



Why do people call crisis helplines? Identifying taxonomies of presenting reasons and discovering associations between these reasons

Journal:	<i>Health Informatics Journal</i>
Manuscript ID	HIJ-18-0429.R2
Manuscript Type:	Original Article
Keywords:	Telephone Crisis Helplines, Suicide, Association Rule Mining, Data Analytics, Computer Science, Public Health, ICD-10, Mental health and well-being helpline, Help-seeking behaviour, Telephony analysis
Abstract:	<p>Objective: To identify the main motivating reasons for contacting a crisis helpline through analysing a large call log data set.</p> <p>Method: Two taxonomies were identified within the call log data from a telephone crisis helpline, categorizing the cited reason for each call. One taxonomy categorized the reasons at a granular level; the other taxonomy used the International Classification of Diseases (ICD-10). Exploratory/confirmatory data analytics were applied to discover insights as to why individuals make contact. Association rule mining identified associations between presenting reasons for client's calls. Risk ratings of calls were compared to examine associations between presenting issue and of risk of suicide. Reasons for contacting the service were assessed across geolocations.</p> <p>Results: Calls with reasons relating to suicide are the most cited reasons for contacting Lifeline and a prominent feature of the discovered association rules. There were significant differences between reasons in both taxonomies concerning risk ratings. Reasons that were most and least consistently cited across geolocations were discovered.</p> <p>Conclusion: Employing two differing taxonomy approaches to analyse call log data reveal prevalence of main presenting reasons for contacting a crisis helpline. Association rule mining using each taxonomy provided insights into the associations between presenting reasons.</p>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Why do people call crisis helplines? Identifying Taxonomies of Presenting Reasons and Discovering Associations Between These Reasons

Abstract

Objective: To identify the most common reasons for contacting a crisis helpline through analysing a large call log data set.

Method: Two taxonomies were identified within the call log data from a Northern Ireland telephone crisis helpline (Lifeline), categorizing the cited reason for each call. One taxonomy categorized the reasons at a fine granular level; the other taxonomy used the relatively coarser International Classification of Diseases (ICD-10). Exploratory data analytic techniques were applied to discover insights as to why individuals contact crisis helplines. Risk ratings of calls were also compared to assess the associations between presenting issue and of risk of suicide as assessed. Reasons for contacting the service were assessed across geolocations. Association rule mining was used to identify associations between the presenting reasons for client's calls.

Results: Both taxonomies show that calls with reasons relating to suicide are the most common reasons for contacting Lifeline and were a prominent feature of the discovered association rules. There were significant differences between reasons in both taxonomies concerning risk ratings. Reasons for calling helplines that are associated with higher risk ratings include those calling with a personality disorder, mental disorders, delusional disorders and drugs (legal).

Conclusion: Employing two differing taxonomy approaches to analyse call log data reveals the prevalence of main presenting reasons for contacting a crisis helpline. Association rule mining using each taxonomy provided insights into the associations between presenting reasons. Practical and research applications are discussed.

Keywords

Association Rule Mining, Call Data Analytics, Mental Health, Suicide, Telephone Crisis Helplines

1. Introduction

Many studies have examined the differing explanations as to why people contact crisis helplines. Individuals may contact crisis helplines to receive social support ^(1,2), to discuss general mental health issues ^(2,3,4), or contact the service on behalf of someone else. Different reasons will influence the type and degree of response that a crisis helpline provides to that caller. The reasons that clients cite may be recorded by the telephone

counsellor, based on the counsellor's subjective interpretation of the call's content and the caller's needs, which may give rise to potential misinterpretation of the nature of the client's current state and level of distress. Callers may cite numerous different reasons within any one phone call, but the crisis counsellors will only record one reason and not all, which will depend on their perception for the reason for calling or their ability to recall the key aspects of the call itself. Lastly, inconsistent recording of these reasons can occur if each main presenting reason is recorded manually by the telephone counsellor, making it difficult to determine the true range of possible reasons for each call within an entire call log. Previous research has identified that telephone counsellors from crisis helplines may struggle to interpret the caller's needs and what type of help the callers require ^(3,5,6).

Call logs provide data which can be analysed to uncover insights about callers' call behaviour and presenting needs, however the inconsistent recording of the reasons for each call can limit the extent to which insights can be achieved through data analytic processes. Attempts to predict future caller behaviour through established machine learning techniques become more difficult as potential reasons arise. Also, there are numerous reasons for calling a helpline that are relatively uncommon, and reported less frequently, thereby hindering the process of building the types of predictive model which would help services better meet callers' needs, because of fewer observations available to train a predictive classifier ⁽⁷⁾.

The objective of this paper is to use two taxonomy approaches to aid the analysis of call log data, specifically the presenting reasons for calling helplines. One taxonomy approach will be a Granular Taxonomy (GT), which contains highly specific reasons identified from the data with duplicate or near-duplicate presenting reasons coalesced. The other taxonomy approach uses the *International Classification of Diseases and Related Health Problems, 10th Revision* ⁽⁸⁾, providing a means to classify the reasons at a broader level. A description of the process for developing the taxonomies is discussed later in this paper.

This paper therefore addresses the following research questions: what are the most prevalent reasons for contacting a crisis helpline, and which are related? Are there differences in risk of suicide between different reasons and if so, which reasons are considered more relevant in relation to risk of suicide? Are the proportions of each reason consistent across geolocations? The following sections provide more information on the Lifeline service and methodological approach for data analytics.

2. Methods

The Lifeline helpline

Lifeline is Northern Ireland's regional telephone crisis helpline which has been in operation since 2008 and is part of Northern Ireland's Protect Life strategy for the

1
2
3 prevention of suicide ⁽⁹⁾, operating 24-hours a day, 7 days a week. Lifeline has provided
4 counselling to over 51,000 clients and has engaged in over 1.24 million phone calls since
5 inception. Staff at Lifeline make and receive calls from clients (50.4% of calls) or to a
6 third party in relation to the client's case (i.e., ambulance/police service, charities; 25.9%
7 of calls), while other calls may be made incorrectly to the service (i.e. hang-ups, silent;
8 23.7% of calls).
9

10
11 Lifeline is operated by qualified counsellors with training in risk management, risk
12 identification and referral, who receive regular training from agencies which deal with
13 specific issues (i.e., homelessness, mental health). All Lifeline counsellors are British
14 Association for Counselling and Psychotherapy (BACP)/ Irish Association for
15 Counselling and Psychotherapy (IACP) registered (or working towards) and had received
16 suicide prevention training and in-house training on procedures that include risk
17 assessment and identification of main presenting reasons. Upon initial contact with the
18 service by a client, current and historical information regarding known risk factors is
19 collected by the Lifeline counsellor, which allows for the counsellor to formulate an
20 appropriate intervention for the individual in crisis. The known risk factors for suicide
21 may be demographic variables, adverse life events, current or previous physical/mental
22 health status and patterns of service use by the client. ⁽¹⁰⁾For more information on the risk
23 factors associated with suicide and the identification of these risk factors within
24 counselling practice, please refer to BACP fact sheet *Working with suicidal clients in the*
25 *counselling professions* ⁽¹¹⁾.
26
27
28

29
30 Lifeline is free for public use, and following initial contact, clients may receive face-to-
31 face counselling sessions. Arrangements can be made for Lifeline to contact clients on a
32 regular basis to check in on the client and to assess if the client has improved from their
33 previous interaction from the service.
34

35 Lifeline data

36
37
38 Call data was recorded on Lifeline's Information Management System for all callers who
39 have interacted with the service. There are many features within the call data, but three
40 features were used in this study:
41

- 42
43 • Main presenting issue: The reason for a client contacting the service is recorded.
44 Some examples could be that the caller is depressed, experiencing financial
45 difficulties, exhibiting suicidal behaviour, etc.
- 46
47 • Risk Rating: Lifeline counsellors conduct a risk assessment on each of their clients.
48 This formulation balances suicide risk factors against protective risk factors within a
49 client's call counselling session. This risk factor is presented as number between 1 to
50 10, with 0 representing missing values, 1-3 representing a low, 4-6 medium to
51 moderate risk of suicide and 7-10 representing a moderate to severe risk of suicide.
52 Rating of 0 have been removed. This rating is assigned by the telephone counsellor,
53 who has specialized training in risk management ⁽¹¹⁾.
- 54
55 • Area in which the caller is from (by Electoral Ward).
56
57
58
59
60

1
2
3
4
5
6 The remaining features in the data that were not used are considered as basic log data
7 features such as timestamps, duration of call, call/event identifier and client identifier.
8 Each record within the call log data has a unique identifier code for each client which is
9 based on the client's phone number. This allows for the client to remain anonymous but
10 still allows crisis helplines to keep trace. Both inbound calls (calls made by the client to
11 the service) and outbound calls (calls made by the service to the client) were used for the
12 data analysis. The call data was subjected to data wrangling processes, making it
13 amenable to user data analytics and machine learning processes which are demonstrated
14 in this paper ⁽¹²⁾.
15

16 17 Data analysis

18
19 R programming language and R Studio (version 3.4.2) was used for the data wrangling
20 process and statistical analyses ⁽¹³⁾. The Rcmdr library was used to conduct statistical
21 analyses while the ggplot2 library was used to create data visualisations. The arules
22 package was used for association rule mining analysis.
23
24

25 Data cleansing, taxonomy development and use

26
27 The data were subjected to a cleansing process, involving the detection and editing of the
28 data containing errors and inconsistencies. The raw Lifeline taxonomy data contained
29 spelling mistakes and multiple reasons that interchangeably describe the same condition.
30 This was due to recording of the reasons from multiple telephone counsellors and
31 different counsellors' perceptions of the content of the call, over the period of the data
32 collection. For instance, "Suicide Ideation" and "Suicidal Ideation" were considered as
33 two individual reasons. Using a heuristic process, duplicated reasons were collapsed into
34 a common reason, narrowing the number of main presenting issues from 226 to 170. It is
35 important to note that this heuristic process does not guarantee that the final result is
36 perfect or optimal. This was labelled as the 'Granular Taxonomy' (GT), representing a
37 list of unique presenting reasons. It was necessary to carry out this step to better
38 understand the distribution of calls against an agreed taxonomy, and to clean the data to
39 remove duplicates, prior to the application of machine learning techniques.
40
41
42

43
44 Next, a taxonomy based on the ICD-10 was created. The ICD-10 is a medical
45 classification manual that was created by the WHO to promote an international standard
46 of diagnostic guidelines for mental and physical illnesses. For each disorder within the
47 ICD-10, a description is provided of the main characteristics, providing a balance of
48 symptoms that is usually required for a diagnosis. Using the broadest medical
49 descriptions provided by the ICD-10 to create a taxonomy with the smallest possible
50 number of categories, each health-related reason (physical and mental) from the GT (of
51 170 reasons) was categorised under the relevant ICD-10 categories (with 26 reasons
52 resulting). An alternative diagnostic manual which could have been used for this task is
53 the Diagnostic and Statistical Manual of Mental Disorders (Fifth edition; DSM-V).
54 However, the DSM-V is used solely for classifying psychiatric disorders only; callers can
55
56
57
58
59
60

1
2
3 contact the service citing mental and physical issues, therefore the ICD-10 is considered
4 more appropriate on this occasion. The WHO provides an online training tool for medical
5 experts and non-medical experts in responsible use of the ICD-10, aiding the
6 development of the ICD-10 taxonomy in this study. It is important to note that it was not
7 possible to categorise all GT reasons under relevant ICD-10 categories. For instance,
8 some reasons were administration related, or were concerned with being a victim or
9 perpetrator of crime. See Appendix 1 for details of ICD-10 and GT taxonomies.
10
11

12 Descriptive statistics

13
14
15 Between 2008-2017, 51,135 clients used the Lifeline service. As mentioned previously,
16 calls were made to/from the service by clients, a third-party organisation or other sources.
17 In this study, only the calls made to and from the clients within the Lifeline data were
18 analysed (N= 636,726 or around 50% of the calls).
19

20 Association rule mining

21
22
23 Association rule mining is a data mining technique which is used to find frequent, co-
24 occurring associations between items within a dataset ⁽¹⁴⁾. Hipp et al ⁽¹⁵⁾ illustrate the
25 principle behind association rule mining, where association is generally expressed as A
26 $\Rightarrow B$, where A and B are items. A dataset of transactions (D), where each transaction (T)
27 that is made contains a set of items. If T contains the item A , then there is some
28 probability that the same transaction will contain the item B also ⁽¹⁵⁾.
29

30
31 Using the *Apriori* algorithm, associations between presenting reasons can be found
32 during a client's tenure with the service, where tenure is defined as the period of time
33 between their first and last interaction within the data set ⁽¹⁶⁾. Calls with no attached
34 motivating reason (termed as "Undeclared"; total number of calls= 125,776 or around
35 20% of eligible calls) were omitted due to the unknown nature of these calls, reducing the
36 number of eligible clients from 51,135 to 46,593.
37

38
39 The Apriori algorithm was applied using both the GT and ICD-10 taxonomies. For the
40 GT taxonomy, the minimum support was set at 0.1, meaning that reasons with a
41 proportion of the total available reasons within the data set of over 10% were included in
42 the process. The confidence level was set at 0.5, meaning that the rule is only included in
43 the output if the rule is correct 50% of the time. For the ICD-10 taxonomy, support level
44 was set at 0.07, meaning that reasons with a proportion of the total available reasons
45 within the data set of over 7% were included in the process. The confidence level was set
46 at 0.75, meaning that the rule is only included in the output if the rule is correct 75% of
47 the time.
48
49

50 Calls at geolocation level

51
52
53 There are 582 electoral ward areas in Northern Ireland, each with an average population
54 of 4,000 (ranging from 700 to 9,500), across around 1,500 households on average per
55 ward ⁽¹⁷⁾. They provide a basis for determining which reasons most and least consistently
56
57
58
59
60

appear across all wards, allowing for the identification of reasons which are most common or highly specific to certain areas. This was calculated by determining the proportion of calls per reason across all wards as a percentage. Standard Deviation (SD) was calculated for each reason for all wards. Reasons which report a high SD would suggest that there is a high variation of calls for that reason, while in contrast, those reasons reporting a low SD would suggest low variation in calls.

3. Results

Frequency of Recorded Reasons

[Insert Figure 1]

Figure 1: Frequency of recorded reasons (GT; top 25 out of 170)

Figure 1 displays the top 25 reasons that were recorded within the GT. The top 5 reasons: the most recorded reason within the GT was “Suicidal Intent” (N= 90,904; 14.3%); the second most recorded reason was “Suicidal Ideation” (N= 82,671; 13%); the third most recorded reason was “Anxiety” (N= 63,474; 10%); the fourth most recorded reason was “Depression” (N= 51,835; 8%); the fifth most recorded reason was “Mental Health” (N=48,614 ; 7.6%).

[Insert Figure 2]

Figure 2: Frequency of recorded reasons (ICD-10)

Figure 2 displays the frequency of ICD-10 reasons within the call data. The top 5 reasons: The most recorded reason under the ICD-10 taxonomy for contacting Lifeline was “Suicide Related Issues” (N=216,305, 34%); the second most recorded reason was “Neurotic Disorders” (N=67,862, 10.66%), the third most recorded reason was “Mood/Affective Disorders” (N=60,634, 9.52%); the fourth most recorded reason was “Miscellaneous/Unspecified” (N=60,595, 9.52%); the fifth most recorded reason was “Alcohol” (N= 26,638, 4.18%).

Risk Ratings

[Insert Figure 3]

Figure 3: Bar plot displaying the distribution of risk rating scores (GT, top 25 reasons out of 170.)

Figure 3 displays the risk ratings of the top 25 reasons within the GT associated with the highest risk rating. Only reasons which had more than 30 calls were included for analysis. The top five reasons with the highest risk ratings were “Suicidal Intent” (Mean= 7.47,

1
2
3 Calls= 4855), “High Risk to Others” (Mean= 7.13, Calls= 47), “Suicide Ideation”
4 (Mean= 6.27, Calls= 71055), “Childhood Abuse” (Mean= 6.22, Calls= 59) and
5 “Transgender” (Mean=6.20, Calls= 143).
6
7

8
9 A Levene’s test that the homogeneity of variances assumptions were violated ($F(80, 256632)= 55.36, p<.001$), therefore a Kruskal-Wallis test was conducted to test whether
10 there were any significant differences between risk ratings within the GT. Risk ratings
11 were significantly affected by reason within the GT ($H(70676)=80, p= <.001$), which
12 means that differences in mean risk ratings observed were less likely to be by chance and
13 that particular reasons are more associated with higher risk ratings than others.
14
15
16
17
18
19
20
21
22

23 [Insert Figure 4]

24
25 Figure 4: Bar plot displaying the distribution of risk ratings (GT, top 25 out of 170, as
26 ranked by volume of calls)
27

28 Figure 4 displays the risk ratings of the top 25 reasons in the GT ranked in descending
29 order by call volume. The top five reasons are “Suicide Ideation” (Mean= 6.27,
30 Calls=71055), “Anxiety” (Mean= 4.29, Calls= 37986), “Mental Health” (Mean=5.25,
31 Calls=32619), “Alcohol” (Mean=4.99, Calls= 19049) and “Depression” (Mean=4.46,
32 Calls=18848).
33
34
35

36
37 [Insert Figure 5]

38 Figure 5: Bar plot displaying the distribution of Risk Rating (ICD-10 taxonomy applied -
39 all reasons included).
40
41

42 Figure 5 displays the distributions of risk ratings across the ICD-10 reasons in descending
43 order. Only reasons which had more than 30 calls were included for analysis. The top
44 five reasons with the highest risk ratings are “Suicide Related Issues” (Mean= 6.27,
45 Calls= 92707), “Drugs: Legal” (Mean= 5.86, Calls= 645), “Personality Disorder”
46 (Mean= 5.83, Calls= 759), “Perpetrator of Crime” (Mean= 5.81, Calls= 190) and
47 “Undeclared” (Mean= 5.31, Calls= 353)
48
49
50
51

52 A Levene’s test showed that homogeneity of variance assumptions were violated
53 ($F(22,257082)=85.52, p<.001$), therefore a Kruskal-Wallis test was conducted to test
54 whether there were any significant differences between risk ratings within the ICD-10
55 taxonomy. Risk ratings between reasons are statistically significant within the ICD-10
56
57
58
59
60

1
2
3 taxonomy ($H(22)=65324$, $p<.001$) and like the GT, the differences in observed risk
4 ratings were less likely to be by chance and that particular reasons are more associated
5 with higher risk ratings than others.
6
7

8 9 Standard deviation of the proportion of reasons across all wards

10
11
12
13 [Insert Figure 6]

14
15
16 Figure 6: SD of proportion of reasons across all wards (Top 25 most variable reasons;
17 GT)
18

19
20
21 Figure 6 shows the SD of the proportion of the top 25 most variable reasons across all
22 wards in descending order. The top 12 reasons recorded the same SD ($SD=4.14$),
23 suggesting that these reasons were most specific to precise wards.
24
25

26 [Insert Figure 7]

27
28 Figure 7: SD of proportion of reasons across all wards (Top 25 least variable reasons;
29 GT)
30

31
32
33 Figure 7 shows the SD of the proportion of the top 25 least variable reasons across all
34 wards in descending order. The reasons that are of particular interests are at the bottom of
35 the figure, as these are the reasons which are most consistently cited across all wards;
36 “Depression” ($SD=0.19$), “Low Mood” ($SD=0.19$), “Family Conflict” ($SD=0.21$),
37 “Information” ($SD=0.21$) and “Family Breakdown” ($SD=0.21$).
38
39

40 [Insert Figure 8]

41
42 Figure 8: SD of proportion of reasons across all wards (ICD-10)

43
44 Figure 8 shows the SD of the proportion of reasons within the ICD-10 taxonomy across
45 all wards in descending order; reasons with the highest variation of calls are at the top
46 and those reasons with the lowest variation of calls are at the bottom. The 5 reasons
47 which were least consistently cited across all wards were “Drugs: Legal” ($SD=1.69$),
48 “Organic Mental Disorders” ($SD=1.28$), “Mental Disorder (non-specified)” ($SD=1.13$),
49 “Delusional Disorders” ($SD=1.11$) and “Sexuality” ($SD=1.08$). The 5 reasons which were
50 most consistently cited across all wards were “Family/Relationship Issues” ($SD=0.19$),
51 “Mood/Affective Disorders” ($SD=0.19$), “Suicide Related Issues” ($SD=0.21$) and
52 “Emotional Issues” ($SD=0.22$).
53
54
55
56
57
58
59
60

Association rule mining: Associations between reasons for contacting Lifeline

The following results have been obtained from association rule mining analysis of GT and ICD-10 taxonomies. The top two rules from each taxonomy are ranked by confidence, support and lift, in turn. A rule can be interpreted as

“if a caller cites A as an antecedent, then the caller will cite B as a consequent in a later call.”

In the context of this research, *confidence* is the measure that relates to the likeliness of the occurrence of the resultant reason in the caller’s call or sequence of calls given that the caller has contacted Lifeline citing a specific reason beforehand. *Support* gives an indication of how frequent a reason or set of reasons appears in all caller call sequences. *Lift* is how much more likely that one reason is cited relative to a typical rate of being cited. If the lift value of a rule is greater than 1, then it would mean that the reasons are cited together more often than what would be expected by chance. Full results from the association rule mining analysis are in Table 1.

Confidence

The leading rule within the GT by confidence is

{Suicide: Concerned about} => {Suicidal Intent} (Confidence: 0.86),

meaning that a caller is likely to cite Suicidal Intent in a later call if the previous call is Suicide: Concerned about.

The rule with the second highest confidence value is

{Information} => {Suicidal Intent} (Confidence: 0.78),

meaning that there are callers who are likely to cite Suicidal Intent in later calls after information seeking in an earlier call.

The leading rule within the ICD-10 taxonomy by confidence is

{Miscellaneous/Unspecified, Mood/Affective Disorders, Neurotic Disorders} => {Suicide Related Issues} (Confidence: 0.88).

In this case, there are three reasons (antecedents) in this rule (which may have occurred in at least 3 earlier calls) with Suicide Related issues as a consequent reason in a later call.

The rule with the second highest confidence value in the ICD-10 taxonomy is

1
2
3
4 $\{Miscellaneous/Unspecified, Mood/Affective Disorders\} \Rightarrow \{Suicide Related$
5 $Issues\}$ (Confidence: 0.84).
6

7
8 This rule has two reasons (antecedents) (which may have occurred in at least 2 earlier
9 calls) with Suicide Related issues as a consequent reason in a later call.
10

11 *Support*

12
13 The leading rule within the GT by support is

14
15 $\{Depression\} \Rightarrow \{Anxiety\}$ (Support: 0.19),
16
17

18 which means that the reasons within this rule are most prevalent within all rules or in all
19 caller call sequences.
20

21 The rule with the second highest support value is

22
23 $\{Suicidal Intent\} \Rightarrow \{Depression\}$ (Support: 0.18).
24
25

26 The leading rule in the ICD-10 taxonomy by support is

27
28 $\{Miscellaneous/Unspecified\} \Rightarrow \{Suicide Related Issues\}$ (Support: 0.24),
29
30

31 meaning that these reasons are the most prevalent in all rules and caller call sequences.
32

33 There are two rules which are tied for the second highest rule by support; using the rule
34 with the higher confidence value, the second highest rule by support is

35
36 $\{Miscellaneous/Unspecified, Mood/Affective Disorders\} \Rightarrow \{Suicide Related$
37 $Issues\}$ (Support: 0.16).
38
39

40 *Lift*

41
42 The leading rule within the GT by lift is

43
44 $\{Suicide: Concerned about\} \Rightarrow \{Suicidal Intent\}$ (Lift: 3.10).
45
46

47 This rule has the largest lift value and can be considered as the most important out of all
48 discovered rules within the data.
49

50 The rule with the second highest lift value is

51
52 $\{Information\} \Rightarrow \{Suicidal Intent\}$ (Lift: 2.79).
53
54

55 The leading rule within the ICD-10 by lift is
56
57
58
59
60

1
2
3
4 *{Miscellaneous/Unspecified, Neurotic Disorders, Suicide Related Issues} =>*
5 *{Mood/Affective Disorders} (Lift: 1.84).*
6
7

8 The rule that has the second highest lift value is
9

10 *{Miscellaneous/Unspecified, Neurotic Disorders} => {Mood/Affective Disorders}*
11 *(Lift: 1.72).*
12
13

14 [Insert Table 1]
15
16

17 Table 1: Full results for association rule analysis of the GT and ICD-10 taxonomy
18 (ranked by confidence)
19
20
21
22
23
24

25 4. Discussion

26
27 This paper sought to uncover insights regarding the reasons why individuals contact crisis
28 helplines, and addresses the following research questions: what are the most common
29 reasons for contacting a crisis helpline as coded by the crisis counsellors, and which are
30 related? Are there differences in terms of risk of suicide attached to different reasons and
31 if so, which reasons are considered more severe in terms of risk of suicide? Are the
32 proportions of each reason consistent across geolocations?
33
34

35 Two different taxonomies were used to help clarify and simplify the presenting reasons.
36 These were GT and ICD-10 taxonomies. Exploratory data analysis showed visible
37 differences in reasons for contacting crisis helplines when each taxonomy is applied.
38 Within the GT, reasons relating to suicide (i.e., “Suicidal Intent”, “Suicide Ideation”)
39 were the most reported reasons for contacting Lifeline, while for ICD-10 the main
40 presenting reason was “Suicide Related Issues”.
41
42

43 “Suicide Related Issues” had the highest mean risk rating of all the reasons within the
44 ICD-10 taxonomy and was second highest in the number of calls. We also see this being
45 reflected within the GT, as “Suicidal Intent”, “Suicidal Ideation” and “History of Suicide
46 Attempts” are the first, third and seventh highest (respectively) in risk rating.
47 Interestingly, “Drugs: Legal” had the second highest mean risk rating in the ICD-10,
48 which constituted of 2 of the 25 top reasons within the GT. However, these reasons
49 account for a relatively small proportion of calls and may only be made by a small
50 number of clients. The reasons which have higher mean risk ratings have been shown to
51 be associated to suicidal behaviour within psychiatry literature. Within ICD-10, calls
52 under the reasons “Personality Disorder”^(18,19), “Delusional Disorders”⁽²⁰⁾ showed a
53 medium to moderate mean risk rating.
54
55
56
57
58
59
60

Another objective was to explore how the *Apriori* association rule mining algorithm could reveal associations between presenting reasons for calling the helpline, employing the two different taxonomies. This study demonstrates that associations exist. The Apriori algorithm identified associations between suicide related reasons when applied to both taxonomies, and that such reasons were widely connected to having a mental illness (i.e., depression, anxiety). The findings from the Apriori algorithm parallel with the wider consensus that mental illnesses such as depression and anxiety are mediating factors in suicidal behaviour indicating that crisis helplines are being contacted for the reasons for which the service was designed^(2,4,21,22). Potentially, this insight can provide early warning signs about a client who could be suicidal and inform crisis helplines as to how their services are used.

Practical and research applications

Many factors contribute towards mental illness and suicidal behaviour at individual and population level. For instance, a study by Bunting et al.⁽²³⁾ found that 33 wards (6% of all wards within Northern Ireland), which had recorded deaths by suicide over a 7-year period, had an excess of 50% more deaths than the societal average. We illustrate the potential of using helpline call data to highlight areas and reasons which are most and least consistently cited across geographical areas at ward level. By determining which reasons are least consistently cited across geolocations, the call data can highlight areas within the country that experience higher levels of specific problems than other areas. For example, the current analysis illustrates that issues concerning legal drug abuse are least consistently cited across geolocations, showing that there are specific areas within the country that experience this issue, which was associated with a high mean risk rating. Perhaps this may be indicative of irresponsible prescribing of medication, solvent abuse or other substances associated with legal highs. The information derived from the call log data could help create a targeted public health intervention for certain issues and potentially lower the risk of dangerous behaviour (self-harm or to others) in the population.

Crisis helplines could benefit from using different taxonomy approaches as it can help provide a consistent structure for categorising reasons for each call. As mentioned earlier, inconsistent recording of reasons as to why individuals contact crisis helplines influences the true range of reasons for contacting a crisis helpline and negatively impacts attempts to “treat” callers who utilise crisis helplines over a prolonged period of time.

From a data analytic perspective, having different taxonomic approaches has numerous advantages. Firstly, employing a broad taxonomy such as the ICD-10 in the case of the current research, can be helpful for reducing the complexity of data analysis tasks when working with large datasets by ‘collapsing’ all the potential outcomes into common categories. In our case, we were able to narrow 226 reasons from the raw data into 170 granular reasons, and then again to 26 reasons using ICD-10. Secondly, if one were to

1
2
3 employ a broad taxonomy approach such as ICD-10 for predictive tasks, it can help to
4 account for reasons which are highly specific but have fewer observations.
5

6 7 Limitations

8
9 The main limitation of this study is that the main presenting reasons for each call is
10 recorded by the crisis counsellor, and the recording of the reason may be influenced by
11 the crisis counsellor's perception of the call content. As mentioned earlier, there can be
12 many reasons cited by a client over the duration of a single call, but only one reason per
13 call is available for analysis. While counsellors will have received appropriate training in
14 coding the main presenting reason for each call, there may be some variation between
15 counsellors. In addition, recall bias may be in effect as it is possible that the crisis
16 counsellors may only record what they have remembered about the call. It was also
17 difficult to verify the accuracy of the data regarding the presenting reasons and the
18 attached risk rating.
19

20
21 The data was collected over a 10-year period, it is therefore likely that telephone
22 counsellors will have received updated training on how to identify risks and how to
23 interpret the call content on a regular basis. This may cause some added variation in how
24 the presenting was recorded over the course of time. Information on types of training
25 (and time of training) that telephone counsellors have received were not available at the
26 time of study, but it would be of interest to investigate whether training (type of training
27 and when it was conducted) has had any impact on how the data is recorded.
28
29

30
31 A large proportion of eligible calls within the dataset did not have a main presenting
32 reason associated with them and were categorised as "Undeclared". This has limited the
33 amount of available data that could be used as to why individuals contacted Lifeline.
34 There are various explanations for this. One such explanation could be that there were
35 instances where urgent intervention was required when a highly suicidal caller had
36 contacted the service, which would take precedence over data collection from the
37 telephone counsellor's end. Another explanation is that a proportion of "Undeclared"
38 calls reflect situations where counsellors failed to record a reason for the call.
39
40

41 42 Future Work

43
44 Future work regarding understanding the reasons why people contact crisis helplines
45 could involve determining which reasons are associated with risk of suicide. However, it
46 is important to note that the prediction of suicide risk has not been possible as it is not
47 recommended as a basis for establishing what type of care an individual should receive;
48 for guidelines, refer to the National Institute for Health and Care Excellence (NICE) ⁽²⁴⁾.
49 Within the current dataset, it is possible to examine which reasons, alongside other
50 characteristics of caller behaviour (i.e., call frequency, caller type, etc.), best predictions
51 of high and low risk rating scores. This would have the potential to provide accurate and
52 early knowledge of which variables are most associated with risk of suicide and therefore
53 allow crisis helplines to act in a more clinical, time-sensitive manner to provide
54 intervention for those at most risk of engaging in suicidal behaviour.
55
56
57
58
59
60

Another line of investigation is to establish whether client risk ratings decrease as they continue to interact with the service. Previous research has found that the level of distress exhibited by the caller to a crisis helpline is significantly lowered from the beginning of their call to the end of the call ⁽²⁵⁾, while another study had found that suicidal callers who contacted a crisis helpline had their overall suicide risk significantly reduced from the beginning to the end of a call ⁽²⁶⁾. However, no study has assessed changes in risk for repeat callers. While crisis helplines are not designed for long term, repeated usage, some callers will contact the service more than once, allowing for temporal analysis of caller outcomes.

Conclusion

This study demonstrated that many actionable insights can be retrieved from crisis helpline call data through the lens of two different taxonomy approaches. Insights from this paper can be used by stakeholders such as crisis helpline services, public health officials and researchers to understand the motivating reasons for contacting crisis helplines, and potentially aid stakeholders in optimising and targeting interventions in the mental health and suicide prevention domains.

In conclusion, many people call helplines for suicide related issues, neurotic disorders, and mood and affective disorders. Using associative rule mining, this research shows that miscellaneous/unspecified calls are related to mood/affective disorders and neurotic disorders. Reasons for calling helplines that are associated with higher risk ratings include those calling with a personality disorder, mental disorders, delusional disorders and drugs (legal). We can also conclude that certain reasons such as drugs (legal), organic mental disorders, mental disorders and sexuality may be subject to regional variation.

References

1. Watson RJ, McDonald J, Pearce DC. An exploration of national calls to Lifeline Australia: social support or urgent suicide intervention?. *British Journal of Guidance & Counselling*. 2006 Nov 1;34(4):471-82.
2. Burgess N, Christensen H, Leach LS, Farrer L, Griffiths KM. Mental health profile of callers to a telephone counselling service. *Journal of Telemedicine and Telecare*. 2008 Apr;14(1):42-7.

- 1
2
3 3. Middleton A, Gunn J, Bassilios B, Pirkis J. Systematic review of research into
4 frequent callers to crisis helplines. *Journal of Telemedicine and Telecare*. 2014
5 Mar;20(2):89-98.
6
7
- 8 4. Bassilios B, Harris M, Middleton A, Gunn J, Pirkis J. Characteristics of people
9 who use telephone counseling: Findings from secondary analysis of a population-
10 based study. *Administration and Policy in Mental Health and Mental Health
11 Services Research*. 2015 Sep 1;42(5):621-32.
12
- 13 5. Kinzel A, Nanson JO. Education and debriefing: Strategies for preventing crises
14 in crisis-line volunteers. *Crisis: The Journal of Crisis Intervention and Suicide
15 Prevention*. 2000;21(3):126.
16
17
- 18 6. Middleton A, Woodward A, Gunn J, Bassilios B, Pirkis J. How do frequent users
19 of crisis helplines differ from other users regarding their reasons for calling?
20 Results from a survey with callers to Lifeline, Australia's national crisis helpline
21 service. *Health & social care in the community*. 2017 May;25(3):1041-9.
22
23
- 24 7. Silla CN, Freitas AA. A survey of hierarchical classification across different
25 application domains. *Data Mining and Knowledge Discovery*. 2011 Jan 1;22(1-
26 2):31-72.
27
- 28 8. World Health Organization. The ICD-10 classification of mental and behavioural
29 disorders: diagnostic criteria for research.
30
31
- 32 9. Northern Ireland Department of Health 2016, Protect Life 2: A Draft Strategy for
33 Suicide Prevention in the North of Ireland, Northern Ireland Department of
34 Health.
35
- 36 10. Ramsey, C., Ennis, E. and O'Neill, S., 2019. Characteristics of Lifeline, Crisis
37 Line, Service Users Who Have Died by Suicide. *Suicide and Life-Threatening
38 Behavior*, 49(3), pp.777-788.
39
40
- 41 11. British Association of Counselling and Psychotherapy. 2018. *Good Practice in
42 Action 042: Fact Sheet: Working with suicidal clients in the counselling
43 professions*. Available at: [https://www.bacp.co.uk/media/2157/bacp-working-
44 with-suicidal-clients-fact-sheet-gpia042.pdf](https://www.bacp.co.uk/media/2157/bacp-working-with-suicidal-clients-fact-sheet-gpia042.pdf) [Accessed 13/01/2019]
45
46
- 47 12. Turkington R, Mulvenna M, Bond R, O'Neill, S., Armour, C., The Application of
48 User Event Log Data for Mental Health and Wellbeing Analysis, 2018,
49 Proceedings of the 32nd International BCS Human Computer Interaction
50 Conference (HCI 2018) Belfast, UK, 4-6 July 2018,
51 <http://dx.doi.org/10.14236/ewic/HCI2018.4/>
52
53
- 54 13. R Core Team (2018). R: A language and environment for statistical computing. R
55 Foundation for Statistical Computing, Vienna, Austria. URL
56
57
58
59
60

14. Agrawal R, Srikant R. Fast algorithms for mining association rules. InProc. 20th int. conf. very large data bases, VLDB 1994 Sep 12 (Vol. 1215, pp. 487-499).
15. Hipp J, Güntzer U, Nakhaeizadeh G. Algorithms for association rule mining—a general survey and comparison. ACM sigkdd explorations newsletter. 2000 Jun 1;2(1):58-64.
16. Turkington, R., Mulvenna, M., Bond, R., O'Neill, S. and Armour, C., 2019, September. User Archetype Discovery By Cluster Analysis of Caller Log Data: Tenure Evolution is Stable as Time Period Reduces. In Proceedings of the 31st European Conference on Cognitive Ergonomics (pp. 196-202). ACM.
17. Northern Ireland Statistics and Research Agency. 2019. *Geography Fact Sheet*. Available at: <https://www.ninis2.nisra.gov.uk/public/documents/NISRA%20Geography%20Fact%20Sheet.pdf> [Accessed 27/11/2019]
18. Brodsky BS, Malone KM, Ellis SP, Dulit RA, Mann JJ. Characteristics of borderline personality disorder associated with suicidal behavior. American Journal of Psychiatry. 1997 Dec 1;154(12):1715-9.
19. Oldham JM. Borderline personality disorder and suicidality. American Journal of Psychiatry. 2006 Jan 1;163(1):20-6.
20. González-Rodríguez, A., Molina-Andreu, O., Navarro Odriozola, V., Gastó Ferrer, C., Penadés, R. and Catalán, R., 2014. Suicidal ideation and suicidal behaviour in delusional disorder: a clinical overview. Psychiatry journal, 2014.
21. Gould MS, Lake AM, Munfakh JL, Galfalvy H, Kleinman M, Williams C, Glass A, McKeon R. Helping callers to the National Suicide Prevention Lifeline who are at imminent risk of suicide: Evaluation of caller risk profiles and interventions implemented. Suicide and Life-Threatening Behavior. 2016 Apr;46(2):172-90.
22. Hunt T, Wilson CJ, Woodward A, Caputi P, Wilson I. Intervention among suicidal men: Future directions for telephone crisis support research. Frontiers in public health. 2018 Jan 19;6:1.
23. Bunting, B., Corry, C., O'Neill, S. and Moore, A., 2018. Death by suicide at the Ward level in Northern Ireland. Psychological medicine, 48(8), pp.1375-1380.

24. National Institute for Health and Care Excellence (2018) *Preventing suicide in community and custodial settings* (NICE Guideline 105). Available at: <https://www.nice.org.uk/guidance/ng105> [Accessed 27/11/2019]
25. Kalafat, J., Gould, M.S., Munfakh, J.L.H. and Kleinman, M., 2007. An evaluation of crisis hotline outcomes. Part 1: Nonsuicidal crisis callers. *Suicide and Life-threatening behavior*, 37(3), pp.322-337.
26. Gould, M.S., Kalafat, J., HarrisMunfakh, J.L. and Kleinman, M., 2007. An evaluation of crisis hotline outcomes part 2: Suicidal callers. *Suicide and Life-Threatening Behavior*, 37(3), pp.338-352.

Appendices

Taxonomy	LHS	RHS	Support	Confidence	Lift	Count
ICD	{Miscellaneous/Unspecified,Mood/Affective Disorders,Neurotic Disorders}	{Suicide Related Issues}	0.11	0.88	1.5	5074
	{Miscellaneous/Unspecified,Mood/Affective Disorders}	{Suicide Related Issues}	0.16	0.84	1.43	7353
	{Miscellaneous/Unspecified,Neurotic Disorders}	{Suicide Related Issues}	0.14	0.82	1.4	6546
	{Miscellaneous/Unspecified,Neurotic Disorders,Suicide Related Issues}	{Mood/Affective Disorders}	0.11	0.78	1.84	5074
	{Miscellaneous/Unspecified}	{Suicide Related Issues}	0.24	0.77	1.3	11051
	{Mood/Affective Disorders,Neurotic Disorders}	{Suicide Related Issues}	0.16	0.75	1.28	7356
	{Miscellaneous/Unspecified,Neurotic Disorders}	{Mood/Affective Disorders}	0.12	0.73	1.72	5771
Granular	{Suicide: Concerned about}	{Suicidal Intent}	0.10	0.86	3.10	4880
	{Information}	{Suicidal Intent}	0.12	0.78	2.79	5404
	{Anxiety,Suicidal Intent}	{Depression}	0.11	0.77	2.01	5202
	{Information}	{Depression}	0.11	0.74	1.92	5117
	{Suicidal Intent}	{Depression}	0.18	0.63	1.66	8238
	{Depression,Suicidal Intent}	{Anxiety}	0.11	0.63	1.61	5202
	{Anxiety,Depression}	{Suicidal Intent}	0.11	0.58	2.08	5202
	{Mental Health}	{Anxiety}	0.11	0.58	1.48	4918
{Mental Health}	{Depression}	0.10	0.56	1.47	4790	

	{Suicidal Intent}	{Anxiety}	0.14	0.52	1.33	6741
	{Depression}	{Anxiety}	0.19	0.50	1.28	8974

Table 1: Full results for association rule analysis of the GT and ICD-10 taxonomy.

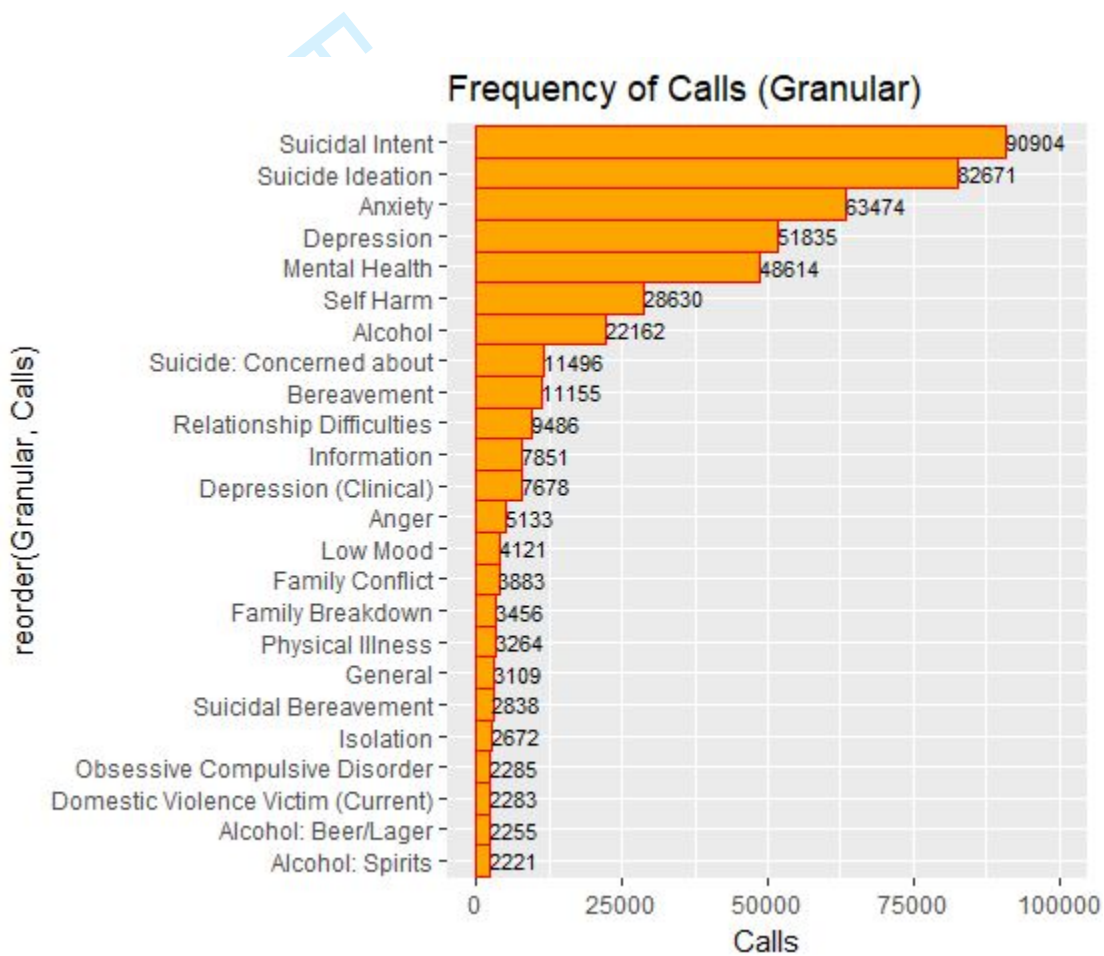


Figure 1: Frequency of recorded reasons (GT; top 25 out of 170)

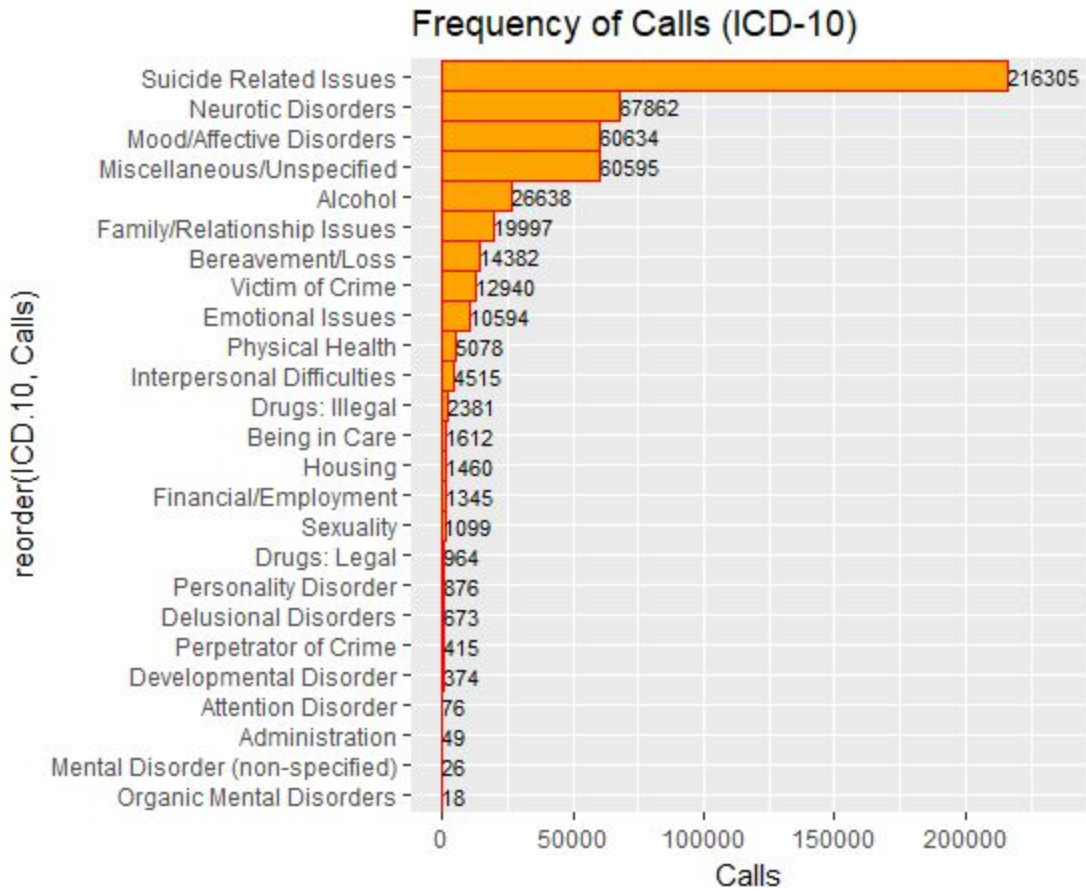


Figure 2: Frequency of recorded reasons (ICD-10)

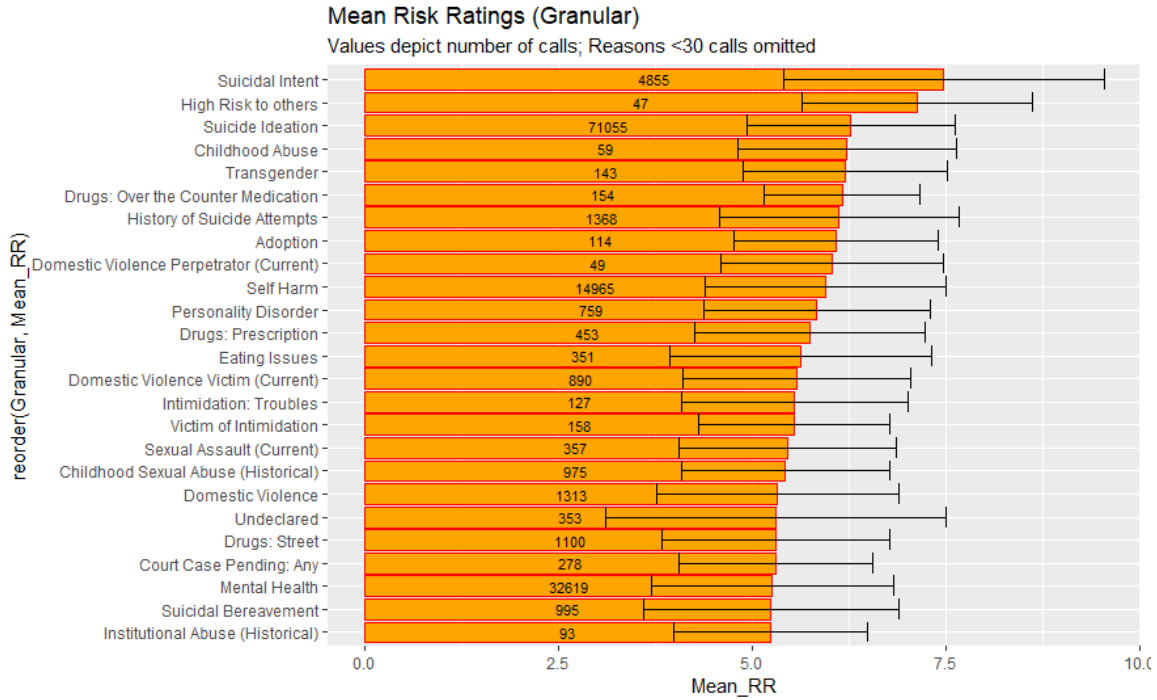


Figure 3: Barplots displaying the distribution of risk rating scores (GT, top 25 reasons out of 170.)

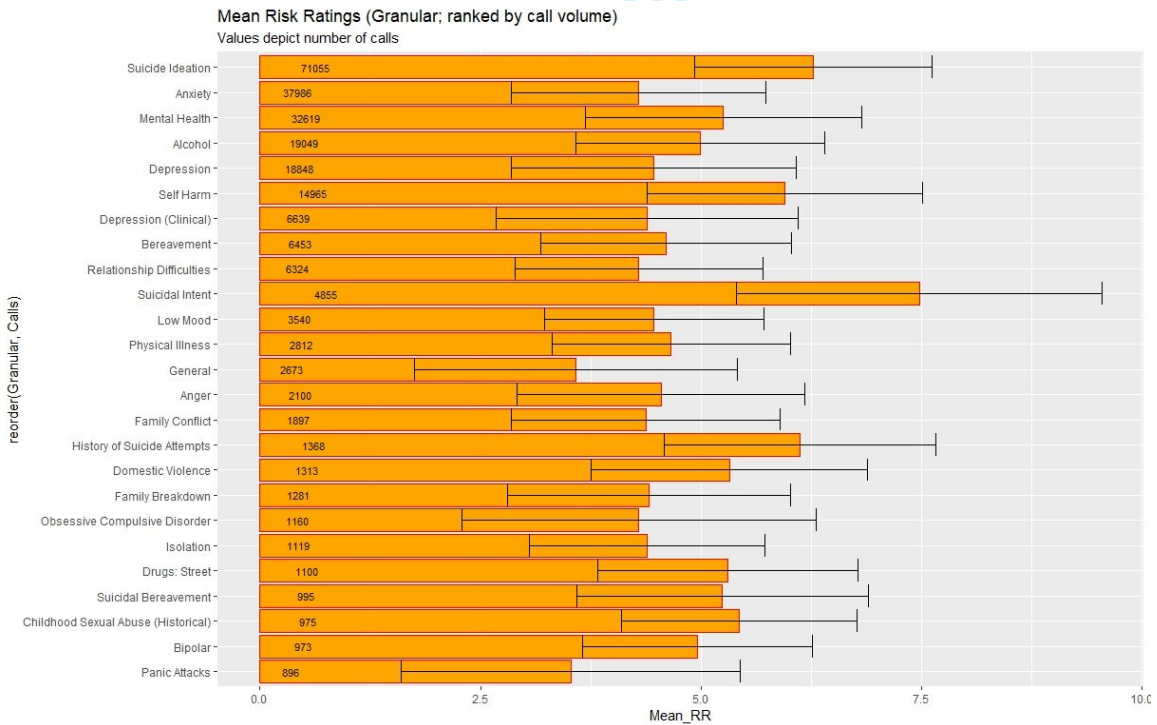


Figure 4: Bar plot displaying the distribution of risk ratings (GT, top 25 out of 170, as ranked by volume of calls)

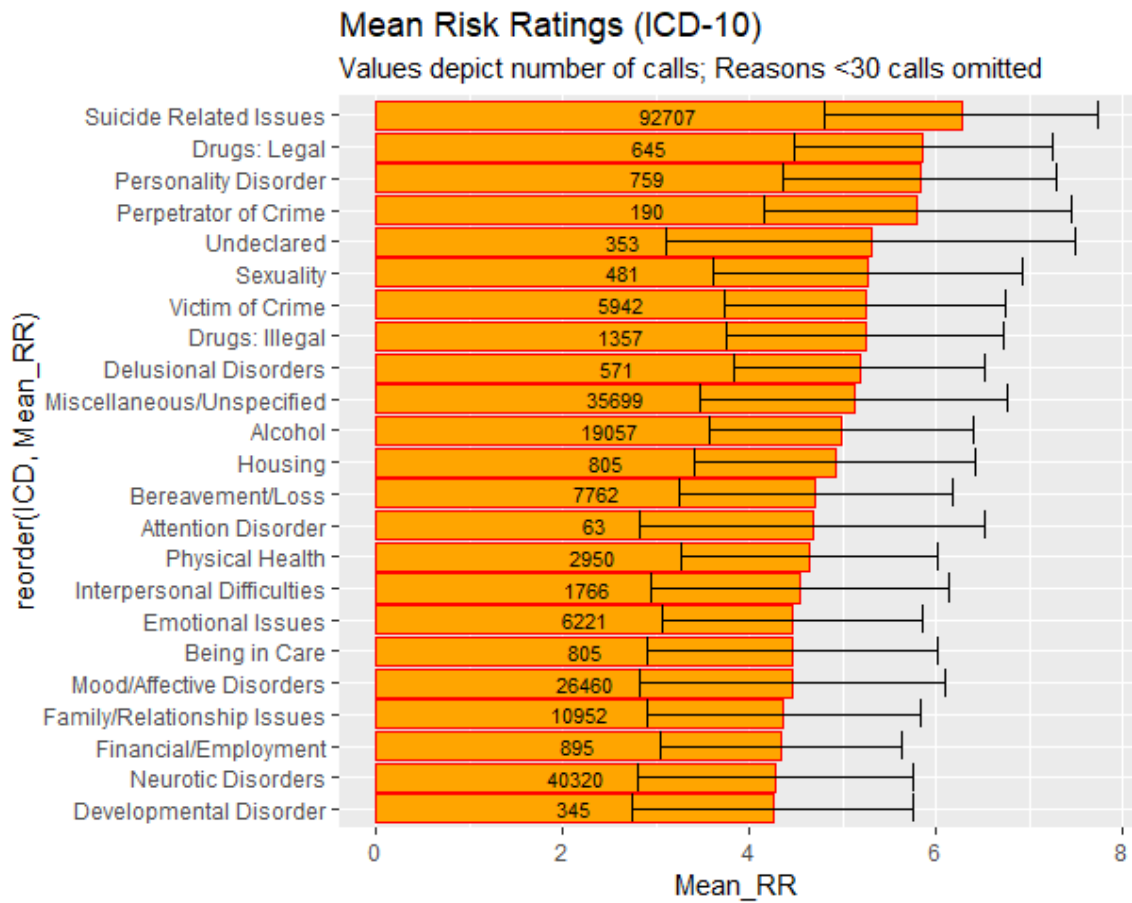


Figure 5: Barplots displaying the distribution of Risk Rating (ICD-10 taxonomy applied).

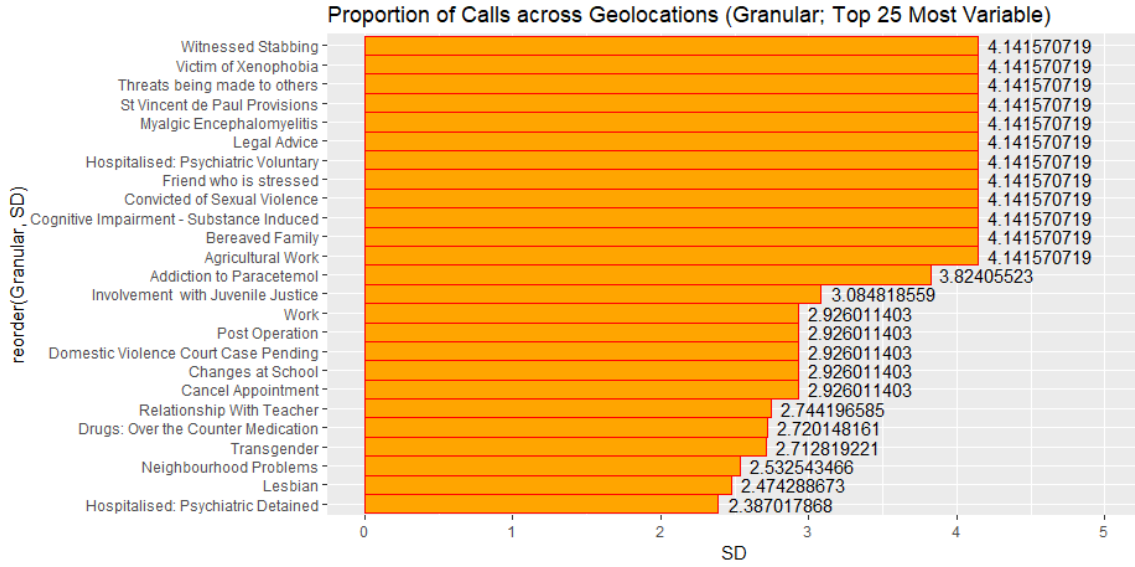


Figure 6: SD of proportion of reasons across all wards (Top 25 Most Variable Granular Reasons)

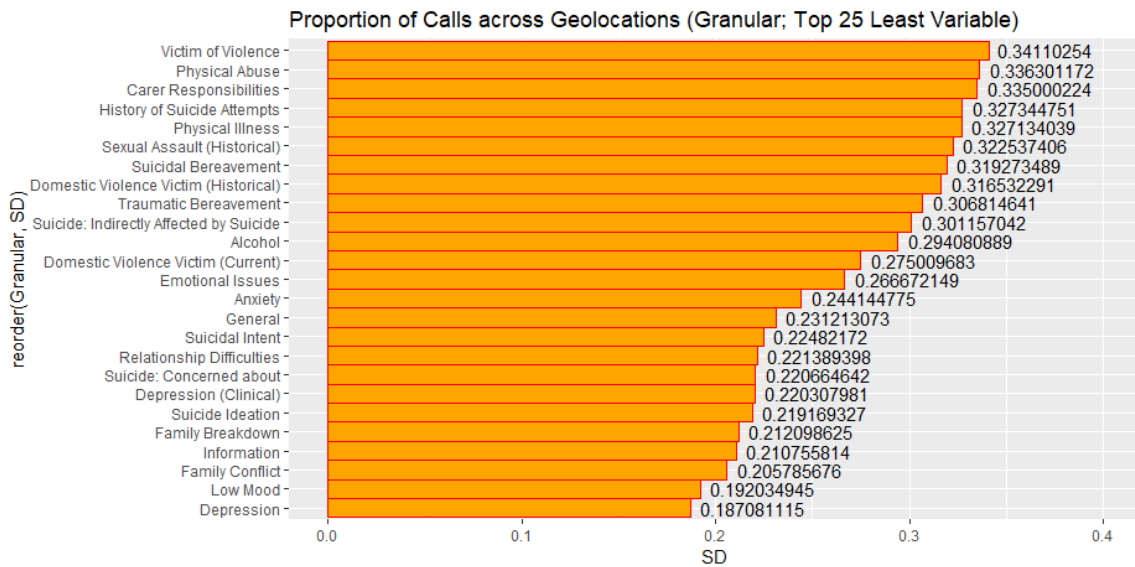


Figure 7: SD of proportion of reasons across all wards (Top 25 Least Variable Granular Reasons)

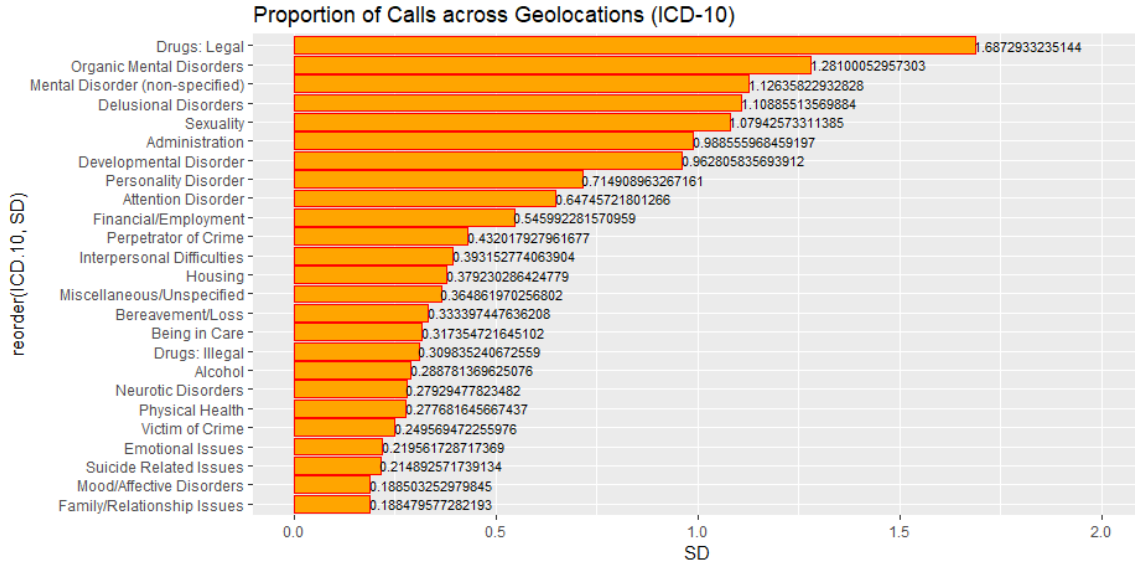


Figure 8: SD of proportion of reasons across all wards (ICD-10)

Appendix 1 – Detail on Granular and ICD-10 Taxonomies

ICD-10 Taxonomy	Granular Taxonomy
Drugs: Legal	Addiction to Paracetamol Drugs: Prescription Drugs: Solvents Legal Highs Medication
Drugs: Illegal	Affected by Substance Abuse Drugs: Street Drugs: Amphetamines Drugs: Cannabis Drugs: Ecstasy Drugs: Cocaine Drugs: Heroin Drugs: LSD Substances: Indirectly Affected
Alcohol	Alcohol Alcohol: Beer/Lager Alcohol: Spirits
Physical Health	Disability Hospitalised: A&E Physical Illness Physically Injured Post Operation Sexual Health Myalgic Encephalomyelitis Insomnia
Mood/Affective Disorders	Bipolar Depression Depression (Clinical)
Neurotic Disorders	Anxiety Obsessive Compulsive Disorder PTSD Trauma Phobia Inc Agoraphobia Panic Attacks
Emotional Issues	Anger Emotional Abuse Emotional Issues Emotionally Injured Neglect Self Esteem Low Mood Chronic Sadness

Sexuality	Bisexual Gay Lesbian Sexuality Transgender Transsexual
Administration	Absence/Administration Cancel Appointment
Attention Disorder	ADHD
Developmental Disorder	Learning Difficulties Aspergers
Personality Disorder	Personality Disorder
Mental Disorder (non-specified)	Hospitalised: Psychiatric Detained Hospitalised: Psychiatric Voluntary Cognitive Impairment: Injury Induced Mental Health
Miscellaneous/Unspecified	General Information Other St Vincent de Paul Provisions Witness Court Case Pending: Any Sexual Violence Court Case Pending
Delusional Disorders	Schizophrenia
Organic Mental Disorder	Dementia
Suicide Related Issues	History of Suicide Attempts Suicidal Intent Suicide: Concerned about Suicide: Indirectly affected by suicide Suicide Ideation Self Harm
Family/Relationship Issues	Family Breakdown Family Conflict Family Illness IVF Abortion Adoption Relationship with Teacher Relationship with parent/carer Pregnancy Relationship Difficulties Ind. Affected by Substance Abuse: Parents/Family Parenting Difficulties Separation Divorce
Being in Care	Being in Care

	Carer Responsibilities Institutional Abuse (Historic)
Bereavement/Loss	Bereavement: Troubles Bereaved Family Bereavement Suicidal Bereavement Traumatic Bereavement
Victim of Crime	Childhood Abuse Childhood Sexual Abuse Childhood Sexual Abuse (Historical) Domestic Violence Domestic Violence Emotional Domestic Violence Physical Domestic Violence Victim (Current) Domestic Violence Victim (Historical) Domestic Violence Witness (Current) Domestic Violence Witness (Historical) Paramilitary Violence Victim Physical Abuse Racism Rape Sexual Assault (Current) Sexual Assault (Historical) Sexual Harassment (Current) Theft Victim: Troubles Victim of Crime Victim of Intimidation Victim of Violence Victim of Xenophobia Witness to Troubles Witnessed Stabbing Intimidation: Troubles
Perpetrator of Crime	Criminal Assault Perpetrator Domestic Violence Perpetrator (Current) Domestic Violence Perpetrator (Historical) High Risk to others Incarceration Involvement with Juvenile Justice Involvement with Probation Perpetrator of Domestic Violence Registered on Sex Offenders Reg. Risk to self or others Threats being made to others Violent Behaviour

	Convicted of Sexual Violence
Financial/Employment	Career Debt Disciplinary Financial Employment Redundancy Gambling Agricultural Work
Interpersonal Difficulties	School Issues Isolation Friend who is stressed Exam Pressure Eating Issues Bullying
Housing	Neighbourhood Problems Housing Security Housing: General Homelessness Leaving Home
Undeclared	Undeclared

Taxonomy Structure