BJPsych The British Journal of Psychiatry (2010) 196, 243–244. doi: 10.1192/bjp.bp.109.068809

Short report

Repeat self-harm: application of hurdle models

Jennifer Bethell, Anne E. Rhodes, Susan J. Bondy, W. Y. Wendy Lou and Astrid Guttmann

Summary

Among those who present to the emergency department for self-harm, many will repeat. Self-harm repetition is an outcome of interest in both observational and intervention studies. However, few such studies analyse the number of repeat self-harm presentations. Here, hurdle models are introduced as a potentially useful statistical method for these analyses. Emergency department data from the Province of Ontario, Canada, are used to illustrate an example of implementing hurdle models and interpreting their results.

Declaration of interest

None.

About 16% of those who present to the emergency department for self-harm (self-poisoning or self-injury, irrespective of suicidal intent)[1] will return within a year.[2] Self-harm repetition increases the risk of suicide[3] and is an outcome of interest in observational[2] and intervention[4] studies. However, many such studies ignore multiple repetition. Often, those who present to hospital are followed-up for subsequent self-harm presentations and then categorised and analysed as repeaters or non-repeaters, disregarding the number of repeat presentations. This approach may obscure important clinical and cost differences. For example, an intervention trial aimed at reducing self-poisoning repetition found, after 2 years, no effect on the proportion of repeaters (21.2% *v.* 22.8% in the intervention and control groups respectively), but when considering the number of repeat presentations, the intervention halved the rate of repetition (rate ratio = 0.49).[5]

Disregarding the number of repeat presentations, despite potential differences, is likely partly attributable to the fact that repetition is not well suited for typical count models (Poisson or negative binomial regression) because of 'excess zeros' (e.g. the 84% of the sample that will not repeat within 1 year). Excess zeros is a source of overdispersion, where the observed variance exceeds that expected under the models' distributional assumptions.[6]

Here, an alternative method to analyse self-harm repetition is proposed: the hurdle model.[7] Hurdle models combine a binary (e.g. logit) model with a zero-truncated count (e.g. Poisson) model. For the self-harm repetition example, the first part tests factors associated with any repetition (repeaters *v.* non-repeaters) and the second part tests factors associated with the number of presentations (among repeaters). Population-based emergency department data are used to illustrate implementing and interpreting hurdle models. Hurdle models are also shown to be more informative than traditional binary analyses, but also adequately fit these data relative to some other count models.

Method

This is a population-based retrospective cohort study of 12- to 17-year-olds presenting to the emergency department for self-harm in Ontario, Canada. Data are from the National Ambulatory Care Reporting System (NACRS), covering a 7-year period (1 April 2002 to 31 March 2009). The data capture every emergency department visit; all legal residents are insured for acute and primary healthcare services and every hospital submitted NACRS emergency department data. The 2006 Ontario population of 12- to 17-year-olds was about 1 million.[8] Ethical approval was obtained from St Michael's Hospital.

Self-harm presentations were identified using ICD–10 criteria (intentional self-harm: X60–84).[9] Index episodes were identified as an individual's first during the study period. Anonymous identifiers on each record allowed follow-up of subsequent presentations. The exposure was in-patient admission resulting from the index episode, chosen to represent an important, well-defined aspect of clinical management. The outcome was repeat self-harm presentation within 1 year of the index episode, calculated from the emergency department or in-patient discharge date (as applicable). Individuals with less than 1 year of follow-up data (index episodes after March 2008 and those who died), were excluded. The data were analysed in SAS (version 9.1.3). First, two binary models were fitted: logistic regression, categorising the outcome as repeater or non-repeater; and survival analysis (Cox regression), using time to first repeat presentation as the outcome. Next, four count models were fitted: Poisson, negative binomial, Poisson hurdle and negative binomial hurdle. The outcome was the count of repeat self-harm presentations, incorporating random effects for hospital-level clustering.[10] Model fit was compared using Akaike and Bayesian information criteria (AIC and BIC),[11] where smaller values are better.

Results

The cohort included 10 937 individuals (8012 (73.3%) girls and 2925 (26.7%) boys), of whom 3546 (32.4%) were admitted at their index episode. Overall, 1325 (12.1%) made at least one repeat self-harm presentation within 1 year of their index episode (classified as repeaters), and this proportion was almost identical in the two exposure groups (12.2% and 12.1% among admitted and non-admitted respectively).

The binary models, logistic regression and survival analysis found no statistically significant association between admission and repetition (odds ratio (OR) 1.01, $P = 0.8309$; hazard ratio 1.01, $P = 0.8614$). The count models' AIC and BIC (Table 1) suggest substantial improvement in model fit from selecting the negative binomial, Poisson hurdle and negative binomial hurdle models over the Poisson model. Both fit indices favour the negative binomial hurdle model, demonstrating their flexibility in accounting for overdispersion from excess zeroes as well as other sources.[12] Interpreting the negative binomial hurdle model, similar to the binary analyses, the logit portion shows admission subsequent to the index episode was not associated with repetition (OR = 1.02, $P = 0.7269$). However, the negative binomial portion shows that, among repeaters, the estimated number of repeat presentations is lower among those admitted ($P = 0.0179$).

Table 1 Count model results for the association between in-patient admission and repeat self-harm presentation(s) within 1 year for 12- to 17-year-olds in Ontario, Canada

Count models	Coefficient (standard error)	*P*	AIC	BIC
Poisson	−0.0537 (0.0495)	0.2791	12317	12327
Negative binomial	−0.0606 (0.0697)	0.3854	10601	10613
Poisson hurdle			10940	10962
Logit	0.0345 (0.0637)	0.5889		
Poisson	−0.2527 (0.0832)	0.0027		
Negative binomial hurdle			10580	10605
Logit	0.0223 (0.0637)	0.7269		
Negative binomial	−0.2854 (0.1194)	0.0179		

AIC, Akaike information criterion; BIC, Bayesian information criterion.

Discussion

These results highlight the importance of considering the number of repeat presentations when studying self-harm. Others have already acknowledged the tendency for self-harm repetition studies to ignore multiple repetition and proposed alternative analyses, including recurrent event survival analysis[13,14] and multinomial logistic regression.[1] Here, we have shown that hurdle models are also an appropriate and useful statistical method. They are more informative than binary analyses because the investigator retains the 'repeaters *v*. non-repeaters' analysis while gaining the second part of the model (the number of subsequent self-harm presentations among repeaters). The hurdle model's two-part results have a similar substantive advantage over conventional count models, as well as the statistical advantage of accounting for overdispersion from excess zeros. Aside from hurdle models, zero-inflated models might also be considered for these data. Both are two-part models, address overdispersion from excess zeros, and tend to produce similar fits to the data, so the decision between them depends on the study's design and purpose.[12] Although hurdle models assume all those in the study sample are at risk of events, the zero-inflated models assume some will not experience any events because they are never at risk. Given that this sample was assembled from individuals who presented to the emergency department for self-harm, the assumption of the former seems more appropriate. The main weakness of hurdle models, however, is that unlike survival analysis, they require uniform follow-up. This drawback was not problematic here; each cohort member had 1 year of follow-up after their index episode and loss to follow-up was considered minimal. Yet, in other studies, accommodating variable follow-up times (e.g. from attrition, timing of analysis, or including fatal events as outcomes) may be an important consideration.

In this example, admission subsequent to a self-harm presentation did not influence the odds of repetition within 1 year, but it was associated with fewer repeat presentations (among repeaters). If this association is causal, it has important clinical and cost implications. It suggests that hospitalisation may be beneficial in reducing future self-harm episodes. Potential explanations involve service access, either during admission or as follow-up. However, without accounting for potential biases, any such substantive interpretations remain speculative. Further work is needed to assess this finding, accounting for severity, self-harm method and service use.

Ultimately, these results demonstrate that when studying self-harm repetition, incorporating the number of repeat presentations can be of value for policy, research and clinical practice. Hurdle models are one way of assessing these patterns.

Jennifer Bethell, MSC, Suicide Studies Unit, St Michael's Hospital, and Dalla Lana School of Public Health, University of Toronto; **Anne E. Rhodes**, PhD, Suicide Studies Unit, St Michael's Hospital, Dalla Lana School of Public Health and Department of Psychiatry, University of Toronto, and Institute for Clinical Evaluative Sciences; **Susan J. Bondy**, PhD, **W. Y. Wendy Lou**, PhD, Dalla Lana School of Public Health, University of Toronto; **Astrid Guttmann**, MDCM, MSc, Department of Paediatrics and Department of Health Policy, Management and Evaluation, University of Toronto, Institute for Clinical Evaluative Sciences, and Hospital for Sick Children, Toronto, Ontario, Canada

Correspondence: Jennifer Bethell, Suicide Studies Unit, St Michael's Hospital, 30 Bond Street, Toronto, Ontario, M5B 1W8, Canada. Email: bethellj@smh.toronto.on.ca

First received 24 May 2009, final revision 4 Nov 2009, accepted 2 Dec 2009

Funding

J.B. is supported by a studentship from the Ontario Mental Health Foundation. A.G. is supported through a salary award from the Canadian Institutes of Health Research. The data were accessed through the Institute for Clinical Evaluative Studies, which is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care.

Acknowledgements

The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement is intended or should be inferred.

References

1 Haw C, Bergen H, Casey D, Hawton K. Repetition of deliberate self-harm: a study of the characteristics and subsequent deaths in patients presenting to a general hospital according to extent of repetition. *Suicide Life Threat Behav* 2007; **37**: 379–96.

2 Owens D, Horrocks J, House A. Fatal and non-fatal repetition of self-harm. Systematic review. *Br J Psychiatry* 2002; **181**: 193–9.

3 Zahl DL, Hawton K. Repetition of deliberate self-harm and subsequent suicide risk: long-term follow-up study of 11 583 patients. *Br J Psychiatry* 2004; **185**: 70–5.

4 Evans J, Evans M, Morgan HG, Hayward A, Gunnell D. Crisis card following self-harm: 12-month follow-up of a randomised controlled trial. *Br J Psychiatry* 2005; **187**: 186–7.

5 Carter GL, Clover K, Whyte IM, Dawson AH, D'Este C. Postcards from the EDge: 24-month outcomes of a randomised controlled trial for hospital-treated self-poisoning. *Br J Psychiatry* 2007; **191**: 548–53.

6 Hilbe JM. *Negative Binomial Regression*. Cambridge University Press, 2007.

7 Mullahy J. Specification and testing of some modified count data models. *J Econometrics* 1986; **33**: 341–65.

8 Statistics Canada. Age and sex for the population of Canada, Provinces, Territories, Census divisions and Census subdivisions, 2006 Census. Statistics Canada, 2007 (http://www12.statcan.ca/english/census06/data/topics/Print.cfm?PID=88989&GID=773551&D1=0&D2=0&D3=0&D4=0&D5=0&D6=0).

9 World Health Organization. *ICD–10: International Statistical Classification of Diseases and Related Health Problems*. WHO, 1992.

10 Min Y, Agresti A. Random effect models for repeated measures of zero-inflated count data. *Stat Modelling* 2005; **5**: 1–19.

11 Lui WS, Cela J. Count data models in SAS. In *Proceedings of the SAS Global Forum 2008 Conference*. SAS Institute, 2008 (http://www2.sas.com/proceedings/forum2008/371-2008.pdf).

12 Rose CE, Martin SW, Wannemuehler KA, Plikaytis BD. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *J Biopharm Stat* 2006; **16**: 463–81.

13 Lilley R, Owens D, Horrocks J, House A, Noble R, Bergen H, et al. Hospital care and repetition following self-harm: multicentre comparison of self-poisoning and self-injury. *Br J Psychiatry* 2008; **192**: 440–5.

14 Reith DM, Whyte IM, Carter GL. Repetition risk for adolescent self-poisoning: a multiple event survival analysis. *Aust N Z J Psychiatry* 2003; **37**: 212–8.

https://doi.org/10.1192/bjp.bp.109.068809 Published online by Cambridge University Press