Clustering suicides: A data-driven, exploratory machine learning approach

Birgit Ludwig[a,e], Daniel König[b], Nestor D. Kapusta[c,*], Victor Blüml[c], Georg Dorffner[d], Benjamin Vyssoki[b]

[a] Clinical Division of General Psychiatry, Department of Psychiatry and Psychotherapy, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria
[b] Clinical Division of Social Psychiatry, Department of Psychiatry and Psychotherapy, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria
[c] Department of Psychoanalysis and Psychotherapy, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria
[d] Section for Artificial Intelligence and Decision Support, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria
[e] Department of Neurology, Medical University of Vienna, Währinger Gürtel 18-20, 1090 Vienna, Austria

ARTICLE INFO

ABSTRACT

Methods of suicide have received considerable attention in suicide research. The common approach to differentiate methods of suicide is the classification into "violent" versus "non-violent" method. Interestingly, since the proposition of this dichotomous differentiation, no further efforts have been made to question the validity of such a classification of suicides. This study aimed to challenge the traditional separation into "violent" and "non-violent" suicides by generating a cluster analysis with a data-driven, machine learning approach. In a retrospective analysis, data on all officially confirmed suicides (N = 77,894) in Austria between 1970 and 2016 were assessed. Based on a defined distance metric between distributions of suicides over age group and month of the year, a standard hierarchical clustering method was performed with the five most frequent suicide methods. In cluster analysis, poisoning emerged as distinct from all other methods – both in the entire sample as well as in the male subsample. Violent suicides could be further divided into sub-clusters: hanging, shooting, and drowning on the one hand and jumping on the other hand. In the female sample, two different clusters were revealed – hanging and drowning on the one hand and jumping, poisoning, and shooting on the other. Our data-driven results in this large epidemiological study confirmed the traditional dichotomization of suicide methods into "violent" and "non-violent" methods, but on closer inspection "violent methods" can be further divided into sub-clusters and a different cluster pattern could be identified for women, requiring further research to support these refined suicide phenotypes.

© 2019 Published by Elsevier Masson SAS.

1. Introduction

Suicide is a major public health issue accounting for over one million deaths per year making it the tenth leading cause of death worldwide [1] and the leading cause of preventable death in the elderly [2]. In 2016 the annual global age-standardized suicide rate of 10.5 per 100,000 population. Despite some progress and extensive research in the field of suicidology, predicting attempts and deaths by suicide still remains a challenge – both in a scientifically controlled setting and in clinical reality [3]. The lack of predictability makes preventive action difficult – both in clinical mental health care settings with individual patients as well as at aggregated levels for risk groups and entire populations. Additionally, the lethality of suicide methods differs greatly, which leads to restriction of lethal or violent means of suicide as the most recommended preventive measure [4–8].

In Austria, suicide rates have been steadily declining since the mid-80 s, from 1990 to 2000 the annual rate was 21.3 per 100,000 persons and from 2000 to 2010 it further decreased to 16.9. The only age-group showing increasing rates were males aged over 80. Women showed a more marked decrease than men [9,10].

Methods of suicide have received considerable attention in suicide research [11] especially regarding prevention and mean's restriction [7,8]. The most common approach to differentiate methods of suicide is the classification into "violent" versus "non-violent" methods, and is commonly referred to as Asberg's criteria

since 1976 [12]. According to these, suicide attempts are considered "violent" when the method of choice is any other than poisoning. Nevertheless, classifications of suicide methods are documented earlier in time, as one of the first publications to explicitly define suicide methods (non-violent methods as poisoning by solid, liquid substances or gas as non-violent and hanging, strangulation, shooting, cutting, piercing instruments and jumping from a high place as violent methods) was Burvill et al.'s [13] study about Australian immigrants and suicide.

Interestingly, since the proposition of this dichotomous differentiation into "violent" versus "non-violent" methods, no further efforts have been made to question the validity of such a classification of suicides. Furthermore, the use of this dichotomous classification is also ambiguous, since various authors subsume different methods under the two categories in current research [14,15].

Several possible limitations of such a classification of suicide should be considered. First, the classification was constructed intuitively and was not based on empirical data analyses. Further, the clinical benefit of this classification is limited as most suicide methods are considered "violent". There is also evidence suggesting that violent suicide attempters constitute a very heterogeneous group in terms of suicide intent [16]. Precise phenotyping is also highly relevant for neurobiological research. One of the proposed explanations for the lack of established biomarkers in psychiatric disorders is the theory that current clinical concepts of psychiatric disorders do not precisely reflect the neurobiological reality [17]. Thus, questioning established psychiatric concepts and proposing novel approaches based on epidemiological data might also move us closer to identify possible neurobiological correlates.

1.1. Aims

In the present study we aimed to challenge the dichotomous classification of "violent" vs. "non-violent" suicides by generating a cluster analysis with a data-driven, machine learning approach. The cluster analysis presented herein aimed to explore novel clusters of suicides.

2. Material and methods

In preparation for this study a literature research on Pubmed was undertaken identifying peer-reviewed publications from January, 1976 to December, 2018 by use of the terms "suicide clusters", "suicide methods", "cluster* suicide methods". Additionally, based on the revealed findings, a more specific literature search with the terms "violent suicide", and "non-violent suicide" was carried out. Articles published in English, French, and German were included. The results of this search were critically examined, culminating in the conclusion that a machine-learning analysis of suicide methods is required to generate data-driven, empirical evidence.

Suicide mortality data including month of death, sex, age, and method of suicide were provided by Statistics Austria for the period of January 1, 1970 to December 31, 2016. Suicide methods were classified into five categories using the *International Statistical Classification of Diseases, 10th Revision*: "Intentional Self-Poisoning" (X60-X69) hereinafter referred to as *poisoning*, "Intentional self-harm by hanging, strangulation, and suffocation" (X70) referred to as *hanging*, "Intentional self-harm by drowning, and submersion "referred to as *drowning*), "Intentional self-harm by handgun discharge" /" Intentional self-harm by rifle, shotgun, and larger firearm discharge" /" Intentional self-harm by other, and unspecified firearm discharge" (X72-X74) referred to as *shooting*, and "Intentional self-harm by jumping from a high place"/ Intentional self-harm by jumping or lying before moving object" (X80 and X81) referred to as *jumping,* and other methods (X75-X79, X82-X84).

Suicide rates were calculated for 20 age groups in 5-year-intervals. The time of the year was recorded as the month in which the suicide occurred. Sex-specific subanalysis was undertaken as previous research suggested significant differences in the distribution of data [18].

To perform clustering, we considered age, and month as spanning a 20 by 12 matrix of absolute and relative frequencies of suicide, for all persons irrespective of sex, as well as for a stratified sample for males and females separately. Calculating values for each of the 5 main suicide methods separately, form distributions of suicide frequencies, which were compared to each other in terms of similarity. As a similarity (or distance) measure between two such distributions we used the χ^2-value over 20 by 12 values, as it would be used for testing whether two such empirical value matrices originate from the same distribution (test for homogeneity) – see supplementary data.

Calculating such matrices for each of the 5 main suicide methods separately leads to distributions of suicide frequencies, which - by using the *dendrogram* function of Matlab (Mathworks, USA) - resulted in hierarchies depicting potential clusters and their distance between each other. For comparisons between clusters or subclusters the differences were re-calculated for the respective distributions of entire subclusters. Given roughly 240 degrees of freedom, any such χ^2-value > 300 would indicate a significant difference between clusters on a significance level of 0.005 (corrected for multiple testing).

Informed consent and institutional review board approval were not required, since the data is open to the public and is being provided in anonymized form by Statistics Austria.

3. Results

3.1. Descriptive data

A total of 77,894 suicides were registered in Austria between January 1, 1970, and December 31, 2016. Of these 56,091 (72%) were male and 21,803 (28%) were female. The median age group for men committing suicide was 50–54 and the median age group of women was 55–59 (see Supplemental Data 2 for graphic depiction). Grouped by suicide method used, 36,053 (46.3%) were due to hanging, 3876 (5%) due to drowning, 11,849 (15.2%) due to shooting, 7188 (9.2%) due to jumping, and 11,968 (15.4%) due to poisoning. A total of 6960 (8.9%) suicides were due to other methods, but excluded from the present analysis because of their heterogeneity and relative low occurrence frequency.

3.2. Clustering data

3.2.1. Total sample
The analysis of our sample revealed two main clusters with poisonings as the first cluster (A_{Total}) and hanging, shooting, drowning, and jumping in the second cluster (B_{Total}) ($\chi^2 = 1241$, p < 0.00001). The second main cluster was further divided into two more subclusters including jumping as the first subcluster, closest to poisoning, and hanging, shooting, and drowning as the second subcluster ($\chi^2 = 648$, p < 0.00001). Thereof, hanging and shooting shared the most similarities within the subcluster, hence they were grouped into a further subcluster division together ($\chi^2 = 510$, p < 0.00001) (Fig. 1).

3.2.2. Male sample
Clusters in the male sample were similar to the total sample. The provided demographic data revealed two main clusters with
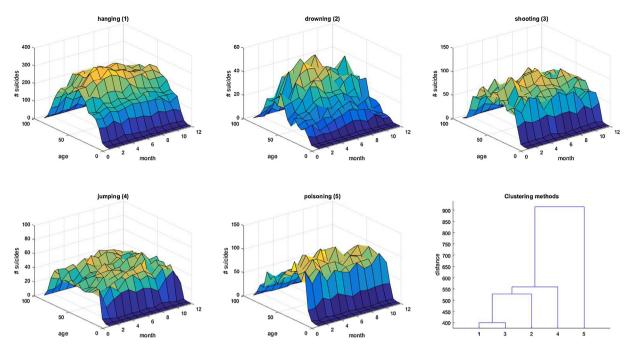
Fig. 1. Cluster analysis of the total sample: Distributions of suicides (both sexes) by month ad age group, plotted as 3d surfaces, and the resulting clustering based on similarities between the distributions.

poisoning as the first cluster (A_{Male}) and hanging, shooting, drowning, and jumping in the second cluster (B_{Male}) ($\chi^2 = 1595$, p < 0.00001). The second main cluster was divided into two more subclusters containing jumping as the first subcluster and hanging, drowning, and shooting and as the second subcluster ($\chi^2 = 543$, p < 0.00001). Therein, hanging and drowning share the most similarities, establishing another subcluster division together ($\chi^2 = 304$, p = 0.0028) (Fig. 2).

3.2.3. Female sample

Two main clusters were identified: hanging and drowning as the first main cluster (A_{Female}) and shooting, poisoning and jumping as the second main cluster (B_{Female}) ($\chi^2 = 732$, p < 0.00001). The second main cluster was further divided into two subclusters: shooting and poisoning shared more similarities establishing the first subcluster and jumping establishing the second subcluster ($\chi^2 = 327$, p = 0.00013) (Fig. 3).
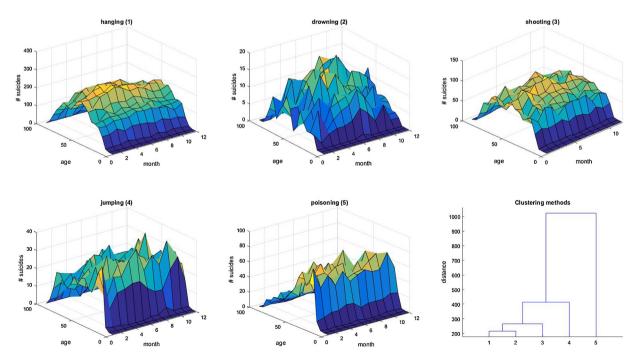

Fig. 2. Cluster analysis of the male sample: Distributions of male suicides by month ad age group, plotted as 3d surfaces, and the resulting clustering based on similarities between the distributions.
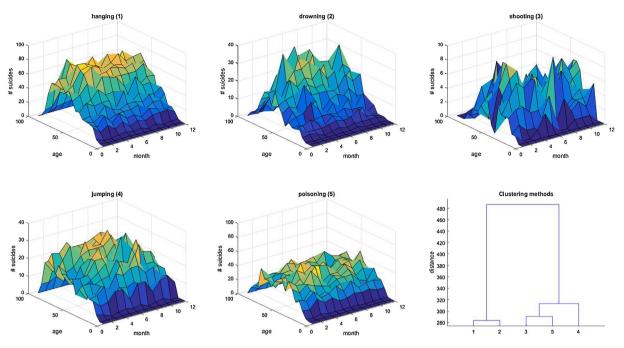
Fig. 3. Cluster analysis of the female sample: Distributions of female suicides by month ad age group, plotted as 3d surfaces, and the resulting clustering based on similarities between the distributions.

4. Discussion

The aim of this data-driven cluster analysis was to find empirical evidence for a classification system of suicide methods. Since the most referenced classification concepts of "violent" versus "non-violent" suicide methods (with poisoning as the only "non-violent" method and all other methods as "violent" methods) had been conceptualized in the 20th century – not based on any empirical data [12,13], a data-driven machine learning approach might be a novel step to refine the concept. Interestingly, the cluster analysis of the total sample supported this dichotomous concept, positioning poisoning as a main cluster opposed to all other methods, just as in the male subsample.

However, when interpreting the data of the total sample in more detail, violent suicide methods could further be divided into subgroups. We first of all acknowledged that the subcluster of hanging and shooting showed fewer in-between groups differences ($\chi^2 = 401$, used as a distance measure) as compared to the other methods (average $\chi^2 = 817$).

From a clinical point of view hanging and shooting mostly take place in private spaces, as opposed to the other methods and they are considered some of the most lethal among suicide methods [19]. Further, both methods require instruments as well as preparation and planning. In this regard this cluster might resemble Lopez-Castroman et al.'s [20] *Well-planned*, Kim et al.'s [21] *planned* and Rapeli et al.'s [22] *definite* categories. However, these defined categories were derived from data on suicide attempters, not from deaths by suicide as it was the case in our study. Furthermore, these studies examined a relatively small number of study participants, diminishing their individual power and thus the reproducibility and validity of the results. When considering the risk of dying by suicide in future attempts, hanging ranks first among all other methods; a previous attempt with hanging is the strongest predictor for a death by suicide in the future [6,23].

On the next level of clustering, a hanging-shooting-drowning subcluster was revealed, in opposition to the second subcluster jumping. Again, one could argue that jumping is different from

hanging, shooting, and drowning in the sense that this method does not require much preparation or any instruments to complete a suicide attempt.

Our results question the definitions of "violent suicide" by Dumais et al. [15] and by Bradvik et al. [14] where drowning and poisoning were both defined as "non-violent" methods. In contrast, in our more detailed analysis of available data these methods were clearly associated with other methods previously defined as "violent" and shared clusters – depending on depth of analysis – with hanging, shooting or jumping.

The male subsample shared many similarities with the total sample. This may also be seen as indicative of the gender gap of completed suicides, as men represent 72% of the total numbers of suicide in the national death registry. Analysis of the distribution of poisoning revealed a ramp-like form with a focus on young men without any seasonal effects. These findings comply with Rapeli et al's [22] *marked intent* cluster with predominantly men attempting suicide by poisoning with marked intent to die [22].

Analysis of the female subsample revealed contrasting results. Hanging and drowning, being in a main cluster, showed a very similar distribution: there was a tendency towards hanging and drowning in elderly women during the later months of the year. Although shooting and poisoning are generally being considered as residing on opposite ends of the violence spectrum, they shared a subcluster in our analysis. The distribution of shooting in the female subsample may not be as representative as the other distributions of methods, because of the relatively small percentage of female suicides by shooting. This might be reflected by the heterogenous, erratic appearance of the three-dimensional figure, representing the distribution of shooting (Fig. 3).

The clinical value of our approach is primarily a typological one. With more elaborated clusters, types of sucide behavior could be better distinguished for clinical purposes. The herein generated knowledge on method choices in addition to specific risk factors might help to develop preventive interventions e.g. treatment procedures for certain patients with a strong emphasis on prescribed medication (to avoid poisonings) or an emphasis on

firearm availability (to avoid shootings). One simple clinical recommendation could be to routinely screen male patients with suicide ideation for firearm possession. Definitely, there is clinical potential for a more elaborated suicide typology, given our limited variables encorporated in the model. Further studies on different comorbidities, or genetic analyses and generally prospective longitudinal studies would be needed to foster this idea.

General limitations of this study include the lack of clinical variables such as intent to die, onset of suicidality, psychopharmacological status, associated psychiatric disorder, isolation, and other demographic variables. Furthermore, the herein presented results might not applicable to countries with differing in variables such as climate, firearm regulations and socio-economic factors. Since not all methods of suicide were included in the cluster analysis due to the lack of sufficiently high numbers, further research is warranted exploring the clustering of suicide methods with additional population data and by including additional less frequent suicide methods.

5. Conclusions

Our data-driven results in this epidemiological study suggest that – based on variables such as sex, age, month of death – suicide methods can be merged into clusters and subclusters. The dichotomous classification of violent vs. non-violent methods could be empirically confirmed in the total sample and in men. However, in our machine learning approach, additional patterns could be identified: the violent cluster showed hanging, shooting, and drowning in one subcluster and jumping as an independent violent subcluster. Furthermore, our results suggest that the traditional concept of violent vs. non-violent suicide methods does not apply to the female sub-sample. Thus, sex differences in suicide research need to be considered systematically.

These findings need to be replicated with clinical variables such as onset of suicidality, severity, and progression of the underlying psychiatric disorder among others, but they hold great potential to advance neurobiological research by allowing to coin new, potentially more precise suicide phenotypes.

References

[1] Turecki G, Brent DA. Suicide and suicidal behaviour. Lancet 2016;387:1227–39.
[2] Boxwell AO. Geriatric suicide: the preventable death. Nurse Pract 1988;13(10–15):8–9.
[3] Yates K, Lang U, Cederlof M, Boland F, Taylor P, Cannon M, et al. Association of psychotic experiences with subsequent risk of suicidal ideation, suicide attempts, and suicide deaths: a systematic review and meta-analysis of longitudinal population studies. JAMA Psychiatry 2019;76:180–9.
[4] Kolves K, McDonough M, Crompton D, de Leo D. Choice of a suicide method: trends and characteristics. Psychiatry Res 2018;260:67–74.
[5] Pirkola S, Isometsa E, Lonnqvist J. Do means matter? Differences in characteristics of Finnish suicide completers using different methods. J Nerv Ment Dis 2003;191:745–50.
[6] Runeson B, Tidemalm D, Dahlin M, Lichtenstein P, Langstrom N. Method of attempted suicide as predictor of subsequent successful suicide: national long term cohort study. BMJ 2010;341:c3222.
[7] Mann JJ, Apter A, Bertolote J, Beautrais A, Currier D, Haas A, et al. Suicide prevention strategies: a systematic review. JAMA 2005;294:2064–74.
[8] Zalsman G, Hawton K, Wasserman D, van Heeringen K, Arensman E, Sarchiapone M, et al. Suicide prevention strategies revisited: 10-year systematic review. Lancet Psychiatry 2016;3:646–59.
[9] Etzersdorfer E, Voracek M, Kapusta N, Sonneck G. Epidemiology of suicide in Austria 1990-2000: general decrease, but increased suicide risk for old men. Wien Klin Wochenschr 2005;117:31–5.
[10] Etzersdorfer E, Klein J, Baus N, Sonneck G, Kapusta ND. Epidemiology of suicide in Austria during 2000–2010: potential years of life lost: time for the national suicide prevention program. Wien Klin Wochenschr 2015;127:308–13.
[11] Varnik A, Kolves K, van der Feltz-Cornelis CM, Marusic A, Oskarsson H, Palmer A, et al. Suicide methods in Europe: a gender-specific analysis of countries participating in the "European Alliance Against Depression. J Epidemiol Community Health 2008;62:545–51.
[12] Asberg M, Traskman L, Thoren P. 5-HIAA in the cerebrospinal fluid. A biochemical suicide predictor? Arch Gen Psychiatry 1976;33:1193–7.
[13] Burvill PW, McCall MG, Reid TA, Stenhouse NS. Methods of suicide of english and welsh immigrants in Australia. Br J Psychiatry 1973;123:285–94.
[14] Bradvik L. Violent and nonviolent methods of suicide: different patterns may be found in men and women with severe depression. Arch Suicide Res 2007;11:255–64.
[15] Dumais A, Lesage AD, Lalovic A, Seguin M, Tousignant M, Chawky N, et al. Is violent method of suicide a behavioral marker of lifetime aggression? Am J Psychiatry 2005;162:1375–8.
[16] Astruc B, Torres S, Jollant F, Jean-Baptiste S, Castelnau D, Malafosse A, et al. A history of major depressive disorder influences intent to die in violent suicide attempters. J Clin Psychiatry 2004;65:690–5.
[17] Mann JJ. Neurobiology of suicidal behaviour. Nat Rev Neurosci 2003;4:819–28.
[18] Callanan VJ, Davis MS. Gender differences in suicide methods. Soc Psychiatry Psychiatr Epidemiol 2012;47:857–69.
[19] Elnour AA, Harrison J. Lethality of suicide methods. Inj Prev 2008;14:39–45.
[20] Lopez-Castroman J, Nogue E, Guillaume S, Picot MC, Courtet P. Clustering suicide attempters: impulsive-ambivalent, well-planned, or frequent. J Clin Psychiatry 2016;77:e711–8.
[21] Kim H, Kim B, Kim SH, Park CHK, Kim EY, Ahn YM. Classification of attempted suicide by cluster analysis: a study of 888 suicide attempters presenting to the emergency department. J Affect Disord 2018;235:184–90.
[22] Rapeli CB, Botega NJ. Clinical profiles of serious suicide attempters consecutively admitted to a university-based hospital: a cluster analysis study. Rev Bras Psiquiatr 2005;27:285–9.
[23] Stenbacka M, Jokinen J. Violent and non-violent methods of attempted and completed suicide in Swedish young men: the role of early risk factors. BMC Psychiatry 2015;15:196.