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

Robin Turkington
Maurice D Mulvenna
Raymond R Bond
Ulster University at Jordanstown, UK

Siobhan O’Neill
Ulster University at Coleraine, UK

Courtney Potts
Ulster University at Jordanstown, UK

Cherie Armour
Edel Ennis
Ulster University at Coleraine, UK

Catherine Millman
Public Health Agency, UK

Abstract
The objective of this study is to identify the most common reasons for contacting a crisis helpline through analysing a large call log data set. Two taxonomies were identified within the call log data from a Northern Ireland telephone crisis helpline (Lifeline), categorising the cited reason for each call. One taxonomy categorised the reasons at a fine granular level; the other taxonomy used the relatively coarser International Classification of Diseases–10. Exploratory data analytic techniques were applied to discover
insights into 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 demonstrate that 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). In conclusion, employing two differing taxonomy approaches to analyse call log data reveals the prevalence of main presenting reasons for contacting a crisis helpline. The 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

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 issues2–4 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. Finally, 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 article is to use two taxonomy approaches to aid the analysis of call log data, especially 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 article.
This article 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.

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

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

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

Lifeline data

Call data were recorded on Lifeline’s Information Management System for all callers who have interacted with the service. There are many features within the call data, but three features were used in this study:

- **Main presenting issue**: The reason for a client contacting the service is recorded. Some examples could be that the caller is depressed, experiencing financial difficulties, exhibiting suicidal behaviour and so on.
- **Risk Rating**: Lifeline counsellors conduct a risk assessment on each of their clients. This formulation balances suicide risk factors against protective risk factors within a client’s call counselling session. This risk factor is presented as a number between 1 to 10, with 0 representing missing values, 1–3 representing a low, 4–6 medium-to-moderate risk of suicide and
7–10 representing a moderate-to-severe risk of suicide. Rating of 0 has been removed. This rating is assigned by the telephone counsellor, who has specialised training in risk management.\textsuperscript{11}

- Area in which the caller is from (by Electoral Ward).

The remaining features in the data that were not used are considered as basic log data features such as timestamps, duration of call, call/event identifier and client identifier. Each record within the call log data has a unique identifier code for each client which is based on the client’s phone number. This allows for the client to remain anonymous, but still allows crisis helplines to keep trace. Both inbound calls (calls made by the client to the service) and outbound calls (calls made by the service to the client) were used for the data analysis. The call data were subjected to data wrangling processes, making it amenable to user data analytics and machine learning processes which are demonstrated in this article.\textsuperscript{12}

**Data analysis**

R programming language and R Studio (version 3.4.2) were used for the data wrangling process and statistical analyses, respectively.\textsuperscript{13} The Rcmdr library was used to conduct statistical analyses, while the ggplot2 library was used to create data visualisations. The arules package was used for association rule mining analysis.

**Data cleansing, taxonomy development and use**

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

Next, a taxonomy based on the ICD-10 was created. The ICD-10 is a medical classification manual that was created by the World Health Organization (WHO) to promote an international standard of diagnostic guidelines for mental and physical illnesses. For each disorder within the ICD-10, a description of the main characteristics is provided, establishing a balance of symptoms that is usually required for a diagnosis. Using the broadest medical descriptions provided by the ICD-10 to create a taxonomy with the smallest possible number of categories, each health-related reason (physical and mental) from the GT (of 170 reasons) was categorised under the relevant ICD-10 categories (with 26 reasons resulting). An alternative diagnostic manual that could have been used for this task is the Diagnostic and Statistical Manual of Mental Disorders (Fifth edition; DSM-V). However, the DSM-V is used solely for classifying psychiatric disorders only, and callers can contact the service citing mental and physical issues; therefore, the ICD-10 is considered more appropriate on this occasion. The WHO provides an online training tool for medical experts and non-medical experts in responsible use of the ICD-10, aiding the development of the ICD-10 taxonomy in this study. It is important to note that it was not possible to categorise all GT reasons
under relevant ICD-10 categories. For instance, some reasons were administration related or were concerned with being a victim or perpetrator of crime. See Supplemental Appendix 1 for details of ICD-10 and GT taxonomies.

Descriptive statistics

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

Association rule mining

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

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

The Apriori algorithm was applied using both the GT and ICD-10 taxonomies. For the GT taxonomy, the minimum support was set at 0.1, meaning that reasons with a proportion of the total available reasons within the data set of over 10 per cent were included in the process. The confidence level was set at 0.5, meaning that the rule is only included in the output if the rule is correct 50 per cent of the time. For the ICD-10 taxonomy, the support level was set at 0.07, meaning that reasons with a proportion of the total available reasons within the data set of over 7 per cent were included in the process. The confidence level was set at 0.75, meaning that the rule is only included in the output if the rule is correct 75 per cent of the time.

Calls at geolocation level

There are 582 electoral ward areas in Northern Ireland, each with an average population of 4000 (ranging from 700 to 9500), across around 1500 households on average per ward. They provide a basis for determining which reasons most and least consistently 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.

Results

Frequency of recorded reasons

Figure 1 displays the top 25 reasons that were recorded within the GT. The top five reasons are as follows: 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%) and the fifth most recorded reason was ‘Mental Health’ (N=48,614; 7.6%).

Figure 2 displays the frequency of ICD-10 reasons within the call data. The top five reasons are as follows: 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%) and the fifth most recorded reason was ‘Alcohol’ (N=26,638; 4.18%).

Risk ratings

Figure 3 displays the risk ratings of the top 25 reasons within the GT associated with the highest risk rating. Only reasons that had more than 30 calls were included for analysis. The top five reasons with the highest risk ratings were ‘Suicidal Intent’ (mean=7.47, calls=4855), ‘High Risk to Others’ (mean=7.13, calls=47), ‘Suicide Ideation’ (mean=6.27, calls=71055), ‘Childhood Abuse’ (mean=6.22, calls=59) and ‘Transgender’ (mean=6.20, calls=143).

Levene’s test showed that the homogeneity of variances assumptions was violated (F(80, 256632)=55.36, p ≈ 0.001); therefore, a Kruskal–Wallis test was conducted to test whether there were any significant differences between risk ratings within the GT. Risk ratings were significantly affected by reason within the GT (H(70,676)=80, p ≈ 0.001), which means that differences in mean risk ratings observed were less likely to be by chance and that particular reasons are more associated with higher risk ratings than others.
Figure 4 displays the risk ratings of the top 25 reasons in the GT ranked in descending order by call volume. The top five reasons are ‘Suicide Ideation’ (mean = 6.27, calls = 71,055), ‘Anxiety’ (mean = 4.29, calls = 37,986), ‘Mental Health’ (mean = 5.25, calls = 32,619), ‘Alcohol’ (mean = 4.99, calls = 19,049) and ‘Depression’ (mean = 4.46, calls = 18,848).

Figure 5 displays the distributions of risk ratings across the ICD-10 reasons in descending order. Only reasons that had more than 30 calls were included for analysis. The top five reasons with the highest risk ratings are ‘Suicide Related Issues’ (mean = 6.27, calls = 92,707), ‘Drugs: Legal’ (mean = 5.86, calls = 645), ‘Personality Disorder’ (mean = 5.83, calls = 759), ‘Perpetrator of Crime’ (mean = 5.81, calls = 190) and ‘Undeclared’ (mean = 5.31, calls = 353).

Levene’s test showed that homogeneity of variance assumptions was violated (F(22,257082) = 85.52, p ≤ 0.001); therefore, a Kruskal–Wallis test was conducted to test whether there were any significant differences between risk ratings within the ICD-10 taxonomy. Risk ratings between reasons are statistically significant within the ICD-10 taxonomy (H(22) = 65,324, p ≤ 0.001), and like the GT, the differences in observed risk ratings were less likely to be by chance and that particular reasons are more associated with higher risk ratings than others.

SD of the proportion of reasons across all wards

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

Figure 7 shows the SD of the proportion of the top 25 least variable reasons across all wards in descending order. The reasons that are of particular interest are at the bottom of the figure, as these are the reasons which are most consistently cited across all wards: ‘Depression’ (SD = 0.19), ‘Low Mood’ (SD = 0.19), ‘Family Conflict’ (SD = 0.21), ‘Information’ (SD = 0.21) and ‘Family Breakdown’ (SD = 0.21).

Figure 8 shows the SD of the proportion of reasons within the ICD-10 taxonomy across all wards in descending order; reasons with the highest variation of calls are at the top and
Figure 3. Bar plot displaying the distribution of risk rating scores (GT, top 25 reasons out of 170).
GT: granular taxonomy.

Figure 4. Bar plot displaying the distribution of risk ratings (GT, top 25 out of 170, as ranked by volume of calls).
GT: granular taxonomy.
those reasons with the lowest variation of calls are at the bottom. The five reasons that were least consistently cited across all wards were ‘Drugs: Legal’ (SD = 1.69), ‘Organic Mental Disorders’ (SD = 1.28), ‘Mental Disorder (non-specified)’ (SD = 1.13), ‘Delusional Disorders’ (SD = 1.11) and ‘Sexuality’ (SD = 1.08). The five reasons that were most consistently cited across all wards were ‘Family/Relationship Issues’ (SD = 0.19), ‘Mood/Affective Disorders’ (SD = 0.19), ‘Suicide Related Issues’ (SD = 0.21) and ‘Emotional Issues’ (SD = 0.22).
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\} \Rightarrow \{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\} \Rightarrow \{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\} \Rightarrow \{Suicide-related issues\} (Confidence: 0.88).

In this case, there are three reasons (antecedents) in this rule (which may have occurred in at least three 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

\{Miscellaneous/unspecified, mood/affective disorders\} \Rightarrow \{Suicide-related issues\} (Confidence: 0.84).

This rule has two reasons (antecedents) (which may have occurred in at least two earlier calls) with suicide-related issues as a consequent reason in a later call.

**Support.** The leading rule within the GT by support is

\{Depression\} \Rightarrow \{Anxiety\} (Support: 0.19),

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

The rule with the second highest support value is

\{Suicidal Intent\} \Rightarrow \{Depression\} (Support: 0.18).
### Table 1. Full results for association rule analysis of the GT and ICD-10 taxonomy (ranked by confidence).

| 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  |
| ICD      | {Miscellaneous/unspecified, mood/affective disorders}                | {Suicide-related issues}         | 0.16    | 0.84       | 1.43 | 7353  |
| ICD      | {Miscellaneous/unspecified, neurotic disorders}                      | {Suicide-related issues}         | 0.14    | 0.82       | 1.4  | 6546  |
| ICD      | {Miscellaneous/unspecified, neurotic disorders, suicide-related issues} | {Mood/affective disorders}       | 0.11    | 0.78       | 1.84 | 5074  |
| ICD      | {Miscellaneous/unspecified}                                         | {Suicide-related issues}         | 0.24    | 0.77       | 1.3  | 11,051|
| ICD      | {Mood/affective disorders, neurotic disorders}                       | {Suicide-related issues}         | 0.16    | 0.75       | 1.28 | 7356  |
| ICD      | {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  |
| Granular | {Information}                                                       | {Suicidal intent}               | 0.12    | 0.78       | 2.79 | 5404  |
| Granular | {Anxiety, suicidal intent}                                          | {Depression}                    | 0.11    | 0.77       | 2.01 | 5202  |
| Granular | {Information}                                                       | {Depression}                    | 0.11    | 0.74       | 1.92 | 5117  |
| Granular | {Suicidal intent}                                                   | {Depression}                    | 0.18    | 0.63       | 1.66 | 8238  |
| Granular | {Depression, suicidal intent}                                        | {Anxiety}                      | 0.11    | 0.63       | 1.61 | 5202  |
| Granular | {Anxiety, depression}                                               | {Suicidal intent}               | 0.11    | 0.58       | 2.08 | 5202  |
| Granular | {Mental health}                                                     | {Anxiety}                      | 0.11    | 0.58       | 1.48 | 4918  |
| Granular | {Mental health}                                                     | {Depression}                    | 0.10    | 0.56       | 1.47 | 4790  |
| Granular | {Suicidal intent}                                                   | {Anxiety}                      | 0.14    | 0.52       | 1.33 | 6741  |
| Granular | {Depression}                                                        | {Anxiety}                      | 0.19    | 0.50       | 1.28 | 8974  |

LHS: left-hand side; RHS: right-hand side; GT: granular taxonomy; ICD: International Classification of Diseases.
The leading rule in the ICD-10 taxonomy by support is

\{Miscellaneous/unspecified\} \geq \{Suicide-related issues\} (Support: 0.24),

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

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

\{Miscellaneous/unspecified, mood/affective disorders\} \geq \{Suicide-related issues\} (Support: 0.16).

**Lift.** The leading rule within the GT by lift is

\{Suicide: concerned about\} \geq \{Suicidal intent\} (Lift: 3.10).

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

The rule with the second highest lift value is

\{Information\} \geq \{Suicidal intent\} (Lift: 2.79).

The leading rule within the ICD-10 by lift is

\{Miscellaneous/unspecified, neurotic disorders, suicide-related issues\} \geq \{Mood/affective disorders\} (Lift: 1.84).

The rule that has the second highest lift value is

\{Miscellaneous/unspecified, neurotic disorders\} \geq \{Mood/affective disorders\} (Lift: 1.72).

**Discussion**

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

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

‘Suicide-Related Issues’ had the highest mean risk rating of all the reasons within the ICD-10 taxonomy and was the second highest in the number of calls. We also see this being reflected
within the GT, as ‘Suicidal Intent’, ‘Suicidal Ideation’ and ‘History of Suicide Attempts’ are the first, third and seventh highest (respectively) in risk rating. Interestingly, ‘Drugs: Legal’ had the second highest mean risk rating in the ICD-10, which constituted of two of the 25 top reasons within the GT. However, these reasons account for a relatively small proportion of calls and may only be made by a small number of clients. The reasons that have higher mean risk ratings have been shown to be associated to suicidal behaviour within the psychiatry literature. Within ICD-10, calls under the reasons ‘Personality Disorder’ and ‘Delusional Disorders’ showed a medium-to-moderate mean risk rating.

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. 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 levels. For instance, a study by Bunting et al. 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 per cent more deaths that 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. First, 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. Second, if one were to employ a broad taxonomy approach such as ICD-10 for predictive tasks, it can help to account for reasons which are highly specific but have fewer observations.
Limitations

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

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

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

Future work

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

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 article 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.

Acknowledgements

The authors would like to thank Contact NI, Lifeline and the Public Health Agency Northern Ireland for providing the service data.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval

Ethical approval for this research was granted by the Ethical Committee at Ulster University.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Department for the Economy for Northern Ireland.

ORCID iDs

Robin Turkington https://orcid.org/0000-0002-8853-6322
Maurice D Mulvenna https://orcid.org/0000-0002-1554-0785
Raymond R Bond https://orcid.org/0000-0002-1078-2232
Siobhan O’Neill https://orcid.org/0000-0002-8786-2118
Courtney Potts https://orcid.org/0000-0002-5621-1611
Edel Ennis https://orcid.org/0000-0002-9677-0725
Catherine Millman https://orcid.org/0000-0002-0257-3721

Supplemental material

Supplemental material for this article is available online.

References

1. Watson RJ, McDonald J and Pearce DC. An exploration of national calls to lifeline Australia: social support or urgent suicide intervention? Br J Guid Counselling 2006; 34(4): 471–482.
2. Burgess N, Christensen H, Leach LS, et al. Mental health profile of callers to a telephone counselling service. J Telemed Telecare 2008; 14(1): 42–47.
3. Middleton A, Gunn J, Bassilios B, et al. Systematic review of research into frequent callers to crisis helplines. *J Telemed Telecare* 2014; 20(2): 89–98.

4. Bassilios B, Harris M, Middleton A, et al. Characteristics of people who use telephone counseling: findings from secondary analysis of a population-based study. *Adm Policy Ment Health* 2015; 42(5): 621–632.

5. Kinzel A and Nanson JO. Education and debriefing: strategies for preventing crises in crisis-line volunteers. *Crisis* 2000; 21(3): 126–134.

6. Middleton A, Woodward A, Gunn J, et al. How do frequent users of crisis helplines differ from other users regarding their reasons for calling? Results from a survey with callers to Lifeline, Australia’s national crisis helpline service. *Health Soc Care Community* 2017; 25(3): 1041–1049.

7. Silla CN and Freitas AA. A survey of hierarchical classification across different application domains. *Data Min Knowl Disc* 2011; 22(1–2): 31–72.

8. World Health Organization. The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research, https://www.who.int/classifications/icd/en/GRNBOOK.pdf

9. Northern Ireland Department of Health. *Protect life 2: a draft strategy for suicide prevention in the North of Ireland*. Belfast: Northern Ireland Department of Health, 2016.

10. Ramsey C, Ennis E and O’Neill S. Characteristics of lifeline, crisis line, service users who have died by suicide. *Suicide Life Threat Behav* 2019; 49(3): 777–788.

11. British Association of Counselling and Psychotherapy. *Good practice in action 042: fact sheet: working with suicidal clients in the counselling professions*, https://www.bacp.co.uk/media/2157/bacp-working-with-suicidal-clients-fact-sheet-gpia042.pdf (2018, accessed 13 January 2019).

12. Turkington R, Mulvenna M, Bond R, et al. The application of user event log data for mental health and wellbeing analysis. In: *Proceedings of the 32nd international BCS human computer interaction conference (HCI 2018)*, Belfast, 4–6 July 2018. https://doi.org/10.14236/ewic/HCI2018.4

13. R Core Team. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2018.

14. Agrawal R and Srikant R. Fast algorithms for mining association rules. In: *Proceedings of the 20th international conference very large data bases (VLDB)*, vol. 1215, Santiago, Chile, 12 September 1994, pp. 487–499. San Francisco, CA: Morgan Kaufmann Publishers.

15. Hipp J, Günther U and Nakhaeizadeh G. Algorithms for association rule mining – a general survey and comparison. *ACM SIGKDD Explor News* 2000; 2(1): 58–64.

16. Turkington R, Mulvenna M, Bond R, et al. 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*, Belfast, 10–13 September 2019, pp. 196–202. New York: ACM.

17. Northern Ireland Statistics and Research Agency. *Geography fact sheet*, https://www.ninis2.nisra.gov.uk/public/documents/NISRA%20Geography%20Fact%20Sheet.pdf (2019, accessed 27 November 2019).

18. Brodsky BS, Malone KM, Ellis SP, et al. Characteristics of borderline personality disorder associated with suicidal behavior. *Am J Psychiatry* 1997; 154(12): 1715–1719.

19. Oldham JM. Borderline personality disorder and suicidality. *American Journal of Psychiatry* 2006; 163(1): 20–26.

20. González-Rodríguez A, Molina-Andreu O, Navarro Odriozola V, et al. Suicidal ideation and suicidal behaviour in delusional disorder: a clinical overview. *Psychiatry* J 2014; 2014: 834901.

21. Gould MS, Lake AM, Munfakh JL, et al. Helping callers to the National Suicide Prevention Lifeline who are at imminent risk of suicide: evaluation of caller risk profiles and interventions implemented. *Suicide Life Threat Behav* 2016; 46(2): 172–190.

22. Hunt T, Wilson CJ, Woodward A, et al. Intervention among suicidal men: future directions for telephone crisis support research. *Front Public Health* 2018; 6: 1.

23. Bunting B, Corry C, O’Neill S, et al. Death by suicide at the ward level in Northern Ireland. *Psychol Med* 2018; 48(8): 1375–1380.

24. National Institute for Health and Care Excellence. *Preventing suicide in community and custodial settings* (NICE Guideline 105), https://www.nice.org.uk/guidance/ng105 (2018, accessed 27 November 2019).

25. Kalafat J, Gould MS, Munfakh JLH, et al. An evaluation of crisis hotline outcomes. Part 1: nonsuicidal crisis callers. *Suicide Life Threat Behav* 2007; 37(3): 322–337.

26. Gould MS, Kalafat J, Harrismunfakh JL, et al. An evaluation of crisis hotline outcomes part 2: suicidal callers. *Suicide Life Threat Behav* 2007; 37(3): 338–352.