

Assessing the risk of self-harm in an adult offender population: an incidence cohort study

Mike Horton,¹ Nat Wright,² Wendy Dyer,^{3,4}
Alex Wright-Hughes,⁵ Amanda Farrin,⁵
Zanib Mohammed,² Jamie Smith,⁶ Tom Heyes,^{7,8}
Simon Gilbody⁹ and Alan Tennant^{1*}

¹Academic Department of Rehabilitation Medicine, Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK

²HMP Leeds Health Care Department, Leeds, UK

³Department of Sociology and Criminology, University of Durham, UK

⁴School of Arts and Social Sciences, Northumbria University, Newcastle upon Tyne, UK

⁵Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds, Leeds, UK

⁶North East Offender Health Commissioning Unit, Durham University Science Park, County Durham, UK

⁷Brampton Primary Care Medical Centre, Rotherham, UK

⁸Spectrum Community Health (CIC), Wakefield, UK, previously at Brampton Primary Care Medical Centre, Rotherham, UK

⁹Department of Health Sciences, University of York, York, UK

*Corresponding author

Declared competing interests of authors: Simon Gilbody is a member of the HTA CET Commissioning Board.

Notes: The correlation matrix is available following application to the authors. This provides the bivariate correlations for the entire data set.

Published October 2014

DOI: 10.3310/hta18640

Scientific summary

Assessing the risk of self-harm in an adult offender population

Health Technology Assessment 2014; Vol. 18: No. 64

DOI: 10.3310/hta18640

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Self-harm, which can be considered to be 'self-poisoning or self-injury, irrespective of the apparent purpose of the act', is common among prisoners, particularly female prisoners. Concerned about the rising incidence of self-harm in prisons, the prison service introduced a care-planning system in 2007 called ACCT (Assessment, Care in Custody, and Teamwork). The ACCT process effectively establishes an assessment and care pathway system (CAREMAP) for those deemed to be at risk. However, it does not incorporate a standardised diagnostic test to estimate the risk of future self-harm.

Objective

To identify one or more established questionnaires, or items from those questionnaires, that could be used to predict the risk of self-harm among prisoners or made into a screening instrument suitable to do so.

Method

We undertook a multistage prospective cohort study in three prisons in northern England in order to identify potential screening instruments and determine whether or not they could predict future self-harm. One prison was for women. Initially, a scoping exercise was undertaken to identify potential screening instruments. The focus was on instruments that could be administered by anyone in a relatively short time and had some evidence that they worked as intended.

A pilot study was used to refine the original study protocol and the final choice of instruments. The selected instruments were administered in a questionnaire pack. All prisoners assigned to an ACCT in the three prisons during the pilot study recruitment period were eligible for inclusion and were approached to consent. The questionnaire pack was administered by experienced prison researchers, or psychologists working in prisons, in a one-off interview. At the end of the interview, the prisoner was asked to comment on the ease of responding to the various instruments. Following this single assessment, all prisoners were followed up until release or for a maximum of 9 months. The follow-up consisted of identifying the number and nature of self-harm events that had occurred. In the case of prisoners who had been transferred, the Global Transfer Report of the NOMIS (National Offender Management Information System) was used to ascertain self-harm status. The information from the pilot was then used to refine the main study protocol and make final decisions about the questionnaires to be included.

In the main study, the chosen questionnaires were to be evaluated for both their robustness within a prison setting and their ability to predict future self-harm. In the former case, the instruments were assessed for (1) unidimensionality through a confirmatory factor analysis; (2) scalability through Mokken scaling, a non-parametric item response theory (IRT) approach; and (3) quantitative structure through fit of their data to the Rasch measurement model, a parametric IRT approach. The last investigation would also examine for differential item functioning where, at the same level of the construct being measured, response to items in the questionnaire would differ by group membership, for example by gender or sentencing status. Also, where items within a questionnaire offered more than two response options (i.e. polytomous items), the approach also allowed an investigation into whether or not the categories were working as intended (i.e. the transfer between categories was properly ordered along the trait being measured). All three approaches present with a variety of statistics to test underlying assumptions,

and whether or not the data fit the model in an appropriate manner, consistent with the attribute being tested. The questionnaire pack also included sociodemographic and sentencing data.

To test the predictive validity of the instruments, an area under the curve (AUC) analysis was used to examine the ability of instruments to predict future self-harm and their associated sensitivity and specificity. Cox proportional hazards regression models were then also used to examine the multivariate predictive ability of the scales and the various sociodemographic and sentencing factors included in the questionnaire and, as such, investigated the hazard rates for different a priori determined risk groups while adjusting for important baseline factors and accounting for the variable time to self-harm. A priori determined risk groups related to cut points associated with the likelihood of self-harm for each of the questionnaires administered as determined via AUC analysis.

The sample size for the study was primarily determined by the need to compare the AUC between each pair of self-harm screening instruments. A secondary requirement was to achieve the relevant degree of precision required by the psychometric analysis (Mokken scale and Rasch analyses) and the Cox proportional hazards regression model. An audit had revealed that approximately 20% of inmates were assigned to an ACCT in any given year. Other work had shown that up to one-quarter of women self-harmed during their current term in prison. Consequently, it was determined that a sample of 405 prisoners would be required to achieve 80% power to detect a difference of 0.1 between a diagnostic test with an AUC of 0.8 and another diagnostic test with an AUC of 0.9 using a two-sided z-test with a 5% significance level, with a self-harm prevalence rate of 20%. Thus, for the comparison of each pair of scales, a sample size of 405 was required. If the questionnaires were to be administered in a block design, as in the pilot, a sample of 840 would provide 420 prisoners who could be compared on any pair of screening instruments. To maximise the follow-up time for subjects, recruitment was to be undertaken prospectively for an 8-month period in each institution (e.g. in one institution approximately 720 ACCTs were opened in the previous year). This was thought to deliberately oversample (by approximately 80%, assessing about 1400 prisoners) to allow recruitment of 840 subjects with sufficient follow-up time for a reliable AUC analysis.

For the Rasch analysis, sample size was primarily concerned with the degree of precision of the estimate of items for any given scale. A sample size of 400 respondents for any given screening instrument would estimate the item difficulty within a scale, with α of 0.01, to within ± 0.3 logits. This is the minimum practical level of stability expected for most variables.

Finally, for the Cox proportional hazards regression analysis, it was estimated that 400 prisoners would provide > 99% power to detect a hazard ratio for self-harm of 2.72 between two a priori risk groups (identified within a diagnostic test), with a standard deviation (SD) of 10.0. This corresponded to an assumption that the risk of self-harm in a prisoner identified as at risk from a diagnostic test was 2.72 times that of a prisoner identified as a non-risk case from the same diagnostic test, and was selected to represent the smallest hazard ratio which could be detected given the available sample size and assumptions. The calculation also assumed a self-harm prevalence rate of 20%, and adjusted for correlation between the risk factor and other covariates.

Ethics approval was granted by the National Research Ethics Committee and the Ministry of Justice, with local approval from each local NHS research and development office.

Results

Scoping and pilot study

An initial search yielded 955 unique journal article records, from which 130 unique potential self-harm or suicide screening measurement instruments were identified. Following the application of the practical and psychometric inclusion criteria, 13 potential screening instruments remained. The majority of scales were removed as a result of inappropriate administration constraints (i.e. clinician-rated scales) or inappropriate

or unspecific scale content (i.e. a scale primarily focused on anger or suicide rather than self-harm). Potential scales were also removed if they were specifically to be administered only after a self-harm event had occurred or if they were deemed to be too long. Following discussions within an expert panel, eight instruments remained.

Of the 72 prisoners recruited to the pilot study, 50 (69%) were male. Age ranged from 18 to 62 years, interquartile range (IQR) was 23–39 years with a median age of 28 years. Once a routine had been established, no problems with the process of administering the questionnaire pack or the general logistics of running the pilot study were reported. The mean administration time of the questionnaire packs was 37 minutes, but the consensus from the prisoners was that they did not find the interview process burdensome or onerous.

At follow-up, 24 (33%) of the prisoners were still housed in the original prison, 26 (36%) had been released, 20 (28%) had been transferred and the status of two (1%) was not known. The mean valid follow-up time was 172 days (SD 100 days). During the follow-up period, 30 (40%) of the prisoners self-harmed, the majority within 6 months.

A number of implications were forthcoming from the pilot study.

1. The data collection process and study logistics worked well, so it was agreed that the process would remain largely the same for the main study.
2. Nevertheless, there was difficulty in trying to conduct all interviews within 72 hours post ACCT, so the time frame was extended up to 2 weeks.
3. Based on participant feedback and the views of the expert panel, a final set of five instruments (from the original eight) were selected for use in the main study. All five instruments were to be administered in a single questionnaire pack, so eliminating the need for a block design and reducing the sample size.
4. As the majority of self-harm events occurred during the first 6 months, the follow-up time was reduced to this length.
5. The follow-up protocol proved very successful, and reduced the need to oversample, such that the final sample size became 359 prisoners to provide 80% power to detect a difference of 0.1 between the AUC for two diagnostic tests, accounting for a 20% loss to follow-up rate. Alternatively, 475 prisoners would provide 90% power to detect such a difference.

Following the pilot, the five instruments for the main study were:

- Prison Screening Questionnaire (PriSnQuest)
- Self-Harm Inventory (SHI)
- Borderline Symptom List-23 (BSL-23)
- Clinical Outcomes in Routine Evaluation System – Outcome Measure (CORE-OM)
- Patient Health Questionnaire (PHQ-9).

Main study

In total, 450 prisoners with a mean age of 31.2 years consented to the study, of whom 26% were female. On average, they left full-time education at 15 years old, with over two-fifths leaving without qualifications of any sort. Almost half (49.4%) had children, but only one in seven (14.3%) reported receiving a visit during the previous 7 days. Just over half of the male prisoners were on remand, compared with just over one-fifth (22.6%) of females. The average tariff of those sentenced was 41 months, of which 14.7 months had been served.

Just over one-third of ACCTs had been initiated because of a known self-harm event and just over one-quarter of prisoners (27.8%) self-harmed during the follow-up period. Taken together, almost half (46.7%) of those entered into the study were reported to have self-harmed, either from their index ACCT, or subsequently.

All screening instruments showed some support for unidimensionality within this setting, and four out of five showed scaling criteria consistent with ordinal scaling, so verifying the validity of cut points (the exception being the CORE-OM). However, many showed gender bias and failure of quantitative structure when their data were fitted to the Rasch measurement model. Although a resolution was made in most cases, latent interval scale estimates also failed to show predictive value when applied within Cox proportional hazards regression models.

The Cox proportional hazards regression model found that several sociodemographic factors significantly predicted a shorter time until self-harm occurred, namely previous self-harm in prison, absence of dependence on alcohol, previous ACCT, age < 30 years and treatment with medication for any mental health problems, with prison also playing an important role. None of the scales themselves was significantly predictive of self-harm after the effect of the sociodemographic factors had been accounted for.

The failure of the candidate screening instruments to predict future self-harm, while disappointing, was not entirely unexpected and was accommodated in the original protocol by moving forward to the selection of individual items. There were 105 items in the candidate instruments, forming an item pool of potential risk indicators, together with other sociodemographic and sentencing criteria (e.g. on remand) which the Cox proportional hazards regression model had shown to be potentially useful predictors.

From this item set, those that were associated with future self-harm (at the 5% level) were considered for a predictive algorithm. It became immediately apparent that there were different indicators for men and women. From an odds ratio perspective, the strongest indicator for subsequent self-harm among men was 'During the last week I have hurt myself by cutting, burning, strangling, head banging, etc.' (4–6 times or daily or more often) and for women it was the item 'Cut yourself on purpose' (have you ever).

Bringing the indicators together in simple gender-specific formative indexes, weighted by their unadjusted odds ratio, gave an AUC for men of 0.716 and of 0.837 for women. For men, this gave a sensitivity of 68% and specificity of 64%, predictive power of a positive test of 40% and predictive power of a negative test of 85%. For women, it gave a sensitivity of 76% and a specificity of 83%, predictive power of a positive test of 68% and predictive power of a negative test of 88%.

It was also possible to create a low–medium–high risk classification for the risk of self-harm. Although the incidence of self-harm among those categorised as low risk is relatively low in both sexes, it is apparent that the male screening is less efficient than the female screening, as just 56.8% of male prisoners classified as high risk subsequently self-harmed, compared with 90% of women. Nevertheless, categorisation by level of risk could contribute to identifying appropriate care pathways and, given the strength of the negative test, support decisions to sign prisoners off from ACCTs. The gender-specific item sets form a single-page screening index which can be administered by any staff within a few minutes.

Conclusions

Just over one-quarter of the 450 prisoners enrolled in the study self-harmed in the follow-up period. As might be expected, females were more likely to self-harm. What was unexpected was a considerable difference in the rate of self-harm between the two male prisons. All the chosen scales failed, at the scale level, to be predictive of self-harm.

Identification of future risk of self-harming behaviour has long been a challenge in prisons, and professionals have often been unfairly criticised for not identifying risk, particularly when a prisoner self-harms following closure of an ACCT. In the case of serious incidents leading to the death of a prisoner, there is a high burden of investigation on prison professionals from their employing organisation, the coroner's inquest and the Prisons and Probation Ombudsman.

Given this, and the fact that gender differences were observed when individual items or sociodemographic factors were considered, potentially useful gender-specific screening instruments were derived, allocating prisoners to low, medium and high risk categories for subsequent self-harm. With high negative predictive values of the test, the instruments may be particularly useful in supporting signing off an ACCT, as well as providing potential guidance on allocation to different care pathways to prevent future self-harm.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: <http://www.nets.nihr.ac.uk/programmes/hta>

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 09/22/46. The contractual start date was in November 2010. The draft report began editorial review in April 2013 and was accepted for publication in February 2014. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health.

© Queen's Printer and Controller of HMSO 2014. This work was produced by Horton *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Editor-in-Chief of *Health Technology Assessment* and NIHR Journals Library

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the HTA Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Professor Aileen Clarke Professor of Public Health and Health Services Research, Warwick Medical School, University of Warwick, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Peter Davidson Director of NETSCC, HTA, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Professor Elaine McColl Director, Newcastle Clinical Trials Unit, Institute of Health and Society, Newcastle University, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Health Sciences Research, Faculty of Education, University of Winchester, UK

Professor Jane Norman Professor of Maternal and Fetal Health, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, University College London, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

Editorial contact: nihredit@southampton.ac.uk