TITLE

The Repeated Episodes of Self-Harm (RESH) score: A tool for predicting risk of future episodes of self-harm by hospital patients

Authors and affiliations

Matthew J. Spittal[1]

Jane Pirkis[1]

Matthew Miller[2]

Gregory Carter[3]

David M. Studdert[1]

1. Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia
2. Department of Health Policy and Management, Harvard School of Public Health, Massachusetts, United States
3. Centre for Translational Neuroscience and Mental Health (CTNMH), Faculty of Health and Medicine, University of Newcastle, Australia.

Corresponding author

Dr Matthew Spittal, Melbourne School of Population and Global Health, The University of Melbourne, Parkville 3010, Victoria, Australia, +61 3 9035 8230 (phone), +61 3 9348 1174 (fax), m.spittal@unimelb.edu.au (email).

Keywords

Deliberate self-harm; Suicide; Epidemiology; Inpatient treatment; Risk assessment

ABSTRACT

Background: Repetition of hospital-treated deliberate self-harm is common. Several recent studies have used emergency department data to develop clinical tools to assess risk of self-harm or suicide. Longitudinal, linked inpatient data is an alternative source of information.

Methods: We identified all individuals admitted to hospital for deliberate self-harm in two Australian states (~350 hospitals). The outcome of interest was a repeated episode of self-harm (non-fatal or fatal) within 6 months. Logistic regression was used to identify a set of predictors of repetition. A risk calculator (RESH: Repeated Episodes of Self-Harm) was derived directly from model coefficients.

Results: There were 84,659 episodes of self-harm during the study period. Four variables – number of prior episodes, time between episodes, prior psychiatric diagnoses and recent psychiatric hospital stay – strongly predicted repetition. The RESH score showed good discrimination (AUC=0·75) and had high specificity. Patients with scores of 0-3 had 14% risk of repeat episodes, whereas patients with scores of 20-25 had over 80% risk. We identified five thresholds where the RESH score could be used for prioritising interventions.

Limitations: The trade-off of a highly specific test is that the instrument has poor sensitivity. As a consequence, the RESH score cannot be used reliably for "ruling out" those who score below the thresholds.

Conclusions: The RESH score could be useful for prioritising patients to interventions to reduce readmission for deliberate self-harm. The five thresholds, representing the continuum from low to high risk, enable a stepped care model of overlapping or sequential interventions to be deployed to patients at risk of self-harm.

INTRODUCTION

Efforts to date to use population-level data to predict suicide at the individual level have failed. The statistical properties of suicide—most notably, its relative rarity—thwart predictive models, even in high risk clinical populations when detailed socio-demographic and clinical information on individuals is available (Goldstein et al., 1991; Large et al., 2011; Sher, 2011). However, several recent studies have demonstrated that the same is not necessarily true of acts of deliberate self-harm (Bilen et al., 2012; Cooper et al., 2006; Steeg et al., 2012). Interest in forecasting self-harm has been spurred by growing recognition of its considerable morbidity burden. Hospital admission rates for self-harm are high and have been increasing in Australia (Spittal et al., 2012), the United States (Ting et al., 2012) and elsewhere (Kõlves et al., 2011); it is a strong risk factor for subsequent suicide (Hawton et al., 2012; Owens et al., 2002) and premature death from other causes (Bergen et al., 2012; Carter et al., 2005b); and repetition of self-harm by certain individuals is common, with international studies suggesting that up to 25 percent of patients hospitalised for self-harm will have had a hospitalisation for the same reason within the previous year (Owens et al., 2002). The much higher incidence of non-fatal episodes of self-harm, relative to fatal ones, opens the way for risk prediction that may guide prevention. Using a large sample of inpatients from Australian hospitals, we aimed to develop a robust risk score for identifying those at risk of repeated self-harm, to test the accuracy of this risk score and to demonstrate the potential impact if applied in the clinical setting.

METHOD

Setting and data sources

We assembled individual-level, linked data on all hospital admissions for self-harm and death records for suicide from two states in Australia (New South Wales and Western Australia) over a seven-year period (2001-07). These two states are geographically distinct and together have approximately 10 million residents (Australian Bureau of Statistics, 2011).

Specifically, we combined morbidity and mortality data from Australia's two premier data linkage agencies, based in New South Wales (NSW) and Western Australia (WA) (Holman et al., 1999; 2008). The Centre for Health Record Linkage in NSW maintains a linkage system that enables health records from core administrative datasets to be linked together at the individual level. This includes morbidity data from public and private hospitals and mortality data from the state death register. The Western Australian Data Linkage Database draws together individual-level health data from a similar range of administrative datasets. The agencies use similar methods to create datasets for specific analyses. Information is assigned to unique individual project-specific identification numbers using a combination of deterministic and probabilistic linkage techniques (Holman et al., 1999).

The agencies identified a cohort of individuals within their respective states who had been admitted to hospital for deliberate self-harm (ICD-10 codes X60-X84) and/or died as a result of deliberate self-harm (same ICD 10 codes). Next, for each individual in the cohort, the agencies extracted all inpatient admission records (including admissions unrelated to self-harm) and death records (including deaths due to causes other than deliberate self-harm).

The hospital admission data was extracted from the Admitted Patients Data Collection in NSW (for the period July 2000 to December 2009) and from the Hospital Morbidity Data Collection in WA (July 2001 to December 2009). Cause-of-death data was extracted from official death registry data in both NSW (January 2000 to December 2007) and WA (July 2000 to December 2007). Federal reporting requirements ensure that these data are coded in the same way and are comparable across states (Health Data Standards Committee, 2008).

Study dataset and variables

We constructed the study dataset at the level of episodes of hospital-treated self-harm ("episodes"). The individual patient identification number allowed us to thread together multiple episodes for any patient who had them, and to observe the dates of each episode in a sequence. We excluded duplicate mortality records from patients who died in hospital.

The outcome variable of interest in our analyses was repeated self-harm, defined as any subsequent episode of hospital-treated deliberate self-harm or suicide occurring within 6 months of any discharge from hospital for deliberate self-harm. Subsequent episodes within 6-months were coded "1"; episodes occurring outside the period were coded "0". To ensure that all living individuals could be observed over the periods of interest, we excluded observations where there was less than 6 months of observation time remaining.

We tested the following variables as candidate predictors of repeated episodes of self-harm: age (<35 years, $\geq$35 years), gender, number of previous episodes (i.e. before the current episode), marital status, method used in the index episode, time interval between index episode and any previous episode (1 to 60 days, 61 days to 1 year, greater than 1 year), and, respectively, diagnosed psychiatric disorders and inpatient psychiatric admission in the year preceding the index episode. Method used in the index episode was coded into eight categories, based on the ICD codes in the first external cause of injury field: poisoning (X60-X66, X68, X69), motor vehicle exhaust gas (X67), hanging (X70), drowning (X71), firearms (X72-X74), cutting/piecing (X78, X79), jumping (X80), and all other methods (X75-X77, X81-X84). Psychiatric disorders were coded into 7 categories, based again on ICD-10 codes: substance misuse disorders (F10-F19), schizophrenia and related diagnoses (F20-F29), mania (F30, F31), depression (F32-F39), anxiety (F40-49), eating disorders (F50), and personality disorders (F51-F59). These diagnosis variables were not mutually exclusive (i.e. patients could have multiple diagnoses). Apart from gender, all variables were time-varying and their values corresponded to the time of the episode.

Statistical analyses

'Best' predictors. To develop a risk prediction tool, we randomly split the sample into a 'test' sample (all data associated with 70% of individuals) and a 'validation' sample (the remaining 30% of individuals). This randomization did not result in any substantial imbalance between the two samples (as shown in **Table A1** of the supplementary appendix), whereas splitting the data according to state would have resulted in some imbalance. We fitted a series of multivariate logistic regression models. The outcome variable in all models was a repeated episode of self-harm occurring within 6 months of the index episode. The predictors were selected from among the variables described above.

Our objective was to assess which combination of covariates best predicted the outcome of interest. This was done in two steps. First, candidate sets of predictors were determined by fitting multivariate logistic regression models to the training sample, and focusing on which predictors had strong and significant relationships with the outcome. Second, the discrimination achieved by each candidate set of predictors was assessed by refitting that predictive model to the validation sample and calculating the area under the receiver-operating curve (AUC), a measure of model discrimination. All analyses were at the episode level and cluster-adjusted robust standard errors were used to account for individuals who had multiple episodes.

Risk score. We constructed a simple weighted scoring algorithm around the set of variables that best predicted repeated episodes, which we dubbed the "RESH" (**R**epeated **E**pisodes of **S**elf-**H**arm) score. The weights assigned to each item in the RESH score were indexed directly to the log odds ratios from the corresponding variables in the multivariate model. Application of the RESH score involves summing weights to produce a total score, ranging from 0 to 25. On each occasion a patient is admitted to hospital for self-harm, a fresh RESH score is calculated.

To assess calibration—that is, how closely RESH scores correlated with patient's actual risk of repeated episodes—we grouped patient admissions into six categories running from low to high across RESH scores, and calculated the actual percentage of subsequent episodes (hospital-treated self-harm or suicide) occurring within 6 months.

Accuracy statistics for the RESH. We calculated positive predictive values, sensitivity and specificity at five different decision cut points of the RESH score (≥ 1, ≥ 8 , ≥ 12, ≥ 16, ≥ 20).

All analyses were undertaken in Stata 13·1 (StataCorp, 2013).

Ethics approval

Approval for this study was granted by human research ethics committees at the University of Melbourne, the Department of Health Western Australia, and the Cancer Institute NSW.

RESULTS

Descriptive statistics

A total of 54 393 unique individuals engaged in 84 659 episodes of self-harm during the study period. The vast majority of episodes resulted in hospitalisations non-fatal self-harm (99%); only 1% of episodes were suicides (730 deaths) (Table 1, left column). The most common methods recorded were poisoning (74·1% of episodes), followed by cutting and piercing (17·8%) and hanging (2·5%).

Twenty-five percent (21 672 / 84 659) of the episodes met our study definition of repeated episodes (i.e. they occurred for individuals who had had at least one other episode in the previous 6 months) (Table 1, right column). Of these repeated episodes, 0·5% resulted in death (400 suicides).

Development of the risk score

Table 2 shows the set of four predictors that constituted the best-fitting model we could identify for predicting a repeated episode. The model produced AUC values of 0·749 (95% CI 0·745 – 0·754) in the training sample and 0·748 (95% CI 0·741 – 0·755) in the validation sample. The odds ratios and confidence intervals shown come from fitting this model to the training sample.

There was a clear dose-response between episode number and risk of future episodes. Compared to individuals with no prior episodes of self-harm, those with one prior episode had 30% higher odds of a subsequent episode within 6 months, and those with six or more prior episodes had 5 times higher odds. The proximity in time between the index episode and the previous episode was a risk factor for repetition, as was a psychiatric stay in the last year (OR, 1·52; 95% CI, 1·42 – 1·62). Of the five psychiatric diagnoses associated with repeated episodes, the presence of an eating disorder (as opposed to not having an eating disorder: OR, 1·81; 95% CI, 1·42 – 2·30) and the presence of a personality disorder (compared with not having a personality disorder: OR, 1·77; 95% CI, 1·62 – 1·93) were the strongest predictors of a subsequent self-harm episode. Inpatient psychiatric treatment within the prior 12 months was associated with increased odds of a subsequent self-harm episode.

Table 2 also shows the weights, derived from the log odds ratios in the best-fitting model, for use in the RESH score. Using these weights, instead of the actual log odds from the multivariable model, did not affect the model's performance on discrimination in the validation sample (AUC = 0·748; 95% CI 0·741 – 0·755).

Calibration of RESH score

In both the training and validation samples, increasing values of the RESH score were associated with increasing risk of self-harm (**Figure 1**). For instance, a patient whose RESH score fell between 0 and 3 had a 14% risk of another episode within 6 months and a patient whose RESH score fell between 20 and 25 had an 80% risk.

RESH score thresholds

Decisions regarding what scores should be used as trigger-points, or thresholds, for mounting interventions inevitably involve trade-offs between sensitivity and specificity; such clinical decisions will also be influenced by the cost, efficacy and intrusiveness of the intervention itself.

To illustrate these trade-offs, **Table 3** shows five possible thresholds on the RESH score that might be used to identify a target group for intervention, and the positive predictive value, sensitivity and specificity of scores at each threshold. At all but the lowest threshold (≥1) the RESH score had poor sensitivity. At a threshold of ≥8, however, the RESH score exhibited fair positive predictive value and excellent specificity, and at a threshold of ≥16 the RESH score was excellent on both measures: 82% of patients with RESH scores of ≥16 would be expected to go on to have another episode of self-harm within 6 months (positive predictive value) and 98% of patients who will not have another episode within 6 months will test have scores below this threshold (specificity).

DISCUSSION

This longitudinal study of hospital admissions for self-harm in two large Australian states shows that it is feasible to identify with reasonable accuracy patients at high-risk of repeated self-harm in the ensuing 6 months. Importantly, we use a definition of repeated self-harm that includes hospital-treated self-harm and suicide deaths, but excludes emergency department presentations without formal admission to the general hospital. The number of previous episodes of self-harm was the strongest predictor of repetition, but the inclusion of several other variables boosted predictive power. The characteristics of the RESH score suggest four out of five patients who record high scores truly are at high risk of future episodes of self-harm (positive predictive value) which might be considered clinically useful; on the other hand, a non-trivial proportion of all patients who will have future episodes will be missed (i.e., they will have misleadingly low scores – poor sensitivity). This will place some limits on the clinical usefulness of the scale. These considerations suggest that the RESH score could be useful in the context of a stepped care model of overlapping interventions. That is, recalculating the score after each re-admission for deliberate self-harm to inform treatment pathways.

We are not the first to develop a risk score for self-harm. Using data from patient presentations to several emergency departments, a UK-based team created a self-harm risk prediction algorithm called the Manchester Self-Harm Rule in 2006 (Cooper et al., 2006), and the ReACT Self-Harm Rule in 2012 (Steeg et al., 2012). A Swedish team developed a similar algorithm in 2012, the Södersjukhuset Self-Harm Rule (Bilen et al., 2012), based on data from two emergency departments in Stockholm. The predictors we identified as distinguishing individuals at high-risk of future episodes of self-harm from those with low-risk are broadly similar to predictors identified in the two most recent of these studies.

The ReACT Self-Harm Rule (Steeg et al., 2012) uses information on recent self-harm within the past year, use of cutting as a method, living alone or being homeless and current treatment for psychiatric disorders. Similarly, the Södersjukhuset Self-Harm Rule (Bilen et al., 2012) uses information on history of self-harm, current psychiatric treatment, use of benzodiazepine as a method and current psychiatric treatment as indicators of self-harm. The ReACT assigns equal weight to all predictors while the Södersjukhuset Self-Harm Rule assigns greatest weight to current psychiatric admission and the next greatest weight to prior history of self-harm. Our approach places the greatest weight on the number of previous self-harm episodes, as this has the strongest association with episode repetition in the model, followed by time since previous episode.

Our study advances previous work in this area in four ways. First, unlike previous studies, which have focused on emergency department presentations, our algorithm draws heavily on data from non-fatal hospital admissions. There are advantages and disadvantages to this alternative approach. Admissions are a smaller subset of all episodes of self-harm than emergency department visits. On the other hand, they are likely to be the most severe and costly episodes, and thus should arguably be especially high priority for prevention. Moreover, the duration and clinical contact associated with inpatient stays may open up some opportunities for intervention that emergency department visits do not.

Second, the predictors incorporated into previous algorithms are not weighted (Manchester/ReACT) or else weighted by direct application of log odds ratios to the predictors (Södersjukhuset). Using unweighted predictors has the advantage of simplicity for users, but results in loss of precision. Asking users to calculate risk by directly applying the logit transformation (Hosmer and Lemeshow, 2000) is likely to be too demanding. We took a middle line between these two approaches. The RESH score weights predictors by assigning them positive integers between 0 and 10, which come directly from the coefficients (log odds) on the same variables in the multivariable predictive model. The resulting tool is an algorithm that should be readily applicable at the beside. Diagnostics statistics suggested that little discrimination was lost in transforming coefficients in this way.

Third, the Södersjukhuset algorithm is based on data from two emergency departments, and the Manchester/ReACT algorithms each drew on presentations to five emergency departments. The RESH score comes from two state-wide health systems consisting of over 350 hospitals (Australian Institute of Health and Welfare, 2013). This difference in scale is reflected in sample size: our algorithm was based on a sample size three times larger than the UK studies and 160 times larger than the Swedish study.

Fourth, and most important, the structure of our predictive algorithm reflects different choices in the inevitable trade-off between sensitivity and specificity. The UK and Swedish self-harm algorithms display excellent sensitivity (i.e. patients who will self-harm again are readily identified as such) but poor specificity and positive predictive value (i.e. many patients who "screen positive" will not in fact self-harm again). Trading off high sensitivity for low specificity is desirable when the objective is to capture as large a proportion of the population with the outcome of interest as possible. This is valuable in some circumstances: for instance, when the tool functions as a first-step in a sequenced screening process, or when the intervention triggered by the tool is cheap, easy to administer and non-intrusive (e.g. information pamphlets, postcards or brief contact letters). On the other hand, high specificity/low sensitivity is a better balance when the algorithm is to be used as a more definitive screening tool, or is designed to target resource-intensive interventions. This is because tests with high specificity are useful for "ruling in" a person if they test positive (Akobeng, 2007). In the clinical context of our study – patients who had been admitted multiple times for self-harm and were at high-risk of readmission – it is difficult to identify anything but a relatively intensive and costly intervention proving efficacious. Nonetheless, the consequence of a highly specific test is that the RESH score will miss detecting many of those who truly have a repeated episode within 6 months, especially as RESH scores increase. This is because as the threshold for intervention increases, the balance between true positives and false positives shifts. For instance, at a threshold of ≥ 8, 44% of those testing positive would have a repeated self-harm episode as would 56% of those who tested negative. At a threshold of ≥ 12, the proportion of true positives reduces to 31% and the proportion of false positives increases to 69%.

The clearest illustration of the difference between the predictive properties of the RESH score and its predecessors is in combining information on the proportion of all patients who screen positive and the probability those patients have a repeated episode (the positive predictive values). With application of a moderate threshold (≥ 12) on the RESH score, 10% of all self-harm admissions would screen positive and 76% of these would experience another episode of deliberate self-harm within 6 months. It would seem feasible to be able to direct a suite of relevant after-care clinical interventions to 10% of the large population of hospital-treated self-harm patients. With application of the ReACT Self-Harm rule, 73-83% of all emergency department presentations for self-harm would be rated as moderate or high-risk, and only 30-37% of these will have another episode within 6 months. Even without considering the nature and cost-effectiveness of a particular intervention needed to be delivered to around 75% of the patients, it is plausible that a moderately effective and relatively expensive treatment and follow-up program may be cost-effective when 3 in 4 "treated" patients are in fact "afflicted" (based on the statistical properties of the RESH score), but implausible when only 1 in 3 are (Manchester/ReACT and Södersjukhuset).

We have intentionally avoided identifying a single threshold for intervention. Instead, we see different thresholds as being linked to different interventions that show promise for addressing the underlying causes of repeated self-harm. In this vein, we have identified five possible thresholds. A low threshold, for instance ≥ 8, may be appropriate for a low-cost, non-intrusive intervention such as postcards (Carter et al., 2005a) or brief contact letters (Motto and Bostrom, 2001). Higher thresholds, for example ≥ 16 or ≥ 20, may be appropriate for resource intensive interventions such as cognitive behavioural therapy or dialectical behaviour therapy targeting self-harm (Carter et al., 2010). Further work is needed to link risk assessment scores to staged interventions.

Limitations

Our study has several limitations, in addition to the sensitivity/specificity trade-off already discussed. First, we relied on inpatient data because in Australia there is no available information on deliberate self-harm in the emergency department setting. It is unknown how many emergency department presentations lead to admission to hospital or how this varies by hospital, location, severity of injury or other clinical characteristics. Nor do we know how inclusion of non-admitted episodes would affect our estimates of risk factors for repetition. Second, we have assumed that patients admitted to hospital represent the most serious cases. However, because people may leave the emergency department prior to assessment or do not disclose their intent, it is possible that some individuals at high risk of deliberate self-harm may go undetected. This would result in an increase the misclassification rates, although the extent of this bias is unknown. Third, the RESH score is derived from a large sample of patients admitted to hospital for deliberate self-harm in two Australian states. Its generalisability—to other settings (e.g. emergency department presentations that do not lead to admissions) and beyond New South Wales and Western Australia—requires further evaluation. Fourth, we did not consider whether prediction could be improved by allowing interactions between risk factors. We chose not to allow interactions because, while they may increase predictive power, this would come at the cost of complexity which could act as a barrier to adoption. Fifth, we were unable to ascertain the 'true' index episode for participants in NSW because no linked data was available prior to July 2000. (This was not the case in WA where we had access to data from the 1970s.) Thus, the effect of the number of prior episodes will be contaminated in a small number of cases where an individual has had a greater number of prior episodes than our data indicates. Finally, we had only a limited set of predictors available to us. We did not have access to reliable measures of, for instance, homelessness, indigenous status or detailed information on mental health diagnoses made outside the hospital system. Inclusion of this information might improve prediction. That said, users of a predictive algorithm at the clinical coalface may well confront more acute data availability issues than we did, rendering impractical all but a simple set of predictors.

Clinical implications

Current approaches to the treatment of deliberate self-harm in Australia and New Zealand already include efforts to assess risk. For example, clinical practice guidelines for the treatment of deliberate self-harm recommend a comprehensive clinical assessment of the risk of further self-harm, with the findings documented in the patient's medical records. These guidelines suggest the need for "organization of general hospital services to provide: emergency department admission; a safe environment; integrated medical and psychiatric management; risk assessment; identification of psychiatric morbidity, and adequate follow up" (Royal Australian and New Zealand College of Psychiatrists Clinical Practice Guidelines Team for Deliberate Self-harm, 2004). We propose that the RESH score be used alongside clinical assessment by psychiatric staff. The score is likely to be most useful in assisting with planning treatment or access to services after discharge from hospital. For example, patients with a high RESH score may be potential candidates for expensive and time-consuming interventions such as dialectical behaviour therapy delivered through outpatient services. However, the score should not replace detailed clinical assessments or clinical judgment and it should not solely determine the type of care delivered to the patient after release from hospital. The limits of risk assessment tools in psychiatry, which are based on grouped data, for predicting individual behaviour have been well documented (Large, 2013; Ryan et al., 2010).

We see two other potential uses for the RESH score. First, within the hospital setting, the RESH score may be useful for assisting medical staff in planning referral routes for patients. Second, because the RESH scoring items are potentially available in the electronic clinical records of each patient, it is possible that the database could provide the treating practitioner with a "red-flag", alerting them to the patient's RESH score. More generally, the score could be used to predict, at the hospital-level, the expected number of re-admissions in a given period of time, based on patient characteristics.

Conclusions

The RESH score performed well in identifying individuals who could be the target of interventions to reduce repeated deliberate self-harm. The ability of the RESH score to grade risk of self-harm means that the tool is flexible, and thresholds can be defined for low-cost, unobtrusive interventions as well as for high-cost, intensive interventions. The algorithm is highly specific and best used for as a tool for ruling in patients when they test positive. Further research is needed to validate the RESH score in other populations, to implement the tool into clinical practice and to match RESH scores to appropriate interventions.

REFERENCES

Akobeng, A.K., 2007. Understanding diagnostic tests 1: sensitivity, specificity, and predictive values. Acta Paediatica 96, 338–341.

Australian Bureau of Statistics, 2011. Australian Demographic Statistics. Catalogue No. 3101.0. Australian Bureau of Statistics, Canberra.

Australian Institute of Health and Welfare, 2013. How many hospitals are there? [WWW Document]. aihw.gov.au. URL http://www.aihw.gov.au/haag09-10/number-of-hospitals/ (accessed 6.13).

Bergen, H., Hawton, K., Waters, K., Ness, J., Cooper, J., Steeg, S., Kapur, N., 2012. Premature death after self-harm: a multicentre cohort study. The Lancet 380, 1568–1574.

Bilen, K., Ponzer, S., Ottosson, C., Castren, M., Pettersson, H., 2012. Deliberate self-harm patients in the emergency department: who will repeat and who will not? Validation and development of clinical decision rules. Emerg Med J.

Carter, G., Clover, K., Whyte, I.M., Dawson, A.H., D'Este, C., 2005a. Postcards from the EDge project: randomised controlled trial of an intervention using postcards to reduce repetition of hospital treated deliberate self poisoning. BMJ 331, 805.

Carter, G., Reith, D.M., Whyte, I.M., McPherson, M., 2005b. Non-suicidal deaths following hospital-treated self-poisoning. Aust N Z J Psychiatry 39, 101–107.

Carter, G., Willcox, C.H.C., Lewin, T.J.T., Conrad, A.M.A., Bendit, N.N., 2010. Hunter DBT project: randomized controlled trial of dialectical behaviour therapy in women with borderline personality disorder. Aust N Z J Psychiatry 44, 162–173.

Cooper, J., Kapur, N., Dunning, J., Guthrie, E., Appleby, L., Mackway-Jones, K., 2006. A clinical tool for assessing risk after self-harm. Ann Emerg Med 48, 459–466.

Goldstein, R.B., Black, D.W., Nasrallah, A., Winokur, G., 1991. The prediction of suicide. Sensitivity, specificity, and predictive value of a multivariate model applied to suicide among 1906 patients with affective disorders. Arch Gen Psychiatry 48, 418–422.

Hawton, K., Bergen, H., Kapur, N., Cooper, J., Steeg, S., Ness, J., Waters, K., 2012. Repetition of self-harm and suicide following self-harm in children and adolescents: findings from the Multicentre Study of Self-harm in England. J Child Psychol Psychiatry.

Health Data Standards Committee, 2008. National health data dictionary, Version 14. Australian Institute of Health and Welfare, Canberra.

Holman, C.D., Bass, A.J., Rouse, I.L., Hobbs, M.S., 1999. Population-based linkage of health records in Western Australia: development of a health services research linked database. Australian and New Zealand Journal of Public Health 23, 453–459.

Holman, C.D.J., Bass, A.J., Rosman, D.L., Smith, M.B., Semmens, J.B., Glasson, E.J., Brook, E.L., Trutwein, B., Rouse, I.L., Watson, C.R., de Klerk, N.H., Stanley, F.J., 2008. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. Aust Health Rev 32, 766–777.

Hosmer, D.W., Lemeshow, S., 2000. Applied logistic regression (Wiley Series in probability and statistics), 2nd ed. Wiley-Interscience Publication.

Kõlves, K., Vecchiato, T., Pivetti, M., Barbero, G., Cimitan, A., Tosato, F., De Leo, D., 2011. Non-fatal suicidal behaviour in Padua, Italy, in two different periods: 1992-1996 and 2002-2006. Soc Psychiatry Psychiatr Epidemiol 46, 805–811.

Large, M., 2013. The relevance of the early history of probability theory to current risk assessment practices in mental health care. History of Psychiatry 24, 427–441.

Large, M., Sharma, S., Cannon, E., Ryan, C., Nielssen, O., 2011. Risk factors for suicide within a year of discharge from psychiatric hospital: a systematic meta-analysis. Australian and New Zealand Journal of Psychiatry 45, 619–628.

Motto, J.A., Bostrom, A.G., 2001. A randomized controlled trial of postcrisis suicide prevention. Psychiatr Serv 52, 828–833.

Owens, D., Horrocks, J., House, A., 2002. Fatal and non-fatal repetition of self-harm. Systematic review. Br J Psychiatry 181, 193–199.

Royal Australian and New Zealand College of Psychiatrists Clinical Practice Guidelines Team for Deliberate Self-harm, 2004. Australian and New Zealand clinical practice guidelines for the management of adult deliberate self-harm. Aust N Z J Psychiatry.

Ryan, C., Nielssen, O., Paton, M., Large, M., 2010. Clinical Decisions in Psychiatry Should Not Be Based On Risk Assessment. Australas Psychiatry 18, 398–403.

Sher, L., 2011. Is it possible to predict suicide? Aust N Z J Psychiatry.

Spittal, M.J., Pirkis, J., Miller, M., Studdert, D.M., 2012. Declines in the Lethality of Suicide Attempts Explain the Decline in Suicide Deaths in Australia. PLoS ONE 7, e44565 EP –.

StataCorp, 2013. Stata: Release 13.1. StataCorp LP, College Station TX.

Steeg, S., Kapur, N., Webb, R., Applegate, E., Stewart, S.L.K., Hawton, K., Bergen, H., Waters, K., Cooper, J., 2012. The development of a population-level clinical screening tool for self-harm repetition and suicide: the ReACT Self-Harm Rule. Psychol Med 42, 2383–2394.

Ting, S.A., Sullivan, A.F., Boudreaux, E.D., Miller, I., Camargo, C.A., 2012. Trends in US emergency department visits for attempted suicide and self-inflicted injury, 1993-2008. Gen Hosp Psychiatry 34, 557–565.

Table 1: Characteristics of all episodes of deliberate self-harm and repeated episodes of self-harm within 6 months

Predictor	**Total episodes N = 84 659**			**Repeat episodes within 6 months *N* = 21 672**	
	n	%		*n*	%
Suicide					
No	83 929	99·1		21 272	0·3
Yes	730	0·9		400	0·5
Gender					
Male	33 318	39·4		7 198	21·6
Female	51 338	60·6		14 473	28·2
Age					
< 35 years	47 145	55·7		12 159	25·8
≥ 35 years	37 496	44·3		9 511	25·4
Marital status					
Never married	47 323	55·9		13 196	27·9
Widowed/divorced/separated	12 000	14·2		3 177	26·5
Married	20 531	24·3		4 306	21·0
Unknown	4 805	5·7		993	20·7
Method					
Poisoning	62 766	74·1		15 182	24·2
Motor vehicle exhaust	1 237	1·5		212	17·1
Hanging	2 135	2·5		439	20·6
Drowning	116	0·1		23	19·8
Firearms	118	0·1		25	21·2
Cutting/piercing	15 057	17·8		4 905	32·6
Jumping	592	0·7		176	29·7
Other methods	2 638	3·1		710	26·9
Time between episodes					
First attempt	54 393	64·2		7 991	14·7
1 to 60 days	14 171	16·7		7 738	54·6
61 days to 12 months	10 138	12·0		4 404	43·4
> 12 months	5 957	7·0		9 530	25·8
Number of prior episodes					
0	54 393	64·2		7 991	14·7
1	12 701	15·0		3 514	27·7
2	5 219	6·2		2 029	38·9
3	2 808	3·3		1 361	48·5
4	1 809	2·1		999	55·2
5	1 261	1·5		771	61·1
6 or more	6 468	7·6		5 007	77·4
Psychiatric diagnoses within last year					
Substance misuse disorder					
Yes	8 631	10·2		3 721	43·1
No	76 028	89·8		17 951	23·6
Psychoses					
Yes	4 501	5·3		1 976	43·9

No	80 158	94·7		19 696	24·6
Mania					
Yes	2 633	3·1		1 303	49·5
No	82 026	96·9		20 369	24·8
Depression					
Yes	12 066	14·3		5 695	47·2
No	72 593	85·7		15 977	22·0
Anxiety disorder					
Yes	10 531	12·4		5 562	52·8
No	74 128	87·6		16 110	21·7
Eating disorder					
Yes	482	0·6		254	52·7
No	84 177	99·4		21 418	25·4
Personality disorder					
Yes	8 500	10·0		5 742	67·6
No	76 159	90·0		15 930	20·9
Psychiatric stay in hospital within last year					
Yes	23 801	28·1		11 566	48·6
No	60 858	71·9		10 106	16·6

Table 2: Logistic regression results for repeated self harm within 6 months and scoring system derived from the log odds ratios, test sample

	Repeat within 6 months OR (95% CI)	Assigned weight
Number of prior episodes		
0 (ref)	1·00	0
1	1·30 (1·19-1·42)	1
2	1·73 (1·56-1·91)	3
3	2·33 (2·06-2·62)	5
4	2·59 (2·26-2·97)	5
5	3·21 (2·72-3·80)	7
6 or more	5·69 (4·96-6·53)	10
Time between episodes *		
1 to 60 days	1·84 (1·68-2·01)	3
61 days to 12 months	1·25 (1·14-1·37)	1
> 12 months (ref)	1·00	0
Psychiatric diagnoses in last 12 months		
Substance misuse disorder		
Yes	1·19 (1·11-1·28)	1
No (ref)	1·00	0
Depression		
Yes	1·23 (1·15-1·31)	1
No (ref)	1·00	0
Anxiety		
Yes	1·34 (1·25-1·44)	2
No (ref)	1·00	0
Eating disorder		
Yes	1·81 (1·42-2·30)	3
No (ref)	1·00	0
Personality disorder		
Yes	1·77 (1·62-1·93)	3
No (ref)	1·00	0
Psychiatric stay in last 12 months		
Yes	1·52 (1·42-1·62)	2
No (ref)	1·00	0

* First attempt not shown because the variable is coded to be perfectly collinear with 0 prior episodes.

Figure 1: Observed risk of deliberate self-harm based on Self-Harm Scores, test and validation samples

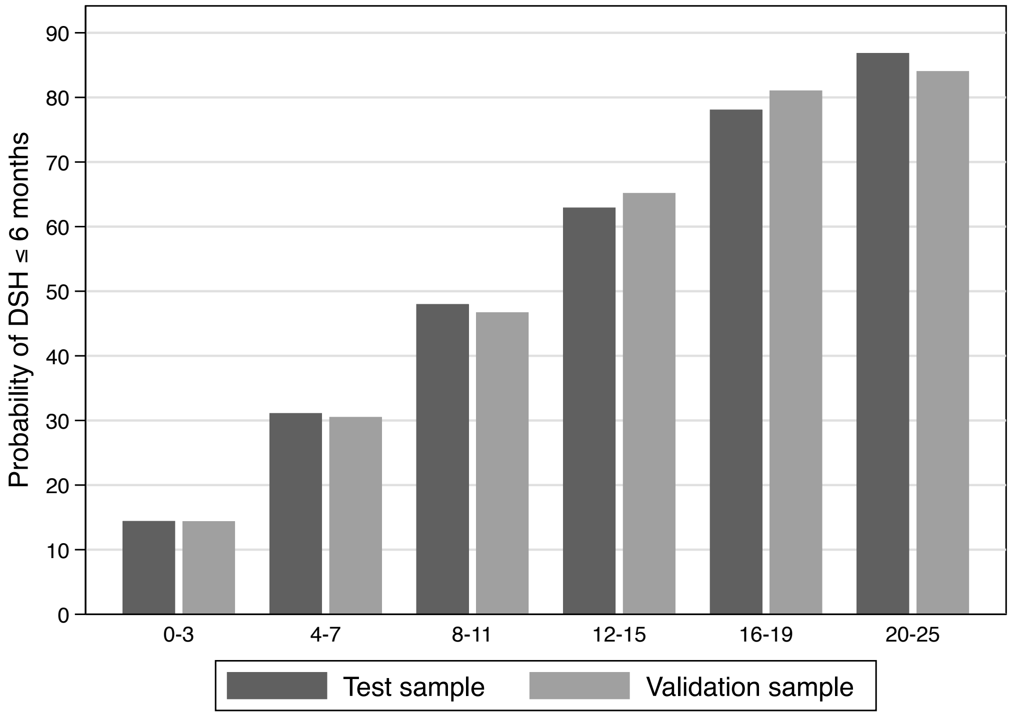

Table 3: Precision for 5 thresholds of the self-harm risk score (validation sample)

Threshold	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Number with score ≥ threshold*
≥ 1	0·74 (0·73 – 0·75)	0·62 (0·62 – 0·63)	0·40 (0·39 – 0·41)	39 981
≥ 8	0·44 (0·43 – 0·45)	0·92 (0·91 – 0·92)	0·64 (0·63 – 0·65)	15 081
≥ 12	0·31 (0·30 – 0·32)	0·97 (0·96 – 0·97)	0·76 (0·74 – 0·77)	8 949
≥ 16	0·21 (0·20 – 0·22)	0·98 (0·98 – 0·99)	0·82 (0·80 – 0·84)	5 462
≥ 20	0·06 (0·05 – 0·06)	1·00 (1·00 – 1·00)	0·84 (0·80 – 0·87)	1 656

* Based on calculations using the total sample