

Executive summary

Outcome measures for adult critical care: a systematic review

JA Hayes¹
NA Black²
C Jenkinson³
JD Young³
KM Rowan^{1*}
K Daly⁴
S Ridley⁵

¹ Intensive Care National Audit and Research Centre, London, UK

² London School of Hygiene and Tropical Medicine, UK

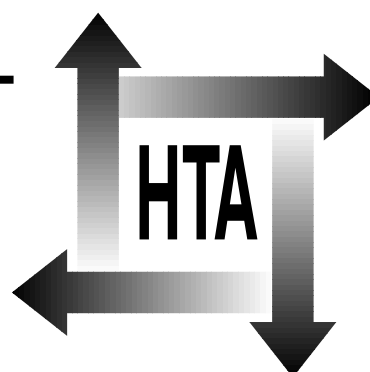
³ University of Oxford, UK

⁴ Guy's Hospital, London, UK

⁵ Norfolk and Norwich Hospital, UK

* Corresponding author

**Health Technology Assessment
NHS R&D HTA Programme**





Executive summary

Objectives

1. To identify generic and disease specific measures of impairment, functional status and health-related quality of life that have been used in adult critical care (intensive and high-dependency care) survivors.
2. To review the validity, reliability and responsiveness of the measures in adult critical care survivors.
3. To consider the implications for future policy and to make recommendations for further methodological research.
4. To review what is currently known of the outcome of adult critical care.

Methods

Data sources

- Searches of electronic databases (MEDLINE, EMBASE, CINAHL, PsycLIT, The Cochrane Library and SIGLE) from 1970 to August 1998.
- Manual searches of five journals (1985–98) not indexed in electronic databases and relevant conference proceedings (1993–98).
- Reference lists of six existing reviews, plus snowballing from reference lists of all relevant articles identified.

Study selection

- Randomised trials, non-randomised trials (cohort studies) and case series that included data on outcomes after discharge from adult (16 years and over) critical care.

Data extraction and synthesis

If reported, the following data were extracted from each paper:

- patient characteristics (age, gender, severity of illness, diagnostic category)
- number of patients eligible for study, follow-up period, number of deaths before follow-up, number and proportion of survivors included in follow-up
- method of presentation of outcome data – proportion normal as defined by reference values, or aggregate value (e.g. mean or median), or aggregate values plus an

indication of variance (e.g. standard deviation or inter-quartile range).

Evidence for three measurement properties was sought for each outcome measure that had been used in at least two studies – their validity, reliability and responsiveness in adult critical care. If the authors did not report these aspects explicitly, an attempt was made to use the data provided to provide these measurement properties. For measures that were used in at least ten studies, information on actual reported outcomes were also extracted.

Results

Measures used in critical care

- Measures of **impairment** were largely confined to the respiratory system so are almost certainly not appropriate for many critical care survivors. They can be categorised as respiratory volumes (e.g. vital capacity), gas flow within the respiratory system (e.g. forced expiratory volume in 1 second (FEV₁)), pulmonary diffusing capacity (e.g. carbon monoxide diffusing capacity) and visualisation of the upper airway (e.g. bronchoscopy). Multiple tests are often performed.
- Eight measures of **physical functional status** were used, five generic and three disease-specific. The most frequently used generic measures were multi-item scales. Two single-item global measures attempted to capture a person's overall activity level or functional status.
- Five multi-item measures of **mental functional status** were used, four generic and one specific to trauma patients. The generic measures were either confined to assessing depressive symptoms or also encompassed a measure of anxiety.
- Measures of **neuropsychological functioning** relate to a person's cognition, attention, ability to process information and memory. Apart from one single-item measure, which focused on communication level, six multi-item measures were used with critical care survivors. Such measures are particularly appropriate for use with survivors of head injury or other

neurological insult and, in that sense, they are disease-specific rather than generic measures.

- Single item measures of **recovery** were frequently used but researchers often invented their own, so there was little consistency in the wording. These measures had five principal foci – return to work, return to own home, degree of recovery, productivity and chronic health status. One multi-item scale was also used.
- Nine measures of **health-related quality of life** were used – although some of these multi-item generic measures encompass functional status also. The three used most extensively were the Sickness Impact Profile/Functional Limitations Profile (SIP/FLP), Perceived Quality of Life (PQOL) scale and Nottingham Health Profile (NHP). In addition, in recent years the Short Form 36 (SF-36) health survey questionnaire was increasingly used.

Assessment of outcome measures

- Overall, few attempts were made to determine the properties of any of the measures when used with critical care survivors and, in many instances, there was little scientific evidence of their properties outside critical care in other patient groups or in the general population. Lack of evidence does not mean these measures necessarily lack validity, reliability or responsiveness but does mean they should be used with caution and with an awareness of their possible inadequacies.
- There was little evidence as to the properties of **impairment** measures in critical care but considerable evidence in other categories of patients. Impairment measures are based on objective assessments using some equipment, the validity and reliability of which should be reported. There was some evidence for the criterion validity of the most commonly used measure of respiratory impairment (FEV_1), in that it correlates with measures of health-related quality of life.
- There was some evidence for the validity and responsiveness of two generic measures of **physical functional status**, Katz's Activities of Daily Living index (ADL) and the Karnofsky Index, but their reliability is unknown. Even less is known about disease-specific measures, although there was some evidence for the construct validity of the American Thoracic Society (ATS) respiratory disease questionnaire and the responsiveness of the New York Heart Association (NYHA) functional classification.
- Similarly, there was only limited information about the properties of the **mental functional**

status measures. There was some evidence for the criterion validity of all generic instruments and the responsiveness of the Centre for Epidemiology Studies Depression Scale.

- The only support for the **neuropsychological functional status** measures was some weak evidence for the criterion validity of the Trailmaking Tests and the Wisconsin Card Sorting Test (WCST).
- Assessment of the properties of measures of **recovery** was restricted to validity. Both the Glasgow Outcome Score and 'return to work' apparently had some construct and criterion validity. There were no published reports of reliability or responsiveness.
- Similarly, there was some evidence for the validity of **health-related quality-of-life** measures but nothing on their reliability or responsiveness in critical care survivors. This mirrors the state of affairs relating to assessment of measurement properties outside critical care. The validity of the SIP, PQOL scale and NHP in critical care appear to be reasonable but information on the SF-36, Spitzer's Quality of Life Index and other, less well-known generic measures was inadequate.

Health of critical care survivors

- Given the concerns expressed above on the limitations of the scientific worthiness of outcome measures used in critical care research, it was impossible to reach a valid and reliable overview of the health of survivors. There were huge differences in outcome between studies. This is not surprising given the variety of patients included, the failure to follow-up all survivors, differences in time of follow-up, lack of independent assessors and, often, poor presentation of data. Such criticism should not be seen as unique to this area of healthcare research.
- Comments (albeit tentative ones) are, therefore, limited to a few broad observations:
 - physical functional status appeared diminished during the first few months but may return to pre-admission levels by 6–12 months. Some degree of dependency in activities of daily living persisted in about half the survivors
 - more than 70% of survivors of working age returned to work, although their work activity may have altered
 - most survivors returned to their own homes within a few months
 - the most frequently diminished areas of health-related quality of life were those relating to work, recreation and sleep.

Conclusions

- The poor current state of knowledge of appropriate outcome measures for adult critical care survivors means that it is impossible to make clear recommendations as to which particular measures should be used. This partly reflects the large number of measures used in critical care research in the past. The evidence indicates that if the research community could agree on a limited list of measures from which to select for any given project, this would at least enable a considerable body of experience and knowledge to be built up around a few measures. In addition, it would allow investigators to make comparisons between studies and facilitate overviews based on secondary research of published results. To aid this, future researchers could confine their selection to the measures below until such time as clearer scientific evidence can distinguish between their relative merits.
- Measures of impairment appear to have limited value except, perhaps, in patients with respiratory disease. Their use in general adult critical care survivors is not recommended.
- Two generic measures of physical functional status appear the most relevant – Katz's ADL and the Karnofsky Index. Two disease-specific measures might also be considered in relevant subgroups: the NYHA functional class in cardiac patients and the ATS respiratory disease questionnaire in respiratory patients.
- Mental functional status is probably best assessed using Profile of Moods States or the Hospital Anxiety and Depression scales, as these cover anxiety in addition to depressive symptoms. In patients who are recovering from trauma, the Impact of Events Scale might also be considered.
- Neuropsychological function needs to be considered in post head-injury patients. There are no clear contenders but, on balance, the Trailmaking Tests and the WCST might be investigated initially.
- Measures of recovery offer few options. The Glasgow Outcome Score is the only multi-item scale available. In addition, standardisation of two single-item measures – return to work and residency or return to own home – would help to establish their usefulness.
- Health-related quality of life offers a wider range of possibilities. The three principal contenders (i.e. those most frequently used in critical care research) are the SIP/FLP, PQOL and NHP. It is suggested that the SF-36 is added to these, as it is being used increasingly often and widely in healthcare research and its measurement properties in other areas have been demonstrated.

Recommendations for further research

There is an urgent need for rigorous assessment of the measurement properties of all measures being used in critical care research. This work should be focussed initially on the leading measures outlined above. All studies that seek to assess the outcome of critical care by means of one of these measures should seek to explore at least one methodological aspect, for example, intra-rater reliability or construct validity. This approach would be more cost-effective than funding purely methodological studies.

Publication

Hayes JA, Black NA, Jenkinson C, Young JD, Rowan KM, Daly K, *et al.* Outcome measures for adult critical care: a systematic review. *Health Technol Assess* 2000;4(24).

NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Acute Sector Panel and funded as project number 95/55/03.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for the recommendations for policy contained herein. In particular, policy options in the area of screening will be considered by the National Screening Committee. This Committee, chaired by the Chief Medical Officer, will take into account the views expressed here, further available evidence and other relevant considerations.

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.

Criteria for inclusion in the HTA monograph series

Reports are published in the HTA monograph series if (1) they have resulted from work either prioritised by the Standing Group on Health Technology, or otherwise commissioned for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Series Editors: Andrew Stevens, Ken Stein and John Gabbay

Monograph Editorial Manager: Melanie Corris

The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
<http://www.ncchta.org>

ISSN 1366-5278