

A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour: update 2002–8

R Whittington,^{1*} JC Hockenhull,² J McGuire,³
M Leitner,⁴ W Barr,¹ MG Cherry,² R Flentje,⁴
B Quinn,¹ Y Dundar² and R Dickson²

¹Health and Community Care Research Unit, University of Liverpool, Liverpool, UK

²Liverpool Reviews and Implementation Group, University of Liverpool, Liverpool, UK

³Clinical Psychology, University of Liverpool, Liverpool, UK

⁴Infotech UK Research (Medical Division of ER&IC Ltd), Cheshire, UK

*Corresponding author

Declared competing interests of authors: none

Published October 2013

DOI: 10.3310/hta17500

Scientific summary

Populations at high risk of engaging in violent behaviour

Health Technology Assessment 2013; Vol. 17: No. 50

DOI: 10.3310/hta17500

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

Scientific summary

Background

This review systematically examines the research literature published in the period 2002–8 on structured violence risk assessment instruments designed for use in mental health services or the criminal justice system. Violence is a major social problem and improved assessment of those who present an above-average risk is an important goal in the overall strategy for addressing the issue. Techniques for formally assessing individual and social risk factors have developed rapidly over the past two decades from a process of unstructured clinical judgement to one of structured assessment based on empirically tested instruments. A vast number of structured risk assessment instruments relating to violence in different populations have been developed over this period and attempts have been made elsewhere to summarise aspects of the literature relating to various instruments. This review adopted much broader inclusion criteria than previously used in order to capture and summarise data on the widest possible range of available instruments.

Objectives

The objectives of the review were to address two questions: (1) what features (i.e. population, instrument, outcome measure and design aspects) are associated with a risk assessment instrument score being significantly associated with a violent outcome? and (2) which risk assessment instruments have the highest level of predictive validity for a violent outcome?

Methods

Data sources

Evidence on the relationship between scores on a structured instrument and the occurrence of violence was identified using both a comprehensive search strategy to interrogate 19 bibliographic databases and the checking of reference lists of identified reviews. The database searches covered the period from 2002 to 2008.

Inclusion criteria

The inclusion criteria for papers were purposefully broad to capture as wide-ranging a selection of relevant studies as possible. Studies had to evaluate a structured risk tool and report an outcome measure of interpersonal violence either directly (e.g. reconviction for a violent offence) or indirectly through a proxy measure (e.g. a validated anger instrument). Participants had to be aged 17 years or over and either have a mental disorder, be an offender, or have committed an indictable offence (without necessarily having been prosecuted, e.g. pre-court diversion schemes, 'dating' violence self-reported purely in the context of the research study).

Data extraction

Data extraction was carried out independently by nine reviewers, with regular meetings to co-ordinate activity and to explicitly cross-check extracted data. Data from each included study ($n = 959$) relating to study design, sample, setting, type of intervention, type of outcome and whether or not a statistically significant outcome was reported were extracted into a predefined Statistical Product and Service Solutions database. Data from a subset of studies ($n = 65$) which reported area under the curve (AUC) statistics were independently extracted into a Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) spreadsheet by one reviewer and cross-checked.

Data synthesis

A series of bivariate analyses using either a chi-squared test or Spearman's rank-order correlation were conducted to explore possible sources of variance in whether or not an instrument was found to have good predictive validity (i.e. scores on the instrument were significantly related to a violent outcome). Data from the AUC subset studies were combined to provide estimates of mean validity.

Results

For the overall set of included studies ($n = 959$), 59% adopted a cross-sectional design and 32% were longitudinal with follow-up of a single group (median follow-up period of 20.2 months).

One-third of longitudinal studies lost 20% or more of their participants between the two time points and the median sample size at end point was 146.

Beyond issues of validity, less than half of the studies (44%) reported data on reproducibility (e.g. test-retest reliability) and less than 2% examined issues of clinical utility (e.g. acceptability to users).

Over half (52%) of the studies were conducted in the USA and over three-quarters (77%) were conducted in the USA, Canada or the UK.

The mean proportion of males in study samples was 79% and the median proportion of participants classified as 'white' and/or 'Caucasian' was 61%. The median age was 34.1 years. There were nearly 100 studies of female-only samples and over 500 studies of male-only samples.

Two-thirds of all studies were conducted with offenders who had either no formal mental health diagnosis (43%) or forensic samples with a formal diagnosis (25%). Fewer studies were conducted with non-indicted perpetrators (17%) and people with a formal mental health diagnosis but no offence history (10%). With mental health samples, mixed diagnoses were the most common group (50%) and with offender samples, mixed offences were most common (34%). Where studies focused on a single diagnostic group, schizophrenia (9%) and personality disorder (9%) were most common. Where studies focused on a single offence category, sexual violence (21%) and domestic violence (21%) were more common than general violence (12%).

The following structured tools were tested in the most studies: Psychopathy Checklist-Revised ($n = 192$ studies); Conflict Tactics Scale ($n = 56$), STATIC-99 ($n = 54$); Historical-Clinical-Risk-20 ($n = 51$); Violence Risk Appraisal Guide ($n = 45$); State-Trait Anger Expression Inventory ($n = 42$); Millon Clinical Multiaxial Inventory ($n = 40$); Minnesota Multiphasic Personality Inventory ($n = 36$); Beck Depression Inventory ($n = 35$) and Barratt Impulsiveness Scale ($n = 34$).

Most studies (82%) started and ended in the same setting. The most common setting was an offenders' institution such as prison (25% of study start points and 20% of study end points).

Most studies (83%) assessed violence outcomes using a scale score (51%) or official data such as reconviction rates (32%).

In terms of crude predictive validity, most studies (78%) reported a statistically significant ($p < 0.05$) relationship between the instrument score and a violent outcome. Only three variables were associated with the tendency to report such a relationship: prospective data collection ($\chi^2 = 4.4$, $p = 0.035$), number of people recruited ($U = 27.8$, $p = 0.012$) and number of participants at end point ($U = 26.9$, $p = 0.04$). With only three relevant variables (including two which were not independent), no multivariate analysis of validity predictors was conducted.

The mean AUC value in the subset of studies reporting such values ($n = 65$) was 0.69 [standard deviation (SD) = 0.08] and AUC values ranged from 0.44 to 0.88. For those instruments tested in more than one study reporting AUC values, the Broeset Violence Checklist achieved the highest mean AUC value (0.81), albeit over a very short (24-hour) period. For those instruments tested in five or more studies reporting AUC values, the relative AUC values were as follows: General Statistical Information on Recidivism (0.73); Violence Risk Appraisal Guide (0.72); Sex Offender Risk Appraisal Guide (0.71); Level of Service Inventory (0.69); Psychopathy Checklist-Revised (0.69); Historical-Clinical-Risk Management-20 (0.69); Psychopathy Checklist Screening Version (0.68); STATIC-99 (0.66); and Rapid Risk Assessment for Sexual Offender Recidivism (0.64).

Conclusions

A very large number of studies examining the relationship between a structured instrument and a violent outcome were published in this relatively short 7-year period. The general quality of the literature is weak in places (e.g. over-reliance on cross-sectional designs, high attrition in longitudinal studies, lack of information on clinical utility, little evidence on cross-cultural transferability, avoidance of follow-up from one setting to another between start and end points, over-reliance on self-report scales for both predictor and outcomes) and a vast range of distinct instruments have been tested to varying degrees. However, there is evidence of some convergence around a small number of high-performing instruments and identification of the components of a high-quality evaluation approach, including AUC analysis. The upper limits (AUC ≥ 0.85) of instrument-based prediction have probably been achieved and are unlikely to be exceeded using instruments alone.

Recommendations for future research

1. The small number of tools that already have demonstrable replicated efficacy should be tested out on a wider range of populations. This expansion should include empirical testing beyond North America and the parts of Europe where they have been extensively tested.
2. There should be a strong case made for expending significant effort on developing and testing any new risk assessment tools given the proliferation of tools developed over the past 15 years. There will always be a tension between a 'one size fits all' philosophy in which three to four dominant instruments with extensive empirical support are seen as suitable for all populations and a 'bottom-up' approach which recognises that many different tools (including those with minimal evidence) are needed to reflect the complexities of variations across populations.
3. Cross-sectional studies and/or studies relying purely on scale scores should be avoided. Too much of the existing literature is based on correlating a predictor and an outcome occurring simultaneously. This prevents the testing of any causal hypotheses and thus does not help in the development of theoretical frameworks for understanding violence. The problem is compounded when both the 'predictor' and the 'outcome' are measured using self-reported experiences recorded on a scale, as opposed to observable hostility or violence. The validity of such scale measures is lower than that of behavioural outcomes.
4. More studies should be conducted prospectively from hospital/prison to the community to examine the potential support of risk assessment tools for discharge/release decisions. While the prevention of intra-institutional violence is important in terms of protecting staff and other patients, it is the transition from hospital or prison to the community which is of most significance for the patient/prisoner and society at large. It is also more challenging to achieve effective prediction when moving from one environment to another and methodologically more difficult to keep track of participants. But clinical decisions on release or discharge are core issues faced by professionals and better research over this transition period is essential.
5. Clinical utility of those instruments with a strong evidence base in terms of predictive validity should be assessed to contextualise this information. While good predictive validity is a core component of an effective instrument, there are a number of other aspects which must be present for the instrument to

be considered entirely effective. These include the availability of a user manual, reasonable cost, available training, specified user competencies (including training and specified qualifications and skill levels), ease of use, appropriate administration time and recognition of protective factors. Some of these aspects can be studied as part of the overall research evaluation of specific tools.

6. The findings from the female-only studies should be examined and summarised separately. The pathways to violence and consequences following from it are likely to be different for females compared with males. Given the identification here of a large literature of female-only samples, there is scope for a powered analysis of this topic on its own to examine differences from the male-only samples.
7. A statistical procedure [similar to Cohen's *d* (e.g. Rice ME, Harris GT. The size and sign of treatment effects in sex offender therapy. *Ann N Y Acad Sci* 2003;**989**:428–40; discussion paper 41–5) for intervention effect size] should be developed for aggregating across AUCs. Mean AUCs were calculated for this study in order to aggregate across replications but this is a rather simplistic approach, especially when there are few studies. Effect size for intervention studies, drawing on means and SDs, is a more robust statistic and the research effort in the area of risk assessment would benefit from a similar approach.

Source of funding

Funding for this study was provided by the National Institute for Health Research Health Technology Assessment and Research for Patient Benefit programmes.

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Five-year impact factor: 5.804

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: www.hta.ac.uk/

This report

The research reported in this issue of the journal was commissioned and funded by the HTA programme on behalf of NICE as project number 08/101/99. The protocol was agreed in May 2009. The assessment report began editorial review in October 2012 and was accepted for publication in March 2013. 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 2013. This work was produced by Whittington *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 Health Sciences, 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

Dr Tom Marshall Reader in Primary Care, School of Health and Population Sciences, University of Birmingham, 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 Honorary Professor, Business School, Winchester University and Medical School, University of Warwick, 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 Professorial Research Associate, 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