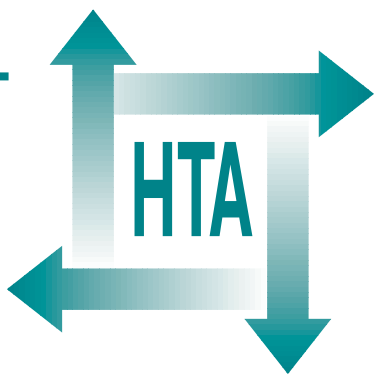
Health Technology Assessment 2000; Vol. 4: No. 24

Review

Outcome measures for adult critical care: a systematic review

JA Hayes
NA Black
C Jenkinson
JD Young
KM Rowan
K Daly
S Ridley

Health Technology Assessment
NHS R&D HTA Programme



Standing Group on Health Technology

Current members

Chair:
Professor Kent Woods
Professor of Therapeutics,
University of Leicester

Professor Martin Buxton
Director & Professor of
Health Economics,
Health Economics
Research Group,
Brunel University

Professor Shah Ebrahim
Professor of Epidemiology
of Ageing, University of Bristol

Professor Francis H Creed
Professor of
Psychological Medicine,
Manchester Royal Infirmary

Professor John Gabbay
Director, Wessex Institute
for Health Research
& Development

Professor Sir John
Grimley Evans
Professor of Clinical Geratology,
Radcliffe Infirmary,
Oxford

Dr Tony Hope
Clinical Reader in Medicine,
Nuffield Department of
Clinical Medicine,
University of Oxford

Professor Richard Lilford
Regional Director of R&D,
NHS Executive West Midlands

Dr Jeremy Metters
Deputy Chief Medical Officer,
Department of Health

Professor Maggie Pearson
Regional Director of R&D,
NHS Executive North West

Mr Hugh Ross
Chief Executive,
The United Bristol
Healthcare NHS Trust

Professor Trevor Sheldon
Joint Director, York Health
Policy Group, University of York

Professor Mike Smith
Faculty Dean of Research
for Medicine, Dentistry,
Psychology & Health,
University of Leeds

Dr John Tripp
Senior Lecturer in Child
Health, Royal Devon and Exeter
Healthcare NHS Trust

Professor Tom Walley
Director,
Prescribing Research Group,
University of Liverpool

Dr Julie Woodin
Chief Executive,
Nottingham Health Authority

Past members

Professor Sir Miles Irving*
Professor of Surgery,
University of Manchester,
Hope Hospital, Salford

Dr Sheila Adam
Department of Health

Professor Angela Coulter
Director, King's Fund,
London

Professor Anthony Culyer
Deputy Vice-Chancellor,
University of York

Dr Peter Doyle
Executive Director, Zeneca Ltd,
ACOST Committee on Medical
Research & Health

Professor John Farndon
Professor of Surgery,
University of Bristol

Professor Charles Florey
Department of Epidemiology
& Public Health, Ninewells
Hospital & Medical School,
University of Dundee

Professor Howard
Glennester
Professor of Social Science
& Administration, London
School of Economics &
Political Science

Mr John H James
Chief Executive,
Kensington, Chelsea &
Westminster Health Authority

Professor Michael Maisey
Professor of
Radiological Sciences,
Guy's, King's & St Thomas's
School of Medicine & Dentistry,
London

Mrs Gloria Oates
Chief Executive,
Oldham NHS Trust

Dr George Poste
Chief Science & Technology
Officer, SmithKline Beecham

Professor Michael Rawlins
Wolfson Unit of
Clinical Pharmacology,
University of Newcastle-
upon-Tyne

Professor Martin Roland
Professor of General Practice,
University of Manchester

Professor Ian Russell
Department of Health Sciences
& Clinical Evaluation,
University of York

Dr Charles Swan
Consultant Gastroenterologist,
North Staffordshire
Royal Infirmary

* Previous Chair

Details of the membership of the HTA panels, the NCCHTA Advisory Group and the HTA Commissioning Board are given at the end of this report.

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Outcome measures for adult critical care: a systematic review

JA Hayes[1]
NA Black[2]
C Jenkinson[3]
JD Young[3]
KM Rowan[1*]
K Daly[4]
S Ridley[5]

[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

Competing interests: none declared.

Published October 2000

This report should be referenced as follows:

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).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE. Copies of the Executive Summaries are available from the NCCHTA website (see overleaf).

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.

ISSN 1366-5278

Published by Core Research, Alton, on behalf of the NCCHTA.
Printed on acid-free paper in the UK by The Basingstoke Press, Basingstoke.

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

Contents

List of abbreviations

ADL	Activities of Daily Living [index]
ANOVA	analysis of variance
APACHE	Acute Physiology and Chronic Health Evaluation
APS	Acute Physiology Score*
ARDS	acute respiratory distress syndrome
ATS	American Thoracic Society
BAEP	brainstem auditory evoked potential
BDI	Beck Depression Inventory
CES–D	Centre for Epidemiological Studies – Depression [scale]
CHE	Chronic Health Evaluation
CI	confidence interval
CPR	cardio-pulmonary resuscitation
CT	computed tomography
D_LCO	carbon monoxide diffusing capacity
DSM III	Diagnostic and Statistical Manual III
EEG	electroencephalogram
FEV_1	forced expiratory volume in 1 second
FIV_1	forced inspiratory volume in 1 second
FLP	Functional Limitations Profile
FVC	forced vital capacity
FRC	functional residual capacity
GCS	Glasgow Coma Scale
GOS	Glasgow Outcome Scale
HAD	Hospital Anxiety and Depression [scale]
ICU	intensive care unit
IES	Impact of Event Scale
ISS	Injury Severity Score
MMEF	mean mid-expiratory flow
MMIF	mean mid-inspiratory flow
MMSE	Mini Mental State Examination
MOF	multiple organ failure*
N/A	not available*
NHP	Nottingham Health Profile
NS	not significant*
NYHA	New York Heart Association
PASAT	Paced Auditory Serial Addition Test
PEF	peak expiratory flow
PEFR	peak expiratory flow rate
PGWB	Psychological and General Well-being [index]
PIF	peak inspiratory flow
POMS	Profile of Mood States
PQOL	Perceived Quality of Life [scale]
RV	residual volume
SAPS	Simplified Acute Physiology Score*
SD	standard deviation
SEM	standard error of the mean
SF-36	Short Form 36 [Health Survey Questionnaire]
SIP	Sickness Impact Profile
TISS	Therapeutic Intervention Scoring System
T_L	transfer function
TLC	total lung capacity
VC	vital capacity
WCST	Wisconsin Card Sorting Test

* Used only in tables

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 (FEV_1)), 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.

Chapter 1

Background and methods

Introduction

The primary aims of healthcare are the reduction of mortality and morbidity, and the maintenance or improvement of functional capacity and quality of life. Traditionally the assessment of critical care has focused largely on mortality, although this is now changing. Assessment of the health of survivors has largely been confined to physiological, radiological and biochemical measurements of impairment. Recently there has been a move away from these objective measures towards subjective measures of functional status and health-related quality of life, with data collected directly from patients.[1] Interest in patients' perspectives in the evaluation of healthcare has led to the development of numerous subjective measures of functional status and health-related quality of life.

For the purposes of this review, adult critical care was confined to care provided in intensive care and high-dependency units. Units that fulfilled a specialised function, such as coronary care, burns and post-anaesthesia care, were excluded but intensive care units (ICUs) restricted to certain groups, such as surgical patients, were included.

The ideal outcome of healthcare is for the patient to return to their pre-existing state or to that expected for a person of the same age and medical condition.[2,3] Whether objective or subjective measures are used, they must provide information that is valid, reliable and responsive (see *Box 1*).

Outcome measures are divided here into impairment, functional status and health-related quality of life or well-being. The relationship between these can be seen in *Figure 1*. Impairment refers to objective measures of anatomical, physiological or biochemical aspects such as haemoglobin concentration or respiratory rate. These are the underlying features of ill-health that can be assessed or measured by another person, rather than the symptoms or problems that patients report. An impairment may or may not affect a person's health or functional status by giving rise to symptoms or limitations in their

BOX 1 Validity, reliability and responsiveness*

Validity

A valid assessment is one that measures what it claims to measure. The evaluation of the validity of a measure usually involves comparison with some standardised criterion or criteria. This is not easy in the social sciences, as there are rarely 'gold standards' against which measures can be compared. However, a number of standard criteria for validity are usually assessed for any properly constructed questionnaire.

- **Face validity** relates to whether the items on a questionnaire appear to be appropriate to the phenomenon being measured and to make sense, as well as being easily understood.
- **Content validity** relates to the choice of, and relative importance given to, items on a questionnaire. It is important that items appropriate to the phenomenon under investigation are chosen and, if they are weighted in some way, that the weights reflect the perceived level of difficulty or health problem.
- **Construct validity** is an important aspect of validity, especially when the variable being measured cannot be observed directly. It refers to when hypotheses are generated and a questionnaire is tested to determine if it actually reflects these prior hypotheses. For example, the construct validity of the SF-36 has been checked to ensure that certain groups (e.g. older, lower social classes, those with illnesses) would gain lower (i.e. worse) scores than other groups (e.g. younger, higher social classes, those without illnesses).
- **Criterion validity** relates to the ability of an instrument to correspond with other measures held up as gold standards. In practice, few studies can truly claim to have evaluated criterion validity, as gold standards are hard to find in this area of research.

Reliability

As with validity, there are a number of methods of assessing reliability.

- The most commonly used method is referred to as **internal reliability or internal consistency** and is measured using Cronbach's α statistic (for items with more than two response categories, such as 'never', 'sometimes', 'always').
- In **test–retest reliability**, the questionnaire is administered on two occasions separated by a few days. Ideally, respondents should not have changed in any way between the two administrations of the questionnaire and, consequently, the results should be almost identical.

Responsiveness

It is essential that evaluative instruments are able to detect change and the level of this change is interpretable in some way. The sensitivity to change or 'responsiveness' of an instrument is a very important criterion to consider when selecting measures.

The effect size statistic is the most commonly cited interpretation of change scores. It is usually calculated by subtracting the mean before treatment from that gained after treatment, and dividing the result by the baseline SD.

**Based on Jenkinson & McGee, 1998* [4]

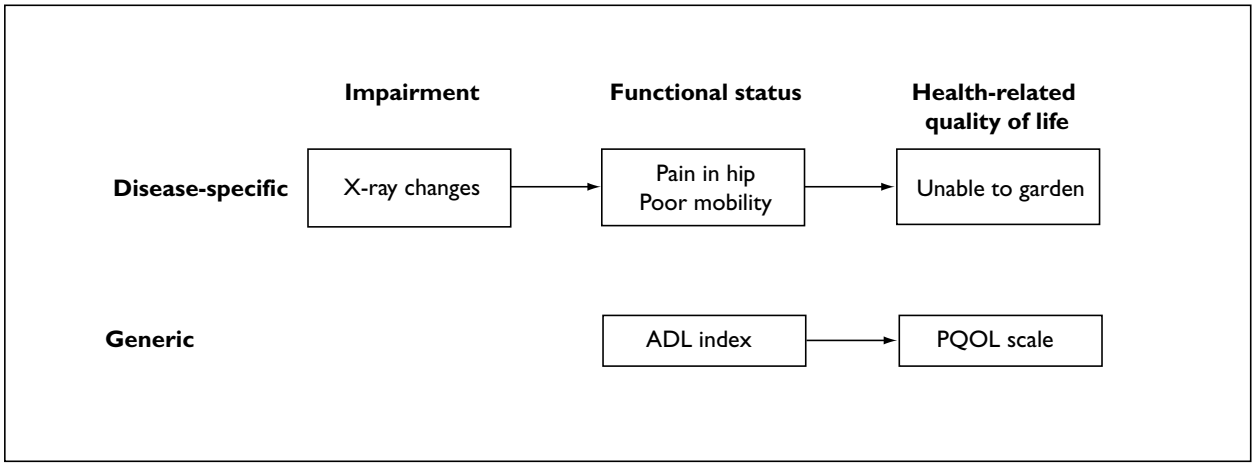


FIGURE 1 *Measures of outcome, taking osteoarthritis of the hip as an example*

ability to function. For example, low haemoglobin may be associated with breathlessness and an inability to walk to the shops. However, people may not report diminished functional status if it is not affecting their quality of life. Thus, an inability to walk to the shops may not diminish quality of life if the person has access to a car or someone else to shop for them. As illustrated in *Figure 1,* measures of each of these three dimensions may either be generic or specific to a particular condition.

As those patients who are admitted to critical care are a heterogeneous group, there is a need for generic outcome measures that can be used across a wide range of medical and surgical patients, as well as condition-specific ones. Some degree of overlap may exist, especially between functional status and health-related quality of life. For example, one of the most widely used generic measures (the Short Form 36 Health Survey Questionnaire (SF-36)) includes questions both on the functional status and the quality of life of respondents.

Most of the measures that are available and have been used in critical care are multi-item scales, that is, they are made up of several (or many) questions/items. Some multi-item scales not only provide a total score indicating the overall aspect being measured, for example, physical functional status, but also generate subscales that provide information on particular aspects, for example, mobility. However, not all measures are multi-item. Some single-item measures exist which generally consist of a global question that attempts to encompass the person's overall state of health. Both multi- and single-item measures are included in this review.

Aims and objectives of the review

The primary aim was to undertake a systematic review of the literature on the measurement properties of outcome measures which have been used with adults following discharge from critical care (intensive care and high-dependency units), in order to advise on the selection of appropriate measures for research and audit. A secondary aim was to report on the actual outcome of patients for those measures for which sufficient data were available.

The objectives were:

- 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 survivors
- to review the validity, reliability and responsiveness of the measures in critical care survivors
- to consider the implications for future policy and to make recommendations for further methodological research
- to review what is currently known about the outcomes of critical care survivors.

Management of the review

The review was designed and managed by a team comprising clinicians (Duncan Young, Kathy Daly, Saxon Ridley) and methodologists (Nick Black, Crispin Jenkinson, Kathy Rowan). Julie Hayes, a research fellow, undertook the literature searching, selection and initial reviewing of papers, and wrote the first draft

of the report. Nick Black, Duncan Young and Crispin Jenkinson undertook the redrafting. The team met on four occasions. At the first meeting, key decisions were made on refining the objectives, agreeing on search strategies and establishing eligibility criteria. At the second meeting, the initial results of the searches were reviewed. For practical reasons, it was decided to confine the review to the 36 measures that had been used on at least two occasions – it would be difficult to 'synthesise' the results from a single study! At the third meeting, it was decided that when addressing the fourth objective (see above), the review would be confined to measures that had been used on at least ten occasions in order to enhance the external validity of the conclusions.

Search strategy

The objective was to undertake a systematic and comprehensive search for all relevant studies in order to avoid bias by exclusion of any studies. This was undertaken using the five approaches detailed below.

The result of the search strategy was the identification of 764 potentially relevant articles and 93 potentially relevant conference abstracts.

1. Searching of electronic databases

Five electronic databases were searched systematically from 1970 (or from the start date of the database) to August 1998. Literature from before 1970 was ignored for two reasons: (i) it was considered that critical care had changed over the past few decades to such an extent that treatments used and outcomes achieved before 1970 would have little relevance today; (ii) there was little concern for, or study of, the functional status and health-related quality of life of survivors before the 1970s.

The following bibliographic databases were searched:

- MEDLINE (National Library of Medicine, USA; the electronic version of *Index Medicus*) using the search software Ovid, CD Plus
- EMBASE (Elsevier Science Publishers BV, The Netherlands; the electronic version of *Excerpta Medica*) using the search software Ovid, CD Plus
- CINAHL (CINAHL Information Systems, USA; Citation Index of the Nursing and Allied Health Literature) using the search software Ovid, CD Plus
- PsycLIT® (American Psychological Association; a subset of PsycINFO®, the electronic version of *Psychological Abstracts*) using the search software SilverPlatter
- The Cochrane Library (Cochrane Collaboration, UK: 1998 Issue 1, CD-ROM version)
- SIGLE (European Association for Grey Literature Exploitation, The Netherlands; System for Information on Grey Literature in Europe) using the search software SilverPlatter.

There was no limitation on the type of study design to be identified in the search. The search strategies developed were designed for maximal retrieval, using indexing terms and free text searching. The authors devised the following search terms, based on their knowledge and experience:

- intensive care, intensive therapy, high dependency, critical care, intermediate care, step-up care; step-down care
- outcome measure, follow-up, health status, functional status, clinical outcome, organ failure
- organ dysfunction; sequelae, quality of life, impairment, morbidity.

The identified studies were then restricted to 'human' and 'age group' > 16 or > 18 years when these options were available. No piloting of the search was performed. To test the repeatability of the search strategy, one database (MEDLINE) was searched twice, first in June 1998 and again in August 1998. During the second search, only newly published papers were identified.

2. Manual searching of journals

The on-line searches were supplemented by manual searching of the following journals that were not covered by the five electronic databases.

- *Critical Care Nursing Quarterly* (1990–95)
- *Intensive Care Nursing* (1985–91)
- *Intensive Care World* (1993–98)
- *International Journal of Intensive Care* (1995–96)
- *New Horizons* (1993–98)

In addition, the following conference proceedings were also searched manually for relevant abstracts.

- World Congress of Intensive and Critical Care Medicine (1985–97)
- European Congress on Intensive Care Medicine (1995–96)
- UK Intensive Care Society Meetings (1994–98)
- Society for Critical Care Medicine Educational and Scientific Symposium (1990–97)
- Australia and New Zealand Intensive Care Society Annual Scientific Meeting (1990–97)
- International Symposium on Intensive Care and Emergency Medicine (1993–98)
- Annual Congress of the European Society of Intensive Care Medicine (1997)
- Annual Meeting of the International Society for Quality of Life Research (1995)

Manual searching was undertaken from the present time working backwards until no further relevant articles or abstracts had been identified for two consecutive years.

3. Reference lists of selected reviews

The reference lists of six existing reviews were checked for additional references (Brooks, *et al.*, 1995;[5] Chelluri, *et al.*, 1995;[6] Brooks, 1996;[7] Heyland, *et al.*, 1998;[8] Schuster, 1998;[9] Thomas & Manara, 1998[10]).

4. Snowballing from references

When selected papers that met the inclusion criteria (see below) had been retrieved, their references were checked for additional references, which were then also retrieved.

5. Experts

A total of 20 researchers known to members of the review group were contacted by mail or email and asked if they were aware of any current research in the field of interest (12 responded). Information was also sought from members of international critical care electronic mailing lists (resulting in responses from eight people) and from seven national and international critical care societies (medical and nursing), of which three responded.

Selection of relevant references

Abstracts of all potentially relevant articles were obtained and considered for inclusion in the review if they met the following criteria:

- the study was based on adult (16 years and over) admissions to critical care, regardless of whether or not some younger cases were included
- data on outcomes after discharge from critical care were included
- study design was a case series, cohort study (non-randomised trial) or randomised trial
- outcome data for at least 20 patients were included.

Of the 764 potential journal articles, 144 met the inclusion criteria. Of these, 121 were identified through MEDLINE and 132 through EMBASE, of which 115 were duplicates. The remaining six papers were identified by snowballing. The 20 papers not in English were translated and included. Three papers were unobtainable.[11–13] The 144 papers provided data on 161 different outcome measures, although for 125 there was only one reported use. The latter are listed in appendix 1. The remaining 36 measures were used on at least two occasions and these formed the basis of the review.

In addition, 92 conference abstracts covering 69 outcome measures were identified, of which 13 were non-standard quality-of-life measures. The conference abstracts provided insufficient information to include them in the synthesis of the evidence.

Data extraction

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

- characteristics of patients (age, gender, severity of illness, diagnostic group)
- number of patients eligible for the study, time of follow-up, 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 (A) or aggregate value (for example, mean or median) (B) or aggregate values plus an indication of variance (for example, standard deviation (SD) or inter-quartile range) (C).

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 (see *Box 1*). If the authors of a paper had not reported on these aspects explicitly, an attempt was made to use the data they had provided to obtain these measurement properties.

Authors frequently reported on associations between functional status and patient characteristics, such as age and severity of illness. However, they rarely provided an explicit hypothesis as to whether or not they expected to see any association, so it was impossible to use such evidence to throw light on the construct validity of the measure in question. We have, therefore, simply reported on the existence, or not, of any such associations but have been unable to judge the measure's construct validity. Similarly, authors reported serial values for some measures over time, for example, follow-up at 3, 6 and 12 months. Again, it was impossible to judge whether stability or change over time indicated responsiveness of the measure or not.

For those measures that had been used in at least ten studies, information on the actual outcomes that had been reported were also extracted for the review.

Structure of the review

Measures of impairment are reported in chapter 2, functional status in chapters 3–6, and health-related quality of life in chapter 7. If a measure included both functional status and health-related quality of life, it has been included under the latter in chapter 7. The characteristics of each measure were considered in the same way:

- a brief description of the measure and its origins
- the measurement properties of the measure when used outside critical care (based on textbooks or review articles)
- the characteristics of the patients (age, gender, severity score and diagnostic group) included in its application in critical care and its mode of administration
- the measurement properties of the measure when used in critical care
- the actual outcomes that have been reported for critical care survivors (for those measures that were used in at least ten studies).

An overview is provided at the end of each chapter. In the final chapter the current state of knowledge on outcome measures in adult critical care is reviewed and suggestions made for their use in clinical practice and recommendations for their use in research and in audit, and for future methodological research.

Chapter 2

Measures of impairment

Measures of impairment have largely been confined to respiratory function and can conveniently be divided into four groups: respiratory volumes, gas flow within the respiratory system, pulmonary diffusing capacity and visualisation of the upper airway. Although the studies have been grouped this way for convenience, in practice multiple tests are usually performed. Often one test gives multiple results; for example, measurement of total lung capacity (TLC) requires the measurement of residual volume (RV) or functional residual capacity (FRC). Thus many studies are included in more than one group.

With the exception of visualisation of the upper airway, pulmonary function tests have been extensively validated outside intensive care. The association between lung disease and heavy industry, and the requirement to detect and grade respiratory impairment for both monitoring and compensation purposes, has led to a well-developed literature on the subject. Documents laying down standards for normal ranges and sources of variability are published regularly by professional organisations. In Europe, the European Respiratory Society documents are most widely used; the most recent was published in the *European Respiratory Journal* in 1993[14] and it is from this document that details of regression equations and reliability were taken for this report. In North America, the American Thoracic Society (ATS) has published a similar document.

Respiratory volumes

Vital capacity and forced vital capacity

The vital capacity (VC) is the volume change measured at the mouth between the positions of full inspiration and complete expiration. Forced vital capacity (FVC) is the volume of gas exhaled during a forced expiration starting from a position of full inspiration and ending at complete expiration. VC and FVC are measured using a spirometer or an integrating pneumotachograph.

VC and FVC are dependent on age, gender, race and body height. Both are reduced in conditions in which lung or chest wall compliance are impeded, such as pulmonary fibrosis or kyphoscoliosis, and are increased in athletes. Values for FVC and VC are either expressed as volumes or as percentages of predicted normal values for the patients' age, gender, race and stature. Alternatively, values may be expressed as a standardised residual (observed minus predicted divided by the residual SD of the regression equation used to generate the normal values), which can be converted to a probability of the observed value being part of a normal population.

Measurement properties outside critical care

Normal values are derived from regression equations specific for the individual's race and gender, containing the variables age and height. The residual SD gives a measure of the variability of the measurement not explained by race, gender, age and height. For Caucasians this is 0.6 l for men and 0.43 l for women. The average within-subject variation for a reference population is 148 ml (expressed as a coefficient of variation, the range is 0.3–11.4%). The absolute accuracy of the machines used to make the measurement should be ± 3.5% or ± 70 ml, whichever is greater.

Application in critical care

Nine papers[15–23] reported on FVC and three on VC.[22,24,25] The mean age of the participants ranged from 25 to 45 years (*Table 1*), and the proportion of males from 31% to 76%. Only two papers reported the severity of illness of the patients. Six studies followed-up patients with acute respiratory distress syndrome (ARDS). Mortality before follow-up ranged from 0% to 72% (*Table 2*). Further attrition was due to loss to follow-up and refusal to participate. As a result, the proportion of available subjects who were followed-up ranged from 24% to 100% (the latter was achieved in only one study). Planned or mean follow-up times ranged from 1 month to 5 years, with only two papers reporting on more than one follow-up.

The mean and either the SD or the standard error of the mean (SEM) of VC and FVC were reported in only two papers.[17,25] The remainder only reported the proportion of patients with 'normal' results.

Measurement properties in critical care

Validity Elliott and colleagues[17] and Peters and colleagues[20] both assessed the construct validity using the variable of age, although this was not explicit in the papers. Both studies found no statistically significant association with age but no conclusions can be drawn in view of the inadequate power of both studies.

Reliability Peters and colleagues[20] reported on re-using the same equipment and protocol but gave no details of the reliability of the instruments.

Responsiveness Ghio and colleagues[19] reported that impaired FVC was found in 76.5% of patients in their study after 1 month and 50% at 1 year.

TLC

This is the volume of gas in the lungs at the end of full inspiration. TLC is almost always calculated by measuring RV by helium dilution or FRC by body plethysmography, and adding VC or inspiratory capacity, respectively. It can also be measured directly radiologically or by body plethysmography. TLC is dependent on race, gender and body height, and is influenced by the same disease processes that alter VC, as well as by obesity.

Measurement properties outside critical care

The variability in TLC measurements depends on the variability of FRC and RV measurements. Within-subject variability averages 110 ml in normal individuals and 376 ml in patients with chronic obstructive airways disease. The measurement device (helium dilution) should have an accuracy of ± 50 ml or 5%. The body plethysmographic method has a coefficient of variation of 5% for within-subject measurements. The absolute accuracy cannot be determined, as FRC measurements require muscular effort and so cannot be simulated. The residual SD for the regression equation for TLC in normal individuals is 0.70l for men and 0.60l for women.

Application in critical care

Details of six papers that reported using this measure are presented in *Tables 1* and *2*.[17,18,20–22,25]

No data on TLC were presented in one paper.[22] In another paper, data were clearly presented, with mean values and SEMs, and in four papers data were presented as the percentages of the predicted values (*Table 2*). Elliott and colleagues[17,18] expressed their data as a percentage of predicted reference values, based on studies of healthy non-smoking adults performed using identical methods in the same laboratory.[26,27] In the study by McHugh and colleagues,[21] predicted values for TLC were calculated using the standards of Miller and colleagues.[28]

Measurement properties in critical care

Validity Construct validity using the variable of age was reported in three papers.[17,20,25] While in two studies no association was found with age, Elliott and colleagues[17] reported a significant

TABLE 1 *Characteristics of populations in studies using respiratory capacity*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Severity score, mean ± SD [median] (range)	Type of patient
Friman, *et al.*, 1976[15 a]	N/A	60	N/A	General
Halevy, *et al.*, 1984[24 b]	39.2 (18–68)	67	N/A	ARDS
Lund, *et al.*, 1985[16 a]	44 (18–87)	76	N/A	Burns
Elliott, *et al.*, 1987[17 ac]	25 (13–42)	31	N/A	ARDS
Elliott, *et al.*, 1988[18 ac]	28 (13–62)	N/A	N/A	ARDS
Ghio, *et al.*, 1989[19 a]	28.5 ± 12.5 (7–61)	51	N/A	ARDS
Peters, *et al.*, 1989[20 acd]	45.1 ± 15 (18–81)	67	N/A	ARDS
McHugh, *et al.*, 1994[21 ac]	41 (19–73)	62	N/A	ARDS
Grotz, *et al.*, 1997[25 bcd]	33.6 ± 2.1	70	ISS 36.8 ± 1.6	MOF; trauma
Jones, *et al.*, 1997[22 abcd]	[54] (17–90)	N/A	APACHE II [15] (3–34)	General
Law, *et al.*, 1997[23 a]	[51] (18–86)	68	N/A	General

[a] *Studies reporting FVC*
[b] *Studies reporting VC*
[c] *Studies reporting TLC*
[d] *Studies reporting RV*

TABLE 2 *Numbers of participants in studies using respiratory capacity*

Study	Number eligible for study (%)	Number of deaths before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean}, (range) (months)	Method of data presentation
Friman, *et al.*, 1976[15]	320	183 (57)	137	10 (7)	N/A	N/A
Halevy, *et al.*, 1984[24]	50	31 (62)	19	18 (95)	(3–30)	A
Lund, *et al.*, 1985[16]	41	24 (59)	17	17 (100)	{2.6} (0.7–4.9)	A
Elliott, *et al.*, 1987[17]	38	0 (0)	38	16 (42)	48	A + C
Elliott, *et al.*, 1988[18]	42	0 (0)	42	30 (71)	6	A
Ghio, *et al.*, 1989[19]	41	N/A	N/A	27 (N/A)	> 12	A
Peters, *et al.*, 1988[20]	166	119 (72)	47	39 (83)	(1–36)	A
McHugh, *et al.*, 1994[21]	216	134 (62)	82	20 (24)	3, 6, 12	A
Grotz, *et al.*, 1997[25]	173	104 (60)	69	50 (72)	{58.8}	C
Law, *et al.*, 1997[23]	109	39 (36)	70	20 (29)	> 6	A
Jones, *et al.*, 1997[22]	146	N/A	N/A	120 (N/A)	6	N/A

effect ($r^2 = 0.42$; $p < 0.01$) using linear regression analysis.

Reliability Peters and colleagues[20] reported using the same equipment and protocol but did not give details of the reliability of the instruments.

Responsiveness over time None of the papers reported on responsiveness.

RV and FRC

The RV is the volume of gas remaining in the lung at the end of a full expiration. The FRC is the volume of gas present in the lungs at the average end expiratory level. Measurement techniques were described above. Both RV and FRC are dependent on age, race, gender and height, and both are reduced by diseases which reduce lung or chest wall compliance. They are also reduced in unconscious patients.

Measurement properties outside critical care

The major sources of variability and their magnitude were discussed above. For a normal Caucasian population, the residual SDs for the regression equations for men are 0.4 l (RV) and 0.6 l (FRC), and for women 0.35 l (RV) and 0.5 l (FRC).

Application in critical care

The characteristics of the patient population studied in the three papers that reported using RV and or FRC are reported in *Tables 1* and *2*.[20,25,27] In only one paper[25] were data on RV and FRC presented as mean values with SEMs. Both Grotz and colleagues[25] and Peters and colleagues[20] reported comparative reference values for their data.

Measurement properties in critical care

None of the papers provided data on the measurement properties of RV or FRC.

Respiratory flow

Forced expiratory and inspiratory volumes in 1 second

The timed forced expiratory volume is the volume of gas exhaled in a specified time from the start of an FVC manoeuvre. By convention, 1 second is used and the measurement is symbolised as FEV_1. It can be measured using a spirometer or an integrating pneumotachograph. It is dependent on age, gender, race and height. It is an extensively used index of airflow limitation and is used to diagnose and monitor diseases such as asthma and chronic obstructive airways disease. FEV_1 is usually presented as a fraction of the FVC (FEV_1/FVC) (see below).

The timed forced inspiratory volume is the volume of air inhaled in a specified time following the beginning of a forced inspiratory manoeuvre. Again, by convention, 1 second is used and the measurement is symbolised as FIV_1. It is a measure of airflow limitation in the same way as FEV_1 but FIV_1 is claimed to be less sensitive to mechanical airway closure (as opposed to bronchospasm).

Measurement properties outside critical care

FEV_1 is an extensively used index with good reproducibility. The SD of repeated measurements

TABLE 3 *Characteristics of populations in studies using FEV_1*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Severity score, mean ± SD [median] (range)	Type of patient
Friman, *et al.*, 1976[15*]	N/A	60	N/A	General
Landercasper, *et al.*, 1984[37]	52 (7–87)	74	N/A	Trauma
Elliott, *et al.*, 1987[17]	25 (13–42)	31	N/A	ARDS
Elliott, *et al.*, 1988[18]	28 (13-62)	N/A	N/A	ARDS
Peters, *et al.*, 1989[20]	45.1 ± 15 (18–81)	67	N/A	ARDS
Ghio, *et al.*, 1989[19]	28.5 ± 12.5 (7–61)	51	N/A	ARDS
McHugh, *et al.*, 1994[21]	41 (19–73)	62	N/A	ARDS
Grotz, *et al.*, 1997[25]	33.6 ± 2.1	70	ISS 36.8 ± 1.6	MOF; trauma
Law, *et al.*, 1997[23]	[51] (18–86)	68	N/A	General
Jones, *et al.*, 1997[22*]	[54] (17–90)	N/A	APACHE II [15] (3–34)	General
Gammie, *et al.*, 1998[38]	Single lung transplant 36 Double lung transplant 38 (14–61)	Single lung transplant 61 Double lung transplant 33	N/A	Lung transplant

**These papers also reported on FIV_1*

on the same patients ranges from 60–270 ml. The long-term variability is, on average, 183 ml. The measurement device should be accurate to ± 3% or ± 50 ml, whichever is the greater. For a normal Caucasian population, the residual SD of the regression equation is 0.5 l for men and 0.38 l for women. No correlation was found between health status (measured using part 1 of the Nottingham Health Profile (NHP)) and FEV_1 in adults with cystic fibrosis,[29] although a substantial correlation was found in patients with chronic obstructive pulmonary disease.[30] Significant correlations between overall SF-36 scores, or scores from some of the nine subscales, and FEV_1 have been reported both for cross-sectional studies and for randomised studies of therapeutic manoeuvres in individuals with asthma and patients with chronic obstructive pulmonary disease.[22,30–36] As well as the SF-36, van der Molen and colleagues[33] reported on the correlation between FEV_1 and two asthma-specific quality-of-life measures (the Asthma Quality of Life Questionnaire and the Living with Asthma Quality of Life Questionnaire), and the Psychological and General Well-being index (PGWB). The correlations between FEV_1 and the quality-of-life measures were all very poor. No data are available on the variability of FIV_1 measurements or their correlation with quality-of-life measures.

Application in critical care

FEV1 was reported as an outcome measure in 11 papers (*Table 3*).[15,17–23,25,37,38] FIV_1 was also reported in two of the papers. Mean ages ranged from 25 years to 52 years and the percentages of men ranged from 31% to 74%. Acute severity was reported in only two studies. General critical care patients were the participants in three studies and patients with ARDS in five.

The numbers of individuals eligible to participate ranged from 38 to 320 (*Table 4*). Mortality before follow-up ranged from 0% to 72%; further attrition was due to loss to follow-up, refusal to participate or inability to return to the hospital. The proportions of available individuals who were followed-up ranged from 7% to 83%. Time points selected for follow-up varied from 1 month to 5 years. Four papers presented data on more than one follow-up.

In six studies the data were presented as the percentages of patients with normal results (*Table 4*). In two papers,[17,18] the authors cited Morris and colleagues[39] for reference values; Ghio and colleagues[19] and McHugh and colleagues[21] cited Crapo and colleagues,[27] and Ghio and colleagues also cited Gardner and colleagues.[40] Only two papers (Gammie, *et al.*, 1998;[38] Grotz, *et al.*, 1997[25]) presented the mean and SEM for FEV_1, and one[15] presented the mean and SD for FIV_1.

Measurement properties in critical care

Validity Peters and colleagues[20] reported no correlation between age and FEV_1. There was no assessment of validity reported in any of the other papers.

TABLE 4 *Numbers of participants in studies using FEV_1*

Study	Number eligible for study	Number of deaths (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean}, (range) (months)	Method of data presentation
Friman, *et al.*, 1976[15]	320	183 (57)	137	10 (7)	N/A	C
Landercasper, *et al.*, 1984[37]	62	13 (21)	49	21 (43)	{60} (6–144)	A
Elliott, *et al.*, 1987[17]	38	0 (0)	38	16 (42)	48	A
Elliott, *et al.*, 1988[18]	38	0 (0)	38	16 (42)	48	A
Peters, *et al.*, 1989[20]	166	119 (72)	47	39 (83)	(1–36)	A
Ghio, *et al.*, 1989[19]	41	N/A	N/A	27 (N/A)	> 12	A
McHugh, *et al.*, 1994[21]	216	134 (62)	82	20 (24)	3, 6, 12	A
Grotz, *et al.*, 1997[25]	173	104 (60)	69	50 (72)	{58.8}	C
Law, *et al.*, 1997[23]	109	39 (36)	70	20 (29)	> 6	A
Jones, *et al.*, 1997[22]	146	N/A	N/A	120 (N/A)	2, 6	N/A
Gammie, *et al.*, 1998[38]	58	25 (43)	33	16 (48)	48	C

Reliability Peters and colleagues[20] reported using the same equipment and protocol repeatedly but gave no details of the reliability of the measurements. None of the other papers reported on the reliability of FEV_1 or FIV_1.

Responsiveness Although serial tests were conducted by several authors, no-one reported on the responsiveness of the measure except Ghio and colleagues,[19] who provided values for FEV_1 at 1 month and 1 year. There were no reports of the responsiveness of FIV_1.

Outcome of critical care survivors

Three papers included data from overlapping groups of surviving patients collected between 1975 and 1986.[17–19] At 1 year following discharge, a group of 21 patients who had suffered from ARDS had a mean FEV_1/FVC ratio of 81% (range 0.64–0.93, 72–107% of predicted value). Three values were less than 80% of that predicted. There was no relationship between the severity of the ARDS and subsequent FEV_1/FVC ratio.[17] In a series of 30 ARDS survivors,[18] symptomatic patients had FEV_1/FVC ratios of 67–87% (51–90% predicted); no data were presented for asymptomatic patients. In a series of 41 survivors of ARDS who were followed-up for 1 year or more, 61% had an FEV_1 that was abnormally low and 33% had an FEV_1/FVC ratio that was abnormally low. The authors used the ATS cut-off of 80% or more of the predicted value to define normality.[19]

Two other series followed survivors of ARDS. In a series of 20 patients only one had an abnormal value for FEV_1/FVC ratio, although 18 had an abnormal FVC following extubation. At 6 months, five patients had an abnormal FEV_1/FVC ratio. In this study,[21] the changes in FVC rather than FEV_1 appear to be the major determinant of the FEV_1/FVC ratio. A study involving 39 survivors of ARDS variably followed-up showed that 79% had abnormal values for FEV_1 at 6 months or less after discharge, reducing to 48% on long-term follow-up. The FEV_1/FVC ratio on long-term follow-up of survivors averaged 80% (SD 8.4%).[20]

Three papers concentrated on patients who had received tracheostomies. In one study of 27 patients the FEV_1/peak expiratory flow (PEF) ratio, rather than the FEV_1/FVC ratio, was used. In patients with no respiratory history prior to ICU admission this was normal at follow-up, in patients with pre-existing chronic obstructive pulmonary disease it was reduced.[23] A study of 120 patients revealed that four had airflow restriction although no values were given.[22] A study of ten patients at various intervals revealed a mean FEV_1/FVC ratio of 76% (SD 4%) compared with a mean predicted value of 75%.[15]

Two studies followed trauma patients. In 50 patients with multiple injuries followed for various periods up to 5 years, the FEV_1/FVC ratio was normal with a mean value of 79%.[25] A study of 21 patients with flail chest followed-up between 6 months and 12 years did not give values but classified the FEV_1/FVC findings into 43% normal, 24% obstructive, 19% restrictive and 14% mixed.[37]

One study followed 38 patients after single and double lung transplants.[38] The FEV_1 decreased from a mean value of 88% at 6 months to 69% at 4 years for single lung transplants. For double lung transplants the FEV_1/FVC ratio was 71% at 6 months but remained static to 4 years.

It is difficult to interpret these results because the pre-morbid pulmonary function of the ICU survivors is unknown and a significant proportion of the patients might be expected to have some chronic respiratory history. When both FEV_1/FVC ratio and FVC were reported, abnormalities of FVC were more common, implying restrictive rather than obstructive pulmonary disease.

Ratio of FEV_1 to FVC

This ratio, usually written as FEV_1/FVC, is used as a guide to assess airway calibre. An obstructive ventilatory defect can be diagnosed as a decreased ratio.[41] When FEV_1 is examined in conjunction with FVC, two disease patterns of restriction and obstruction can be distinguished. In restrictive disease, such as pulmonary fibrosis, both FEV_1 and FVC are reduced but the ratio between them remains normal or increases. In obstructive pulmonary disease, such as asthma, FEV_1 is reduced out of proportion to reductions in FVC and, thus the ratio FEV_1/FVC is low. FEV_1/FVC has the advantage over FEV_1 alone in that the effect of FVC on FEV_1 is eliminated. The ratio is also virtually independent of body size and stature, but is age dependent.

Measurement properties outside critical care

The within-subject variability of FEV_1/FVC depends on the variability of the individual components, as described above. FEV_1/FVC ratios are often used to assess the response to bronchodilator drugs, a significant response (i.e. less than a 5% chance of the change in the variable being due to chance) being given as a 7.7–10.5% change in FEV_1 and a 5.2–10.7% change in FVC.

In studies of normal Caucasian populations, the residual SD of the regression equation (which only includes age as a variable) is 7.2% for men and 6.5% for women. The accuracy of the measurement devices is described above and on page 9.

Application in critical care

Eight papers were identified which had reported using FEV_1/FVC.[15,17–21,24,42] The mean reported ages in the seven studies reporting this variable ranged from 25 years to 45 years (*Table 5*). The percentages of male participants in these papers ranged from 31% to 67% in the seven studies that reported this variable. Acute severity scores were not reported in any of the studies. Six studies reported on patients with ARDS.

The numbers of participants in the studies are given in *Table 6*. The numbers of those eligible to participate ranged from 38 to 320. Mortality before follow-up was reported in seven papers and ranged from 0% to 72%; further attrition was due to loss to follow-up, and refusal and inability to attend outpatient appointments. The proportions of available individuals who were followed-up ranged from just 7% to 95%. Follow-up time was reported in seven papers. Two papers assessed patients at 6 months[18,21] and two assessed patients at multiple time points.[19,21] Halevy and colleagues[24] also reported using serial testing but did not specify the time frames of these; instead they presented a range of months from which patients were assessed.

In three papers the data was presented clearly as mean values with SDs or SEMs and in three as percentages of predicted values.

TABLE 5 *Characteristics of populations in studies using FEV_1/FVC*

Study	Mean age ± SD (range) (years)	Male (%)	Type of patient
Aass, 1975[42]	(1 day–89 years)	59	General
Friman, *et al.*, 1976[15]	N/A	60	General
Halevy, *et al.*, 1984[24]	39.2 (18–68)	67	ARDS
Elliott, *et al.*, 1987[17]	25 (13–42)	31	ARDS
Elliott, *et al.*, 1988[18]	28 (13–62)	N/A	ARDS
Ghio, *et al.*, 1989[19]	28.5 ± 12.5 (7–61)	51	ARDS
Peters, *et al.*, 1989[20]	45.1 ± 15 (18–81)	67	ARDS
McHugh, *et al.*, 1994[21]	41 (19–73)	62	ARDS

TABLE 6 *Number of participants in studies using FEV_1/FVC*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (range) (months)	Method of data presentation
Aass, 1975[42]	79	36 (46)	43	32 (74)	(9–48)	N/A
Friman, *et al.*, 1976[15]	320	183 (57)	137	10 (7)	N/A	N/A
Halevy, *et al.*, 1984[24]	50	31 (62)	19	18 (95)	(3–30)	A
Elliott, *et al.*, 1987[17]	38	0 (0)	38	16 (42)	48	C
Elliott, *et al.*, 1988[18]	42	0 (0)	42	30 (71)	6	C
Ghio, *et al.*, 1989[19]	41	N/A	N/A	27 (N/A)	> 12	A
Peters, *et al.*, 1989[20]	166	119 (72)	47	39 (83)	(1–36)	C
McHugh, *et al.*, 1994[21]	216	134 (62)	82	20 (24)	1, 3, 6	A

Measurement properties in critical care

Validity Some idea of construct validity could be gleaned from two papers that used linear regression to examine the association with age.[17,20] No correlation was found.

Reliability Peters and colleagues[20] reported using the same equipment but did not give details of the reliability of the instruments.

Responsiveness Halevy and colleagues[24] reported using serial tests but did not present any statistical analysis. Ghio and colleagues[19] reported changes in impairment between 1 month and 1 year but did not test the statistical significance of the changes.

Flow volume curves (or loops)

Flow volume loops are an x/y plot of lung volume against airflow during a maximal inspiration from RV to TLC and then a maximal expiration back to RV. The test is commonly used to diagnose extrathoracic airflow obstruction; in critical care it is usually used to determine if cannulation of the upper airway has produced stenosis. When used to diagnose extrathoracic airway obstruction, the test produces no numeric result and interpretation is based on pattern recognition.

The flow volume loop can also be used to measure the peak expiratory flow rate (PEFR) and the maximal mid-expiratory and inspiratory flow rates (MMEF and MMIF; see below).

Measurement properties outside critical care

As flow volume loops do not produce a numerical result, normal ranges and variability cannot be given. A study in which two observers reported loops on two occasions gave between-observer kappa scores of 0.58 and 0.68; the agreement within-observers over time values were 0.5 and 0.46, respectively.[43]

Application in critical care

The mean ages of the participants ranged from 28 to 58 years (*Table 7*).[16,17,22,44] The numbers eligible to participate ranged from 41 to 150 (*Table 8*). Mortality before follow-up ranged from 0% to 59% in the three papers reporting on this variable. The proportions of available individuals who were followed-up ranged from 69% to 100%. Only one paper reported serial follow-up times. Reported follow-up times ranged from 1 month to 96 months.

Stauffer and colleagues[44] presented data for assessment of airway narrowing as the number and percentage of survivors with stenosis, and

TABLE 7 *Characteristics of populations in studies using flow volume curves*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Severity score [median] (range)	Type of patient
Stauffer, *et al.*, 1981[44]	58 (17–88)	73	N/A	General
Lund, *et al.*, 1985[16]	44 (18–87)	76	N/A	Burns
Elliott, *et al.*, 1988[18]	28 (13–62)	N/A	N/A	ARDS
Jones, *et al.*, 1997[22]	[54] (17–90)	N/A	APACHE II: [15] (3–34)	General

TABLE 8 *Numbers of participants in studies using flow volume curves*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time [median] (range) (months)	Method of data presentation
Stauffer, *et al.*, 1981[44]	150	86 (57)	64	44 (69)	1, 3, 6	A
Lund, *et al.*, 1985[16]	41	24 (59)	17	17 (100)	[2.6] (0.7–4.9)	A
Elliott, *et al.*, 1988[18]	42	0 (0)	42	30 (71)	[48] (6–96)	N/A
Jones, *et al.*, 1997[22]	146	N/A	N/A	120 (N/A)	6	N/A

Lund and colleagues[16] presented their data as the numbers of individuals with pathological or with normal findings. No data were presented in the remaining papers.

Measurement properties in critical care

There was no evidence for the assessment of validity, reliability or responsiveness over time in any of the papers reporting on this measurement.

MMEF or MMIF rate

The maximal expiratory flow at a specified lung volume is the expiratory flow achieved at the designated lung volume during a forced expiratory manoeuvre starting from TLC. The volume used is usually 25%, 50% or 75% of FVC. The average expiratory flow rate between 25% and 75% of FVC is commonly used and is reported to be sensitive to small changes in airflow limitation. This is usually termed the MMEF. It is normally determined from a flow volume loop. A similar measurement can be made on inspiration (MMIF).

Measurement properties outside critical care

Although widely used, within-subject variability for MMEF and MMIF appears not to have been reported. In adult Caucasian populations MMEF is dependent on age, gender and height. The residual SDs for men and women are 1.04 and 0.88 litres per second, respectively. The MMEF is inversely correlated with the Sickness Impact Profile (SIP).[45]

Application in critical care

Three studies have reported using mid-expiratory flow rates: Halevy and colleagues[24] and Peters and colleagues[20] reported on MMEF; Law and colleagues[23] reported on MMEF and MMIF. The mean reported ages of patients ranged from 39.2 to 45.1 years (see *Table 1*). The percentages of men were similar for all three papers. None of the papers reported on acute severity scores.

The numbers of individuals eligible to participate ranged from 50 to 166 (*Table 2*). Mortality before follow-up ranged from 36% to 72%, with further attrition due to loss to follow-up, refusal and inability to return for clinical follow-up appointments. The proportions of those individuals followed-up from those available ranged from 29% to 95%. Follow-up times varied between the studies.

Halevy and colleagues[24] and Law and colleagues[23] presented data as the percentage of predicted MMEF for each participant. Peters and colleagues[20] presented the percentage of survivors with reduced MMEF relative to the reference values of Cherniack and Raber.[46]

Measurement properties in critical care

Validity In only one of the papers was the effect of age on outcome assessed;[20] it was reported as non-significant.

Reliability and responsiveness over time There was no evidence of any assessment of reliability or responsiveness.

PEF or PEFR

PEF or PEFR is the maximal flow during a forced expiratory VC manoeuvre starting from full inspiration. It can be measured using a spirometer, a pneumotachograph or a peak-flow meter. It reflects the calibre of central airways in healthy individuals and the peripheral airways in those with obstructive pulmonary disease. It is effort-dependent and the precise value depends on the exact definition of the timing of the measurement. However, because of the ease with which it can be made, this measurement is very widely used. It is used to assess both disease severity and response to treatment. It is gender, race, height and age dependent.

Measurement properties outside critical care

PEFRs show a diurnal variation in values, with a mean variation of 6.3%.[47] The within-subject

diurnal variation is considerably higher in patients with lung disease (17.2%). The accuracy of the measurement device should be within 3% of the reference value. PEFR measurements in Caucasian adults have a residual SD from the regression equation of 1.2 and 0.9 λ per second for men and women, respectively.

Iwasaki and colleagues[48] and van der Molen and colleagues[33] reported significant correlations between daily PEF measurements and symptom scores (r^2 between 0.26 and 0.45). The daily PEF variability was found to be useful in the treatment of asthmatic patients. Van der Molen and colleagues[33] also reported on the correlation between PEF and two asthma-specific quality-of-life measures, the Asthma Quality of Life Questionnaire and the Living with Asthma Quality of Life Questionnaire, as well as two generic questionnaires, the SF-36 and the PGWB index. The correlations between these quality-of-life measures and PEF were generally poor.

Application in critical care

Only two studies reported on PEF.[22,23] The median age of the patients was 51–54 years (see *Table 1*). Jones and colleagues[22] reported acute severity scores and in both studies general critical care patients had been recruited. The numbers eligible for participation were 109 and 146, and the numbers followed-up were 20 and 120 (see *Table 2*). Jones and colleagues[22] followed-up patients at 6 months after discharge and Law and colleagues[23] at more than 6 months, although no mean values or ranges are provided. Law and colleagues presented data for the percentage of predicted PEF for each patient and the ratio of PEF to peak inspiratory flow (PIF). No data were available from Jones and colleagues.[22]

Measurement properties in critical care

There was no evidence of the measurement properties of the measure in either paper.

Carbon monoxide diffusing capacity

The measurement of diffusing capacity quantifies the degree of diffusion limitation between the alveoli and the pulmonary capillaries. Diffusing capacity is normally measured using carbon monoxide (DLCO) but can also be determined using oxygen or nitric oxide. It is usually expressed as either:

- the transfer function (T_L), which is the rate of gas uptake for a given alveolar pulmonary capillary gas tension gradient; this is also referred to as the DLCO
- the transfer coefficient, which is the transfer function per unit alveolar volume.

The diffusing capacity is usually determined with a single timed breath-hold, measuring the initial and end-tidal concentrations of the indicator gas (carbon monoxide). DLCO is significantly reduced by thickening or oedema of alveolar walls and significantly but spuriously increased by alveolar haemorrhage.

Measurement properties outside critical care

The short-term variability in measurements is between 4.4 and 5.5% in healthy individuals. Between-subject variability in normal individuals is affected by smoking, inspired oxygen fraction, severe anaemia or polycythaemia, age, stature, body mass index, and race. The regression equations (for transfer function) have a residual SD of 1.4 and 1.2 $mmol.min^{-1}.kPa^{-1}$ for men and women, respectively. The DLCO correlates well with other indicators of reduced oxygen diffusing capacity such as peak oxygen uptake.[49] The measurement devices used should have an accuracy of ± 1% or ± 2% (carbon monoxide analysers).

Statistically significant correlations (r^2 = –0.24 to –0.36) were reported between a number of respiratory disease-specific questionnaires (St George's Respiratory Questionnaire, the Breathing Problems Questionnaire and the Chronic Respiratory Disease Questionnaire) and the transfer coefficient.[50]

Application in critical care

The use of DLCO was reported in seven papers (*Table 9*).[17–21,25,37] The mean reported ages of participants ranged from 25 years to 45.1 years and the percentages of men ranged from 31% to 74%. Acute severity scores were reported in only one study.[25] A majority of the papers related to patients with ARDS. The remainder reported on trauma patients who had been admitted to critical care.

The numbers of individuals eligible to participate ranged from 38 to 216 (*Table 10*). Mortality before follow-up ranged from 0% to 72%, further attrition being due to loss to follow-up, refusal to participate and inability to return for outpatient clinic appointments. The proportions of available individuals who were followed-up ranged from

TABLE 9 *Characteristics of populations in studies using DLCO*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD	Type of patient
Landercasper, *et al.*, 1984[37]	52 (7–87)	74	N/A	Trauma
Elliott, *et al.*, 1987[17]	25 (13–42)	31	N/A	ARDS
Elliott, *et al.*, 1988[18]	28 (13–62)	N/A	N/A	ARDS
Ghio, *et al.*, 1989[19]	28.5 ± 12.5 (7–61)	51	N/A	ARDS
Peters, *et al.*, 1989[20]	45.1 ± 15 (18–81)	67	N/A	ARDS
McHugh, *et al.*, 1994[21]	41 (19–73)	62	N/A	ARDS
Grotz, *et al.*, 1997[25]	33.6 ± 2.1	70	ISS 36.8 ± 1.6	MOF; trauma

TABLE 10 *Numbers of participants in studies using DLCO*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} (range) (months)	Method of data presentation
Landercasper, *et al.*, 1984[37]	62	13 (21)	49	20 (41)	{60} (6–144)	A
Elliott, *et al.*, 1987[17]	38	0 (0)	38	16 (42)	48	C
Elliott, *et al.*, 1988[18]	42	0 (0)	42	30 (71)	6	A
Peters, *et al.*, 1989[20]	166	119 (72)	47	39 (83)	(1–36)	A
Ghio, *et al.*, 1989[19]	41	N/A	N/A	13 (N/A) 27 (N/A)	1 > 12	A
McHugh, *et al.*, 1994[21]	216	134 (62)	82	20 (24)	Extubation 3, 6, 12	A
Grotz, *et al.*, 1997[25]	173	104 (60)	69	50 (72)	{58.8}	C

24% to 83%. In only one study had baseline data at endotracheal extubation been obtained and there was a wide range of follow-up times, from 1 month to 144 months. Serial follow-up was reported in only one paper.[21]

Mean and SEM values were reported in two studies.[17,25] In the remaining five papers the percentages of predicted values were presented. Reference values were provided by Grotz and colleagues,[25] who used the methods described by Quanjer and colleagues,[14] McHugh and colleagues[21] and Peters and colleagues,[20] who reported using the methods of Gaensler and Wright,[51] and Elliott and colleagues,[17,18] who reported using the methods of Crapo and colleagues.[27]

Measurement properties in critical care

Validity Construct validity was assessed in two papers,[17,20] both of which reported no significant association with age. Criterion validity, using the SIP, was reported by McHugh and colleagues;[21] no significant correlations were found at 3, 6, and 12 months follow-up using the Spearman rank order correlation test.

Reliability Peters and colleagues[20] and Elliott and colleagues[17,18] reported using the same equipment and protocol but did not give details of the reliability of the instruments.

Responsiveness McHugh and colleagues' paper[21] was the only one in which statistical differences in their serial pulmonary function tests using repeated measures analysis of variance (ANOVA) were reported.

Visualisation of the upper airway

Two techniques are used to visualise the upper airway: X-rays and endoscopy (bronchoscopy or laryngoscopy). X-rays are generally used to investigate the trachea, endoscopy can be used to examine the entire airway from mouth to segmental bronchi. Bronchoscopy can be performed using either a rigid bronchoscope or a flexible fibre-optic device to enable visualisation of the airway. Rigid bronchoscopy normally requires general anaesthesia, whereas flexible bronchoscopy can be performed under local anaesthesia.[52]

Laryngoscopy and laryngotracheoscopy are similar forms of endoscopy to visualise specific parts of the patient's upper airway. Bronchoscopy and laryngoscopy cannot be used to quantify stenoses; X-rays can be used to make qualitative or quantitative measurements of stenosis. The main use of upper airway visualisation following critical care is to determine the site and severity of tracheal (subglottic) stenosis following prolonged endotracheal intubation or tracheostomy.

Measurement properties outside critical care

There are no available data on measurement properties outside critical care for assessing tracheal stenosis. For post-transplant bronchial stenoses, axial computed tomography (CT) or 3-dimensional reconstructions are equally effective at assessing the diameter and length of stenosis.[53]

Application in critical care

Six papers were identified which reported using X-rays and eight which reported using endoscopic techniques to assess stenoses and upper airway obstruction (*Table 11*).[15,16,23,24,37,42,44,54–58] Friman and colleagues[15] used chest X-rays taken in the supine position using a radiopaque indicator to highlight the trachea. Dane and King[54] also took X-rays to highlight the trachea and, in addition, took X-rays during the bronchoscopic procedure. Landercasper and colleagues[37] used X-rays to identify evidence of pleural thickening, rib callus and localised pulmonary fibrosis. Walz and colleagues[58] obtained X-rays normal to the frontal and sagittal planes to identify stenosis.

Mean ages of participants ranged from 39 years to 58 years, men made up about 60–70% of patients, and most studies included general critical care patients. Acute severity scores were not reported in any of the papers; two studies predated the development of acute severity scores. The follow-up proportions ranged from 7% to 100% (*Table 12*). Most studies followed-up patients several months after admission to critical care, although some reported outcomes several years later. Most studies reported the proportion of survivors with normal findings; no data were presented in two papers.

Measurement properties in critical care

Validity There was no evidence of an attempt to assess the validity of the measure.

Reliability Halevy and colleagues[24] used two radiologists to assess chest X-rays independently, although no details of their level of agreement are reported.

Responsiveness X-rays were taken at multiple time points although these were not specified.[58] Although serial bronchoscopies were performed in two studies, there were no evaluations of changes over time.

TABLE 11 *Characteristics of populations in studies using visualisation*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Type of patient
Aass, 1975[42] ab	(1 day–89)	59	General
Dane & King, 1975[54] cb	43.5 (4–74)	68	General
Friman, *et al.*, 1976[15] b	N/A	60	General
Stauffer, *et al.*, 1981[44] c	58 (17–88)	73	General
Sellery, *et al.*, 1978[55] a	N/A	N/A	General
Halevy, *et al.*, 1984[24] b	39.2 (18–68)	67	ARDS
Landercasper, *et al.*, 1984[37] b	52 (7–87)	74	Trauma
Lund, *et al.*, 1985[16] c	44 (18–87)	76	Burns
Winkler, *et al.*, 1994[56] d	[62] (30–88)	60.5	General
Lengas, *et al.*, 1996[57] c	42 (9–83)	74	General
Law, *et al.*, 1997[23] d	[51] (18–86)	68	General
Walz, *et al.*, 1998[58] b	45.1 ± 17.9 (11–77)	68	General

[a] *Laryngoscopy*
[b] *X-ray*
[c] *Bronchoscopy*
[d] *Laryngotracheoscopy*

TABLE 12 *Numbers of participants in studies using visualisation*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean ± SD} [median] (range) (months)	Method of data presentation
Aass, 1975[42] [bc]	79	36 (46)	43	35 (81)[b]/ 22 (51)[c]	(9–48)	A
Dane & King, 1975[54] [bc]	40	15 (38)	25	25 (25)	1, 3	A
Friman, *et al.*, 1976[15] [b]	320	183 (57)	137	10 (7)	N/A	N/A
Stauffer, *et al.*, 1981[44] [c]	150	86 (57)	64	8 (13)	1, 3, 6	N/A
Sellery, *et al.*, 1978[55] [a]	N/A	N/A	N/A	21 (N/A)	Up to 36	A
Halevy, *et al.*, 1984[24] [b]	50	31 (62)	19	18 (95)	(3–30)	A
Landercasper, *et al.*, 1984[37] [b]	62	13 (21)	49	21 (43)	[60] (6–144)	A
Lund, *et al.*, 1985[16] [c]	41	24 (59)	17	17 (100)	[2.6] (0.7–4.9)	A
Winkler, *et al.*, 1994[56] [d]	71	45 (63)	26	14 (54)	6	A
Lengas, *et al.*, 1996[57] [c]	38	0 (0)	38	38 (100)	8	A
Law, *et al.*, 1997[23] [d]	109	39 (36)	70	20 (29)	> 6	A
Walz, *et al.*, 1998[58] [b]	326	209 (64)	117	106 (91)	(6–33) {9.9 ± 5.6}	A

[a] *Laryngoscopy*
[b] *X-ray*
[c] *Bronchoscopy*
[d] *Laryngotracheoscopy*

Hepatic, renal and haematological measures

Several measures of impairment of liver, kidney and other organ functions were reported (*Table 13*).[25,38,59–61] Such measures were generally only reported in one study and in none of the studies were the measurement properties in critical care investigated.

Summary

- The use of measures of impairment to assess the outcome of critical care has largely been confined to the respiratory system. Reporting of impairment of other systems (hepatic, renal) has been restricted to single studies in which specific measures were used and from which, therefore, it is difficult to draw conclusions.
- Few attempts have been made to investigate the measurement properties of the respiratory instruments that have been used. Evidence of construct validity was absent for measures of respiratory volume (VC, FVC, TLC), in that they were not associated with age (in four out of five studies), although this may be due either to the small sample sizes or to the somewhat restricted age distributions of the samples. The same is true of respiratory flow measures (FEV_1, FVC, MMEF), as demonstrated in three studies, and in two studies for diffusing capacity (D_LCO). The only other evidence of measurement properties was the absence of criterion validity of D_LCO (no association with the SIP).
- Each measure of impairment was organ-specific and was therefore not deemed appropriate for the follow-up of general critical care patients. They may be appropriate for the follow-up of specific subgroups such as those with ARDS. When these measures are used, supportive evidence for validity and reliability of equipment and methods should be provided. The European Respiratory Society publishes clear guidelines on the performance of pulmonary function tests.
- The measurement properties of pulmonary function tests (excluding techniques to visualise the upper airway) outside critical care are very well established and the sources of variability not related to disease severity have been established and quantified. In addition, for two disease groups (chronic obstructive pulmonary disease and asthma) there is a body of evidence suggesting that a physiological measure of disease severity (FEV_1) is correlated with two health-related quality-of-life measures, NHP and SF-36.

TABLE 13 *Measures of hepatic, renal and other metabolic functions*

Function	Study
Hepatic function	
general	Martinelli, *et al.*, 1995[59]
bilirubin	Grotz, *et al.*, 1997[25]
aspartate aminotransferase	Nordback & Auvinen, 1985[60]
serum glutamyltransferase	Doepal, *et al.*, 1993[61]
alkaline phosphatase	Nordback & Auvinen, 1985[60]
C-peptide	Nordback & Auvinen, 1985[60]
Renal function	
general	Martinelli, *et al.*, 1995[59]
creatinine	Grotz, *et al.*, 1997;[25] Gammie, *et al.,* 1998[38]
urea	Grotz, *et al.*, 1997[25]
Glucose metabolism	
blood glucose	Doepel, *et al.*, 1993[61]
glucose tolerance	Nordback & Auvinen, 1985[60]
Haematological	
haemoglobin	Doepal, *et al.*, 1993;[61] Nordback & Auvinen, 1985[60]
erythrocyte sedimentation rate	Nordback & Auvinen, 1985[60]
full blood count	Grotz, *et al.*, 1997[25]
glycohaemoglobin A1 and A1C	Doepal, *et al.*, 1993[61]
reticulocyte count	Grotz, *et al.*, 1997[25]
leucocyte count	Doepal, *et al.*, 1993;[61] Nordback & Auvinen, 1985[60]

- It is difficult to interpret the results on FEV_1 because the pre-morbid pulmonary function of ICU survivors is unknown and a significant proportion of these patients might be expected to have some chronic respiratory history. When both the FEV_1/FVC ratio and FVC were reported, abnormalities of FVC were more common, implying restrictive rather than obstructive pulmonary disease.

Chapter 3

Measures of physical functional status

Outcome measures assessing functional status were identified whose use had been reported on more than one occasion. They were grouped into four categories:

(i) physical functional status
(ii) mental functional status
(iii) neuropsychological functional status
(iv) extent of recovery.

In this chapter, physical functional status is considered.

Eight measures have been used to assess the physical functional status of critical care survivors. Five are generic measures (Katz's Activities of Daily Living (ADL) index, Karnofsky Index, Barthel Index, activity level, functional state) and three are disease-specific measures (New York Heart Association (NYHA) questionnaire, ATS respiratory questionnaire, walk test). Of the five generic measures, three are multi-item scales and two are based on a single global item or question. The generic measures are discussed first.

Katz's ADL index

The ADL index was developed by Katz and colleagues[62] in 1963 to describe the functional status of elderly patients for clinical purposes. Based on the observation of a large number of patients with fractured hips, it has subsequently been used to assess outpatient treatment for patients with rheumatoid arthritis and stroke.[63,64] The index ranks individuals according to their performance of six functions: bathing, dressing, toileting, transferring, continence and feeding, which is then expressed as a grade (from A (independent) to G (dependent)) in each of the six functions. The index was developed for completion by an observer. The developers claim that the index is a useful tool in the study of prognosis and the effectiveness of treatment, a survey instrument and a means of acquiring additional knowledge about the ageing process.

Measurement properties outside critical care

There is little evidence of the measurement properties of this index. Inter-observer reliability was established by the developers of the measure[62] when they reported discordance between observers in only 5% of cases. There is little available evidence reporting the validity of the index. The ADL index showed weak to moderate correlations with other scales examining mobility ($r^2 = 0.25$) and house confinement ($r^2 = 0.14$).[65] The fact that it produces a style index means that it is of limited value in the information it provides.

Application in critical care

Eleven papers were identified which had utilised Katz's ADL index (*Table 14*).[66–76] Four studies used a modified version of the index. The mean reported ages of the participants ranged from 30 years to 89 years, and the percentages of male participants ranged from 41% to 80%. Acute severity scores were reported in seven studies. The majority of studies reported on general patients but three dealt specifically with neurological and trauma patients.

The numbers of individuals eligible for participation in the studies ranged from 63 to 3619 (*Table 15*). The proportions of available individuals who were followed-up ranged from 21% to 100%, with complete follow-up achieved in only one study. Follow-up times ranged from 1 month to a mean of 64 months. Two studies had more than one follow-up period. Three studies acquired baseline data on admission for pre-admission status. Despite Katz's ADL index being designed to be observer-completed, in seven studies methods of administration other than those suggested by the developers of the index were selected: five used telephone interviews,[69,71–74] two used mailed questionnaires,[67,76] and only one used face-to-face interviews.[68] In addition to face-to-face interviews, Kass and colleagues[68] also reviewed patients' medical records and interviewed staff at nursing homes where the patients resided.

A variety of methods were used to present the data from these studies (*Table 15*). Mean values with SDs were presented in four studies and, in five, the numbers and percentages of patients who had gained functional independence were presented.

TABLE 14 *Characteristics of populations studied using Katz's ADL index*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Levy, *et al*., 1985[66] (modified)	[61]	61	N/A	Neurological
Ridley & Wallace, 1990[65]	N/A	N/A	N/A	General
Kass, *et al*., 1992[68] (modified)	89.4 ± 3.5 (85–102)	41	N/A	General
Rockwood, *et al*., 1993[69] (modified)	N/A	63	< 65 years: APS, 15 ± 9 APACHE II, 17 ± 9 > 65 years: APS, 15 ± 8 APACHE II, 21 ± 9	General
Chelluri, *et al*., 1993[70]	65–74: 69 ± 0.3 SEM 75+: 81 ± 0.5 SEM	44 (65–74 years) 52 (75+ years)	65–74 years: APACHE II, 18 ± 0.9 SEM APS, 13 ± 0.9 SEM TISS, 28 ± 1.8 SEM 75+ years: APACHE II, 20 ± 0.8 SEM APS, 14 ± 0.8 SEM TISS, 32 ± 1.8 SEM	General
Holbrook, *et al*., 1994[71*]	30 ± 13.1 (18–69)	74	ISS: 15 ± 10.1 (5–43)	Trauma
Broslawski, *et al*., 1995[72]	77.3 ± 6.9 (65–92)	50	APACHE II: 16.3 ± 6.8	General
Wu, *et al*., 1995[73]	62.1	55	N/A	General
Dardaine, *et al*., 1995[74]	78 ± 0.7	59	SAPS: 15 ± 0.6	General
Cho & Wang, 1997[75]	60.5 (14–87)	80	APACHE II: 13.5 ± 5.6 (2–29) APACHE III: 42.6 ± 18 (3–115) GCS: 4.9 ± 2.2 (3–19)	Neurological
Battistella, *et al*., 1998[76] (modified)	85 ± 3.9 (77–99)	N/A	ISS: 9.4 ± 7.7	Trauma

**Holbrook and colleagues[71] reported the use of the functional disability score*

TABLE 15 *Numbers of participants in studies using Katz's ADL index*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} (range) (months)	Method of data presentation
Levy, *et al.,* 1985[66]	210	121 (58)	89	19 (21)	12	A
Ridley & Wallace, 1990[67]	385	129 (34)	256	156 (61)	(12–36)	N/A
Kass, *et al*., 1992[68]	105	67 (64)	38	36 (95)	12	C
Rockwood, *et al*., 1993[69]	884	305 (35)	579	430 (74)	12	A
Chelluri, *et al*., 1993[70]	97	59 (61)	38	32 (84)	1, 6, 12	A
Holbrook, *et al*., 1994[71]	63	N/A	N/A	42 (N/A)	3	N/A
Broslawski, *et al.,* 1995[72]	68	18 (40)	50	27 (54)	6	A
Wu, *et al*., 1995[73]	3619	1306 (36)	2313	1746 (75)	2	C
Dardaine, *et al*., 1995[74]	110	74 (67)	36	27 (75)	18	C
Cho & Wang, 1997[75]	200	55 (28)	145	145 (100)	{26.4} (12–36)	C
Battistella, *et al*., 1998[76]	279	132 (47)	147	93 (63)	{64.8}	A

Measurement properties in critical care

Validity In only one paper, by Wu and colleagues,[73] had criterion validity been assessed by using the Duke Activity status ($r^2 = 0.24$; $p < 0.0001$) and the SIP ($r^2 = 0.32$; $p < 0.0001$).

Broslawski and colleagues[72] assessed construct validity using age, length of ICU stay and Acute Physiology and Chronic Health Evaluation (APACHE) II score on admission, none of which were reported as being significantly associated with ADL index score. Wu and colleagues[73] also assessed construct validity using the variables of age ($p < 0.01$), gender ($p < 0.001$), and the Glasgow Coma Scale (GCS) ($p < 0.01$) using the chi-squared test in a regression. Rockwood and colleagues[69] reported an association with age using regression, although the variable accounted for less than 2% of the variance.

Reliability There was no evidence of reliability assessment in any of the eleven papers.

Responsiveness A significant change between pre- and post-critical care ADL scores ($p < 0.01$) was reported by Wu and colleagues,[73] who also examined the predicted versus actual ADL scores. In their analysis, the area under the Receiver Operator Characteristics curve was 0.70, providing some support for discriminatory characteristics of the measure.

Outcome of critical care survivors

Studies employing the ADL have been undertaken for elderly survivors,[68–70,74,76] for all survivors from general ICUs,[72,73] for survivors from a neurosurgical ICU,[75] for survivors from a trauma centre,[71] and for survivors with a specific diagnosis of cerebral hypoxia-ischaemia.[66]

For elderly ICU survivors, Chelluri and colleagues[70] reported that, for those aged 65–74 years ($n = 43$), functional ability was significantly decreased at 1 month compared with pre-admission but returned to pre-admission status by 6 months. For survivors aged 75+ years ($n = 54$), functional ability was unchanged at 1 year follow-up compared with pre-admission status. Rockwood and colleagues,[69] who compared ICU survivors aged less than 65 years ($n = 478$) with those over 65 years ($n = 406$), reported no evidence of a higher level of functional impairment in the older group at 1 year, except for more assistance with bathing. In multiple regression analysis, age – although strongly associated with ADL – accounted for less than 2% of the variation. Dardaine and colleagues,[74] in a population of ICU survivors aged 70+ years and ventilated for at least the first 24 hours in ICU ($n = 36$), reported dependency for ADL of 22% at 18 months compared with 13% pre-admission. Battistella and colleagues[76] reported long-term follow-up (mean 5.4 years) for ICU survivors aged 70+ years ($n = 93$) as 35% reporting complete independence and 57% reporting complete or only moderate dependence (no difficulties in performing 12 out of 14 ADLs).

Turning to all ICU survivors, Broslawski and colleagues[72] reported, for 27 ICU survivors at 6 months, decreased function in six survivors, with 21 showing similar or improved function. The strongest predictors of decreased function were ICU and total length of stay in hospital. Neither age nor severity of illness correlated with decreased functional ability. In the largest study ($n = 1746$), Wu and colleagues[73] reported that the proportion of survivors with severe functional limitations had, at 2 months, nearly tripled to 34%. The proportion with severe functional limitations at 2 months increased directly with the admission's pre-admission dependency. Eight variables were independent predictors of severe functional limitation at 2 months: ADLs pre-admission; reported quality of life pre-admission; Duke Activity status pre-admission; disease group; physiology score; GCS score; age; and days in hospital before recruitment to the study.

Cho and Wang[75] reported long-term follow-up (mean 2.2 years) for 122 neurosurgical ICU survivors: 48% were independent; 8% were dependent in one activity; 10% in two; 7% in three; and 27% in four or more. APACHE III was a better predictor of functional outcome than APACHE II and the GCS.

Holbrook and colleagues[71] reported 3-month follow-up for 42 trauma patients (aged 18 years or more, GCS score 12 or more on admission, length of stay 24 hours or more) using a modification of the ADL index. The mean pre-admission score, 29, was higher than the mean score at 3 months, 17. Most patients reported improved function at follow-up. Finally, Levy and colleagues[66] reported 12-month follow-up for admissions to ICU with cerebral hypoxia-ischaemia ($n = 210$). At 12 months, 23% had severe disability (dependency in all ADLs) and 78% were worse with no recovery (continued coma until death) or persistent vegetative state.

Karnofsky Index

The Karnofsky Index was originally developed as a measure of overall health status in lung cancer patients.[78] Scores range from 0 (dead) to 100 (normal). Although it was not designed to assess quality of life, it is frequently misused for this purpose.[79] The Karnofsky Index emphasises physical performance and dependency. The index is therefore weighted heavily towards the physical rather than psychological dimensions, with scores being assigned by a clinician rather than the patient.

Measurement properties outside critical care

Criterion validity was established by Mor and colleagues,[80] using Katz's ADL index and the Spitzer Quality of Life Scale. They also reported that construct validity was adequate. Some studies, such as the US National Heart Transplantation Study,[81] have demonstrated changes in scores before and after organ transplantation, whereas other studies have reported inconsistent results in the correlation between the Karnofsky Index and treatment response.[82,83] Wide discrepancies have been reported between physicians' ratings using the Karnofsky Index and patients' ratings using the SIP.[81] Yates and colleagues[84] claimed that the index was not appropriately scaled, a view supported by Schipper and colleagues.[85] Poor inter-rater agreement has been reported by Hutchinson and colleagues[86] and Mercier and colleagues.[87]

Application in critical care

Five studies have used this index.[88–92] The mean reported ages of patients ranged from 40 years to 83 years (*Table 16*). The proportions of men reported in the three papers detailing this variable ranged from 42% to 67%. Four of the five papers reported acute severity scores. The types of patient varied. The numbers of individuals eligible for participation in these studies ranged from 15 to 292 (*Table 17*). Mortality before follow-up ranged from 28% to 70%; further attrition was due to loss to follow-up and refusal to participate. The proportions of available patients who were followed-up ranged from

TABLE 16 *Characteristics of populations studies using the Karnofsky Index*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD [median] (range)	Type of patient
Yinnon, *et al.*, 1989[88]	54 (15–102)	62	16 ± 8	General
Kumar, *et al.*, 1995[89]	Group 1: 83.2 ± 2.2 (80–87) Group 2: 83 ± 2.0 (80–89)	Group 1: 60 Group 2: 42	N/A	Cardiac surgery
Singh, *et al.*, 1997[90]	48 (28–68)	N/A	APACHE II: 14.5 APS: 7.9	Liver transplant
Weinert, *et al.*, 1997[91]	40 ± 12	67	Lung injury score: 2.4 ± 0.54 (1.25–3.25)	Acute lung injury
Kocher & de Torrenté, 1998[92]	67	56	APACHE II: [9.0] (0–52)	General

TABLE 17 *Numbers of participants in studies using the Karnofsky Index*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean ± SD} [median] (range) (months)	Method of data presentation
Yinnon, *et al.*, 1989[86]	126	71 (56)	55	52 (95)	6	C
Kumar, *et al.*, 1995[87]	Group 1: 15 Group 2: 53	Group 1: 7 (47) Group 2: 15 (28)	Group 1: 8 Group 2: 38	Group 1: 8 (100) Group 2: 38 (100)	Group 1: {76.4 ± 3.6} Group 2: {18.6 ± 3.0}	C
Singh, *et al.*, 1997[88]	50	9 (30)	41	41 (100)	6	B
Weinert, *et al.*, 1997[89]	69	35 (51)	34	24 (71)	{19} [15] (6–41)	A
Kocher & de Torrenté, 1998[90]	292	203 (70)	89	45 (51)	(12–74)	A

51% to 100%. Two studies achieved 100% follow-up and one achieved 95%. In two papers,[88,90] baseline data were reported that had been obtained on admission to critical care.

Singh and colleagues[90] provided no details of how they administered the Karnofsky Index. Despite the index being designed for completion by a clinician, data were collected in this way in only one study.[88] Kumar and colleagues[89] administered the index by telephone interview, and Weinert and colleagues[89] and Kocher and de Torrenté[92] used the postal method.

A variety of methods were used to present the data in these studies (*Table 17*). In two papers[88,89] the data was presented as the mean and SD, and in one paper[90] only the mean index score was reported.

Measurement properties in critical care

Validity Only one paper addressed the issue of construct validity. Kocher and de Torrenté[92] presented positive correlations with age and APACHE II scores. Weinert and colleagues[91] reported high correlations for criterion validity with the physical component of the SF-36 ($r^2 = 0.56$) and a health satisfaction scale ($r^2 = 0.62$), which ranged from worst imaginable to best imaginable health.

Reliability Inter-rater reliability was assessed in one study by comparing patient responses with those from relatives or clinicians, but no analyses were reported.[89]

Responsiveness Sensitivity over time was reported by Yinnon and colleagues,[88] who found no significant differences between admission and 6-month follow-up data. Kumar and colleagues[89] detailed significant changes between the pre-surgical and the follow-up Karnofsky Index category. Critical care survivors had a significantly ($p < 0.01$) lower Karnofsky score (worse health) at 6-month follow-up compared with patients who had not needed admission to critical care.

Barthel Index

The Barthel Index[93] is based on observed functions and was developed to compare physical functional status before and after an intervention, and to indicate potential nursing requirements. The developers designed the index for use with long-term hospitalised patients, especially those with musculoskeletal or neuromuscular disorders.[94] The index is completed by a therapist or other observer and is a rating scale that takes approximately 30 seconds to complete. It comprises nine dimensions: feeding, mobility from bed to chair, personal toilet, getting on/off the toilet, bathing, walking on level surface, going up/down stairs, dressing and continence. If the patient does not meet the criterion, they are given a score of zero. The index is suitable only for use with institutionalised patients, for whom it was designed. The scoring system elicits a score of zero if the patient is totally dependent on all activities and a score of 100 if they are fully independent.

Measurement properties outside critical care

Mattison and colleagues[95] reported correlations of –0.69 and 0.65 between the Barthel Index and the PULSES* scale and the Edinburgh Rehabilitation Status Scale,[96] respectively. Other studies in which the index was compared with the PULSES scale reported correlations ranging from –0.74 to –0.90.[97] Factor analysis has confirmed that the index reflects a single domain, and Wade and Langton-Hewer[98] reported that it is sensitive to detecting a patient's recovery. Cronbach's α coefficients of 0.95 were reported by Sherwood and colleagues,[99] supporting its internal consistency. Test–retest reliability was found to be high in a study of severely disabled patients.[97]

Application in critical care

Three papers were identified that reported using the Barthel Index.[100–102] Two papers reported the mean ages of the participants and the percentages of men participating but only one reported an acute severity score (*Table 18*). Two studies reported on the outcomes of trauma patients who had been in critical care and one reported on general critical care patients.

The numbers of patients eligible for participation ranged from 118 to 558 (*Table 19*). Mortality before follow-up ranged from 0% to 52%. Further attrition was due to loss to follow-up and refusal to participate. The proportions of available individuals who were followed-up ranged from 68% to 100%, two studies achieving the latter. In one study,[102] the index was administered at two time points, one of which was a baseline measurement obtained at discharge from critical care. There was no consistency between studies in the times of follow-up.

* PULSES, Physical condition, Upper limb function, Lower limb function, Sensory component, Excretory function, mental and Status [profile].

TABLE 18 *Characteristics of populations studied using the Barthel Index*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Mean severity score	Type of patient
Day, *et al.*, 1994[100]	N/A	N/A	ISS: 25	Trauma
Gobiet, 1995[101]	23 (6–80)	66	N/A	Trauma; neurological
Neundörfer, *et al.*, 1996[102]	56.7 ± 18.8	56	N/A	General

TABLE 19 *Numbers of participants in studies using the Barthel Index*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (range) (months)
Day, *et al.*, 1994[100]	118	36 (31)	82	56 (68)	24–51
Gobiet, 1995[101]	558	0 (0)	558	558 (100)	36
Neundörfer, *et al.*, 1996[102]	422	219 (52)	203	203 (100)	(18–30)

Gobiet[101] did not specify how the Barthel Index was administered. Day and colleagues[100] and Neundörfer and colleagues[102] administered the index by post. As the index was designed for administration by an observer or therapist, the mode selected in these studies may be inappropriate.

There was considerable variation in data presentation. Day and colleagues[100] presented categorical data as the number of subjects scoring < 60, 61–69 and 70–100. Gobiet[101] presented data as the numbers and percentages of individuals with Barthel Index scores of < 30, < 35, < 70, < 80, and 80 and above. Scores in each category were detailed by Neundörfer and colleagues.[102] None of the papers presented mean values or SDs.

Measurement properties in critical care

Validity Only one paper[102] provided any information on construct validity. The authors showed that once other factors such as pre-existing diseases and complications were taken into account, the Barthel Index score was not age-related.

Reliability and responsiveness None of the papers made any reference to the assessment of reliability or responsiveness over time in critical care survivors.

Activity levels

This global measure comprises five levels: normal activity; limited activity; still ill; in hospital; and dead. The questionnaire was first reported, in this form, by Madsen and Eriksen.[103]

Measurement properties outside critical care

Activity levels have been used in several studies but there is no evidence relating to reliability and validity outside critical care.

Application in critical care

Four studies have used this measure,[104–107] of which one did not provide information on the characteristics of the patients (*Table 20*). The remaining three papers all derive from the same study. The percentages of men in this sample ranged from 50% to 69%. Acute severity scores were reported in one paper only.[104] Study populations were either from general critical care or were a specific group of patients (alcohol-dependent).

The numbers of patients eligible for participation ranged from 26 to 1308 (*Table 21*). Mortality before follow-up was 43% to 50%, further attrition being primarily due to loss to follow-up and refusal to participate. Complete follow-up of available individuals was reported in three papers. Schuster[107] reported on long-term follow-up (over 12 months), while in the other papers outcomes were reported at 3, 6, 9 and 12 months.

Schuster[107] did not provide any information as to how the measure was administered. The other papers reported that the questionnaire was administered by post. For all papers, data were presented as the numbers of patients or the numbers and percentages of patients in each category of the measure.

TABLE 20 *Characteristics of populations studied using activity levels*

Study	Mean age ± SD [median] (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Jensen, *et al.*, 1988[104]	56.2	69	APS: 19.1 (4–36) TISS: 32 (9–52)	Alcohol-dependent
Dragsted & Qvist, 1989[105]	[60]	51	N/A	General
Dragsted, 1991[106]	Men: [61] Women: [58]	50	N/A	General
Schuster, 1991[107]	N/A	N/A	N/A	General

TABLE 21 *Numbers of participants in studies using activity levels*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (range) (months)
Dragsted & Qvist, 1989[105]	1308	558 (43)	750	750 (100)	3, 6, 9, 12
Jensen, *et al.*, 1988[104]	26	13 (50)	13	13 (100)	3, 6, 9, 12
Dragsted, 1991[106]	1308	558 (43)	750	750 (100)	3, 6, 9, 12
Schuster, 1991[107]	1308	N/A	N/A	N/A	(12–60)

TABLE 22 *Characteristics of populations studies using functional states*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score mean ± SD (range)	Type of patient
Cullen, *et al.*, 1976[108]	59	65	TISS: 43 ± 1.0	General
Cullen, *et al.*, 1984[109]	62	56	N/A	General
Slayter, *et al.*, 1986[110]	43.7 (0.25–82)	58	TISS: 88 ± 129.9	General
Nolla-Salas, *et al.*, 1993[111]	76 ± 3.7 (70–85)	N/A	APS: 11.2 ± 5.1 (1–23)	General

Measurement properties in critical care

Validity Both Dragsted and Qvist[105] and Dragsted[106] reported a correlation between age and activity levels. In one paper[105] this was reported as significant ($p < 0.001$).

Reliability and responsiveness The reliability of this outcome measure was not assessed in any paper. Although several time points were used in three of the papers, the sensitivity of the measure over time was not reported.

Functional state measures

This is a global measure, in which patients are assigned to one of four broadly defined health states: freely ambulatory; limited activities; bedridden – self-care; bedridden – no self-care.

Measurement properties outside critical care

This measure was devised for use in ICU patients, so no information is available on its properties in other settings.

Application in critical care

The mean reported ages of the patients in these studies ranged from 44 years to 76 years and the percentages of men recruited ranged from 56% to 65% (*Table 22*).[108–111] Acute severity was reported in three studies. The numbers of eligible subjects ranged from 97 to 231 (*Table 23*). Mortality before follow-up ranged from 11% to 73%, further attrition being due to loss to follow-up or refusal to participate. A high percentage (90–100%) of patients were alive at the time of follow-up and were assessed. There were two studies[108,109] in which the measure was administered on more than one occasion. The range of follow-up times was 1–12 months.

TABLE 23 *Numbers of participants in studies using functional states*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (months)
Cullen, *et al.*, 1976[108]	231	169 (73)	62	62 (100)	1, 3, 6, 12
Cullen, *et al.*, 1984[109]	206	138 (69)	68	61 (90)	1, 6, 12
Slayter, *et al.*, 1986[110]	100	11 (11)	89	89 (100)	1
Nolla-Sallas, *et al.*, 1993[111]	97	51 (53)	46	44 (96)	6

Cullen and colleagues[108] invited patients to outpatient appointments for assessment. If they could not attend they were contacted by mail or telephone. Slayter and colleagues[110] did not specify how they administered the outcome measure. In three studies,[108,109,111] the outcome measure was administered by telephone interview. All four papers presented data as the numbers and percentages of patients in each category of the measure.

Measurement properties in critical care

Validity and reliability None of the papers presented any evidence of validity or reliability with regards to this outcome measure.

Responsiveness Although Cullen and colleagues[108,109] examined outcome at various time points, they did not assess the responsiveness of the measure.

NYHA functional class

The functional class grades of the Criteria Committee of the NYHA[112] are a functional and therapeutic classification for the prescription of physical activity for patients with cardiac disease. It is a disease-specific assessment that is frequently used in clinical trials of patients with congestive cardiac disease. The classifications, which are usually made by clinicians,[113,114] range from I to IV, as follows:

- class I – no limitation on activities, no symptoms and no difficulties with normal physical activity
- class II – slight limitation on physical activities, where physical activity may result in fatigue, breathlessness or anginal pain
- class III – marked limitations on physical activity, less than ordinary physical activity results in fatigue, breathlessness and/or anginal pain
- class IV – any physical activity results in discomfort and the symptoms of cardiac insufficiency may be present even at rest.

Measurement properties outside critical care

The NYHA classification has been found to correlate poorly with a measure of impairment (exercise testing), which suggests poor discriminative ability. Agreement with a measure of impairment (treadmill exercise test) was also poor (51%). In addition, high inter-observer variability was reported by Goldman and colleagues[113] and Wiklund and colleagues.[115,116] Two physicians agreed in only 56% of cases.

Application in critical care

This measure was used in five studies.[38,89,117–119] The mean reported ages of patients ranged from 36 years to 83 years (*Table 24*). The percentages of men participating ranged from 42% to 75%. Data for acute severity scores were presented in only one paper.[119] Four papers reported on follow-up of patients who had undergone cardiac surgery; the remaining paper reported on follow-up of patients who had undergone lung transplantation.

The numbers of patients eligible for participation ranged from 15 to 116 (*Table 25*). Mortality before follow-up ranged from 23% to 50%. The proportions of available patients followed-up ranged from 64% to 100%, the latter achieved in one study only. Follow-up periods varied from study to study, with a mean ranging from 27 months to 76 months.

Although the scale was designed for completion by a doctor, two studies used telephone interviews[89,119] and another[118] used postal questionnaires. Lee and colleagues[117] and Gammie and colleagues[38] did not specify how they administered the measure.

The authors of two papers[89,118] presented mean values and Kumar and colleagues[89] also presented SDs. Data were generally presented as the numbers or percentages of patients in each NYHA functional class.

TABLE 24 *Characteristics of populations studied using NYHA scale*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD	Type of patient
Lee, *et al.*, 1993[117]	50.8 ± 12.9 (22–72)	71	N/A	Cardiac surgery
Kumar, *et al.*, 1995[89]	Group 1: 83.2 ± 2.2 (80–87) Group 2: 83.0 ± 2.0 (80–89)	Group 1: 60 Group 2: 42	N/A	Cardiac surgery
McHugh, *et al.*, 1997[118]	77.4	47	N/A	Cardiac surgery
Trouillet, *et al.*, 1996[119]	60.5 ± 12.2	62	SAPS*: 9.7 ± 4 GCS: 13.1 ± 3	Cardiac surgery
Gammie, *et al.*, 1998[38]	(14–61) Single lung: 36 Double lung: 38	 Single lung: 67 Double lung: 75	N/A	Lung transplant

**SAPS, simplified acute physiology score*

TABLE 25 *Numbers of participants in studies using NYHA scale*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean ± SD} [median] (range) (months)	Method of data presentation
Lee, *et al.*, 1993[117]	28	14 (50)	14	9 (64)	{27 ± 20} (1–53)	A
Kumar, *et al.*, 1995[89]	Group 1: 15 Group 2: 53	Group 1: 7 (47) Group 2: 15 (28)	Group 1: 8 Group 2: 38	Group 1: 8 (100) Group 2: 38 (100)	Group 1: {76.4 ± 3.6} Group 2: {18.6 ± 3.0}	C
McHugh, *et al.*, 1997[118]	97	17 (18)	80	75 (94)	{34.8}	B
Trouillet, *et al.*, 1996[119]	116	27 (23)	89	59 (66)	[81]	A
Gammie, *et al.*, 1998[38]	58	25 (43)	33	32 (97)	48	A

Measurement properties in critical care

Validity No details of the validity of the NYHA classification were presented in any of the papers.

Reliability No assessment of reliability was reported in any of the papers reporting this measure.

Responsiveness Kumar and colleagues[89] reported significant changes in NYHA class from pre-surgery to follow-up. None of the other papers assessed responsiveness.

ATS respiratory disease questionnaire

This questionnaire comprises ten sections and was first reported by Ferris.[120] The questions require 'yes' or 'no' responses, with an occasional 'not applicable', and cover the following areas: cough, phlegm, episodes of cough and phlegm, wheeze, breathlessness, chest colds and chest illness, past illness, occupational history, smoking and family history. It is interviewer- or self-completed.

Measurement properties outside critical care

The ATS respiratory disease questionnaire has been subjected to rather few assessments of its measurement properties, given the frequency with which it has been used. Osterman and colleagues[121] provided some evidence as to its criterion validity by demonstrating its association with FEV_1 measurements and the presence of respiratory symptoms. Similarly, Hopp and colleagues[122] showed that the ATS questionnaire could predict the presence of increased bronchial reactivity.

Application in critical care

The characteristics of the patient populations studied in the four papers that reported using

TABLE 26 *Characteristics of populations studied using the ATS respiratory disease questionnaire*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Elliott, *et al.*, 1987[17]	25 (13–42)	31	N/A	ARDS
Elliott, *et al.*, 1988[18]	28 (13–62)	N/A	N/A	ARDS
Ghio, *et al.*, 1989[19]	28.5 ± 12.5 (7–61)	51	N/A	ARDS
Weinert, *et al.*, 1997[91]	40 ± 12	67	Lung injury score: 2.4 ± 0.54 (1.25–3.25)	Acute lung injury

TABLE 27 *Numbers of participants in studies using the ATS respiratory disease questionnaire*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} [median] (range) (months)	Method of data presentation
Elliott, *et al.*, 1987[17]	38	0 (0)	38	16 (42)	48 (12–96)	A
Elliott, *et al.*, 1988[18]	42	0 (0)	42	30 (71)	6	A
Ghio, *et al.*, 1989[19]	41	N/A	N/A	25 (N/A)	(1.25–32.3)	A
Weinert, *et al.*, 1997[91]	69	35 (51)	34	24 (71)	{19} [15] (6–41)	A

the ATS respiratory disease questionnaire are presented in *Table 26*.[17–19,91] The mean reported ages of the participants ranged from 25 years to 40 years. Three studies reported the percentages of men participating; these ranged from 31% to 67%. Lung injury scores were reported in one study.

The numbers of patients studied ranged from 38 to 69 (*Table 27*). Mortality before follow-up ranged from 0% to 51%. The proportions of available patients who were followed-up ranged from 42% to 71%. The time points selected for follow-up ranged from 1 month to 96 months. Weinert and colleagues[91] administered the ATS questionnaire by post while in the remaining papers the mode of administration was not specified. The authors reported the numbers of individuals with specific problems.

Measurement properties in critical care

Validity Only Ghio and colleagues[19] presented data for construct validity by correlating the questionnaire with age at onset of ARDS ($p = 0.07$) and gender ($p = 0.53$). Criterion validity was assessed by comparison with measures of impairment but no significant association was found.

Reliability and responsiveness No attempts were made to assess either of these properties.

Walk test

There are 1-, 6- and 12-minute walk tests, during which the patient is asked to cover as much ground as possible in the allotted time. Following this, they are asked to assess their level of dyspnoea on a visual analogue scale which ranges from 'extremely short of breath' (0) to 'no shortness of breath' (10).[123] The test is used principally with patients suffering chronic obstructive pulmonary disease.[124,125]

Measurement properties outside critical care

There is little evidence of the measurement properties of the walk test, despite its common usage. Eakin and colleagues[124] reported moderate correlations with lung function and Roul and colleagues[126] reported a correlation of $r^2 = 0.42$ ($p < 0.01$) between the walk test and rate of utilisation of oxygen. Weaver and colleagues[127] reported concurrent validity between the walk test and a Pulmonary Functional Status Scale ($r^2 = 0.38$; $p < 0.001$).

Application in critical care

Three studies have employed the walk test.[25,128,129] Only Grotz and colleagues[25] presented data for mean age, percentage of participating men and acute severity score (*Table 28*). Two studies reported on the follow-up of general critical care patients and one provided data for patients

TABLE 28 *Characteristics of populations studied using the walk test*

Study	Mean age ± SD (years)	Male (%)	Severity score, mean ± SD	Type of patient
Weir & Waldmann, 1994[128]	N/A	N/A	N/A	General
Waldmann & Gaine, 1996[126]	N/A	N/A	N/A	General
Grotz, *et al.*, 1997[25]	33.6 ± 2.1	70	ISS 36.8 ± 1.6	MOF; trauma

admitted for trauma injuries with multiple organ failure. Details of the numbers of subjects participating in the studies are presented in *Table 29*. Two studies administered the walk test on more than one occasion and followed-up the patients at 2, 6 and 12 months.

In all studies, the walk test was observed. Only Grotz and colleagues[25] presented data as mean values and SEMs. Neither of the other papers presented any data on this measure.

Measurement properties in critical care

There was no evidence in any of the three papers relating to the assessment of validity, reliability or responsiveness over time.

Summary

- Five generic and three disease-specific measures of physical functional status have been used to assess the outcome of critical care.
- The measurement properties of the two most commonly used generic measures (Katz's ADL and the Karnofsky indexes) have been subject to some limited investigation in critical care survivors. In both cases there is some evidence of their construct and criterion validity and of their responsiveness. Reliability has not been investigated. The properties of three other generic measures (Barthel Index, activity level, functional state) have received little attention, so it is not possible to comment of their usefulness in critical care.
- The three disease-specific measures that have been used were each developed for particular patient groups – the NYHA functional class for cardiac disease, the ATS respiratory disease questionnaire for respiratory disease, and the walk test for chronic obstructive airways disease. Their measurement properties in critical care survivors are either unknown (criteria validity, reliability) or of limited scope – there is some evidence for construct validity for the ATS respiratory questionnaire and for the responsiveness of the NYHA functional status measure.
- Given the lack of adequate information on the properties of all the measures that have been used, no one measure stands out as superior to the others. Relative to other generic measures, the two multi-item scales about which most is known are Katz's ADL index and the Karnofsky Index. It seems prudent to use these in future critical care research so that their measurement properties can be fully investigated.
- Disease-specific measures may be appropriate when studying relevant patient groups.
- Survivors of intensive care were found to be limited in their daily activities (Katz's ADL index) during the early months of follow-up but some studies found that this returned to pre-admission levels by 6–12 months. Two longer-term studies found some degree of dependency in 52% of neurosurgical survivors at 2 years and in 65% of elderly survivors at about 5 years.

TABLE 29 *Numbers of participants in studies using the walk test*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} (months)
Weir & Waldmann, 1994[128]	N/A	N/A	N/A	N/A	2, 6, 12
Waldman & Gaine, 1996[129]	N/A	N/A	N/A	N/A	2, 6, 12
Grotz, *et al.*, 1997[25]	173	104 (60.1)	69	50 (72)	{58.8}

Chapter 4

Measures of mental functional status

Four generic measures of mental functional status (affect or mood) have been used (Profile of Mood States (POMS), Centre for Epidemiological Studies – Depression (CES–D) scale, Beck Depression Inventory (BDI), Hospital Anxiety and Depression scale (HAD)), and one disease-specific measure (Impact of Event Scale (IES)). All of the measures are multi-item and each is described and reviewed in turn.

POMS

POMS uses a list of adjectives rather than symptoms to describe mood[130,131] and was originally developed to assess mood in psychiatric outpatient clinics. It is a self-administered, 65-item (adjectives) measure of present mood state. A total score can be obtained, as can six subscale scores of affect or mood: tension–anxiety, depression–dejection, anger–hostility, vigour–activity, fatigue–inertia and confusion–bewilderment. Respondents rate the 65 adjectives on a five-point intensity scale based on how they have felt throughout the previous week. Scores range from zero for 'not at all' to four for 'extremely'. The higher the score, the greater the disturbance experienced, with the exception of the vigour–activity subscale.[132] A 72-adjective POMS also exists[133] that assesses bipolar mood states, and there is also an abbreviated version comprising 30 items with the same six subscales as the longer version.[131]

Measurement properties outside critical care

McNair and colleagues[130] reported adequate construct, content and factorial validity of the measure, and component analysis of the fatigue–inertia and vigour–activity subscales confirmed their factor structure.[134] In those studies which have used the POMS for assessment of patients with cancer, validity, reliability and sensitivity over time has been established, as well as population norms for this patient group.[130,135,136] Test–retest reliability values reported by the developers of the measure were 0.70 for tension–anxiety, 0.74 for depression–dejection, 0.71 for anger–hostility, 0.65 for vigour–activity, 0.66 for fatigue–inertia and 0.68 for confusion–bewilderment. Internal consistency was reported as ranging between 0.90 and 0.92 for anxiety and 0.90 and 0.95 for depression.[130]

Application in critical care

Five papers were identified in which POMS was used as an outcome measure in critical care patients (*Table 30*).[90,137–140] The mean age of the participants was reported in only two papers. The percentages of men who participated in these studies ranged from 49% to 84%. Two studies reported acute severity.

The numbers of patients eligible for participation ranged from 50 to 228 (*Table 31*). Mortality before follow-up ranged from 0% to 30%. Complete follow-up of available patients was achieved in three of the four studies that reported this variable. In four studies, the POMS questionnaire was

TABLE 30 *Characteristics of populations studied using POMS*

Study	Mean age ± SD (years)	Male (%)	Severity score, mean ± SD	Type of patient
Stanton, *et al.,* 1983[137]	N/A	84	N/A	Cardiac surgery
Stambrook, *et al.*, 1990[138]	Severe: 36 ±14.1 Moderate: 43 ±18.2 Mild: 39.5 ± 17.1	80	GCS: severe, 6.2 ± 1.97 moderate, 13.3 ± 2.0 mild, 13.2 ± 2.6	Head injury
Sawdon, *et al.*, 1995[139]	(0.3–94)	49	N/A	General
Jones, *et al.*, 1994[140]	N/A	N/A	N/A	General
Singh, *et al.*, 1997[90]	48 (28–68)	N/A	APACHE II: 14.5 APS: 7.9	Liver transplant

TABLE 31 *Numbers of participants in studies using POMS*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (range) (months)
Stanton, *et al.*, 1983[137]	228	0 (0)	228	228 (100)	6
Stambrook, *et al.*, 1990[138]	131	0 (0)	131	131 (100)	(4–98)
Sawdon, *et al.*, 1995[139]	100	20 (20)	80	57 (71)	6
Jones, *et al.*, 1994[140]	N/A	N/A	N/A	28 (N/A)	2, 6
Singh, *et al.*, 1997[90]	50	9 (30)	41	41 (100)	6

administered at 6 months and in one paper[140] the questionnaire was administered after hospital discharge on two occasions.

In two studies the POMS questionnaire was mailed.[137,139] Singh and colleagues[90] stated that the questionnaire was self-completed but did not report on the mode of administration, and Stambrook and colleagues[138] provided no information on the mode of administration.

In three studies no data from the POMS questionnaire were presented. Jones and colleagues[140] presented data as the distribution of scores across the POMS categories, and Singh and colleagues[90] presented the mean values for total POMS score and for each dimension, although no SD or SEM was provided.

Measurement properties in critical care

Validity Only two papers assessed the validity of POMS. Both Stanton and colleagues[137] and Jones and colleagues[140] reported on criterion validity, although this was not explicit in the text of either paper. Stanton and colleagues[137] reported a weak correlation of $r^2 = 0.04$ ($p < 0.01$) between POMS and 'return to work'. Jones and colleagues[140] found no correlation between POMS and the Whiston Hospital questionnaire (a measure of health-related quality of life – see page 76), although they did report correlations between the Perceived Quality of Life (PQOL) questionnaire[141] and the fatigue–inertia dimensions ($r^2 = 0.46$; $p = 0.01$) and hostility–anxiety dimensions ($r^2 = 0.48$; $p < 0.003$) of POMS.

Reliability and responsiveness There has been no assessment of reliability or responsiveness.

CES–D scale

The CES–D scale was developed by Radloff.[142,143] It is used to assess the frequency and severity with which symptoms of depression were experienced over the previous week. CES–D was designed as a diagnostic tool and is a 20-item, self-report scale with six main symptom areas: depressed mood, feelings of guilt/worthlessness, sense of helplessness/hopelessness, psychomotor retardation, loss of appetite and sleep disturbance. Scores on the CES–D scale range from zero to 60, higher scores being indicative of higher symptom frequency and severity. A score of 16 has been identified as indicating depressive symptomology.[144,145] In addition, a shortened version comprising eight items[146] and a modified (19-item) children's version have also been developed.

Measurement properties outside critical care

Criterion validity for this measure was established by comparison with the Hamilton Depression Scale ($r^2 = 0.19$–0.31).[147] The CES–D scale is also reported to correlate well with the BDI ($r^2 = 0.66$) and to demonstrate sensitivity to change over time.[148] Radloff[143] reported that the CES–D scale was able to discriminate well between the general population and psychiatric inpatients, and moderately well between levels of severity within patient groups.

During the developmental stages of the CES–D scale, Radloff[143] reported a wide range of inter-item ($r^2 = 0.001$–0.53), item-scale ($r^2 = 0.08$–0.62) and inter-scale ($r^2 = 0.1$–0.55) correlations. Internal consistency was reported as 0.85 for the general population and 0.90 for the patient population, and test–retest reliability ranged from 0.67 at 4 weeks to 0.32 at 12 months. Lyness and colleagues[149] noted the ability of the CES–D scale to discriminate between major and minor depression.

Application in critical care

The mean ages of the participants in the four studies that have used the CES–D scale ranged from 30 years to 81 years, and the percentages of men ranged from 44% to 74% (*Table 32*).[70,71,91,150]

TABLE 32 *Characteristics of populations studied using the CES–D scale*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD [median] (range)	Type of patient
Bedell, *et al.*, 1983[150]	70 (18–101)	54	N/A	CPR
Chelluri, *et al.*, 1993[70]	Group I: 69 ± 0.3 SEM Group II: 81 ± 0.5 SEM	Group I: 44 Group II: 52	APACHE II: Group I: 18 ± 0.9 SEM [18.8] Group II: 20 ± 0.8 SEM [19.5]	General
Holbrook, *et al.*, 1994[71]	30 ± 13.1 (18–69)	74	ISS: 15 ± 10.1 (5–43)	Trauma
Weinert, *et al.*, 1997[91]	40 ± 12	67	Lung injury score: 2.4 ± 0.54 (1.25–3.25)	Acute lung injury

TABLE 33 *Numbers of participants in studies using the CES–D scale*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} [median] (range) (months)	Method of data presentation
Bedell, *et al.*, 1983[150]	294	205 (70)	89	38 (43)	6	B
Chelluri, *et al.*, 1993[70]	97	59 (61)	38	24 (63)	1, 6, 12	C
Holbrook, *et al.*, 1994[71]	63	N/A	N/A	42 (N/A)	3	C
Weinert, *et al.*, 1997[91]	69	35 (51)	34	24 (71)	{19} [15] (6–41)	C

Acute severity scores were reported in three papers. The numbers of individuals eligible for inclusion ranged from 63 to 294 (*Table 33*). Mortality before follow-up ranged from 51% to 70%, further attrition being primarily due to loss to follow-up or refusal to participate. The proportions of available individuals who were followed-up ranged from 43% to 71%. In three studies, baseline data were obtained either on admission or at discharge from critical care and the follow-up periods ranged from 1 month to 41 months, with only one study using more than one follow-up.[70]

Chelluri and colleagues[70] did not specify how they administered the CES–D scale to their patients. In the other studies face-to-face interviews were employed, and Holbrook and colleagues[71] also made use of telephone interviews when patients were unable to attend face-to-face interviews. Mean scores were reported in all studies and Chelluri and colleagues,[70] Holbrook and colleagues,[71] and Weinert and colleagues[91] also provided SEMs or SDs for their data.

Measurement properties in critical care

Validity Only Weinert and colleagues[91] reported on criterion validity between the CES–D scale and a number of other scales: a correlation of $r^2 = 0.70$ with a life satisfaction scale, and $r^2 = 0.88$ with the mental component of the SF-36. None of the other papers presented data to support the testing of the validity of the CES–D scale.

Reliability None of the papers addressed this issue.

Responsiveness Bedell and colleagues[150] reported significant differences between CES–D scores ($p < 0.0001$) at hospital discharge (clinical depression) and those 6 months later, which had changed to within normal ranges for the population. Although Holbrook and colleagues[71] had data pertaining to pre-discharge and follow-up, they did not address the issue of responsiveness.

HAD scale

The HAD scale is a self-assessment instrument developed by Zigmond and Snaith[151] to measure mood disorders of anxiety and depression in non-psychiatric patients. It is designed for use by adults, although a children's version also exists. The scale comprises 14 items that are divided into two subscales, for which the patient rates each item on a four-point scale. One subscale contains seven items on depression and the other seven

items on anxiety. Patients are asked to assess their emotional state over the 'past week'.

The developers designed the scale to detect anhedonic depression and high scores on the HAD scale may therefore indicate that antidepressant therapy may be required. The HAD scale differs from other self-assessment scales in that it avoids the inclusion of items such as loss of appetite and insomnia which, although symptoms of anxiety and depression, may also be present in an individual suffering from physical illness. The focus of the scale is thus on psychological rather than somatic representations of mood disorder. The HAD scale is brief and easy to administer. The developers note that care has been taken to distinguish between the concepts of anxiety and depression, and the HAD scale incorporates clear guidance on the significance of a subscale score (normal, pre-morbid, morbid). The scale is not derived from factor analysis but from clinical experience.

Measurement properties outside critical care

Zigmond and Snaith[151] assessed the validity of the HAD scale with psychiatric outpatients and hospital staff. They reported that severity ratings correlated highly with psychiatric assessments ($r^2 = 0.49$ for depression; $r^2 = 0.55$ for anxiety). Aylard and colleagues[152] reported correlations ranging from 0.44 to 0.59 between the HAD and other recognised anxiety and depression scales. Lewis and Wessely[153] noted that the HAD scale was equal to the General Health Questionnaire in its ability to detect minor psychiatric disorders. Sensitivity to change was reported in a study of patients with neuroses.[154]

The structure of the scale attempts to overcome response bias by alternating the order of responses and by providing four options so that respondents would not opt for a middle grade. The developers reported correlations ranging from 0.41 to 0.76 for internal consistency.[151]

Application in critical care

There were three reports of the HAD scale being used as an outcome measure in following-up critical care patients, two published[15,16] and one personal communication (Eddleston & colleagues; personal communication, 1999) (*Table 34*). The mean ages of the patient populations ranged from 49 years to 56 years and the percentages of men were remarkably consistent, ranging from 57% to 59%. Acute severity scores were reported in two studies and all of the papers reported on general critical care patients.

The numbers of individuals eligible for inclusion ranged from 78 to 7988 (*Table 35*). Mortality before

TABLE 34 *Characteristics of populations studied using the HAD scale*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Rowan, 1992[155]	55.4 ± 0.7 16–90	58	APACHE II: 15.3 ± 0.3 (0–49)	General
Dixon, *et al.,* 1997[156]	56.2 ± 21	59	N/A	General
Eddleston, *et al.,* 1999*	49 ± 11.6	57	APACHE II: 18.8 ± 6.2	General

**Personal communication*

TABLE 35 *Numbers of participants in studies using the HAD scale*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean ± SD} (range) (months)
Rowan, 1992[155]	7988	N/A	N/A	2986 (N/A)	{6.7 ± 4.4} (1.1–22)
Dixon, *et al.*, 1997[156]	78	27 (36)	51	51 (100)	1
Eddleston, *et al.*, 1999*	370	144 (39)	226	143 (63)	3

**Personal communication*

follow-up ranged from 36% to 39% in the two papers in which this variable was reported, further attrition being primarily due to loss to follow-up and refusal to participate. The proportions of individuals who were followed-up ranged from 63% to 100%. There was a lack of consistency in time of follow-up and none of the papers reported serial administrations of the measure.

In two studies[155,156] the questionnaire was administered by postal survey and in one (Eddleston and colleagues; personal communication, 1999.) by telephone. Data from the HAD scale were presented in only two papers. Rowan[155] presented the percentages of patients who were depressed and anxious, and the numbers and percentages responding to the component questions of the HAD scale. Eddleston and colleagues presented data for the numbers of males and females scoring 0–7, 8–10 and 11 or more for the anxiety and depression dimensions.

Measurement properties in critical care

Validity Only Rowan[155] reported on the criterion validity of the HAD scale using summary health questions ($r^2 = 0.22$; $p < 0.0001$ for anxiety); ($r^2 = 0.42$; $p < 0.0001$ for depression). She reported that all dimensions of the NHP were significantly correlated with anxiety and depression ($p < 0.0001$).

Reliability and responsiveness Reliability and responsiveness were not addressed in any of the papers.

BDI

The BDI is a clinically derived scale developed by Beck and colleagues.[157] It covers a number of different dimensions: sadness, pessimism, discouragement, sense of failure, dissatisfaction, guilt, expectation of punishment, self-dislike, self-accusation, suicidal ideation, crying, irritability, social withdrawal, indecisiveness, body image distortion, work retardation, insomnia, fatiguability, anorexia, weight loss, somatic preoccupation, and loss of libido. The full scale comprises 21 items with four choices in the form of statements, which are ranked in order of severity. Patients are asked to respond to the statement that best represents how they have been feeling over the previous week, including the day on which they are questioned. The total scores of the BDI range from zero to 63, with 63 being the most severe score. The BDI can be self-administered or can be administered by interview; it takes about 10–15 minutes to complete.

Measurement properties outside critical care

Beck and colleagues[157] reported the level of agreement between the BDI and ratings by psychiatrists to be 56%. Beck[158] also reported correlations of $r^2 = 0.19$ and 0.31 between the BDI and other scales that measure depression (including the Hamilton Depression Scale)[159] and between the BDI and clinical ratings of $r^2 = 0.24$ and above.[160] The BDI is also sensitive to type of depression and can distinguish depressive symptoms from anxiety. Joseph and Lewis[161] reported a correlation with the depression–happiness scale of $r^2 = 0.31$, and Suarez-Mendoza and colleagues[162] reported a correlation of $r^2 = 0.69$ with the HAD scale.

Beck and colleagues[157] reported high internal consistency and a split-half reliability of 0.86. Test–retest reliability over 2–5 weeks was cited as 0.90 by Beck and colleagues,[157,158] and as ranging from 0.73 to 0.90 by Gallagher and colleagues,[163] who also reported an internal consistency of 0.86. More recent studies have indicated that the test–retest reliability of the BDI is rather low.[164]

Application in critical care

Two studies reported using the BDI.[90,165] The mean ages of patients eligible for participation in these studies ranged from 46 years to 48 years (*Table 36*). Only one study reported the percentage of men who participated in the study, and acute severity scores were also

TABLE 36 *Characteristics of populations studied using the BDI*

Study	Mean age ± SD (range) (years)	Male (%)	Mean severity score	Type of patient
Riether, *et al.*, 1992[165]	45.8 ± 12.1	67	N/A	Liver and heart transplant
Singh, *et al.*, 1997[90]	48 (28–68)	N/A	APACHE II: 14.5 APS: 7.9	Liver transplant

only reported in only one study. Both studies were reporting on outcomes in organ transplantation patients.

The numbers of patients who were eligible for participation in these studies ranged from 50 to 61 (*Table 37*). Mortality before follow-up was reported, in only one paper, as 30%. Singh and colleagues[90] achieved complete follow-up of available patients. However, these follow-up data were not given by Riether and colleagues.[165] Follow-up periods ranged from 3 months to 12 months, with only Riether and colleagues[165] following-up survivors on several occasions.

Riether and colleagues[165] administered the BDI during a face-to-face interview; Singh and colleagues[90] stated that the questionnaire was self-completed but did not report on the mode of administration. Riether and colleagues[165] presented data as mean values and SDs; Singh and colleagues[90] presented mean values only.

Measurement properties in critical care

Validity Riether and colleagues[165] assessed the criterion validity of the BDI using data from an electroencephalogram (EEG), a clinical measure of impairment which records electrical activity in the brain. A correlation of $r^2 = 0.18$ ($p < 0.001$) was reported between the two measures. Singh and colleagues[90] did not report any assessment of validity in their study.

Reliability There was no evidence on the reliability of the BDI in critical care.

Responsiveness Although Riether and colleagues[165] reported administering the BDI on multiple occasions, they did not address the issue of responsiveness.

IES

The IES is a 15-item, disease-specific measure which assesses levels of subjective post-traumatic psychological distress and which provides specific measures of event intrusion and event-related avoidance, the two key elements of post-traumatic stress disorder.[166] Patients specify the frequency with which they have had intrusion- or avoidance-related thoughts in the previous 7 days on a Likert scale ranging from zero (not at all) to 5 (often). The scores for the intrusion component of the scale range from zero to 35 and, for the avoidance component, zero to 40. The higher the score, the greater the level of distress indicated.

Measurement properties outside critical care

Internal consistencies with Cronbach's α ranging from 0.78 to 0.91 have been reported.[166,167] Criterion validity against the Trauma Symptom Inventory and the Los Angeles Symptom Checklist have also been established by Briere and Elliott,[168] who also presented normative data in their paper.

Application in critical care

The mean ages of the patients in the two studies that used this measure[169,170] were 36 years and 37 years and the proportions of men were 66% and 68%, respectively (*Table 38*). Both papers

TABLE 37 *Numbers of participants in studies using the BDI*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number followed-up	Follow-up time (months)
Riether, *et al.*, 1992[165]	51 heart & 61 liver transplants	N/A	N/A	Heart: 14, 14, 7 (N/A) Liver: 17, 17, 10 (N/A)	3, 6, 12
Singh, *et al.*, 1997[90]	50	9 (30)	41	41 (100)	6

TABLE 38 *Characteristics of populations studied using the IES*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Landsman, *et al.*, 1990[169]	36 (14–80)	66	ISS: 18.3 (4–50)	Trauma
Richmond, *et al.*, 1998[170]	37.4 ± 16.8	68	ISS: 15.5 ± 9.9	Trauma

TABLE 39 *Numbers of participants in studies using the IES*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time [median] (range) (months)
Landsman, *et al.*, 1990[169]	137	0 (0)	137	137 (100)	[15] (3–39)
Richmond, *et al.*, 1998[170]	228	N/A	N/A	109 (N/A)	3

reported acute severity scores and in both the groups studied were trauma patients.

The numbers of patients who were eligible were 137 and 228 in the two studies using this measure (*Table 39*). The actual numbers of patients who were followed-up were 137 and 109, respectively. Neither paper administered the outcome measure on more than one occasion. Landsman and colleagues[169] administered the IES by mail and Richmond and colleagues[170] used telephone interviews. Both studies provided mean scores for the IES but only one provided SDs.[170] This study also presented the range of scores and data for the total IES and for the intrusion and avoidance subscales.

Measurement properties in critical care

Validity Both papers presented data for criterion validity but in neither was there a significant association with injury severity and brief symptom inventory[169] or with extremity injury.[170] Richmond and colleagues[170] did, however, find a statistically significant correlation between the intrusion score and the SIP.

Reliability Richmond and colleagues[170] reported on internal consistency with Cronbach's $\alpha = 0.88$, but neither paper presented any data on inter-rater or intra-rater reliability.

Responsiveness Neither paper referred to the issue of responsiveness of the IES.

Summary

- Four generic and one disease-specific measure of mental functional status were used to assess the outcome of critical care.
- The generic measures all encompass depressive symptoms and two, POMS and the HAD scale, also cover anxiety. Researchers have demonstrated the criterion validity of each of these measures but there is no evidence of construct validity or reliability and there is evidence of responsiveness for only one measure, the CES–D.
- Given the similar level of knowledge of the properties of the four generic measures, there are no strong reasons to select one in preference to another on these grounds. The shorter length of the HAD (14 items) might make it preferable to the CES–D (20 items), BDI (21 items) and POMS (65 items).
- The only disease-specific measure, the IES, was designed for use in trauma patients to measure the level of post-traumatic distress. Lack of information on its measurement properties precludes drawing any conclusion as to its value.

Chapter 5

Measures of neuropsychological functioning

Seven measures of neuropsychological functioning have been used in critical care survivors. These measures are concerned with the cognitive functioning and skills of respondents. Six of the measures are multi-item scales (Trail-making Tests A & B, Wisconsin Card Sorting Test (WCST), Wechsler Memory Scale, Bentons' Visual Retention Test, Mini-Mental State Examination (MMSE), Paced Auditory Serial Addition Test (PASAT)) and only one is a single-item global measure (communication level). In all cases, unless otherwise stated, the mode of administration of these tests was by face-to-face interview and was often part of a larger battery of tests.

Trailmaking Tests A and B

The Trailmaking Test[171] was developed to assess attention, perceptual speed, cognitive flexibility and visual memory. It is a two-part instrument. Trailmaking Test A is a test of simple visual motor attention that is scored in seconds. Trailmaking Test B is a similar visual motor attention test that requires subjects to shift attention between two sets of stimuli; the latter test is also measured in seconds. An impaired score on the Trailmaking Test B is indicative of cognitive impairment.

Measurement properties outside critical care

Jones and colleagues[172] reported criterion validity between the Mental Alternation Test and Trail-making Test A ($r^2 = 0.28$; $p < 0.001$) and Test B ($r^2 = 0.29$; $p < 0.001$).

Application in critical care

Four papers reported using the Trailmaking Tests (*Table 40*).[137,165,173,174] The mean reported ages of participants ranged from 23 years to 46 years and the percentages of men ranged from 67% to 86%. Only two papers reported acute severity scores. Each paper reported on different patient populations within critical care.

The numbers of individuals eligible for participation ranged from 46 to 228 (*Table 41*). Mortality before follow-up ranged from 6% to 28% in the two papers reporting this variable, further attrition being due to refusal to participate and inability to complete the test. The proportions of available individuals who were followed-up were 30% and 100% in the two papers reporting this variable. Three papers reported administering the Trailmaking Tests at 6 months and two papers at 12 months. Two papers administered the Trailmaking Tests at multiple time points. In two papers the data were presented as a mean and SD and in one paper as a mean and range. Stanton and colleagues[137] presented no data related to the outcome measurement.

Measurement properties in critical care

Validity Two papers presented evidence of criterion validity, although this was not explicit in the text of the paper. Stanton and colleagues[137] correlated the Trailmaking Tests with return to work ($r^2 = 0.03$; $p < 0.05$), and Riether and colleagues[165] reported correlations of $r^2 = 0.20$ ($p < 0.001$) for Trailmaker Test A and $r^2 = 0.12$ ($p < 0.01$) for Test B with a measure of impairment (EEG). Uzzell and colleagues[173] and McKee and colleagues[174] cited no evidence for the assessment of validity.

TABLE 40 *Characteristics of populations studied using Trailmaking Tests*

Study	Mean age ± SD (years)	Male (%)	Severity score, mean ± SD	Type of patient
Stanton, *et al.*, 1983[137]	N/A	84	N/A	Cardiac surgery
Uzzell, *et al.*, 1987[173]	Mild: 24.5 ± 4.9 Severe: 23.2 ± 7.0	86	GCS: mild, 13.8 ± 1.4 severe, 5.2 ± 1.4	Head injury
Riether, *et al.*, 1992[165]	45.8 ± 12.1	67	N/A	Liver and heart transplant
McKee, *et al.*, 1997[174]	37.5	76	ISS: 33.2	Trauma

TABLE 41 *Numbers of participants in studies using the Trailmaking Tests*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (months)	Method of data presentation
Stanton, *et al.*, 1983[137]	228	14 (6)	214	214 (100)	6	N/A
Uzzell, *et al.*, 1987[173]	54	N/A	N/A	43 (N/A)	6	C
Riether, *et al.*, 1992[165]	51 heart & 61 liver transplants	N/A	N/A	Heart: 14 14 7 (N/A) Liver: 17 17 10 (N/A)	3 6 12 3 6 12	C
McKee, *et al.*, 1997[174]	46	13 (28)	33	10 (30)	12	C

Reliability and responsiveness There was no evidence for the assessment of reliability or responsiveness in any of the four papers reviewed.

WCST

The WCST is a measure of concept formation and cognitive flexibility.[175] It requires formulation of hypotheses by the subject of the test with regards to sorting strategies and how to test them. It also requires the ability to maintain and shift sets appropriately. Errors are made in the test by sorting cards into an incorrect category. If an individual continues to sort the cards into the previously correct category despite being informed that this is no longer correct, the errors are referred to as perseverative. Non-perseverative errors occur when sorting is made into an incorrect category (not the previously correct one).

Scoring is in terms of the number of categories successfully completed (maximum of six), the number of correct sorts (cards correctly sorted but not totalling up to a complete category of ten consecutive correct cards), the total number of errors, the number of perseverative and non-perseverative errors, the number of unique errors (cards sorted according to a category generated by the individual and other than the three types of errors described above).

Measurement properties outside critical care

The WCST was reported by Watkins and colleagues[176] to correlate with age for the total number of errors ($r^2 = 0.15$; $p = 0.002$) and for perseverative errors ($r^2 = 0.16$; $p = 0.001$). Zihl and colleagues[177] demonstrated that the WCST showed discriminant ability between patients with schizophrenia and patients with affective disorders for both total errors ($F = 45.76$; $p < 0.001$) and perseverative errors ($F = 25.63$; $p < 0.001$).

Application in critical care

The mean ages of participants in the three studies that have used this measure ranged from 30 years to 46 years (*Table 42*).[165,178,179] The percentages of men participating ranged from 67% to 92%. Only one paper[179] presented data for acute severity scores. There were two reports on the follow-up of patients who had received a head injury, while Riether and colleagues[165] presented data on the follow-up of liver- and heart-transplant patients.

TABLE 42 *Characteristics of populations studied using the WCST*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD	Type of patient
Riether, *et al.*, 1992[165]	45.8 ± 12.1	67	N/A	Liver and heart transplant
Anderson, *et al.*, 1994[178]	Severe: 30 (15–63) Moderate: 34 (15–68)	Severe: 86 Moderate: 92	N/A	Head injury
Lannoo, *et al.*, 1998[179]	33 ± 15	79	GCS: 6.5 ± 2.7	Head injury

TABLE 43 *Numbers of participants in studies using the WCST*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (months)
Riether, *et al.*, 1992[165]	51 heart & 61 liver transplants	N/A	N/A	Heart: 14 14 7 Liver: 17 17 10	3 6 12 3 6 12
Anderson, *et al.*, 1994[178]	88	20 (23)	68	61 (90)	6
Lannoo, *et al.*, 1998[179]	43	0 (0)	43	43 (100)	6, 12, 24

The number of individuals eligible for participation in the studies ranged from 43 to 88 (*Table 43*). Mortality before follow-up ranged from 0% to 23% in the two papers reporting on this variable, further attrition being due to loss to follow-up and an inability to complete the task. The proportions of available individuals who were followed-up ranged from 90% to 100%. Three papers reported follow-up data for 6 months and two papers for 12 months. Two papers presented data for three follow-ups. Follow-up periods ranged from 3 months to 24 months. All papers presented data as mean values and SDs. In addition, Anderson and colleagues[178] also presented data for error and perseverative scores.

Measurement properties in critical care

Validity Riether and colleagues[165] reported a non-significant correlation between EEG data (a measure of neurological impairment) and WCST. Anderson and colleagues[178] reported a correlation of $r^2 = 0.14$ for errors and $r^2 = 0.16$ for perseverative errors and the Glasgow Outcome Scale (GOS).

Reliability and responsiveness There was no evidence for the reliability or responsiveness of the WCST.

Wechsler Memory Scale

The Wechsler Memory Scale was developed to assess short- and long-term memory deficits and, according to the developer, should only be administered by those with specific training.[180,181] It is most frequently used as a diagnostic and screening tool alongside other neuropsychological batteries of tests. The original version comprised seven components, which formed three factors: orientation, memory/learning and attention/concentration. The new revised version comprises 12 subtests grouped under five memory scores: verbal memory, visual memory, general memory, delayed recall, attention/concentration. The battery of tests takes approximately 45–60 minutes to administer.

Measurement properties outside critical care

The Wechsler Memory Scale has been extensively tested for validity, reliability and sensitivity to change.[181] A user manual is available. Factor analysis yielded two main factors, which corresponded with general memory and learning, and attention/concentration. Test–retest correlation coefficients range from 0.41 to 0.90, and inter-rater reliability coefficients were 0.99 for logical memory and 0.97 for visual reproduction. The test may not be appropriate for administration to the elderly.[182]

Application in critical care

Three papers reported using the Wechsler Memory Scale (*Table 44*).[150,173,183] The mean ages of the patients ranged from 23 years to 70 years. The percentages of men participating ranged from 54% to 86%. Only one paper reported acute severity scores. Two papers[173,183] presented their data clearly with mean scores and SDs.

The numbers of individuals participating in studies using the Wechsler Memory Scale are shown in *Table 45*.

Measurement properties in critical care

Validity, reliability and responsiveness There was no attempt to assess the validity, reliability or responsiveness of the Wechsler Memory Scale in any of the three papers discussed here.

TABLE 44 *Characteristics of populations studied using the Weschler Memory Scale*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD	Type of patient
Bedell, *et al.,* 1983[150]	70 (18–101)	54	N/A	CPR
Uzzell, *et al.*, 1986[183]	N/A	83	N/A	Head injury
Uzzell, *et al.*, 1987[173]	Mild: 24.5 ± 4.9 Severe: 23.2 ± 7.0	86	GCS: mild, 13.8 ± 1.4 severe, 5.2 ± 1.4	Head injury

TABLE 45 *Numbers of participants in studies using the Weschler Memory Scale*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (range) (months)
Bedell, *et al.*, 1983[150]	294	205 (70)	89	38 (43)	6
Uzzell, *et al.*, 1986[183]	103	61 (59)	42	42 (100)	(2.7–5.9)
Uzzell, *et al.*, 1987[173]	54	N/A	N/A	43 (N/A)	6

TABLE 46 *Characteristics of populations studied using Benton's test for visual retention*

Study	Mean age ± SD (years)	Male (%)	Severity score, mean ± SD	Type of patient
Alexandre, *et al.*, 1983[188]	N/A	N/A	N/A	Neurological
Lannoo, *et al.*, 1998[179]	33 ± 15	79	GCS: 6.5 ± 2.7	Head injury

TABLE 47 *Numbers of participants in studies using Benton's test for visual retention*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (months)
Alexandre, *et al.*, 1983[188]	178	76 (43)	102	100 (98)	12, 24
Lannoo, *et al.*, 1998[179]	43	0 (0)	43	43 (100)	6, 12, 24

Benton's test for visual retention

This test of short-term visual memory was first described by Benton.[184,185] Participating individuals are presented with one of 15 different abstract designs. After the design is removed, the subject must identify the original from four alternatives. Error scores are compared to expected standard scores for educational background and age. Error scores of four or more above the expected score are classified as defective.

Measurement properties outside critical care

There is limited information available relating to the measurement properties of the test. Crookes and McDonald[186] noted that the test was able to discriminate between early dementia and depression. Weiss[187] discussed the equivalence of three different forms of the visual retention test.

Application in critical care

Of the two papers found,[179,188] only one provided any details of mean age, percentage of male participants and acute severity score (*Table 46*). The numbers of individuals eligible for participation ranged from 43 to 178 (*Table 47*). Mortality before follow-up ranged from 0% to 43%. Almost complete follow-up was achieved. Both studies followed-up patients at 12 months and 24 months, and both administered the measure at multiple time points. Neither study provided

details on the method of administration of the instrument. Lannoo and colleagues[179] presented data as mean scores and SDs. Alexandre and colleagues presented no data.[188]

Measurement properties in critical care

Neither of these studies attempted to assess the validity, reliability or responsiveness of Benton's test of visual retention for survivors of critical care.

MMSE

The MMSE is a brief test of cognitive mental state,[189] which has the ability to distinguish between organic and functional psychiatric illness. The MMSE was originally designed for use with neurogeriatric patients. It takes approximately 5–10 minutes to administer and comprises two parts (verbal and performance). The maximum score on the verbal subscale is 21 and, on the performance subscale, 9. The maximum overall score is therefore 30 with lower scores representing cognitive malfunction, the cut-off score being either 23 or 24. The MMSE encompasses orientation, registration, attention and calculation, recall and language. The test is interviewer-administered.

Measurement properties outside critical care

Folstein and colleagues[189] assessed criterion validity against the Wechsler Adult Intelligence Scale and found significant associations between the two measures. The majority of validation studies, which were evaluated by Nelson and colleagues,[190] reported associations between the subscales of the MMSE and clinical descriptions using both the Diagnostic and Statistical Manual (DSM) III diagnoses and radiological investigations.

Folstein and colleagues[189] reported responsiveness in a group of elderly people who had affective disorders with or without cognitive malfunction. Inter-rater and test–retest (0.83–0.98) reliability were reported as satisfactory.

Molloy and Standish[191] noted that, as the guidelines for application of the MMSE are brief, the administration and scoring of the test can vary considerably between individuals, thus diminishing the test's reliability. Some items have to be altered depending on where the test is being administered (home, hospital clinic) and as there are no clear guidelines on time limits, administrators of the test may not be sure how long they should wait for a response.

Application in critical care

The mean ages of patients in the two studies that used this measure[165,192] were 36 years and 46 years (*Table 48*). The percentages of men ranged from 67% to 81%, and only one paper presented acute severity scores. One study was based on trauma patients and the other on liver- or heart-transplant patients.

The numbers of individuals participating in studies using the MMSE are shown in *Table 49*.

Frutiger and colleagues[192] presented data as the number and percentage scoring below the cut-off score of 24, which indicated cognitive malfunction, while Riether and colleagues[165] presented a mean score with SD.

Measurement properties in critical care

There was no assessment of validity, reliability or responsiveness in either paper.

PASAT

PASAT assesses speed of information processing.[193] A random series of digits ranging from 1 to 9 are presented orally to the subject, who is instructed to add pairs of digits such that each number is added to the one immediately preceding it. To be correct, the response must be made before the presentation of the next stimulus item. Normalised data for healthy adults have been reported by Wiens and colleagues.[194]

TABLE 48 *Characteristics of populations studied using the MMSE*

Study	Mean age ± SD (range) (years)	Male (%)	Mean severity score	Type of patient
Frutiger, *et al.*, 1991[192]	35.6 ± 16.88 (5–84)	81	ISS: 29.3	Trauma
Riether, *et al.*, 1992[165]	45.8 ± 12.1	67	N/A	Liver and heart transplant

TABLE 49 *Number of participants in studies using the MMSE*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (range) (months)
Frutiger, *et al.*, 1991[197]	233	56 (24)	177	91 (51)	(60–96)
Riether, *et al.*, 1992[165]	51 heart & 61 liver transplants	N/A	N/A	Heart: 14	3
				14	6
				7 (N/A)	12
				Liver: 17	3
				17	6
				10 (N/A)	12

Measurement properties outside critical care

Litvan and colleagues[195] reported criterion validity between the PASAT and the Rey auditory verbal learning test ($p < 0.01$).

Application in critical care

Mean ages of the participants ranged from 30 years to 34 years in the two studies that used this measure (*Table 50*).[178,179] The percentages of men participating ranged from 79% to 92%. Only Lannoo and colleagues[179] presented data for acute severity scores (using the GCS). Both papers provided data for patients with head injuries. The numbers of patients eligible for participation were 43 and 88 (*Table 51*). Mortality before follow-up ranged from 0% to 23%, further attrition being due to loss to follow-up and inability to complete the task. The proportions of available patients who were followed-up ranged from 90% to 100%. Both studies administered the PASAT at 6 months and Lanoo and colleagues[179] administered it on three occasions. Both presented their data as mean scores with SDs.

Measurement properties in critical care

Validity Anderson and colleagues[178] assessed the criterion validity of the PASAT using a relatives' questionnaire, although their intention to assess this was not made explicit in the text of the paper. No evidence of validity was reported by Lannoo and colleagues.[179]

Reliability and responsiveness There was no assessment of the reliability or responsiveness of PASAT.

Communication level

This is a single-item global question, which has been used by several authors, although its measurement properties do not appear to have been established in any prior research. It categorises patients as: fully alert; communicates well

TABLE 50 *Characteristics of populations studied using the PASAT*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Anderson, *et al.*, 1994[178]	Severe: 30 (15–63) Moderate: 34 (15–68)	Severe: 86 Moderate: 92	N/A	Head injury
Lannoo, *et al.*, 1998[179]	33 ± 15	79	GCS: 6.5 ± 2.7	Head injury

TABLE 51 *Numbers of participants in studies using the PASAT*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (months)
Anderson, *et al.*, 1994[178]	88	20 (23)	68	61 (90)	6
Lannoo, *et al.*, 1998[179]	43	0 (0)	43	43 (100)	6, 12, 24

TABLE 52 *Characteristics of populations studied using communication level*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Cullen, *et al.*, 1976[108]	59	65	TISS: 43 ± 1.0	General
Cullen, *et al.*, 1984[109]	62	56	N/A	General
Slayter, *et al.*, 1986[110]	43.7 (0.25–82)	58	TISS: 88 ± 129.9	General
Nolla-Salas, *et al.*, 1993[111]	76 ± 3.7 (70–85)	N/A	APS: 11.2 ± 5.1 (1–23)	General

TABLE 53 *Numbers of participants in studies using communication level*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (months)
Cullen, *et al.*, 1976[108]	231	169 (73)	62	62 (100)	1, 3, 6, 12
Cullen, *et al.*, 1984[109]	206	138 (69)	68	61 (90)	1, 6, 12
Slayter, *et al.*, 1986[110]	100	11 (11)	89	89 (100)	1
Nolla-Salas, *et al.*, 1993[111]	97	51 (53)	46	44 (96)	6

but not as well as before illness; communicates inadequately; or comatose. It was developed by Cullen and colleagues.[108]

Application in critical care

The mean ages of the participants in the four studies ranged from 44 years to 76 years (*Table 52*).[108–111] The percentages of male participants ranged from 56% to 65% and three studies presented data for acute severity score. All four studies followed-up general critical care survivors.

The number of patients who were eligible for participation ranged from 97 to 231 (*Table 53*). Mortality before follow-up ranged from 11% to 73%, with further attrition being due to loss to follow-up. The proportions of available patients who were followed-up ranged from 90% to 100%. Three studies administered the questionnaire at 1 month and two at 6 and 12 months. Two studies administered the questionnaire at multiple time points.

Cullen and colleagues[108] invited patients back to outpatient appointments for assessment. If they were unable to attend they, or their carers, were then contacted by mail or telephone. Cullen and colleagues[109] relied on telephone interviews; Slayter and colleagues[110] and Nolla-Salas and colleagues[111] did not specify how they administered the measure. All four papers presented data on the numbers and percentages of patients in each response category.

Measurement properties in critical care

Validity and reliability There was no attempt in any of the four studies to assess the validity or reliability of the measure.

Responsiveness Although Cullen and colleagues,[108,109] examined outcome at multiple time points, they did not address the statistical differences in terms of responsiveness over time.

Summary

- Six generic multi-item measures of neuropsychological functioning have been used to assess the outcome of critical care, particularly but not exclusively in patients who have suffered head injuries.
- The measures relate to cognition, attention, information processing and memory. There is almost no evidence available on the properties of the measures when used in critical care; there is some weak evidence for the criterion validity of the Trailmaking Tests A and B and the WCST.
- Assessment of neuropsychological functioning may be of only limited value in general critical

care patients and should perhaps be used as a disease-specific measure, confined to patients who have suffered a head injury or other central neurological insult.

- One single-item global measure (communication level) was used in four studies but there is no information available on its measurement properties.

Chapter 6

Measures of recovery

Six measures of the extent of recovery have been used, only one of which, the GOS, is a multi-item scale. The other five are single-item global measures, three of which have a standard question structure or response categories (Chronic Health Evaluation (CHE), degree of recovery, productivity) and two do not (return to work, residence).

GOS

The GOS, a multi-item instrument, was originally developed as a means of describing overall social outcome in neurological and head-injured patients 6 months after the ictus.[196] It has four outcome categories. A 'good recovery' is one that equates to resumption of pre-injury activities with no or minimal neurological deficits or apparent changes in personality. A 'moderate disability' is associated with the ability to function independently at a reduced level as a result of personality, intellectual, or physical differences when compared with pre-injury status. Jennett[197] described this group as 'independent but disabled'. 'Severe disability' is characterised by an inability to function independently and an additional requirement for substantial care at home or in an institution as a result of physical or intellectual impairment. The final category is 'vegetative state', which is when patients show no evidence of meaningful responsiveness. Death is an additional category that is sometimes included.

Measurement properties outside critical care

There is limited evidence to support the measurement properties of the GOS outside critical care research. Jennett[197] reported inter-observer reliability of 95% agreement for 150 patients. Maas and colleagues[198] presented data for inter-rater and intra-rater agreement in scoring the GOS and found substantial discrepancies between groups, which does not support the reliability of the GOS.

Application in critical care

Twenty-nine studies have used the GOS as an outcome measure in their research following up critical care patients.[25,66,101,102,173,174,178,179,183,192,199–217] The characteristics of the patient populations studied in these papers are shown in *Table 54*. In three studies[66,201,204] a modified version of the GOS was used.

The mean ages of the those eligible for participation ranged from 22 years to 57 years. The percentages of men participating in the studies ranged from 41% to 92% in the 24 papers reporting on this variable. Acute severity-of-illness scores were presented in 14 studies. A majority of studies (21) reported data on patients who had been admitted for head injury or for neurological reasons, five reported on general trauma patients and four were based on general critical care patients.

The numbers of patients eligible for participation ranged from 37 to 980 (*Table 55*). Mortality before follow-up ranged from 0% to 61% in the 26 papers reporting this variable, with further attrition being due to loss to follow-up. Follow-up of available patients ranged from 21% to 100%, the latter achieved in 18 of the 26 studies reporting on this variable. In 14 studies the GOS was administered at 6 months and in seven studies (24%) data were presented for 12 months. Ten studies reported GOS values at more than one time.

Eleven studies did not specify the mode of administration of the GOS.[66,101,200,203,205,206,208,210–212,214] Of those that did, Neundörfer and colleagues[102] used mailed questionnaires, Zarén and Hedstrand[201] and Zarén and Bergström[204] used telephone interviews, while the rest used face-to-face interviews.

No data were presented by Neundörfer and colleagues,[102] Fernandez and colleagues,[210] and Vazquez-Mata and colleagues.[212] The most frequent mode of presentation was the number and/or percentage of patients in each category of the GOS. Mean scores were reported by Zarén and Hedstrand[201] and means and SDs were reported by Grotz and colleagues.[25]

Measurement properties in critical care

Validity Fifteen papers did not report any details relating to the validity of the GOS. Details of the

TABLE 54 *Characteristics of populations studied using the GOS*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Severity score, mean ± SD [median] (range)	Type of patient
Levy, *et al*., 1981[199]	59	51	N/A	Neurological
Levy, *et al*., 1985[66*]	[61]	61	N/A	Neurological
Uzzell, *et al*., 1986[183]	N/A	83	N/A	Neurological
Alberico, *et al*., 1987[200]	38.6	79	GCS: 5.4	Neurological
Uzzell, *et al*., 1987[173]	Mild: 24.5 ± 4.9 Severe: 23.2 ± 7.0	86	GCS: mild, 13.8 ± 1.4 severe, 5.2 ± 1.4	Head injury
Zarén & Hedstrand, 1987[201*]	50 ± 19.2 (15–92)	56	N/A	General
Stocchetti, *et al*., 1988[202]	N/A	61	N/A	Head injury
Nordström, *et al*., 1989[203]	[21] (7–61)	N/A	N/A	Neurological
Zarén & Bergström, 1989[204*]	54	58	N/A	General
Judson, *et al*., 1990[205]	[20] (10–70)	73	ISS: [33] (16–59)	Neurological
Marshall, *et al*., 1991[206]	29.5 [25]	(3:1)	N/A	Neurological
Frutiger, *et al*., 1991[192]	35.6 ± 16.9 (5–84)	81	ISS: 29.3	Trauma
Wärme, *et al*., 1991[207]	Group 1: 37 Group 2: 36	N/A	N/A	Neurological
Fearnside, *et al*., 1993[208]	22.4	76	N/A	Neurological
Anderson, *et al*., 1994[178]	Severe: 30 (15–63) Moderate: 34 (15–68)	Severe: 86 Moderate: 92	N/A	Head injury
Jones, *et al*., 1994[209]	Severe: 34 Moderate: 37 Minor: 43	84	ISS: [25]	Head injury
Gobiet, 1995[101]	23 (6–80)	66	N/A	Trauma; neurological
Neundörfer, *et al*., 1996[102]	56.7 ± 18.8	56	N/A	General
Fernandez, *et al*., 1996[210]	N/A	N/A	N/A	General
Cruz, 1996[211]	31 (15–72)	N/A	GCS: 5.7 (3–8)	Neurological
Vazquez-Mata, *et al*., 1996[212]	31.2 ± 0.9	78	APACHE II: 13.5 ± 0.4 ISS: 23.6 ± 0.6	Trauma
Heinzelmann, *et al*., 1996[213]	39 ± 16	N/A	ISS: 29.5 ± 7.6 GCS: 10.5 ± 4.0	Neurological
McKee, *et al.,* 1997[174]	37.5 (16–63)	76	ISS: 33.2 (16–57)	Trauma
Pohlmann-Eden, *et al.,* 1997[214]	39.6 ± 19.3 (3–75)	69	GCS: 6.6 ± 3.1 [7]	Neurological
Enblad & Persson, 1997[215]	53 (female) 49 (male)	41	N/A	Neurological
Resnick, *et al*., 1997[216]	31.1	81	GCS: 5.2	Neurological
Grotz, *et al*., 1997[25]	33.6 ± 2.1	70	ISS: 36.8 ± 1.6	MOF; trauma
Cruz, 1998[217]	Group 1: 30 ± 9 Group 2: 29 ± 8	N/A	GCS: Group 1, 5.5 ± 1 Group 2, 5.6 ± 1.2	Neurological
Lannoo, *et al*., 1998[179]	33 ± 15	79	GCS: 6.5 ± 2.7	Neurological

**Modified version*

TABLE 55 *Numbers of participants in studies using the GOS*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time [median] (range) (months)
Levy, *et al.*, 1981[199]	500	305 (61)	195	60 (31)	1, 3, 6, 12
Levy, *et al.*, 1985[66]	210	121 (57)	89	19 (21)	1, 3, 6, 12
Uzzell, *et al.*, 1986[183]	103	61 (59)	42	42 (100)	(2.7–5.9)
Alberico, *et al.*, 1987[200]	230	104 (45)	126	126 (100)	(3–96)
Uzzell, *et al.*, 1987[173]	54	N/A	N/A	43 (N/A)	6
Zarén & Hedstrand, 1987[201]	980	259 (26)	721	717 (99)	1, 6, 12
Stocchetti, *et al.*, 1988[202]	354	145 (41)	209	209 (100)	6
Nordström, *et al.*, 1989[203]	587	261 (45)	326	326 (100)	6
Zarén & Bergstrom, 1989[204]	980	317 (32)	663	663 (100)	1, 6, 12
Judson, *et al.*, 1990[205]	100	36 (36)	64	64 (100)	6
Marshall, *et al.*, 1991[206]	746	243 (33)	503	503 (100)	(0.4–39) [22]
Frutiger, *et al.*, 1991[192]	233	56 (24)	167	167 (100)	60–96
Wärme, *et al.*, 1991[207]	121	N/A	N/A	N/A	6
Fearnside, *et al.*, 1993[208]	315	97 (31)	218	218 (100)	6
Anderson, *et al.*, 1994[178]	88	20 (23)	68	61(90)	6
Jones, *et al.*, 1994[209]	124	26 (21)	98	98 (100)	[12] (11–13)
Gobiet, 1995[101]	558	0 (0)	558	558 (100)	36
Neundörfer, *et al.*, 1996[102]	422	219 (52)	203	203 (100)	(18–30)
Fernandez, *et al.*, 1996[210]	578	0 (0)	578	578 (100)	6
Cruz, 1996[211]	205	36 (17.5)	169	159 (94)	6
Vasquez–Mata, *et al.*, 1996[212]	351	N/A	N/A	351 (N/A)	12, 24
Heinzelmann, *et al.*, 1996[213]	139	13 (9)	126	126 (100)	0
McKee, *et al.*, 1997[174]	46	13 (28)	33	20 (61)	32 (12–59)
Pohlmann–Eden, *et al.*, 1997[214]	42	24 (57)	18	18 (100)	3
Enblad & Persson, 1997[215]	61	18 (30)	43	43 (100)	14 [13.8] (2–30)
Resnick, *et al.*, 1997[216]	37	6 (16)	31	31 (100)	6, 12
Grotz, *et al.*, 1997[25]	173	104 (60.1)	69	50 (72)	58.8
Cruz, 1998[217]	353	62 (17.6)	291	291 (100)	6
Lannoo, *et al.*, 1998[179]	43	0 (0)	43	43 (100)	6, 12, 24

assessment of construct and criterion validity in the other 14 studies are shown in *Table 56*. In nine studies, construct validity was considered by examining the relationship between GOS and age. A statistically significant association was found in five studies, in two no significant association was found, and in two no indication was provided. Only one study reported on association with gender and found no significant relationship. These findings provide some support for the validity of the GOS in critical care survivors.

Criterion validity was assessed in nine studies, although in two[192,208] there was no indication of the statistical significance of the finding. There was evidence of significant association between the GOS and various measures of impairment (somatosensory evoked potential, brainstem auditory

TABLE 56 *Assessment of validity of the GOS*

Study	Construct validity	Criterion validity
Levy, *et al.,* 1985[66]	Age, gender (χ^2, NS)	N/A
Alberico, *et al.*, 1987[200]	Age ($p < 0.001$, correlation)	N/A
Zarén & Hedstrand, 1987[210]	Age (χ^2, $p < 0.001$)	N/A
Nordström, *et al.*, 1989[203]	Age	N/A
Zarén & Bergstrom, 1989[204*]	Age ($p = 0.007$)	TISS ($p = 0.08$); logistic regression
Judson, *et al.*, 1990[205]	N/A	Sensory evoked potential ($\chi^2 = 50.7$, $p < 0.0001$) GCS ($\chi^2 = 12.1$, $p < 0.01$)
Marshall, *et al.*, 1991[296]	Age (no value)	Intracranial diagnosis ($p < 0.001$) GCS ($p < 0.001$)
Frutiger, *et al.*, 1991[192]	N/A	Working status at 5 years (no value)
Wärme, *et al.*, 1991[207]	Age ($p < 0.007$) ANOVA	ANOVA GCS motor score ($p < 0.003$)
Fearnside, *et al.*, 1993[208]	N/A	Abnormal motor response, CT scan logistic regression (no value)
Anderson, *et al.*, 1994[178]	N/A	Logical memory (immediate): 0.43, $p < 0.01$ Logical memory delayed: 0.36, $p < 0.01$ Rey Osterrieth (copy): 0.34, $p < 0.01$ Rey Osterrieth (recall): 0.42, $p < 0.01$ PASAT (number correct): 0.59, $p < 0.001$ PASAT (longest string): 0.41, $p < 0.01$ WCST (errors): 0.38, $p < 0.01$ WCST (perseverative errors): 0.40, $p < 0.01$
Jones, *et al.*, 1994[209]	Age (logistic regression, NS)	GCS: intracranial pressure
Neundörfer, *et al.*, 1996[102]	Age ($p < 0.001$)	N/A
Pohlmann-Eden, *et al.,* 1997[214]	N/A	Somatosensory evoked potential ($r = -0.70$, $p = 0.0001$) BAEP # ($r = -0.50$, $p = 0.0001$) GCS (NS)

**Modified GOS*

evoked potential (BAEP), intracranial diagnosis, GCS) and of neuropsychological functional status (memory, PASAT, WCST). These findings provide limited evidence of the criterion validity of GOS.

Reliability Only one study investigated intra-rater reliability[192] but no details were reported.

Responsiveness Fernandez and colleagues[210] used the GOS as a reference measure to assess the responsiveness of their own quality-of-life measure (see page 75) (weighted kappa 0.56; $p < 0.001$).

Outcome of critical care survivors

Moderate to good recovery ranged from 7% to 61% in the 15 studies reporting this outcome.[66,101,173,174,178,179,192,199,200,203,206,209,214–216] In one study that retrospectively assessed pre-critical care health status,[201] 88% of survivors were able to live independently 12 months after critical care compared with 94% 3 months before admission. Judson and colleagues[205] reported a favourable outcome in 55% of subjects. Zarén and Bergström[204] reported that there had been an improvement in functional status in 12% of their patients and a decrease in 22%.

Return to work

Return to work is a proxy measure of functional status. As such, identification of the measurement properties outside critical care was not sought. It is a single-item global measure that has been used in many studies but does not appear to have a consistent format.

Application in critical care

In all, 34 papers reported using return to work when following-up critical care patients (*Table 57*).[25,60,61,88,91,107,137,138,155,173,174,201,218–239] The mean ages of the participants ranged from 23 years to 74 years and the percentages of male participants ranged from 29% to 86%. Acute severity scores were reported in 16 studies. General critical care survivors were followed-up in 13 papers and trauma survivors in eight. The remaining papers reported on more disease-specific populations who had required critical care treatment.

The numbers of patients eligible for participation ranged from 40 to 7988 (*Table 58*). Mortality before follow-up ranged from 0% to 68% in the 25 papers reporting this variable, with further attrition being due to loss to follow-up and refusal to participate. The proportions of available patients followed-up ranged from 38% to 100%, with the latter achieved in five of the 26 papers reporting this variable. Six studies followed-up patients after 6 months and six papers at 12 months. Multiple follow-ups were reported in nine of the papers.

Return to work data were collected in a variety of ways. Ten papers obtained the data by mail, eight used telephone interviews and nine used face-to-face interviews. Grotz and colleagues,[25] Nordback and Auvinen,[60] and Kivioja and colleagues[232] invited survivors to attend outpatient appointments; Alho and Rokkanen[220] retrieved data on the basis of outpatient reports and medical notes; and Pessi[219] reported using both case notes and questionnaires. Eight papers did not report the mode of administration. All papers presented results as numbers and percentages of survivors by employment status.

Measurement properties in critical care

Validity Construct validity was assessed in relation to age in six studies (*Table 59*). Three studies found a statistically significant relationship and three failed to do so. One study[228] also examined the relationship with gender, marital status and education. They found a significant association with education but not with gender or marital status. Criterion validity was assessed by comparison with measures of impairment (GCS), mental functional status (POMS), neuropsychological function (Trailmaking Test A), and health-related quality of life (SIP). Statistically significant associations were found with each measure (except for one dimension of POMS in the study by Stambrook and colleagues[138]).

Reliability None of the papers reported on reliability.

Responsiveness The responsiveness of the measure was not systematically assessed in any paper.

Outcome of critical care survivors

Hurel and colleagues[237] noted that employment status remained unchanged in 80% of subjects. However, there was a significant change in the level of work activity from before to after critical care. There was an increase in the number of retired survivors and those on sick leave, and a decrease in the number of survivors in full- or part-time employment. Kriwanek and colleagues[239] reported that worsened employment status was noted by 25% of respondents in their study.

MacKenzie and colleagues[228] reported that, at 1 year, 17% of their patients were unable to work and Alho and Rokanen[220] reported that 68% of their patients who were capable of returning to work had done so. Broome and colleagues[235] reported that 80% of individuals had been working before admission to critical care and, of these, 70% had returned to work 10 months, on average, after their discharge from hospital. Rowan[155] noted that 35% of survivors reported working full- or part-time before admission but, of these, 42% were not working 6 months after discharge from critical care. According to Kivioja and colleagues,[232] 72% of those who had been in work before admission to critical care had been able to return to work, the majority of whom were in full-time employment. Parno and colleagues[223] presented data indicating that 54% of patients younger than 41 years of age, 39% of patients aged between 41 years and 65 years, and 4% of those older than 65 years had returned to work 2 years after hospitalisation. Mundt and colleagues[230] reported that whereas 50% of their sample had been employed beforehand, only 36% were employed at follow-up. McKee and colleagues[174] reported that 60% of those followed up were able to return to work. According to Sage and colleagues,[225,227] only 18% and 11% of respondents, respectively, reported that their employment status had been affected by their health. Of those patients who had been in a post-anoxic unconscious state for at least 30 days, only 5% returned to full employment and 78% still required 24-hour nursing care. Of those below retirement age, 65% were working at least part-time before admission compared with 56% at 1 year after admission. In contrast, the employment rate increased from 38% to 40% between before

TABLE 57 *Characteristics of populations studied using return to work*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Pessi, 1973[219]	50 ± 21.8 (1–93)	65	N/A	Surgical
Alho & Rokkanen, 1973[220]	N/A	72	N/A	Trauma
Bürgisser & Ritz, 1982[221]	54.5 ± 17.8 (14–89)	66	N/A	General
Stanton, *et al*., 1983[137]	N/A	84	N/A	Cardiac surgery
Bergner, *et al*., 1984[222]	N/A	N/A	N/A	CPR
Parno, *et al*., 1984[223]	55 ± 4.6 [58]	N/A	N/A	General
Nordback & Auvinen, 1985[60]	48 (33–65)	N/A	N/A	Pancreatitis
Goldstein, *et al*., 1986[224]	N/A	N/A	N/A	General
Sage, *et al*., 1986[225]	58.4	50	APACHE II: 13.3	General
Bosatra, *et al*., 1987[226]	Group 1: 48 ± 19 Group 2: 47 ± 19	Group 1: 35.5 Group 2: 29	N/A	Medical
Sage, *et al*., 1987[227]	74.7 ± 0.7 (65–96)	50	N/A	General
Uzzell, *et al*., 1987[173]	Mild: 24.5 ± 4.9 Severe: 23.2 ± 7.0	86	GCS: mild, 13.8 ± 1.4 severe, 5.2 ± 1.4.	Head injury
Zarén & Hedstrand, 1987[201]	50 ± 19.2 (15–92)	56	N/A	General
MacKenzie, *et al*., 1988[228]	(18–35)	78	N/A	Trauma
Ritz, 1988[229]	56 (8–86)	63	N/A	General
Yinnon, *et al*., 1989[88]	54 (15–102)	62	APACHE II: 16 ± 8	General
Mundt, *et al*., 1989[230]	59 ± 18.2	57	N/A	General
Gaillard, *et al*., 1990[231]	[32] (10–55)	60	N/A	Suicide attempt
Stambrook, *et al*., 1990[138]	Severe: 36 ± 14.1 Moderate: 43 ± 18.2 Mild: 39.5 ± 17.1	80	GCS: severe, 6.2 ± 1.97 moderate, 13.3 ± 2.0 mild, 13.2 ± 2.6	Head injury
Kivioja, *et al*., 1990[232]	45 (22–80)	N/A	ISS: 38.9 ± 1.2 (17–66)	Trauma
Schuster, 1991[107]	N/A	N/A	N/A	General
Rowan, 1992[155]	55.4 ± 0.7 (16–90)	58	APACHE II: 15.3 ± 0.3 (0–49)	General
Morris, *et al*., 1991[233]	36	N/A	ISS: 33	Trauma
Doepal, *et al*., 1993[61]	49 (26–90)	68	N/A	Pancreatitis
Sazbon, *et al*., 1993[234]	38 (2–80)	63	N/A	Neurological
Bell & Turpin, 1994[218]	54 (19–83)	52	APACHE II: 12 (0–29)	General
Broome, *et al*., 1996[235]	Group 1: 50 Group 2: 53 Group 3: 50	73 68.5 55	APACHE: 9.0 N/A N/A	Pancreatitis
Brenneman, *et al*., 1997[236]	37	65	ISS: 25	Trauma
Grotz, *et al*., 1997[25]	33.6 ± 2.1	70	ISS: 36.8 ± 1.6	MOF; trauma
McKee, *et al*., 1997[174]	37.5	76	ISS: 33.2	Trauma
Hurel, *et al*., 1997[237]	51.6 (17.9)	56	SAPS: 11.6 ± 4.7	General
Weinert, *et al*., 1997[91]	40 ± 12	67	Lung injury score: 2.4 ± 0.54 (1.25–3.25)	Acute lung injury
Schelling, *et al*., 1998[238]	(18–85)	51	Lung injury score (2.75–3.75)	ARDS
Kriwanek, *et al*., 1998[239]	53 (23–79)	65	APACHE II: (6–33)	Pancreatic surgery

TABLE 58 *Number of patients in studies reporting return to work*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} [median] (range) (months)
Pessi, 1973[219]	1001	334 (33)	667	667 (100)	12
Alho & Rokkanen, 1973[220]	258	41 (15.9)	217	202 (93)	> 24
Bürgisser & Ritz, 1982[221]	330	98 (27)	232	89 (38)	36
Stanton, *et al.,* 1983[137]	228	14 (6)	214	214 (100)	6
Bergner, *et al*., 1984[222]	472	N/A	N/A	424 (N/A)	6
Parno, *et al*., 1984[223]	558	182 (33)	376	216 (57)	24
Nordback & Auvinen, 1985[60]	40	15 (38)	25	24 (96)	(60–132)
Goldstein, *et al*., 1986[224]	2213	518 (23)	1695	1695 (100)	{8.2}
Sage, *et al*., 1986[225]	341	83 (25)	258	140 (54)	(15–20)
Bosatra, *et al*., 1987[226]	689	N/A	N/A	367 (N/A)	2, 8
Sage, *et al*., 1987[227]	134	37 (28)	97	59 (61)	18
Uzzell, *et al*., 1987[173]	54	N/A	N/A	43 (N/A)	within 16
Zarén & Hedstrand, 1987[201]	980	259 (26)	721	717 (99)	1, 6, 12
MacKenzie, *et al*., 1988[228]	597	N/A	N/A	479 (N/A)	6, 12
Ritz, 1988[229]	1508	N/A	N/A	N/A	3, 12, 36
Yinnon, *et al*., 1989[88]	126	71 (56)	55	52 (95)	6
Mundt, *et al*., 1989[230]	1345	84 (6)	1261	887 (70)	6
Gaillard, *et al*., 1990[231]	160	57 (36)	103	46 (45)	[42]
Stambrook, *et al*., 1990[138]	131	0 (0)	131	131 (100)	(4–98)
Kivioja, *et al*., 1990[232]	213	91 (43)	122	92 (75)	(60–240)
Schuster, 1991[107]	1308	N/A	N/A	N/A	(12–60)
Rowan, 1992[155]	7988	N/A	N/A	2986 (N/A)	{6.7 ± 4.4} (1.1–22)
Morris, *et al*., 1991[233]	114	16 (14)	98	88 (90)	{31.2}
Doepal, *et al*., 1993[61]	67	23 (34)	44	37 (84)	[74.4] (12–168)
Sazbon, *et al*., 1993[234]	100	68 (68)	32	18 (56)	(12–72)
Bell & Turpin, 1994[218]	95	N/A	N/A	60 (N/A)	3
Broome, *et al*., 1996[235]	Group 1: 40 Group 2: 89 Group 3: 47	8 (20) N/A N/A	32 N/A N/A	22 (69) N/A N/A	51 39.5 28.3
Brenneman, *et al*., 1997[236]	439	2 (4)	437	195 (45)	12
Grotz, *et al*., 1997[25]	173	104 (60)	69	50 (72)	58.8
McKee, *et al*., 1997[174]	46	13 (28)	33	18 (55)	12
Hurel, *et al*., 1997[237]	329	N/A	N/A	223 (N/A)	6
Weinert, *et al*., 1997[91]	69	35 (51)	34	24 (71)	{19} [15] (6–41)
Schelling, *et al*., 1998[238]	192	90 (47)	102	80 (78)	[48]
Kriwanek, *et al*., 1998[239]	147	55 (37)	92	92 (100)	24

and after critical care in the study conducted by Yinnon and colleagues.[88] Doepel and colleagues[61] reported 84% normal working capacity before illness and a reduction to 65% at follow-up.

According to Morris and colleagues,[233] 61% of survivors were working at the time of injury and, of these, only 28% returned to work. In the study by Brenneman and colleagues,[236] 51% of trauma victims had returned to employment 1 year later, as had 55% of patients with severe head injuries studied by Stambrook and colleagues.[138] In their sample of patients with acute pancreatitis, Nordback and Auvinen[60] reported an increasing number of survivors who had to retire during the follow-up period on ill-health grounds. Some 40% of patients surviving cardiopulmonary resuscitation were unemployed before admission but this had increased to 51% at the follow-up interview. Goldstein and colleagues[224] noted that 65% of survivors were employed either full- or part-time at follow-up. Gaillard and colleagues[231] reported that 85% had returned to work after a median of 7.5 months. Ritz[229] reported that 90% had returned to work by follow-up at 1 year. Of those questioned at 2 months, 18% were unable to return to professional activity but this had reduced to 7% after 8 months.[226] Schuster[107] reported that whereas 65% were employed 3 months pre-injury, this was reduced to 56% at 12 months post-injury.

Overall, a fairly consistent pattern emerges. In the 16 studies that reported on the change in employment status, 11 found that over 70% of those in work before their critical care episode had returned to work. In some studies, the proportion resuming work was as high as 85–90%. In contrast, in some studies work resumption proportions were as low as 45–55%. These were mostly trauma victims.

Residence

Residence is a proxy measure that assumes that those patients who have not reached a sufficient level of functional recovery are more likely to be institutionalised or to live at home with assistance than those who have made a better recovery. No attempt was made to identify the measurement properties of this single-item global measure outside critical care. Like 'return to work', this measure lacks a standard format.

Application in critical care

There were 18 studies that reported using place of residence as an outcome measure (*Table 60*).[70,74,76,88,100,109–111,118,201,241–248] The mean ages of the participants ranged from 44 years to 88 years and the percentages of male participants ranged from 47% to 75%. Acute severity scores were reported in 13 papers. A total of 11 papers reported on the outcome of general patients and four on trauma survivors; the remainder examined the outcome of more specific populations within critical care.

The numbers eligible for participation ranged from 34 to 1832 (*Table 61*). Mortality before

TABLE 59 *Assessment of validity of 'return to work'*

Study	Construct validity	Criterion validity
Alho & Rokkanen, 1973[220]	Age ($p < 0.001$)	N/A
Stanton, *et al.*, 1983[137]	N/A	POMS ($r = –0.02$, $p < 0.01$) Trailmaking Test A ($r = –0.16$, $p < 0.05$)
Zarén & Hedstrand, 1987[201]	Age (χ^2, $p < 0.001$)	N/A
MacKenzie, *et al.*, 1988[228]	Age, gender, prior marital status (NS) Education ($p < 0.01$)	N/A
Stambrook, *et al.*, 1990[138]	Age (NS)	GCS ($r = 0.23$, $p < 0.05$) POMS: tension-anxiety (NS); confusion ($r = –0.24$, $p < 0.05$); depression ($r = –0.20$, $p < 0.005$) SIP: physical subscale ($r = –0.38$, $p < 0.01$); psychosocial subscale ($r = –0.30$, $p < 0.01$) Katz adjustment scale: relative verbal expansiveness ($r = –0.29$, $p < 0.01$)
Morris, *et al.*, 1991[233]	Age ($p < 0.005$);	Functional status ($p < 0.001$)
Grotz, *et al.*, 1997[25]	Age (χ^2, NS)	N/A

TABLE 60 *Characteristics of populations studied using 'place of residence'*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Campion, *et al.*, 1981[241]	N/A	N/A	N/A	General
Cullen, *et al.*, 1984[109]	62	56	N/A	General
Goldstein, *et al.*, 1984[242]	N/A	N/A	N/A	Medical
Oreskovich, *et al.,* 1984[243]	75	50	APACHE: 19 (4–61)	Trauma
McLean, *et al.*, 1985[244]	N/A	N/A	APS: 16.1 ± 7.8	Respiratory
Slayter, *et al.*, 1986[110]	43.7 (0.25–82)	58	TISS: 88 ± 129.9	General
Zarén & Hedstrand, 1987[201]	50 ± 19.2 (15–92)	56	N/A	General
Yinnon, *et al.*, 1989[88]	54 (15–102)	62	APACHE: 16 ± 8	General
Mahul, *et al.*, 1991[245]	76.4 ± 4.55	50	SAPS: 14.5 ± 5.3	General
Goins, *et al.*, 1991[246]	44.9 ± 19.85	75	ISS: 34 ± 15.9 GCS: 11.9 ± 3	Trauma
Chelluri, *et al.*, 1992[247]	88 ± 3	52	APACHE II: 18 ± 5 APS: 10 ± 5 TISS: 28 ± 10	General
Nolla-Salas, *et al.*, 1993[111]	76 ± 3.7 (70–85)	N/A	APS: 11.2 ± 5.1 (1–23)	General
Chelluri, *et al.*, 1993[70]	75	49	APACHE II: [18.8]	General
Day, *et al.*, 1994[100]	N/A	N/A	ISS: 25	Trauma
Dardaine, *et al.*, 1995[74]	78 ± 0.7	59	SAPS: 15 ± 0.6	General
McHugh, *et al.,* 1997[118]	77.4 ± SEM 0.3	47	N/A	Cardiac surgery
Douglas, *et al.*, 1997[248]	61.4 ± 19.9	53	APACHE III: 66.5 ± 25.3	General
Battistella, 1998[76]	85 ± 3.9 (77–79)	N/A	ISS: 9.4 ± 7.7	Trauma

TABLE 61 *Numbers of participants in studies using 'place of residence'*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean ± SD} (range) (months)
Campion, *et al.*, 1981[241]	1832	497 (32)	1335	1283 (96)	(6–18)
Goldstein, *et al.*, 1984[242]	1256	368 (29)	888	888 (100)	{8.1}
Cullen, *et al.*, 1984[109]	206	138 (69)	68	61 (90)	1, 6, 12
Oreskovich, *et al.,* 1984[243]	100	15 (15)	85	85 (100)	12+
McLean, *et al.*, 1985[244]	49	20 (41)	29	14 (48)	(12–24)
Zarén & Hedstrand, 1987[201]	980	259 (26)	721	717 (99)	1, 6, 12
Slayter, *et al.*, 1986[110]	100	11 (11)	89	89 (100)	1
Yinnon, *et al.*, 1989[88]	126	71 (56)	55	52 (95)	6
Mahul, *et al.*, 1991[245]	295	103 (48)	192	106 (55)	1, 6, 12
Goins, *et al.*, 1991[246]	87	15 (17)	72	72 (100)	3, 12
Chelluri, *et al.*, 1992[247]	34	13 (38)	21	21(100)	{18 ± 10} (1–32)
Nolla-Salas, *et al.*, 1993[111]	N/A	N/A	N/A	44 (N/A)	N/A
Chelluri, *et al.*, 1993[70]	97	59 (61)	38	38 (100)	12
Day, *et al.*, 1994[100]	118	36 (31)	82	56 (68)	(24–51)
Dardaine, 1995[70]	110	66 (60)	44	36 (82)	6, 12, 18
McHugh, *et al.*, 1997[118]	97	17 (18)	80	78 (98)	{34.8}
Douglas, *et al.*, 1997[248]	58	29 (50)	29	27 (93)	6
Battistella, 1998[76]	279	132 (47)	147	93 (63)	{64.8 ± 13.2}

follow-up ranged from 11% to 69%. The proportions of available participants who were followed-up ranged from 48% to 100%, the latter being achieved in six studies. Six studies reported on follow-up at 6 months and six at 12 months. Ten papers reported measuring this outcome at several time points.

Six papers identified place of residence by mailed enquiry,[76,100,118,241,242,245] one used face-to-face interviews,[88] and the mode of administration was not specified in six papers.[70,110,111,243,246,248] The nine remaining studies used telephone interviews. All papers presented data as the percentages of patients in each residential category, such as 'home', 'nursing home' or 'hospital'.

Measurement properties in critical care

Validity Two papers made reference to the assessment of construct validity[118,201] using the ages of patients, although neither paper explicitly tested for this type of validity. There were no attempts to assess criterion validity.

Reliability There was no evidence for the assessment of reliability.

Responsiveness Although Zarén and Hedstrand,[201] Battistella and colleagues,[76] and Day and colleagues[100] had data for before and after critical care episodes, none used the data to assess responsiveness.

Outcome of critical care survivors

According to Goins and colleagues,[246] 57% of survivors went to a rehabilitation centre and 26% returned to their own homes. Slayter and colleagues[110] reported 65% of patients had returned home by follow-up at 1 month. Zarén and Hedstrand[201] noted that most subjects had been living at home before admission to critical care (90%) and were again doing so 1 year after discharge (86%). Yinnon and colleagues[88] observed that the residence of 92% was unchanged following critical illness. According to McLean and colleagues,[244] 71% of those who left critical care were able to live at home by follow-up. Only 18% of patients in the study by Douglas and colleagues[248] were living independently at home after 6 months. Battistella and colleagues[76] reported that 94% of survivors had been living at home before admission to critical care; this had decreased to 69% at discharge and risen again to 83% at follow-up. Similarly, in the study by Mahul and colleagues,[245] 56% of survivors had returned home by 1 month follow-up and 88% by 1 year. Oreskovich and colleagues[243] noted that whereas beforehand 96% of patients had been living at home independently, by follow-up this had been drastically reduced to 8%, and 72% still required full nursing care at 1 year. In contrast, by 2 or more years later, only 14% required nursing or hospital care, with 85% discharged to their own homes from hospital and 93% at home.

Overall, the vast majority of critical care survivors have returned home within a few months. Three studies found much lower proportions[243,246,248] for reasons which are unclear.

CHE

Knaus and colleagues first reported CHE used as a single-item global measure of functional status.[249] It was used as a means to determine pre-admission health status and as a component of the APACHE II score. CHE ranges from A–D (A, no functional limitations; B, mild-to-moderate limitation of activity; C, serious but not incapacitating restriction of activity; D, severe restriction, including persons who are bedridden or institutionalised due to illness). CHE is rarely used outside critical care, as it is part of the APACHE II score. As it was designed for use within critical care, its properties outside this area are not available.

Application in critical care

The mean ages of the participants in the five studies that have used this measure[74,104,119,245,250] ranged from 50 years to 78 years (*Table 62*) and the percentage of male participants ranged from 50% to 69%. Four papers reported acute severity. Three papers reported following-up general critical care survivors, one alcohol-dependent individuals and one patients following cardiac surgery.

The numbers eligible for participation ranged from 26 to 295 (*Table 63*). Mortality before follow-up ranged from 23% to 67%. The proportions of available participants who were followed-up ranged from 54% to 100%. Four studies administered the CHE at several time points.

Data were primarily collected from telephone interviews in four of the five studies.[74,119,245,250] Le Gall and colleagues[250] and Mahul and colleagues[245] also used postal and face-to-face interviews, respectively, to maximise the response rate while Jensen and colleagues[104] only used mailed questionnaires. The primary method of data presentation was to report the numbers and percentages of individuals in each group.

TABLE 62 *Characteristics of populations studied using CHE*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Le Gall, *et al.*, 1982[250]	50 (15–82)	58	N/A	General
Jensen, *et al.*, 1988[104]	56.2	69	APS: 19.1 (4–36) TISS: 32 (9–52)	Alcohol-dependent
Mahul, *et al.*, 1991[245]	76.4 ± 4.55	50	SAPS: 14.5 ± 5.3	General
Dardaine, 1995[74]	78 ± 0.7	59	SAPS: 15 ± 0.6	General
Trouillet, *et al.*, 1996[119]	60.5 ± 12.2	62	SAPS: 9.7 ± 4 GCS: 13.1 ± 3	Cardiac surgery

TABLE 63 *Numbers of participants in studies reporting on CHE*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time [median] (months)
Le Gall, *et al.*, 1982[250]	228	116 (51)	112	112 (100)	3, 6, 9, 12
Jensen, *et al.,* 1988[104]	26	13 (50)	13	13 (100)	3, 6, 9, 12
Mahul, *et al.*, 1991[245]	295	103 (48)	192	103 (54)	1, 6, 12
Dardaine, 1995[74]	110	74 (67)	36	36 (100)	6, 12, 18
Trouillet, *et al.*, 1996[119]	116	27 (23)	89	59 (66)	[81]

Measurement properties in critical care

Validity and reliability There was no evidence for the assessment of validity or reliability. Knaus and colleagues[249] reported that inter-observer reliability had been assessed although there is no statistical report in this paper.

Responsiveness Mahul and colleagues[245] assessed the difference between the CHE obtained on admission and 1 year later to provide improved, same or worsened functional status but no statistical analysis is provided for the assessment of responsiveness. Le Gall and colleagues[250] reported a statistically significant change from before admission to follow-up ($\chi^2 = 23.15$; $p < 0.001$).

Degree of recovery

This is a measure that has been used in several studies, although its psychometric properties do not appear to have been established in any prior research. It comprises four response categories: full recovery, progressing to full recovery, partial recovery at best, no improvement. It appears to have first been used by Cullen and colleagues.[108]

Application in critical care

Four papers were identified which reported using degree of recovery.[108–111] The mean ages of participants ranged from 44 years to 76 years (*Table 64*) and the percentage of male participants ranged from 56% to 65%. Three papers reported acute severity scores. The numbers eligible for participation ranged from 100 to 231 (*Table 65*). Mortality before follow-up ranged from 11% to 73%. In two studies,[108,109] the measure was administered on several occasions.

Cullen and colleagues[108] invited patients back to outpatient clinics for assessment. If they could not attend they were contacted by mail or telephone. Slayter and colleagues[110] and Nolla-Salas and colleagues[111] did not specify how they measured degree of recovery. Cullen and colleagues[108,109] and Slayter and colleagues[110] reported on the numbers and percentages of patients in each category. Nolla-Salas and colleagues[111] reported only the numbers of patients in each category.

Measurement properties in critical care

Validity and reliability None of the papers reported evidence of its validity or reliability.

TABLE 64 *Characteristics of populations studied using degree of recovery and productivity*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Cullen, *et al.*, 1976[108]	59	65	TISS: 43 ± 1.0	General
Cullen, *et al.*, 1984[109]	62	56	N/A	General
Slayter, *et al.*, 1986[110]	43.7 (0.25–82)	58	TISS: 88 ± 129.9	General
Nolla-Salas, *et al.*, 1993[111]	76 ± 3.7 (70–85)	N/A	APS: 11.2 ± 5.1 (1–23)	General

TABLE 65 *Numbers of participants in studies using degree of recovery and productivity*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (months)
Cullen, *et al.*, 1976[108]	231	169 (73)	62	62 (100)	1, 3, 6, 12
Cullen, *et al.*, 1984[109]	206	138 (69)	68	61 (90)	1, 6, 12
Slayter, *et al.*, 1986[110]	100	11 (11)	89	89 (100)	1
Nolla-Salas, *et al.*, 1993[111]	N/A	N/A	N/A	44 (N/A)	6

Responsiveness Although Cullen and colleagues[108,109] examined outcome at multiple time points, they did not address the statistical differences in terms of responsiveness over time.

Responsiveness Although Cullen and colleagues[108,109] examined outcome at multiple time points, they did not address the statistical differences in terms of responsiveness over time.

Productivity

Productivity refers to the extent to which an individual can engage in 'activities'. It comprises six response categories: as productive as before illness, limited, active with assistance, independent self-care, no self-care, hospitalised or nursing home.[108] This measure has been used in several studies although its measurement properties do not appear to have been established by any prior research.

Application in critical care

The characteristics of the patients in the four studies that used this measure[108–111] have already been described above and appear in *Tables 64* and *65*.

Measurement properties in critical care

Validity and reliability There was no attempt in any of the four papers to assess validity or reliability.

Summary

- Six measures of recovery have been used, only one of which is a multi-item scale (GOS). The others are all global measures based on a single item or question, two of which lack both a standard, uniform structure and response categories (return to work, residence).
- There is some limited evidence of the construct and criterion validity of the GOS and the use of return to work in critical care survivors. The validity of the other four measures has not been investigated.
- There is no information available on the reliability and responsiveness of any of these measures in critical care.
- The outcome of critical care survivors varied considerably between studies, partly dependent on the nature and severity of the reasons for the patients' admission. The majority of survivors (over 70%) were able to return to work and return to their own homes.

Chapter 7

Measures of health-related quality of life

Nine health-related quality-of-life measures have been used to assess the outcome of critical care survivors.

SIP

The SIP was developed in the USA by Bergner and colleagues[251,252] as a measure of perceived health status across a large number of health problems and diseases in different demographic and cultural groups. The SIP focuses on the resultant impact that sickness has on functional status and quality of life.

The SIP was developed on the basis of a literature review and the perceptions of both ill and healthy lay people as well as health professionals.[253] During its development, the SIP was administered to a wide range of in- and outpatients, home-care patients with chronic disease, critical care patients, and hip replacement and arthritis patients. The SIP can be self- or interviewer-administered and takes 20–30 minutes to complete. The questionnaire comprises 136 questions which are either affirmed or not by respondents. It assesses: work, recreation, emotion, affect, home life, sleep, rest, eating, ambulation, mobility, communication and social functioning. Scores can be obtained for the 12 dimensions as well as physical and psychosocial summary scores and a global aggregated score (all measured in the range 0–100) with a lower score representing good health. In the UK, modifications to the SIP have resulted in the Functional Limitations Profile (FLP).[254]

Measurement properties outside critical care

Good results for test–retest reliability (0.88–0.92) and internal consistency (0.81–0.97) were reported by Bergner and colleagues.[253] Deyo and colleagues reported test–retest reliability (0.91),[255] as did de Bruin and colleagues (0.75–0.85);[256] the latter also assessed internal consistency (0.91–0.95) and inter-rater reliability (0.87–0.92).

Criterion validity has been established by Bergner and colleagues,[253] using the ADL index[62] with a correlation of $r^2 = 0.41$, and Hall and colleagues,[257] using the RAND Mental Health Index ($r^2 = 0.10$–0.29). Bergner and colleagues also established construct (discriminant) validity. Bowling[94] reported the results of several studies which suggested that the SIP was not particularly responsive to changes in health status, and de Bruin and colleagues[256] have queried the construct validity of SIP measures because the results of factor analysis have varied considerably between studies.

Applications in critical care

Twenty papers have used the SIP or the anglicised version, the FLP.[21,73,107,138–141,165,170,172,222,225,227,230,248,258–262] The mean ages of the participants ranged from 36 years to 75 years (*Table 66*). The percentages of male participants ranged from 50% to 80%. Acute severity scores were reported in almost half the papers. Eleven studies were based on general critical care patients.

The numbers of participants eligible for follow-up ranged from 58 to 6424 (*Table 67*). Mortality before follow-up ranged from 0% to 66%. The proportions of available participants who were followed-up ranged from 7% to 100%. In ten studies the SIP was administered at 6 months after critical care and in four at 12 months. Four studies used several time points.

In nine studies the SIP was administered by mail,[107,140,172,225,227,230,260–262] in six during face-to-face interviews,[73,139,141,165,258,259] and in four by telephone interview.[73,170,225,227] Four studies[21,138,222,248] provided no clear details of how the SIP was administered, although Bergner and colleagues[222] and Stambrook and colleagues[138] did note that the SIP was administered during an interview.

A variety of methods were used to present the data from these studies. No SIP data were presented in four papers. Patrick and colleagues[141] represented all the categories of the SIP graphically and both Sage and colleagues[227] and Miranda[260] tabulated all categories. In ten papers a mean SIP score was presented with SD or 95% confidence interval (CI). In eight papers data were presented consistently with mean total SIP scores and physical and psychosocial dimension scores.

TABLE 66 *Characteristics of populations studied using the SIP/FLP*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD [median] (range)	Type of patient
Bergner, *et al.*, 1984[222]	N/A	N/A	N/A	CPR
Sage, *et al.*, 1986[225]	58.4	50	APACHE II: 13.3 TISS: 104	General
Sage, *et al.*, 1987[227]	74.5	50	N/A	General
Danis, *et al.*, 1988[258]	68.9 ± 8.4	52	APACHE II: 13.0 ± 7	Medical; respiratory
Patrick, *et al.*, 1988[141]	69 ± 8.0	52	APACHE II: 13.0 ± 6	Medical; respiratory
Mundt, *et al.*, 1989[230]	59 ± 18.2	57	N/A	General
Stambrook, *et al.*, 1990[138]	Severe: 36 ± 14.1 Moderate: 43 ± 18.2 Mild: 39.5 ± 17.1	80	GCS: severe, 6.2 ± 1.97 moderate, 13.3 ± 2.0 mild, 13.2 ± 2.6	Head injury
Schuster, 1991[107]	N/A	N/A	N/A	General
Hulsebos, *et al.*, 1991[259]	47 ± 23.5 (1–92)	63	N/A	General
Riether, *et al.*, 1992[165]	45.8 ± 12.1	67	N/A	Liver and heart transplant
Jones, *et al.*, 1993[172]	N/A	58	APACHE II: [12] (2–21)	General
Jones, *et al.*, 1994[140]	N/A	N/A	N/A	General
Miranda, 1994[160]	64 ± 19	64	APACHE: 20.1 ± 8.6 TISS: 29.3 ± 11.9	CPR
McHugh, *et al.*, 199[421]	41 (19–73)	62	N/A	ARDS
Tian & Miranda, 1995[261]	60.1 ± 15	N/A	APACHE II: 9.7± 5.2	General
Sawdon, *et al.*, 1995[139]	(0.3–94)	50	N/A	General
Wu, *et al.*, 1995[73]	62.1	55	N/A	General
Douglas, *et al.*, 1997[248]	61.4 ± 19.9	53	APACHE III: 66.5 ± 25.3	General
Grady, *et al.*, 1998[262]	53 ± 9 (24–71)	80	N/A	Heart transplant
Richmond, *et al.*, 1998[170]	37.4 ± 16.8	68	ISS: 15.5 ± 9.9 (1–51)	Trauma

Measurement properties in critical care

Validity Some form of validity was reported on in 11 papers although in none of them was this an explicitly stated purpose of the research (*Table 68*). Jones and colleagues[172] reported a correlation between the physical and psychosocial dimensions of the SIP ($r^2 = 0.64$; $p < 0.0001$). Factor analysis was conducted by Tian and Miranda,[261] who reported that the structure of the SIP, as applied to survivors of critical care, was similar to the structure as described by the developers of the instrument.

Construct validity was assessed using the variable of age in four studies. Weak associations were found between age and several dimensions of SIP,[222,225] and with total SIP score.[261] One study did not find even a weak association.[260] Direct and proxy measures of severity (Therapeutic Intervention Scoring System (TISS), APACHE II, Injury Severity Score (ISS), length of stay) were either weakly or not associated with SIP scores. One study[170] considered pre-admission employment status, income and educational level and demonstrated only weak associations. Overall, these findings supported the construct validity of SIP in critical care survivors.

Criterion validity was assessed using a variety of measures. SIP scores were significantly associated with one measure of impairment (EEG) but not another (D_LCO). There was also no association with the level of social networks that an individual had but there was an association with employment status. There was also evidence of the criterion

TABLE 67 *Numbers participants in studies using the SIP/FLP*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} [median] (range) (months)
Bergner, *et al.*, 1984[222]	472	N/A	N/A	424 (N/A)	6
Sage, *et al.*, 1986[225]	337	83 (25)	254	140 (55)	(15–20)
Sage, *et al.*, 1987[227]	156	37 (28)	119	59 (50)	{18} (16–20)
Danis, *et al.*, 1988[258]	193	83 (43)	110	69 (63)	N/A
Patrick, *et al.*, 1988[141]	160	70 (44)	90	69 (77)	[19]
Mundt, *et al.*, 1989[230]	1345	84 (6)	1261	887 (70)	6
Stambrook, *et al.*, 1990[138]	131	0 (0)	131	131 (100)	(4–98)
Schuster, 1991[107]	1308	N/A	N/A	N/A	(12–60)
Hulsebos, *et al.*, 1991[259]	330	74 (22)	256	157 (61)	N/A
Riether, *et al.*, 1992[165]	112	N/A	N/A	17 (N/A)	3, 6, 12
Jones, *et al.*, 1993[172]	216	69 (32)	147	49 (33)	6
Jones, *et al.*, 1994[140]	N/A	N/A	N/A	60 (N/A) 44 (N/A)	6 12
Miranda, 1994[260]	477	313 (66)	164	69 (42) 12 (7)	6 24
McHugh, *et al.*, 1994[21]	216	134 (62)	82	20 (24)	3, 6, 12
Tian & Miranda, 1995[261]	6424	N/A	N/A	3655 (N/A)	12
Sawdon, *et al.*, 1995[139]	100	29 (29)	71	57 (80)	6
Wu, *et al.*, 1995[73]	3619	1306 (36)	2313	1746 (75)	2
Douglas, *et al.*, 1997[248]	58	29 (50)	29	6 (21)	6
Grady, *et al.*, 1998[262]	269	N/A	N/A	219 (N/A)	6
Richmond, *et al.*, 1998[170]	228	N/A	N/A	109 (N/A)	3

validity of the SIP in the observed associations with the IES, PQOL and a non-standard quality-of-life measure.[172]

Reliability Jones and colleagues[172] reported an inter-rater reliability of 0.85 ($p < 0.0001$) between patients' and relatives' responses to the SIP. Cronbach's α, as a measure of internal consistency, was cited as 0.94 by Richmond and colleagues[170] and it is assumed that this figure related to the overall SIP score, although this was not made explicit. None of the other papers reported assessments of reliability.

Responsiveness Miranda[260] reported no significant differences in SIP scores administered at two different time points. Significant differences were found between pre-admission and follow-up using χ^2 analyses for housework, leisure activity, and social contact ($p < 0.01$).

Outcome in critical care survivors

From the nine papers that reported each dimension of the SIP in detail, consistent findings were that the five dimensions which had the highest scores (poorest quality of life) at follow-up were work, home life, recreation, sleep, and rest.

PQOL

The PQOL[141] is a cognitive measure of quality of life and satisfaction. It comprises 11 items with scores ranging from zero to 100, with higher scores indicating higher satisfaction. The items in the questionnaire encompass health, thinking, happiness, family, help, community, leisure, income, respect, meaning and work. The average of the question scores provides a summated rating score. This was designed as a measure

TABLE 68 *Assessment of validity of the SIP/FLP*

Study	Construct validity	Criterion validity
Bergner, *et al.*, 1984[222]	Age: household management (r = 0.19), mobility (r = 0.17), ambulation (r = 0.23), physical dimension (r = 0.17)	N/A
Sage, *et al.*, 1987[225]	Age: physical dimension (r = 0.41); linear regression	N/A
Patrick, *et al.*, 1988[141]	N/A	Total SIP with PQOL (r = –0.49, p = 0.0001); psychosocial dimension with PQOL (r = –0.48, p = 0.0001); physical dimension with PQOL (r = –0.33, p = 0.004)
Stambrook, *et al.*, 1990[138]	N/A	Physical dimension with employment status (r = –0.38, p < 0.01); psychosocial dimension with employment status (r = –0.30, p < 0.05).
Hulsebos, *et al.*, 1991[259]	Length of stay (NS)	N/A
Riether, *et al.*, 1992[165]	N/A	Total SIP with EEG (r = –0.39, p < 0.001); physical dimension with EEG (r = –0.35, p < 0.01); psychosocial dimension with EEG (r = –0.31, p < 0.01); Pearson correlation
Jones, *et al.*, 1993[172]	N/A	FLP and own scale (r = 0.7, p < 0.0001)
Miranda, 1994[260]	Age; TISS; APACHE II (NS); factor analysis	N/A
McHugh, *et al.*, 1994[21]	N/A	SIP with D_LCO (NS)
Tian & Miranda, 1995[261]	Age; TISS; APACHE II (< 0.23); factor analysis	N/A
Richmond, *et al.*, 1998[170]	ISS; employment; income; years of education: (> 0.2); regression analysis; injury type (χ^2, NS)	SIP with social network (NS); with IES (> 0.2); linear regression

for medical patients in critical care so no attempt has been made to review its measurement properties in other fields.

Application in critical care

A total of 11 studies have used the PQOL (*Table 69*).[67,70,118,139–141,172,237,258,263,264] The mean ages of participants ranged from 23 years to 77 years and the percentages of male participants from 47% to 58%. Seven papers reported an acute severity score.

The numbers eligible for participation ranged from 83 to 385 (*Table 70*). Mortality before follow-up was reported in nine papers, ranging from 18% to 61%. The proportions of available patients who were followed-up ranged from 39% to 77%. Six studies administered the PQOL at 6 months after discharge from critical care and three after 12 months. Only Jones and colleagues[140] and Chelluri and colleagues[70] reported administering the PQOL at multiple time points, and the studies by Thiagarajan and colleagues[263] and Chelluri and colleagues[70] were the only ones in which patients were asked to recall their quality of life before their illness.

The PQOL was administered by mail in nine studies, one study used telephone interviews,[139] three employed face-to-face interview techniques[141,258,263] and only one[70] did not report the method of administration.

Seven studies reported their data using means and SDs, SEMs or 95% CIs,[70,118,141,237,258,263,264] four did not report any details of the PQOL, and one[172] illustrated the data in a scatter plot.

Measurement properties in critical care

Validity There was no evidence for the construct validity of the PQOL in critical care. Criterion validity has been assessed in four studies. Patrick and colleagues[141] demonstrated correlations between the PQOL and the SIP (r^2 = 0.24; p = 0.0001) and the PGWB index (r^2 =0.29;

TABLE 69 *Characteristics of populations studied using the PQOL scale*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Severity score, mean ± SD [median] (range)	Type of patient
Danis, *et al.*, 1988[258]	68.9 ± 8.4	52	APACHE II: 13.0 ± 7	Medical; respiratory
Patrick, *et al.*, 1988[141]	69 ± 8.0	52	APACHE II: 13.0 ± 6	Medical; respiratory
Thiagarajan, *et al.*, 1994[263]	22.5 (20–29)	N/A	APACHE: 6 (4–8)	Multiple trauma
Ridley & Wallace, 1990[67]	N/A	N/A	N/A	General
Jones, *et al.*, 1993[172]	N/A	58	APACHE II: [12] (2–21)	General
Chelluri, *et al.,* 1993[70]	75	49	APACHE II: [18.8]	General
Ridley, *et al.*, 1994[264]	53 (95% CI, 47 to 59)	N/A	APACHE II: 11.8 (95% CI, 9.4 to 14.2)	General
Jones, *et al.,* 1994[140]	N/A	N/A	N/A	General
Sawdon, *et al.*, 1995[139]	(0.3–94)	49	N/A	General
McHugh, *et al.*, 1997[118]	77.4 [0.3]	47	N/A	Cardiac
Hurel, *et al.*, 1997[237]	51.6 ± 17.9	56	SAPS: 11.6 ± 4.7	General

TABLE 70 *Numbers of participants in studies using the PQOL scale*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} [median] (range) (months)
Danis, *et al.*, 1988[258]	193	83 (43)	110	69 (63)	N/A
Patrick, *et al.*, 1988[141]	160	70 (44)	90	69 (77)	[19]
Ridley & Wallace, 1990[67]	385	129 (34)	256	156 (61)	(12–36)
Jones, *et al.*, 1993[172]	85	12 (14)	73	49 (67)	6
Chelluri, *et al.*, 1993[70]	97	59 (61)	38	24 (63)	1, 6, 12
Thiagarajan, *et al.*, 1994[263]	83	16 (19)	67	42 (63)	{10.4} (9.6–13.3)
Ridley, *et al.*, 1994[264]	90	33 (37)	57	41 (72)	12
Jones, *et al.*, 1994[140]	N/A	N/A	N/A	13 (N/A)	2, 6, 12
Sawdon, *et al.*, 1995[139]	100	20 (20)	80	57 (71)	6
McHugh, *et al.*, 1997[118]	97	17 (18)	80	31 (39)	{34.8}
Hurel, *et al.*, 1997[237]	329	N/A	N/A	223 (N/A)	6

$p = 0.001$); Jones and colleagues[140,172] obtained a value of $r^2 = 0.46$ with the Whiston Hospital questionnaire ($p < 0.00001$) and the POMS ($p = 0.01$); and Hurel and colleagues[237] compared the PQOL and the NHP and obtained a z score of 9.853 ($p = 0.0001$).

Reliability Patrick and colleagues[141] reported internal consistency for the PQOL (Cronbach's $\alpha = 0.88$). Apart from that, no evidence was presented for the reliability of the PQOL.

Responsiveness Although Chelluri and colleagues[70] assessed PQOL at more than one time point, they did not address responsiveness over time. Ridley and colleagues[264] reported no significant changes in PQOL scores between those obtained on admission to ICU and follow-up.

Outcome in critical care survivors

Hurel and colleagues[237] noted that a high health-related quality of life was perceived by 24% of patients. Patients were highly satisfied

with the help they received from family and relatives, whereas satisfaction with happiness, recreation, income and professional activity was low. Thiagarajan and colleagues[263] noted that patients reported significant decreases from their pre-critical care levels in their overall health, happiness, ability to think and to pursue leisure activities, income and employment. There was a decrease of 14% in mean PQOL score from pre- to post-trauma (although the pre-trauma value was obtained retrospectively).

NHP

The NHP was developed in the UK[265] and is based on lay perceptions of functional status and quality of life. It is intended to indicate perceptions of physical, social and emotional health problems. It was developed as a result of interviewing large numbers of lay people about the effects of illness on behaviour. The NHP was designed to measure the experience of ill health. It is concise (taking under 10 minutes to complete) and can be easily administered. It focuses on negative rather than positive experiences. Both population and individual group norms exist[265] for comparative purposes. A book outlining the development of the measure is available,[266] as is a user manual.[267]

The NHP is in two parts.

- Part I measures perceived or subjective functional status by requiring a yes or no answer to 38 statements associated with six dimensions: physical mobility, pain, sleep, energy, emotional reactions and social isolation. Each dimension has a potential score in the range 0–100 where zero indicates good health and 100 indicates poor health. More recent work has suggested that by eliminating a number of the questions and reducing the questionnaire to 24 items, it is possible to calculate a global score which could be used in quality-adjusted life-years cost–benefit analysis,[268] although usage in this form is rare.
- Part II focuses on quality of life and asks the individual about the effects of his/her functional health status on seven areas of daily life: work, looking after the home, social life, home life, sex life, interests, hobbies and holidays. Bowling[94] noted that Part II is no longer recommended by the developers, because of difficulties associated with its measurement properties.

As the developers of the NHP decided to focus only on the severe extremes of ill-health, the result is that data obtained from the NHP may be highly skewed, with a majority of respondents scoring zero or very low scores in most, if not all, of the dimensions.[269] Minor illnesses are not easily detected by the NHP and, as a result, minor improvements in health are less likely to be detected over time.

Measurement properties outside critical care

The following examples provide illustrations of the extensive testing for validity, reliability and responsiveness over time that has been conducted on the NHP. Hunt and colleagues[266] established face, content and criterion validity in physical, social and emotional dimensions during the development of the profile. Criterion validity was also established by Doll and colleagues,[270,271] using patient-reported health measures. Responsiveness has also been reported by several authors.[266,273–275] Jenkinson and colleagues[276] reported a correlation of 0.61 ($p < 0.0001$) between the General Health Questionnaire and the emotional reactions scale of the NHP for rheumatoid arthritis and migraine sufferers. Fitzpatrick and colleagues[274] also identified significant correlation coefficients between the NHP and the Arthritis Impact Measurement Scale for the dimensions of pain (0.55; $p < 0.001$), physical mobility (0.79; $p < 0.001$) and emotion (0.58; $p < 0.001$), and between the emotional scale of the NHP and the BDI (0.54; $p < 0.001$).

Hunt and colleagues[267] reported test–retest reliability coefficients ranging from 0.77 to 0.85 for Part I of the NHP, and 0.44–0.86 for Part II for osteo-arthritis patients, and 0.75–0.88 and 0.55–0.89, respectively, for patients with peripheral vascular disease. Results for Part II were not as acceptable, which explains the withdrawal of this part of the questionnaire.

Application in critical care

Eleven studies have reported on the NHP as an outcome measure in following-up critical care patients (*Table 71*).[119,139,155,156,218,237,263,277–280] One study used a simplified version of the NHP. The mean ages of the participants ranged from 23 years to 64 years and the percentages of male participants ranged from 49% to 68%. Acute severity scores were reported in nine papers. Seven papers presented data from general critical care patients, with the remainder referring to specific patient groups.

The numbers eligible for participation in these studies ranged from 78 to 7988 (*Table 72*).

TABLE 71 *Characteristics of populations studied using the NHP*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Shiell, *et al.*, 1990[277]	51	N/A	APACHE: 8.0	General
Rowan, 1992[155]	55.4 ± 0.7 (16–90)	58	APACHE II: 15.3 ± 0.3 (0–49)	General
Bell & Turpin, 1994[218]	54 (19–83)	52	APACHE II: 12 (0–29)	General
Thiagarajan, *et al.*, 1994[263]	22.5 (20–29)	N/A	APACHE: 6 (4–8)	Multiple trauma
Sawdon, *et al.*, 1995[139]	(0.3–94)	49	N/A	General
Munn, *et al.,* 1995[278]	56	N/A	APACHE II: [11.8]	General
Trouillet, *et al.*, 1996[119]	60.5 ± 12.2	62	SAPS: 9.7 ± 4 GCS: 13.1 ± 3	Cardiac surgery
Dixon, *et al.*, 1997[156]	56.2 ± 21.0	59	APACHE: 18 (0–35)	General
Gopal, *et al.,* 1997[279]	56.9 (13.4–81)	67	APACHE II: 25 (15–41)	MOF; renal failure
Hurel, *et al.*, 1997[237]	51.6 ± 17.9	56	SAPS: 11.6 ± 4.7	General
Nielsen, *et al.*, 1997[280]	64.4 (33–83)	68	N/A	Cardiac surgery; MOF

TABLE 72 *Numbers of participants in studies using the NHP*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean ± SD} [median] (range) (months)
Shiell, *et al.*, 1990[277]	200	69 (35)	131	82 (63)	6
Rowan, 1992[155]	7988	N/A	N/A	2986 (N/A)	{6.7 ± 4.4} (1.1–22)
Bell & Turpin, 1994[218]	95	32 (34)	63	60 (95)	3
Thiagarajan, *et al.*, 1994[263]	83	16 (19)	67	42 (63)	{10.4} (9.6–13.3)
Munn, *et al.*, 1995[278]	1947	612 (31)	1335	768 (58)	6
Sawdon, *et al.*, 1995[139]	100	20 (20)	80	57 (71)	6
Trouillet, *et al.*, 1996[119]	116	27 (23)	89	54 (61)	[81]
Dixon, *et al.*, 1997[156]	78	27 (36)	51	23 (45)	1
Gopal, *et al.*, 1997[279]	250	165 (66)	85	35 (41)	[33.6] (2–63.6)
Hurel, *et al.*, 1997[237]	329	N/A	N/A	223 (N/A)	6
Nielsen, *et al.*, 1997[280]	96	46 (48)	50	47 (94)	> 12

Mortality before follow-up ranged from 19% to 66%. The proportions of available participants who were followed-up ranged from 41% to 95%. Four papers reported following-up participants at 6 months and one paper reported on follow-up at several time points. The mode of administration and parts of the NHP administered (Part I and/or Part II) varied (*Table 73*).

A variety of methods were used to present the data from these studies. Seven studies presented mean values and two presented mean values along with SDs or 95% CIs. Sawdon and colleagues[139] and Dixon and colleagues[156] did not present any details relating to the findings of the NHP.

Measurement properties in critical care

Validity Three papers assessed construct validity. Two considered patient age and found no statistically significant association.[155,278] Rowan[155] and Hurel and colleagues[237] also found no association with patient gender.

TABLE 73 *Mode of administration of the NHP and presentation of results*

Study	Postal	Telephone interview	Face-to-face interview	Method of data presentation
Shiell, *et al.*, 1990[277]	Parts I and II	–	–	B
Rowan, 1992[155]	Parts I and II	–	–	C
Bell & Turpin, 1994[218]	Part I	–	–	B
Thiagaran, *et al.*, 1994[263]	–	–	Part I and II	C
Sawdon, *et al.*, 1995[139]	–	–	Not specified which part used	N/A
Munn, *et al.,* 1995[278]	Part II	–	–	B
Trouillet, *et al.*, 1996[119]	–	Part I	–	B
Dixon, *et al.*, 1997[156]	–	–	Part I and II	N/A
Gopal, *et al.*, 1997[279]	Part I and II	–	–	B
Hurel, *et al.*, 1997[237]	Part I	–	–	B
Nielson, *et al.,* 1997[280]	Part I and II	–	–	B

Criterion validity was explored in four studies. Rowan[155] used a global quality-of-life question and the HAD scale, Thiagarajan and colleagues[263] and Hurel and colleagues[237] used the PQOL, and Dixon and colleagues[156] used a satisfaction scale.

Reliability and responsiveness None of the papers assessed reliability or the responsiveness of the NHP for follow-up of critical care patients.

Outcome in critical care survivors

Hurel and colleagues[237] reported high aggregate scores which were indicative of a high quality of life in 39% of respondents. A normal social life was reported by 55% of participants; pain and limitations to physical functioning were infrequently reported. Energy, sleep and emotional reactions showed the most severe alterations.

Rowan[155] reported that survivors scored highest on the energy dimension followed by sleep. Pain and social isolation had the lowest scores. The status of survivors of ICU as measured by the NHP was found to be similar to that of two other patient groups, patients with stroke and patients with multiple sclerosis, the only difference being in the physical dimension for the stroke patients.

Munn and colleagues[278] reported that 32% of participants felt their employment was affected by ill health and 48% felt their capacity to do housework was affected; social and home life were affected in 46% and 22%, respectively. Other problems were reported in sex life (36%), hobbies (48%) and holidays (41%). Shiell and colleagues[277] noted that the main areas of difficulty for patients at follow-up were holidays (46%), social life (45%), and hobbies/interests (44%). Other problems were reported in home and sex life, both of which had 34% of patients reporting problems, and employment, for which 31% reported difficulties. Significant differences were reported between critical care survivors and a control group in the dimensions of emotional reactions ($p < 0.05$), energy ($p < 0.05$) and physical mobility ($p < 0.05$) in Part I of the NHP, and in housework ($p < 0.05$), sex life ($p < 0.01$) and hobbies ($p < 0.01$) in Part II.[280] Thiagarajan and colleagues[263] noted that Part I of the NHP suggested problems with energy (31%) and emotional reactions (21.3%), whereas physical mobility (14.9%) and social isolation (13.2%) were not so important; Part II highlighted problems with hobbies (69%) and employment (62%). The highest scores reported by Trouillet and colleagues[119] were mobility, social isolation and emotional reaction, the lowest were for sleep, energy and bodily pain. Gopal and colleagues[279] reported that the greatest difficulties were associated with mobility, limited energy, bodily pain and interrupted sleep.

SF-36

The SF-36 is a generic measure that is the product of two large-scale studies conducted in the USA: the Health Insurance Experiment and the Medical Outcomes Study. The SF-36 is a self-administered questionnaire which comprises

eight dimensions: physical functioning (ten items); social functioning (two items); role limitations due to physical problems (four items); role limitations due to emotional problems (three items); general mental health (five items); energy/vitality (four items); bodily pain (two items); general health perceptions (five items). The questionnaire takes the respondent approximately 5–10 minutes to complete. Item scores for each dimension are summated and transformed using a scoring algorithm into a scale ranging from 0% (poor health) to 100% (good health). There is an anglicised version and population norms for UK studies are also available.

Measurement properties outside critical care

Construct validity has been established by Brazier and colleagues.[281] The Medical Outcome Study researchers reported the criterion validity of each dimension of the SF-36. In the UK, criterion validity has been established by Wright and colleagues,[282] Jenkinson and colleagues,[283] and Brazier and colleagues.[281] Ware and colleagues[2] also reported correlations between the physical functioning subscale and the equivalent subscales of the SIP and the NHP (r^2 = 0.27–0.72). There is some discrepancy in the reporting of discrimination between different disease groups, as summarised by Bowling.[79] Good internal consistency and test–retest reliability were reported by Brazier and colleagues,[281] with internal consistency values ranging from 0.60 to 0.81. Jenkinson and colleagues[284] reported Cronbach's α coefficients ranging from 0.76 to 0.90. Ware and colleagues[2] reported test–retest reliability coefficients ranging from 0.43 to 0.90.

One criticism of the SF-36 is that it may be prone to ceiling effects, whereby respondents maximally affirm all items in a given scale and consequently gain the maximum scale score.[281] Consequently, if a respondent's health deteriorates further, this would not be detected by the instrument. Ceiling effects have been found to be most common on the role emotional and role physical dimensions. A modified version of the instrument (SF-36, version II) has been developed which appears to overcome this problem,[285] but the SF-36 version II has not been widely used to date.

Application in critical care

Nine studies, one of which was a personal communication from Eddleston and colleagues (1999), were identified in which the SF-36 had been used as an outcome measure in critical care patients (*Table 74*).[91,235,236,238,239,286–288] It should be noted that the study conducted by Chrispin and colleagues[286] has been included for its value in assessing the measurement properties of the SF-36 within critical care, even though it was administered prior to discharge and thus is not a follow-up measure.

TABLE 74 *Characteristics of populations studied using the SF-36*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD (range)	Type of patient
Broome, *et al.*, 1996[235]	Group 1: 50 Group 2: 53 Group 3: 50	73 69 55	APACHE: 9.0 N/A N/A	Pancreatitis
Chrispin, *et al.*, 1997[286]	61.9 (59–64.2)	68	APACHE II: 10 (8–11)	General
Brenneman, *et al.*, 1997[236]	37	65	ISS: 25	Trauma
Weinert, *et al.*, 1997[91]	40 ± 12	67	Lung injury score: 2.4 ± (1.25–3.25)	Acute lung injury
Ridley, *et al.*, 1997[287]	N/A	N/A	N/A	General
Kriwanek, *et al.*, 1998[239*]	53 (23–79)	65	APACHE II: (6–33)	Pancreatic surgery
Schelling, *et al.*, 1998[238]	(18–85)	51	Lung injury score: (2.75–3.75)	ARDS
Eddleston, *et al.*, 1999[†]	49 ± 11.6	57	APACHE II: 18.8 ± 6.2	General
Davidson, *et al.*, 1999[288]	[40.6] (15–81)	N/A	APACHE III: [66] (27–109) ISS: 24 (4–43)	ARDS

[*] *Kriwanek and colleagues only administered questions 1–19, 21–22, 25, 28 and 35*
[†] *Personal communication*

The mean ages of the participants ranged from 37 years to 62 years and the percentages of participating males ranged from 55% to 73%. Acute severity scores were reported in eight papers. Eddleston and colleagues (personal communication, 1999), Ridley and colleagues,[287] and Chrispin and colleagues[286] presented data from the follow-up of general critical care patients; the remaining papers referred to more specific groups of patients within the critical care population.

The numbers eligible for participation in the studies ranged from 40 to 439 (*Table 75*). Mortality before follow-up ranged from 0% to 51%. The proportions of available participants who were followed-up ranged from 45% to 100%.

In four published studies[235,236,238,288] and in the study by Eddleston and colleagues (personal communication, 1999), the SF-36 was administered by telephone interview; the remaining four studies used either mailed questionnaires[91,287] or face-to-face interviews.[239,286] Five studies presented data as mean values, four of which also presented SDs or 95% CIs.

Measurement properties in critical care

Validity Ridley and colleagues[287] reported on the construct validity of the SF-36 when used in critical care using the variable of pre-morbid employment status ($p < 0.001$). Davidson and colleagues[288] used a linear regression model reporting no significant effects of age, acute severity or co-morbid disease. Weinert and colleagues[91] reported a significant correlation between age and the physical component of the SF-36 ($r^2 = 0.14$). Chrispin and colleagues[286] used a general linear model to identify a significant difference in score distribution due to age ($F = 6.3$; $p < 0.001$) and gender ($F = 9.2$; $p < 0.001$).

Criterion validity was explored by Weinert and colleagues,[91] who reported significant correlations between the Karnofsky Index and the physical component ($r^2 = 0.56$) and the mental component ($r^2 = 0.37$) of the SF-36.

Reliability Three papers assessed the reliability of the SF-36 in the context of critical care. Chrispin and colleagues[286] reported Cronbach's α values of 0.87 for social function, 0.77 for mental health and 0.93 for physical functioning. Schelling and colleagues[238] reported values of Cronbach's α of 0.70 for vitality to 0.98 for emotional and role function. Most multi-item scales had Cronbach's α values ranging between 0.93 and 0.98, indicating good reliability. Internal reliability (Cronbach's α) ranging from 0.85 to 0.93 was reported by Weinert and colleagues[91] for the subscales of the SF-36.

Responsiveness Ridley and colleagues[287] obtained data at two time points and reported differences in specific dimensions of the SF-36. They presented

TABLE 75 *Numbers of participants in studies using the SF-36*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} [median] (range) (months)	Method of data presentation
Broome, *et al.*, 1996[235]	Group 1: 40 Group 2: 89 Group 3: 47	8 (20) N/A N/A	32 N/A N/A	22 (69) N/A N/A	{51} {39.5} {28.3}	B
Chrispin, *et al.*, 1997[286]	336	170 (50)	166	166 (100)	Prior to discharge from critical care	C
Brenneman, 1997[236]	439	2 (4)	437	195 (45)	12	C
Ridley, *et al.*, 1997[287]	166	29 (18)	137	95 (69)	6	C
Weinert, *et al.*, 1997[91]	69	35 (51)	34	24 (71)	{19} [15] (6–41)	C
Schelling, *et al.*, 1998[238]	192	90 (47)	102	80 (78)	[48]	C
Kriwanek, *et al.*, 1998[239]	147	55 (37)	92	92 (100)	24	A
Eddleston, *et al.*, 1999*	370	144 (39)	226	143 (63)	3	C
Davidson, *et al.*, 1999[288]	102	0 (0)	102	77 (75)	[23]	C

*Personal communication

data for the mean change over time from ICU discharge to 6-month follow-up with 95% CIs. Significant increases were reported in mental health, vitality, social functioning and reduction in bodily pain scores, although no values were provided.

Rosser's disability and distress categories

Rosser's disability and distress categories measure the degree of disability and also the distress experienced. It is composed of four dimensions: general mobility, usual activity, self-care, social and personal relationships.[289] With the exception of the general mobility dimension, which requires a statement to be selected, all the other items require a yes/no response. General mobility and usual activity are coded 1–6 and 1–4, respectively. For the remainder of the dimensions, a positive response is allocated a score of 1 (the range of scores is therefore 0–4 for each category). The maximum score is 18. An allocation method then enables the scores to be categorised between I and VII, where I is the least and VII the most disabled.

Measurement properties outside critical care

Gater and colleagues[290] have reported a correlation of 0.55 between Rosser's disability categories and the total NHP score. Kind and Gudex[291] reported similar findings.

Application in critical care

The mean ages of patients ranged from 23 years to 53 years (*Table 76*).[67,263,264,277,292] Only Shiell and colleagues[277] presented data on the percentages of male participants. Three papers presented acute severity scores. The populations were general critical care patients in all but one paper.

The numbers of subjects eligible for participation ranged from 83 to 385 (*Table 77*). Mortality before follow-up ranged from 19% to 40%. The proportions of available participants followed-up ranged from 61% to 88%. The time of follow-up varied between studies.

Postal methods of administration were used in all studies. In addition, Thiagarajan and colleagues[263] administered the questionnaire during face-to-face interviews, if necessary, and Kerridge and colleagues[293] administered the measure by telephone interview if required.

TABLE 76 *Characteristics of populations studied using Rosser's disability categories*

Study	Mean age (range) (years)	Male (%)	Severity score, mean [median] (range)	Type of patient
Shiell, *et al.*, 1990[277]	51	60	APACHE: [8.0]	General
Ridley & Wallace, 1990[67]	N/A	N/A	N/A	General
Thiagarajan, *et al.*, 1994[263]	22.5 (20–29)	N/A	APACHE: 6 (4–8)	Multiple trauma
Ridley, *et al.*, 1994[264]	53 (95% CI, 47 to 59)	N/A	APACHE II: 11.8 (95% CI, 9.4 to 14.2)	General
Kerridge, *et al.*, 1995[292]	N/A	N/A	N/A	General

TABLE 77 *Number of participants in studies using Rosser's disability categories*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time {mean} (range) (months)	Method of data presentation
Shiell, *et al.*, 1990[277]	200	69 (35)	131	82 (63)	6	A
Ridley & Wallace, 1990[67]	385	129 (34)	256	156 (61)	(12–36)	A
Thiagarajan, *et al.*, 1994[263]	83	16 (19)	67	42 (63)	{10.4} (9.6–13.3)	A
Ridley, *et al.*, 1994[264]	90	33 (37)	57	41 (72)	12	A
Kerridge, *et al.*, 1995[292]	248	98 (40)	150	Distress: 122 (81) Disability: 132 (88)	36	B

Some authors presented the data as percentages of patients, others as the numbers of patients in each category. Others referred to those who had shown an improvement or deterioration in scores. Only Kerridge and colleagues[293] reported mean scores for Rosser's disability and distress categories.

Measurement properties in critical care

Validity and reliability None of the papers attempted to assess the validity or reliability of Rosser's disability categories in the follow-up of critical care survivors.

Responsiveness Ridley and colleagues[264] found no change in responses in 61% of participants in their study between critical care admission and follow-up.

Spitzer's quality-of-life index and uniscale

This index was developed by Spitzer and colleagues[294] for use by clinicians in relation to chronically ill patients and those suffering from cancer. The components of quality of life were derived from a series of three panels of 43 individuals comprising cancer patients and their relatives, patients with chronic diseases and their relatives, healthy individuals aged 20–59 years and 60+ years, physicians, nurses, social workers, other health professionals and members of the clergy. Those factors rated as highly important were then compiled into the first draft of the quality of life index, which was then tested on outpatients. The final version of the quality of life index comprised the following dimensions: activity, performance of activities of daily living, perception of health, support from family and friends, and outlook on life.

Respondents select items in terms of the applicability of the statements to them. The index also includes a visual analogue scale, on which the respondent and the interviewer are asked to mark an 'X' on a line that rates quality of life from 'lowest ... to highest quality of life'. There is a maximum score of ten. The scale can be summed to give an overall single score or each item can be presented separately; it is then referred to as the uniscale. The test is very brief, taking on average 1 minute to complete. Spitzer and colleagues[294] caution against using the index as a means of assessing quality of life in a 'healthy' population.

Measurement properties outside critical care

Spitzer and colleagues[294] established content validity through a literature review and by asking lay people and physicians to assess the scope of the instrument. Mor and colleagues[80] reported a correlation between Spitzer's quality-of-life index and the Karnofsky Index ($r^2 = 0.39$). The item correlations with the Karnofsky Index ranged from $r^2 = 0.02$–0.32. Gough and colleagues[295] reported a correlation between the overall single score and the Karnofsky Index of $r^2 = 0.36$. The developers of the index claim that it can distinguish between healthy individuals and those with varying degrees of illness, thus justifying their claim of discriminant validity. In contrast, Slevin and colleagues[296] stated that the index was not sensitive enough to discriminate between different stages of treatment in women with breast cancer.

In terms of reliability of the measure, Spitzer and colleagues[294] reported a coefficient of 0.77 for internal consistency and an inter-rater reliability coefficient of 0.81. Another study[296] did not report such a high degree of reproducibility. Ratings between patients and physicians were positively correlated ($r^2 = 0.37$), as were ratings between social workers and patients ($\tau = 0.72$) according to Spitzer and colleagues,[294] and Gough and colleagues.[293] Test–retest reliability of 0.81 ($p < 0.001$) has been reported by Churchill and colleagues.[297]

Application in critical care

Four papers have reported using Spitzer's quality-of-life index[110,262,298,299] and a further four papers have reported using the uniscale[92,207,225,227] (see *Table 78*). Six papers reported that the mean ages of participants ranged from 44 years to 75 years. The percentages of male participants ranged from 50% to 80% in the seven studies in which it was reported. Four papers reported acute severity scores. Six papers reported on general critical care populations, the other two being based on surgical or cardiac transplant populations.

The numbers eligible for participation in these studies ranged from 100 to 1308 (*Table 79*). Mortality before follow-up ranged from 11% to 70%. The proportion of individuals available for follow-up ranged from 51% to 100%. In two papers,[110,299] data were obtained for baseline quality of life either on admission to, or at discharge from critical care. Two papers reported outcome at 6 months[262,299] and two at 12 months.[298,299] Five papers clearly reported

TABLE 78 *Characteristics of populations studied using Spitzer's quality-of-life index or uniscale*

Study	Mean age ± SD [median] (range) (years)	Male (%)	Severity score, mean ± SD [median] (range)	Type of patient
Sage, *et al.*, 1986[225*]	58.4	50	APACHE II: 13.3	General
Slayter, *et al.*, 1986[110]	43.7 (0.25–82)	58	TISS: 88 ± 129.9	General
Sage, *et al.*, 1987[227*]	74.7 ± 0.7 (65–96)	50	N/A	General
Frede & Lanter, 1990[298]	N/A	52	N/A	Surgical
Schuster, 1991[107*]	N/A	N/A	N/A	General
Konopad, *et al.*, 1995[299]	55 ± 20 [58]	54	APS: 12 ± 6 APACHE II: 16 ± 7	General
Grady, *et al.*, 1998[262] (modified)	53 ± 9	80	N/A	Heart transplant
Kocher & de Torrenté, 1998[92*]	67	56	APACHE II: [9.0] (0–52)	General

[*] *Spitzer's uniscale only*

TABLE 79 *Numbers of participants in studies using Spitzer's quality-of-life index and uniscale*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (range) (months)	Method of data presentation
Sage, *et al.*, 1986[225]	341	83 (25)	258	140 (54)	(15–20)	C
Slayter, *et al.*, 1986[110]	100	11 (11)	89	89 (100)	1	A
Sage, *et al.*, 1987[227]	134	37 (28)	97	59 (61)	18	C
Frede & Lanter, 1990[298]	164	80 (49)	84	74 (88)	12, 24	A
Schuster, 1991[107]	1308	N/A	N/A	N/A	(12–60)	C
Konopad, *et al.*, 1995[299]	504	126 (25)	378	293 (78)	6, 12	A
Grady, *et al.*, 1998[262]	269	N/A	N/A	219 (N/A)	6	A
Kocher & de Torrenté, 1998[92]	292	203 (70)	89	45 (51)	(12–74)	A

that the mode of administration was by mail.[92,107,225,262,299] Konopad and colleagues[299] administered the questionnaire using face-to-face or telephone interviews, as appropriate. Sage and colleagues[225,227] administered questionnaires by telephone interviews. Slayter and colleagues[110] provided no information as to the mode of administration. A variety of methods were used to present the data in these studies. Three papers reported data as mean values with SDs.

Measurement properties in critical care

Validity Konopad and colleagues[299] examined construct validity through the variable of age, although this was not explicitly examined in the text of the paper. None of the remaining papers that reported using Spitzer's quality-of-life index attempted to assess the validity of the measure.

Reliability None of the papers reported assessing reliability.

Responsiveness Konopad and colleagues[299] reported a significant decrease in levels of activity between baseline and follow-up at 12 months ($p < 0.001$), and yet perceived health was reported to have improved at follow-up ($p < 0.05$). None of the remaining three papers made reference to assessment of responsiveness. Slayter and colleagues[110] assessed pre-admission quality of life with respect to 1 month before

admission (retrospectively obtained at follow-up) and also reviewed quality of life 1 month after discharge ($\chi^2 = 101.62$; $p < 0.0005$). There was no evidence for the assessment of the responsiveness of the uniscale in critical care patients.

PGWB

The PGWB schedule[300,301] was designed to measure feelings of well-being and distress rather than a broader concept of quality of life. It is often, however, used as a proxy measure of quality of life.[79] It consists of 18 or 22 questions and utilises a 6-point response scale for intensity or frequency (for 14 items) and a 0–10 rating scale defined by adjectives at each end (for four items). Total scores of 0–60 reflect severe psychological distress, 61–72 represent moderate psychological distress, and 73–110 represent positive psychological well-being. The domains that are examined include: anxiety, depression, positive well-being, self-control, general health and vitality. Respondents are asked to reflect on their feelings over the last month. The questionnaire is self-administered and takes approximately 10–12 minutes to complete.

Measurement properties outside critical care

Moderate to strong correlations ($r^2 = 0.22$–0.81) have been reported with interviewers' ratings of depression.[302] Factor analysis has identified three factors that account for 51% of the variation. Dupuy[301] reported these as: anxiety, tension and depression; health and energy; and positive well-being or life satisfaction. Becker and colleagues[303] reported criterion validity between the PGWB and the SF-36 and HAD scale. They reported correlations for depression, anxiety, mental health and vitality with the subscales of the SF-36 and HAD, which ranged from 0.69 to 0.79 ($p < 0.0001$). Test–retest reliability coefficients have ranged between 0.50 and 0.86.[304,305] An internal reliability coefficient of 0.93 was also reported by Monk.[304] Becker and colleagues[303] also reported on test–retest reliability, which they reported as $r^2 = 0.27$–0.44.

Application in critical care

The mean ages of participants ranged from 61 years to 69 years (*Table 80*) and the percentage of male participants was about 52%. Acute severity scores were presented in all three papers.[141,248,258] Douglas and colleagues[248] reported on the outcome of general critical care patients, the two remaining papers followed-up medical or respiratory patients. The numbers eligible for participation ranged from 58 to 193 (*Table 81*). Mortality before follow-up ranged from 43% to 50%. The proportions of available participants followed-up ranged from 21% to 77%, although the former included only six patients.

Both Danis and colleagues[258] and Patrick and colleagues[141] administered the PGWB using face-to-face interviews. Douglas and colleagues[248] did not specify the mode of administration. Mean PGWB scores were presented in all

TABLE 80 *Characteristics of populations studied using the PGWB index*

Study	Mean age ± SD (years)	Male (%)	Severity score, mean ± SD	Type of patient
Danis, *et al.*, 1988[258]	68.9 ± 8.4	52	APACHE II: 13.0 ± 7	Medical; respiratory
Patrick, *et al.*, 1988[141]	69 ± 8.0	52	APACHE II: 13.0 ± 6	Medical; respiratory
Douglas, *et al.*, 1997[248]	61.4 ± 19.9	53	APACHE III: 66.5 ± 25.3	General

TABLE 81 *Numbers of participants in studies using the PGWB index*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time [median] (months)
Danis, *et al.*, 1988[258]	193	83 (43)	110	69 (63)	N/A
Patrick, *et al.*, 1988[141]	160	70 (44)	90	69 (77)	[19]
Douglas, *et al.*, 1997[248]	58	29 (50)	29	6 (21)	6

three papers. In addition, SDs were presented by Danis and colleagues[258] and Patrick and colleagues.[141]

Measurement properties in critical care

Validity There was no testing of construct validity but Patrick and colleagues[141] examined criterion validity; they reported a modest correlation ($r^2 = 0.29$; $p = 0.001$) between the PGWB and the PQOL.

Reliability and responsiveness Assessment of the reliability and responsiveness of the PGWB was not undertaken in any of the three studies.

Fernandez's questionnaire

Fernandez and colleagues[210] developed a questionnaire which comprised 15 items grouped into three subscales evaluating physiological activities, normal daily activities and emotional state. The authors acknowledged that the ADL index[62] was an inspiration for the development of some of the items in their own scale. Scores ranged from zero to 29 with a score of zero signifying normality and increasing scores signifying decreased quality of life. The questionnaire only took 5–10 minutes to complete and could be completed by a close family member if a patient was too ill to respond. Face validity was established by Fernandez and colleagues[210] via a committee of experts who had used a previous version of the questionnaire. They selected aspects they wished to explore and the facets of quality of life that would best explore those aspects.

Measurement properties outside critical care

This questionnaire does not appear to have been used outside of critical care settings.

Application in critical care

Three papers reported on the use of this questionnaire (*Table 82*).[210,212,306] Only one study reported the age and gender of the patients.[306] Two studies were based on general patients, the third on trauma patients. All three studies were large (351–716 patients) with high rates of follow-up at 6 or 12 months after discharge (*Table 83*). Mean values with SDs were presented in all three papers.

Measurement properties in critical care

Validity Two studies[212,306] reported on construct validity. Both found a moderate correlation with age, APACHE II scores and, in one of the studies, with the ISS. There is some evidence of criterion validity from Fernandez and colleagues,[210] who reported a significant correlation with the GOS.

Quality of life as measured by this scale and the GOS decreased 6 months after discharge

TABLE 82 *Characteristics of populations studied using the Fernandez or the Whiston Hospital questionnaires*

Study	Mean age ± SD (range) (years)	Male (%)	Severity score, mean ± SD [median] (range)	Type of patient
Vazquez-Mata, *et al.*, 1992[212*]	N/A	N/A	APACHE II: 14.65 ± 0.25	General
Fernandez, *et al.*, 1996[210*]	N/A	N/A	N/A	General
Vazquez-Mata, *et al.*, 1996[306*]	31.2 ± 0.86	78	APACHE II: 13.5 ± 0.4 ISS: 23.6 ± 0.6	Trauma
Jones, *et al.*, 1993[172†]	N/A	58	APACHE II: [12] (2–21)	General
Jones, *et al.*, 1994[140†]	N/A	N/A	N/A	General
Weir & Waldmann, 1994[128†]	N/A	N/A	N/A	General
Sawdon, *et al.*, 1995[139†]	(0.3–94)	49 4 children	N/A	General
Griffiths, *et al.*, 1997[307†]	Study group: 59 (22–89) Control group: 64.5 (22–75)	N/A	APACHE II: study group, 17 (11–34) control group, 13 (11–31)	General

**Fernandez's questionnaire*
†Whiston Hospital questionnaire

TABLE 83 *Numbers of participants in studies using the Fernandez or the Whiston Hospital questionnaires*

Study	Number eligible for study	Mortality (%) before follow-up	Number available for follow-up	Number (%) followed-up	Follow-up time (months)
Vazquez-Mata, *et al.*, 1992[212]*	716	213 (30)	503	422 (84)	12
Fernandez, *et al.*, 1996[210]*	578	0 (0)	578	578 (100)	6
Vazquez-Mata, *et al.*, 1996[306]*	351	0 (0)	351	351 (100)	12, 24
Jones, *et al.*, 1993[172]†	85	12 (14)	73	49 (67)	6
Jones, et al., 1994[140]†	N/A	N/A	N/A	N/A	2, 6, 12
Weir & Waldmann, 1994[128]†	N/A	N/A	N/A	N/A	2, 6, 12
Sawdon, *et al.*, 1995[139]†	100	20 (20)	80	57 (71)	6
Griffiths, *et al.*, 1997[307]†	84	46 (55)	38	38 (100)	6

**Fernandez's questionnaire*
†Whiston Hospital questionnaire

from the ICU. The authors[210] compared the magnitude of the differences between the two scales reporting a weighted kappa index of 0.56 ($p < 0.001$).

Reliability Fernandez and colleagues[210] reported internal consistency for the global quality-of-life scale (Cronbach's $\alpha = 0.85$) and for the physiological activities (Cronbach's $\alpha = 0.66$), physical capacities (Cronbach's $\alpha = 0.81$) and emotion (Cronbach's $\alpha = 0.82$) subscales. The α coefficient ranged between 0.824 and 0.852 when each item in turn was removed from the scale. Vazquez Mata and colleagues[212] reported the internal consistency as $\alpha = 0.67$. They also reported on inter-observer reliability (> 0.9 for global scale and physical capacities subscale, emotional subscale 0.77, basic physiological activities subscale 0.61) and intra-observer reliability (> 0.9 for global scale and physical capacities subscale, emotion subscale 0.84, basic physiological activities subscale 0.93).

Fernandez and colleagues[210] also examined the reliability between the responses of patients and those of close relatives; they reported correlation coefficients of > 0.9 for global scale and physical capacities subscale, and 0.82 and 0.76 for the emotion and basic physiological activities subscales, respectively. Correlation coefficients of > 0.95 for global scale and physical capacities subscale, and 0.81 and 0.86 for the emotion and basic physiological activities subscales, respectively, were reported for the reproducibility between direct and telephone interviews. Inter-observer reliability was also reported as being 92% by Vazquez Mata and colleagues,[212] patient–relative reliability was reported as 85%, and doctor–patient reliability as 80%.

Responsiveness Vasquez Mata and colleagues[306] reported significant correlations between the quality-of-life scores at 1 and 2 years ($r^2 = 0.908$; $p < 0.0001$). There was a weak correlation between baseline and 1 year ($r^2 = 0.167$; $p < 0.0001$), and admission and 2 years ($r^2 = 0.198$; $p < 0.0001$). Vasquez Mata and colleagues[212] reported a deterioration in quality of life between the first and second administration of the questionnaire ($p < 0.01$).

Whiston Hospital questionnaire

The Whiston Hospital questionnaire has been used in several studies and appears to have first been reported by Jones and colleagues.[172] It comprises a pre-morbid health questionnaire and a reworded follow-up questionnaire, which is loosely based on the FLP. The information collected includes previous/current health, mobility, work, leisure, and contact with friends or relatives. The questionnaire is scored on a scale of zero for good health and 1, 2, or 3 for increasingly poor health, with a maximum possible total score of 25.

Measurement properties outside critical care

The Whiston Hospital questionnaire was designed to follow-up critical care patients and thus there is no evidence of its measurement properties outside this area.

Application in critical care

Five papers reported on the use of this questionnaire (*Table 82*).[128,139,140,172,307] Only two papers reported on patients' ages and two on gender mix. All five studies included general critical care patients. Three studies were relatively small in size (with 84, 85 and 100 patients eligible for study); the sizes of the other two studies are unknown (*Table 83*).

Griffiths and colleagues[307] presented data for mean values with SDs for both the control and experimental group in their study. Jones and colleagues[172] presented data as a scatter plot and Weir and Waldmann[128] presented only qualitative data.

Measurement properties in critical care

Validity Two studies[140,172] provided some information on construct and criterion validity (*Table 84*). There was no significant correlation between the score and the APACHE II score on admission. Reasonably high correlations with the FLP and PQOL suggest criterion validity exists (although the former association is not surprising, given the origins of the questionnaire) and another study found no significant association with POMS.

Reliability and responsiveness There was no assessment of responsiveness in those papers which reported using the Whiston Hospital questionnaire.

Summary

- For the nine quality-of-life measures, data presentation in terms of mean values and SDs was good overall, with 39 papers presenting mean values and 31 presenting SDs. Data that supported evidence for validity was available in 19 papers. However, evidence to support reliability and responsiveness over time is much more restricted, being reported in only five papers (this is mirrored in the data available for the measurement properties of the quality-of-life measures outside critical care, where validity is well documented but evidence for reliability is poor).
- The SIP has been widely tested for reliability and validity in a variety of populations, including critical care patients. Although not explicit in their intentions, some papers did provide evidence of internal consistency, inter-rater reliability, construct and concurrent validity in the critical care populations examined. In addition, there does appear to be consistency in the results from various reports on different population groups within intensive care. However, the questionnaire is long and can take 20–30 minutes to complete. Consequently, this means that shorter measures may be more practicable in many studies.
- Criterion validity of the PQOL appears to have been substantiated in a number of studies using the NHP and the SIP. Reliability was reported only once and was high. The PQOL is a simple and easy-to-administer questionnaire, which may have some role to play in assessing the quality of life of critical care survivors. Further work is required to substantiate the reliability of the measure. The PQOL appears to be an appropriate measure for use with ICU patients.
- The NHP can be completed in a short time. The instrument was designed to assess the severe end of ill-health and, consequently, it manifests

TABLE 84 *Assessment of validity of the Fernandez and the Whiston Hospital questionnaires*

Study	Construct validity	Criterion validity
Vazquez-Mata, *et al.*, 1992[212*]	Age: (r = 0.39) APACHE II: (r = 0.12)	Quality of life (good, normal, bad)
Fernandez, *et al.*, 1996[210*]	N/A	GOS: $p < 0.0001$ for global scale and all subscales of questionnaire
Vazquez-Mata, *et al.*, 1996[306*]	Age: (r^2 = 0.289) ISS: (r^2 = 0.388) APACHE II: (r^2 = 0.327)	GOS: no value given
Jones, *et al.*, 1993[172†]	APACHE II: (r = –0.19; p = 0.174)	FLP: (r = 0.7; $p < 0.0001$) PQOL (r = –0.678; $p < 0.00001$)
Jones, *et al.*, 1994[140†]	N/A	POMS questionnaire (NS)

**Fernandez's questionnaire*
†Whiston Hospital questionnaire

floor effects in some patient groups, which may mean that small changes over time cannot be detected. The validity, reliability and responsiveness of the NHP have been established in a number of different patient populations. Criterion validity has been established in critical care patients, although no evidence of reliability or responsiveness is reported in any of the papers assessed.

- The measurement properties of the SF-36 are well established outside critical care. The evidence available suggests that the instrument may be appropriate for critical care patients. It appears to be acceptable to patients and to have good reliability and validity, although its responsiveness is unclear.
- There has been a lack of evaluation of the measurement properties of Spitzer's quality-of-life index as an outcome measure in critical care. More extensive research is required in order for this measure to be recommended for use in following-up critical care patients. The uniscale is a simple and easy-to-administer measure of quality of life. However, there is a lack of evidence about its measurement properties in critical care.
- The Spanish quality-of-life questionnaire reported by Vazquez Mata and colleagues[212,306] and Fernandez and colleagues[210] has undergone extensive testing for reliability and validity. As yet it has not been tested in populations in the UK and may warrant further investigation.
- The Whiston Hospital questionnaire is largely adapted from the FLP and has been reworded to allow for pre-admission quality of life to be assessed. As there is more evidence available for the psychometric properties of the SIP and FLP, it is suggested that they are used in preference to the Whiston Hospital questionnaire.
- Far more research into the measurement properties of health status instruments used in this patient group is needed. To date, selection of measures has been *ad hoc* and unsupported by evidence. This can lead potentially to inappropriate measures being used and inaccurate and misleading data being collected.
- It is difficult to summarise the health-related quality of life of critical care survivors. The results that have been reported vary considerably, depending on case mix and the length of follow-up. In addition, relating the findings with one measure to those with another is problematic.

Chapter 8

Conclusions and recommendations for research

This review had four objectives, each of which is considered in turn;

- 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 survivors
- to review the validity, reliability and responsiveness of the measures in critical care survivors
- to consider the implications for future policy and to make recommendations for further methodological research
- to review what is currently known about the outcome of critical care survivors.

Measures used in critical care

- Measures of **impairment** have largely been confined to the respiratory system so are almost certainly not appropriate for many critical care survivors. These measures can be categorised as respiratory volumes (e.g. vital capacity), gas flow within the respiratory system (e.g. FEV_1), pulmonary diffusing capacity (e.g. D_LCO) and visualisation of the upper airway (e.g. bronchoscopy). Often, multiple tests are performed.
- Eight measures of **physical functional status** have been used, five of them generic and three disease-specific (NYHA functional class, ATS respiratory questionnaire, walk test). The generic measures most frequently used have been multi-item scales (Katz's ADL scale, Karnofsky Index, Barthel Index). Two single-item global measures have attempted to capture a person's overall activity level or functional status.
- Five multi-item measures of **mental functional status** have been employed, four of them generic and one specifically for trauma patients (IES). The generic measures are either confined to assessing depressive symptoms (CES–D, BDI) or also encompass a measure of anxiety (POMS, HAD).
- Measures of **neuropsychological functioning** are concerned with a person's cognition, attention, ability to process information and memory. Apart from one single-item measure that focuses on communication level, there have been six multi-item measures used with critical care survivors (Trailmaking Tests, WCST, Wechsler Memory Scale, Benton's Test for Visual Retention, PASAT, MMSE). These measures are particularly appropriate for use in survivors of head injury or other neurological insult. In that sense, they are disease-specific rather than generic measures.
- Single item measures of **recovery** have frequently been used but researchers have often invented their own so there is little consistency in the wording. These measures have had five principal foci – return to work, return to own home, degree of recovery, productivity and chronic health status (CHE). One multi-item scale, the GOS, has also been used.
- Nine measures of **health-related quality of life** have been used – although some of these multi-item generic measures encompass functional status also. The three most extensively employed have been the SIP/FLP, PQOL and NHP. In addition, in recent years, the SF-36 has increasingly been used. Other less commonly used measures are Rosser's disability and distress categories, Spitzer's quality-of-life index and uniscale, the PGWB index, Fernandez's questionnaire and the Whiston Hospital questionnaire.

Measurement properties

- Overall, few attempts have been made to determine the properties of any of the measures when used with the survivors of critical care. In addition, in many instances, there is 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 is little evidence as to the properties of the **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 is some evidence as to the criterion validity of the most commonly used measure of respiratory impairment, FEV_1, in that it correlated with measures of health-related quality of life.

- There is some evidence as to the validity and responsiveness of two generic measures of **physical functional status**, Katz's ADL and the Karnofsky Index. However, the reliability of these measures is unknown. Even less is known about the disease-specific measures, though there is some evidence as to the construct validity of the ATS respiratory disease questionnaire and the responsiveness of the NYHA functional classification.
- Similarly, there is only limited information about the properties of the **mental functional status** measures. There is some evidence as to the criterion validity of all the generic instruments and the responsiveness of the CES–D.
- The only support for the **neuropsychological functional status** measures is some weak evidence as to the criterion validity of the Trailmaking Tests and the WCST.
- Assessment of the properties of measures of **recovery** has been restricted to validity. Both the GOS and return to work appear to have some construct and criterion validity. No reports of reliability or responsiveness have been published.
- There is evidence as to the validity of **health-related quality-of-life** measures but rather little as to their reliability or responsiveness in critical care survivors. This mirrors the state of affairs as regards assessment of measurement properties outside critical care. The validity of the SIP, PQOL and NHP in critical care appears to be reasonable but there is inadequate information on the SF-36, Spitzer's quality-of-life index and the other, less well-known, generic measures.

Implications for policy and recommendations for research

- 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 state of affairs partly reflects the large number of measures that have been used in critical care research in the past. The first recommendation, therefore, is that the research community should 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 these proposals, it is suggested that future researchers confine their selection to the following measures, until such time as clearer scientific evidence can distinguish between their relative merits. A crude summary of their measurement properties is provided in *Table 85*.
- Measures of **impairment** appear to have been of limited value except, perhaps, in those patients with respiratory disease. Their use in studies of 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 sub-groups (NYHA functional class in cardiac patients and the ATS respiratory disease questionnaire in respiratory patients).
- **Mental functional status** is probably best assessed using the POMS or HAD scale, as these cover anxiety in addition to depressive symptoms. In patients who are recovering from trauma, the IES 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 GOS is the only multi-item scale available. In addition, standardisation of two single item measures – one on return to work and one on residency or return to own home – would help to establish their usefulness.
- **Health-related quality of life** offers a greater range of possibilities than most of the other categories of outcome. The three principal contenders, in that they have been used most frequently in critical care research, are the SIP/FLP, PQOL and NHP. To these, it is suggested, the SF-36 should be added as this measure is being used increasingly often and widely in healthcare research, and its measurement properties in other areas have been demonstrated.
- As has already been stated, there is an urgent need for rigorous assessment of the measurement properties of all instruments being used

TABLE 85 *Extent of knowledge as to the measurement properties of prioritised outcome measures*

	Reliability	Construct validity	Criterion validity	Responsiveness
Physical functional status				
Katz's ADL	NK	±	+	+
Karnofsky's index	NK	+	+	+
NYHA functional class	NK	NK	NK	+
ATS respiratory disease questionnaire	NK	+	+	NK
Mental functional status				
POMS	NK	NK	±	NK
HAD scale	NK	NK	+	NK
IES scale	NK	NK	±	NK
Neurophysiological function				
Trailmaking tests	NK	NK	+	NK
WCST	NK	NK	+	NK
Measures of recovery				
GOS	NK	+	+	NK
Return to work	NK	+	+	NK
Residency/return home	NK	NK	NK	NK
Health-related quality of life				
SIP/FLP	+	+	+	+
PQOL scale	+	NK	+	NK
NHP	NK	NK	NK	NK
SF-36	+	+	+	NK

+, some evidence; ±, inconsistent evidence; NK, not known

in critical care research. This work should be focussed initially on the leading measures outlined above. All studies that assess the outcome of critical care by means of one of these measures should also seek to explore at least one methodological characteristic of the measure used (e.g. intra-rater reliability, construct validity). This approach would be more cost-effective than funding studies that are purely methodological in intent. Researchers should consider comparing two or more equivalent measures in head-to-head comparisons.

Health of critical care survivors

- Given all the concerns expressed above about limitations to the scientific worthiness of the outcome measures that have been used in critical care research, it is impossible to arrive at a valid and reliable overview as to the health of survivors. As will be apparent from the literature, huge differences in outcome exist between studies. This is not surprising given the variety of patients included, the failure to follow-up all survivors, differences in the time of follow-up, the lack of independent assessors, and the often poor presentation of the data. Such criticisms should not be seen as unique to this area of healthcare research.
- Thus our comments are confined (albeit that they are tentative) to a few broad observations:
  - physical functional status appears to be 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 persist in about half the survivors
  - over 70% of survivors of working age return to work, although their work activity may have altered
  - the majority return to their own homes within a few months
  - the areas of quality of life that are most frequently diminished are those relating to work, recreation and sleep.
- In conclusion, this review highlights the limitations of research on the outcomes of critical care. In order for evaluations of healthcare interventions to have any value, it is essential that outcome measures produce accurate and meaningful data. The evidence provided here suggests that limited consideration has been given to the choice of outcome measures and interpretation of subsequent

results. Further investigations are required to provide a greater understanding of outcome measurement in this area. Without a greater knowledge of the operating characteristics of measures used in this field, results will remain ambiguous and difficult to interpret, and consequently of limited value to policy-makers, clinicians and patients.

Acknowledgements

This study was commissioned by the NHS HTA programme.

The authors would like to thank Keryn Vella for her administrative and secretarial help and support.

The views expressed in this report are those of the authors, who are also responsible for any errors.

The authors are indebted to the referees for their perseverance in reading the report and the quality of their comments.

References

(excluding conference abstracts)

1. Geigle R, Jones SB. Outcomes measurement: a report from the front. *Inquiry* 1990;**27**:7–13.

2. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston, MA: Health Institute, New England Medical Center; 1993.

3. Silver GA. Paul Anthony Lembke: a pioneer in medical care evaluation. *Am J Psychol* 1990;**80**:342–8.

4. Jenkinson C, McGee H. Health status measurement. Abingdon: Radcliffe Medical; 1998.

5. Brooks R, Bauman A, Daffurn K, Hillman K. Post-hospital outcome following intensive care. *Clin Intensive Care* 1995;**6**:127–35.

6. Chelluri L, Grenvik A, Silverman M. Intensive care for critically ill elderly: mortality, costs and quality of life. *Arch Intern Med* 1995;**155**:1013–22.

7. Brooks N. Quality of life after intensive care. *Nurs Crit Care* 1996;**1**:90–5.

8. Heyland D, Guyatt G, Cook D, Meade M, Juniper E, Cronin L, *et al.* The frequency and methodological rigour of quality of life assessments in the critical care literature. *Crit Care Med* 1998;**26**:591–8.

9. Schuster HP. Outcome nach intensivtherapie. *Med Klin* 1998;**93**:91–8.

10. Thomas EO, Manara AR. A review of percutaneous tracheostomy. *Br J Intensive Care* 1998;**8**(4):122–30.

11. Hernandez GM, Pages T, Tejero RJ, Soler D. Life after the ICU. *Rev Enfermeria* 1983;**6**(62):54–5.

12. Zarén B. Prognosis of outcome from intensive care patients [PhD thesis]. Acta Universitatis Upsaliensis, 1987.

13. Zumbotel V, Hentsch S, Kollig E. Long term prognosis of the critically ill intensive care patient. *Langenbecks Arch Chir* 1996;**113**:289–92.

14. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault J-C. Lung volumes and forced ventilatory flows. Report of working party on standardization of lung function tests, European Community for Steel and Coal. Official statement of the European Respiratory Society. *Eur Respir J* 1993;**6** Suppl 16:5–40.

15. Friman L, Hedenstierna G, Schildt B. Stenosis following tracheostomy. A qualitative study of long-term results. *Anaesthesia* 1976;**31**:479–93.

16. Lund T, Goodwin CW, McManus WF, Shirani KZ, Stallings RJ, Mason AD, *et al.* Upper airway sequelae in burn patients requiring endotracheal intubation or tracheostomy. *Ann Surg* 1985;**201**:374–82.

17. Elliott CG, Rasmusson BY, Crapo RO, Morris AH, Jensen RL. Prediction of pulmonary function abnormalities after adult respiratory distress syndrome (ARDS). *Am Rev Respir Dis* 1987;**135**:634–8.

18. Elliot CG, Rasmusson BY, Crapo RO. Upper airway obstruction following adult respiratory distress syndrome. An analysis of 30 survivors. *Chest* 1988;**94**:526–30.

19. Ghio AJ, Elliott CG, Crapo RO, Berlin SL, Jensen RL. Impairment after adult respiratory distress syndrome: an evaluation based on American Thoracic Society Recommendations. *Am Rev Respir Dis* 1989;**139**:1158–62.

20. Peters JI, Bell RC, Prihoda TJ, Harris G, Andrews C, Johanson WG. Clinical determinants of abnormalities in pulmonary functions in survivors of the adult respiratory distress syndrome. *Am Rev Respir Dis* 1989;**139**:1163–8.

21. McHugh LG, Milberg JA, Whitcomb ME, Schoene RB, Maunder RJ, Hudson LD. Recovery of function in survivors of the adult respiratory distress syndrome. *Am J Respir Crit Care Med* 1994;**150**:90–4.

22. Jones C, Macmillan RR, Harris C, Griffiths RD. Severe tracheal stenosis associated with reintubations. *Clin Intensive Care* 1997;**8**:122–5.

23. Law RC, Carney AS, Manara AR. Long-term outcome after percutaneous dilational tracheostomy. *Anaesthesia* 1997;**52**:51–6.

24. Halevy A, Sirik Z, Adam YG, Lewinsohn G. Long-term evaluation of patients following the adult respiratory distress syndrome. *Respir Care* 1984;**29**:132–7.

25. Grotz M, Hohensee A, Remmers D, Wagner TOF, Regel G. Rehabilitation results of patients with multiple injuries and multiple organ failure and long-term intensive care. *J Trauma* 1997;**42**:919–26.

26. Crapo RO, Morris AH. Standardised single breath normal values for carbon monoxide diffusing capacity. *Am Rev Respir Dis* 1981;**123**:185–9.

27. Crapo RO, Morris AH, Gardner RM. Reference spirometric values using techniques and equipment that meets ATS recommendations. *Am Rev Respir Dis* 1981;**123**:659–64.

28. Miller A, Thorton JC, Warshaw R, Anderson H, Teirstein A, Selikoff I. Single breath diffusing capacity in a representative sample of the population of Michigan, a large industrial state. *Am Rev Respir Dis* 1983;**127**:270–7.

29. Congleton J, Hodson ME, Duncan-Skingle F. Do Nottingham Health Profile scores change over time in cystic fibrosis? *Respir Med* 1998;**92**:268–72.

30. Preito L, Alonso J, Ferrer M, Anto JM. Are results of the SF-36 health survey and the Nottingham Health Profile similar? A comparison in COPD patients. *J Clin Epidemiol* 1997;**50**:463–73.

31. Bousquet J, Knani J, Dhivert H, Richard A, Chicoye A, Ware JE, *et al.* Quality of life in asthma. Internal consistency and validity of the SF-36 questionnaire. *Am J Respir Crit Care Med* 1994;**149**:371–5.

32. Okamoto LJ, Noonan M, DeBoisblanc BP, Kellerman DJ. Flucticasone propionate improves quality of life in patients with asthma requiring oral corticosteriods. *Ann Allergy Asthma Immunol* 1996;**76**:455–61.

33. van der Molen T, Postma DS, Schreurs AJ, Bosveld HE, Sears MR, Meyboom de Jong B. Discriminative aspects of two generic and two asthma specific instruments: relation with symptoms, bronchodilator use and lung function in patients with mild asthma. *Qual Life Res* 1997;**6**:353–61.

34. Alonso J, Prieto L, Ferrer M, Vilagut G, Broquetas JM, Roca J, *et al.* Testing the measurement properties of the Spanish version of the SF-36 Health Survey among male patients with chronic obstructive pulmonary disease. *J Clin Epidemiol* 1998;**51**:1087–94.

35. Leidy KN, Chan KS, Coughlin C. Is the asthma quality of life questionnaire a useful measure for low-income asthmatics? *Am J Respir Crit Care Med* 1998;**158**:1082–90.

36. Ware JE, Kemp JP, Buchner DA, Singer AE, Nolop KB, Goss TF. The responsiveness of disease specific and generic health measures to changes in the severity of asthma among adults. *Qual Life Res* 1998;**7**:235–44.

37. Landercasper J, Cogbill TH, Lindesmith LA. Long-term disability after flail chest injury. *J Trauma* 1984;**24**:410–13.

38. Gammie JS, Keenan RJ, Pham SM, McGrath MF, Hattler BG, Khoshbin E, *et al.* Cardiac and pulmonary replacement. Single- versus double-lung transplantation for pulmonary hypertension. *J Thorac Cardiovasc Surg* 1998;**115**:397–403.

39. Morris H, Kanner RE, Crapo RO, Gardner RM. Clinical pulmonary function testing: a manual of uniform laboratory procedures. Salt Lake City: Intermountain Thoracic Society; 1984.

40. Gardner RM, Hankinson JL, Clausen JL, Crapo RO, Johnson RL, Epler GR. Standardization of spirometry: 1987 update. *Am Rev Respir Dis* 1987;**136**:1285–98.

41. Cotes JE. Lung function. Assessment and application in medicine. 5th ed. London: Blackwell Scientific; 1993.

42. Aass AS. Complications to tracheostomy and long-term intubation: a follow-up study. *Acta Anaesthesiol Scand* 1975;**19**:127–33.

43. Bright P, Millar MR, Franklyn JA, Sheppard MC. The use of a neural network to detect upper airway obstruction caused by goitre. *Am J Respir Crit Care Med* 1998;**157**:1885–91.

44. Stauffer JL, Olson DE, Petty TL. Complications and consequences of endotracheal intubation and tracheotomy. *Am J Med* 1981;**70**:65–76.

45. Kennedy SM, Desjardins A, Kassam A, Ricketts M, Chan-Yeung M. Assessment of respiratory limitation in activities of daily life among retired workers. *Am J Resp Crit Care Med* 1994;**149**:575–83.

46. Cherniack RM, Raber MB. Normal standards for ventilatory function using automated wedge spirometer. *Am Rev Respir Dis* 1972;**106**:38–46.

47. Casale R, Pasqualetti P. Cosinor analysis of circadian peak expiratory flow variability in normal subjects, passive smokers, heavy smokers, patients with chronic obstructive pulmonary disease and patients with interstitial lung disease. *Respiration* 1997;**64**:251–6.

48. Iwasaki Y, Kubota Y, Yokomura I, Ueda M, Hashimoto S, Mizobuchi K, *et al.* Usefulness and early detection of acute exacerbation by peak expiratory flow rate in the management of asthma. *Jpn J Thorac Dis* 1997;**35**:1040–6.

49. Kraemore MD, Kubo SH, Rector TS, Brunsvold N, Bank AJ. Pulmonary and peripheral vascular factors are important determinants of peak exercise oxygen uptake in patients with heart failure. *J Am Coll Cardiol* 1998;**21**:641–8.

50. Hajiro T, Nishimura K, Tsukino M, Ikeda A, Koyama H, Izumi T. Comparison of discriminative properties among disease-specific questionnaires for measuring health related quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;**157**:785–90.

51. Gaensler EA, Wright GW. Evaluation of respiratory impairment. *Arch Environ Health* 1966;**12**:146–89.

52. Oh TE. Intensive care manual. Sydney: Butterworths; 1990.

53. McAdams HP, Palmer SM, Erasmus JJ, Patz EF, Connolly JE, Goodman PC, *et al.* Bronchial anastomotic complications in lung transplant recipients: virtual bronchoscopy for non-invasive assessment. *Radiology* 1998;**209**:689–95.

54. Dane TEB, King EG. A prospective study of complications after tracheostomy for assisted ventilation. *Chest* 1975;**67**:398–404.

55. Sellery GR, Worth A, Greenway RE. Late complications of prolonged tracheal intubation. *Can Anaesth Soc J* 1978;**25**:140–3.

56. Winkler W-B, Karnik R, Seelmann O, Havlicek J, Slany J. Bedside percutaneous dilational tracheostomy with endoscopic guidance: experience with 71 ICU patients. *Intensive Care Med* 1994;**20**:476–79.

57. Lengas A, Patelli M, Poletti V, Spiga L. Efficacy of bronchoscopic Nd:YAG-laser for laryngotracheal stenosis due to prolonged intubation. *Chirurg* 1996;**67**:730–3.

58. Walz MK, Peitgen K, Thurauf N, Trost HA, Wolfhard U, Sander A, *et al.* Percutaneous dilational tracheostomy – early results and long term outcome of 326 critically ill patients. *Intensive Care Med* 1998;**24**:685–90.

59. Martinelli L, Rinaldi M, Pederzolli C, Pederzolli N, Goggi C, Mantovani V, *et al.* Different results of cardiac transplantation in patients with ischemic and dilated cardiomyopathy. *Eur J Cardiothorac Surg* 1995;**9**:644–50.

60. Nordback IH, Auvinen OA. Long-term results after pancreas resection for acute necrotising pancreatitis. *Br J Surg* 1985;**72**:687–9.

61. Doepel M, Eriksson J, Halme L, Kumpulainen T, Höckerstedt K. Good long term results in patients surviving severe acute pancreatitis. *Br J Surg* 1993;**80**:1583–6.

62. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of illness in the aged. The index of ADL: a standardized measure of biological and psychosocial function. *JAMA* 1963;**185**:914–19.

63. Katz S, Ford AB, Chinn AB, *et al.* Prognosis after strokes: long term course of 159 patients with stroke. *Medicine* 1966;**45**:236–46.

64. Katz S, Vignos PJ, Moskowitz RW, Thompson HM, Svec KH. Comprehensive outpatient care in rheumatoid arthritis: a controlled study. *JAMA* 1968;**206**:1249.

65. Katz S, Downs TD, Cash HR, *et al.* Progress in the development of an index of ADL. *Gerontologist* 1970;**10**:20–30.

66. Levy DE, Caronna JJ, Singer BH, Lapinski RH, Frydman H, Plum F. Predicting outcome from hypoxic-ischemic coma. *JAMA* 1985;**253**:1420–6.

67. Ridley SA, Wallace PGM. Quality of life after intensive care. *Anaesthesia* 1990;**45**:808–13.

68. Kass JE, Castriotta RJ, Malkakoff F. Intensive care unit outcome in the very elderly. *Crit Care Med* 1992;**20**:1666–71.

69. Rockwood K, Noseworthy TW, Gibney RTN, Konopad E, Shustack A, Stollery D, *et al.* One-year outcome of elderly and young patients admitted to intensive care units. *Crit Care Med* 1993;**21**:687–91.

70. Chelluri L, Pinsky MR, Donahoe MP, Grenvik A. Long-term outcome of critically ill elderly patients requiring intensive care. *JAMA* 1993;**269**:3119–23.

71. Holbrook TL, Hoyt DB, Anderson JP, Hollingsworth-Fridlund P, Shackford SR. Functional limitaton after major trauma: a more sensitive assessment using the quality of well-being scale – the trauma recovery pilot project. *J Trauma* 1974;**36**:74–8.

72. Broslawski GE, Elkins M, Algus M. Functional abilities of elderly survivors of intensive care. *J Am Osteopath Assoc* 1995;**95**:712–17.

73. Wu AW, Damiano AM, Lynn J, Alzola C, Teno J, Landefeld CS, *et al.* Predicting future functional status for seriously ill hospitalized adults: the SUPPORT prognostic model. *Ann Intern Med* 1995;**122**:342–50.

74. Dardaine V, Constans T, Lasfargues G, Perrotin D, Giniès G. Outcome of elderly patients requiring ventilatory support in intensive care. *Aging Clin Exp Res* 1995;**7**:221–7.

75. Cho DY, Wang YC. Comparison of the APACHE III, APACHE II and Glasgow Coma Scale in acute head injury for prediction of mortality and functional outcome. *Intensive Care Med* 1997;**23**:77–84.

76. Battistella FD, Din AM, Perez L. Trauma patients 75 years and older: long-term follow-up results justify aggressive management. *J Trauma* 1998;**44**:618–23.

77. Deleted.

78. Karnofsky DA, Abelmann WH, Craver LF, *et al.* The use of nitrogen mustards in the palliative treatment of carcinoma. *Cancer* 1948;**1**:634–56.

79. Bowling A. Measuring disease. 2nd ed. Buckingham: Open University Press; 1997.

80. Mor V, Laliberte L, Morris JN, Wiemann M. The Karnofsky performance status scale: an examination of its reliability and validity in a research setting. *Cancer* 1984;**53**:2002–7.

81. Evans RW, Manninen DL, Overcaste TD, *et al.* The national heart transplantation study: final report. Washington, DC: Battelle Human Affairs Research Center; 1984.

82. Bakker W, van Oosterom AT, Aaronson NK, van Breukelen FJ, Bins MC, Hermans J. Vindesine, cisplatin, and bleomycin combination chemotherapy in non-small cell lung cancer: survival and quality of life. *Eur J Cancer Clin Oncol* 1986;**22**:963–70.

83. Splinter TAW. Chemotherapy in advanced non-small cell lung cancer. *Eur J Cancer* 1990;**2**:287–95.

84. Yates JW, Chalmer B, McKegney FP. Evaluation of patients with advanced cancer using the Karnofsky performance status. *Cancer* 1980;**45**:2220–4.

85. Schipper H, Clinch J, McMurray A, Levitt M. Measuring the quality of life of cancer patients: the functional living index – cancer: development and validation. *J Clin Oncol* 1984;**2**:472–83.

86. Hutchinson TA, Boyd NF, Feinstein AR, Gonda A, Hollomby D, Rowat B. Scientific problems in clinical scales as demonstrated in the Karnofsky Index of Performance Status. *J Chron Dis* 1979;**32**:661–6.

87. Mercier M, Schraub S, Bransfield DD, Fournier J. Patient acceptance and differential perceptions of quality of life measures in a French oncology setting. *Qual Life Res* 1992;**1**:53–61.

88. Yinnon A, Zimran A, Hershko C. Quality of life and survival following intensive medical care. *Q J Med* 1989;**71**:347–57.

89. Kumar P, Zehr KJ, Chang A, Cameron DE, Baumgartner WA. Quality of life in octogenarians after open heart surgery. *Chest* 1995;**108**:919–26.

90. Singh N, Gayowski T, Wagener MM. Intensive care unit management in liver transplant recipients: beneficial effect on survival and preservation of quality of life. *Clin Transplant* 1997;**11**:113–20.

91. Weinert CR, Gross CR, Kangas JR, Bury CL, Marinelli WA. Health related quality of life after acute lung injury. *Am J Respir Crit Care Med* 1997;**156**:1120–8.

92. Kocher A, De Torrenté A. Survie et qualité de vie après un séjour aux soins intensifs. *Rev Med Suisse Romande* 1998;**118**:215–20.

93. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J* 1965;**14**:61–5.

94. Bowling A. Measuring health: a review of quality of life measurement scales. 2nd ed. Buckingham: Open University Press; 1998.

95. Mattison PG, Aitkin RCB, Prescott RJ. Rehabilitation status – the relationship between the Edinburgh Rehabilitation Status Scale (ERSS), Barthel Index and PULSES profile. *Int Disabil Studies* 1991;**13**:9–11.

96. Affleck JW, Aitkin RC, Hunter JA, McGuire RJ, Roy CW. Rehabilitation status. A measure of medico-social dysfunction. *Lancet* 1988;**i**:230–3.

97. Granger CV, Dewis LS, Peters NC, Sherwood CC, Barrett JE. Stroke rehabilitation: analysis of repeated Barthel Index measures. *Arch Phys Med Rehabil* 1979;**60**:14–17.

98. Wade DT, Langton-Hewer R. Functional abilities after stroke: measurement, natural history and prognosis. *J Neurol Neurosurg Psychiatry* 1987;**50**:177–82.

99. Sherwood SJ, Morris J, Mor V, Gutkin C. Compendium of measures for describing and assessing long term care populations. Boston, MA: Hebrew Rehabilitation Center for the Aged; 1977.

100. Day RJ, Vinen J, Hewitt-Falls E. Major trauma outcomes in the elderly. *Med J Aust* 1994;**160**:675–8.

101. Gobiet W. Influence of polytrauma on rehabilitation and outcome in patients with severe head injury. *Zentralbl Chir* 1995;**120**:544–50.

102. Neundörfer B, Hilz MJ, Wimbauer M. Course of disease and prognosis of neurological intensive care unit patients with special respect to age. *Fortsch Neurol Psychiatr* 1996;**64**:285–91.

103. Madsen M, Eriksen LS. Epidemiology of acute myocardial infarctions in the Copenhagen area. *Ugeskr Laeger* 1979;**141**:1152–4.

104. Jensen NH, Dragsted L, Christensen JK, Jørgensen JC, Qvist J. Severity of illness and outcome of treatment in alcoholic patients in the intensive care unit. *Intensive Care Med* 1988;**15**:19–22.

105. Dragsted L, Qvist J. Outcome from intensive care: III. A five year study of 1308 patients: activity levels. *Eur J Anaesthesiol* 1989;**6**:385–96.

106. Dragsted L. Outcome from intensive care IV: a five year study of 1308 patients. Long-term outcome. *Eur J Anaesth* 1990;**7**:51–62.

107. Schuster HP. Intensive care in old age. *Med Klin* 1991;**86**:473–81.

108. Cullen DJ, Ferrara LC, Briggs BA, Walker PF, Gilbert J. Survival, hospitalisation charges and follow-up results in critically ill patients. *N Engl J Med* 1976;**294**:982–7.

109. Cullen DJ, Keene R, Waternaux C, Kunsman JM, Caldera DL, Peterson H. Results, charges, and benefits of intensive care for critically ill patients: update 1983. *Crit Care Med* 1984;**12**:102–6.

110. Slatyer MA, James OF, Moore PG, Leeder SR. Costs, severity of illness and outcome in intensive care. *Anaesth Intensive Care* 1986;**14**:381–9.

111. Nolla-Salas J, Vazquez A, Carrasco G, Marrugat J, Solsona JF. Elderly patients admitted at ICU: analysis of the reasons for admittance, evolution and quality of life at medium term. *Med Intensiva* 1993;**17**:33–9.

112. New York Heart Association: Criteria Committee. Nomenclature and criteria for diagnosis of disease of the heart and blood vessels. Boston: Little, Brown; 1964.

113. Goldman L, Cook EF, Mitchell N, Flatley M, Sherman H, Cohn PF. Pitfalls in the serial assessment of cardiac functional states: How a reduction of 'ordinary' activity may reduce the apparent degree of cardiac compromise and give a misleading impression of improvement. *J Chronic Dis* 1982;**35**:763–71.

114. Olsson G, Lubsen J, van Es GA, Rehqvist N. Quality of life after myocardial infarction: effect of long term metoprolol on mortality and morbidity. *BMJ* 1986;**292**:1491–3.

115. Wiklund I, Lindvall K, Swedberg K, Zupkis RV. Self assessment of quality of life in severe heart failure: an instrument for clinical use. *Scand J Psychol* 1987;**28**:220–5.

116. Wiklund I, Gorkin L, Pawitan Y, Schron E, Schoenberger J, Jared LL, *et al.* Methods for assessing quality of life in the cardiac arrhythmia suppression trial (CAST). *Qual Life Res* 1992;**1**:187–201.

117. Lee WA, Gillinov AM, Cameron DE, Thompson S, Birenbaum I, Baumgartner WA, *et al.* Centrifugal ventricular assist device for support of the failing heart after cardiac surgery. *Crit Care Med* 1993;**21**:1186–91.

118. McHugh GJ, Havill JH, Armistead SH, Ullal RR, Fayers TM. Follow up of elderly patients after cardiac surgery and intensive care unit admission, 1991 to 1995. *N Z Med J* 1997;**110**:432–5.

119. Trouillet JL, Scheimberg A, Vuagnat A, Fagon JY, Chastre J, Gibert C. Long-term outcome and quality of life of patients requiring multidisciplinary intensive care unit admission after cardiac operations. *J Thorac Cardiovasc Surg* 1996;**112**:926–34.

120. Ferris BG. Epidemiology standardisation project. *Am Rev Respir Dis* 1978;**118**:10–36.

121. Osterman JW, Brochu D, Theriault G, Greaves IA. Evaluation of the ATS respiratory diseases questionnaire among French-speaking silicon carbide workers. *Can J Public Health* 1990;**81**:66–72.

122. Hopp RJ, Biven RE, Degan FA, Bewtra AK, Townley RG. The usefulness of questionnaire-derived information to predict the degree of nonspecific bronchial hyper-responsiveness. *Allergy Proc* 1995;**16**:129–34.

123. Guyatt GH, Sullivan MJ, Thompson PJ, Fallen EL, Pugsley SO, Taylor DW, *et al.* The six minute walking test: a new measure of exercise capacity in patients with chronic heart failure. *Can Med Assoc J* 1985;**132**:919.

124. Eakin EG, Kaplan RM, Ries AL. Measurement of dyspnoea in chronic obstructive pulmonary disease. *Qual Life Res* 1993;**2**:181–91.

125. Johnson AN, Cooper DF, Edwards RHT. Exertion of stair climbing in normal subjects and in patients with chronic obstructive bronchitis. *Thorax* 1977;**32**:711–16.

126. Roul G, Germain P, Bareiss P. Does the 6-minute walk test predict the prognosis in patients with NHYA class II or III chronic heart failure? *Am Heart J* 1998;**136**:449–57.

127. Weaver TE, Narsavage GL, Guilfoyle MJ. The development and psychometric evaluation of the pulmonary functional status scale: an instrument to assess functional status in pulmonary disease. *J Cardiopulm Rehabil* 1998;**18**:105–11.

128. Weir L, Waldmann CS. Long term follow up of ITU patients. *Auditorium* 1994:40–1.

129. Waldmann C, Gaine M. The intensive care follow-up clinic. *Care Crit Ill* 1996;**12**:118–21.

130. McNair DM, Lorr M, Droppleman LF. Manual for the Profile Of Mood States. San Diego, Cal: Educational and Industrial Testing Services; 1971.

131. McNair DM, Lorr M, Droppleman LF. EdITS Manual for the Profile of Mood States. San Diego, Cal: Educational and Industrial Testing Service; 1992.

132. Derogatis LR. The psychological adjustment to illness scale. *J Psychosom Res* 1986;**30**:77–91.

133. Lorr M, McNair D, Fisher S. Evidence for bipolar mood states. *J Pers Assess* 1982;**46**:432–6.

134. Norcross JC, Guadagnoli E, Prochaska JO. Factor structure of the Profile Of Mood States (POMS): two partial replications. *J Clin Psychol* 1984;**40**:77.

135. McCorkle R, Quint-Benoliel J. Symptom distress, current concerns and mood disturbances after diagnosis of threatening disease. *Soc Sci Med* 1983;**17**:431–8.

136. Cassileth BR, Lusk EJ, Brown LL, *et al.* Psychosocial status of cancer patients and next of kin: normative data from the Profile of Mood States. *J Clin Oncol* 1985;**3**:99–105.

137. Stanton BA, Jenkins CD, Denlinger P, Savageau JA, Weintraub RM, Goldstein RL. Predictors of employment status after cardiac surgery. *JAMA* 1983;**249**:907–11.

138. Stambrook M, Moore AD, Peters LC, Deviaene C, Hawryluk GA. Effects of mild, moderate and severe closed head injury on long-term vocational status. *Brain Injury* 1990;**4**:183–90.

139. Sawdon V, Woods I, Proctor M. Post-intensive care interviews: implications for future practice. *Intensive Crit Care Nurs* 1995;**11**:329–32.

140. Jones C, Griffiths RD, Macmillan R, Palmer TEA. Psychological problems occurring after intensive care. *Br J Intensive Care* 1994;**4**(February):46–53.

141. Patrick DL, Danis M, Southerland LI, Hong G. Quality of life following intensive care. *J Gen Intern Med* 1988;**3**:218–23.

142. Radloff LS. Sex differences in depression: the effects of occupation and marital status. *Sex Roles* 1975;**1**:249–65.

143. Radloff LS. The CES-D scale: a self report depression scale for research in the general population. *Appl Psychol Measure* 1977;**1**:385–401.

144. Eaton WW, Kessler LG. Rates of symptoms of depression in a national sample. *Am J Epidemiol* 1981;**114**:528–38.

145. Radloff LS, Locke BZ. The community mental health assessment survey and the CES-D scale. In: Weissman MM, Myers JK, Ross CE, editors. Community surveys of psychiatric disorders. New Brunswick, NJ: Rutgers University Press; 1985.

146. Shumaker SA, Schron E ,Ockene J, editors. The handbook of health behavior change. New York: Springer-Verlag; 1990.

147. Shaver PR, Brennan KA. Measures of depression and loneliness. In: Robinson JP, Shaver PR, Wrightsman LS, editors. Measures of personality and social psychological attitudes. San Diego, Cal: Academic Press; 1991.

148. Weissman MM, Prusoff B, Newberry PB. Comparisons of CES-D, Zung and Beck self report depression scales. Rockville, MD: Center for Epidemiological Studies, National Institute of Mental Health; 1975. Technical report ADM 42-47-83.

149. Lyness JM, Noel TK, Cox C, King D, Conwell Y, Caine ED. Screening for depression in elderly primary care patients. *Arch Intern Med* 1997;**157**:449–54.

150. Bedell SE, Delbanco TL, Cook EF, Epstein FH. Survival after cardiopulmonary resuscitation in the hospital. *N Engl J Med* 1983;**309**:569–76.

151. Zigmond AS, Snaith RP. The Hospital Anxiety And Depression Scale. *Acta Psychiatr Scand* 1983;**67**:361–70.

152. Aylard PR, Gooding JH, McKenna PJ, Snaith RP. A validation study of three anxiety and depression self assessment scales. *Psychosom Res* 1987;**31**:261–8.

153. Lewis G, Wessely S. General Health questionnaire and the Hospital Anxiety and Depression Scale. *Br J Psychiatry* 1990;**157**:860–4.

154. Tyrer P, Seivewright N, Murphy S, Ferguson B, Kirigolon D, Barezak P, *et al.* The Nottingham study of neurotic disorder: comparison of drug and psychological treatments. *Lancet* 1988;**ii**:235–40.

155. Rowan KM. Outcome comparisons of intensive care units in Great Britain and Ireland using APACHE II method [DPhil thesis]. Oxford: University of Oxford; 1992.

156. Dixon JJ, Manara AR, Willats SM. Patient and relative satisfaction with intensive care. Importance of duration and quality of life. *Clin Intensive Care* 1997;**8**:63–8.

157. Beck AT, Ward C, Mendelson M, *et al.* An inventory for measuring depression. *Arch Gen Psychiatry* 1961;**4**:561–71.

158. Beck AT. Depression: causes and treatment. Philadelphia: University of Pennsylvania Press; 1970.

159. Hamilton M. Rating scale for depression. *J Neurol Neurosurg Psychiatry* 1970;**23**:56–62.

160. Beck AT, Steer RA, Garbin MG. Psychometric properties of the BDI: twenty-five years of evaluation. *Clin Psychol Rev* 1988;**8**:77–100.

161. Joseph S, Lewis CA. The depression–happiness scale: reliability and validity of a bipolar self-report scale. *J Clin Psychol* 1998;**4**:537–44.

162. Suarez-Mendoza AA, Cardiel MH, Caballero-Uribe CV, Ortga-Soto HA, Marquez-Marin M. Measurement of depression in Mexican patients with rheumatoid arthritis: validity of the Beck Depression Inventory. *Arthritis Care Res* 1997;**10**:194–9.

163. Gallagher D, Nies G, Thompson LW. Reliability of the Beck Depression Inventory with older adults. *J Consult Clin Psychol* 1982;**50**:152–3.

164. Ahava GW, Iannone C, Grebstein L, Schirling J. Is the Beck Depression Inventory reliable over time? An evaluation of multiple test–retest reliability in a non-clinical college student sample. *J Pers Assess* 1998;**70**:222–31.

165. Riether AM, Smith SL, Lewison BJ, Cotsonis GA, Epstein CM. Quality-of-life changes and psychiatric and neurocognitive outcome after heart and liver transplantation. *Transplantation* 1992;**54**:444–50.

166. Horowitz M, Wilner N, Alvarez W. Impact of Events Scale: a measure of subjective distress. *Psychosom Med* 1979;**41**:209.

167. Zilberg NJ, Weiss DS, Horowitz MJ. Impact of Events Scale: a cross validation study and some empirical evidence supporting a conceptual model of stress response syndromes. *J Consult Clin Psychol* 1982;**50**:407.

168. Briere J, Elliott DM. Clinical utility of the impact of event scale: psychometrics in the general population. *Assessment* 1998;**5**:171–80.

169. Landsman IS, Baum CG, Arnkoff DB, Craig MJ, Lynch I, Copes WS, *et al.* The psychosocial consequences of traumatic injury. *J Behav Med* 1990;**13**:561–81.

170. Richmond TS, Kauder D, Schwab CW. A prospective study of predictors of disability at 3 months after non-central nervous system trauma. *J Trauma* 1998;**44**:635–43.

171. Reitan RM. Trail-making manual for administration, scoring and interpretation. Indianapolis: Indiana University Medical Center; 1958.

172. Jones C, Hussey R, Griffiths RD. A tool to measure the change in health status of selected adult patients before and after intensive care. *Clin Intensive Care* 1993;**4**:160–5.

173. Uzzell BP, Langfitt TW, Dolinskas CA. Influence of injury severity on quality of survival after head injury. *Surg Neurol* 1987;**27**:419–29.

174. McKee MD, Schemitsch EH, Vincent LO, Sullivan I, Yoo D, Gruen GS. The effect of a femoral fracture on concomitant closed head injury in patients with multiple injuries. *J Trauma* 1997;**42**:1041–5.

175. Heaton RK. Wisconsin Card Sorting Test manual. Odessa, Florida: Psychological Assessment Resources; 1981.

176. Watkins KE, Hewes DK, Connelly A, Kendall BE, Kingsley DP, Evans JE, *et al.* Cognitive deficits associated with frontal lobe infarction in children with sickle cell disease. *Dev Med Child Neurol* 1998;**40**:536–43.

177. Zihl J, Groon G, Brunnauer A. Cognitive deficits in schizophrenia and affective disorders: evidence for a final common pathway disorder. *Acta Psychiatr Scand* 1998;**97**:351–7.

178. Anderson SI, Taylor R, Jones PA, Miller JD. A comparison of neuropsychological and functional outcome, and uptake of rehabilitation services, following severe and moderate head injury. *Clin Rehabil* 1994;**8**:301–6.

179. Lannoo E, Colardyn F, de Deyne C, Vandekerckhove T, Jannes C, de Soete G. Cerebral perfusion pressure and intracranial pressure in relation to neuropsychological outcome. *Intensive Care Med* 1998;**24**:236–41.

180. Wechsler D. A standardised memory scale for clinical use. *J Psychol* 1945;**19**:87–95.

181. Wechsler D. Wechsler Memory Scale – Revised (WMS-R). Sidcup, UK: Psychological Corporation; 1988.

182. Kane RA, Kane RL. Assessing the elderly. Lexington, MA: Lexington Books; 1988.

183. Uzzell BP, Obrist WD, Dolinskas CA, Langfitt TW. Relationship of acute CBF and ICP findings to neuropsychological outcome in severe head injury. *J Neurosurg* 1986;**65**:630–5.

184. Benton AL. The revised visual retention test. Iowa City: University of Iowa; 1955.

185. Benton AL. Revised visual retention test: clinical and experimental application. New York: Psychological Corporation; 1974.

186. Crookes TG, McDonald KG. Benton's visual retention test in the differentiation of depression and early dementia. *Br J Soc Clin Psychol* 1972;**11**:66–9.

187. Weiss AA. Equivalence of three alternative forms of Benton's visual retention test. *Percept Mot Skills* 1974;**38**:623–35.

188. Alexandre A, Colombo F, Nertempi P, Benedetti A. Cognitive outcome and early indices of severity of head injury. *J Neurosurg* 1983;**59**:751–61.

189. Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state': a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;**12**:189–98.

190. Nelson A, Fogal BS, Faust D. Bedside cognitive screening instruments: a critical assessment. *J Nerv Ment Dis* 1986;**174**:73–83.

191. Molloy DW, Standish TI. A guide to the standardized Mini-Mental State Examination. *Int Psychogeriatrics* 1997;**9**(Suppl 1):87–94.

192. Frutiger A, Ryf C, Bilat C, Rosso R, Furrer M, Cantieni R, *et al.* Five years' follow-up of severely injured ICU patients. *J Trauma* 1991;**31**:1216–26.

193. Gronwall DM. Paced auditory serial-addition task: a measure of recovery from concussion. *Percept Mot Skills* 1977;**44**:367–73.

194. Wiens AN, Fuller KH, Crossen JR. Paced Auditory Serial Addition Test adult norms and moderator variables. *J Clin Exp Neuropsychol* 1997;**19**:473–83.

195. Litvan I, Grafman J, Vendrell P, Martinez JM. Slowed information processing in multiple sclerosis. *Arch Neurol* 1988;**45**:281–5.

196. Jennett B, Bond M. Assessment of outcome after severe brain damage. *Lancet* 1975;**i**:480–4.

197. Jennett B. Disability after severe head injury: observations on the use of the Glasgow Outcome Scale. *J Neurol Neurosurg Psychiatry* 1981;**44**:285–93.

198. Maas AIR, Braakman R, Schouten HJA, Minderhoud JM, van Zomeren AH. Agreement between physicians on assessment of outcome following severe head injury. *J Neurosurg* 1983;**58**:321–5.

199. Levy DE, Bates D, Caronna JJ, Cartlidge NEF, Knill-Jones RP, Lapinski RH, *et al.* Prognosis in nontraumatic coma. *Ann Intern Med* 1981;**94**:293–301.

200. Alberico AM, Ward JD, Choi SC, Marmarou A, Young HF. Outcome after severe head injury: relationship to mass lesions, diffuse injury and ICP course in pediatric and adult patients. *J Neurosurg* 1987;**67**:648–56.

201. Zarén B, Hedstrand U. Quality of life among long-term survivors of intensive care. *Crit Care Med* 1987;**15**:743–7.

202. Stocchetti N, Serioli T, Mergoni M. Intensive treatment of head injured patient: analysis of 354 cases. *Agressologie* 1988;**29**:249–51.

203. Nordström CH, Sundbärg G, Messeter K, Schalén W. Severe traumatic brain lesions in Sweden: part 2. Impact of aggressive neurosurgical intensive care. *Brain Injury* 1989;**3**:267–81.

204. Zarén B, Bergström R. Survival compared to the general population and changes in health status among intensive care patients. *Acta Anaesthesiol Scand* 1989;**33**:6–12.

205. Judson JA, Cant BR, Shaw NA. Early prediction of outcome from cerebral trauma by somatosensory evoked potentials. *Crit Care Med* 1990;**18**:363–8.

206. Marshall LF, Gautille T, Klauber MR, Eisenberg HM, Jane JA, Luerssen TG. The outcome of severe closed head injury. *J Neurosurg* 1991;**75**:S28–36.

207. Wärme PE, Bergström R, Persson L. Neurosurgical intensive care improves outcome after severe head injury. *Acta Neurochir* 1991;**110**:57–64.

208. Fearnside MR, Cook RJ, McDougall P, McNeil RJ. The Westmead Head Injury Project outcome in severe head injury. A comparative analysis of pre-hospital, clinical and CT variables. *Br J Neurosurg* 1993;**7**:267–79.

209. Jones PA, Andrews PJD, Midgley S, Anderson SI, Piper IR, Tocher JL, *et al.* Measuring the burden of secondary insults in head-injured patients during intensive care. *J Neurosurg Anesthesiol* 1994;**6**:4–14.

210. Fernandez RR, Sanchez Cruz JJ, Vazquez Mata GV. Validation of a quality of life questionnaire for critically ill patients. *Intensive Care Med* 1996;**22**:1034–42.

211. Cruz J. Relationship between early patterns of cerebral extraction of oxygen and outcome from severe acute traumatic brain swelling: cerebral ischaemia or cerebral viability? *Crit Care Med* 1996;**24**:953–6.

212. Vasquez Mata G, Riviera Fernandez R, Gonzalez Carmona A, Delgado-Rodriguez M, Torres Ruiz JM, Raya Pungaire A, *et al.* Factors related to quality of life 12 months after discharge from an intensive care unit. *Crit Care Med* 1992;**20**:1257–62.

213. Heinzelmann M, Platz A, Imhof HG. Outcome after acute extradural haematoma, influence of additional injuries and neurological complications in the ICU. *Injury* 1996;**27**:345–9.

214. Pohlmann-Eden B, Dingethal K, Bender HJ, Koelfen W. How reliable is the predictive value of SEP (Somatosensory Evoked Potentials) patterns in severe brain damage with special regard to the bilateral loss of cortical responses? *Intensive Care Med* 1997;**23**:301–8.

215. Enblad P, Persson L. Impact on clinical outcome of secondary brain insults during the neurointensive care of patients with subarachnoid haemorrhage: a pilot study. *J Neurol Neurosurg Psychiatry* 1997;**62**:512–16.

216. Resnick DK, Marion DW, Carlier P. Outcome analysis of patients with severe head injuries and prolonged intracranial hypertension. *J Trauma* 1997;**42**:1108–11.

217. Cruz J. The first decade of continuous monitoring of jugular bulb oxyhemoglobin saturation: management strategies and clinical outcome. *Crit Care Med* 1998;**26**:344–51.

218. Bell D, Turpin K. Quality of life at three months following admission to intensive and coronary care units. *Clin Intensive Care* 1994;**5**:276–81.

219. Pessi TT. Experiences gained in intensive care of surgical patients. A prospective clinical study of 1001 consecutively treated patients in a surgical intensive care unit. *Ann Chir Gynaecol Fenniae* 1973;**62** Suppl 185:1–72.

220. Alho A, Rokkanen P. Recovery of patients after intensive care for blunt injuries. *Ann Chir Gynaecol Fenniae* 1973;**62**:64–8.

221. Bürgisser C, Ritz R. Follow-up of intensive care medical patients. *Schweiz Med Wochenschr* 1982;**112**:1283–6.

222. Bergner L, Bergner M, Hallstrom AP, Eisenberg M, Cobb LA. Health status of survivors of out-of-hospital cardiac arrest six months later. *Am J Public Health* 1984;**74**:508–10.

223. Parno JR, Teres D, Lemeshow S, Brown RB, Avrunin JS. Two-year outcome of adult intensive care patients. *Med Care* 1984;**22**:167–76.

224. Goldstein RL, Campion EW, Thibault GE, Mulley AG Skinner E. Functional outcomes following medical intensive care. *Crit Care Med* 1986;**14**:783–8.

225. Sage WM, Rosenthal MH, Silverman JF. Is intensive care worth it? An assessment of input and outcome for the critically ill. *Crit Care Med* 1986;**14**:777–82.

226. Bosatra MG, Bobbio Pallavicini F, Verde G, Bausardo MG, Braschi A, Emmi V, *et al.* Follow-up of the psychophysical sequelae in critical care patients discharged from a multidisciplinary intensive therapy unit. *Minerva Anestesiol* 1987;**53**:571–7.

227. Sage WM, Hurst CR, Silverman JF, Bortz II WM. Intensive care for the elderly: outcome of elective and nonelective admissions. *J Am Geriatr Soc* 1987;**35**:312–18.

228. MacKenzie EJ, Siegel JH, Shapiro S, Moody M, Smith RT. Functional recovery and medical costs of trauma: an analysis by type and severity of injury. *J Trauma* 1988;**28**:281–95.

229. Ritz R. Model of a prospective follow-up study of patients managed in intensive care. *Schweiz Med Wochenschr* 1988;**118**:637–42.

230. Mundt DJ, Gage RW, Lemeshow S, Pastides H, Teres D, Avrunin JS. Intensive care unit patient follow-up. Mortality, functional status, and return to work at six months. *Arch Intern Med* 1989;**149**:68–72.

231. Gaillard M, Martel S, Reynaud P, Mercadal L, Liger C, Herve C. Serious suicide: short and long term survey. Analysis of quality of life. *Agressologie* 1990;**31**:749–52.

232. Kivioja AH, Myllynen PJ, Rokkanen PU. Is the treatment of the most severe multiply injured patients worth the effort? A follow-up examination 5 to 20 years after severe multiple injury. *J Trauma* 1990;**30**:480–3.

233. Morris JA, Sanchez AA, Bass SM, MacKenzie EJ. Trauma patients return to productivity. *J Trauma* 1991;**31**:827–33.

234. Sazbon L, Zagreba F, Ronen J, Solzi P, Costeff H. Course and outcome of patients in vegetative state of nontraumatic aetiology. *J Neurol Neurosurg Psychiatry* 1993;**56**:407–9.

235. Broome AH, Eisen GM, Harland RC, Collins BH, Meyers WC, Pappas TN. Quality of life after treatment for pancreatitis. *Ann Surg* 1996;**223**:665–72.

236. Brenneman FD, Redelmeier DA, Boulanger BR, McLellan BA, Culhane JP. Long term outcomes in blunt trauma: who goes back to work? *J Trauma* 1997;**42**:778–81.

237. Hurel D, Loirat P, Saulnier F, Nicolas F, Brivat F. Quality of life 6 months after intensive care: results of a prospective multicentre study using a generic health status scale and a satisfaction scale. *Intensive Care Med* 1997;**23**:331–7.

238. Schelling G, Stoll C, Haller M, Briegel J, Manert W, Hummel T, *et al.* Health-related quality of life and posttraumatic stress disorder in survivors of the acute respiratory distress syndrome. *Crit Care Med* 1998;**26**:651–9.

239. Kriwanek S, Armbruster C, Dittrich K, Beckerhinn P, Schwarzmaier A, Redl E. Long-term outcome after open treatment of severe intra-abdominal infection and pancreatic necrosis. *Arch Surg* 1998;**133**:140–4.

240. Gaillard M, Pasquier C, Guerrini P, Margenet A, Leroux V, Herve C. Short and long term survey of 250 polytrauma patients. *Agressologie* 1990;**31**:633–6.

241. Campion EW, Mulley AG, Goldstein RL, Barnett GO, Thibault GE. Medical intensive care for the elderly. A study of current use, costs and outcomes. *JAMA* 1981;**246**:2052–6.

242. Goldstein RL, Campion EW, Mulley AG, Thibault GE. Nursing home patients admitted to a medical intensive care unit. *Med Care* 1984;**22**:854–62.

243. Oreskovich MR, Howard JD, Copass MK, Carrico CJ. Geriatric trauma: injury patterns and outcome. *J Trauma* 1984;**24**:565–9.

244. McLean RF, McIntosh JD, Kung GY, Leung DMW, Byrick RJ. Outcome of respiratory intensive care for the elderly. *Crit Care Med* 1985;**13**:625–9.

245. Mahul P, Perrot D, Tempelhoff G, Gaussorgues P, Jospe R, Ducreux JC, *et al.* Short- and long-term prognosis, functional outcome following ICU for elderly. *Intensive Care Med* 1991;**17**:7–10.

246. Goins WA, Reynolds HN, Nyanjom D, Dunham CM. Outcome following prolonged intensive care unit stay in multiple trauma patients. *Crit Care Med* 1991;**19**:339–45.

247. Chelluri L, Pinsky MR, Grenvik ANA. Outcome of intensive care of the "oldest-old" critically ill patients. *Crit Care Med* 1992;**20**:757–61.

248. Douglas SL, Daly BJ, Brennan PF, Harris S, Nochomovitz M, Dyer MA. Outcomes of long term ventilator patients: a descriptive study. *Am J Crit Care* 1997;**6**:99–105.

249. Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE – acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981;**9**:591–7.

250. Le Gall JR, Brun-Buisson C, Trunet P, Latournerie J, Chantereau S, Rapin M. Influence of age, previous health status, and severity of acute illness on outcome from intensive care. *Crit Care Med* 1982;**10**:575–7.

251. Bergner M, Bobbitt RA, Kressel S, Pollard WE, Gilson BS, Morris J. The Sickness Impact Profile: conceptual formulation and methodology for the development of a health status measure. *Int J Health Serv* 1976;**6**:393–415.

252. Bergner M, Bobbitt RA, Pollard WE, Martin DP, Gilson BS. The Sickness Impact Profile: validation of a health status measure. *Med Care* 1976;**14**:57–67.

253. Bergner M, Bobbitt RA, Carter WB, Gilson BS. The Sickness Impact Profile: development and final revision of a health status measure. *Med Care* 1981;**19**:787–805.

254. Patrick D, Peach I, editors. Disablement in the community. Oxford: Oxford University Press, 1989.

255. Deyo RA, Inui TS, Leininger JD, Overman SS. Measuring functional outcomes in chronic disease: a comparison of traditional scales and self administered questionnaire in patients with rheumatoid arthritis. *Med Care* 1983;**21**:180–92.

256. de Bruin AF, de Witte LP, Stevens F, Kiederiks JPM. Sickness Impact Profile: the state of the art of a generic functional status measure. *Soc Sci Med* 1992;**35**:1003–14.

257. Hall J, Hall N, Fisher E, Killer D. Measurement of outcomes of general practice: comparison of three health status measures. *Fam Pract* 1987;**4**:117–23.

258. Danis M, Patrick DL, Southerland LI, Green ML. Patients' and families' preferences for medical intensive care. *JAMA* 1988;**260**:797–802.

259. Hulsebos RG, Beltman FW, Reis Miranda D, Spangenberg JFA. Measuring quality of life with the Sickness Impact Profile: a pilot study. *Intensive Care Med* 1991;**17**:285–8.

260. Miranda DR. Quality of life after cardiopulmonary resuscitation. *Chest* 1994;**106**:524–30.

261. Tian ZM, Miranda DR. Quality of life after intensive care with the Sickness Impact Profile. *Intensive Care Med* 1995;**21**:422–8.

262. Grady KL, Jalowiec A, White-Williams C. Quality of life 6 months after heart transplantation compared with indicators of illness severity before transplantation. *Am J Crit Care* 1998;**7**:106–16.

263. Thiagarajan J, Taylor P, Hogbin E, Ridley S. Quality of life after multiple trauma requiring intensive care. *Anaesthesia* 1994;**49**:211–18.

264. Ridley SA, Biggam M, Stone P. A cost-utility analysis of intensive therapy: II. Quality of life in survivors. *Anaesthesia* 1994;**49**:192–6.

265. Hunt SM. Nottingham Health Profile. In: Wenger NK, Mattson ME, Furberg, editors. Assessment of quality of life in clinical trials of cardiovascular therapies. New York: Le Jacq; 1984.

266. Hunt SM, McEwan J, McKenna SP. Measuring health status. London: Croom Helm; 1986.

267. Hunt S. McKenna S. Nottingham Health Profile users' manual: revised edition. Manchester: Galen Research and Consultancy, 1991.

268. McKenna S, Hunt S, Tennant A. The development of a patient-completed index of distress from the Nottingham Health Profile: a new measure for use in cost utility studies. *Br J Med Econ* 1993;**6**:13–24.

269. Kind P. Carr-Hill R. The Nottingham Health Profile: a useful tool for epidemiologists? *Soc Sci Med* 1987;**25**:905–10.

270. Doll HA, Black NA, Flood AB, McPherson K. Patient-perceived health status before and up to 12 months after transurethral resection of the prostate for benign prostatic hyperthrophy. *Br J Urol* 1993;**71**:297–305.

271. Doll HA, Black NA, Flood AB, McPherson K. Criterion validation of the Nottingham Health Profile: a study of surgery for benign prostatic hypertrophy. *Soc Sci Med* 1993;**37**:115–22.

272. McKenna S, McEwan J, Hunt SM, Papp E. Changes in the perceived health of patients recovering from fractures. *Public Health* 1984;**98**:97–102.

273. O'Brien BJ. Assessment of treatment in heart disease. In: Teeling-Smith G, editor. Measuring health: a practical approach. Chichester: John Wiley, 1988.

274. Fitzpatrick R, Newman S, Lamb R, Shuday M. Social relationships and psychological well-being in rheumatoid arthritis. *Soc Sci Med* 1988;**27**:399–403.

275. Caine N, Harrison SCW, Sharples LD, Wallwork J. Prospective quality of life before and after coronary artery bypass graft. *BMJ* 1991;**302**:511–16.

276. Jenkinson C, Fitzpatrick R, Argyle M. The Nottingham Health Profile: an analysis of its sensitivity in differentiating illness groups. *Soc Sci Med* 1988;**27**:1411–14.

277. Shiell AM, Griffiths RD, Short AIK, Spiby J. An evaluation of the costs and outcome of adult intensive care in two units in the UK. *Clin Intensive Care* 1990;**1**:256–62.

278. Munn J, Willatts SM, Tooley MA. Health and activity after intensive care. *Anaesthesia* 1995;**50**:1017–21.

279. Gopal I, Bhonagiri S, Ronco C, Bellomo R. Out of hospital outcome and quality of life in survivors of combined acute multiple organ and renal failure treated with continuous venovenous hemofiltration/hemodiafiltration. *Intensive Care Med* 1997;**23**:766–72.

280. Nielsen D, Sellgren J, Ricksten SE. Quality of life after cardiac surgery complicated by multiple organ failure. *Crit Care Med* 1997;**25**:52–7.

281. Brazier J, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, *et al.* Validating the SF-36 health survey questionnaire: new outcome measures in primary care. *BMJ* 1992;**305**:160–4.

282. Wright SJ, Stein A, Boyle H, Moorhouse J, Walls J. Health-related quality of life in end stage renal failure. *Qual Life Res* 1994;**3**:63.

283. Jenkinson C, Coulter A, Wright L. Short Form 36 (SF36) health survey questionnaire: normative data for adults of working age. *BMJ* 1993;**306**:1437–40.

284. Jenkinson C, Wright L, Coulter A. Quality of life measurement in health care: a review of measures and population norms for the UK SF-36. Oxford: University of Oxford, Health Services Research Unit, Department of Public Health and Primary Care; 1993.

285. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Evaluation of the SF-36 version II in the United Kingdon. *J Epidemiol Community Health* 1999;**53**:46–50.

286. Chrispin PS, Scotton H, Rogers J, Lloyd D, Ridley SA. Short Form 36 in the intensive care unit: assessment of acceptability, reliability and validity of the questionnaire. *Anaesthesia* 1997;**52**:15–23.

287. Ridley SA, Chrispin PS, Scotton H, Rogers J, Lloyd D. Changes in quality of life after intensive care: comparison with normal data. *Anaesthesia* 1997;**52**:195–202.

288. Davidson TA, Caldwell ES, Curtis JR, Hudson LD, Steinberg KP. Reduced quality of life in survivors of acute respiratory distress syndrome compared with critically ill control patients. *JAMA* 1999;**281**:354–60.

289. Kind P, Rosser RM, Williams A. Valuation of quality of life: some psychometric evidence. In: Jones-Lee MW, editor. The value of life and safety. Amsterdam: North Holland Publishing; 1982. p.159–70.

290. Gater RA, Kind P, Gudex C. Quality of liaison psychiatry. A comparison of patient and clinician assessment. *Br J Psychiatry* 1995;**166**:515–20.

291. Kind P, Gudex CM. Measuring health status in the community: a comparison of methods. *J Epidemiol Community Health* 1994;**48**:86–91.

292. Kerridge RK, Glasziou PP, Hillman KM. The use of "quality adjusted life years" (QALYs) to evaluate treatment in intensive care. *Anaesth Intensive Care* 1995;**23**:322–31.

293. Kerridge R, Brooks R, Hillman K. Quality of life after intensive care. In: Vincent J, editor. Yearbook of intensive care and emergency medicine. Berlin: Springer-Verlag; 1994. p.827–38.

294. Spitzer WO, Dobson AJ, Hall J, Chesterman E, Levi J, Shepherd R, *et al.* Measuring quality of life of cancer patients: a concise QL-index for use by physicians. *J Chronic Dis* 1981;**34**:585–97.

295. Gough IR, Furnival CM, Schilder L, Grove W. Assessment of the quality of life of patients with advanced cancer. *Eur J Cancer Clin Oncol* 1983;**19**:1161–5.

296. Slevin ML, Plant H, Lynch D, *et al.* Who should measure quality of life, the doctor or the patient? *Br J Cancer* 1988;**57**:109–12.

297. Churchill DN, Torrance GW, Taylor DW, Barnes CC, Ludwin D, Shimizu A, *et al.* Measurement of quality of life in end stage renal disease: the time trade-off approach. *Clin Invest Med* 1987;**10**:14–20.

298. Frede KE, Lanter G. Surgical intensive care in advanced age. *Helvetica Chir Acta* 1990;**57**:903–7.

299. Konopad E, Noseworthy TW, Johnston R, Shustack A, Grace M. Quality of life measures before and one year after admission to an intensive care unit. *Crit Care Med* 1995;**23**:1653–9.

300. Dupuy HJ. Report of the national conference on evaluation in alcohol, drug abuse and mental health programs: utility of the National Center for Health Statistics General Well-Being Schedule in the assessment of self representations of subjective well-being and distress. Washington DC: ADA MHA; 1974.

301. Dupuy HJ. Self representations of general psychological well-being of American adults. In: Proceedings of the American Public Health Association meeting, Los Angeles, October 1978.

302. Fazio AF. A concurrent validation study of the NCHS General Well-being Schedule. *Vital Health Stat 2* 1977;(73):1–53.

303. Becker N, Bondegaard Thomsen A, Olsen AK, Sjogren P, Bech P, Eriksen J. Pain epidemiology and health related quality of life in chronic non-malignant pain patients referred to a Danish multidisciplinary pain centre. *Pain* 1997;**73**:393–400.

304. Monk M. Blood pressure awareness and psychological well-being in the health and nutrition examination survey. *Clin Invest Med* 1981;**4**:183–9.

305. Edwards D, Yarvis R, Mueller D, Zingale H, Wagman W. Test taking and the stability of adjustment scales: can we assess patient deterioration? *Eval Q* 1978;**2**:275–91.

306. Vasquez Mata G, Riviera Fernandez R, Perez Aragon A, Gonzalez Carmona A, Fernandez Mondejar E, Navarrete Navarro P. Analysis of quality of life in polytraumatised patients two years after discharge from an intensive care unit. *J Trauma* 1996;**41**:326–32.

307. Griffths RD, Jones C, Palmer TEA. Six month outcome of critically ill patients given glutamine supplemented parenteral nutrition. *Nutrition* 1997;**13**:295–302.

308. Ciaglia P, Graniero KD. Percutaneous dilatational tracheostomy: results and long term follow-up. *Chest* 1992;**2**:464–67.

309. Santini F, Dyke C, Edwards S, Stavri G, Feccia M, Khan H, *et al.* Pulmonary autograft versus homograft replacement of the aortic valve: a prospective randomised trial. *J Thorac Cardiovasc Surg* 1997;**113**:894–900.

310. Potgieter PD, Rosenthal E, Benatar SR. Immediate and long-term survival in patients admitted to a respiratory ICU. *Crit Care Med* 1985;**13**:798–802.

311. Kaukinen L. Results of intensive care in abdominal surgery patients. *Ann Chir Gynaecol* 1982;**71**:168–71.

312. Spicher JE, White DP. Outcome and function following prolonged mechanical ventilation. *Arch Intern Med* 1987;**147**:421–5.

313. Searle JF. The outcome of mechanical ventilation: report of a five-year study. *Ann R Coll Surg Engl* 1985;**67**:187–9.

314. Hill AT, Hopkinson RB, Stableforth DE. Ventilation in a Birmingham intensive care unit 1993–1995: outcome for patients with chronic obstructive pulmonary disease. *Respir Med* 1998;**92**:156–61.

315. Thoner J. Outcome and costs of intensive care. A follow-up study on patients requiring prolonged mechanical ventilation. *Acta Anaesthesiol Scand* 1987;**31**:693–8.

316. Bams JL, Miranda DR. Outcome and costs of intensive care. *Intensive Care Med* 1985;**11**:234–41.

317. Løes Ø, Smith-Erichsen N, Lind B. Intensive care: cost and benefit. *Acta Anaesthesiol Scand* 1987;**31** Suppl 84:3–19.

318. Chassin MR. Costs and outcomes of medical intensive care. *Med Care* 1982;**20**:165–79.

319. Bingham HG, Hudson D, Popp J. A retrospective review of the burn intensive care unit admissions for a year. *J Burn Care Rehabil* 1995;**16**:56–8.

320. Martens PB, Ho G Jr. Septic arthritis in adults: clinical features, outcome, and intensive care requirements. *J Intensive Care Med* 1995;**10**:246–52.

321. McLauchlan GJ, Anderson ID, Grant IS, Fearon KC. Outcome of patients with abdominal sepsis treated in an intensive care unit. *Br J Surg* 1995;**82**:524–9.

322. Marquette CH, Saulnier F, Leroy O, Wallaert B, Chopin C, Demarcq JM, *et al.* Long term prognosis of near-fatal asthma: a 6-year follow-up study of 145 asthmatic patients who underwent mechanical ventilation for a near-fatal attack of asthma. *Am Rev Respir Dis* 1992;**146**:76–81.

323. Holdgaard HO, Pedersen J, Schurizek BA, Melsen NC, Juhl B. Complications and late sequelae following nasotracheal intubation. *Acta Anaesthesiol Scand* 1993;**37**:475–80.

324. Vale FL, Burns J, Jackson AB, Hadley MN. Combined medical and surgical treatment after acute spinal cord injury: results of a prospective pilot study to assess the merits of aggressive medical resuscitation and blood pressure management. *J Neurosurg* 1997;**87**:239–46.

325. Ng KKP, Howard RS, Fish DR, Hirsch NP, Wiles CM, Murray NMF, *et al.* Management and outcome of severe Guillain-Barré syndrome. *Q J Med* 1995;**88**:243–50.

326. Benzer H, Mutz N, Pauser G. Psychosocial sequelae of intensive care. *Int Anesthesiol Clin* 1983;**21**:169–78.

327. Schnaper N. The psychological implications of severe trauma: emotional sequelae to unconsciousness. A preliminary study. *J Trauma* 1975;**15**:94–8.

328. Asbury AJ. Patients' memories and reactions to intensive care. *Care Crit Ill* 1985;**1**(2):12–13.

329. Havill JH, Walker L, Sceats JE. Three hundred admissions to the Waikato hospital intensive therapy unit – survival, costs, and quality of life after two years. *N Z Med J* 1989;**102**:179–81.

330. Gefke K, Schroeder TV, Thisted B, Olsen PS, Perko MJ, Agerskov K, *et al.* Abdominal aortic aneurysm surgery: survival and quality of life in patients requiring prolonged postoperative intensive therapy. *Ann Vasc Surg* 1994;**8**:137–43.

331. Jacobs CJ, van der Vliet JA, van Roozendaal MT, van der Linden CJ. Mortality and quality of life after intensive care for critical illness. *Intensive Care Med* 1988;**14**:217–20.

332. Winckler C, Oksenhendler G, Coulombe G, Hubscher C, Lecointre F, Petit J. Complex consequences of abdomino-pelvic abdominal surgery: mid term survival (1 year) – apropos of 236 cases. *Anesth Analg Reanim* 1978;**35**:1175–84.

333. Horn J, Geier HJ, Seefried G, Platt D. Intensive care medicine in geriatric patients as seen by the internist. *Fortschr Med* 1992;**110**(13):33–6.

334. Fakhry SM, Kercher KW, Rutledge R. Survival, quality of life, and charges in critically ill surgical patients requiring prolonged ICU stays. *J Trauma* 1996;**41**:999–1007.

335. Hervé C, Gaillard M, Martel S, Huguenard P. Serious suicides: short and long-term results. One hundred and fifty six cases with a follow-up period of 1 to 7 years. *Acta Anaesthesiol Belg* 1984; **35** Suppl:353–9.

336. Miranda AF, Miranda S. Quality of life and long-term survival after intensive care discharge. *Med J Malaysia* 1991;**46**:66–71.

337. Treiman RL, Levine KA, Cohen JL, Cossman DV, Foran RF, Levin PM. Aneursymectomy in the octogenarian. A study of morbidity and quality of survival. *Am J Surg* 1982;**144**:194–7.

338. Rustom R, Daly K. Quality of life after intensive care. *Br J Nurs* 1993;**2**:316–20.

339. Söderlind K, Rutberg H, Olin C. Late outcome and quality of life after complicated heart operations. *Ann Thorac Surg* 1997;**63**:124–8.

340. Maurette P, Dabadie P, Dartigue JF, Mignonsin D, Erny P. Results of intensive care in elderly patients. *Ann Fr Anesth Reanim* 1987;**6**:95–9.

341. Rattray J, Johnston M, Wildsmith JAW. Assessment of individual quality of life in survivors of intensive care. *Nurs Crit Care* 1998;**3**:220–6.

342. Daffurn K, Bishop GF, Hillman KM, Bauman A. Problems following discharge after intensive care. *Intensive Crit Care Nurs* 1994;**10**:244–51.

343. Schilling G, Scheer JW, Laubach W, Klapp BF. Psychopathology, coping and defense in intensive-care patients. *Forschr Neurol Psychiatr* 1994;**62**:233–40.

344. Friedman BC, Boyce W, Bekes CE. Long-term follow-up of ICU patients. *Am J Crit Care* 1992;**1**:115–17.

345. Pauser G, Benzer H, Bunzel B, Kubinger KD, Riegler R. Intensive medicine and what follows? Psycho-social follow-up study of former critical care patients. *Anaesthesist* 1984;**33**:189–95.

346. Rohrer MJ, Cutler BS, Wheeler HB. Long-term survival and quality of life following ruptured abdominal aortic aneursym. *Arch Surg* 1988;**123**:1213–17.

Appendix 1

Outcome measures with only one reported use

The papers listed below provided data on 125 different outcome measures for which there was only one reported use. The review was confined to measures that had been used at least twice since results from a single study could not be synthesised.

Measures of impairment

Respiratory

Anterior-posterior linear air tomograms	Stauffer, *et al.*, 1981[44]
Chest expansion	Landercasper, *et al.*, 1984[37]
Expired ventilation	Landercasper, *et al.*, 1984[37]
Full blood count	Grotz, *et al.*, 1997[25]
Lung resistance	Grotz, *et al.*, 1997[25]
Maximum vital capacity at 40 breaths/minute	Friman, *et al.*, 1976[15]
Maximal voluntary ventilation	Landercasper, *et al.*, 1984[37]
Physical/respiratory capacity	Friman, *et al.*, 1976[15]
Pulmonary function tests (non-specific)	Waldmann & Gaine, 1996[129]
	Weir & Waldmann, 1994[128]
	Ciaglia & Graniero, 1992[308]
Radiographs	Friman, *et al.*, 1976[15]
Xeroradiograms	Lund, *et al.*, 1985[16]

Cardiac

Chest X-ray	Santini, *et al.*, 1997[309]
Coronary artery disease	Martinelli, *et al.*, 1995[59]
ECG	Santini, *et al.*, 1997[309]
Doppler ECG	Santini, *et al.*, 1997[309]
Hypertension	Martinelli, *et al.*, 1995[59]
Left ventricular ejection fraction	Grady, *et al.*, 1998[262]
Right and left heart haemodynamics	Martinelli, *et al.*, 1995[59]

Haematological

Erythrocyte sedimentation rate	Nordback & Auvinen, 1985[60]
Glycohaemoglobin A1 and A1C	Doepel, *et al.*, 1993[61]
Haemoglobin	Doepel, *et al.*, 1993[61] Nordback & Auvinen, 1985[60]
Leucocyte levels	Doepel, *et al.*, 1993[61] Nordback & Auvinen, 1985[60]
Thrombocyte levels	Grotz, *et al.*, 1997[25]

Measures of functional status

Generic

Ability to use stairs	Bergner, *et al.*, 1984[222]
Category test	McKee, *et al.*, 1997[174]
Degree of recovery	Potgieter, *et al.*, 1985[310] Kaukinen, 1982[311] Spicher & White, 1987[312]
Degree of disability	Alho & Rokkanen, 1973[220] Searle, 1985[313] Kivioja, *et al.*, 1990[232] Gobiet, 1995[101]
Dependence	Hill, *et al.*, 1998[314] Ritz, 1988[229]
Exercise tolerance	Hill, *et al.*, 1998[314]
Functional independence measure	Grotz, *et al.*, 1997[25]
Functional limitation scale	Le Gall, *et al.*, 1982[250]
Activity level	Kriwanek, *et al.*, 1998[239] MacKenzie, *et al.*, 1988[228] Morris, *et al.*, 1991[233]

Activities of daily living	Mahul, *et al.,* 1991[245] Parno *et al.,* 1984[223] Munn, *et al.,* 1995[278]
Functional capacity	Hill, *et al.,* 1998[314] Bell & Turpin, 1994[218] Schuster, 1991[107] McLean, *et al.,* 1985[244]
Degree of limitations	Rowan, 1992[155] Bosatra, *et al.,* 1987[226] Thoner, 1987[315]
Health state	Weinert, *et al.,* 1997[91]
Health status	Bams & Miranda, 1985[316]
Instrumental activities of daily living (modified)	Battistella, *et al.,* 1998[76]
Philadelphia Geriatric Center activities of daily living	Bedell, *et al.,* 1983[150]
Job description inventory	Stanton, *et al.,* 1983[137]
Katz's adjustment scale – relatives' form	Stambrook, *et al.,* 1990[138]
Locomotion score	Grotz, *et al.,* 1997[25]
Mental activity	Bürgisser & Ritz, 1982[221]
Overall performance categories	Løes, *et al.,* 1987[317]
Quality of well-being scale	Holbrook, *et al.,* 1994[71]
RAND — physical limitations scale	Chassin, 1982[318]
Range of motion	Bingham, *et al.,* 1995[319]
Response to exercise	Ciaglia & Graniero, 1992[308]
Return of joint function	Martens & Ho, 1995[320]
Self-reported global health assessment	Brenneman, *et al.,* 1997[236]
Sexual activity	Yinnon, *et al.,* 1989[88]
Sleep index	Yinnon, *et al.,* 1989[88]
State trait anxiety inventory	Riether, *et al.,* 1992[165]
Strength	Bingham, *et al.,* 1995[319]
Subjective fitness	Kivioja, *et al.,* 1990[232]
Symptoms	Gaillard, *et al.,* 1990[231,240]
Tegner activity score	Grotz, *et al.,* 1997[25]
WHO performance score	McLauchlan, *et al.,* 1995[321]

Neuropsychological tests

Psychometric tests (not stated)	Jones C, *et al.,* 1994[140]
Auditory verbal learning	Lannoo, *et al.,* 1998[179]
Binary choice	Lannoo, *et al.,* 1998[179]
California verbal learning test (list A)	Riether, *et al.,* 1992[165]
Complex figure	Lannoo, *et al.,* 1998[179]
Controlled oral word association	Lannoo, *et al.,* 1998[179]
Corsi's test for spatial memory	Alexandre, *et al.,* 1983[188]
Dot cancellation	Lannoo, *et al.,* 1998[179]
Frenchay aphasia screen test	Anderson, *et al.,* 1994[178]
Kimuras test for spatial memory	Alexandre, *et al.,* 1983[188]
Logical memory subtest of Wechlser memory scale	Anderson, *et al.,* 1994[178]
National adult reading test	Anderson, *et al.,* 1994[178]
Reitan–Indiana aphasia screening test	Uzzell, *et al.,* 1986[183]
Rey Osterrieth complex figure	Anderson, *et al.,* 1994[178]
Rey's test for verbal learning	Alexandre, *et al.,* 1983[188]
Russell's revised Wechsler memory scale	Uzzell, *et al.,* 1986[183]
Stroop interference	Lannoo, *et al.,* 1998[179]
Temporal orientation test	Anderson, *et al.,* 1994[178]
Token test	Alexandre, *et al.,* 1983[188]
Visual reaction time	Lannoo, *et al.,* 1998[179]
WAIS	Uzzell, *et al.,* 1986[183]
WAIS digit forward and backwards	Lannoo, *et al.,* 1998[179]

Wechsler Bellevue form 1	Alexandre, *et al.,* 1983[188]
Wechsler memory form	Alexandre, *et al.,* 1983[188]

Respiratory

Symptoms	Sellery, *et al.,* 1978[55] Jones, *et al.,* 1997[22]
British medical research gradation of dyspnoea	Landercasper, *et al.,* 1984[37]
Chronic respiratory questionnaire	Weinert, *et al.,* 1997[91]
Dyspnoea	Halevy, *et al.,* 1984[24] Peters, *et al.,* 1989[20]
Questionnaire for asthma	Marquette, *et al.,* 1992[322]
Pain, dyspnoea, satisfaction with scar, voice changes, difficulty swallowing	Law, *et al.,* 1997[23] Ciaglia & Graniero, 1992[308] Walz, *et al.,* 1998[58] Hill, *et al.,* 1998[314]
Symptoms from nose, ears, larynx, trachea	Holdgaard, *et al.,* 1993[323]
Voice, breathing	Winkler, *et al.,* 1994[56]

Neurological

American Spinal Injury Association impairment Scale of spinal cord injuries	Vale, *et al.,* 1997[324]
Disability grading in Guillan–Barre syndrome	Ng, *et al.,* 1995[325]

Cardiac

Cardiovascular symptoms	Kumar, *et al.,* 1995[89]
Congestive cardiac functional class	Kumar, *et al.,* 1995[89]
Heart transplant symptom checklist	Grady, *et al.,* 1998[262]

Other

Abdominal/ other symptoms	Doepel, *et al.,* 1993[61]
Bowel function	Nordback & Auvinen, 1985[60]

Measures of quality of life

Attitude to ICU (necessity)	Benzer, *et al.,* 1983[326]
Brief symptom inventory	Landsman, *et al.,* 1990[169]
Coping scale	Singh, *et al.,* 1997[90]
Emotional sequelae	Schnaper, 1975[327]
Family environment scale	Landsman, *et al.,* 1990[169]
Impression of ICU stay	Chelluri, *et al.,* 1992[247]
ICU quality-of-life questionnaire	Eddleston, *et al.;* personal communication, 1999
Index of well-being	Weinert, *et al.,* 1997[91]
Perception of ICU stay	Chelluri, *et al.,* 1993[70]
Health satisfaction and beliefs	Rockwood, *et al.,* 1993[69]
Heart transplant stressor scale	Grady, *et al.,* 1998[262]
Impact message inventory	Riether, *et al.,* 1992[165]
Jalowiec coping scale	Grady, *et al.,* 1998[262]
Linear analogue self-assessment	Yinnon, *et al.,* 1989[88]
Memories, dreams	Asbury, 1985[328]
Personal adjustment and role skills	Landsman, *et al.,* 1990[169]
Psychological sequelae	Kerridge, *et al.,* 1995[292]
Quality of life (non-specific)	Hill, *et al.,* 1998[314] Chelluri, *et al.,* 1992[247] Wu, *et al.,* 1995[73] McLean, *et al.,* 1985[244] Havill, *et al.,* 1989[329] Gefke, *et al.,* 1994[330] Jacobs, *et al.,* 1988[331] Kerridge, *et al.,* 1995[292] Kumar, *et al.,* 1995[89] Winckler, *et al.,* 1978[332] Gaillard, *et al.,* 1990[231,240] Horn, *et al.,* 1992[333] Ritz, 1988[229] Fakhry, *et al.,* 1996[334] Herve, *et al.,* 1984[335] Miranda & Miranda, 1991[336] Treiman, *et al.,* 1982[337] Rustom & Daly, 1993[338]

	Söderlind, *et al.,* 1997[339] Maurette, *et al.,* 1987[340] Singh, *et al.,* 1997[90] Rattray, *et al.,* 1998[341]
Present health, employment, functional state recollection of ICU, sequelae	Daffurn, *et al.,* 1994[342]
Psychosocial responses	Schilling, *et al.,* 1994[343]
Recollection, relationships, sex, sleeping, eating, work, use of medical resources	Friedman, *et al.,* 1992[344]
Psychological, economic and social factors	Pauser, *et al.,* 1984[345]
Rating question form	Grady, *et al.,* 1998[262]
Self evaluation of life functional scale	Rohrer, *et al.,* 1988[346]
Social support	Holbrook, *et al.,* 1994[71] Kumar, *et al.,* 1995[89]
Willingness to undergo same surgery again	McHugh, *et al.,* 1997[118]
Willingness to undergo cardio-pulmonary resuscitation again	Bedell, *et al.,* 1983[150]
Irritability, medication	Bergner, *et al.,* 1984[222]

Health Technology Assessment panel membership

This report was identified as a priority by the Acute Sector Panel.

Acute Sector Panel

Current members

Chair:
Professor Francis H Creed
University of Manchester

Professor Clifford Bailey
University of Leeds

Ms Tracy Bury
Chartered Society of Physiotherapy

Professor Collette Clifford
University of Birmingham

Dr Katherine Darton
M.I.N.D.

Mr John Dunning
Papworth Hospital, Cambridge

Mr Jonathan Earnshaw
Gloucester Royal Hospital

Mr Leonard Fenwick
Freeman Group of Hospitals, Newcastle-upon-Tyne

Professor David Field
Leicester Royal Infirmary

Ms Grace Gibbs
West Middlesex University Hospital NHS Trust

Dr Neville Goodman
Southmead Hospital Services Trust, Bristol

Professor Mark Haggard
MRC Institute of Hearing Research, University of Nottingham

Professor Robert Hawkins
University of Manchester

Dr Duncan Keeley
General Practitioner, Thame

Dr Rajan Madhok
East Riding Health Authority

Dr John Pounsford
Frenchay Hospital, Bristol

Dr Mark Sculpher
University of York

Dr Iqbal Sram
NHS Executive, North West Region

Mrs Joan Webster
Consumer member

Past members

Professor John Farndon*
University of Bristol

Professor Senga Bond
University of Newcastle-upon-Tyne

Professor Ian Cameron
Southeast Thames Regional Health Authority

Ms Lynne Clemence
Mid-Kent Health Care Trust

Professor Cam Donaldson
University of Aberdeen

Professor Richard Ellis
St James's University Hospital, Leeds

Mr Ian Hammond
Bedford & Shires Health & Care NHS Trust

Professor Adrian Harris
Churchill Hospital, Oxford

Dr Gwyneth Lewis
Department of Health

Mrs Wilma MacPherson
St Thomas's & Guy's Hospitals, London

Dr Chris McCall
General Practitioner, Dorset

Professor Alan McGregor
St Thomas's Hospital, London

Professor Jon Nicholl
University of Sheffield

Professor John Norman
University of Southampton

Professor Michael Sheppard
Queen Elizabeth Hospital, Birmingham

Professor Gordon Stirrat
St Michael's Hospital, Bristol

Dr William Tarnow-Mordi
University of Dundee

Professor Kenneth Taylor
Hammersmith Hospital, London

continued

* Previous Chair

continued

Diagnostics and Imaging Panel

Current members

Chair:
Professor Mike Smith
University of Leeds

Dr Philip J Ayres
Leeds Teaching Hospitals NHS Trust

Dr Paul Collinson
St George's Hospital, London

Dr Barry Cookson
Public Health Laboratory Service, Colindale

Professor David C Cumberland
University of Sheffield

Professor Adrian Dixon
University of Cambridge

Mr Steve Ebdon-Jackson
Department of Health

Mrs Maggie Fitchett
Association of Cytogeneticists, Oxford

Dr Peter Howlett
Portsmouth Hospitals NHS Trust

Professor Alistair McGuire
City University, London

Dr Andrew Moore
Editor, *Bandolier*

Dr Peter Moore
Science Writer, Ashtead

Professor Chris Price
London Hospital Medical School

Dr William Rosenberg
University of Southampton

Mr Tony Tester
South Bedfordshire Community Health Council

Dr Gillian Vivian
Royal Cornwall Hospitals Trust

Dr Greg Warner
General Practitioner, Hampshire

Past members

Professor Michael Maisey*
Guy's & St Thomas's Hospitals, London

Professor Andrew Adam
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Dr Pat Cooke
RDRD, Trent Regional Health Authority

Ms Julia Davison
St Bartholomew's Hospital, London

Professor MA Ferguson-Smith
University of Cambridge

Dr Mansel Haeney
University of Manchester

Professor Sean Hilton
St George's Hospital Medical School, London

Mr John Hutton
MEDTAP International Inc., London

Professor Donald Jeffries
St Bartholomew's Hospital, London

Dr Ian Reynolds
Nottingham Health Authority

Professor Colin Roberts
University of Wales College of Medicine

Miss Annette Sergeant
Chase Farm Hospital, Enfield

Professor John Stuart
University of Birmingham

Dr Ala Szczepura
University of Warwick

Mr Stephen Thornton
Cambridge & Huntingdon Health Commission

Dr Jo Walsworth-Bell
South Staffordshire Health Authority

Methodology Group

Current members

Chair:
Professor Martin Buxton
Health Economics Research Group, Brunel University

Professor Doug Altman
ICRF/NHS Centre for Statistics in Medicine, University of Oxford

Dr David Armstrong
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Professor Nicholas Black
London School of Hygiene & Tropical Medicine

Professor Ann Bowling
University College London Medical School

Dr Mike Clarke
UK Cochrane Centre, Oxford

Professor Paul Dieppe
MRC Health Services Research Collaboration, University of Bristol

Professor Mike Drummond
Centre for Health Economics, University of York

Dr Vikki Entwistle
University of Aberdeen

Professor Ewan Ferlie
Imperial College, London

Professor Ray Fitzpatrick
University of Oxford

Mrs Jenny Griffin
Department of Health

Professor Jeremy Grimshaw
University of Aberdeen

Dr Stephen Harrison
University of Leeds

Mr John Henderson
Department of Health

Professor Richard Lilford
R&D, West Midlands

Professor Theresa Marteau
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Dr Henry McQuay
University of Oxford

Dr Nick Payne
University of Sheffield

Professor Maggie Pearson
NHS Executive North West

Dr David Spiegelhalter
Institute of Public Health, Cambridge

Professor Joy Townsend
University of Hertfordshire

Ms Caroline Woodroffe
Standing Group on Consumers in NHS Research

Past members

Professor Anthony Culyer*
University of York

Professor Michael Baum
Royal Marsden Hospital

Dr Rory Collins
University of Oxford

Professor George Davey Smith
University of Bristol

Professor Stephen Frankel
University of Bristol

Mr Philip Hewitson
Leeds FHSA

Mr Nick Mays
King's Fund, London

Professor Ian Russell
University of York

Professor David Sackett
Centre for Evidence Based Medicine, Oxford

Dr Peter Sandercock
University of Edinburgh

Dr Maurice Slevin
St Bartholomew's Hospital, London

Professor Charles Warlow
Western General Hospital, Edinburgh

* Previous Chair

Pharmaceutical Panel

Current members

Chair:
Professor Tom Walley
University of Liverpool

Dr Felicity Gabbay
Transcrip Ltd

Dr Peter Golightly
Drug Information Services,
NHS Executive Trent

Dr Alastair Gray
Health Economics
Research Centre,
University of Oxford

Professor Rod Griffiths
NHS Executive
West Midlands

Mrs Jeanette Howe
Department of Health

Professor Trevor Jones
ABPI, London

Ms Sally Knight
Lister Hospital, Stevenage

Dr Andrew Mortimore
Southampton & SW Hants
Health Authority

Mr Nigel Offen
NHS Executive Eastern

Dr John Reynolds
The Oxford Radcliffe Hospital

Mrs Marianne Rigge
The College of Health,
London

Mr Simon Robbins
Camden & Islington
Health Authority, London

Dr Frances Rotblat
Medicines Control Agency

Dr Eamonn Sheridan
St James's University Hospital,
Leeds

Mrs Katrina Simister
National Prescribing Centre,
Liverpool

Dr Ross Taylor
University of Aberdeen

Past members

Professor Michael Rawlins*
University of Newcastle-
upon-Tyne

Dr Colin Bradley
University of Birmingham

Professor Alasdair
Breckenridge
RDRD, Northwest Regional
Health Authority

Ms Christine Clark
Hope Hospital, Salford

Mrs Julie Dent
Ealing, Hammersmith &
Hounslow Health Authority,
London

Mr Barrie Dowdeswell
Royal Victoria Infirmary,
Newcastle-upon-Tyne

Dr Tim Elliott
Department of Health

Dr Desmond Fitzgerald
Mere, Bucklow Hill, Cheshire

Professor Keith Gull
University of Manchester

Dr Keith Jones
Medicines Control Agency

Dr John Posnett
University of York

Dr Tim van Zwanenberg
Northern Regional
Health Authority

Dr Kent Woods
RDRD, Trent RO,
Sheffield

Population Screening Panel

Current members

Chair:
Professor Sir John
Grimley Evans
Radcliffe Infirmary, Oxford

Mrs Stella Burnside
Altnagelvin Hospitals Trust,
Londonderry

Mr John Cairns
University of Aberdeen

Professor Howard Cuckle
University of Leeds

Dr Carol Dezateux
Institute of Child Health,
London

Mrs Anne Dixon-Brown
NHS Executive Eastern

Professor Dian Donnai
St Mary's Hospital,
Manchester

Dr Tom Fahey
University of Bristol

Mrs Gillian Fletcher
National Childbirth Trust

Dr JA Muir Gray
National Screening
Committee, NHS Executive
Oxford

Professor Alexander Markham
St James's University Hospital,
Leeds

Dr Ann McPherson
General Practitioner,
Oxford

Dr Susan Moss
Institute of Cancer Research

Mr John Nettleton
Consumer member

Mrs Julietta Patnick
NHS Cervical
Screening Programme,
Sheffield

Dr Sarah Stewart-Brown
Health Service Research Unit,
University of Oxford

Past members

Dr Sheila Adam*
Department of Health

Professor George Freeman
Charing Cross & Westminster
Medical School, London

Dr Mike Gill
Brent & Harrow
Health Authority

Dr Anne Ludbrook
University of Aberdeen

Professor Theresa Marteau
Guy's, King's &
St Thomas's School of
Medicine & Dentistry,
London

Professor Catherine Peckham
Institute of Child Health,
London

Dr Connie Smith
Parkside NHS Trust,
London

Ms Polly Toynbee
Journalist

Professor Nick Wald
University of London

Professor Ciaran Woodman
Centre for
Cancer Epidemiology,
Manchester

* Previous Chair

continued

continued

Primary and Community Care Panel

Current members

Chair:
Dr John Tripp
Royal Devon & Exeter Healthcare NHS Trust

Mr Kevin Barton
East London & City Health Authority

Professor John Bond
University of Newcastle-upon-Tyne

Dr John Brazier
University of Sheffield

Ms Judith Brodie
Cancer BACUP

Mr Shaun Brogan
Ridgeway Primary Care Group, Aylesbury

Mr Joe Corkill
National Association for Patient Participation

Dr Nicky Cullum
University of York

Professor Pam Enderby
University of Sheffield

Dr Andrew Farmer
Institute of Health Sciences, Oxford

Dr Jim Ford
Department of Health

Professor Richard Hobbs
University of Birmingham

Professor Allen Hutchinson
University of Sheffield

Dr Aidan MacFarlane
Independent Consultant

Professor David Mant
Institute of Health Sciences, Oxford

Dr Chris McCall
General Practitioner, Dorset

Dr Robert Peveler
University of Southampton

Professor Jennie Popay
University of Salford

Dr Ken Stein
North & East Devon Health Authority

Past members

Professor Angela Coulter*
King's Fund, London

Professor Martin Roland*
University of Manchester

Dr Simon Allison
University of Nottingham

Professor Shah Ebrahim
Royal Free Hospital, London

Ms Cathy Gritzner
King's Fund, London

Professor Andrew Haines
RDRD, North Thames Regional Health Authority

Dr Nicholas Hicks
Oxfordshire Health Authority

Mr Edward Jones
Rochdale FHSA

Professor Roger Jones
Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Mr Lionel Joyce
Chief Executive, Newcastle City Health NHS Trust

Professor Martin Knapp
London School of Economics & Political Science

Dr Phillip Leech
Department of Health

Professor Karen Luker
University of Liverpool

Dr Fiona Moss
Thames Postgraduate Medical & Dental Education

Professor Dianne Newham
King's College London

Professor Gillian Parker
University of Leicester

Dr Mary Renfrew
University of Oxford

Ms Hilary Scott
Tower Hamlets Healthcare NHS Trust, London

National Coordinating Centre for Health Technology Assessment, Advisory Group

Current members

Chair:
Professor John Gabbay
Wessex Institute for Health Research & Development

Dr Sheila Adam
Department of Health

Professor Nicholas Black
London School of Hygiene and Tropical Medicine

Professor Martin Buxton
Health Economics Research Group, Brunel University

Mr Harry Cayton
Alzheimer's Disease Society

Professor Angela Coulter
The King's Fund, London

Professor Paul Dieppe
MRC Health Services Research Collaboration, University of Bristol

Professor Mike Drummond
Centre for Health Economics, University of York

Professor Shah Ebrahim
MRC Health Services Research Collaboration, University of Bristol

Ms Lynn Kerridge
Wessex Institute for Health Research & Development

Professor Jos Kleijnen
NHS Centre for Reviews and Dissemination, University of York

Dr Ruairidh Milne
Wessex Institute for Health Research & Development

Ms Kay Pattison
Research & Development Directorate, NHS Executive

Professor James Raftery
Health Economics Unit, University of Birmingham

Professor Ian Russell
Department of Health Sciences & Clinical Evaluation, University of York

Dr Ken Stein
North & East Devon Health Authority

Professor Andrew Stevens
Department of Public Health & Epidemiology, University of Birmingham

Professor Kent Woods
Department of Medicine & Therapeutics, University of Leicester

Past member

Dr Paul Roderick
Wessex Institute for Health Research & Development

HTA Commissioning Board

Current members

Chair:
Professor Shah Ebrahim
Professor of Epidemiology of Ageing, University of Bristol

Professor Doug Altman
Director, ICRF Medical Statistics Group, Centre for Statistics in Medicine, University of Oxford

Professor John Bond
Director, Centre for Health Services Research, University of Newcastle-upon-Tyne

Mr Peter Bower
General Manager and Independent Health Advisor, Thames Valley Primary Care Agency

Ms Christine Clark
Honorary Research Pharmacist, Hope Hospital, Salford

Professor Martin Eccles
Professor of Clinical Effectiveness, University of Newcastle-upon-Tyne

Dr Mike Gill
Regional Director of Public Health, NHS Executive South East

Dr Alastair Gray
Director, Health Economics Research Centre, University of Oxford

Professor Mark Haggard
Director, MRC Institute of Hearing Research, University of Nottingham

Dr Jenny Hewison
Senior Lecturer, Department of Psychology, University of Leeds

Professor Alison Kitson
Director, Royal College of Nursing Institute

Dr Donna Lamping
Senior Lecturer, Department of Public Health, London School of Hygiene & Tropical Medicine

Professor Alan Maynard
Joint Director, York Health Policy Group, University of York

Professor David Neal
Joint Director, York Health Policy Group, University of York

Professor Jon Nicholl
Director, Medical Care Research Unit, University of Sheffield

Professor Gillian Parker
Nuffield Professor of Community Care, University of Leicester

Dr Tim Peters
Reader in Medical Statistics, Department of Social Medicine, University of Bristol

Professor Martin Severs
Professor in Elderly Health Care, University of Portsmouth

Dr Sarah Stewart-Brown
Health Service Research Unit, University of Oxford

Professor Ala Szczepura
Director, Centre for Health Services Studies, University of Warwick

Dr Gillian Vivian
Consultant, Royal Cornwall Hospitals Trust

Professor Graham Watt
Department of General Practice, University of Glasgow

Professor Kent Woods
Professor of Therapeutics, University of Leicester

Dr Jeremy Wyatt
Senior Fellow, Health Knowledge Management Centre, University College London

Past members

Professor Ian Russell*
Department of Health Sciences & Clinical Evaluation, University of York

Professor Charles Florey*
Department of Epidemiology & Public Health, Ninewells Hospital & Medical School, University of Dundee

Professor David Cohen
Professor of Health Economics, University of Glamorgan

Mr Barrie Dowdeswell
Chief Executive, Royal Victoria Infirmary, Newcastle-upon-Tyne

Dr Michael Horlington
Head of Corporate Licensing, Smith & Nephew Group Research Centre

Professor Sir Miles Irving
Professor of Surgery, University of Manchester, Hope Hospital, Salford

Professor Martin Knapp
Director, Personal Social Services Research Unit, London School of Economics & Political Science

Professor Theresa Marteau
Director, Psychology & Genetics Research Group, Guy's, King's & St Thomas's School of Medicine & Dentistry, London

Professor Sally McIntyre
MRC Medical Sociology Unit, Glasgow

Professor David Sackett
Centre for Evidence Based Medicine, Oxford

Dr David Spiegelhalter
MRC Biostatistics Unit, Institute of Public Health, Cambridge

Professor David Williams
Department of Clinical Engineering, University of Liverpool

Dr Mark Williams
Public Health Physician, Bristol

* Previous Chair

Feedback

The HTA programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

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