, by victim status.

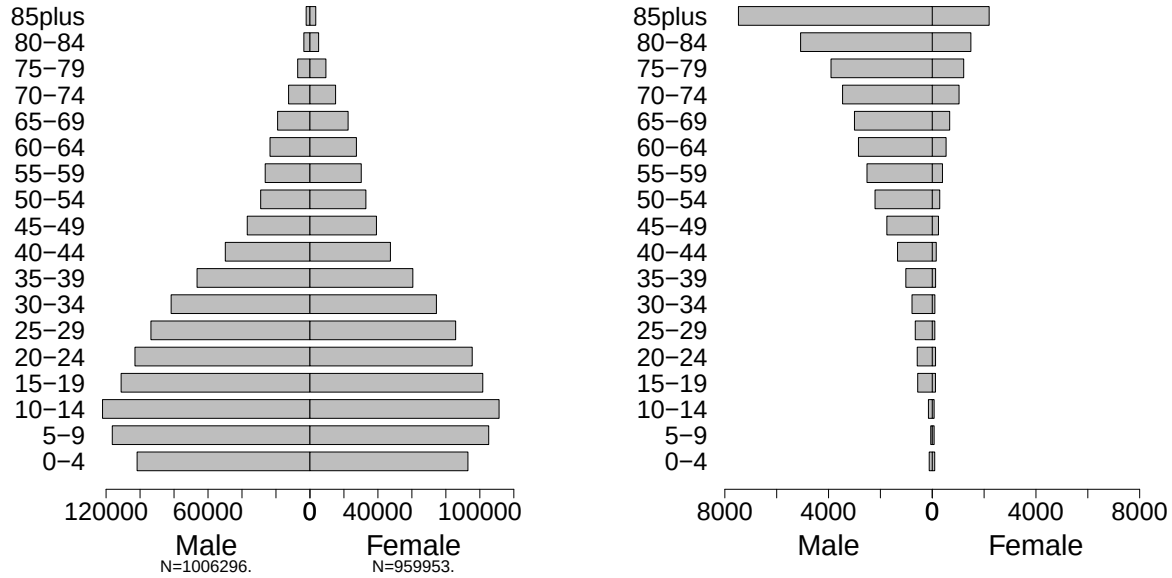
killing, disappearance, or death caused by war) by subtracting the birth year from the year of the violation. Figure 4.2 reports the frequency distribution of war victims by status, age, and sex, respectively.

As seen previously (see Table 4.6), war victims are mostly male, regardless of victim status. 72.9% of reported male civilian victims are aged between 15 and 74 years, and 87.4% victims who were members of armed formations are reported to be 20 to 49 years old.

The reported age-sex information of civilian victims that the KMB database provides can be adjusted by the age-sex proportions of the Kosovo population. A comparison to the population information provides us with a measure of the rate of civilian victimization according to victims' age and sex characteristics. A population adjustment is a ratio that compares the proportion of reported victims in a given age-sex category to the proportion of that particular age-sex category in the reference population. We further multiply the population-adjustment ratio by 100,000. This provides us with the rate of victims in a given age-sex group per 100,000 people.

To conduct the population adjustment, we rely on the estimated age-sex distribution of the entire population living in Kosovo as projected for the year 1998.[6] We use the 1998

[6]This data was obtained from the International Database of the U.S. Census Bureau. It is the result of a projection of the Kosovo population from a 1981 through a 2008 pilot census. This census data is the best available to date.



(a) Projected age-sex distribution of Kosovo population, in 1998.

(b) Number of victims per 100,000, by sex and age.

Figure 4.3: Population adjusted rate of civilian war victims, by sex and age.

projection as this is the last year prior to an intense eruption of violence in the region. The results of the population-adjustment analysis are shown in Figure 4.3.

In Figure 4.3(a), the 1998 projection of the Kosovo population is presented in terms of the expected age-sex distribution. The vast majority of the population is projected to be aged 39 or younger, while the age distribution across both sexes appears symmetric. A population that is by the majority younger than 40 years as in the given case is said to produce an inverse v-shaped age-sex pyramid.

Figure 4.3(b) reports the comparison of the age-sex distribution of reported civilian victims to the proportion of these age-sex groups in the full Kosovo reference population, per 100,000 Kosovo inhabitants. As can be seen, the number of victims steadily increases with an individual's age for both sexes, but overall is much higher for males. The per 100,000 victimization rate for male individuals aged 80 years or older by far outnumbers that of victims in other age-sex groups in the Kosovo population.

4.8 Victim place of birth

The KMB database reports the administrative district in which a war victim was born. For victims born inside the territory of Kosovo, this administrative district corresponds

to a Kosovo municipality. Municipality information (i.e., place of birth, residence, and violation in this section and Sections 4.9 and 4.11 below) is entered in both Albanian and Serbian languages. For victims born outside of Kosovo, the place of birth corresponds to the country in which a victim was born. Country names are denoted in English.

Table 4.8 summarizes information on where war victims were born. For 243 war victims (1.8%), information on the exact place of birth is missing. Furthermore for 18 war victims (0.1%), it is only established that they were born in 'Kosovë/Kosovo', while information on the exact municipality is missing.

Table 4.8: Reported war victims, by birth municipality/country and status.

Birth place/Status	Civilian	(%)	MOAF	(%)	Total	(%)
Skënderaj/Srbica	1075	10.4	170	5.3	1245	9.2
Gjakovë/Djakovica	966	9.4	216	6.7	1182	8.7
Gllogoc/Glogovac	800	7.8	186	5.8	986	7.3
Rahovec/Orahovac	814	7.9	113	3.5	927	6.9
Serbia	263	2.6	658	20.5	921	6.8
Pejë/Peć	599	5.8	197	6.1	796	5.9
Prizren	517	5.0	157	4.9	674	5.0
Podujevë/Podujevo	542	5.3	107	3.3	649	4.8
Suharekë/Suva Reka	504	4.9	104	3.2	608	4.5
Vushtrri/Vučitrn	459	4.5	72	2.2	531	3.9
Deçan/Dečani	306	3.0	223	6.9	529	3.9
Klinë/Klina	396	3.8	104	3.2	500	3.7
Mitrovicë/Kosovska Mitrovica	395	3.8	95	3.0	490	3.6
Prishtinë/Priština	381	3.7	94	2.9	475	3.5
Malishevë/Malishevë	354	3.4	70	2.2	424	3.1
Istog/Istok	211	2.0	77	2.4	288	2.1
Kaçanik/Kačanik	152	1.5	93	2.9	245	1.8
Unknown	213	2.1	30	0.9	243	1.8
Lipjan/Lipljan	201	2.0	32	1.0	233	1.7
Gjilan/Gnjilane	185	1.8	48	1.5	233	1.7
Ferizaj/Uroševac	143	1.4	47	1.5	190	1.4
Shtime/Štimlje	115	1.1	41	1.3	156	1.2
Obiliq/Obilić	114	1.1	15	0.5	129	1.0
Montenegro	76	0.7	42	1.3	118	0.9
Kamenicë/Kosovska Kamenica	79	0.8	29	0.9	108	0.8

Continued on next page.

Table 4.8: Reported war victims, by birth municipality/country and status.

Viti/Vitina	66	0.6	36	1.1	102	0.8
Fushë Kosovë/Kosovo Polje	78	0.8	7	0.2	85	0.6
Croatia	57	0.6	23	0.7	80	0.6
Bosnia-Herzegovina	47	0.5	26	0.8	73	0.5
Shtërpcë/Štrpce	35	0.3	8	0.2	43	0.3
Albania	13	0.1	26	0.8	39	0.3
Dragaš/Dragash	27	0.3	10	0.3	37	0.3
Leposaviq/Leposavić	24	0.2	10	0.3	34	0.3
Macedonia	23	0.2	11	0.3	34	0.3
Zubin Potok/Zubin Potok	22	0.2	6	0.2	28	0.2
Zveçan/Zvečan	18	0.2	3	0.1	21	0.2
Kosovë/Kosovo	15	0.1	3	0.1	18	0.1
Novobërdë/Novo Brdo	11	0.1	2	0.1	13	0.1
Germany	3	0.0	4	0.1	7	0.1
France	0	0.0	4	0.1	4	0.0
Slovenia	2	0.0	1	0.0	3	0.0
Austria	1	0.0	2	0.1	3	0.0
Russia	0	0.0	3	0.1	3	0.0
USA	0	0.0	3	0.1	3	0.0
Australia	1	0.0	1	0.0	2	0.0
Italy	1	0.0	1	0.0	2	0.0
China	1	0.0	0	0.0	1	0.0
Algeria	0	0.0	1	0.0	1	0.0
Switzerland	0	0.0	1	0.0	1	0.0
Total	10305	100.0	3212	100.0	13517	100.0

4.9 Victim place of residence

The database of the Kosovo Memory Book provides information on the last location where a victim was living until their death or disappearance, i.e., a war victim's place of residence. For residential locations inside Kosovo, the municipality is denoted. For locations outside of Kosovo, the country of residence is listed.

In Table 4.9, information on a war victim's place of residence (municipality/country) is summarized by victim status. For 70 war victims (0.5%), information on the exact place

of residence is missing. Furthermore for 203 war victims (1.5%), it is only established that they were residing in 'Kosovë/Kosovo', while information on the exact municipality is missing.

Table 4.9: Reported war victims, by municipality/country of residence and status.

Home residence/Status	Civilian	(%)	MOAF	(%)	Total	(%)
Gjakovë/Djakovica	1004	9.7	210	6.5	1214	9.0
Skënderaj/Srbica	1006	9.8	153	4.8	1159	8.6
Gllogoc/Glogovac	790	7.7	175	5.4	965	7.1
Serbia	221	2.1	729	22.7	950	7.0
Rahovec/Orahovac	820	8.0	105	3.3	925	6.8
Pejë/Peć	671	6.5	192	6.0	863	6.4
Prizren	535	5.2	163	5.1	698	5.2
Suharekë/Suva Reka	520	5.0	103	3.2	623	4.6
Podujevë/Podujevo	490	4.8	92	2.9	582	4.3
Mitrovicë/Kosovska Mitrovica	475	4.6	89	2.8	564	4.2
Prishtinë/Priština	455	4.4	102	3.2	557	4.1
Vushtrri/Vučitrn	466	4.5	74	2.3	540	4.0
Klinë/Klina	434	4.2	100	3.1	534	4.0
Deçan/Dečani	306	3.0	201	6.3	507	3.8
Mališevo/Malishevë	326	3.2	64	2.0	390	2.9
Istog/Istok	218	2.1	71	2.2	289	2.1
Kaçanik/Kačanik	155	1.5	90	2.8	245	1.8
Gjilan/Gnjilane	196	1.9	41	1.3	237	1.8
Lipjan/Lipljan	201	2.0	29	0.9	230	1.7
Ferizaj/Uroševac	158	1.5	56	1.7	214	1.6
Kosovë/Kosovo	95	0.9	108	3.4	203	1.5
Fushë Kosovë/Kosovo Polje	185	1.8	15	0.5	200	1.5
Shtime/Štimlje	122	1.2	44	1.4	166	1.2
Obiliq/Obilić	111	1.1	15	0.5	126	0.9
Viti/Vitina	75	0.7	31	1.0	106	0.8
Kamenicë/Kosovska Kamenica	50	0.5	23	0.7	73	0.5
Unknown	59	0.6	11	0.3	70	0.5
Montenegro	20	0.2	23	0.7	43	0.3
Shtërpcë/Štrpce	31	0.3	8	0.2	39	0.3
Leposaviq/Leposavić	12	0.1	16	0.5	28	0.2

Continued on next page.

Table 4.9: Reported war victims, by municipality/country of residence and status.

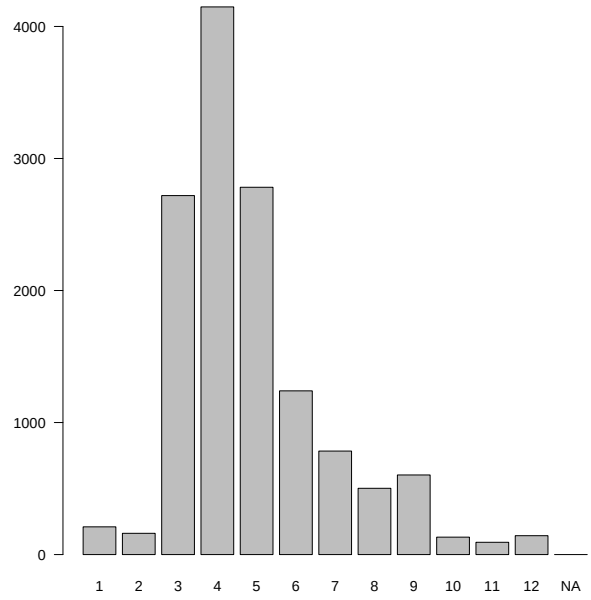
Albania	3	0.0	25	0.8	28	0.2
Zubin Potok/Zubin Potok	21	0.2	6	0.2	27	0.2
Croatia	23	0.2	2	0.1	25	0.2
Dragaš/Dragash	13	0.1	8	0.2	21	0.2
Novobërdë/Novo Brdo	14	0.1	3	0.1	17	0.1
Zveçan/Zvečan	10	0.1	5	0.2	15	0.1
Macedonia	6	0.1	7	0.2	13	0.1
Germany	3	0.0	7	0.2	10	0.1
Bosnia-Herzegovina	3	0.0	2	0.1	5	0.0
USA	0	0.0	4	0.1	4	0.0
Russia	1	0.0	2	0.1	3	0.0
Switzerland	0	0.0	2	0.1	2	0.0
China	1	0.0	0	0.0	1	0.0
Algeria	0	0.0	1	0.0	1	0.0
France	0	0.0	1	0.0	1	0.0
Great Britain	0	0.0	1	0.0	1	0.0
Italy	0	0.0	1	0.0	1	0.0
Sweden	0	0.0	1	0.0	1	0.0
Ukraine	0	0.0	1	0.0	1	0.0
Total	10305	100.0	3212	100.0	13517	100.0

4.10 Date of violation

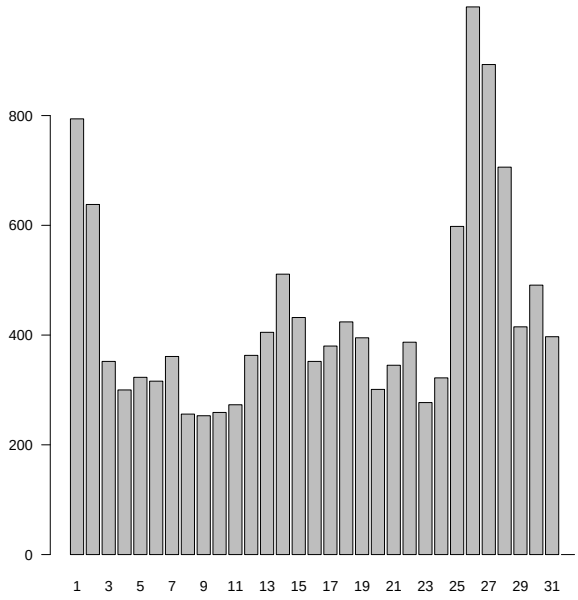
The KMB database reports the date a war victim experienced a violation during the three-year observation period of 1998 to 2000. In the case of 'killings,' this date denotes when the violation took place. In the case of 'disappearances,' this date denotes when the disappeared person was last seen or in contact with.

Date of the violation information is reported in the format 'month/day/4-digit year.' This date was converted to the format 'year-month-day' for the purpose of this analysis. The earliest date is '1998-01-01,' while the latest date in the database is '2000-12-16.' Figure 4.4 reports the distribution of violation date values with regard to month, day, and year information, respectively.

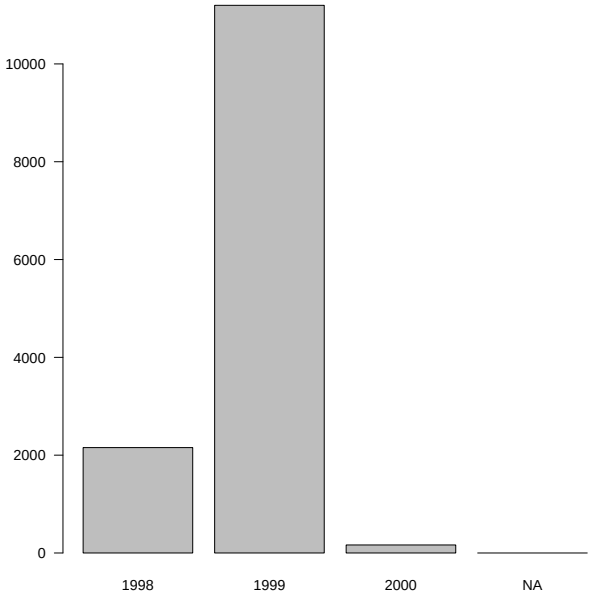
Similar to date of birth information (cf. Section 4.7), KMB staff coded missing viola-



(a) Month counts

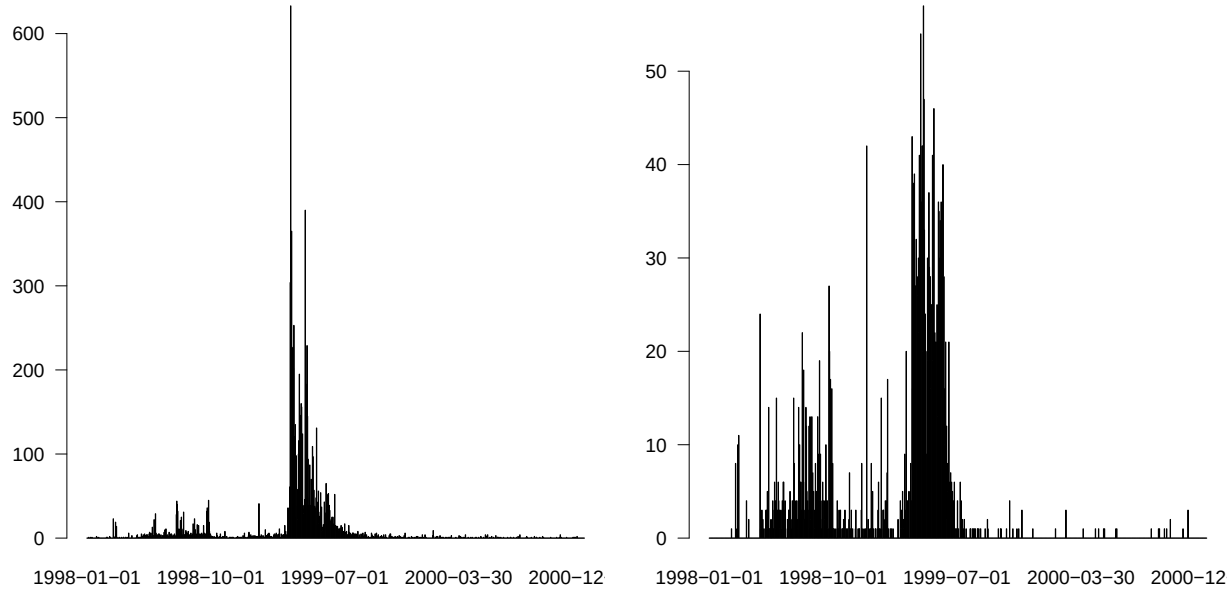


(b) Day counts



(c) Year counts

Figure 4.4: Distribution of war victims' date of violation values.



(a) Civilians (b) Members of armed formations

Figure 4.5: Distribution of violation dates, by victim status and day.

tion date information by assigning a '1' to unknown day and month information, respectively. However, KMB staff reports that the number of cases with missing information on the day and/or month of the violation is very low. Looking at the distribution of month, day, and year information, we see larger shares of war victims with violation dates on the first, second, 26th, 27th, and 28th days of a given month. HLC staff explains that the day distribution is mostly driven by mass events on certain dates that caused larger numbers of war victims. For the remaining few cases with missing day and/or month information, we recommend KMB staff correct this coding practice as advised in Section 4.7 above. There is no missing information in this database field. I.e., for 0 war victims date of violation information is completely missing.

The distribution of violation dates depending on victim status is shown in Figure 4.5. As can be seen in both plots, the majority of war victims regardless of status is reported for the first half of 1999 during March and June of that year.

4.11 Place of violation

The KMB database reports the location where a death or disappearance occured within the territory of Kosovo, Serbia, Albania, Macedonia, or Montenegro. In the case of 'killings,' this place denotes where the violation occured. In the case of 'disappearances,'

this place denotes where a victim was last seen or in contact with. In the case of 'deaths caused by war,' this place denotes where a person died. For violation locations inside Kosovo, information on the municipality is given. For locations outside of Kosovo, information on the country is provided.

In Table 4.10, information on the location of deaths and disappearances involving war victims is given by victim status. For 0 war victims (0%), information on the exact place of the violation is missing. For 17 war victims (0.1%), it is only established that the violation occured in 'Kosovë/Kosovo,' while information on the exact municipality is missing. The lack of missing information combined with a small degree of uncertain information on violation locations attests to how well the circumstances of every death and disappearance have been researched by KMB analysts.

Table 4.10: Reported war victims, by violation municipality/country and status.

Violation place/Status	Civilian	(%)	MOAF	(%)	Total	(%)
Gjakovë/Djakovica	1135	11.0	528	16.4	1663	12.3
Skënderaj/Srbica	1006	9.8	235	7.3	1241	9.2
Gllogoc/Glogovac	818	7.9	253	7.9	1071	7.9
Rahovec/Orahovac	832	8.1	119	3.7	951	7.0
Pejë/Peć	721	7.0	161	5.0	882	6.5
Prizren	597	5.8	265	8.3	862	6.4
Prishtinë/Priština	614	6.0	141	4.4	755	5.6
Suharekë/Suva Reka	490	4.8	152	4.7	642	4.7
Vushtrri/Vučitrn	493	4.8	82	2.6	575	4.3
Mitrovicë/Kosovska Mitrovica	465	4.5	97	3.0	562	4.2
Deçan/Dečani	253	2.5	271	8.4	524	3.9
Podujevë/Podujevo	413	4.0	106	3.3	519	3.8
Mališevo/Malishevë	330	3.2	92	2.9	422	3.1
Istog/Istok	271	2.6	129	4.0	400	3.0
Klinë/Klina	269	2.6	67	2.1	336	2.5
Serbia	196	1.9	120	3.7	316	2.3
Kaçanik/Kačanik	164	1.6	100	3.1	264	2.0
Lipjan/Lipljan	226	2.2	32	1.0	258	1.9
Gjilan/Gnjilane	207	2.0	31	1.0	238	1.8
Ferizaj/Uroševac	164	1.6	45	1.4	209	1.5
Fushë Kosovë/Kosovo Polje	166	1.6	7	0.2	173	1.3
Shtime/Štimlje	106	1.0	62	1.9	168	1.2

Continued on next page.

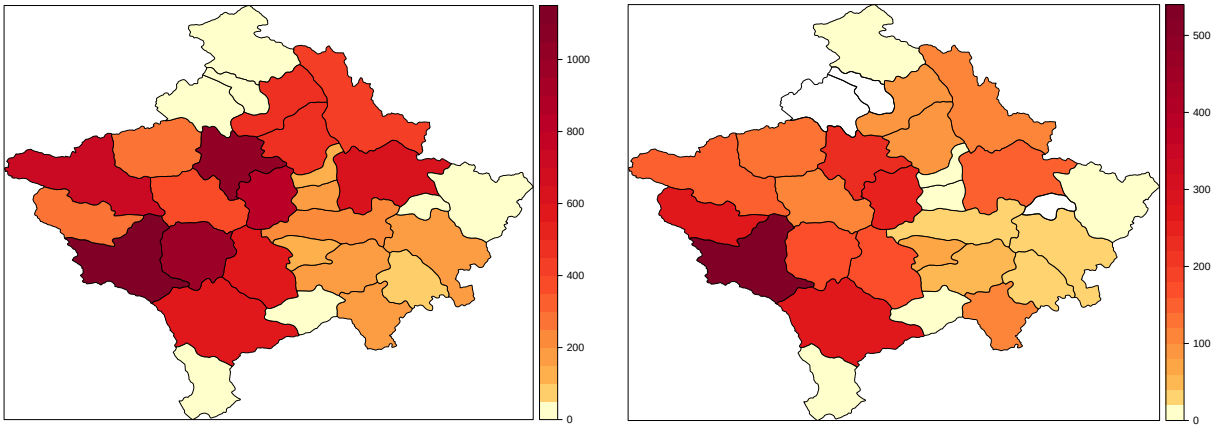
Table 4.10: Reported war victims, by violation municipality/country and status.

Obiliq/Obilić	102	1.0	18	0.6	120	0.9
Viti/Vitina	73	0.7	26	0.8	99	0.7
Kamenicë/Kosovska Kamenica	45	0.4	17	0.5	62	0.5
Montenegro	21	0.2	27	0.8	48	0.4
Zubin Potok/Zubin Potok	33	0.3	0	0.0	33	0.2
Albania	18	0.2	6	0.2	24	0.2
Shtërpcë/Štrpce	18	0.2	4	0.1	22	0.2
Dragaš/Dragash	11	0.1	9	0.3	20	0.1
Zveçan/Zvečan	19	0.2	0	0.0	19	0.1
Kosovë/Kosovo	9	0.1	8	0.2	17	0.1
Novobërdë/Novo Brdo	9	0.1	0	0.0	9	0.1
Leposaviq/Leposavić	6	0.1	1	0.0	7	0.1
Macedonia	5	0.0	1	0.0	6	0.0
Total	10305	100.0	3212	100.0	13517	100.0

The municipalities that experienced the largest number of violations were 'Gjakovë/Djakovica,' 'Skënderaj/Srbica,' 'Gllogoc/Glogovac,' and 'Rahovec/Orahovac.' Outside of Kosovo, 316 lethal violations occured in Serbia, 48 in Montenegro, 24 in Albania, and 6 in Macedonia.

Figure 4.6 provides a spatial representation of the spread of reported violations for civilian victims and members of armed forces, respectively, across municipalities in Kosovo.[7]

[7]Please note that in the graphic maps we use the administrative setup of Kosovo municipalities prior to the UNMIK reform in 2000 (see Section 6.2.2 for further details). All violation-municipality information in the KMB war victims register was converted accordingly.



(a) Civilians (b) Members of armed formations

Figure 4.6: Distribution of war victims in Kosovo 1998–2000, by victim status and municipality.

Chapter 5

Sources of the Kosovo Memory Book database

The KMB team provided us with information on the sources that document every individual registered in the Kosovo Memory Book database. In this chapter, we provide summary information, as well as an evaluation of the source documents on which the Kosovo Memory Book is built.

5.1 Number of documents per individual record

The most important quantitative criterion for assessing the quality and accuracy of the KMB database is the number of sources that is consulted per victim in order to verify both the identity of individuals, as well as the circumstances under which individuals lost their life or disappeared. The KMB team follows a precise protocol of record verification which they explained to us in interviews, in supplementary documentation, and in response to follow-up questions.[1]

For an individual to be included in the war victims register, a connection of the death or disappearance to the war has to be established. Therefore, sufficient reliable information on the circumstances of the death or disappearance has to be available. The KMB team obtains this information by collecting witness statements, consulting court judgments, as well as by evaluating other sources such as reports by human rights organizations, the Serbian police or the military. When a connection to the war is confirmed, an individual is listed in the 'war victims' register. For as long as such connection cannot

[1]For an in-depth explanation of the KMB methodology, please refer to the document prepared by Nataša Kandić (HLC) that accompanies this evaluation report.

be confirmed, an individual is listed as a 'potential victim.' When it is established that a connection to the war did not exist, an individual is listed in the 'not war' register.[2]

In addition to determining each victim's category, information provided in the available source documents is used to fill in the various fields described in Chapter 4. These are, for example, a victim's combatant/civilian status, the type of the violation, personally identifying information, as well as the date and location of the violation.

According to KMB verification rules, at least two independent sources have to confirm a connection to the war, unless a court judgment is available on a given victim case. Sources are considered independent of each other when no ties exist between them. For example, if a source is found to be based on another original source, these two sources are not considered independent of each other. Witness statements provided by members of the same family are not considered independent either. To maintain source independency during field research missions, KMB staff may interview various witnesses at different points in time to cross-check the obtained information ex- post.

To investigate the 18,001 unique individuals contained in the database, the KMB team reviewed and examined a total of 30,009 source documents. Within a given source document, often more than one victim was mentioned. To establish the KMB database, every source document was therefore thoroughly evaluated with regard to all the individuals and their respective fates listed, described, or referenced within. The resulting total number of 152,575 *victim-source connections* that were examined by the KMB team testifies to the dimension of in-depth research that was undertaken to establish the Kosovo Memory Book database. In the following analysis, we review these 'victim-source connections' in further detail as these most accurately represent the magnitude of research undertaken.

In Table 5.1, the number of sources that report an individual, i.e., the victim-source connections, are summarized across victim categories. Please note that in the scope of this evaluation, source dependency could not be controlled for. We relied on the KMB team's statement that at least two independent sources are required to establish war victims, unless a court judgment exists. In the table below, the stated number of victim-source connections therefore does not indicate whether all available documents are indeed independent of each other. Rather, Table 5.1 indicates the total of source-references per individual, be they dependent or independent of each other.

[2]KMB staff also include individuals in the KMB database as a precautionary measure. Family members may state to HLC/HLC-Kosovo researchers that a relative died during the war, but in a way unrelated to the conflict. HLC/HLC-Kosovo add such individual cases to the 'not war' register to keep a record for future reference. This may prove helpful if they find other sources to mention that individual as a war victim.

Table 5.1: Number of sources per individual, by victim category.

Number/Category	War	(%)	Potential	(%)	Not war	(%)	Total	(%)
1	0	0.0	1151	71.8	475	16.5	1626	9.0
2	111	0.8	217	13.5	995	34.5	1323	7.3
3	225	1.7	120	7.5	578	20.1	923	5.1
4	373	2.8	54	3.4	372	12.9	799	4.4
5	612	4.5	39	2.4	196	6.8	847	4.7
6	1010	7.5	14	0.9	130	4.5	1154	6.4
7	1287	9.5	6	0.4	79	2.7	1372	7.6
8	1460	10.8	2	0.1	25	0.9	1487	8.3
9	1366	10.1	0	0.0	14	0.5	1380	7.7
10	1232	9.1	0	0.0	2	0.1	1234	6.9
11	1122	8.3	0	0.0	9	0.3	1131	6.3
12	980	7.3	0	0.0	2	0.1	982	5.5
13	804	5.9	0	0.0	2	0.1	806	4.5
14	619	4.6	0	0.0	2	0.1	621	3.4
15	495	3.7	0	0.0	0	0.0	495	2.7
16	466	3.4	0	0.0	0	0.0	466	2.6
17	342	2.5	0	0.0	0	0.0	342	1.9
18	266	2.0	0	0.0	0	0.0	266	1.5
19	225	1.7	0	0.0	0	0.0	225	1.2
20	149	1.1	0	0.0	0	0.0	149	0.8
21	99	0.7	0	0.0	0	0.0	99	0.5
22	73	0.5	0	0.0	0	0.0	73	0.4
23	61	0.5	0	0.0	0	0.0	61	0.3
24	40	0.3	0	0.0	0	0.0	40	0.2
25	31	0.2	0	0.0	0	0.0	31	0.2
26	20	0.1	0	0.0	0	0.0	20	0.1
27	15	0.1	0	0.0	0	0.0	15	0.1
28	10	0.1	0	0.0	0	0.0	10	0.1
29	4	0.0	0	0.0	0	0.0	4	0.0
30	3	0.0	0	0.0	0	0.0	3	0.0
31	8	0.1	0	0.0	0	0.0	8	0.0
32	3	0.0	0	0.0	0	0.0	3	0.0
33	3	0.0	0	0.0	0	0.0	3	0.0
35	2	0.0	0	0.0	0	0.0	2	0.0
37	1	0.0	0	0.0	0	0.0	1	0.0
Total	13517	100.0	1603	100.0	2881	100.0	18001	100.0

Furthermore, we depict the information contained in Table 5.1 in graphical form in Figure 5.1, by victim category. These histograms help visualize the frequency distribution of the number of source connections per victim.

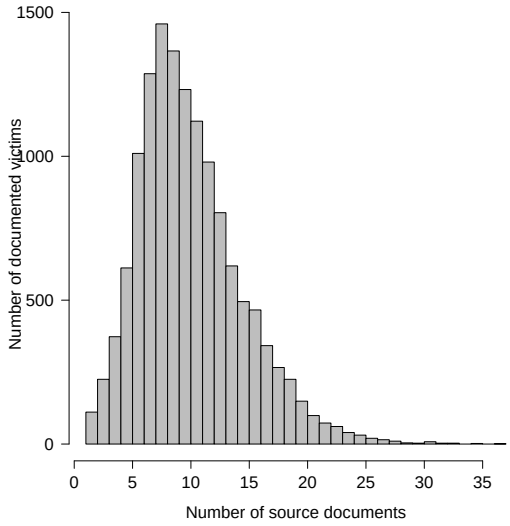
As can be seen in Table 5.1, as well as Figure 5.1, the number of source documents reviewed per war victim ranges between 2 and 37 individual documents. The most common number of sources per war victim is 8.

For potential victims, in turn, the number of available sources ranges between 1 and 8, with the most common number of available source documents being 1. The latter indicates the preliminary nature of possible violations listed in the registry of potential victims. The exact circumstances of the relevant deaths and disappearances are currently unclear for these records regardless of how many sources may be reporting on them. At the time of writing, the KMB team is still in the process of examining these potential victims.

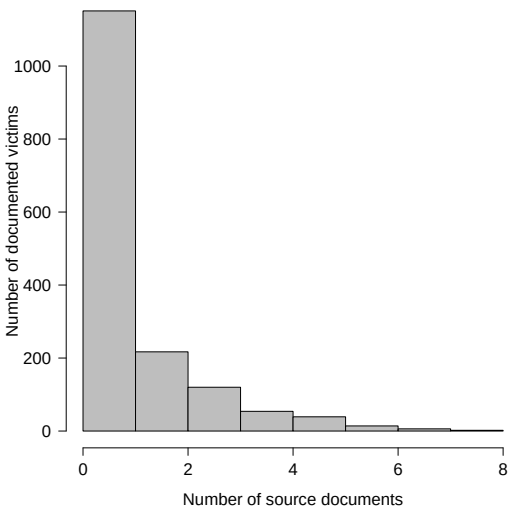
As can be seen in Figure 5.1(a), as well as Table 5.1, it is not very likely that war victims are reported by two sources only. Rather, it is most common that 8 sources report on a given war victim. Juxtaposing this with the fact that for most potential victims, only 1 independent source is reporting on these victims to date, it is unlikely that these sources are reporting on cases of 'true' war victims. As we will discuss later in Section 6.4, we therefore regard it highly unlikely that the current size of the register of potential victims suggests that the war victims register is incomplete. It is possible that there may be several tens of victims among these potential victims that constitute cases of 'true' war victims, but unless more social information becomes available this will be impossible to verify by KMB staff.

Finally, for individuals identified as 'not war' victims, the number of source documents reviewed ranges between 1 and 14, with the most common number of documents being 2. In addition, Table 5.2 below breaks down the type of death for the 475 'not war' victims with only one source connection each. As can be seen, 290 individuals were found to be alive. For 185 individuals, in turn, the circumstances of the death are either unknown, or the death was classified by KMB staff as an accident or as a natural death.

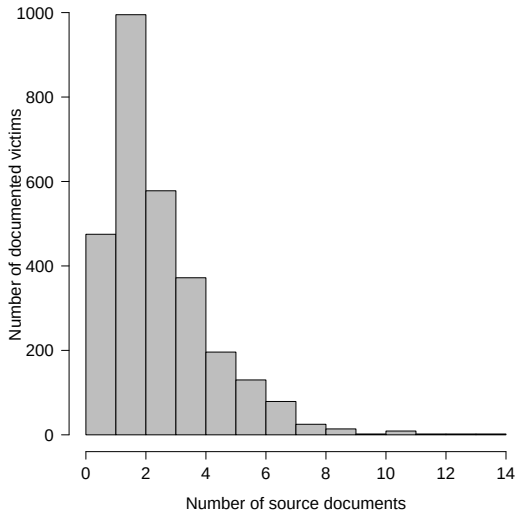
We note that there may be some inconsistency with regard to how victims were confirmed to be connected to the war versus not being connected to the war. One could argue that for verification rules to be applied consistently across victim categories, victims that are not found to be alive have to be listed as 'potential victims' for as long as only one source is reporting on them. A connection to the war would be rejected only when sufficient evidence from at least two independent sources becomes available that allows



(a) War victims



(b) Potential victims



(c) Not war victims

Figure 5.1: Frequency distribution of source document counts per victim, by victim category.

to classify potential victims as 'not war' victims.

Table 5.2: Reported 'not war' victims with one source connection, by violation.

Violation	Victims	Percent of victims
Found alive	290	61.1
Unknown	166	34.9
Death by accident	12	2.5
Natural death	7	1.5
Total	475	100

5.2 Descriptive summary of source documents

The KMB's process of data collection changed over time, depending on the availability of and access to various types of sources. To maximize the quantity and accuracy of the information, KMB staff reviewed a vast variety of source types to obtain information on war victims.

In Table 5.3, we provide a summary of all source documents by type. As can be seen, more than a third of all source documents underlying the KMBD are statements. This amount underlines the magnitude of independent research that is undertaken to create the KMB. In the effort of confirming and verifying war victims, as well as clarifying circumstances of deaths and disappearances, KMB staff collect statements of relatives and witnesses in the field.

Resulting from the fact that witness statements are the most important source used to create the KMBD, the majority of internal source documents are of primary character (Table 5.4). According to KMB terminology, primary sources provide "first-hand evidence of victims." They originate from immediate participants or observers of violations, and can include autobiographies, memoirs, and oral histories that were recorded after a violation occurred. Secondary sources, in turn, are "materials that process, analyze, evaluate and/or interpret information contained within primary or other secondary sources."[3]

Next in Table 5.5, we summarize the origin or author of source documents in terms of the seven main categories that were established according to KMB methodology. As can be seen, the largest number of source documents originates from the group of non-governmental organizations (NGOs), civil society organizations (CSOs), and private persons. Witnesses, human rights groups, associations of victims' families, war veterans, or

[3]Definitions and descriptions of terms can be found in the accompanying methodology document by Nataša Kandić (HLC).

Table 5.3: Source documents, by type.

Type	Documents	Percent of documents
Statement	10769	35.9
Photograph	6256	20.8
Verification	3703	12.3
Media item	1868	6.2
Memorial	1832	6.1
Confirmation	1613	5.4
Personal document	1313	4.4
In memoriam	759	2.5
Report	558	1.9
Records	167	0.6
Verdict	163	0.5
Listing	151	0.5
Transcript	144	0.5
Decision	114	0.4
Indictment	102	0.3
Book	101	0.3
Medal	56	0.2
Questionnaire	54	0.2
Criminal complaint	48	0.2
Notion	43	0.1
Letter/Fax	42	0.1
Official note	34	0.1
Video recording	30	0.1
Order	21	0.1
Request	20	0.1
Submission	17	0.1
Diary	15	0
Appeal	12	0
Drawing/sketch/scheme	3	0
Map	1	0
Total	30009	100

Table 5.4: Source documents, by character.

Character	Documents	Percent of documents
Primary	18062	60.2
Secondary	11947	39.8
Total	30009	100

citizens, as well as local or religious communities form part of this category. The second largest group of source documents originates from civil state authorities such as Serbian state institutions, UNMIK, and EULEX, as well as municipal authorities, health and educational agencies.

Table 5.5: Source documents, by group.

Group	Documents	Percent of documents
NGOs, CSOs, private persons	22298	74.3
Civil state bodies	2713	9
Media and publishers	1979	6.6
International state, civil, other institutions	1214	4
Military and police authorities	962	3.2
Judicial institutions	767	2.6
Unknown	50	0.2
Political parties, companies, etc.	26	0.1
Total	30009	100

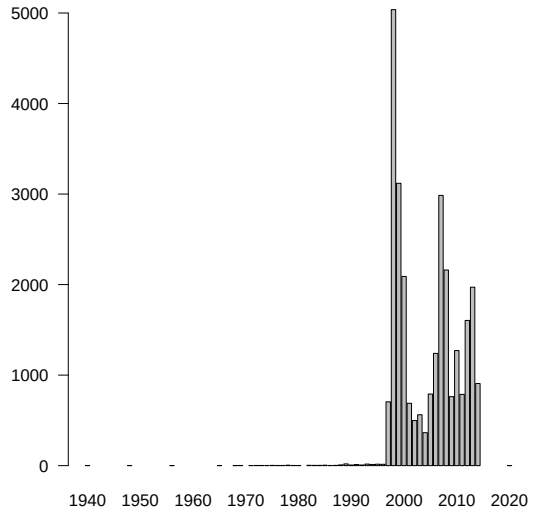
A major strength of HLC's and HLC-Kosovo's data collection effort is the multitude of languages in which source material was evaluated. This is important because relevant information could be published by local, regional, and international actors, during as well as after the conflict. As can be seen in Table 5.6, the range of languages includes all regional languages that are prevalent in the territory of observation, i.e., Albanian, Serbian, Bosnian, and Croatian, plus English, or combinations of all former in case source documents were published multi-lingually. A little over half of all source documents was published in Albanian. For 6,275 documents information on the language is missing. One reason for this is that many photographs do not contain language. Furthermore, KMB staff did not enter language information for approximately 150 memorial type documents; these missing fields will be corrected in the next version of the data.[4]

HLC and HLC-Kosovo did not restrict themselves to any specific period when the 30,009 source documents underlying the KMBD were created. Source material was required to confirm both the identity of victims, as well as the occurrence of a war-related violation. Therefore, the original dates of when a source document was created cover a considerable time period for all victim categories, as can be seen in Figure 5.2.[5]

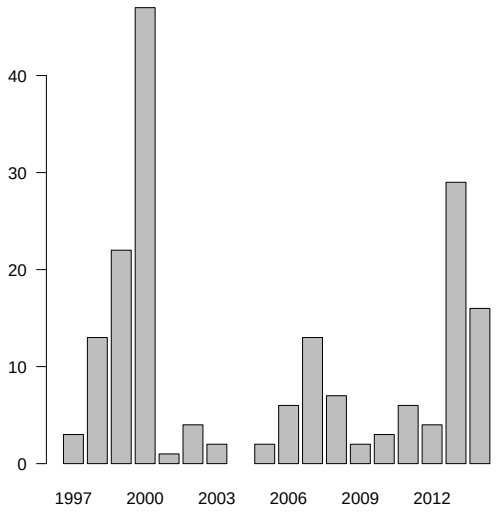
In Table 5.7, the different types of sources that were evaluated are summarized by victim category and in terms of victim-source connections. As can be seen, the most

[4]Email conversation with Predrag Miletic, November 26, 2014.

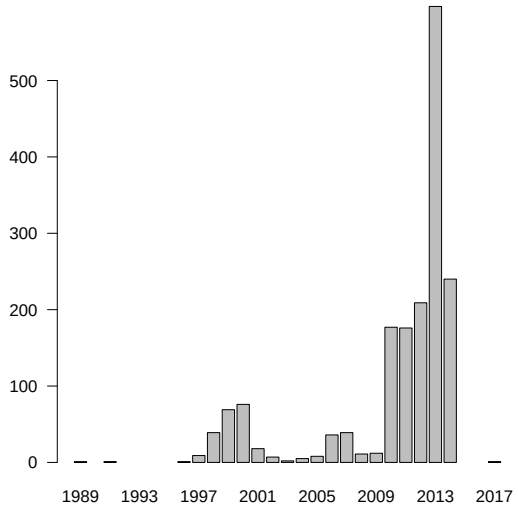
[5]Note that among war victims, 1 source document has a creation date in the year 2020, while among 'not war' victims 1 source document has a creation date in 2017. We suspect that these are data entry errors requiring correction.



(a) War victims



(b) Potential victims



(c) Not war victims

Figure 5.2: Distribution of date of origin of KMB internal source documents, by victim category and year.

Table 5.6: Source documents, by language.

Language	Documents	Percent of documents
Albanian	12010	40
Information missing	6275	20.9
Bosnian-Croatian-Serbian (BCS)	6094	20.3
Albanian/BCS	3318	11.1
English/Albanian/BCS	1098	3.7
English	958	3.2
English/Albanian	148	0.5
English/BCS	88	0.3
Other	20	0.1
Total	30009	100

common type of victim-source connections across all victim categories are victims reported in books. This is due to the fact that many long victim listings were published in book format, such as the victim enumerations by the Democratic League of Kosovo (LDK), *Martinsen*, and *Osmani* (see Section 6.1 for a more detailed description of these three source documents). For war victims, the second and third most common type of victim-source connections are witness statements and victim lists.

Table 5.7: Type of victim-source connections, by victim category.

Type/Category	War	(%)	Potential	(%)	Not war	(%)	Total	(%)
Book	42836	30.3	1050	42.0	2965	34.8	46851	30.7
Statement	25934	18.3	115	4.6	1226	14.4	27275	17.9
Listing	25120	17.7	404	16.2	1570	18.4	27094	17.8
Report	13305	9.4	573	22.9	653	7.7	14531	9.5
Media item	8221	5.8	200	8.0	417	4.9	8838	5.8
Photograph	6529	4.6	7	0.3	41	0.5	6577	4.3
Memorial	4940	3.5	64	2.6	61	0.7	5065	3.3
Verification	3396	2.4	21	0.8	1422	16.7	4839	3.2
Questionnaire	2351	1.7	3	0.1	8	0.1	2362	1.5
Confirmation	1618	1.1	8	0.3	23	0.3	1649	1.1
In memoriam	1499	1.1	24	1.0	5	0.1	1528	1.0
Personal document	1284	0.9	9	0.4	30	0.4	1323	0.9
Records	1230	0.9	1	0.0	3	0.0	1234	0.8
Transcript	844	0.6	11	0.4	40	0.5	895	0.6
Verdict	798	0.6	0	0.0	20	0.2	818	0.5
Indictment	527	0.4	0	0.0	6	0.1	533	0.3
Decision	184	0.1	0	0.0	4	0.0	188	0.1
Official note	150	0.1	1	0.0	2	0.0	153	0.1
Criminal complaint	137	0.1	1	0.0	9	0.1	147	0.1
Diary	115	0.1	1	0.0	1	0.0	117	0.1
Video recording	114	0.1	1	0.0	2	0.0	117	0.1

Continued on next page.

Table 5.7: Type of victim-source connections, by victim category.

Order	116	0.1	0	0.0	0	0.0	116	0.1
Request	72	0.1	0	0.0	3	0.0	75	0.0
Letter/Fax	66	0.0	1	0.0	5	0.1	72	0.0
Notion	55	0.0	3	0.1	6	0.1	64	0.0
Medal	59	0.0	0	0.0	0	0.0	59	0.0
Drawing/sketch/scheme	21	0.0	0	0.0	1	0.0	22	0.0
Submission	19	0.0	0	0.0	0	0.0	19	0.0
Appeal	13	0.0	0	0.0	0	0.0	13	0.0
Map	1	0.0	0	0.0	0	0.0	1	0.0
Total	141554	100.0	2498	100.0	8523	100.0	152575	100.0

Chapter 6

Evaluation of database completeness

In this chapter, we assess to what extent the KMB database may represent a complete enumeration of all war victims related to armed conflict in Kosovo. We compared the KMB database to other sources that also report on victims in Kosovo between 1998 and 2000. In Section 6.1, we present the different data sources that were available for cross-comparison and evaluation. In Section 6.2, we compare the temporal trend and spatial distribution of violence in each of the various available data sources reporting on violence in Kosovo between 1998 and 2000. In Section 6.3, we describe the process and outcomes of our analysis of record capture. Finally in Section 6.4, we discuss whether the remaining number of potential victims challenges the KMB's completeness.

6.1 Other data sources for comparison

Additional to the KMB database, ten data sources are available to cross-compare the reporting of victims during armed conflict in Kosovo, 1998–2000. The way information on victims was collected differs largely across these sources. In particular, it differs in terms of the chosen method(s) for obtaining and evaluating information, the timing and regional spread, as well as, the intended scope of the data collection effort. In the following, each data source will be briefly presented.[1]

The non-profit *American Bar Association/Central and East European Law Initiative* (**ABA**/CEELI) conducted interviews with ethnic Albanian refugees in refugee camps or private homes in Kosovo, Albania, Macedonia, Yugoslavia, Poland, and the United States. International teams on behalf of the International Tribunal for the Former Yu-

[1]KMB team members kindly provided information on some of the data sources outlined below.

goslavia (ICTY) conducted *exhumations* (**EXH**) in 24 of the 29 municipalities in Kosovo.[2] The non-governmental organization *Human Rights Watch* (**HRW**) interviewed ethnic Albanian refugees as they left Kosovo via borders to Albania, Macedonia, or Montenegro between March and June 1999. Additional interviews were conducted in various regions throughout Kosovo between June and December 1999. The *Organization for Security and Cooperation in Europe* (**OSCE**) conducted a Kosovo Verification Mission starting in October 1998, in addition to running an OSCE Mission in Kosovo since June 1999. Between March and June 1999, OSCE staff interviewed ethnic Albanians in refugee camps, private homes or communal places in Albania and Macedonia, but not inside Kosovo proper. After the establishment of the OSCE Mission in Kosovo in June 1999, observers resumed the collection of statements from victims and witnesses inside Kosovo.[3]

Three sources were shared with HRDAG in 2007 by the ICTY Office of the Prosecutor to provide additional data for expert testimony prepared for the case of Milutinovic et al. (IT-05-87). The *International Committee of the Red Cross* (**ICRC**) compiled a list of missing persons based on applications and inquiries of family members, providing personal information, as well as where and when someone was last seen.[4] The UNMIK *Office of Missing Persons and Forensics* (**OMPF**) operated between 2002 and 2010 to clarify the fate of missing individuals by way of exhumations, autopsies, and forensic analysis. It provides identifying information on individuals who were exhumed.[5] The *International Commission on Missing Persons* (**ICMP**) produced a list of missing persons who, after being refered to ICMP by OMPF or the Serbian government, were identified with the help of DNA analysis.[6]

The *Democratic League of Kosovo* (**LDK**) is a Kosovar political party. In 2002, it published a book on the consequences of the war in Kosovo that reported, among other things, on casualties and damages in relation to the armed conflict.[7] LDK states that information was collected by local LDK committees in the field who visited every family.

[2]The exhumations continued until at least 2004. The data used in this report, however, only includes victims identified through April 2001.

[3]Ball, P. and J. Asher (2002): '*Statistics and Slobodan: Using Data Analysis and Statistics in the War Crimes Trial of Former President Milosevic*,' *Chance*, 15(4), 19–20.

[4]The list available to HRDAG for analysis is an extraction of the ICRC database as of April 1, 2005. The ICRC database has been updated since.

[5]HRDAG obtained the consolidated OMPF list as of October 27, 2006. The consolidated list is based on information collected by OMPF and the ICTY, supplemented with information from the UNMIK police, OSCE, and associations of families of missing persons. The OMPF database is nowadays accessible via the online database of the ICTY.

[6]Note that ICMP does not provide information on the circumstances of a death. The list provides information on the exhumation, the adequacy of the DNA profile, personally identifying details of victims, and whether remains were returned to the family. According to the KMB team and based on KMB research, some individuals listed in the ICMP register may not be connected to the war.

[7]LDK (2002): "*Pasojat e luftës në Kosovë, 1998–1999.*"

In 2010, *Josef Martinsen* (**Martinsen**) published a register of dead or missing civilians from the Kosovo war between 1998–1999.[8] Data was collated from a number of sources: the ICRC, the ICTY, OMPF, HLC, HRW, No Peace Without Justice, OSCE, as well as, the Council for Defence of Human Rights and Freedoms. The author also conducted field research and interviews from 2001 to 2004, and from 2006 to 2009.

In 2012, *Jusuf Osmani* (**Osmani**) published a register of civilian victims who were killed or missing as a result of war crimes and human rights violations committed by Serbian authorities in Kosovo, as well as of fallen fighters.[9] Data was collected by way of investigations, as well as by consulting a variety of sources such as, for example, municipal assemblies, national governmental authorities, ICRC, HLC, and victim associations.[10]

In Table 6.1 below, the various lists are briefly presented in order of magnitude regarding the number of reported civilian victims for the period 1998 to 2000.[11] As can be seen, ABA reports the fewest and LDK the most victims. The third column ('Operation') provides the period of data collection underlying each source available for comparison. While some of the data sources became available to HRDAG shortly after violence in Kosovo ended (e.g., ABA, EXH, HRW, OSCE), other data collection efforts continued until recently or even through the present (e.g., *Martinsen*, *Osmani*, HLC). Variation in terms of when data collection started and ended is likely to be linked to the total number of victims reported. With more time available for data collection, the number of victims that can be identified and reviewed increases. Furthermore, later data sources may have been able to access earlier victim enumerations in order to compare them to their own lists.

In the final column of Table 6.1, we note whether KMB staff had access to a given data source and were able to compare it to the KMB register. This information is important because when a data source is accessible to the HLC and HLC-Kosovo, it is likely to be fully absorbed into the KMB by way of the research and record confirmation procedures that are being followed. This information is further relevant to our evaluation of the KMB's completeness below.

The HLC and HLC-Kosovo did not have access to approximately half of these lists. The ABA list was not public, and therefore it was not shared with the HLC or HLC-Kosovo. KMB staff only had access to HRW's published documents and reports. HRW

[8] Josef Martinsen (2010): '*What happened in Kosovo?, 1998–1999.*'

[9] Jusuf Osmani (2012): "*Krimet e Serbisë në Kosovë, 1998–1999.*"

[10] Spreadsheets with all victim records as originally listed in LDK, *Martinsen*, and *Osmani* were given to us by the HLC. HRDAG also transcribed the victim records listed in *Martinsen*'s publication.

[11] Note that in some of these datasets, the violation date for some victims can be missing. It was then assumed that the respective victim was reported for the time period 1998–2000.

publications are deemed very valuable to KMB researchers for their information on circumstances of violations and events, as well as some names of victims that are mentioned. The HLC and HLC-Kosovo, however, had no access to HRW's full lists of victims underlying the publications. The EXH list was neither public, nor shared with HLC/HLC-Kosovo. KMB researchers only knew about exhumations if told by victim families. The HLC and HLC-Kosovo used the public OSCE reports 'As Seen, As Told' in Serbian language, but never had access to the actual OSCE database underlying those reports. The OSCE reports do not mention the names of victims, but provide information on cases and circumstances. Facts on circumstances helped KMB staff to recognize individual cases, to then compare those to the KMB register.

The ICMP list presents a special case. KMB staff had no access to this register because this organization does not publish names of victims.[12] However, it may be that ICMP cases are contained in the consolidated list by the OMPF as the work of these two organizations was closely connected.

To the following data sources the HLC and HLC-Kosovo do have access: A consolidated OMPF list became publicly available after the ICTY included it in its online database in 2009. Furthermore, the KMB team has full access to the listings and publications by the ICRC,[13] as well as the publications by the LDK, *Martinsen*,[14] and *Osmani*. KMB staff also uses ICRC and OMPF information to declare individuals as 'not war' victims in case individuals are declared as found alive.

6.2 Cross-source comparison

In this section, we compare the reporting of victims across available sources with regard to time (Section 6.2.1) and space (Section 6.2.2). The goal is to identify whether other data sources report victims for specific temporal or spatial units that are not contained in the KMB database. If we were to find victims reported for specific times or places that are not also covered by HLC/HLC-Kosovo, this would be an indicator that the KMB is not complete.

Please note that for the purposes of completeness evaluation, we use the entire KMB

[12]The KMB team obtains information on ICMP-listed victims indirectly when victims' families share death certificates. When KMB staff know the names of victims and a few other personally identifying details, they can query ICMP's online database by using this information.

[13]ICRC does not provide information about the circumstances of individual victims. KMB staff requires additional sources to determine a connection with the war.

[14]While for this report we only use Martinsen's 2010 publication, HLC also evaluated another of his works – Martinsen (2005): '*The Wells of Death, Nine Weeks in Spring of 1999.*'

Table 6.1: Lists of civilian victims in Kosovo, by magnitude of total number of reported victims for the period 1998–2000.

Source name	Abbr.	Operation	N(victims)	Accessed by HLC?
American Bar Association	ABA	1999	727	No
Human Rights Watch	HRW	1999	906	Publications, not database
International Commission for Missing Persons	ICMP	2003–2006	1726	No, maybe via OMPF
Exhumations	EXH	1999–2001	2123	No
International Committee of the Red Cross	ICRC	1998–2005	2888	Yes
Organisation for Security and Cooperation in Europe	OSCE	1998–1999	3338	Publications, not database
Office of Missing Persons and Forensics	OMPF	2002–2006	4547	Yes
Jusuf Osmani	Osmani	1999-2012	9374	Yes
Humantarian Law Centre and HLC-Kosovo	HLC	1998–present	10305	–
Josef Martinsen	Martinsen	2001–2009	10505	Yes
Democratic League of Kosovo	LDK	1998–2002	11287	Yes

database (18,001 individuals) in this comparison against the other ten data sources, i.e., the sum of all war, potential, and 'not war' victims for a given temporal or spatial unit. This is being done because we believe that HLC has conducted the most comprehensive effort to research the circumstances of every victim.

Based on their work, KMB researchers may have determined that certain individuals were members of armed formations. However, it is possible that the ten data sources available for comparison may list such individuals as civilian war victims. Furthermore, for a lack of information on the circumstances of a death, KMB staff may currently list a certain record as a 'potential victim' even though the victim may be reported as a war victim in other sources. Similarly, KMB researchers may have classified a certain individual as a 'not war' victim given detailed information on the circumstances of a death that do not prove an actual connection to the war, or because someone was found alive later on. Such individuals may still be listed as war victims in some of the ten data sources available for comparison to the KMB.

6.2.1 Reporting over time

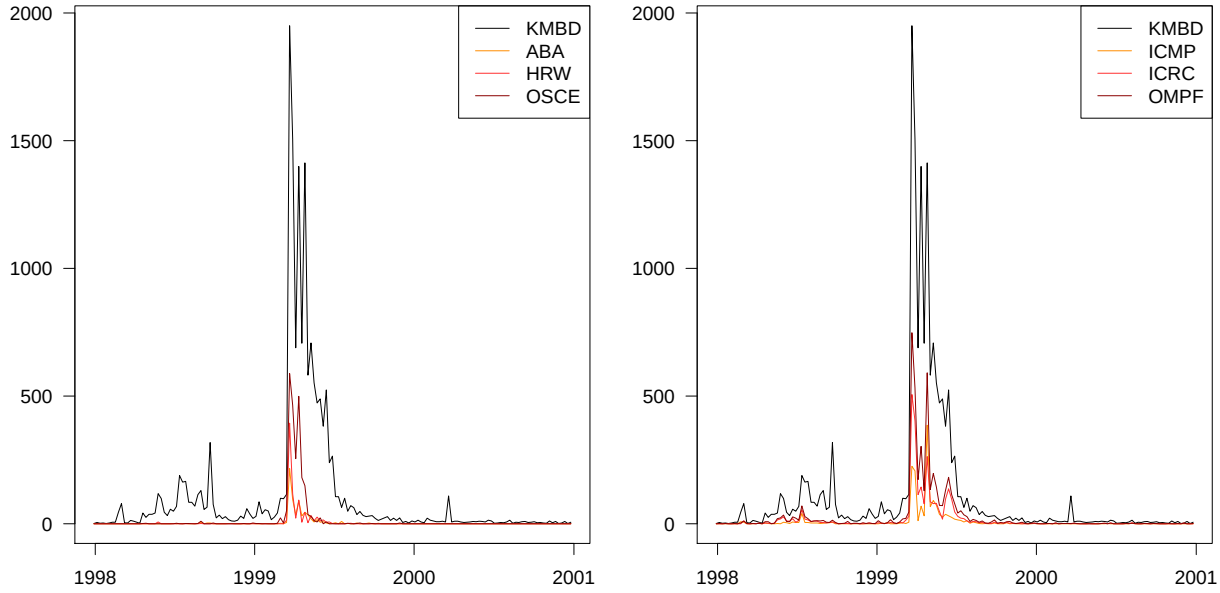
First, we perform a comparison of the number of reported victims over time. Nine of the above mentioned data sources available for comparison report a violation date.[15]

In Figure 6.1, we compare two groups of sources against the KMBD with regard to the number of weekly reported victims for the period 1998–2000, respectively. The first group in Figure 6.1(a) are the three data sources that became available to HRDAG soon after conflict in Kosovo ended. HLC/HLC-Kosovo never gained access to these victim listings. As can be seen, the three lines for ABA, HRW, and OSCE always stay below the black line that indicates the weekly counts of individuals listed in the KMBD. These sources therefore do not suggest weekly victim counts that exceed those documented in the KMB data.

The same observation can be made by looking at Figure 6.1(b). Among this second group of sources, ICRC and OMPF became publicly available at a later time. To date, the KMB has not gained direct access to the ICMP list of missing persons. The three data sources of this second group report weekly counts of victims that are of higher magnitudes than the first group examined. Yet, we observe again that the victim counts always remain below the black KMBD line.

Next in Figure 6.2, we compare the third and final group of LDK, *Martinsen* and

[15]The exhumations data (EXH) do not form part of the temporal comparison. In this source, the reported date indicates when human remains were found.



(a) ABA, HRW and OSCE vs. KMBD

(b) ICMP, ICRC and OMPF vs. KMBD

Figure 6.1: Reported victims by data source and week, part 1.

Osmani data against the KMB victim register. These data sources invested a comprehensive effort in collecting information on all human losses in Kosovo. Note that KMB staff has access to all three data sources and was therefore able to fully absorb the reported victim records into their own database. Not surprisingly, the respective victim enumerations come very close to the KMB's weekly counts, which is why we present one source comparison per plot.

Again, we find that overall the black KMBD line exceeds the weekly counts reported in the other data sources. However, as can be seen in Figure 6.2(a), in early 1998 and 1999, as well as the second half of 1999, the red LDK line exceeds that of weekly victim counts documented in the KMB register. It is likely that this observation is due to date encoding errors in the LDK data. We found a substantial number of LDK victim records that had day and month information transposed in the violation date field. For example, a true violation date of May 8 could mistakenly be presented as August 5. Given that the KMB database reports correct violation dates, this could explain the discrepancy between the two timelines we observe here.

From our comparison of the number of weekly reported victims across data sources we conclude that the KMB register does not miss any victim counts for the period 1998–2000. None of the sources across the three groups considered reports weekly temporal counts that exceed those documented in the KMBD. An exception to this is the LDK data. We have, however, reason to believe that the weekly counts reported by this data

source result from incorrect date encodings.

6.2.2 Reporting over space

Second, we compared the number of reported victims for the location of death. Eight of the above mentioned data sources available for comparison report a place of violation.[16]

Please note that in the spatial analysis, we use the administrative division of Kosovo into districts and municipalities that was in place prior to the administrative reform by the United Nations Interim Administration Mission in Kosovo (UNMIK) in the year 2000. This administrative reform led to the formation of new districts. All municipality information across data sources had to be converted to the pre-reform administrative setup for the purposes of this comparison. A number of sources that we used in the comparison used the pre-reform municipality division (i.e., ABA, EXH, HRW, and OSCE). We are unable to convert these data sources to the municipal organization after UNMIK's administrative reform in 2000, so we chose the pre-reform organization that was in place at the time of most of the events in the KMB.

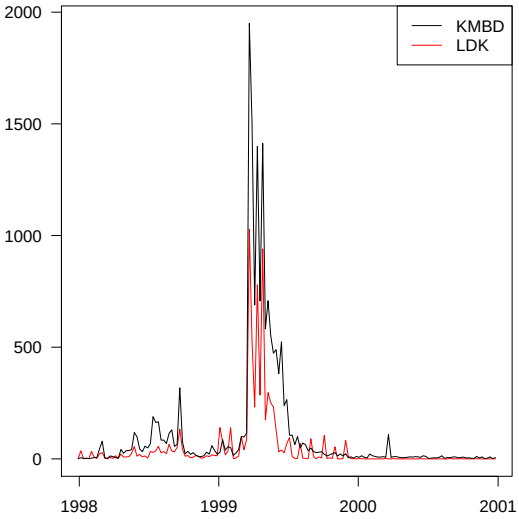
In Figure 6.3, we present the reported victim counts for the period 1998–2000 by municipality and data source. The more victims are reported for a given municipality, the darker that area in the respective municipality. Therefore, if other data sources were to report victim counts that exceed those reported in the KMB data, we would see darker shaded areas in their municipality maps compared to the KMBD map. As can be seen, across all data sources the KMBD reports the same or more victims for nearly every municipality.

There is one exception: in the municipality of Kaçanik/Kačanik there are more reported deaths for the OSCE than for the KMBD (469 for OSCE versus 330 for KMBD). Based on our analysis of the patterns of 'singleton' records found in non-KMB databases but not in the KMB (see Section 6.3 for details), we believe the OSCE reports inaccurate records for Kaçanik/Kačanik.

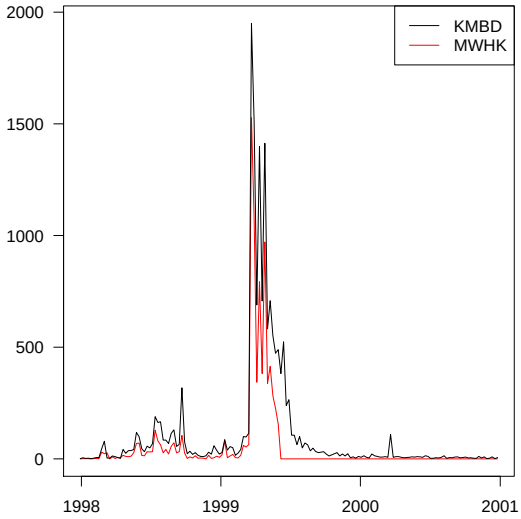
From our cross-source comparison of the number of reported victims per municipality we conclude that the KMB register is unlikely to exclude any valid victim records. None of the other data sources reports convincing victim counts across the Kosovo municipalities that exceed those documented in the KMB database.

The cross-source comparison in this section is not sufficient to evaluate the complete-

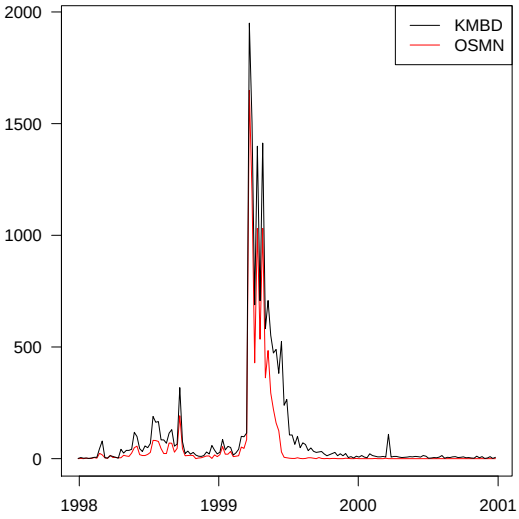
[16]The exhumations data (EXH) do not form part of the spatial comparison. In this data, the reported location indicates where human remains were found. The LDK list does not provide information on the municipality in which a violation occurred.



(a) LDK vs. KMBD



(b) Martinsen vs. KMBD



(c) Osmani vs. KMBD

Figure 6.2: Reported victims by data source and week, part 2.

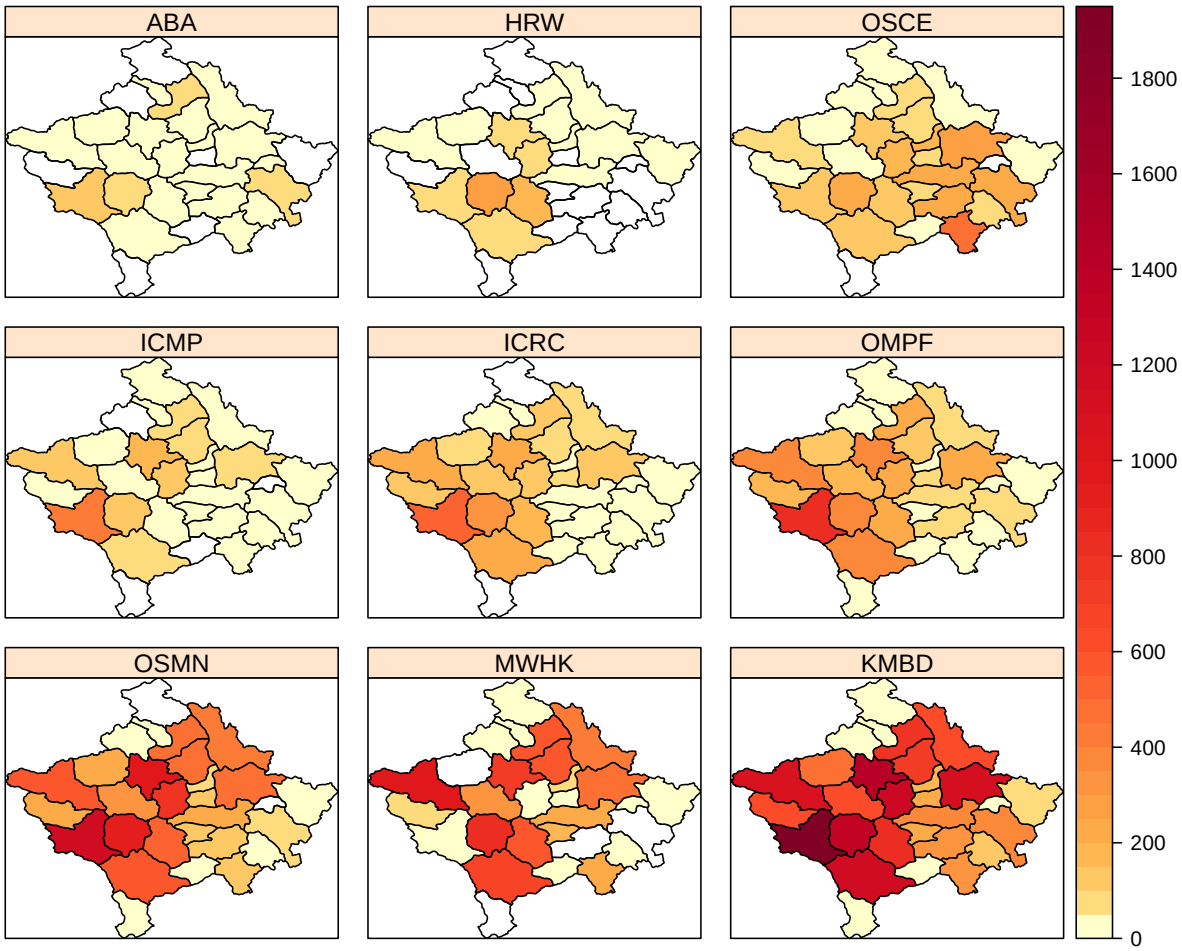


Figure 6.3: Reported victims in Kosovo 1998–2000, by data source and municipality.

ness of the Kosovo Memory Book. Count comparisons do not tell us whether all individual victims reported in one data source truly refer to a subset of all individuals reported in the KMB database. The logical misuse of statistical data to draw an inference about individuals based on group-level information is known as *ecological fallacy*. To conclude that because the KMB reports more victims in each location and period than any other source, it must therefore be a superset of all the other sources, would be an ecological fallacy. The findings in this section are a necessary but insufficient condition with which to conclude that the KMB is a complete list of victims. We address this problem in the next section.

6.3 Evaluation of record capture

To accurately compare victim records in other data sources against the database of the Kosovo Memory Book, we linked victim records across all the available data sources on violence in Kosovo. This cross-validation process of 'record linkage' is also called 'matching.' The goal of matching is to identify victim records in non-KMB lists that cannot be linked to any entries in the KMBD. If such records were identified, it would suggest that the KMB database is incomplete and missing war victims had to be added. If it was found however that all victims in other sources are also registered with HLC/HLC-Kosovo, the KMB database can be regarded complete.

It is possible that there are some victims that have not yet been documented by the KMB team. The matching test conducted here is an informal assessment of how many victims are likely to remain undocumented. We were unable to find convincing evidence of any records that are not captured in the KMB database, even when considering data sources to which the HLC and HLC-Kosovo did not have access. Therefore, we believe that there are relatively few additional killings that the HLC and HLC-Kosovo have not yet documented. In the following, we outline the process of cross-validation.

In the matching, we included several additional lists to supplement the non-KMB lists presented in Section 6.1. One is the KMB list of 'not war' victims because these individuals could still be documented as war victims in other data sources. Another is an enumeration of individuals by the ICRC for whom search requests were canceled by families, or who had been found alive. A third list of closed cases originates from the OMPF denoting individuals who were found alive, or cases that were closed for a lack of further information on a supposed victim. It was important to include these lists along with the KMB lists of victims of armed-group membership or potential status. Data sources that were produced shortly after the conflict ended may still report individuals whose cases were later closed. Some sources may erroneously list combatants as civilian victims, with an individual's status clarified by KMB researchers at some later point. Other victims may be regarded as potential cases until the KMB team determines that reliable information on circumstances establishes a connection to the war.

The following sources were included in the matching process:

1. KMB: three lists of a) war victims, b) potential victims, and c)'not war' victims
2. ICRC: two lists of a) victims, and b) closed cases and canceled requests
3. OMPF: two lists of a) victims, and b) found alive or closed cases

4. ABA
5. EXH
6. HRW
7. ICMP
8. LDK
9. *Martinsen*
10. OSCE
11. *Osmani*

Every field that provided information on a victim and the connected violation was compared (see Sections 4.1 to 4.11 in Chapter 4). To facilitate this comparison across sources, fields were standardized. Compared to every other source listed above, the KMB database provides the most detail on every case. This complicated the cross-source record comparison process because relative to the KMB, other sources provide insufficient or inaccurate detail on reported victims. The quality of data sources under comparison differed largely as a result. The KMB's most comprehensive information on every victim case had to be compared to lists of much lower quality. What follows is a list of data coding issues we encountered in non-KMB lists that complicated our record comparison:

- Spelling variations of individual and location names caused by transcriptions of Albanian and Serbian information into English field values by non-native speakers. Common spelling issues were noticed. For example, 'K,' 'Q,' and 'C' were often interchanged. 'GJ' was interchanged with 'XH' or 'XJ.' 'D' and 'T' were sometimes also inter-changeable.
- Errors in name fields as some datasets coded nicknames, diminutives, maiden names, or titles (e.g., 'Commander') in relevant name information. Some data projects erroneously reversed information on father's and last name.
- Data fields that refer to other concepts. For example, location and date information had to be deleted in EXH because it denoted where and when a body was found instead of where and when a violation occurred.

- Data fields that refer to several concepts simultaneously. Location information in *Martinsen* had to be deleted because it could denote different types of locations (i.e., where a violation occurred, or where a body was found, or where victims were born or lived) in contrast to the clear parsing of location information in the KMB enumeration (cf. Sections 4.8, 4.9, and 4.11).

- Missing or incorrect sex encodings. To remedy sex-missingness and inaccuracy in other sources, we used first name information to correct or fill in missing sex wherever possible. To obtain information on the sex of a victim based on their Serbian or Albanian name, Google, and Facebook searches were conducted.

- Incorrect date encodings. In particular, LDK has month and day information transposed in date fields. For example, LDK lists some victims with a violation date of December 4, 1999, when the relevant incident is known to have happened on April 12, 1999. We controlled for this date encoding issue with the help of an algorithm that compared all dates with both month-day versions to each other across all data sources.

- Missing field information that could not be filled in. Some data sources have entirely or partially missing information on individuals' ethnicity, date of birth, or date and place of violation. This provides insufficient individually identifying detail to successfully link records across data sources.

The process of record linkage across the available data sources was both automated using HRDAG's semi-supervised machine learning approach to matching,[17] as well as matches performed by hand.

After multiple rounds of automated matching, it became clear that the superior quality in the KMB database, combined with the high rates of missing information and inaccuracies in the other lists could not be reconciled in an automated matching process. The precision that is inherent in the KMB data but missing in other data collection projects created a machine learning problem because different degrees of data precision resulted in low matching recall and precision. For example, rules to link records of low information to obvious matches in the KMB list had to be very permissive, enabling many records to match to each other. This means such rules had to allow for considerable discrepancy

[17]The semi-automated matching approach that was undertaken here is very similar to the project described in Appendix A in Price, M., J. Klingner, and P. Ball (2013): 'Preliminary Statistical Analysis of Documentation of Killings in the Syrian Arab Republic.' The Benetech Human Rights Program, commissioned by the United Nations Office of the High Commissioner for Human Rights (OHCHR). January 2, 2013. https://hrdag.org/wp-content/uploads/2013/02/Benetech-final-SY-report.pdf.

in the way names were spelled, as well as allow for field missingness in non-KMB lists. Such permissive matching rules, however, were then inadequate to accurately differentiate between more precise victim records of very similar information that denote different people, such as is the case with members of the same family.[18]

As a result of the varying degrees of information quality across the different lists, the automated match process did not link every record from other lists to the 'master data.' Note that the master data were defined as the compilation of all records from the KMB register, i.e., all war, potential, and 'not war' victims, as well as all records from the closed lists of ICRC and OMPF that are not deaths. Records that could not be linked to the master data were denoted as 'singletons.' Upon inspection of the automated matching outcome, we suspected these singletons to represent *false negatives*, i.e., true match pairs that the matcher inaccurately denoted as non-match pairs. False negatives are the opposite of *true negatives*, i.e., true non-match pairs that the matcher correctly identifies as non-match pairs. Finding true negatives would suggest incompleteness of the KMB database. A result of false negatives suggests that the automated matching process underperforms, i.e., it 'under-matches.'

In a next step, we therefore proceeded with evaluating singletons by hand. In this hand-matching step, human analysts closely reviewed singleton records and compared them to the master data. Common name spelling and name errors were checked, while values in date and location fields were fully matching, which suggested that these records should be linked. Google and Facebook searches were performed twice to validate questionable names in singleton records. Singleton records often showed a combination of spelling variations paired with missingness in other relevant fields, such as information on the date of birth, place and date of a violation.

In a further step, we passed samples of singleton records that we were unable to link to the master data by automated record linkage or hand-matching to the HLC for a second round of hand-matching. HLC Belgrade staff was able to link nearly all of these records to the KMB register.[19] In this process, a search of the database's source and judgment layers was particularly helpful (see Chapter 3).[20] For example, some victims were listed by their maiden names in non-HLC lists, and these earlier names were found in the KMB's 'source' layer of information. The victims' true, married names were in the

[18]Semi-automated matching accurately identified approximately 300 duplicates within the KMB database. All duplicate records were merged as a result.

[19]Note that in this final round of hand-matching, the closed lists of ICRC and OMPF were not consulted.

[20]Jule Krüger was on-site with the HLC in November 2014 while part of the remaining singletons were hand-matched by KMB staff in this final round of record linkage.

database's judgment layer, and the preservation of all the information enabled the final linkage.

In Table 6.2, we provide a summary of the outcomes from our three-step record linkage process. In the first column, the respective data source under cross-comparison is listed. In the second column ('A-singletons'), we list the number of singletons that remained after HRDAG's automated matching approach. In the third column ('HRDAG (Sample)'), we denote the type and size of the sample selected from these singletons that was examined in a first round of hand-matching performed by HRDAG. The fourth column ('Matched (%)') denotes the outcome and match rate of this first hand-matching step in terms of the number of singletons that was successfully matched to the master data. In the fifth column ('H-singletons I'), we summarize the number of singletons that remained after this first round of hand-matching. The next column ('HLC (Sample)') gives the size and type of sample selected from these singletons to be examined by the HLC in a second hand-matching step. The final two columns ('Matched' and 'H-singletons II') summarize the outcome of this second hand-matching step in terms of the number of records HLC was and was not able to link to the KMB, respectively. Furthermore, we place data sources in the upper half of the table if KMB researchers had access to them, and in the lower half of the table if they did not have access to these victim enumerations (see Section 6.1 and Table 6.1 for details).

As can be seen in Table 6.2, the only singletons that remained after the third match step come from data sources that the HLC and HLC-Kosovo did not have access to (see the lower half of Table 6.2). No singletons remain for sources to which KMB researchers had full access (see the upper half of the table). This result shows that the KMB research team put forth a remarkably thorough effort at absorbing all records from accessible sources, without omitting a single record.

With regard to the remaining singletons from data sources that were not accessed by the HLC or HLC-Kosovo, we are not convinced that these represent plausible records still missing from the KMB database. For example, from the remaining ICMP singletons, one matches a witness (not a victim) who the KMB lists as wounded but not killed during the conflict.[21] Similarly, of the two HRW singletons, none have date of birth information, one record has several possible matches in the KMB database because it is imprecise; while another has a name which does not seem to be from the region, although the suggested incident is very well documented. The EXH singletons only provide name and sex information of victims.[22] Of the remaining OSCE singletons, none has date of birth

[21] The other ICMP record may be a potential victim that will require further research.

[22] For one EXH singleton, several KMB records could be matched as the first name is a nickname for

Table 6.2: Outcomes of three-step record linkage process, and singleton analysis.

Source	A-singletons[a]	HRDAG[b] (Sample)[c]	Matched	(%)	H-singletons I[d]	HLC (Sample)	Matched	H-singletons II[e]
ICRC	63	63 FE	58	92.1	5	5 FE	5	0
OMPF	125	100 RS	53	53	47	10 RS	10	0
OSMN	621	–	–	–	621	60 RS	60	0
MWHK	732	100 RS	64	64	36	10 RS	10	0
LDK	1269	100 RS	59	59	41	10 RS	10	0
ICMP	42	42 FE	39	92.6	3	3 FE	1	2
HRW	125	125 FE	67	53.6	58	10 RS	8	2
ABA	207	207 FE	117	56.5	90	15 RS	8	7
EXH	272	208 AS[f]	146	70.2	62	10 RS	7	3
OSCE	814	200 RS	111	55.5	89	15 RS	12	3

[a]Singletons remaining from automated matching.
[b]Number of singletons reviewed by HRDAG.
[c]The type of sample can be a full enumeration (FE), a random sample (RS), or an arbitrary sample (AS).
[d]Singletons resulting after HRDAG's first round of hand-matching.
[e]Singletons remaining after HLC's second round of hand-matching.
[f]A random sample of 100 EXH singletons was drawn, but an additional, arbitrary selection of 108 singleton records was also examined.

information, one has a missing violation date, and for one individual there are 13 victims in the relevant location with that last name, but none of the statements from families mention an individual with the particular first name. Of the ABA singletons, none has date of birth information, and one has a missing violation location. The ABA records are the hardest to match to the KMB due to their imprecision.[23]

In the process of matching using the semi-automated approach, as well as by hand, we linked every record that had sufficient information to the KMB database. There are still singletons for which we cannot determine whether they are contained in the KMB register. The imprecision and missing information that characterizes such records suggests that these singletons contain inadequate information. For example, the name information may not resemble any known regional names or possible spelling variations (i.e., no known names could be guessed from these records), suggesting wildly misspelled names. Or insufficient detail beyond the name is available, such as missing birth date, location, and date of an alleged violation. As a result, we do not believe that the remaining singleton records imply that the KMB is incomplete.

Similarly, the automated record linkage process was slightly inaccurate. Although the errors were small, the KMB database is very precise, so the errors became obvious. We found that the non-matching records often had many missing fields. We learned to be skeptical of records that do not link to the KMB war victim register because when we reviewed non-matching records, we found that they were not records missing from the KMB, as they seemed at first. Instead, they were either highly imprecise, with many missing fields; or they were inaccurate, and the KMB team reported that they were considered as not reliable.

As a result of our analysis of records that seemed not to appear in the KMB database—we did not find any singleton records with adequate information—we cannot analyze the overlap of the KMBD with the other data systems. We conclude that the KMBD is a superset of the other databases. The reported killings are either found in the KMBD, or rejected as insufficient or inaccurate. No pattern of systematic victim exclusion could be identified. Put differently, our findings from cross-comparing available data sources on victims in Kosovo suggest that the KMB database of war victims in Kosovo 1998–2000 by the HLC and HLC-Kosovo represents a nearly complete enumeration.

several different names.

[23] See Sections 5.1 and 6.4 for a related discussion on the number of sources for potential victims.

6.4 Potential victims

In this section, we consider whether the amount of remaining 1,603 potential victims suggests that the KMB database's war victims register is still incomplete.

As outlined in Section 5.1 and shown in Table 5.1, the majority of potential victims (1,151 individuals, or 71.8% of all potential victims) have only 1 source reporting on them. This is quite different from the confirmed war victims, for whom on average 8 source documents are available.

Furthermore, there is substantial uncertainty surrounding the potential victims. For 706 potential victims (44%) the date of violation is January 1, 1999. Given the KMB team's date coding practice (see Section 4.7), this date means that for most if not all of these victims, the exact date of the violation during the year of 1999 is unknown. Similarly, for the place of violation, for 408 potential victims (25.5%) the place of the violation is "Kosovë/Kosovo," meaning the violation is supposed to have happened in Kosovo but the exact location, i.e., the municipality, town, or village, is unknown. For 250 potential victims (15.6%), both the date and location information is jointly missing (note that these are contained in the previous counts of missing dates and missing locations).

According to the KMB's protocol for record verification, in order for an individual to be included in the war victim register, sufficient reliable evidence must exist to establish a connection to the war. During a visit to the HLC Belgrade office in November 2014, together with a KMB analyst, we reviewed eight potential cases on whom seven or eight sources are reporting (see Table 5.1). This review showed that all of the relevant sources merely 'list' these victims, without providing details on the war connection. At HLC-Kosovo, three to four staff are currently working on the verification of victims that are listed in the potentials register by way of collecting more statements from relatives or witnesses in the field.

In one data review pass, KMB researchers reviewed 36 of the 1,603 potential victims analyzed in this report (see Table 4.1). Of these 36 potential victims, only two (2) were confirmed as additional war victims. Of the remainder, 17 were determined to be 'not war' victims, and 17 records were linked to existing war victim dossiers by way of correcting name or other field information. It is impossible to know whether this laborious, meticulous effort will be representative of the remaining 1,567 potential victims. However, it is unlikely that the remaining potential victims will be radically different from this sample. This is another reason—in addition to the low number of reporting sources, uncertain information, as well as missing information on circumstances—to be confident that the KMB is nearly complete.

The amount of uncertainty associated with potential victims shows that the KMB team have reached a point where we are primarily left with records that are essentially rumors. It is possible that there are several tens, perhaps one hundred victims that are still missing from the KMB's list of confirmed war victims. Unless more social knowledge becomes available, however, it may be impossible to ever resolve these potential cases and establish whether they represent true war victims.

Chapter 7

Conclusions and recommendations

The most important conclusion from our analysis is that the Kosovo Memory Book database documents all or nearly all the human losses during conflict in Kosovo during the period 1998–2000. By "nearly all," we mean that in our opinion, it is very unlikely that there are more than a few tens of undocumented deaths. This conclusion is based on several analyses and findings, including a comparison with ten other databases in which no new records were found; a statistical analysis in which the KMBD was found to have more records than any other database in every period and for each municipality; and a companion analysis by Professor Michael Spagat which shows that the KMB is consistent with two independent probability-based estimates of the total human losses.

We believe that there is a larger lesson in the final records of 'potential victims' in the KMBD. Even with work of dozens of researchers and analysts spanning more than a decade, it is probably impossible to create an exact, precise list which includes only the true war victims, and at the same time all of the true war victims. The stock of social knowledge about the past simply does not extend to every single one of the events in which victims suffered. However, it is indeed possible to get very close to an exact list. Slight imprecisions in the reporting do not impede our analysis of the larger patterns in the conflict. The KMBD is a rich source for historical memory and other transitional justice reflections, for statistical analysis of the conflict, and most importantly, to be a permanent record of the names and lives of the victims lost in the conflict.

7.1 Recommendations

We have a small number of technical recommendations for the HLC and HLC-Kosovo:

- We recommend that KMB staff differentiate dates between missing information and the first day of the month and first month of the year (see Sections 4.7 and Section 4.10). Missing days are currently recorded as '1,' and missing months are recorded as '1.' These should have some other value (e.g,. -1 or NA).
- Name information should be further standardized, as noted in Section 4.3.
- The current ethnicity field should be divided into ethnicity and citizenship, as noted in Section 4.6.
- We recommend that the HLC and HLC-Kosovo reconsider the determination of 'not war' victims who have only one source document when a victim was not found alive. Perhaps these victims could be maintained on the potential victim list until more information becomes available that allows staff to confirm or reject the connection with the war in Kosovo.
- Correct information on date of origin, fill in missing information on source language. See Section 5.2 for more discussion.

These are relatively minor improvements and standardizations, and they will contribute to improving an already excellent database.

7.2 Directions for future research

There are a number of interesting directions for possible future research for this project:

- To systematically research and code the circumstances of the death. The database currently only records the type of violation, i.e., death or disappearance. A future version of the data could provide systematic information on the cause of death, such as the weapon used.
- To find information about the perpetrator, i.e., in terms of the responsible armed formation (see Table 4.4), of each killing would be enormously beneficial for subsequent researchers' use of this data, in particular for transitional justice purposes.
- Another related area would be to systematically document the movement of armed forces and the use of airstrikes in the conflict.

- To group victims into events, linking victims who died and disappeared within the same event, to examine the number of victims per event. The creation of events would transform the database from a list of victims to a narrative of events, a step toward a definitive, comprehensive history of the conflict.

- With a much larger project, to document the non-lethal violations that preceded each lethal violation, including threats, imprisonment, torture, sexual assaults, the destruction of property, and other violence. These violations would need the same details about the date and location of the violence, and the perpetrators. Information about the non-lethal violations is another step toward a comprehensive history of the conflict, and it would enable researchers to understand repertoires and trajectories of violence.

We congratulate the HLC and HLC-Kosovo on an extraordinary and remarkable project. Few conflicts have received the sustained and professional attention that the HLC and HLC-Kosovo have given to the human losses in Kosovo 1998–2000. The world benefits from this knowledge. Above all, we acknowledge the victims who will now always be remembered.

About HRDAG

The Human Rights Data Analysis Group is a non-profit, non-partisan organization[1] that applies scientific methods to the analysis of human rights violations around the world. This work began in 1991 when Patrick Ball began developing databases for human rights groups in El Salvador. HRDAG grew at the American Association for the Advancement of Science from 1994–2003, and at the Benetech Initiative from 2003–2013. In February 2013, HRDAG became an independent organization based in San Francisco, California; contact details and more information are available on HRDAG's website and Facebook page.

HRDAG is composed of applied and mathematical statisticians, computer scientists, demographers, and social scientists. HRDAG supports the protections established in the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and other international human rights treaties and instruments. HRDAG scientists provide unbiased, scientific results to human rights advocates to clarify human rights violence. The human rights movement is sometimes described as "speaking truth to power:" HRDAG believes that statistics about violence need to be as true as possible, with the best possible data and science.

HRDAG Field Consultant Jule Krüger conducted most of the meetings and interviews with the HLC, and did most of the data analysis and writing. HRDAG Executive Director Patrick Ball reviewed the database design and conceptual issues with the database, and assisted with data analysis, writing, and editing the report. HRDAG Data Management Consultant Michelle Dukich did the hand matching, helped with the cleaning and canonicalization of name, sex, and location fields, and transcribed the *Martinsen* data. In her work, she was assisted by HRDAG Research Assistant Christopher Dukich. HRDAG Computer Engineering Consultant Scott Weikart assisted with HRDAG's automated record linkage. HRDAG Communications Consultant Christine Grillo assisted with editing the report. HRDAG is very grateful to HLC's Founder, Nataša Kandić, and to KMB's Predrag Miletic for answering countless questions. We are further grateful to Tim Thomay for his assistance in weaving long tables from *R* into LaTeX.

[1] Formally, HRDAG is a fiscally sponsored project of Community Partners.

missions beyond the scope of this license may be available at https://hrdag.org.