A REVIEW OF
DISEASE-SPECIFIC
QUALITY OF LIFE
MEASUREMENT
SCALEs

Second Edition

Ann Bowling

Open University Press
Buckingham · Philadelphia
## CONTENTS

PREFACE xiii
ACKNOWLEDGEMENTS xvi
LIST OF ABBREVIATIONS xvii

### 1 HEALTH-RELATED QUALITY OF LIFE: CONCEPTUAL MEANING, USE AND MEASUREMENT

Background: the ‘quality of life’ 1
What is quality of life and health-related quality of life? 2
Health-related quality of life and health 7
Quality of life assessment and research on health care outcomes 9
Utility assessments and quality of life assessment 11
Who should rate quality of life? 13
Patients’ preferences 14
Disease-specific, domain-specific and generic measurement scales 16
Methodological issues 19

### 2 CANCERS

The importance of measuring the quality of life of cancer patients 23
What has been measured in studies of the quality of life of cancer patients? 23
What should be measured? 24
Recommended measurement scales 24
Review articles on quality of life measurement and cancer 25
The domains of measurement 25
The measurement scales 30
Scales measuring physical functioning, pain and symptoms
- Functional status scales used with cancer patients
  - The Karnofsky Performance Index (KPI)
  - Other versions of the Karnofsky Scale
  - The WHO Functional Scale
  - The Zubrod Scale (or the Eastern Cooperative Oncology Group Performance Status Rating Scale: ECOG PSR)
- Pain scales
  - The McGill Pain Questionnaire (MPQ)
  - Other symptom scales
  - Symptom Distress Scale
  - Lasry Sexual Functioning Scale for Breast Cancer Patients
  - Summary of other symptom scales
  - McGill Nausea Questionnaire (MNQ)
  - The WHO Symptom Checklist
  - Medical Research Council (MRC) UK Scale
  - The Qualitator
  - Disease-specific quality of life scales
    - EORTC Modular Approach
    - Rotterdam Symptom Checklist
    - Functional Living Index – Cancer (FLIC)
    - Functional Living Index – Emesis (FLIE)
    - Cancer Inventory of Problem Situations (CIPS) and the Cancer Rehabilitation Evaluation System (CARES)
    - Visual Analogue Scales (VASs)
    - Spitzer Quality of Life (QL) Index
    - Linear Analogue Self-Assessment (LASA) Scale
    - Ontario Cancer Institute/Royal Marsden LASA Scale
    - Padilla Quality of Life (QL) Scale and variants, including the Multidimensional Quality of Life Scale – Cancer (MQOLS-CA)
    - Holmes and Dickerson
    - Global Quality of Life Scale (Coates)
    - Quality of Life Index
    - VAS for bone marrow transplant patients
    - Other scales
    - Functional assessment of cancer therapy – Lung Quality of Life Instrument (FACT–L)
    - Breast Cancer Chemotherapy Questionnaire (BCCQ)
    - European Neuroblastoma Study Group Quality of Life Assessment Form – Children (QLAF–C)
    - Brief details of other scales
    - Cancer Leukaemia Group B Studies (CALGB)
    - Ability Index
    - Burge Quality of Life Severity Scale
    - Anamnestic Comparative Self-Assessment (ACSA)
    - World Health Organization Quality of Life Assessment Instrument (WHOQOL)
    - Time Without Symptoms and Toxicity (TWiST)
    - Palliative care
    - Support Team Assessment Schedule (STAS)
    - McGill Quality of Life Questionnaire (MQOL)
    - Conclusion

---

**Table of Contents**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scales measuring physical functioning, pain and symptoms</td>
<td>30</td>
</tr>
<tr>
<td>Functional status scales used with cancer patients</td>
<td>30</td>
</tr>
<tr>
<td>The Karnofsky Performance Index (KPI)</td>
<td>30</td>
</tr>
<tr>
<td>Other versions of the Karnofsky Scale</td>
<td>33</td>
</tr>
<tr>
<td>The WHO Functional Scale</td>
<td>33</td>
</tr>
<tr>
<td>The Zubrod Scale (or the Eastern Cooperative Oncology Group Performance Status Rating Scale: ECOG PSR)</td>
<td>33</td>
</tr>
<tr>
<td>Pain scales</td>
<td>34</td>
</tr>
<tr>
<td>The McGill Pain Questionnaire (MPQ)</td>
<td>34</td>
</tr>
<tr>
<td>Other symptom scales</td>
<td>35</td>
</tr>
<tr>
<td>Symptom Distress Scale</td>
<td>35</td>
</tr>
<tr>
<td>Lasry Sexual Functioning Scale for Breast Cancer Patients</td>
<td>36</td>
</tr>
<tr>
<td>Summary of other symptom scales</td>
<td>36</td>
</tr>
<tr>
<td>McGill Nausea Questionnaire (MNQ)</td>
<td>36</td>
</tr>
<tr>
<td>The WHO Symptom Checklist</td>
<td>37</td>
</tr>
<tr>
<td>Medical Research Council (MRC) UK Scale</td>
<td>37</td>
</tr>
<tr>
<td>The Qualitator</td>
<td>37</td>
</tr>
<tr>
<td>Disease-specific quality of life scales</td>
<td>37</td>
</tr>
<tr>
<td>EORTC Modular Approach</td>
<td>37</td>
</tr>
<tr>
<td>Rotterdam Symptom Checklist</td>
<td>42</td>
</tr>
<tr>
<td>Functional Living Index – Cancer (FLIC)</td>
<td>44</td>
</tr>
<tr>
<td>Functional Living Index – Emesis (FLIE)</td>
<td>46</td>
</tr>
<tr>
<td>Cancer Inventory of Problem Situations (CIPS) and the Cancer Rehabilitation Evaluation System (CARES)</td>
<td>47</td>
</tr>
<tr>
<td>Visual Analogue Scales (VASs)</td>
<td>49</td>
</tr>
<tr>
<td>Spitzer Quality of Life (QL) Index</td>
<td>49</td>
</tr>
<tr>
<td>Linear Analogue Self-Assessment (LASA) Scale</td>
<td>52</td>
</tr>
<tr>
<td>Ontario Cancer Institute/Royal Marsden LASA Scale</td>
<td>54</td>
</tr>
<tr>
<td>Padilla Quality of Life (QL) Scale and variants, including the Multidimensional Quality of Life Scale – Cancer (MQOLS-CA)</td>
<td>55</td>
</tr>
<tr>
<td>Holmes and Dickerson</td>
<td>57</td>
</tr>
<tr>
<td>Global Quality of Life Scale (Coates)</td>
<td>57</td>
</tr>
<tr>
<td>Quality of Life Index</td>
<td>58</td>
</tr>
<tr>
<td>VAS for bone marrow transplant patients</td>
<td>59</td>
</tr>
<tr>
<td>Other scales</td>
<td>59</td>
</tr>
<tr>
<td>Functional assessment of cancer therapy – Lung Quality of Life Instrument (FACT–L)</td>
<td>61</td>
</tr>
<tr>
<td>Breast Cancer Chemotherapy Questionnaire (BCCQ)</td>
<td>60</td>
</tr>
<tr>
<td>European Neuroblastoma Study Group Quality of Life Assessment Form – Children (QLAF–C)</td>
<td>61</td>
</tr>
<tr>
<td>Brief details of other scales</td>
<td>62</td>
</tr>
<tr>
<td>Cancer Leukaemia Group B Studies (CALGB)</td>
<td>62</td>
</tr>
<tr>
<td>Ability Index</td>
<td>62</td>
</tr>
<tr>
<td>Burge Quality of Life Severity Scale</td>
<td>63</td>
</tr>
<tr>
<td>Anamnestic Comparative Self-Assessment (ACSA)</td>
<td>63</td>
</tr>
<tr>
<td>World Health Organization Quality of Life Assessment Instrument (WHOQOL)</td>
<td>63</td>
</tr>
<tr>
<td>Time Without Symptoms and Toxicity (TWiST)</td>
<td>64</td>
</tr>
<tr>
<td>Palliative care</td>
<td>64</td>
</tr>
<tr>
<td>Support Team Assessment Schedule (STAS)</td>
<td>65</td>
</tr>
<tr>
<td>McGill Quality of Life Questionnaire (MQOL)</td>
<td>66</td>
</tr>
<tr>
<td>Conclusion</td>
<td>68</td>
</tr>
</tbody>
</table>
### Quality of life research and psychiatric and psychological morbidity

- Measuring quality of life in institutions
- Community indicators of quality of life
- Measuring quality of life in individuals: the domains of measurement
- Subjective accounts and the value of patients’ ratings

### THE MEASUREMENT SCALES

#### Part I Symptom scales
- Background
- Symptom questionnaires
- Hopkins Symptom Checklist (SCL) and the Symptom Checklist-90
- Brief Symptom Inventory (BSI)
- Symptom Rating Test (SRT)
- Center for Epidemiological Studies Depression Scale (CES-D)
- Rand Depression Screener
- Hospital Anxiety and Depression Scale (HADS)
- Goldberg’s General Health Questionnaire (GHQ)
- Beck Depression Inventory (BDI)
- Hamilton Rating Scale for Depression
- Zung Self-rating Depression Scale
- Quality of Life in Depression Scale (QLDS)
- State–Trait Anxiety Inventory (STAI)
- Profile of Mood States (POMS)
- Scales for children
- Conclusion to Part I

#### Part II Quality of life instruments
- Condition-specific quality of life scales
- Quality of Life Scale (QLS)
- Lehman Quality of Life Interview
- Variations
- Lancashire Quality of Life Profile
- SmithKline Beecham Quality of Life Scale (SBQOL)
- Oregon Quality of Life Self-Report Questionnaire and semi-structured interview rating version
- Wisconsin Quality of Life Index for Mental health (W-QLI)
- Index of Health-Related Quality of Life
- Health of the Nation Outcome Scales for Severe Mental Illness (HONOS–SMI)
- Team for the Assessment of Psychiatric Services (TAPS) Measures
- General Well-Being Schedule (GWBS)
- Scales for assessing needs
- Key Informant Survey Scales
- MRC Needs for Care Assessment
- Camberwell Assessment of Need (CAN)
- Services Needed, Available, Planned, Offered, Rendered Instrument
- Patient attitude scales
Social Problem Questionnaire 126
Client Satisfaction Questionnaire (CSQ) 127
Satisfaction with Life Domains Scale 129
Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q) 129
Conclusion to Part II 130

Part III Role functioning and related instruments 131
Role functioning, performance and social behaviour scales 131
Social Functioning Schedule 132
Social Role Performance Schedule (SRPS) 133
Social Behaviour Schedule (SBS) 134
Rehabilitation Evaluation (REHAB) 136
Social Behaviour Assessment Schedule (SBAS) 137
WHO Disability Assessment Schedules (WHO-DAS) – Psychiatric 138
Global Assessment Scale (GAS) 140
Global Assessment of Functioning (GAF) Scale 141
Adjustment and adaptation scales for use with psychiatrically and/or physically ill people 141
Social Dysfunction Rating Scale 142
Structured and Scaled Interview to Assess Maladjustment (SSIAM) 142
Vineland Adaptive Behaviour Scale 144
Social Adjustment Scale-II (SAS-II) and Self-Report Version 146
Katz Adjustment Scales 147
Psychosocial Adjustment to Illness Scale (PAIS) 148
Global Adjustment to Illness Scale (GAIS) 149
Acceptance of Illness Scale 150
Reintegration to Normal Living (RNL) Index 150
Stress and coping scales for use with psychiatrically and/or physically ill people 151
Ways of Coping Scale 153
Health and Daily Living Form 155
Stress in Life Coping Scale 156
COPE 157
Control over life 158
Rotter’s Internal–External Locus of Control Scale 159
Multidimensional Health Locus of Control Scales 160
Generalized Self-Efficiency Scale (GSES) 161
Conclusion to Part III 162

4 RESPIRATORY CONDITIONS 163
Quality of life in respiratory disease sufferers 163
The domains of measurement 163
Measurement scales 168
Symptom-specific scales: dyspnoea 168
The Fletcher Scale and the MRC Dyspnoea Grade and Respiratory Symptoms Questionnaire 169
American Thoracic Society Respiratory Questionnaire and Grade of Breathlessness Scale 171
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horsley Respiratory Symptoms Questionnaire</td>
<td>171</td>
</tr>
<tr>
<td>American Lung Association Severity of Disability (COAD)</td>
<td>172</td>
</tr>
<tr>
<td>Feinstein’s Index of Dyspnoea</td>
<td>172</td>
</tr>
<tr>
<td>Severity of Symptoms Visual Analogue Scale (VAS)</td>
<td>173</td>
</tr>
<tr>
<td>Oxygen-Cost Diagram</td>
<td>173</td>
</tr>
<tr>
<td>The 6- and 12-Minute Walking Tests and Stair Climbing</td>
<td>174</td>
</tr>
<tr>
<td>Borg Ratio of Perceived Exertion</td>
<td>174</td>
</tr>
<tr>
<td>Mahler Baseline and Transition Dyspnoea Index</td>
<td>175</td>
</tr>
<tr>
<td>Disease-specific quality of life measures</td>
<td>176</td>
</tr>
<tr>
<td>Quality of life assessment in COPD/COAD patients: disease-specific scales</td>
<td>176</td>
</tr>
<tr>
<td>Guyatt’s McMaster Chronic Respiratory Questionnaire (CRQ)</td>
<td>176</td>
</tr>
<tr>
<td>St George’s Respiratory Questionnaire (SGRQ)</td>
<td>179</td>
</tr>
<tr>
<td>Chronic Disease Assessment Tool (CDAT)</td>
<td>180</td>
</tr>
<tr>
<td>Breathing Problems Questionnaire (BPQ)</td>
<td>180</td>
</tr>
<tr>
<td>Quality of life assessment in asthma patients: disease-specific scales</td>
<td>182</td>
</tr>
<tr>
<td>Living with Asthma Questionnaire</td>
<td>182</td>
</tr>
<tr>
<td>Outcome Measures in Ambulatory Care (Asthma and Diabetes) (OMAC)</td>
<td>183</td>
</tr>
<tr>
<td>Asthma Quality of Life Questionnaire</td>
<td>184</td>
</tr>
<tr>
<td>Asthma Symptom Checklist</td>
<td>186</td>
</tr>
<tr>
<td>Summary of other scales which await full psychometric testing</td>
<td>187</td>
</tr>
<tr>
<td>Asthma Self-efficacy Scale</td>
<td>187</td>
</tr>
<tr>
<td>Attitudes to Asthma Scale</td>
<td>187</td>
</tr>
<tr>
<td>Life Activities Questionnaire for Adult Asthma (LAQAA)</td>
<td>188</td>
</tr>
<tr>
<td>Conclusion</td>
<td>188</td>
</tr>
</tbody>
</table>

# Neurological Conditions

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measuring health-related quality of life in neurology</td>
<td>189</td>
</tr>
<tr>
<td>Cerebrovascular accident (CVA/stroke)</td>
<td>189</td>
</tr>
<tr>
<td>The domains of measurement</td>
<td>190</td>
</tr>
<tr>
<td>Measurement scales</td>
<td>193</td>
</tr>
<tr>
<td>Stroke scales</td>
<td>193</td>
</tr>
<tr>
<td>Rankin Handicap Scale</td>
<td>193</td>
</tr>
<tr>
<td>National Institute of Health (NIH) Stroke Scale</td>
<td>194</td>
</tr>
<tr>
<td>Canadian Neurological Scale (CNS)</td>
<td>195</td>
</tr>
<tr>
<td>Hemispheric Stroke Scale</td>
<td>195</td>
</tr>
<tr>
<td>Disease-specific measures of quality of life: functional ability</td>
<td>196</td>
</tr>
<tr>
<td>Barthel Index</td>
<td>196</td>
</tr>
<tr>
<td>Modifications of the Barthel Index</td>
<td>198</td>
</tr>
<tr>
<td>Granger’s Modified Barthel Index</td>
<td>198</td>
</tr>
<tr>
<td>Functional Independence Measure (FIM) and Functional Assessment Measure (FAM)</td>
<td>198</td>
</tr>
<tr>
<td>Gompertz’s Modified Barthel Index</td>
<td>199</td>
</tr>
<tr>
<td>Nottingham Extended Activities of Daily Living Questionnaire</td>
<td>199</td>
</tr>
<tr>
<td>Frenchay Activities Index</td>
<td>200</td>
</tr>
<tr>
<td>Rivermead Mobility Index (RMI)</td>
<td>201</td>
</tr>
<tr>
<td>Disease-specific measures of quality of life: stroke</td>
<td>202</td>
</tr>
<tr>
<td>Kudos Battery for Stroke Patients</td>
<td>202</td>
</tr>
<tr>
<td>London Handicap Scale</td>
<td>202</td>
</tr>
</tbody>
</table>
### Epilepsies

- The domains of measurement
- What instruments have been used?
- Measurement scales
- A severity measurement scale
- Disease-specific quality of life scales: epilepsy
- Washington Psychosocial Seizure Inventory (WPSI)
- Epilepsy Surgery Inventory (ESI-55)
- Katz Adjustment Scales (revised for use with epilepsy)
- Health-Related Quality of Life Model
- General tests of neurophysiological functioning
- Trail-Making Test
- Wechsler Scales
- Wechsler Memory Scale (WMS)
- Wechsler Adult Intelligence Scale (WAIS)
- Scales for children
- Tests for cognitive impairment in elderly people
- Abbreviated Mental Test Score (AMTS)
- Mini Mental State Examination (MMSE)
- Mental Status Questionnaire (MSQ)
- Older Americans' Resources and Services Schedule (OARS) mental health measures
- Brief Cognitive Functioning Scale (SIP sub-scale)
- Geriatric Mental State (GMS) and Comprehensive Assessment and Referral Evaluation (CARE)
- Conclusion

### RHEUMATOLOGICAL CONDITIONS

- Quality of life in people with joint disorders
- The domains of measurement
- The measurement scales
- Disease severity and functional classifications
- Disease-specific scales: physical functioning
- Stanford Arthritis Center Health Assessment Questionnaire (HAQ)
- Arthritis Impact Measurement Scales (AIMS)
- Functional Status Index (FSI)
- Western Ontario and McMaster Universities Arthritis Index (WOMAC)
- Office of Population Censuses and Surveys (OPCS) Disability Scale
- Rheumatoid Arthritis (RA) Impact on the Homemaker Questionnaire
- Broader disease-specific quality of life scales
- MACTAR (McMaster–Toronto Arthritis) Patient Function Preference Questionnaire
- Hörnquist’s Quality of Life Status and Change
- Rand Joint Problems Battery
- Summaries of other scales
- Arthritis Helplessness Index (AHI)
- EURIDISS (European Research on Incapacitating Diseases and Social Support)
- Conclusion
7 CARDIOVASCULAR DISEASES

Heart disease
Trends in measuring quality of life in people with heart disease
Most and least frequently used indicators
What should be measured?
The domains of measurement
The measurement scales
Categorizations and measures of physical activity
New York Heart Association (NYHA) Functional Classification Scale
Olsson Ranking Scale
Canadian Cardiovascular Society (CCS) Functional Classification for Angina Pectoris
Specific Activity Scale (SAS)
Rankin Handicap Scale
Symptom questionnaires
Fletcher Questionnaire and the Rose (WHO) Questionnaires: The London School of Hygiene Dyspnoea Questionnaire and Cardiovascular Questionnaire (angina of effort)
Feinstein’s Index of Dyspnoea
Ratio Property Scale
Disease-specific quality of life scales: heart disease
Rand Congestive Heart Failure (Shortness of Breath/Enlarged Heart/Heart Failure)
   Battery and Rand Chest Pain (Angina) Battery
Guyatt’s Chronic Heart Failure Questionnaire (CHQ)
Seattle Angina Questionnaire (SAQ)
Minnesota Living with Heart Failure (LIhFE) Questionnaire
Quality of Life in heart failure Questionnaire (QLQ-CHF)
Ferrans and Powers’ Quality of Life Index (CARDIAC)
Rehabilitation Questionnaire
Circulatory disease
Trends in measuring quality of life in people with circulatory disease: hypertension
What should be measured?
Disease-specific quality of life scales: hypertension
Palmer’s Symptom Checklist and Battery
Bulpitt’s Hypertension Questionnaire and Batteries
Quality of Life Impairment Scale – Hypertension
Rand Health Insurance Study (HIS) Blood Pressure Battery
Conclusion

8 OTHER DISEASE- AND CONDITION-SPECIFIC SCALES

Other developments in the measurement of disease-specific quality of life
Popular generic measures used to supplement disease-specific scales
Disease-specific measures of quality of life
Diabetes
Outcome Measures in Ambulatory Care (OMAC) (Asthma and Diabetes)
Diabetes Impact Measurement Scales (DIMIS)
Quality of Life, Status and Change (QLSC)
Rand Diabetes Mellitus Battery
Diabetes Quality of Life Measure (DQOL) 282
Renal disease 283
Functional classification 283
The use of generic scales and domain-specific batteries 283
Kidney Disease Quality of Life Questionnaire (KDQoLQ) 284
Haemodialysis Quality of Life Questionnaire (HQLQ) 285
Leicester Uraemic Symptom Scale (LUSS) and Battery 285
Quality of Life Assessment 286
Bowel diseases 287
Inflammatory Bowel Disease Questionnaire (IBIQ) 287
Rand Surgical Conditions Battery: haemorrhoids 288
Human Immunodeficiency Virus (HIV) and Acquired Immune Deficiency Syndrome (AIDS) 288
Medical Outcomes Study HIV Health Survey 30-Item Form (MOS-HIV) 289
HIV Overview of Problems–Evaluation System (HOPES) 290
AIDS Health Assessment Questionnaire 291
Varicose veins 291
Varicose Veins Questionnaire 291
Rand Surgical Conditions Battery: Varicose Veins 292
Back pain 293
Clinical Back Pain Questionnaire 294
Children 295
Older people 296
A popular core (generic) measure for use in disease-specific studies of adults: the Short Form-36 (SF-36) 298
Individualized measures of quality of life 302
Schedule for the Evaluation of Individual Quality of Life (SEIQoL) 303
Patient Generated Index (PGI) 305

APPENDIX: A SELECTION OF USEFUL SCALE DISTRIBUTORS AND ADDRESSES 307
REFERENCES 313
INDEX 390
Whom would the just man fail to greet, in order to stop an injustice?
What medicine tastes too nasty to save a dying man?
How much meanness would you not commit if the Aim is to stamp out meanness?
If you’d found a way to alter this planet, what would You Refuse to do?


The aim of this volume is to introduce the key literature on the psychometric properties of measures of disease-specific quality of life, including the symptom scales often used alongside them. In addition to updating the reviews published in the first edition of this book, the opportunity has also been taken to include additional scales. These include the Living with Heart Failure Questionnaire; the Heart Failure Questionnaire; the Seattle Angina Questionnaire; the Camberwell Assessment of Need, Quality of Life Enjoyment and Satisfaction Questionnaire; the Quality of Life in Depression Scale; the London Handicap Scale; the Breathing Problems Questionnaire; the Medical Outcomes Study HIV Health Survey; the McGill Quality of Life Questionnaire; the McGill Nausea Questionnaire; the Support Team Assessment Schedule; and the Kidney Disease Quality of Life Questionnaire. Individualized measures are also included. For example, the Schedule for the Evaluation of Individual Quality of Life and the Patient Generated Index. The final chapter on measurement issues in the first edition of Measuring Disease has been removed from this edition, as these are covered more fully in the author’s Research Methods in Health (Open University Press 1997).

There has been a rapidly increasing interest in both generic and disease-specific measures of quality of life. Other current reviews of a broad range of measurement scales are those by Wilkin et al. (1992); McDowell and Newall (1996); Spilker et al. (1990); Spilker (1996) and Salek (1998). Jenkinson and McGee (1998) have also described the principles for designing disease-specific measures. Databases of research on the related concept of happiness have also been compiled (Veenhoven 1994). The rapid growth of an industry of standardized scale development in relation to quality of life reflects the international emphasis on the provision of effective, evidence based health care, and the measurement of the outcome of care in the broadest sense. It has also led to a renewed concern about the content validity of the measures produced, and hence to interest in using individualized measures. This wide range of approaches is included here in the second edition of Measuring Disease. The scale literature was updated by using
electronic database searches of the literature, including EMBASE, Ageline, Cancer, Medline, CINAHL and PsychINFO, manual searches of key journals and correspondence with scale developers. The general rationale for the inclusion of scales in this volume is the existence of a body of evidence to support their psychometric properties. In some cases, where this body of evidence is not extensive or where testing is still in progress, scales have been included because of their potential or continuing popularity. It is, of course, impossible to include every scale that has been developed, and decisions have had to be taken over including some and excluding others where their evidence base is still emerging. The scales are presented in chapters which relate to some of the most common conditions across all age groups. Areas where scale development is less well developed, and/or there is less of a tradition of outcome measurement, are grouped together in a final chapter along with brief reviews of commonly recommended or increasingly popular core modules (standardized and individualized).

THE STRUCTURE OF THE CHAPTERS

Each chapter follows a similar, but not identical pattern, which is dependent upon the amount and type of research on health-related quality of life that has been carried out in that field. Each chapter on condition- and disease-specific scales begins with an introduction to the methods of measuring quality of life in that area, and also references the commonly used generic and domain-specific (e.g. depression) health-related quality of life scales. Generic and several domain-specific scales have been reviewed by the author elsewhere, in Measuring Health (Open University Press 1997).

Disease- and condition-specific quality of life scales often require further supplementation with disease-specific symptom items, or other sensitive indicators of the presence of perceptions of ill health. These range from self-reporting of symptoms and unwanted treatment effects, as in cancer, to assessments of grip strength among people with joint problems, to the reporting of symptoms of anxiety and depression in the case of psychiatric conditions and psychological morbidity. Although they are not quality of life measures, it is important to include perceived symptoms and discomfort in batteries of measures purporting to measure health-related quality of life, and thus they are presented here before the reviews of quality of life measures.

RECOMMENDATIONS

The strengths, weaknesses and coverage of each scale are presented in the text. Readers need to acquaint themselves with the range of available scales in their field of interest and select the ones which are most appropriate for the aims of their investigation. All scales have their good and bad points. Not all scales have been fully tested, and scale development is often ongoing. For this reason, no ‘quasi-scientific’ league table of ‘best buys’ has been attempted, although readers will find text comments, and a concluding section at the end of each chapter with some recommendations, which should help them to make up their own minds about whether a particular scale is appropriate for their study.

Finally, it is easier to criticize a scale than to construct one. Researchers are encouraged to use existing scales, or adapt them where necessary, rather than design them from scratch. Where scales have not been fully tested, users are encouraged to test further the reliability and validity of selected instruments when carrying out their own research. In this way, a better, and continually updated, body of knowledge will be developed.

COPYRIGHT

Many scales are only available for purchase commercially. This is particularly true of those developed in North America. In many cases, the purchase price simply covers administration costs and the cost of a manual and questionnaire; in others, the cost of the manual and scale can be several hundred pounds. Potential users are advised to contact the authors or distributors of scales for details of permission of use and, where applicable, purchase.

The author would like to echo the plea of Wilkin et al. (1992): in view of the difficulty that can be experienced in obtaining scales, authors should be
encouraged by publishers to either reproduce their questionnaire (if short) in a major article on the scale’s psychometric properties, or publish the name and address of the scale distributor. This would save a great deal of unnecessary correspondence, searching electronically and manually for likely references and scanning of the catalogues of the different distributors.
Whenever we seemed
To have found the answer to a question
One of us untied the string of the old rolled-up
Chinese scroll on the wall, so that it fell down and
Revealed to us the man on the bench who
Doubted so much.

(‘The Doubter’, Bertolt Brecht, c. 1937. Reprinted
Bertolt Brecht, Poems 1913–1956. London:
Methuen. Reproduced with permission.)

BACKGROUND: THE ‘QUALITY
OF LIFE’

In the developed world, with the emphasis on
affluence and with the increasing longevity of
populations, there is a general interest in how to
achieve the ‘goodness’ of life, sometimes called
life satisfaction or quality of life. But what is
quality of life? Dictionary definitions of quality
indicate that it implies the degree of excellence of
a characteristic. But the quality of life may mean
different things to different people. The theoret-
cal definitions of related concepts of happiness,
life satisfaction, well-being, the ‘good life’ and
‘quality of life’ have attracted much conceptual
confusion, and have preoccupied a wide range of
disciplines, dating far back to Aristotle (384–22
bc) and early Greek philosophy (see Ross 1947;
Veenhoven 1991; Argyle et al. 1995).

The following extract is taken from a dialogue
between Serafin and Summerchild in Michael
Frayn’s (1991) novel, A Landing on the Sun, in
which the plot centres on a government investiga-
tion of quality of life. It is used here, along with
the discussion that follows it, to illustrate the inef-
fable nature and topicality of ‘quality of life’:

...Now...To recapitulate: ‘the quality of
life’, as you understand it, is some property
which is in one way or another promoted
or enhanced by washing machines...
[p. 81]...Does the class of things which to
your mind promote[s]...the quality of life
stretch far enough from the kitchen door to
include the family television set? [p. 82]...
Washing machines, and similar sorts of
domestic machinery, are intended to save
our time and labour. Now no-one could
claim that a television set saves time. On the
contrary – it uses up the time we’ve saved
with the dishwasher and the vacuum
cleaner. And I think one might go on to
argue that far from saving labour it creates
it...[p. 83]...They enjoy watching it.
Let’s leap from the television to the central
heating. Do they have central heating
because they enjoy it? [p. 83]...Do you
find the word ‘motor car’ on your list?
[p. 85]...The list that you apparently keep
inside your head of things that enhance the quality of life [p. 85].

(Frayn 1991)¹

Walter and Shannon (1990) described the current interest in ‘quality of life’ in the developed world, ranging from current concerns for the environment to the marketing of the products we buy, and to the evaluation of the benefit–burden ratio involved in medical treatments. More globally, quality of life as an outcome indicator has been added to social, as well as health, service programme development (Department of Health and Social Security 1989). It was added to the World Health Organization’s (WHO) international Healthy Cities Programme, now in its third phase (World Health Organization 1998), and defined as spanning the visual arts, recreation, employment, transport, housing, environmental and conservation issues, health and other indicators of what has been labelled as ‘the social temperature’. The Organization for Economic Cooperation and Development (OECD) long ago agreed on a list of quality of life related social concerns of member states, including health, command over goods and services, employment and quality of working life (Andrews 1973), and the national press regularly publishes comparative annual statistics on indicators of quality of life across the world. The salience of quality of life across disciplines has resulted in the emergence of a generic Quality of Life Questionnaire (Evans and Cope 1989), and quality of life questionnaires for use in health care evaluation (Chubon 1987; Patrick et al. 1988; Siegrist et al. 1993).

WHAT IS QUALITY OF LIFE AND HEALTH-RELATED QUALITY OF LIFE?

Quality of life

Quality of life research, then, spans a range of topics, from quality of life in the last year of life (Lawton et al. 1990) to quality of life in urban environments (Rogerson et al. 1989). As was illustrated earlier, quality of life is an amorphous concept that has a usage across many disciplines – geography, literature, philosophy, health economics, advertising, health promotion and the medical and social sciences (e.g. sociology and psychology). It is also a vague concept; it is multidimensional and theoretically incorporates all aspects of an individual’s life.

Social science and the ‘good life’

There are several meanings of the term ‘quality of life’ in social research, which range from individual fulfilment and satisfaction with life – the satisfaction of basic human needs, the ability to lead a ‘normal’ life – to the quality of the external environment (Edlund and Tancredi 1985; Fowlie and Berkeley 1987; Rogerson et al. 1989). Quality of life has been defined as the ‘output’ of the inputs of the physical and the spiritual (Liu 1974); as the degree to which a person accomplishes life goals (Cella and Cherin 1987); and even quantified crudely as a formula in which quality of life (QL) is a product of one’s natural endowment (NE) and the effort made on one’s behalf by the family (H) and society (S), such that $QL = NE \times H \times S$ (Shaw 1977). Grant et al. (1990) defined quality of life as ‘a personal statement of the positivity or negativity of attributes that characterize one’s life’. Quality of life is still taken to refer to equally amorphous conceptions of the ‘goodness of life’ (Zautra and Goodhart 1979).

Experiential social indicators research, which developed during the 1960s onwards, focused on the importance of measuring subjective well-being (see Elster and Roemer 1993). The concern with subjective indicators led to the first large surveys of life satisfaction, happiness, quality of life and the ‘good life’ among adults and among older people (Gurin et al. 1960; Bradburn and Caplowitz 1965; Bradburn 1969; Andrews 1973; Andrews and Withey 1976; Campbell et al. 1976; Lawton 1983). This research has increasingly indicated that, in contrast to subjective variables, objective, social background variables account for relatively little of the variance in happiness, life satisfaction and well-being, thus leading to more emphasis on the importance of subjective feelings of independence, control and autonomy as predictors of well-being (Larson 1978; Inglehart and Rabier 1986; Mroczek and Kolarz 1998). Issues of definition and measurement were the subject of critical investigation


Lawton (1983) first proposed a theoretical model of quality of life as 'the good life', defined as psychological well-being, perceived quality of life, behavioural competence and the 'objective' environment. It was increasingly held that, in the developed world where human needs have generally been met, quality of life is the extent to which pleasure and satisfaction have been obtained (Andrews 1974). Quality of life has been referred to as an affective response to one's role situation and values (Andrews and Withey 1976), as the discrepancy between desired and actual circumstances (Krupinski 1980), and as the well-being or 'ill-being' of people and/or their environment (Bubolz et al. 1980). The gerontological literature on the topics of 'successful ageing', 'positive ageing' and 'quality of older age' makes a similar point, focusing largely on life satisfaction and morale (Neugarten et al. 1961; Havighurst 1963; Williams and Wirths 1965; Bradburn 1969; Lawton 1975; Andrews and Withey 1976; Palmore 1979). There has been a rapid expansion of literature on correlates of life satisfaction as a proxy for quality of life (Bowling et al. 1991; Shahtahmasebi et al. 1992); and life satisfaction has become a key variable in analyses of health status and mortality in old age (Bowling and Browne 1991; Grundy et al. 1992; Bowling and Farquhar 1993; Bowling et al. 1994, 1996a). The work of social gerontologists also has roots in philosophy. As long ago as 44 bc, Cicero ([44 bc] 1979), argued that old age contains many opportunities for positive change and productive functioning, and should not be confused with illness. Quality of life research in gerontology has taken the theoretical debate further with an emphasis on the additional importance of feelings of cognitive efficacy, social competence and productivity, personal control and motivation (Baltes and Baltes 1990; Day 1991).

In non-experiential social indicators research, quality of life encompasses all external, or objective, circumstances of life – for example housing, leisure activities, work, the environment and so on (Campbell et al. 1976; Wingo and Evans 1978; Kaplan 1985). Environmental research has focused on non-experiential objective background characteristics of neighbourhoods, while also attempting to incorporate subjective public values and levels of satisfaction and preferences. Rogerson et al. (1989) and Rogerson (1995) reviewed the concept of quality of life, and pointed out that no definitive list of criteria of quality of life has yet been developed, and neither has an acceptable weighting system for the incorporation of subjective and objective indicators. The research conducted by Rogerson et al. (1989) in human geography was based on a national opinion survey of 1200 respondents in Britain. They were asked to rate 20 dimensions of quality of life in terms of their degree of importance in influencing their choice of where to live, on a scale of 5 (very important) to 1 (indicating minimal significance). The top five items were violent crime, non-violent crime, health provision, pollution levels and cost of living. The bottom five related to travel to work time, leisure facilities, quality council housing, access to council housing and cost of private or rented housing. Thus, in research on quality of life in cities, health care ranks third in importance. This research is consistent, and overlaps with housing surveys of neighbourhood satisfaction and community needs assessment exercises (Percy-Smith and Sanderson 1992; Burrows and Rhodes 1998).

It has been argued that human needs are the foundations for quality of life and that quality of life is the degree of satisfaction of those needs – for example, physical, psychological, social, activity, marital and structural (Hörnquist 1982). Needs satisfaction is the theoretical basis of many of the quality of life scales developed for use in mental health. This theory is reminiscent of Maslow's (1954, 1962a, 1962b) hierarchy of need (physiological, safety and security, social and belonging, ego, status and self-esteem, and self-actualization), and the argument that once basic biological and survival needs have been met, emotional and social needs become more prominent. Gap (or relative deprivation) theorists argue that in the developed world perceptions of quality of life are less likely to be related to basic needs, where these are largely met, but more to one's expectations in life, and to social comparisons with past achievements, and also to comparisons with others (Michalos 1986). Quality of life could be defined in relative terms of what one has lost, or lacks, rather than what one has. Calman (1984) proposed that quality of life is the difference, at a
particular period in time, between the hopes and expectations of the individual and their present experience. So quality of life is influenced by past experience, present circumstances and aspirations for the future. Aristotle (384–22 BC) stated, in relation to man’s perception of happiness: ‘When he falls ill he says that it is his health, and when he is hard up he has that it is money’. In relation to health, quality of life has also been defined in terms of the difference between reality, or the perception of reality, and expectations (Calman 1984; Presant 1984). However, the relativity argument has been criticized on the grounds of conceptual confusion (Veenhoven 1991) and inconsistent literature in support of it (Headey and Wearing 1992). Also, gap theories are really explanations of perceptions of quality of life, and do not increase knowledge about the constituents of quality of life as such. There has been little attempt to examine these assumptions critically (Bauer 1966a; Ziller 1974) although they provide the conceptual background to many health-related quality of life scales. For example, gap theory (Calman 1984) underpins the Patient Generated Index (PGI) (Ruta et al. 1994b, 1999) and the needs satisfaction model underpins several mental health scales, in addition to the more subjective life satisfaction model (e.g. Lehman et al.’s (1982) Quality of Life Interview).

**Philosophical approaches**

Overlapping with this research is the history of the exploration of happiness dating from the work of early to present-day philosophers (e.g. Aristotle, referred to earlier) and social scientists (Morgan 1934; Ross 1947; Barschak 1951; Tatarskievicz 1975; Gallup 1976; Mason and Faulkenberry 1978; Andrews 1981; Parducci 1984; Veenhoven 1988, 1989, 1991). Bentham ([1834] 1983) introduced the dimension of psychological hedonism as the governing principle of human conduct in the form of well-being and its measurement, which he defined as ‘the difference in value between the sum of pleasures of all sorts and the sum of pains of all sorts which a man experienced in a given period of time’. He attempted to show how the moral theory of utilitarianism could be developed into a calculus of pleasures whereby the effects of actions could be judged, and the right policy thus identified: the end of human conduct is the greatest happiness of the greatest number. It was thus expected that utilitarianism would provide a rational foundation for social and legal policy. This utilitarian focus on pleasure and happiness was derived from the translation of Aristotle’s concept of *eudaimonia* or ‘human flourishing’ as pleasure or happiness, rather than as excellence. Bentham’s argument, that utility is identified with the amount of satisfaction of desires, allows that in some circumstances the happiness of some people should be purchased at the cost of the undeserved and uncompensated misery of others. Contemporary utilitarianism focuses more on satisfaction than on happiness as the relevant outcome. Current philosophical views of the ‘good life’ relate not only to hedonism, but to the satisfaction of desires or preferences, and the realization of normative ideals (Brock 1993; Scanlon 1993). This influence can be seen in current measures of health-related quality of life, particularly in mental health with the predominant model of quality of life as needs satisfaction. The focus of Bentham’s utilitarianism on the greatest happiness of the greatest number also underpins attempts by policy makers and health economists to justify health service rationing policies (particularly age-based rationing policies) (A. Williams 1997). The use of Quality Adjusted Life Years (QALYs) as indicators of health gain in health service resource allocation is frequently defended on utilitarian grounds, in that they appear to facilitate the greatest health gain, given scarce resources, for the greatest number. The ageism inherent in this formula, and the problems of classic utilitarianism, are rejected by its defenders, who argue that QALY values are derived from surveys of the public and therefore reflect the aggregated preferences of the population (Williams 1994; Edgar et al. 1998). However, the basis of such logic was early on branded by John Stuart Mill (1861) as the worst tyranny – ‘the tyranny of the majority’, legitimized in democracy and expressed in legislation by elected tyrants. Mill argued therefore that society has a moral duty to correct this majoritarian bias and uphold the principle of equality (he actually argued that our unelected legal judges should have this duty). His own ‘Utilitarianism’ (Mill 1861) was an attempt to elaborate on this theory given the conflict in Bentham’s utilitarianism between the demands of justice and the demand to maximize general happiness.
Individual meaning

Phenomenologists argue that quality of life is dependent upon the interpretation, and perceptions, of the individual (Ziller 1974). Cohen (1982) has also pointed out that the simple listing of quality of life domains is not a satisfactory way of measuring quality of life because it is unknown whether all important domains have been included. Researchers who construct health-related quality of life measurement scales are seldom philosophically sophisticated or concerned with competing accounts of the good life. To some extent, the need to develop valid measures for use with large and varied samples of people necessitates compromises and the simplification of issues of philosophical importance (Brock 1993). Rosenberg (1992) has argued that the psychometric translation of quality of life into components such as emotional status, social interaction, economic status, health status and physical capacity, while incorporating the multidisciplinary nature of human beings, does not capture their subjectivity. He argued that hermeneutic thinking should be introduced into modern medicine, so that a naturalistic concept of mankind is presented along with a concept of the human being as a self-reflective individual responsible for their own actions. Ziller (1974) has also argued that the approach to quality of life is ‘through the eye of the experiencer; that is, a phenomenological approach’. The health-related quality of life scales designed by Guyatt et al. (1987a, 1987c, 1989a, 1989d), O’Boyle and colleagues (O’Boyle et al. 1990, 1992, 1993; Hickey et al. 1999), and Ruta and his colleagues (Ruta et al. 1994a, 1994b; Garratt and Ruta 1999), which take account of the individual’s perspective, are exciting developments which are countering the trend towards the pre-definition of quality of life by the researcher (Joyce et al. 1999). O’Boyle’s work, for example, takes individual meaning into account, and has involved the application of the techniques of human judgement analysis (Brehmer and Joyce 1988) to assess health-related quality of life (O’Boyle et al. 1989; Hickey et al. 1999). With this technique respondents are asked to list, rate and weight the five areas of life (‘cues’) that they judge to be the most important to their overall quality of life (the technique is known as the Schedule for the Evaluation of Individual Quality of Life or SIQoL). This is described more fully in Chapter 8.


The general conclusion is that perception and achievement of quality of life is dependent on an individual’s preferences and priorities in life. The meaning of the concept of quality of life is thus arguably dependent on the user of the term, their understanding of it, and their position and agenda in the social and political structure (Edlund and Tancredi 1985): ‘Quality of life is a vague and ethereal entity, something that many people talk about, but which nobody very clearly knows what to do about’ (Campbell et al. 1976). Adequate measurement should therefore reflect these elements and be preference weighted (Diamond and Becker 1999). Individualized measures are a step in this direction.

Is health a domain of quality of life?

The empirical evidence appears to justify including health as a dimension of quality of life. Research on valued states of existence, from the early days of social indicators research to the present day, has reported that health is among the most valued states, and among the most important areas of life and of quality of life nominated by people (Rokeach 1973; Kaplan 1985; Bowling 1995, 1996a, 1996b; Farquhar 1995). It also increases in value and priority with older age (Bowling 1996b).

What is health-related quality of life?

In relation to health, health status is increasingly referred to as quality of life, and, so as to narrow down its operationalization in research studies, quality of life is referred to as health-related quality of life. Quality of life in relation to health is
rarely explicitly defined in published studies (Gill and Feinstein 1994), but often implicitly defined from a functionalist perspective of society, which relates to the ability to perform activities of daily living and fulfill role obligations (necessary for the functioning of society as a whole). It has traditionally been viewed negatively in this context, in terms of ‘dis’abilities, and measured accordingly (e.g. with activities of daily living and physical and role functioning scales).

The use of the concept health-related quality of life in scientific research requires more precise definition of the concept itself. From a health (or disease) perspective, quality of life has been said to refer to the social, emotional and physical well-being of patients following treatment (Greer 1984), mirroring WHO’s (World Health Organization 1947, 1948, 1958) definition of health (see below), and to the impact of disease and treatment on disability and daily functioning (Kaplan 1985). It focuses on the impact of perceived health status on the ability to lead a fulfilling life (Bullinger et al. 1993). It is a double-sided concept, incorporating positive as well as negative aspects of well-being and life, and it is multidimensional, incorporating social, psychological and physical health. It is also, ultimately, a personal and a dynamic concept for, as health status deteriorates, perspectives on life, roles, relationships and experiences change (Sherwood et al. 1977; Morris et al. 1986).

Taking these definitions into account, health-related quality of life in scientific research requires more precise definition of the concept itself. From a health (or disease) perspective, quality of life has been said to refer to the social, emotional and physical well-being of patients following treatment (Greer 1984), mirroring WHO’s (World Health Organization 1947, 1948, 1958) definition of health (see below), and to the impact of disease and treatment on disability and daily functioning (Kaplan 1985). It focuses on the impact of perceived health status on the ability to lead a fulfilling life (Bullinger et al. 1993). It is a double-sided concept, incorporating positive as well as negative aspects of well-being and life, and it is multidimensional, incorporating social, psychological and physical health. It is also, ultimately, a personal and a dynamic concept for, as health status deteriorates, perspectives on life, roles, relationships and experiences change (Sherwood et al. 1977; Morris et al. 1986).

Taking these definitions into account, health-related quality of life is defined here as optimum levels of mental, physical, role (e.g. work, parent, carer, etc.) and social functioning, including relationships, and perceptions of health, fitness, life satisfaction and well-being. It should also include some assessment of the patient’s level of satisfaction with treatment, outcome and health status and with future prospects. It is distinct from quality of life as a whole, which would also include adequacy of housing, income and perceptions of immediate environment. Quality of life as a whole has also been divided into subjective and objective areas: it ‘encompasses what a person is capable of doing (functional status), access to resources and opportunities to use these abilities to pursue interests, and sense of well-being. The former two dimensions are often referred to as objective quality of life and the latter as subjective quality of life’ (Lehman et al. 1995).

The theoretical framework of health-related quality of life, then, is largely based on a multi-dimensional perspective of health as physical, psychological and social functioning and well-being, along the lines of the WHO’s (World Health Organization 1947, 1948) definition of health: a ‘state of complete physical, mental and social well-being and not merely the absence of disease or infirmity’. WHO (World Health Organization 1984) has since added ‘autonomy’ to this list. Ware (1987) has argued that five health concepts are inherent in this definition: physical health, mental health, social functioning, role functioning and general well-being. He restricts his definition because the goal of health care is to maximize the health component of quality of life. Health status may influence quality of life without determining it.

WHO (World Health Organization 1991) has a working party on quality of life under its umbrella: the World Health Organization Quality of Life Group (WHOQOL Group 1991, 1993a, 1993b, 1993c, 1994a), in relation to its international investigation into health-related quality of life (Sartorius 1993), and has provided a definition of quality of life which also takes individual perception and relationship to the environment into account:

Quality of life is defined as an individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person’s physical health, psychological state, level of independence, social relationships, and their relationships to salient features of their environment.

(WHOQOL Group 1993b: 3)

This definition underpins the development of the WHOQOL instrument for measuring quality of life that can be used in a variety of cultural settings (WHOQOL Group 1998b; Skevington 1999).

**Unresolved questions**

There are many vexing and unresolved questions in quality of life research. For example, as Schipper (1983) has asked: How can the quality of life of,
HEALTH-RELATED QUALITY OF LIFE

say, a city and a country dweller be compared? Does their quality of life change in comparable ways in response to illness? Are some parameters of quality of life weighted differently by different groups of people, and in relation to different illnesses? The issue is not straightforward, and a multitude of additional questions are raised when attempting to consider the impact of an illness on an individual: What does the condition prevent the patient from doing? What do they miss doing as a result of the condition? What are their fears for the future? How do they cope with anger and frustration related to their condition? How does mood affect functioning and social relationships? Are there any positive consequences of the illness (e.g. bringing relatives closer together)? How do people spend their time, and how does this change when they are afflicted by certain illnesses (e.g. less time spent on leisure and domestic activities and social roles)? What is the impact of the condition on other family members? How have they adjusted, and have they been restricted?

Rarely does one health status measure encompass any of these items, and researchers are forced to resort to lengthy, and sometimes impractical, batteries of several measurement scales.

HEALTH-RELATED QUALITY OF LIFE AND HEALTH

A concept of health

A concept of health-related quality of life must rest on a concept of health as well as of quality of life. A measure of health status should also be based on a concept of health. A medical conception of health is freedom from disease and abnormalities; a sociological view can be defined in terms of the possession of acceptable levels of mental and physical fitness in order to perform one’s social role in society. As with the concept of quality of life, there is also a humanistic view of health in which optimal autonomy, self-mastery and a positive perception of life are central components or influences (Heyrman and van Hoeck 1993).

The concept of positive health has been described elsewhere (Bowling 1997a). Briefly, most health status instruments measure deviations away from a state of health and are really measuring ill health, or the absence of illness and disease. They rarely reflect the global definition of WHO (World Health Organization 1948), although this is the theoretical underpinning, explicit or implicit, in more recent attempts at scale development. There are multiple influences upon patient outcome, and these require a broad model of health. The non-biological factors which may influence or mediate recovery include patient expectations, self-esteem, self-mastery, motivation, coping, adherence to therapy, socio-economic status, availability of health care, social support networks, and individual cultural beliefs and health behaviours.

A negative conception and measure of health is more appropriate when measuring severely ill populations, but less appropriate in general population surveys. Negative definitions of, and measurements of, health status will tell us little about the health of the population. Positive health is therefore an increasingly popular concept, encompassing not just the absence of disease, but feelings of mental and physical well-being, full functioning, physical fitness, ability to cope, social support, adjustment and efficiency of mind and body. Collectively, these positive states have been referred to in the literature variously as positive health, social health and health-related quality of life. Even in disease-specific studies, where negative measures of health are appropriate, a more balanced scale including positive measures should also be used in order to assess outcome in relation to degrees of wellness as well as illness.

The public’s view of health

Dubos (1959) argued that health and disease cannot be defined merely in terms of anatomical, physiological or mental attributes, and that ‘their real measure is the ability of the individual to function in a manner acceptable to himself and to the group of which he is part’. An absolute definition of health is not possible to construct and even the WHO (World Health Organization 1947, 1948; WHOQOL Group 1993b) definitions can only be viewed as relative concepts (see above). People define their health variously, depending on socio-demographic factors and on their culture (D’Houtard and Field 1984; Currer and Stacey 1986). Their reported definitions include health as not being ill, as absence of disease, as behaviour, as role functioning,
as physical fitness, energy and vitality, as social relationships and as emotional well-being. Wright (1990) has summarized lay definitions of health as health as being, health as doing and health as having. Wright (1997) also pointed to the debate about whether health is a ‘state’ or a ‘trait’ (constructs which, he points out, are not mutually exclusive). He refers to the application of various constructs which he summarizes as ‘dispositional resilience’, such as ‘hardiness’, ‘sense of coherence’ and ‘dispositional optimism’ (Kobasa 1979; Scheier and Carver 1985; Antonovsky 1993). Wright argued that ‘dispositional resilience’ may influence health outcomes.

The concept of health is inevitably subject to cultural relativism (Heyrman and van Hoeck 1993), and while health may be a social goal common to all groups, the salience of health to individuals must be assessed relative to other goals. The place of health in one’s value system will be reflected in one’s definition of health. This has been amply demonstrated by the classic studies of Koos (1954) and Herzlich (1973), as well as by more recent research.

In a study of several thousand people attending health check-ups in Nancy in France, D’Houtard et al. (1990) reported that the most common and consistent definitions of health were ‘hygiene, living conditions, to feel well in one’s skin, to know oneself well, work, luck, to be at the top of one’s form, personal unfolding and not to feel one’s body’. These were comparable with findings from similar research carried out in Nijmegen in the Netherlands. D’Houtard and Field (1984) reported the results of a study of 4000 people from a health centre in Lorraine, France in more detail. When asked an open-ended question about what health meant to them, the most common replies were ‘not to be sick’, ‘to be at the top of one’s form’ (more than 400 mentions), ‘good physical equilibrium’, ‘good mental equilibrium’, and ‘joy of living’ (300–400 mentions).

In the British national Health and Lifestyle Survey of 9003 adults living in randomly sampled households, Cox et al. (1987) reported that the concepts of health most often used for describing what health is in ‘someone else’ were, among female respondents in all age groups, ‘never ill, no disease, never see a doctor’, ‘fit, strong, energetic, physically active’, ‘has healthy habits’ (e.g. not smoking, taking exercise, taking care of health) and ‘able to do a lot, work, socially active’ (in order of frequency mentioned). Among male respondents in all age groups, the most common definitions were ‘fit, strong, energetic, physically active’, ‘never ill, no disease, never see a doctor’, ‘has healthy habits’ (e.g. not smoking, taking exercise, taking care of health) and ‘able to do a lot, work, socially active’ (in order of frequency mentioned). When asked about definitions of health used for describing ‘what it is like to be healthy oneself’, the most common response of more than half the males and females in all age groups was ‘feel psychologically fit’ (e.g. good, happy, able to cope), followed in order of frequency by ‘fit, strong, energetic, physically active’, ‘able to do a lot, work, get out and about’ and ‘never ill, no disease, never see a doctor’. Respondents with A level education and above, particularly females, were more likely to mention ‘physically fit, strong’.

Others have defined health in the context of interference with the performance of normal social roles and activities (Wright 1990), or of ability to work (Twaddle 1969). The various lay definitions of health reported by Baumann (1961) pointed to three main orientations: a general feeling of well-being, the absence of illnesses and the ability to perform social roles. Definitions vary according to age, sex, level of education (Cox et al. 1987), cultural group (Bowling 1994b) and socio-economic group, with those in the lower socio-economic groups defining health more negatively (Blaxter and Patterson 1982) and more likely to perceive the causes of health as being outside their control (Blaxter 1983; Pill and Stott 1985, 1988; Coulter 1987; D’Houtard et al. 1990). On the other hand, Cox et al.’s (1987) national Health and Lifestyle Survey in Britain reported that those in poor economic circumstances were not significantly more likely than those who were better off to associate poverty with ill health. They also reported that, while poverty and prosperity were seen generally as important determinants of health for society at large, it was rarely mentioned as a cause of ill health in the context of respondents’ own lives. This pattern of responses was also found in relation to individual behaviours (diet, smoking, exercise). These beliefs and definitions are pertinent to measuring subjective health status, and have
HEALTH-RELATED QUALITY OF LIFE

implications for the development of cross-cultural instruments.

These examples from the research literature on health beliefs illustrate that some groups of people do include social roles and social functioning in their definitions of health. However, for some decades there was much criticism of the previously described WHO (World Health Organization 1947, 1948) definition of health as a complete state of mental, physical and social well-being. This was criticized as utopian and unrealistic, particularly given the many studies of community morbidity which imply that having unreported (to the doctor) symptoms is a normal condition of the population (Dunnell and Cartwright 1972). The lack of realism in the WHO definition, and the impossibility – even undesirability – of achieving ‘perfect health’ is emphasized by Dubos’ view of health:

Every manifestation of existence is a response to stimuli and challenges, each of which constitutes a threat if not adequately dealt with. The very process of living is a continual interplay between the individual and his environment, often taking the form of a struggle resulting in injury or disease. The more creative the individual the less he can hope to avoid danger, for the stuff of creation is made up of responses to the forces that impinge on his body and soul. Complete and lasting freedom from disease is but a dream remembered from imaginings of a Garden of Eden designed for the welfare of man.

(Dubos 1959: 1–3)

Indeed, ‘health is more than just a biomedical phenomenon; it involves a social human being functioning in a social environment with social roles he must fulfill’ (Lerner 1973) – hence the importance of measuring health-related quality of life.

Gradually, definitions of health, like health-related quality of life, have moved away from a total disease model to one which incorporates health and well-being. The most widely used measures of broader health status, or health-related quality of life, reflect this definition, and incorporate physical functioning, psychological well-being and social support and activity items. Also, in recognition of the emphasis on the positive as well as the negative evaluations of health and consequences of illness, Hyland and Kenyon (1992) have developed the Satisfaction with Illness Scale, and Argyle et al. (1989, 1995) have developed the Oxford Happiness Scale.

QUALITY OF LIFE ASSESSMENT AND RESEARCH ON HEALTH CARE OUTCOMES

Why are we measuring health-related quality of life?

Life expectancy at birth in the developed world has increased over the past 150 years, although most of the increase has taken place during the first half of this century. Expectation of life, and expectations of a morbidity-free life at older ages, has also increased and has led to international attempts to measure health expectancy and disability-free life expectancy (Bone 1992; Robine et al. 1992; Colzve 1996). Debate focuses on whether the extra years of life are spent in good or poor health quality, with one side arguing, with some evidence, that chronic morbidity is being compressed into an increasingly shorter period before death, and the other arguing that there is an expansion of morbidity in old age. This interest in health expectancy, as opposed to simply analysing mortality rates, coincides with the more positive view of health measurement.

Purchasers of health care are increasingly expected to allocate health care resources on the basis of the evidence of the effectiveness of health care interventions – i.e. on the basis of health outcomes of intervention in the broadest sense. Health outcome indicators, defined as indicators of change in health status (Donabedian 1985) (a definition that does not take account of stabilizing and palliative treatments), have traditionally included information on avoidable mortality, survival rates, standardized mortality ratios, adverse reactions, complications, symptom relief, pain and physical and biochemical markers of recovery. Purchasing debates in health care now focus more on health care costs in relation to ‘health gain’ or broader health benefits from the treatments and interventions that are contracted for (Øvretveit 1993; see also overview by Normand and Bowling 1998).

Quality of life as a measure of outcome redirects attention towards consideration of the impact of...
the condition, and the treatment, on the patient’s emotional and physical functioning and lifestyle. Quality of life indicators help to answer the question of whether the treatment leads to a life worth living, by providing a more patient-led baseline against which the effects of the intervention can be evaluated. The concept is not new to clinicians; many early health-related quality of life measures were developed by clinicians some decades ago, although they were often crude and limited to functioning (Karnofsky et al. 1948; Steinbrocker et al. 1949; Katz and Akpom 1976a, 1976b).

**The fashion for health-related quality of life assessment**

Quality of life assessment as a supplement to more objective clinical indicators is becoming more topical in view of the increasing questioning of the effectiveness and appropriateness of many existing medical treatments and methods of organizing health services (Brook et al. 1983, 1990; Winslow et al. 1988; Enthoven 2000). It also represents a paradigm shift in the approach to the operationalization and measurement of health outcomes (O’Boyle 1997). The measurement of the health outcome of clinical interventions has become a cornerstone of health services research; it is essential for the assessment of the effectiveness and appropriateness of health care interventions (Brook 1990). Quality of life assessment is increasingly popular among pharmaceutical companies, with most reporting that they have used some type of quality of life instrument in their clinical trials of drugs (Luco et al. 1989). Information about broader patient outcomes is also required in order to guide and empower patients about appropriate health care.

Health care outcome measures must, however, relate to the objectives and the known risks and benefits of the intervention or care (Frater 1992). The Department of Health (1992) suggested the following should be incorporated in outcome assessment: survival rates, symptoms and complications, health status and quality of life, the experiences of patients and their carers, and the costs and use of resources. As their report continued: ‘Many health technologies are intended to improve general health and the quality of life, so it is important to measure patients’ subjective experiences of illness and the care they receive’.

Quality of life assessment has become an industry in itself. The term was introduced by Medline as a heading in 1975, and accepted as a concept by Index Medicus in 1977; this was followed by acknowledgement and acceptance by various scientific bodies (Bech 1992). Since the 1970s, there has been an explosion of interest in the subject, with an increasing number of citations of quality of life in the medical literature (de Haes and van Knippenberg 1985, 1987; Cella and Tulsky 1990; Aaronson et al. 1991b). There has also been a proliferation of study groups, conferences, bibliographies and special journal issues (e.g. Advances in Nursing Science in 1985; Journal of Chronic Diseases in 1987; Psychotherapy and Psychosomatics in 1990; Medical Care in 1990; WHOQoL Group 1994a and Social Science and Medicine in 1995). In 1991, a journal entitled *Quality of Life Research* was published by Rapid Communications of Oxford, devoted to the study of health-related quality of life. Bardelli and Saracci (1978) reviewed all the clinical studies published in six major cancer journals between 1956 and 1976, and reported that in less than 5 per cent an attempt had been made to measure some aspect of quality of life. The situation changed from the late 1970s and the number has since exploded, although the methodology has frequently been relatively crude (Najman and Levine 1981; Fayers and Jones 1983; Aaronson et al. 1988; Hollandsworth 1988; Falotico-Taylor et al. 1989; Padilla et al. 1992b). A search of Medline by O’Boyle (1997) using the term ‘quality of life’ yielded eight publications in which the term was mentioned in 1974, but this had increased to 1482 for 1996. Medline showed many more than this for 2000. However, such a global, and fashionable, key search term is relatively useless as it is unclear what relevance most publications had to quality of life.

Some investigators remain sceptical about using patient-based indicators in outcomes or descriptive research because of their subjective nature (e.g. self-reported health status and perceptions of well-being), quite apart from the poor quality of their operationalization and measurement (Muldoon et al. 1998). However, it is important to include patient-based indicators in assessments of outcome precisely because they do not necessarily correlate with objective measures of patients’ level of physical functioning (Evans et al. 1985). There are many
examples of studies which have reported low levels of agreement among doctors in relation to objective indicators (Wigton 1988), as well as between doctors’ and patients’ assessment of outcome. Subjective indicators (based on self-ratings) of quality of life, and health-related quality of life are increasingly popular due to the recognition of the importance of patient satisfaction and of how individuals feel, rather than what statistics imply they ought to feel. Clinical indicators of outcome are no longer sufficient alone, particularly in view of the debate about quality of life in terminal conditions (see Jennett 1976). The argument in favour of measuring quality of life as an outcome of clinical interventions has been neatly summed up by Ebbs et al. (1989). They argued that, in the case of chronic conditions with only partial or temporary amelioration of symptoms, a comprehensive evaluation is needed in order to determine the merits of treatment. Sullivan (1992) also pointed out that with incurable conditions the realistic goal of care is to provide a life that is as comfortable, functional and satisfying as possible. This argument can be extended to intervention for more acute conditions where the benefits of the treatment are complex and/or uncertain. The primary goals of medical treatment and care are to increase survival and to add quality to the survival.

UTILITY ASSESSMENTS AND QUALITY OF LIFE ASSESSMENT

Utility measurement by health economists deals with valuations of health states – it is concerned with the preference for (desirability of) a particular health state or outcome, and results in a summary score which equals a single value. This single value facilitates comparisons of treatments in cost-utility studies. It is conceptually and technically different from broader quality of life outcome assessment. There is some blurring of the boundaries between the measures used by health economists and those used by investigators of health-related quality of life treatment outcomes. This has occurred because the latter investigators often use measures primarily designed to measure utilities (and which were designed to produce a single value of a health state to inform cost-effectiveness analysis) in health research. Also, some investigators of existing broader, multi-dimensional health status scales are attempting to adapt them to produce a single value for use in utility studies (Brazier et al. 1996). The more recently developed Euroqol, designed to produce a single value, is popularly used in broader outcome assessment because of its relative brevity (Euroqol Group 1990; Kind 1996). Each of these approaches is presented briefly next in order to clarify the distinction between them and the broader measures of health and disease-specific related quality of life reviewed in this volume and in Measuring Health (Bowling 1997a).

Utility assessments

In response to the interest of health policy makers in cost-benefit formulas for purchasing decisions, a main focus among health economists has been on the development of utility assessment. The researchers initially responsible for applying utility theory to the health care field include Kaplan, Bush and their colleagues in the USA (Kaplan and Bush 1982; Kaplan et al. 1984), Weinstein and his colleagues also in the USA (Weinstein et al. 1980), Williams and colleagues in England (Kind et al. 1982) and Torrance and his colleagues in Canada (Torrance et al. 1972, 1982; Torrance 1986, 1987). A generic utility questionnaire collects data on quality of life in a way that enables it to be applied to a pre-scaled matrix of health state preferences (values).

Quality Adjusted Life Years (the QALY)

Cost–utility studies need a common measurement of health outcome. With the QALY the cost of an intervention is related to the number of QALYs, a concept first introduced by Weinstein and Stason (1976). A QALY is a year of full life quality. Poor health may reduce the quality of a year (e.g. from 1 to 0.5). In QALYs, improvements in the length and quality of life are amalgamated into one single index. Each life year is quality adjusted with a utility value, where 1 = full health. The utility value aims to reflect the health-related quality of life. Different types of medical interventions are then compared by calculations of costs per gained QALY (Williams 1985b; Office of Health Economics 1989). QALYs are not really measures of
quality of life but measures of units of benefit from a medical intervention, aiming to reflect the change in survival with a weighting factor for quality of life.

QALYs have been severely criticized (Carr-Hill 1989; Carr-Hill and Morris 1991) for their lack of validation. There are inadequate data on their conceptual basis, and the data on reliability and validity that do exist are based either on unrepresentative ‘convenience’ samples or random samples of the population with high rates of non-response, partly because of the complexity of the task. The application of QALYs to health policy making has led to ethical concerns, particularly given their crude state, over reliance on valuations by health professionals and non-representative groups of people (Grimley Evans 1992). There is no evidence that the judgements determining the QALY for a particular condition bear any relationship to real judgements faced by patients suffering from that condition.

QALYS can be derived using several different methods (e.g. the Rosser Index of Disability (Rosser and Watts 1972), standard gamble, trade-off and rating scale techniques (Torrance et al. 1972, 1982; Torrance 1986, 1987)).

The approach of Kaplan and Bush
Kaplan and Bush (1982) and Kaplan et al. (1984) adopted a slightly different approach. They placed individuals with given health states into categories of mobility, physical activity and social activity, and then classified the symptoms and health problems that they had on a given day. Four hundred case descriptions were then compiled to illustrate the combinations of functional levels, symptoms or problems. The scale includes death. Random samples of the public gave preference ratings to the descriptions, and weights were derived for each level of mobility, physical activity, social activity and symptom or problem. It is based on a measure of functional status. A utility value was assigned to each functional level, and questionnaire responses were used to assign subjects to one of a number of discrete function states. Kaplan’s Index of Well-being Scale is a single-score scale that has been developed out of this methodology, and has been used widely as a health status measure and methodological tool for health care decision making (Kaplan et al. 1976; Bush 1984). The scale quantifies the health output of a treatment in terms of years of life, adjusted for the changes in quality that it is responsible for.

Rosser Index of Disability
In the UK the basic measurement tool underlying the calculation of the QALY is the Rosser Index of Disability (Kind et al. 1982; Williams 1985a; Williams and Kind 1992), which was originally developed as an indicator of hospital performance (Rosser and Watts 1972; Rosser 1992). The Rosser Index is based on the concept of a health index, and subjects are graded into one of eight areas of disability, from none to unconscious, with each state being graded on four levels of distress, from none to severe. States are scored on a scale ranging from 0 at death to 1 = healthy (with negative values for states valued to be worse than death, such as vegetative states). The scaling and valuation of health state techniques of the scale have been clearly described and reviewed by Wilkin et al. (1992). Kind and Gudex (1991) have also developed a survey questionnaire – the Health Measurement Questionnaire – designed to collect self-reported information capable of being processed to yield Rosser disability/distress categories.

Standard gamble, trade-off and rating scale techniques
In Canada, the McMaster group’s approach was that QALYs should capture the subject’s preferences concerning the length and quality of life, and proposed three approaches to the derivation of utility values (Torrance et al. 1972, 1982; Torrance 1986, 1987). The first is the standard gamble technique, in which subjects are asked to choose between their own health state and a gamble that they might die immediately or achieve full health status for the remainder of their lives (e.g. in relation to a specific treatment choice). The second is the time trade-off technique (Torrance et al. 1972), in which an individual is asked to consider a health state that is to last for a fixed period of time. A new procedure will give the individual normal health for a shorter period of time, but they will probably die or be severely disabled at the end of that time. The person is asked to “trade-off” the time with reduced
capacity for living with normal health for a shorter period of time. The time spent in normal health is varied until the point of indifference is found. Subjects may also be asked to evaluate the equivalences between the numbers of people helped by different treatments – how many people in state B must be helped to provide a benefit that is equivalent to helping one person in state A (see review by Kaplan et al. 1993a). The time trade-off technique has been reported to be more reliable and valid than the standard gamble technique (Dolan et al. 1993).

Third, Torrance et al. (1982) developed a rating scale approach, specifying six attributes that should be included in a health state: physical function, emotional function, sensory function, cognitive function, self-care and pain. Each of these attributes is given several levels of gradation, and the characteristics of a given health state would include a description of functioning, self-care and pain associated with that state. The descriptions can be shown on video or in written vignettes. The ratings assigned to the descriptions are placed on a visual analogue scale ranging from 0 (death/least desirable) to 100 (healthy/most desirable). Multiple Attribute Theory is then used to determine the value for each level of the attributes and the utility value of an associated health state.

The Euroqol

The aim of the Euroqol is to provide a standardized non-disease-specific survey instrument for describing health-related quality of life, and to generate a single index value for each health state (Euroqol Group 1990). It contains five questions relating to physical functioning, mental health and pain, and a self-rating of health on a ‘thermometer’ VAS. While its construct and convergent validity is adequate, it has suffered from moderate to low response rates in a number of population surveys, is highly skewed and has relatively poor sensitivity (Brazier et al. 1992, 1993a, 1993b; Bowling 1998), particularly in relation to disease-based outcomes research (Casellas et al. 2000; Selai et al. 2000). It has inherent methodological design faults in the wording and range of its response scales, and it is possible that the length of the thermometer scale (0 to 100) is biasing. It is frequently criticized (Carr-Hill 1992; Jenkinson and McGee 1998). However, it is widely used internationally, and the development and refinement of the instrument is still in progress (Dolan et al. 1995).

Limitations of utility assessments of quality of life states

The disadvantages of all these methods are their cost, the requirements for skilled interviewers and complexity (leading to reliance on non-random or unrepresentative samples of the public). Although the Euroqol can be self-administered and has been widely used in postal approaches, inconsistencies in responses have been found suggesting that some members of the public find it too difficult (Carr-Hill 1992).

One of the main debates surrounding the use of utility assessment techniques is who should provide the utility values – the general public, health care providers and/or patients and their families themselves? While patients’ assessments are undoubtedly important, proponents of obtaining utility values of the public and health professionals argue that patients’ values may change over the course of an illness and thus their utility values would not be stable. The ethical concern is that the judgements of quality of life and utility assessments are poorly understood (Cohen 1990). A review by Kaplan et al. (1993a) has pointed to evidence from psychology experiments that suggest that methods commonly used for economic analysis do not represent underlying true preferences (Kahneman and Tversky 1983).

WHO SHOULD RATE QUALITY OF LIFE?

Discrepancies between doctors’ and patients’ assessments

Patients own first-hand views are also essential in health-related quality of life outcome assessments. Research suggests that there are wide discrepancies between patients’ and doctors’ ratings of outcome after specific therapies (Orth-Gomer et al. 1979; Thomas and Lyttle 1980; Jachuck et al. 1982; Slevin et al. 1988). The case against observer ratings of another person’s quality of life, which is then taken as a proxy indicator, has been made by several empirical studies comparing doctors’ and
patients’ ratings, and by the increasing literature on discrepancies in treatment preferences. This literature serves to indicate that the patient’s feelings, values and opinions cannot be assumed. Jachuck et al. (1982) reported that, while all treating doctors in their study rated their patients’ quality of life as having improved after they started anti-hypertensive treatment, three quarters of the patients’ relatives thought that it was worse; 8 per cent of the patients felt worse and 44 per cent felt the same (the remainder felt they had improved). Poor correlations between professionals’ and patients’ own self-assessments were also reported by Padilla et al. (1981). Large discrepancies in assessments between patients and doctors have been reported by Slevin et al. (1988), who compared their assessments using the Karnofsky Performance Scale, the Spitzer Quality of Life Index and the Linear Analogue Self-Assessment (LASA) Scale.

Relatively few investigators have based their operational definitions and measurement decisions on what the public say are the relevant domains of quality of life. While some generic and disease-specific measures of health-related quality of life have included lay persons’ perceptions of the effects of symptoms on their lives (e.g. Bergner et al. 1981; Hunt et al. 1986), investigators have more commonly turned to the existing literature as their starting point (Stewart and Ware 1992). Some scales, particularly in mental health and oncology, bypass the patient’s perspective altogether and base ratings on the perspective of a staff member or a ‘significant other’ (e.g. relative or friend).

In health care, the debate sometimes revolves around whether the quality of life assessment should be made by the patient or by a health professional (e.g. a doctor). Objections to physician ratings include the argument that while patients may judge their quality of life to be low, they may nevertheless value their lives as precious (Brock 1993). Some clinicians object to patients’ ratings on the grounds of their subjectivity. This subjectivity, as it reflects the patient’s point of view, should be viewed as their strength. There is now general recognition among health services researchers that measures of health outcome should incorporate the patient’s perspective, not simply in terms of whether or not the treatment was a success, but more globally in relation to perceived mental and physical well-being as a consequence of an intervention. A person can feel ill without medical science being able to detect any apparent disease. A person’s ill health is indicated by feelings of pain and discomfort, or perceptions of change in usual functioning and feeling. It has been known for many years that the utilization of health services is more closely associated with the perception of symptoms and people’s feelings than with their actual medical condition (Mechanic 1962; Goldberg and Huxley 1980). Measures of health outcome need to take account of both the traditional disease model (pathological abnormality indicated by a set of symptoms and signs) and the patient’s perspective.

The implication is that patients should complete a questionnaire about their quality of life themselves, or the questionnaire should be administered to them by a trained interviewer. ‘Significant others’ (e.g. relatives) and health care professionals should only complete ratings where their perspective is also required, and where the patient is too frail or ill to be questioned. However, a number of scales require a health professional to complete a questionnaire on behalf of the patient. These include Spitzer et al.’s (1981) Quality of Life Index, the rating scales of dependency, work and school performance, family and non-family relationships used by Horowitz and Cohen (1968), Taylor and Falconer (1968) and Rausch and Crandall (1982) in their studies of patients with temporal lobe seizures, and various other physical performance indicators, such as the Barthel Index (Mahoney and Barthel 1965) and the Karnofsky Performance Scale (Karnofsky et al. 1948). On the other hand, although patient-based, self-report questionnaires are the ideal to be aimed for, it is not denied that there is also a practical need for supplementary, indirect measures of quality of life when the patient is too frail or ill to respond. The criticisms of these have arisen because some investigators have selected the latter rather than the former for primary routine use.

PATIENTS’ PREFERENCES

Patients’ preference assessments and decision making

Some consideration should be given to patients’ preferences for treatments in relation to their likely effect on their health and health-related quality of
life. Few outcome scales incorporate this aspect of the patient’s view, but it is arguably important in understanding patients’ perceptions of, and satisfaction with, their health outcome following treatment, and their preferences for different types of invasive or non-invasive treatments.

A preference is an attempt to weigh up, consider and express a consistent value for alternative choices of action (Till et al. 1992). A study of people’s preferences involves the assessment of their attitudes (preferences) towards specific options, after informed deliberation on the risks and benefits, and in the light of people’s own value systems (Entwistle et al. 1998). The preferences of people for alternative treatments need to be measured in order to empower patients by taking their views into account, thereby improving opportunities for choice, and leading to a more rounded, patient-based body of knowledge on clinical appropriateness. Such preference assessments are distinct from the preference assessments of different health states in the development of utility measures (Blischke et al. 1975).

There have been few insightful attempts to identify aspects of procedures that make them more (or less) acceptable to patients (Lippman et al. 1985), and no attempt to identify the social and cultural variations in the views and preferences expressed, and the values underlying them. Research on preferences requires careful design to avoid complexity (Frogberg and Kane 1989), and question framing effects (McNeil et al. 1982), and to sample people for whom the options presented are relevant, while in a hypothetical context. One technique of eliciting people’s preferences, and their intensity, is conjoint analysis, which has been used mainly in market and transport research, and more recently in health research on preferences for treatments, as well as in the development of health status measures (Harwood et al. 1994; Ryan 1996; Ryan and Hughes 1997; Ryan et al. 1998). With conjoint analysis, respondents are presented with hypothetical scenarios, and asked to express preferences between alternatives. Although it can be complex and taxing for respondents, it has been shown to be acceptable to people and achieves good response rates in postal as well as interview surveys (Ryan 1996).

Understanding risk is likely to be important for assessing preferences and for both patient and clinical decision making (Wennberg et al. 1988). People will vary in their understanding of personal risks to health and in risk-taking attitudes, and this is likely to influence their preferences for either conservative management or invasive intervention. A post-operative mortality risk of 5 per cent, or even 25 per cent, may not be unacceptably high to the patient with risk-taking attitudes, especially when the prognosis without surgery is poor in relation to quality of life and/or survival. Psychological research on risk locates this concept within the individual and focuses on the concept of perceptions of control (Conner and Norman 1995), although the use of untested measures has been widespread, and it has under-emphasized the effects of social structure (e.g. family, friends, socio-economic status, workplace) on behaviour. Most health research on risk taking and perceived control relates to preventive behaviours and lifestyle among younger people, and has focused on levels of ‘unrealistic optimism’ or ‘optimism–pessimism bias’ (Weinstein 1984). Little is known about most patients’ understanding and perceptions of the risk to their quantity and quality of life imposed by their condition and its treatment. Nothing is known about the effects of perceived risk, control, optimism bias, value attached to health, health status and social characteristics on preferences for treatment. Such variables are relevant to patients’ perceptions of their treatment and outcome, although it should also be noted that not all patients may want to exercise a choice, and not all patients who do make choices opt for the choice expected by clinicians (Wolberg et al. 1987; Wilson et al. 1988; Fallowfield and Hall 1991; Wolberg 1991).

In relation to decision making about quality vs quantity of survival, attitudes diverge within patient groups. Decision making is never clear-cut, as a brief review of some of the literature by Byrne (1992) emphasizes. While research on patients across disease conditions has reported that people generally agree about which health states are worse than death (e.g. permanent coma, severe dementia, loss of essential functional ability such as being able to feed oneself – Pearlman et al. 1993), there is a great deal of variation among doctors in relation to treatment preferences (Mackillop et al. 1992), as well as discrepancy between clinicians and patients on the treatment of choice (Mackillop et al. 1986,
DISEASE-SPECIFIC, DOMAIN-SPECIFIC AND GENERIC MEASUREMENT SCALES

It is important to clarify the different types of instrument. A generic scale is useful when the relevant variables are covered and when investigators wish to make comparisons of results between different diseases and conditions. A domain-specific scale is required when the area covered is of particular relevance to the study and its hypotheses, and where generic and disease-specific scales neglect that area. A disease-specific scale is used when disease or condition-related attributes need to be assessed, and greater sensitivity to the clinical condition under consideration is required. It is often recommended that a generic measure should be used alongside a disease-specific measure in clinical trials, in order to address both clinical and broader policy questions, and to detect unexpected positive or negative effects of interventions (Bombardier et al. 1995; Fitzpatrick et al. 1998). This does carry the problems of respondent and research burden, the need for an increased number of statistical tests thus increasing the likelihood of obtaining statistical significance by chance (Fitzpatrick et al. 1998). Fitzpatrick et al. (1998) have described the requirements for selecting and judging the appropriateness of measurement scales, from psychometric criteria and their clinical significance to the importance of ensuring that a body of evidence exists to justify the selection of measures that have social significance (i.e. measures which include domains and items that reflect areas of importance to patients).

Disease-specific scales

Disease-specific measurement scales have the aim of being more clinically and socially significant in relation to specific conditions – of being able to discriminate more finely between patients’ levels of severity of condition, and of being more sensitive to their clinical outcomes. In choosing a measurement instrument, or set of measures, key questions to consider are whether a disease-specific and/or a generic measure is needed, and whether either requires supplementation with single domain measures that are important to the study aims (e.g. depression). There is little point in utilizing a health status measure alone if it is unlikely to detect the effects of the treatment in question, or the symptoms specific to the condition. In addition, some measure of disease severity will also be required. Several severity and co-morbidity indexes have been developed (Kellerman and Hackman 1988; Parkerson et al. 1993a), although these are often fairly crude (Linn et al. 1968) and clinical investigators prefer to rely on biomedical indicators. Most health services researchers work closely with clinicians to ensure that the appropriate disease-specific outcome and severity indicators have been included in their batteries.

Clearly, criteria for assessing outcome of care will vary for different disease syndromes. A universal questionnaire to elicit the relevant information for a number of conditions would require a questionnaire of enormous length (Goligher 1987). Disease-specific quality of life scales are needed not simply for greater brevity, but to ensure sensitivity to sometimes small, but clinically significant changes in health status and levels of disease severity (for good illustrations of this point in relation to head and neck cancer, see Berg et al. 1976; Dhillon et al. 1982; Morris 1990). The domains of focus in relation to outcome assessment across the different diseases and conditions vary considerably, and this diversity is reflected in the disease-specific health-related quality of life scales. For example, in psychiatry a main area of focus is on the arrangements for care, and consequently measurement scales reflect a needs-based approach (e.g. housing, money, shelter and so on, from the perspective of both patient and carer). In this sense the scale used will reflect the tradition of social indicators research. In oncology, the toxicity of treatment and the aggressive nature of the condition are of importance; in relation to respiratory and cardiovascular conditions, the symptoms and the restrictions they impose on everyday life and activities are relevant; in rheumatology it is limitations on activities and functioning, as well as pain, that are the focus for measurement. These variations also reflect the evidence that different...
areas of life are affected by different conditions (Bowling 1996a, 1996b).

However, while different procedures require appropriate disease-specific quality of life outcome measures, there can still be a central core of quality of life questions common to a wide range of disease areas. It is easy to assume that disease-specific measures per se will be more sensitive indicators of patient outcome than generic measures. However, a study of osteoarthritis patients by Kantz et al. (1992) reported that while knee-specific function measures and pain scales were more specific than the generic health status scale, the SF-36 (Ware et al. 1993), among patients with other co-morbid conditions, the SF-36 physical functioning sub-scale, plus knee-specific adaptations of that scale were just as specific as the disease-specific scales. Moreover, the disease-specific measures failed to distinguish between treated and untreated patients. A combined approach to outcome assessment, using both disease-(or condition-) specific and generic measures is preferable where a broad disease-specific quality of life instrument has not been satisfactorily developed. The generic measure of choice across many diseases is increasingly the SF-36. As Ware (1993) concluded, because the SF-36 is short, well tested, and population norms exist, it may constitute a good generic core for use along with disease-specific outcome measures. The potential use of this scale as a generic core will be discussed more fully within the appropriate chapters of this book (see Chapter 8 for review).

**Domain-specific scales**

As the reviews of disease-specific indicators of quality of life will show, they can often be criticized for being too narrow in focus, while neglecting the measurement of important outcome and modifying variables (e.g. social support, adjustment, coping, life satisfaction, self-esteem, depression and other domains). Some investigators supplement their disease-specific scales with domain-specific measures. Generic and domain-specific measures have been reviewed by several authors (e.g. Wade 1992; Wilkin et al. 1992; McDowell and Newell 1996; Bowling 1997a).

The domain-specific areas of interest will vary according to how the condition and its treatment affect the patient. Thus measures of psychiatric status will, for some diseases and conditions, necessitate the inclusion of a memory test, as well as a depression scale. The measurement of physical functioning may be restricted to global categories ranging from fully functioning to bed-bound for more dramatic conditions and interventions where great changes are expected, but may need to be more refined and sensitive (i.e. at the 'less restrictive' end of the scale) in the case of more moderate cases.

Other domains of potential relevance may include occupational and social role functioning, including maintenance of social relationships and activities (Kaplan 1985; Bowling 1994a, 1997a; Bowling et al. 1994). Psychological well-being can be important, and includes happiness/satisfaction, self-mastery, expectations, personal control, adjustment and coping ability (influencing or modifying variables), well-being, emotional support and interaction (outcome and modifying variables), self-concept and self-esteem, body image and somatic comfort (Lewis 1982, 1989; Young and Longman 1983; Schipper et al. 1984; Schipper and Levitt 1985, 1986; Ferrans and Ferrell 1990).

Some argue that a definition of quality of life should reflect the pre-illness situation of the patient; others argue that it should reflect the attainment of an ideal quality (Calman 1984; Schipper and Levitt 1986). There is, of course, an individual judgement involved and people differ in their values. In medicine, the goal is to return patients to normal lives, rather than attain the ideal, and measurement scales should reflect this (Selby and Robertson 1987). Measures of quality of life should therefore reflect the range of normal activities that have been potentially affected by the condition and treatment. It is possible to obtain good response and completion rates with a comprehensive questionnaire (Sadura et al. 1992).

The specialized scales measuring specific domains are often long, and have to remain fairly lengthy in order to retain their sensitivity and psychometric properties. A caution must be made here in relation to the use of batteries of several scales, for example as a supplement to a brief disease-specific scale. Although disease- and condition-specific scales of quality of life are easy to criticize for their brevity, and their often narrow scope, no study can hope to measure every relevant domain,
and most disease-specific scales are limited to the most essential. Domain-specific scales should be used when the area is of particular interest to the investigator, and the disease-specific (or generic) scale selected for use neglects that domain. They should not be included for the sake of having a comprehensive battery, without any theoretical or methodological justification. The danger is the production of a battery that contains too many – and also overlapping – items, which becomes tiring to respondents, expensive to administer and analyse, and produces an overwhelming amount of data which may not always be helpful.

**Generic scales**

Measures which implicitly or explicitly aim to tap health-related quality of life are usually referred to as broader measures of health status. They should encompass the dimensions of physical, mental and social health. Investigators have tended to supplement generic health status measures with specific disease items. They have used generic measures in order to make comparisons with other conditions, to broaden their outcome indicators, and because of the slow development of disease-specific questionnaires. On the other hand, generic instruments have an important constraint as they are unable to identify the condition-specific aspects of a disease that are essential for the measurement of outcome (Hutchinson and Fowler 1992). Generic measures will always require supplementation with disease-specific measures in order to detect important clinical changes (Guyatt et al. 1986). McKenna (1993) has argued that the role of generic measures in health services and clinical research will diminish as more disease-specific measures are produced which can focus directly on the research issues. As McKenna points out, the use of disease-specific measures avoids asking irrelevant questions of respondents and maximizes the chance of detecting clinically significant changes, which is essential in clinical and policy-orientated research. However, it is unlikely that they will obviate the need for generic measures, or at least a generic core, as long as comparisons across disease groups within and between specialities are required. The most popular generic measures include the Rand batteries, in particular the increasingly used SF-36 (Carter et al. 1976; Deyo et al. 1982; Bergner 1984, 1993; Deyo 1984a; Stewart and Ware 1992; De Bruin et al. 1993; Ware et al. 1993), the Sickness Impact Profile (Bergner et al. 1981), the Nottingham Health Profile (Hunt 1984; Hunt et al. 1986) and the McMaster Health Index Questionnaire (Chambers et al. 1976, 1987; Chambers 1993). A popular and promising generic measure that is being developed for use in primary medical care is the Dartmouth Coop Function Charts (Nelson et al. 1987, 1990; McHorney et al. 1992; see review by Wilkin et al. 1992). However, while health-related quality of life measures are increasingly used in health services research, several reviews of clinical trials have indicated that they are still underused by clinicians (e.g. Guyatt et al. 1989b).

**Satisfaction with care and outcome**

This domain is listed separately because it is an important and neglected component of disease-specific and generic scales or batteries of scales. A person’s degree of satisfaction with their health status and outcome, and degree of fulfilment of expectations of the treatment, should be included. It is arguably an outcome as well as a process indicator of the quality of care. It is usually neglected because the detection of dissatisfaction has foiled all patient satisfaction questions and questionnaires. Reviews of the problems inherent in measuring patient satisfaction can be found in Roberts and Tugwell (1987), Ware and Hays (1988), Cartwright (1989), Fitzpatrick (1990), Wilkin et al. (1992) and van Campen et al. (1992). Davies and Ware (1991) have developed a promising Consumer Satisfaction Survey Questionnaire (including a ‘Visit-specific Satisfaction Questionnaire’) and user manual. The ‘visit-specific’ questions were developed from the Rand Medical Outcomes Study Questionnaires. Readers are referred to these sources where their measurement scales of choice do not address patient satisfaction issues.

**Computerized item banking**

The issue of which domains to include in batteries of measurement scales may be less problematic and taxing in the future if computerized item banking is developed, as in educational research (Uttaro and Lehman 1999). With this, items measuring all possible life areas of importance to people and also
those that are important to measure clinically are entered onto a computer database; the computer is programmed to select the relevant questions for each respondent based on their responses to a set of initial screening questions about their condition, its effects on their life, their circumstances and their perceptions of the important areas of life. The aim will be to reduce the length and burden of the questionnaire for respondents, while retaining its sensitivity and a scoring system that reflects individuals’ concerns and enables comparisons between people.

METHODOLOGICAL ISSUES

Other texts and specialized review articles have described research methodologies, question and item response design, scaling methods and a wide range of statistical issues – thus the principles of research will not be repeated here. Readers are referred elsewhere for information on these methodological issues (e.g. Webb et al. 1966; Oppenheim 1968; Blalock and Blalock 1971; Moser and Kalton 1971; Andrews and Crandall 1976; Pocock 1983; Sudman and Bradburn 1983; Barker and Rose 1984; Kleinman 1986; Armitage and Berry 1987; Cella and Cherin 1987; Selby and Robertson 1987; Bellamy 1989; Spector and Thompson 1991; Fletcher et al. 1992b; Streiner and Norman 1995; McDowell and Jenkinson 1996; McDowell and Newell 1996; Bowling 1997b).

However, an introduction to some of the basic psychometric concepts, as well as operationalization and level of data, is required for the interpretation of the descriptions of scale reliability and validity in the text.

Operationalization

There has always been considerable debate among social scientists about ‘operationalism’, and the extent to which the gap between theory and empirical research can be bridged (for a discussion, see Blalock and Blalock 1971). However, for the concept of quality of life to have any value in descriptive or health outcome research, it must be decided what is to be measured and the agreed concepts need to be defined and translated into an observable form. This is essential if quality of life is to be researched in any coherent and scientific way – and necessary before a quality of life scale can be said to have content validity.

Scepticism has often greeted attempts to measure quality of life in clinical studies (Muldoon et al. 1998). This is unsurprising given that most investigators in the past have not defined or operationalized their terms (Kaplan and Anderson 1990). Many assume that the concept refers to physical functioning, and/or psychological and mental status or symptom levels. Many scales purporting to be, or used as, quality of life scales are, in fact, simply health status scales, and they are also frequently single-domain scales (e.g. level of physical functioning or mental health). Because most so-called ‘quality of life’ instruments, and their proxy measures, do not attempt to measure the philosophical, external or environmental aspects of quality of life, but focus on health-related areas, Fitzpatrick et al. (1998) have consequently labelled the term ‘quality of life’ as misleading, and rejected its use in favour of the term ‘patient-based outcome measures’. Annas (1990) also argued that the term ‘quality of life’ is so misused that it should be banished from our lexicon.

Reliability and validity

Measuring the quality of life of patients is inevitably difficult. Questions about the sensitivity, reliability, validity and generalizability of data continue to be raised because of the complex nature of diseases, treatments (Smart and Yates 1987), expected recovery times (Bardsley and Coles 1992), and, of course, the concept of quality of life itself.

Whichever measure, or battery of measures is selected, all instruments should satisfy the criteria for reliability and validity. Unfortunately, many instruments have been developed ‘ad hoc’ and many studies are based on instruments with weak psychometric properties, or little evidence of their reliability and validity (Coste et al. 1995). Validity is the extent to which an instrument measures what is intended. Reliability is the extent to which measurements on the same respondent are similar on repeated applications of the measure at different times. Another important property of quality of life scales when used as outcome measures is their responsiveness to change, particularly clinically important changes. There is an unresolved debate about
whether responsiveness is an aspect of validity (Hays and Hadhorn 1992). Several texts describe these concepts in more detail (e.g. Carmines and Zeller 1979; Kline 1986; Streiner and Norman 1995; Bowling 1997b). They are now described briefly here.

Reliability

Reliability refers to the ability to produce consistent results, and consistent results on different occasions, when there is no evidence of change. This is tested by test–retest reliability (administration of the scale on different occasions to the same population), internal consistency (measurement of the same concept by different scale items), and inter- and intra-rater reliability (consistency of a measure when administered by different interviewers or the same interviewer on different occasions). It is important that the setting of the administration is the same on each occasion to prevent contamination (Ziebland and Fitzpatrick 1992).

A few of the appropriate statistical techniques for assessing reliability are referred to below, in order to clarify the chapters in this volume where values are presented. Readers are referred to the references and to appropriate methodological and statistical texts for further details, and for details of other tests as well as appropriate levels of statistical significance in relation to sample size (e.g. Pocock 1983; Armitage and Berry 1987).

Internal consistency involves testing for homogeneity. Tests assess the extent to which individual items are inter-correlated and the extent to which they correlate with overall scale scores. Internal scale consistency is usually tested with inter-item correlations and with Cronbach’s alpha (Cronbach 1951), based on the average correlation among the items and the number of items in the instrument (values range from 0 to 1). Although some regard 0.70 as the minimally acceptable level for internal consistency reliability, others accept the level ≥0.50 as an indicator of good internal consistency, as well as of test–retest reliability (Cronbach 1951; Ebel 1951; Helmstater 1964; Cronbach et al. 1972; Nunnally 1978). It is accepted practice to reject scale items below $r = 0.20$ (Kline 1986). Streiner and Norman (1995) cite an alpha of 0.85 or above as indicating adequate internal consistency.

The items of the measure can also be split into two halves and the alphas for the alternative forms compared (the Spearman–Brown formula uses this correlation to estimate reliability after adjusting for the number of scale items); or the scale can be divided into two groups and coefficients can be computed for each half and compared. Comparable coefficients confirm the consistency of the responses (Zeller and Carmines 1980; Williams and Wood-Dauphinee 1989).

The interpretation of alpha is that, for example, if Cronbach’s alpha is high (e.g. 0.80), then the implication is that the responses are consistent, and the sum of the item responses yields a score for the underlying dimensions that the item represents. If the items are adequately sampled from the quality of life domain, the sum of the responses should give a better indication of quality of life for the respondent than responses to any one item. A low coefficient alpha, on the other hand, indicates that the item does not come from the same conceptual domain (Williams and Wood-Dauphinee 1989).

Other tests of reliability include inter-rater agreement (the concordance, or reliability, of scores by different raters on a single occasion) and intra-rater agreement – the reliability of the rating by the same rater (of the same subjects) on different occasions (repeat testings, or test–retest). If the measure is categorical, then the most appropriate statistical method to employ is Cohen’s (1968) Kappa Test of Concordance. Kappa has a value of 0 if agreement is no better than chance, a negative value if worse than chance, and a value of unity (1.0) if agreement is perfect. Fleiss (1981a, 1981b) suggested that less than 0.40 is poor agreement, 0.40–0.59 is fair agreement, 0.60–0.74 is good agreement, and 0.75–1.00 is excellent agreement. Spearman’s Rho and Kendall’s coefficient of concordance, W, may also be considered (the latter in cases where more than two sets of ranked data are involved). However, simple correlation analysis is less appropriate for studying concordance as it makes no allowance for chance agreement. Pearson’s product–moment correlation is used for comparing continuous scores. Intra-class correlation coefficients compare the variance between subjects, the variance between raters and the variance between times with error variance (Shrout and Fleiss 1979; Fleiss 1986). An intra-class correlation coefficient of, for example, 0.80 or more indicates that the scale is highly reliable.

The stability of a measure over a period of time is measured by repeated administrations of the test
on subjects and domains not expected to change in their scores over a carefully defined time period. Test–retest reliability is generally assessed by correlations of the measure administered on the different occasions. A high level of consistency in response is desired. A test–retest correlation of 0.85 or more indicates that the instrument has an acceptably low level of random measurement error (McDowell and Newell 1996). However, the observed associations can sometimes be difficult to interpret – a low correlation may reflect a genuine change in health status, rather than poor reliability. Some investigators also feel that correlations are a weak measure of test–retest reliability, and recommend the use of confidence intervals to assess the size of the difference between the scores (Bland and Altman 1986; Ruta et al. 1994a).

Validity
Validity refers to the extent to which the scale measures the underlying concept of interest. Scales also need to have predictive validity (to be sensitive to real changes in condition) and to have social and clinical validity (relevance). That is, they should tap areas of social importance to the patient and should be responsive to the different disease stages and fluctuations due to the condition and its treatment over time (Aaronson et al. 1993). Validity is measured by testing for face validity (at face value, does it appear to be measuring what it is intended to measure; is it unambiguous and appropriate?), content validity (does the scale tap all relevant concepts of the attribute to be measured?) and criterion validity (the extent to which the measure correlates with a ‘gold standard’). Validity is often difficult to assess because there is no objective ‘gold standard’ measure of quality of life. Criterion validity encompasses concurrent validity (correlations with an existing measure of the same construct) and predictive validity (correlations against other measures to assess predictive powers). Quality of life scales are often reliant on predictive validity to demonstrate their psychometric properties. There is also construct validity where hypotheses are generated and the scale tested against a measure central to the hypothesis.

Construct validity is usually divided into convergent validity (tests for correlations with other indicators intending to measure the same concept) and discriminant validity (lack of correlations with unrelated indicators) (see Campbell and Fiske 1959).

Hyland (1992a, 1993) has cautioned against reliance on claims for content validity by scale developers. He pointed out that quality of life comprises several connected constructs, and any questionnaire may reflect one or more of these constructs. In the absence of an agreed theoretical description of quality of life, validity can be difficult to assess. Moreover, standardized questionnaires, with standardized scoring and weighting procedures, impose an external value system on respondents. Such measures may often satisfy conventional tests of reliability and validity, but may not be relevant to all individuals and thus lack content and social validity (Hickey et al. 1996). Research using individualized measures has indicated that areas of relevance to people are not always represented in standardized scales (e.g. in relation to people with AIDS, important areas which emerged related to spirituality, sexuality and issues relating to death; in relation to older people nominated areas included religion, independence, finances and happiness) (Browne et al. 1994; Hickey et al. 1996).

Factor structure
Questions that deliberately tap different dimensions within a scale cannot be expected to necessarily have high item–total or full item–item correlations. Therefore, factor analysis should be used to identify the separate factors within the scale. Factor analysis is a technique which defines a small number of underlying dimensions (factors that account for a high proportion of the common variance of the items), and in so doing demonstrates whether the items in the scales group together in a consistent and coherent way. Exploratory factor analysis is usually used in scale development in order to identify and discard items that are not correlated with the items of interest. Factor analysis is also used to confirm that scale items principally load onto that factor and correlate weakly with other factors. While small samples may be used in analysis, ultimately a confirmatory factor analysis should include a larger sample for assessment of stability (Comrey 1978; Loo 1983). A factor is considered important, and its items worthy of retention in the scale, if its
eigenvalue (a measure of its power to explain variation between subjects) exceeds a certain value. Jolliffe and Morgan (1992) state that this value should be 1.1, although 1.5 appears to be commonly taken.

**Translation and cultural equivalence**

Elaborate guidelines are available which stress the need for rigorous assessment of cultural equivalence when questionnaires are translated, and in the back translation (translation back to original language) procedures (Bullinger et al. 1993). There remains concern that such procedures may still fail to identify cultural values not initially anticipated in the culture where the instrument was first developed, and therefore may have little content validity (Fitzpatrick et al. 1998). The innovatory approach of the WHOQOL Group (1998a, 1998b) in their recent development of WHO’s Quality of Life Assessment instrument, intended for international use, was therefore to develop the concepts and items for inclusion in 15 study sites across the world. The disadvantage of this approach is the consequently long length of the scale (100 items).

**Response formats**

There are several possible response formats for use with closed questions. Most responses are usually in a categorical dichotomous format (e.g. yes/no), a categorical Likert scale (Likert 1952) (e.g. a series of ordered responses, such as ‘strongly agree, agree, neither agree nor disagree, disagree, strongly disagree’), or a visual analogue scale (VAS) (respondents are asked to place a mark on a numbered or unnumbered line, usually 10cm in length, on which opposing statements or descriptions are placed at each end). A range of other response scales exist (Thurstone 1927a, 1927b; Guttman 1944; Osgood et al. 1957; Cantril 1965; Andrews and Withey 1976). They have been described more fully by Bowling (1997b). The choice of response format is important as this determines the type of statistical techniques that can be applied.

**Measurement scales vs single item questions**

The superiority of multi-item measurement scales over single-item questions has been demonstrated by Davies and Ware (1981) and Manning et al. (1982). The quality of data collection is improved by the use of standardized, well-tested scales, with good psychometric properties. In addition, account must be taken of the type of scoring that the instrument is based on and whether it yields:

- Nominal data: numbers are used simply for classification (e.g. ‘died’, ‘survived’).
- Ordinal scale data: scale items stand in some kind of relation to each other (e.g. ‘very difficult’ through to ‘not very difficult’).
- Interval scale data: the characteristics of an ordinal scale but the distances between any two numbers on the scale are of a known size (e.g. temperature).
- Ratio scale data: the characteristics of an interval scale with the addition of a true – not arbitrary – zero point (e.g. weight).

The full definitions of these concepts can be found in most research method textbooks (e.g. Blalock and Blalock 1971; Moser and Kalton 1971). Their consideration is crucial because different scales are suitable for providing different types of data and for answering different types of question. In addition, interval and ratio scales are required if powerful statistical analysis is envisaged. Some measurement scales, which technically provide combined nominal and ordinal scale data, have formulas for re-coding, summing and transforming response categories, using a scoring algorithm, into a scale ranging from 0 (worst) to 100 (best) (e.g., Ware et al. 1993, 1995). This allows the results to be reported as means, rather than frequency distributions, and for the use of more powerful statistical techniques. This is justified on the grounds that the treatment of such data at interval level has minimal effects on most statistical procedures, although this is not without controversy (Julious et al. 1995; Lamping et al. 1998a).