Patient-reported Health Instruments Group

Instruments for Children and Adolescents: a Review

Louise J. Schmidt
Andrew M. Garratt
Ray Fitzpatrick

Report to the Department of Health
July 2001

National Centre for Health Outcomes Development

health outcome indicators

PATIENT-REPORTED HEALTH INSTRUMENTS GROUP
(formerly the Patient-assessed Health Outcomes Programme)

INSTRUMENTS FOR CHILDREN AND ADOLESCENTS: A REVIEW

A STRUCTURED REVIEW OF PATIENT-REPORTED HEALTH INSTRUMENTS

Louise J. Schmidt
Andrew M. Garratt
Ray Fitzpatrick

Patient-reported Health Instruments Group
National Centre for Health Outcomes Development (Oxford site)
Unit of Health-Care Epidemiology
Department of Public Health
University of Oxford

July 2001

———————

This report should be referenced as follows:

Schmidt LJ, Garratt AM, Fitzpatrick R. *Instruments for Children and Adolescents: a Review* Report from the Patient-reported Health Instruments Group *(formerly the Patient-assessed Health Outcomes Programme)* to the Department of Health, July 2001.

Copies of this report can be obtained from:

Dr Kirstie Haywood
Co-Director, Patient-reported Health Instruments Group
National Centre for Health Outcomes Development (Oxford site)
Unit of Health-Care Epidemiology
Department of Public Health
University of Oxford
Old Road
Headington
Oxford OX3 7LF

tel: +44 (0)1865 227157
e-mail: kirstie.haywood@uhce.ox.ac.uk

Alternatively, it can be downloaded free of charge from the PHIG website:

http://phi.uhce.ox.ac.uk/

LIST OF CONTENTS

Executive Summary

Background

This report presents a review of multi-dimensional generic measures of child/parent-reported health outcomes (encompassing functional, health status, and health-related quality of life measures) for use with general populations of children and adolescents. It also highlights the major methodological issues to be considered when carrying out such assessments in this population. The review will provide information to guide potential users in the selection and appropriate use of instruments.

Research Aims

1. to highlight methodological issues in assessing subjective health outcomes among children and adolescents;

2. to identify published reviews of such instruments;

3. to identify relevant generic measures;

4. to present the existing evidence on the properties of relevant generic instruments, including reliability and validity;

5. to make recommendations regarding the selection of individual instruments.

Methods

Relevant literature was identified using the PHIG database which has been designed to capture electronically-held references relating to self-reported health outcome measures. In addition, key sources were hand-searched. The PHIG database was searched for references relating to children or adolescents; the abstracts were then assessed against inclusion criteria.

After retrieving relevant references, the following information was extracted:

- the purpose and content of the instrument;

- instrument development and scoring;

- population samples in which the instrument was developed and tested;

- measurement properties: reliability, validity and acceptability.

Key Findings

The literature search identified 10 reviews of instruments for use with children or adolescents, none of which focussed on applications at the population level. One comprehensive and systematic review of measures for children with chronic diseases was identified.

Sixteen generic and multi-dimensional instruments which had been evaluated in a general population of children or adolescents were identified. Three of these had been developed in

the UK. Most instruments cover the three main areas of physical, social and mental health and well-being; some also address school achievement, family functioning and risk-taking behaviour.

Several child-completed instruments were identified for use with young children (from the age of six), although parent-completed measures were common for this and younger age-groups. For older children (aged 11 and over), the majority of instruments identified were self-completed. Four parent-completed instruments can be used with children under one year old, whilst child-completed instruments have been developed for children as young as four.

Only five instruments have reported data on both internal consistency and test-retest reliability in general populations. All except two instruments have undergone some testing for construct validity. Various formats, including storybook pictures or computer presentations, have been used in an attempt to reduce the response burden on children.

The major methodological issues to be considered when measuring child/parent-reported health outcomes are as follows:

- there is a lack of standardisation in the conceptualisation and operationalisation of health-related quality of life in the young;

- instruments developed for use with adults are less likely to be appropriate for use with children and adolescents;

- population-based approaches tend to broaden concepts of health and well-being to include school achievement, family functioning and self-esteem;

- domains measured by instruments need to be developmentally and culturally appropriate;

- children are likely to be able to provide self-reports if the instrument is appropriate to their abilities, although the exact age from which this is possible is subject to debate and may vary according to domains;

- data from proxies (usually parents) is likely to differ from that gathered from children themselves. Further investigation of this is required, using instruments that allow for parallel child- and parent-reporting.

Key conclusions and recommendations

All instruments require further validity and/or reliability testing in UK populations, and this should take place alongside any application of instruments. We recommend that difficulty be assessed whenever instruments are administered.

In choosing a particular instrument, the nature and design of the instrument should be assessed against the prospective application. One needs to be clear whether a parent- or child-response is preferred, which domains are of most relevance, and what degree of prior testing of the instrument is acceptable.

For younger populations, the CHQ-PF50 has been the most extensively evaluated but is available as a parent-completed measure only. Two of the UK measures are child-completed measures designed for young children, although at present insufficient evidence is available

for their psychometric properties. Where younger children are asked to complete measures, this should, ideally, be accompanied by parallel proxy assessments (usually by parents). For this reason the new CHIP-CE seems particularly promising, as child- and parent-completed versions are available for children from a young age, although validity evidence is not expected to be presented until the Autumn of 2001.

For older children, the weight of evidence suggests the CHQ-CF87 and the CHIP-AE. The main drawback to both of these instruments is their length, although a shortened version of the CHQ-CF87 is under development.

If one is interested in health service utilisation and the uptake of services, the UK-developed Warwick Child Health and Morbidity Profile would be the most appropriate.

Chapter 1: INTRODUCTION

a) Aim of the review

The main aim of this review is to identify and evaluate multi-dimensional, generic instruments for measuring health-related quality of life (also referred to as health status, functioning and well-being) which have been assessed for use in measuring the subjective health outcomes of children at the population level. This review considers self-completed instruments and those completed by proxies (usually parents) on behalf of their children. As a basis for this review, the major methodological issues in evaluating health-related quality of life in children and adolescents are first summarised, existing reviews of instruments in this field are then identified and described.

b) Child/parent-reported health outcome measures

Self-reported health outcome measures aim to measure subjective quality of life, health-related quality of life or health status from the viewpoint of the population or patient group themselves. In the case of children or adolescents,[1] this can be achieved either by asking the children themselves for their responses, or by using proxy raters, usually parents. The various terms used to refer to instruments of this nature (e.g. quality of life, health-related quality of life, health status) can be differentiated, although in practice there is little consistency in the use of these terms, or agreement as to what they mean (Fitzpatrick et al., 1998). However, the common feature of such instruments is that they measure health from the subjective viewpoint of the individual concerned (Fayers and Machin, 2000).

Self-reported health outcome measures developed as a result of several trends (Fitzpatrick et al., op.cit). First, it was increasingly recognised that traditional biomedical outcomes needed to be supplemented by measures that took the patient's experience and concerns into account, particularly with regard to chronic diseases, where the intention is often to improve functioning and general quality of life rather than to cure. Second, it is increasingly considered appropriate and desirable for patients' preferences and wishes to be taken into account in decision-making concerning their health care. Third, health care budget holders face rising pressure on resources which has led to the growing use of cost-effectiveness evaluation, requiring evidence of benefits perceived by patients, professionals and society as a whole.

To date, most applications of health-related quality of life measures have been in clinical trials, where data from patients has often supplemented clinical indicators of morbidity in assessing the outcomes of interventions. However, such assessments are potentially relevant also at the population level, where they can be used to evaluate specific or general population-level interventions, such as health-promotion initiatives. Instruments based on a broad definition of quality of life capable of capturing a variety of outcomes are likely to be more relevant for the evaluation of population-level multi-sectoral initiatives.

A number of key issues have an important bearing on the scope of the review that follows. First, instruments can be classified as disease-specific or generic. Disease-specific measures, as their name suggests, have been developed specifically for use with patients who have particular conditions or illnesses. Generic instruments, by contrast, are designed to measure aspects of health which are of universal importance. They are therefore suitable for use across different patient populations, and are potentially applicable also to healthy populations.

[1] Hereinafter 'children' will be used to refer to both children and adolescents

Another key issue is whether instruments assess single dimensions of quality of life, such as physical functioning, or whether they assess multiple dimensions of quality of life, such as physical health, mental health, and social well-being. In the general literature on self-reported health outcomes, several existing generic instruments focus principally on physical functioning and are not likely to be relevant to generally healthy populations. Although this was often the case with some of the early instruments, the content of many generic instruments has since been expanded to include social and emotional aspects of health, as well as existential issues (Fayers and Machin, op.cit.).

Instruments assessing multiple domains can be further grouped into those that produce a profile of scores relating to different 'dimensions' of health, or those that combine the domains into a single index or score of health. From the literature on measuring health outcomes in children, the CHIP-AE and the Comprehensive Quality of Life Scale are examples of profile and single index instruments, respectively.

Finally, instruments can be administered in different ways: from self-completion questionnaires to interviews. In the field of children's health outcome measures, there is more innovation in the administration format, arising from the desire to make completion of the instrument enjoyable and easy. Examples include the Exqol which consists of computer presentations and the Generic Children's Quality of Life Measure which uses a storybook format.

c) Conceptual and methodological issues

i. Definitional issues

As is the case for measures applied to adults, there is no uniform consensus on the theoretical framework defining health-related quality of life in children (Levi and Drotar, 1998). A lack of standardisation in both the conceptualisation and the operationalisation of health-related quality of life assessment has produced a large number of instruments (Landgraf and Abetz, 1996). One review identifies confusion in the definition of health-related quality of life, as shown by the overlap between quality of life and functional status measures (Eiser and Morse, 2001 a & b).

The question of whether children have the same underlying concept of quality of life as adults, and whether instruments devised for use with adults can (with some adaptation) be considered appropriate for use among children, is unresolved. A recent review found three studies where adult measures had been used directly with children, with little or no adaptation made for this specific population. In a further 11 studies (using 9 separate measures), adult measures were used as a model for work with children (Eiser and Morse, op.cit.).

Since the goal of adult functioning is to be self-sufficient and economically productive, adult-based measures of functioning are not likely to be relevant to children (Kozinetz et al., 1999; Pal, 1996). In the literature on adults, quality of life is often defined as the gap between expectations and reality, but children's immaturity may mean their expectations are limited (Colver and Jessen, 2000). It has been suggested that children's and adults' conceptions of health and illness differ, in that children view health and illness as separate entities rather than as lying on a continuum (Colver and Jessen, op.cit.).

Operational definitions of health-related quality of life in available instruments for children fall into the categories of functioning, health status (including well-being) and preference- or

utility-based measures, with little comparison between these different assessment methods (Levi and Drotar, op.cit.). Most instruments use a simple functional concept of health, comprising a list of activities grouped into physical, psychological and social domains, although Starfield and Lindstrom use other models (Pal, op.cit.).

The lack of an agreed theoretical framework as to the nature of health-related quality of life in children also means there is a lack of consensus concerning the domains of quality of life that should be measured to reflect children's views. As a result, there is variability in both the number and the definition of domains covered by existing instruments (Eiser and Morse, op.cit.). One review found quality of life was rarely assessed in a multi-dimensional fashion (Bullinger and Ravens-Sieberer, 1995). In another assessment, symptoms and pain, together with motor functioning, cognitive functioning, social functioning, autonomy and emotional functioning, were found to be the most prevalent domains (Vogels et al., 1998).

Even within domains, there can be variations of emphasis. In measuring physical quality of life, the emphasis may be on physical symptoms, self-care, participation in physical activities, or distress caused by limitations (Eiser and Morse, op.cit.). There seems to be increasing recognition that, since the health and behaviour of children is extremely sensitive to the social context in which they live, instruments should take account of this - although they often fail to do so (Pal, op.cit.).

A further complication of measuring health-related quality of life in children may be that domains are more intertwined than for adults: for example, cognitive development may precede social interaction (Schor, 1998). Population-based approaches to child health attempt to broaden the construct of health and well-being on which many disease-specific measures are based, by including the aspects of school achievement, family support and self-esteem (Raphael, 1996).

It can be difficult to compare instruments when their theoretical framework and the domains they assess vary. This has led to the suggestion that instruments should be assessed in terms of their intentions and conceptualisation/theorisation of health-related quality of life (Pantell and Lewis, 1987). These concepts and assumptions need to be made explicit, particularly in order to enable construct validity testing, which seeks to determine whether the instrument measures what it claims to measure (see section d) below).

ii. Developmental issues

A second major area of discussion in the literature concerns the different ways in which children develop, and the different speed at which this can occur from child to child. As development is not always linear (Pantell and Lewis, 1987), how do we know that 'outcomes' are really outcomes and not indicators of development? Although some commentators consider there is a lack of agreement on appropriate functioning, especially given that societal values and expectations are constantly changing (ibid.), others consider the primary milestones of children as they develop from a young age to adolescence are adequately documented in the developmental literature (Landgraf and Abetz, op.cit.).

It is important to consider whether the concepts inherent in instruments are developmentally appropriate, and whether items are appropriate for gender, age and culture (ibid.). The operationalisation of constructs such as body image and self-esteem may vary across cultures, so this would need to be considered in choosing an instrument. If one wishes to monitor health in longitudinal studies, a possible solution is to use items that are not overly age-related, so that children of different ages can complete the same instrument (Erling, 1999).

iii. Self-reports by children

In principle, children are able to provide self-reports of their health-related quality of life or health status if an instrument appropriate to a child's abilities is chosen (Bullinger and Ravens-Sieberer, op.cit.), although this assertion has not been thoroughly tested (Kozinetz et al., op.cit.). There may be differences between the ages at which children can self-report on different domains. For instance, children as young as five may be able to provide self-reports of pain, whilst the age of nine or ten may be more appropriate for subjective concepts such as behaviour and self-esteem (Landgraf and Abetz, op.cit.). There may even be differences between groups of children: children with chronic illnesses may be better at providing self-reports than healthy children of a similar age, due to their greater contact with health services (Kozinetz et al., op.cit.; Colver and Jessen, op.cit.).

Potential problems with children providing self-reports include position biases (tendency to select first answer), acquiescent response bias (tendency to agree with questionnaire, regardless of content), limited understanding of negatively-worded items, and problems with perceiving time periods (Kozinetz et al., op.cit.; Pantell and Lewis, op.cit.; Connolly and Johnson, 1999). If a written questionnaire format is used, one needs to be sure that the children have the necessary cognitive and reading skills to understand the item. Different administrative formats, such as drawings, may help to lessen the burden of readability by not requiring children to understand written questions (Finkelstein, 1998), although assessments of children's abilities to understand the concepts behind the drawings would still be necessary.

There has been little evaluation to date of the different modes of administration and readability of instruments (Landgraf and Abetz, op.cit.). Ideally difficulty should be assessed whenever instruments are administered (Eiser and Morse, op.cit.).

iv. Reliability and validity of proxy reports

Given that it may not always be possible for children to provide self-reports, one may need to consider the possibility of other people (proxies) providing data on their behalf. One review found that few studies actually used self-reported methods; instead, parents and clinical staff assessments accounted for 90% of assessments (Bullinger and Ravens-Sieberer, op.cit.). This was a far higher proportion than found in the present review, probably due to the inclusion of disease-specific measures and clinician-rating scales.

It has been suggested that agreement between parent and child is more likely for functional status items and less likely where the items are more subjective, e.g. getting on with others, where parents have less access to information, e.g. making friends, or where the subject matter is considered sensitive, e.g. family functioning (Pantell and Lewis, op.cit.). A systematic review found 14 studies where child- and parent-responses could be compared. Although there was evidence that agreement is closer for physical functioning compared with social and emotional domains, differences between the instruments made it difficult to draw conclusions (Eiser and Morse, op.cit.).

Whether discrepancies in the information provided represent real differences of opinion between proxies and children, or whether children are less able to evaluate more subjective domains is unclear, although there is evidence that both factors may contribute (Pantell and Lewis, op.cit.).The reasons why differences between proxy and self-ratings arise need to be examined further. Parents could be influenced by knowledge of other children, their

expectations and hopes for the child, additional life stresses, and their own mental state (Eiser and Morse, op.cit.). From the limited evidence available, no simple relationship was found between agreement and variables like age, gender and illness (ibid.).

The choice of proxy requires careful consideration. Where self-report is unavailable and depending on what one is measuring, it may be wise to look further afield for proxies, to include teachers and, for older children, possibly peers (Colver and Jessen, op.cit.). Few instruments are designed for parallel child- and parent-reporting. If proxies are used, allowing for self-report of the proxy's own health would enable the relationship between self-perceived health and proxy-reported health to be examined (Connolly and Johnson, op.cit.). There is evidence that fathers rate children as having fewer behavioural and psychological problems but since by default mothers almost always complete instruments, this issue has not been fully assessed (Landgraf and Abetz, op.cit.).

d) Criteria for assessing measures

The criteria by which self-reported health outcome measures can be evaluated have been summarised as: appropriateness, validity, reliability, responsiveness, precision, interpretability, acceptability, and feasibility (Fitzpatrick et al., op.cit.).

The first and most fundamental point to consider is the appropriateness of an instrument, i.e. whether it measures what have been identified as the most important outcomes for the purposes of the evaluation. Specifically, one would want to consider whether the instrument contains all of the domains of relevance, and the appropriateness of child- or proxy-report for the particular information to be collected. An appropriate measure is also, in a general sense, one that fulfils the other criteria listed above.

Before an instrument can be recommended for application, its measurement properties of validity, reliability, and responsiveness should be assessed. Validity concerns whether an instrument is measuring what is intended, and can be assessed using both qualitative and quantitative methods. It is not a fixed property of instruments ascertainable from a single experiment, rather it should be assessed in relation to each application of an instrument. Face and content validity are matters of qualitative judgement; this relies on information such as whether the patient or population group targeted by the instrument was included in generating its content, and whether the items chosen are considered adequately to cover the domains of the instrument. An assessment of internal validity is closely allied to the item/domain relationship and, using statistical methods such as factor analysis, seeks to assess whether the items said to measure the same construct do actually group together.

Construct validation includes comparisons with other instruments, relating the instrument scores to clinical and socio-demographic variables, and looking at relationships between domain scores within the instrument. Prior hypotheses should always be made against which results can be assessed and conclusions can be drawn. The statistical methods usually involve correlation but if groups are being compared, t-tests or equivalent non-parametric equivalents are used.

Reliability looks at whether an instrument is consistent in its measurements, either internally or over time. Cronbach's alpha, a test of internal consistency, assesses the overall level of correlation between items within a scale and can be used with multi-item scales. Standards for the reliability coefficient are dependent on whether the instrument is intended for use with

groups, for which a reliability coefficient of 0.7 is recommended, or individual patients, for which the more stringent criterion of 0.9 is recommended.

Test-retest reliability is designed to take account of variation in information generated by the instrument over time. It assesses the level of association between two sets of instrument scores from the same group of patients on two different occasions. There is no real agreement on the length of time between administrations of test and retest questionnaires, but it should not be so short that patients can recall their previous responses, nor so long that their health may have changed. Ideally, there should be some attempt to assess whether there have been actual health changes between the two administrations. In practice, this is often achieved by including a health transition question. A reliability coefficient of 0.7 for group data is commonly cited, although some set higher standards.

Responsiveness refers to the ability of an instrument to measure significant changes in health. This is an important property of any instrument used for measuring outcomes. Responsiveness is assessed by looking at changes in instrument scores for groups whose health is known to have changed, and is commonly used in patient populations. It is, however, unclear how this criterion would be evaluated in a generally healthy population; as a result, generic measures at the population level have rarely been evaluated for responsiveness.

The precision of an instrument's scores, a related issue, can be indicated by (a) the range of response options available (at one extreme, a 'yes/no' response is likely only very crudely to indicate levels of health-related quality of life) and (b) the existence of ceiling (maximum score) or floor (minimum score) effects. If responses are concentrated at either end of a score's potential range, the instrument is likely to be poor at differentiating responses, whether between respondents or over time.

Interpretability considers the degree to which the scores generated can be considered meaningful. To date it is not possible to compare the self-reported health outcome measures on the criterion of interpretability.

An instrument is more likely to be acceptable to a patient or population group if it measures what they consider to be the most important aspects of health-related quality of life. This is often achieved by ensuring that representatives of the population of interest are involved in generating the items included in the instrument. Proxy indicators for acceptability include response and completion rates.

Lastly, instruments must be feasible if their uptake is to be encouraged. Unfortunately, information on the time and resources needed for the application of instruments is often lacking.

Chapter 2: METHODS

a) Search strategy

The PHIG database was used to search for records containing the terms 'child*' or 'adolesc*'. This database was constructed using thorough and extensively evaluated search criteria, designed to retrieve all references relating to the development and testing of self-reported health outcome measures as well as methodological and review papers. The PHIG search strategy is shown in Appendix 1. In developing the database, the following electronic databases were searched: Embase, Medline, Biological Abstracts, PsychInfo, AMED, Econlit, Sociological Abstracts, British Nursing Index, PAIS International, the Royal College of Nursing database, SIGLE, and Cinahl. The journal "Quality of Life Research" was hand-searched, as were the following sources:

- Salek, S. (1998). Compendium of Quality of Life Instruments. New York: Wiley.
- Tamburini, M. Researcher's Guide to the Choice of Quality of Life Assessment in Medicine - *http://www.qlmed.org*
- Bowling, A (1995). Measuring Disease. Buckingham: Open University Press.
- Shumaker and Berzon, eds. (1995). The International Assessment of Health-Related Quality of Life: Theory, Translation, Measurement and Analysis.
- Spilker, B. (1996). Quality of Life and Pharmacoeconomics in Clinical Trials. 2nd Ed. Philadelphia: Lippincott- Raven.
- McDowell, I. and Newell, C. (1996) Measuring health: a guide to rating scales and questionnaires. 2nd Ed. Oxford University Press, New York.

Individual abstracts generated by the database search were examined to assess whether the reference met the criteria for inclusion in the review. If this was the case, a copy of the full article was retrieved for evaluation, and the reference lists of these papers were also scanned to identify other relevant papers.

b) Inclusion/exclusion criteria

To be included in the review, an instrument had to be a generic, multi-dimensional instrument evaluated in a general population of children under 18 years of age. Also included were reviews of such instruments, and papers addressing methodological or conceptual issues associated with measuring health-related quality of life in this population.

The review excluded studies focussing solely on the evaluation of instruments in groups of patients with particular illnesses or conditions. Also excluded were dimension-specific instruments containing only one domain of health-related quality of life, such as physical functioning. Finally, an instrument was excluded if no English-language version of the instrument had been evaluated. Excluded instruments are listed in Appendices 2 and 3.

c) Data extraction

Instruments identified as meeting the inclusion criteria were summarised and evaluated against the criteria shown in Table I. The criterion of responsiveness was not included because there has been no evaluation of this measurement property in general populations of children. Feasibility issues were also not addressed in the studies identified. Table II summarises previously published reviews, while Table III presents the instruments meeting the inclusion criteria for this review. Table IV shows the dimensions and the number of items in each instrument; data relating to the populations involved in instrument evaluations are

shown in Table V. Tables VI and VII summarise the evidence on reliability and validity, respectively. Other issues, such as item generation and scoring, are covered in the summaries of each instrument contained in Chapter 4.

Table I: *Inclusion criteria*

Instrument description	**Population description**	**Measurement properties**
Purpose	Age, sex, ethnicity, socio-economic status	Acceptability (response rates and completion rates)
Development of instrument	Proxy or self-completion	Validity (face, content and construct validity)
Number of items	Setting of evaluation	Reliability (test-retest and internal consistency)
Dimensions covered	Country of evaluation	
Scoring		

Chapter 3: RESULTS

a) Search results

The PHIG database includes 3,921 publications concerned with the development and testing of self-reported health outcome measures. Of these, 232 (6%) relate to instruments developed or evaluated for use with children, although many were developed for use with specific disease groups and were therefore excluded from this review.

i. Reviews identified

The search identified ten reviews of instruments measuring health-related quality of life in children, although most of these focussed on groups of patients with specific diseases rather than general child populations. These reviews are listed in Table II. One was a comprehensive systematic review of generic and disease-specific instruments of health-related quality of life for use with chronically ill children, whether by self-report or proxy raters (Eiser and Morse, op.cit.). Although Eiser and Morse did not focus on the use of such instruments in general population surveys, it is useful in identifying generic instruments which have been evaluated in chronically ill child populations.

ii. Instruments identified

The database search identified 16 instruments which met the inclusion criteria; these are listed in Table III. Searching the reference list of published reviews generated one additional relevant measure: the Generic Children's Quality of Life Measure (Collier, 1997). Instruments which failed to meet the inclusion criteria are listed in Appendices 2 and 3. The most common reason for exclusion was that the instrument had undergone evaluation with disease-specific groups only. Other reasons include not focussing on a child or adolescent population, or being restricted to just one dimension. Five instruments were excluded because there had been no evaluation of an English-language version of the instrument.

b) Nature of the reviews

None of the ten reviews focussed specifically on the application of health-related quality of life instruments in general populations of children, although all of them included generic instruments which had been evaluated in disease-specific groups. Two methodologically thorough reviews were identified; these provided details of the databases searched, the search terms used, and the inclusion or exclusion criteria used (Connolly and Johnson, op.cit.; Eiser and Morse, op.cit.).

Eiser and Morse also recommended particular instruments for use, as fulfilling certain psychometric criteria. Of the four reviews producing recommendations for use, three concur that the Child Health Questionnaire is the best available instrument in terms of data on its psychometric properties (Eiser and Morse, op.cit.; Colver and Jessen, op.cit.; Kozinetz et al., op.cit.). The Health Utilities Index and the PedsQL are both singled out by two of the reviews making recommendations; however, these instrument are excluded from the present review, since no evaluations with general child or adolescent populations could be found.

Analysis of the instruments

i. Content

The content of the instruments is summarised in Table IV. Regarding the theoretical basis of instruments, several of them employ a simple functional concept of health: viz. a list of activities grouped into physical, psychological and social domains. Two instruments, the CHIP-AE and the Quality of Life Profile-Adolescent Version, were developed from a more complex theoretical basis. The Warwick Child Health and Morbidity Profile takes a different approach by including items on health service contacts and utilisation of services.

The shortest instrument is the Dartmouth COOP charts, which comprises six items, whilst the longest is the Pediatric HealthQuiz containing 375 items. Two instruments contain over 100 items (Pediatric HealthQuiz and CHIP-AE) but the majority consist of less than 40 items.

In terms of domains covered, all the instruments explicitly cover physical health or functioning and most cover mental or psychosocial health (except the Pictorial Scale of Perceived Competence and Acceptance, Children's Health Rating Scale and Warwick Child Health and Morbidity Profile). Four instruments explicitly consider school functioning or achievements (CHQ, CHIP, COOP charts, and Exqol), whilst seven instruments address family functioning (CHQ, CHIP, COOP charts, Exqol, Generic Children's Quality of Life Measure, Pediatric HealthQuiz, and Pictorial Scale of Perceived Competence and Social Acceptance for Young Children). Four instruments contain items eliciting data on risk-taking behaviour (CHIP, Instrument for Monitoring Adolescent Health Issues, Pediatric HealthQuiz, and Juvenile Health and Wellness Survey), whilst five inquire about symptoms or specific disorders (CHQ-PF50, CHIP, Pediatric HealthQuiz, Exqol, and the Juvenile Health and Wellness Survey). There are several versions of the FS II (R) containing different age-appropriate behavioural items.

ii. Populations

Four instruments (Child Health Status Questionnaire, Pediatric HealthQuiz, Warwick Child Health and Morbidity Profile, and the FS II(R)) can be used with children under one year old; these are, naturally, parent-completed measures. Child-completed instruments are reported to be suitable for children as young as four (Pictorial Scale of Perceived Competence and Social Acceptance for Young Children). Three child-completed instruments (Exqol, Generic Children's Quality of Life Measure, and Child's Health Self-Concept Scale) are designed for use with schoolchildren aged from six to around 13, as is the new CHIP-CE. The CHQ is a parent-completed instrument designed for use with parents of children aged 5-13; there is also a version of the Child Health Status Questionnaire designed for parents of children in this age-group. The Children's Health Rating Scales are child-completed and designed for use with children aged 9-12.

Most child-completed instruments identified (five in total) are applicable for use in children approaching adolescence or teenagers, ranging from 10 to 18 years (Juvenile Wellness and Health Survey, and the child-completed version of the CHQ) to 14-21 years (COOP charts). Four instruments (CHQ,[2] Exqol, Generic Children's Quality of Life Measure, and Warwick Child Health and Morbidity Profile) were either developed or tested in a UK population.

[2] The UK evaluation of the CHQ was, however, undertaken with a population of chronically ill children, although evaluation with a general population is currently underway.

iii. Reliability

As shown in Table VI, all except three instruments (Instrument for monitoring adolescent health issues, Pediatric HealthQuiz, and Warwick Child Health and Morbidity Survey) have been assessed for internal consistency reliability. Fewer instruments (8/16), have been assessed for test-retest reliability. Only five instruments have to date evaluated both types of reliability in a general population of children: CHIP, CHQ, Child's Health Self-Concept Scale, ComQol, and COOP charts.

iv. Validity

Table VII summarises the data available on the construct validity of each instrument in a general population of children. There were four main methods of assessing construct validity in the evaluations identified. First, the instruments were compared with other instruments measuring similar constructs; this type of validity testing was used for nine instruments. Second, the individual's responses were compared with a proxy response; this occurred in five cases. Third, comparisons were made between sub-groups of respondents whose scores were likely to vary: for instance, scores from the general population were compared with scores from a patient group likely to exhibit worse health, or else the scores of sub-groups defined by demographic variables such as sex or age were compared; this was carried out for four instruments. Fourth, domains or items within instruments expected to show particular relationships with each other were compared; this was explicitly assessed for seven instruments. Since it is common for evaluations to include a large volume of data on relationships between variables which may be difficult to interpret, data is reported here only where explicit hypotheses were used. For two instruments, construct validity was not assessed (Instrument for monitoring adolescent health issues, Pediatric HealthQuiz).

It was rare that children only were involved in the generation of items included in the instrument. Often, children constituted one source of item generation together with information from other sources, such as the published literature and health professionals. When children or parents were not involved in generating items, instruments tended to be piloted with them to assess difficulty of items.

Table II: *Reviews of instruments*

Review	Aim	Databases and search terms	Inclusion criteria	Evaluative criteria	Instruments identified	Recommendations
Bullinger & Ravens-Sieberer, 1995		Searched Medline, Embase, Psyndex, PsycInfo, Psycom, Cancerlit, Aidsline, Bioethicsline & Somed 1964-1995. Used terms 'quality of life' & 'child'. Weighting system used but not clear what.		Descriptive: age, respondent, number of scales, target population, reliability, validity, sensitivity.	Measures of function: NIE Functional Status Index, Functional Status (II)R, Batelle Developmental Inventory, Vineland Adaptive Behavior Scales, Play Performance Scale for Children, Wee-FIM, Pediatric Evaluation of Disability Inventory. Generic quality of life instruments: Ontario Child Health Study, CHQ, Children's Health Rating Scales, Quality of Well-being scale, General Health Rating Index, NICQL, QoL index for Nordic countries.	
Colver & Jessen, 2000	To identify generic measures in English which either have been or could be used in neonatal follow-up studies.	Not explicit		Descriptive: mode of administration, age, respondent, reliability, validity.	Functional health status instruments: Vineland Adaptive Behaviour Scale, WeeFIM, HUI Mark II & III. Measures of health status/QOL: CHQ, Children's Quality of Life Scale, PedsQL, TACQOL, KINDL, Adolescent Quality of Life Profile.	Comments that the best instrument is the CHQ, although the PedsQL is shorter & has the advantage of seeking the views of children from age 5. The HUI3 is useful for economic evaluations.
Connolly & Johnson, 1999	To provide an overview of generic HRQOL measures used in paediatric populations.	Searched Medline, HealthSTAR & Embase 1980-1988. Used terms 'quality of life', 'paediatrics', 'child*' & 'adolescent'.	Instruments focus on measurement of health-related quality of life for use in paediatric populations with evidence of its use & results.	Descriptive: domains, respondent, age, number of items, mode of administration, country/language, translations, population, reliability, validity.	CHIP-AE, CHQ, COOP, DUCATQOL, Functional Status (II)R, KINDL, Nordic Quality of Life Questionnaire for Children, Ontario Child Health Study, Rand Health Status Measures for Children, TACQOL, WCHMP, 16D, 17D. Preference-based measures: HUI Mark II & III, QWB Scale.	
Eiser & Morse, 2001 a & b	To identify currently available generic & disease-specific measures of quality of life for children with chronic diseases.	Searched Medline, BIDS ISI Science Citation Index, BIDS ISI Social Science Citation Index, PsycInfo, Cochrane Controlled Trials Register & meta-Register of Controlled Trials for English language papers 1980-1999. Used terms 'functional status', 'health status', 'quality of life', 'chronic diseases', 'illness' & individual chronic diseases. Hand searching & checking of reference lists.	Included if measure of quality of life, health status or well-being in children aged 18 or under with a chronic disease. Measures had to include some reliability or validity data & be used by child, proxy or both.	Descriptive: respondent, age, number of domains, number of items, reliability, validity, origin.	CHIP, CHQ, Child Quality of Life Questionnaire, COOP, Exeter Quality of Life Measure, Functional Status (II)R, Generic Health Questionnaire, How Are You?, KINDL, Nordic Quality of Life Questionnaire for Children, Pediatric Quality of Life Questionnaire, Perceived Illness Experience, Quality of Life Profile-Adolescent Version, SIP, TACQOL, Warwick Child Health & Morbidity Profile, HUI Mark II & III, 16D, 17D, Quality of Well-Being.	Three instruments fulfil basic psychometric criteria: CHQ, Pediatric Quality of Life Questionnaire, HUI Mark II (though the last two are not designed to assess the full range of functioning).
Kozinetz et al., 1999	To identify reliable & valid instruments for measuring the health status of children with special care needs in the clinical setting.	Searched Medline 1966-1988 for English language papers using terms 'health status', 'quality of life', 'outcome assessment', 'functional status' & 'patient satisfaction'.		Descriptive: purpose, respondent, timing of use, reliability/validity, mode of administration, clinical use.	Measures of health status: Rand Health Status Measures for Children, HUI, CHIP-AE, HUI Mark II, CHQ. Four measures of satisfaction with care. Measures of satisfaction with health status: Feeling Thermometer, Standard Gamble. Functional status measures: Basic Gross Motor Assessment, Functional Independence Measure for Children (WeeFIM), Functional Status II(R), Play Performance Scale for Children. Measure of family health status: Impact on Family Scale.	

Review	Aim	Databases and search terms	Inclusion criteria	Evaluative criteria	Instruments identified	Recommendations
Landgraf & Abetz, 1996	To identify instruments developed & validated specifically for paediatric populations.	Extensive search of psychological & medical literature using terms 'quality of life', 'health status indicators', 'generic health surveys', 'health outcomes', 'outcomes assessment' & 'activities of daily living'.		Descriptive: purpose, age, respondent, mode of administration, number of items, psychometric results.	CHIP, COOP, Functional Status II(R), Health Institute's Child Health Assessment Project, Rand Health Status Measures for Children, National Health Interview Survey, Ontario Child Health Study, Quality of Well-Being Scale.	
Levi & Drotar, 1998				Descriptive: domain, age, respondent, specific conditions.	CHIP, CHQ Rand Health Status Measures for Children, HUI Mark II, Quality of Well-Being scale. Six functional status measures: Child Health Assessment Questionnaire, Functional Disability Inventory, WeeFIM, Functional Status II(R), PEDI, Play Performance Scale for Children.	
Marra et al., 1996	To identify recent work in producing multi-dimensional measures of HRQOL for children & adolescents.			Descriptive: domains	Rand Health Status Measures for Children, CHIP-AE, Functional Status (II)R, MAHS, Vineland Adaptive Behaviour Scale, Paediatric Evaluation of Disability Inventory, Play Performance Scale for Children.	CHIP, FS II (R), Rand Health Status Measures for Children
Pal, 1996		No terms given but searched Medline, Embase & SciSearch 1979-1995.	Instruments assessed according to criterion of 'child-centredness' & extent to which child considered part of 'family unit within a social network'; had to be 'generalisable' & have 'appropriate underlying assumptions'.	Descriptive: age, dimensions, method of administration, psychometric characteristics, scoring, statistical issues & practicality.	Rand Health Status Measures for Children, Functional Status II(R), MASC, CHIP-AE, Nordic Quality of Life Questionnaire, Child Quality of Life Questionnaire, FSQ, instruments by Austin (1994) & Schmidt (1993).	
Spieth & Harris, 1995		No details of search terms or databases used.	Measures included if covered four core components of QoL: disease status, functional status, psychological functioning, social functioning.	Descriptive: domains, respondent, age, number of items, psychometric properties, disease-specific populations.	Play Performance Scale for Children, Quality of Well Being Scale, Rand Health Status Measures for Children, CHIP.	

Table III: *Instruments*

Instrument	Evaluative papers	Aim/intended application of measure	Child/parent-report
Child Health & Illness Profile/CHIP-AE	Starfield et al., 1993, 1995, 1996; Riley et al., 1998 a & b	To document state of health in adolescent populations, identify differences in health of sub-populations, assess impact of health service interventions on health, make initial assessment of adolescents for screening services.	Child-report
Modified CHIP-AE	Chen & Chen, 1999	The modified CHIP-AE is specifically modified for assessing adolescent health behaviours to inform school health programme planning.	Child-report
Child Health Questionnaire/CHQ	Landgraf & Abetz, 1997, 1998; Landgraf et al., 1998; Waters et al., 1999, 2000	To measure & compare health of general & specific groups of children; to evaluate treatments.	Parent report - Landgraf 1998; Waters et al., 1999, 2000 Child-report - Landgraf & Abetz, 1997; Waters et al., 1999
Child's Health Self-Concept Scale/CHSCS	Hester, 1984	Potential use for nursing research & practice.	Child-report
Children's Health Rating Scales	Maylath, 1990	Self-report of general health in children for group comparisons or multivariate analyses.	Child-report
Child Health Status Questionnaire	Eisen et al., 1979; Diaz et al., 1986	Measure of child health status suitable for testing hypotheses about health care financing & health status.	Parent-report & child-report - Diaz et al., 1986
Comprehensive Quality of Life Scale/ComQOL	Gullone & Cummins, 1999	Assessment tool covering subjective & objective domains of life for research & applied purposes.	Child-report
Dartmouth COOP Functional Health Assessment Charts	Wasson et al., 1994	Survey instrument for evaluating treatment outcomes & detecting important problems, for use in the classroom or physician's office.	Child-report
Exeter Quality Life Measure/ Exqol	Eiser et al., 2000	Computer-delivered measure of quality of life for children based on experience with chronically ill children.	Child-report
Functional Status II(R)	Stein & Jessop, 1990	Can measure health status of children across wide age-range; especially suitable for children with chronic physical conditions who are not disabled.	Parent-report
Generic Children's Quality of Life Measure/GCQ	Collier, 1997; Collier et al., 2000	Allows comparison between chronically ill children & the general child population.	Child-report
Instrument for monitoring adolescent health issues	Stanton et al., 2000	Survey instrument to monitor health status & health-related behaviour in secondary school students.	Child-report
Juvenile Wellness & Health Survey/JWHS-76	Steiner et al., 1998	School-based screening tool to assess general & mental health in adolescents.	Child-report
Pediatric HealthQuiz	Goldbloom et al., 1999	Screen for potential child health problems, including psychosocial, accident prevention & home safety issues. Could be used at population level, or for evaluation of interventions, especially preventative.	Parent-report
Pictorial Scale of Perceived Competence & Social Acceptance for Young Children	Harter & Pike, 1984	Scores may be useful in determining behaviour & motivations, & for assessing sub-groups of children under different types of stress.	Child-report
Quality of Life Profile-Adolescent Version	Raphael et al., 1996	To assess coping & functioning, identify service needs, develop health-enhancing environments, assess effects of illness & treatment.	Child-report
Warwick Child Health & Morbidity Profile	Spencer & Coe, 1996	Measure of health & morbidity suitable for research, service-planning, measuring cross-sectional & longitudinal health & morbidity.	Parent-report

Table IV: *Instrument dimensions (number of items)*

Child Health & Illness Profile/CHIP-AE	**Child Health Questionnaire (parent-completed short-form)**	**Child Health Questionnaire (child-completed)**	**Child's Health Self-Concept Scale**	**Children's Health Rating Scales**	**Child Health Status Questionnaire**	**Comprehensive Quality of Life Scale/ComQOL**	**Dartmouth COOP Functional Health Assessment Charts**	**Exeter Health-related Quality Life Measure/Exqol***
Satisfaction with health (overall health & self-esteem) (12)	General health perceptions (6)	General health perceptions (12)	Psychosocial (13)	Current health quality (3)	Physical health (13 for 5-13 yrs, 5 for 0-4 yrs)	Material well-being (5)	Physical (1)	Symptoms (sleep, aches, food allergies, sickness) (4)
Discomfort (physical & emotional symptoms, limitations of activity) boys (44) girls (45)	Physical functioning (6)	Physical functioning (9)	Physical health (8)	Current illness state (3)	Mental health (12 for 5-13 yrs)	Health (5)	Emotional (1)	Social well-being (2)
Achievement (academic & work performance) (11)	Bodily pain (2)	Bodily pain (2)	Healthiness (3)	Current comparative health (3)	Social relations (3 for 5-13 yrs)	Productivity (5)	School work (1)	School achievements (1)
Risks (individual risks, threats to achievement, peer influences) (39)	Role/social-physical (2)	Role/social-physical (3)	Values (5)	Resistance to illness (5)	General health (7 for 0-13 yrs)	Intimacy (5)	Social support (1)	Physical activity (3)
Resilience (family involvement, problem-solving, physical activity) (20)	Role/social-emotional-behavioural (3)	Role/social-emotional (3)	Energy (5)	Health outlook (3)	Satisfaction with development (4 for 0-4 yrs)	Safety (5)	Family communications (1)	Worry (1)
Disorders (conditions) (45)	Mental health (5)	Role/social-behaviour (3)				Place in community (5)	Health habits (1)	Family relationships (1)
Home safety & health (not expected to behave as scale) (12)	Behaviour (6)	Mental health (16)				Emotional well-being (5)		
[CHIP taxonomy: satisfaction, discomfort, risks & resilience]	Self-esteem (6)	Behaviour (17)						
[Modified CHIP-AE excludes limitations of activity, work performance, home safety & health, recurrent disorders, long-term medical & surgical disorders, & psychosocial disorders]	Parental impact-emotional (3) Parental impact-time (3) Family activities (6) Family cohesion (1) Change in health (1)	Self-esteem (14) Family activities (6) Family cohesion (1) Change in health (1)						

Functional Status II(R)	**Generic Children's Quality of Life Measure/GCQ**	**Instrument for monitoring adolescent health issues****	**Juvenile wellness & health survey/JWHS-76**	**Pediatric HealthQuiz**	**Pictorial Scale of Perceived Competence & Social Acceptance for Young Children**	**Quality of Life Profile-Adolescent Version**	**Warwick Child Health & Morbidity Profile**
General health (15)	General affect (worry, happiness) (6)	Tobacco use	General risk taking (17)	Medical (pregnancy, perinatal health, child development, past illnesses, operations, accidents, symptoms, family history) (200)	Cognitive competence (6)	Physical being (6)	General health status (1)
Hospitalisations (3)	Peer relationships (5)	Alcohol use	Mental health problems (10)	Preventative (family relationships, nutrition, preventive health care, psychosocial issues such as mental illness, behavioural & educational problems) (175)	Physical competence (6)	Psychological being (6)	Acute minor illness status (1)
Age-specific behaviour >I year-old (5), 1 year old (13), >2 years old (23) [short version:14 items for all]	Attainments (4)	Other substance abuse	Sex-related risks (17)		Peer acceptance (6)	Spiritual being (6)	Behavioural status (1)
		Leisure					
	Relationship with parents (4)	Sun exposure	Eating & dietary problems (7)		Maternal acceptance (6)	Physical belonging (6)	Accident status (1)
	General satisfaction (1)	Injury	General health problems (11)			Social belonging (6)	Acute significant illness status (1)
	Support (2)	Dietary habits	*Other (14)****			Community belonging (6)	Hospital admission status (1)
	Health/appearance (3)	Exercise & fitness				Practical becoming (6)	Immunization status (1)
		Sexual health				Leisure becoming (6)	Chronic illness status (1)
		Mental health				Growth becoming (6)	Functional health status (1)
		Violence					Health-related quality of life (1)
		Safety					

* dimensions yet to be proposed by instrument's author; grouped in this report as a guide only

** not possible to group items on the information given

*** items do not form coherent factor

Table V: *Population evaluations*

Instrument	Study	Population	Mean age (range)	Sex/ethnicity/socio-economic status
Child Health & Illness Profile (CHIP-AE)	Starfield et al., 1993	121 adolescents: acutely or chronically ill & healthy USA	(11-17)	more girls than boys & more black adolescents than white
	Starfield et al., 1995	3451 middle & high-school students USA	(11-17)	53% female, 10-98% white (3 samples), urban & rural communities
	Starfield et al., 1996	877: sub-sample from Starfield et al., 1995, plus 3 samples of chronically ill children USA	14.3	54% female, 88% African American, mean socio-economic status score 77
	Riley et al., 1998 a & b	4019: amalgamation of previous samples (Starfield et al., 1993 & 1995) USA	14.0-14.6 across samples (11-17)	48-57% female, 3-89% minority, mean socio-economic status score 53-77.
	Chen & Chen, 1999	338 schoolchildren USA	(14-17)	72% female, 99% African American, urban area
Child Health Questionnaire	Landgraf et al., 1998	100 asthmatic children* UK	8.9 (5-13)	46% female, 78% white
		411 general population children USA	11.5 (4-19)	45% female, 82% white, 50% with at least some college education
	Landgraf & Abetz, 1997	278 schoolchildren USA	13 (10-15)	58% female, estimated 92% African-American
	Waters et al., 2000	5414 schoolchildren Australia compared against Landgraf et al., 1998 sample	11.58 (5-18)	49.6% female
	Waters et al., 1999	249 parents of schoolchildren (primary & secondary schools) compared against Landgraf et al., 1998 sample Australia	8.8 (5-12) & 13.9 (12-18)	37.5% & 52% female, 38% of primary school parents were from overseas, socio-economic diversity
		171 schoolchildren (secondary school) compared against Landgraf & Abetz, 1997 sample	13.9 (12-18)	52% female, 17% born overseas
Child's Health Self-concept Scale (CHSCS)	Hester, 1984	681children USA	9.45** (7-13)	51% female, rural communities
Children's Health Rating Scales	Maylath, 1990	1201 schoolchildren USA	4^{th}-6^{th} graders (9-12)	male & female; schools covering rural, metro, suburban & town areas
Child Health Status Questionnaire	Eisen et al., 1979	2152 children USA	6.3 (0-13)	48% female, 77.5% white, low income families slightly over-sampled
	Diaz et al., 1986	120 children with high, average & low use of medical services USA	11.2	about 1/3 white, 16% fathers in professional occupation
Comprehensive Quality of Life Scale (ComQOL)	Gullone & Cummins, 1999	264 schoolchildren Australia	14.9 (12-18)	44% female, included students of Asian origin, socio-economic status normally distributed
Dartmouth COOP Functional Health Assessment Charts	Wasson et al., 1994	658 adolescents USA	Median 15 (12-21)	54% female; 60% non-Hispanic whites, 29% Hispanic, 6% black, 5% other ethnicity
Exqol	Eiser et al., 2000	69 children UK	7.49 (6-11)	100% white, 41% male, range of social backgrounds

* this is included since it is the only study to assess the American-to-English translation of the CHQ, albeit with a chronically ill population

** the validity & test-retest samples were slightly younger: 9.03 & 8.92, respectively

Instrument	Study	Population	Mean age (range)	Sex/ethnicity/socio-economic status
Functional Status II(R)	Stein et al., 1990	276 healthy children USA	(0-16)	11% mothers white, 30% without health insurance
Generic Children's Quality of Life Measure (GCQ)	Collier, 1997 Collier et al., 2000	71 & 91 schoolchildren UK 720 schoolchildren UK	(7-11) 10.3 (6-14)	both sexes, mixed inner-city & non-affluent urban areas 52% girls, schools from different socio-economic districts
Instrument for monitoring adolescent health issues	Stanton et al., 2000	479 schoolchildren Australia	years 9 to 11 (secondary school)	schools from different socio-economic districts
Juvenile Wellness & Health Survey (JWHS-76)	Steiner et al., 1998	1769 high school students USA	15.9 (10-18)	48% girls, 60% white, suburban areas, modal socio-economic status upper middle class
Pediatric HealthQuiz	Goldbloom et al., 1999	100 attendees at paediatric ambulatory care centres, USA	(1 month-12 years)	90% female, 31% black, 13% had not completed high school
Pictorial Scale of Perceived Competence & Social Acceptance for Young Children	Harter & Pike, 1984	90 pre-school children, 56 at kindergarten, 65 first-graders, 44 second-graders USA	4.45, 5.54, 6.32 (6-7), 7.41 (7-8)	approx. 50% female, middle-class neighbourhood, 96% white
Quality of Life Profile-Adolescent Version	Raphael et al., 1996	160 adolescents Canada	17.4 (14-20)	62% female, racially homogenous, mean socio-economic status score 47.20
Warwick Child Health & Morbidity Profile	Spencer & Coe, 1996	47 attendees child health clinic (CHC), 30 attendees child development unit (CDU), 51 attendees paediatric outpatient department (OPD) UK	20, 33 & 24 months* (0-5 years)	43% from most deprived areas, 10% names indicating Indian origin

* studies 1, 2 & 3, respectively

Table VI: *Reliability of the instruments*

Instrument	**Internal consistency** (Cronbach's alpha unless otherwise stated)	**Test-retest** (Pearson correlation coefficients unless otherwise stated)
Child Health & Illness Profile/CHIP-AE	0.41-0.92 [excluding one particularly low alpha of 0.02] (Starfield et al., 1993)* 0.40-0.93 [across samples] (Starfield et al., 1995) 0.59-0.90 (Starfield et al., 1998) 0.56-0.83 modified CHIP-AE (Chen & Chen, 1999)	0.53-0.87 (Starfield et al., 1995)
Child Health Questionnaire	0.61-0.94 parent (UK, Landgraf et al., 1998) 0.59-0.93 parent (US, Landgraf et al., 1998) 0.66-0.93 parent (Australia, Waters et al., 1999) 0.60-0.93 parent (Australia, Waters et al., 2000) 0.63-0.89 child (USA, Landgraf & Abetz, 1997) 0.75-0.90 child (Australia, Waters et al., 1999)	no interim health event reported: ICC 0.49-0.78; Spearman 0.54-0.73; interim health event reported: ICC 0.08-0.77; Spearman 0.18-0.77 (parent, Australia, Waters et al., 2000)
Child's Health Self-Concept Scale/CHSCS	0.70 (Hoyt reliability coefficients ranged 0.48-0.80, total 0.86)	0.44-0.58
Children's Health Rating Scales	0.83 (range 0.78-0.85 across age groups)	-
Child Health Status Questionnaire	0.53-0.87 (across dimensions & age groups)	-
Comprehensive Quality of Life Scale/ComQOL	0.75-0.83 (across dimensions, age & sex)	0.40-0.88
Dartmouth COOP Functional Health Assessment Charts	0.60-0.94	0.71 to >0.80
Exeter Health-related Quality Life Measure/Exqol	exceeded 0.64 for all scales	-
Functional Status II(R)	0.84-0.94 (for ill & healthy samples combined, across age ranges, short & long forms)	
Generic Children's Quality of Life Measure/GCQ	0.74 (perceived-self score) 0.78 (quality of life score)	-
Instrument for monitoring adolescent health issues	-	0.27-0.99 (across content areas & ages)
Juvenile Wellness & Health Survey/JWHS-76	0.57-0.80	-
Pediatric HealthQuiz	-	Medical Peds: 67-80% agreement Prevent Peds: 79-89% agreement (range across different administration formats)
Pictorial Scale of Perceived Competence & Social Acceptance for Young Children	0.85-0.89 (total scale, across age groups), range across dimensions 0.50-0.85 (across ages)	-
Quality of Life Profile-Adolescent Version	0.94 (0.67-0.74)	-
Warwick Child Health & Morbidity Profile	-	0.50-0.86 (weighted kappa)

* An earlier version of the CHIP, subsequently revised

Table VII: *Validity of the instruments**

Instrument	Inter-instrument relationships	Proxy ratings	Demographic variables	Intra-instrument relationships
Child Health & Illness Profile/CHIP-AE	State-Trait Anxiety Inventory: children & emotional discomfort scale r=0.67 Children's Depression Inventory & emotional discomfort scale r=0.68 Family Assessment Device: general functioning scale & family involvement scale r=0.59 Discriminant validity shown by CDI & self-esteem scale r=-0.40 (Starfield et al., 1995)	Parent & child agreement ranged r=0.16-0.51 (Starfield et al., 1995)	Most expected differences in score between healthy & ill groups were observed; differences relating to sex, ethnicity & age. (Starfield et al., 1993) Reported academic performance & actual grades ranged 0.34-0.54. Sex differences as predicted for satisfaction, physical fitness, risky behaviour & social relationships; older adolescents engaged in more risky behaviour; some differences for socio-economic status. (Starfield et al., 1995) Differences in score between acutely ill & healthy teenagers found in 5/20 sub-domains & between chronically ill & healthy teenagers in 12/20 sub-domains. Substantial differences between health status of acutely & chronically ill teenagers. (Starfield et al., 1996)	All sub-domains expected to correlate moderately (r=–0.002 to 0.56) (Starfield et al., 1995) Range of sub-domain relations within each domain r=0.17-0.74 (Riley et al., 1998 a & b)
CHIP taxonomy			No statistically significant differences between profiles for socio-economic status. Eight profile distributions differ significantly by age. Boys more likely to be in profile-types reflecting high risk-taking; girls more often in profiles reflecting dissatisfaction, discomfort & the worst health. Adolescents in two-biological-parent families significantly more likely to have good health; youths with a mental disorder significantly more likely to be in the worst profile-types.	Achievement was worst for those with poor health & risk-taking behaviours. Those in poor health had worse disorders scores.
Modified CHIP-AE			Expected, significant gender differences in all domains except resilience; fewer significant age effects.	Correlations between domains ranged –0.11 to –0.42.
Child Health Questionnaire	Behaviour scale & separate behaviour item r=-0.50 Mental health & reports of anxiety r=0.35 Mental health & reports of depression r=-0.31 Behaviour scale & factored composition of anxiety, behaviour, depression & sleep r=-0.40 (all significant) (Australia, Waters et al., 2000)		8/9 CHQ scales able to discriminate between the schoolchildren & two groups of children with chronic diseases (but children with attention deficit hyperactivity disorder reported better scores than the healthy children on 3 scales). As age increased, children produced significantly worse scores on bodily pain, mental health & behaviour scales. (Landgraf & Abetz, 1997)	
Child's Health Self-Concept Scale		Parents & teachers completed replicas of the CHSCS. Multi-trait multi-method approach used. Some support for convergent validity; no support for discriminant validity.		
Children's Health Rating Scales	Four individual items developed in the Rand Health Insurance Experiment r=-0.22 to 0.53 (all significant & in expected direction)		Significant difference between mean score of paediatric asthma patients & general children: t=1.60 at the 0.10 level of significance	

Instrument	Inter-instrument relationships	Proxy ratings	Demographic variables	Intra-instrument relationships
Child Health Status Questionnaire		Parent-completed HSQ showed highly statistically significant differences for general health ratings, anxiety & depression. Children's responses similar but not statistically significant. (Diaz et al., 1986)	Functionally limited children reported to have significantly worse health status as measured by all scales & illness counts; proxy's own health status ratings generally significantly associated with rating of child's health status. (Eisen et al., 1979)	Almost all associations in the hypothesised direction (gamma coefficients); general health dimensions interrelated median = 0.37; mental health dimensions = 0.56; general health ratings correlated significantly with almost all adult ratings of own health; general health ratings & mental health scales correlated lower than mental health scales & social relations (Eisen et al., 1979)
Comprehensive Quality of Life Scale/ComQOL	No consistent pattern between Fear Survey Schedule for Children-II & subjective QOL (contrary to hypothesis). Fear Survey Schedule for Children II correlated with subjective QOL as hypothesised r=-0.14 to -0.32. Fear & anxiety correlated with objective QOL (contrary to expectations) r=-0.13 to -0.47			
Dartmouth COOP Functional Health Assessment Charts	Items measuring similar constructs, r=0.52 to 0.74. Items not expected to correlate strongly r=0.04-0.67. Higher chart scores corresponded with higher yield of detected problems (75% of respondents indicating use of drugs in a survey responded 'all the time' to health habit chart)		Health habits chart scores significantly associated with recognised 'at risk' behaviour of 138 adolescents exhibiting behavioural problems	
Exeter Health-related Quality Life Measure/Exqol			Significant differences in score between general & chronically ill children ($F=5.94$, $p<0.05$)	
Functional Status II(R) (short version only)	Separate global evaluation of health question, r=-0.29	FS II(R) scores correlated moderately in expected direction with clinical ratings	Means for well children significantly higher than for ill children for every scale & age group. Days in bed in past 2 weeks, r=-0.58 Days absent in past 2 weeks, r=-0.28 Hospitalisations in past 6 months, r=-0.13 Days hospitalised in past 6 months, r=-0.10	
Generic Children's Quality of Life Measure/GCQ	General 'happy with life' question, correlations significant & ranged r=0.31-0.51 (1997 & 2000 studies) r=0.58 (perceived self) & r=0.50 (quality of life)			

Instrument	Inter-instrument relationships	Proxy ratings	Demographic variables	Intra-instrument relationships
Juvenile Wellness & Health Survey/JWHS-76	Coping Response Inventory - Youth Form: approach coping had significant negative correlations with 4/5 dimensions r=0.04 to –0.21; avoidance coping had significant positive correlations with all dimensions r=0.12-0.18		Socio-economic status had significant negative correlations with 4/5 dimensions r=-0.12 to -0.13 ($p<0.001$); for 4/5 dimensions, girls had significantly higher mean scores than boys, though boys were expected to score more highly in areas of general risk; older subjects reported higher general & sexual risk-taking behaviours, as expected.	All between-dimension correlations were significant & in expected direction. Indicators of deception in the questionnaire correlated significantly & positively with higher risk on all dimensions r=0.05 to 0.10 ($p<0.05$)
Pictorial Scale of Perceived Competence & Social Acceptance for Young Children	Correlation between maternal acceptance scale & authors' depression/cheerfulness measure was 0.48	Child & teacher judgments range 0.06 (social acceptance: non-significant) to 0.37 (cognitive competence $p<0.001$)	Mean cognitive competence scores of children kept back a year at school significantly lower than scores of those promoted. Perceived peer acceptance scores of children who had recently joined the school significantly lower than others. Physical competence scores of children born pre-term significantly lower than full-term infants.	Two social acceptance scales intercorrelated most highly (0.62-0.80), two competence scales less so (0.43-0.56)
Quality of Life Profile-Adolescent Version	Hypothesised correlations: Satisfaction with Life index, r = 0.51; Rosenberg Self-Esteem measure r = 0.51; Social Support index r=0.32-0.52; Life Chances Questionnaire r=0.24-0.37. Little differentiation among correlations with various sub-scales, contrary to expectations. Health status correlated 0.30 ($p<0.01$) with overall QOL (range 0.06-0.36 across sub-scales). All coefficients significant.		Only one sub-scale related to socio-economic status, contrary to expectations.	
Warwick Child Health & Morbidity Profile		Health records highly correlated with parent-reporting; all reports of chronic illness confirmed; parent-reporting inconsistent with immunization status in 10% of children; hospital admission status wrongly reported in two cases; global parent report versus medical judgment (where paed-iatricians had no access to child but only to parents' second-tier responses) range 0.70-0.95 (weighted kappa).		Chronic illness & functional impairment, health level & experience of acute significant/frequent minor/chronic illnesses &/or hospital admissions, acute significant illness &/or chronic illness & hospital admission, chronic illness & loss of health-related quality of life, chi-square analyses all $p<0.001$

* validity has not been evaluated for the Instrument for monitoring adolescent health issues or the Pediatric HealthQuiz

Chapter 4: INSTRUMENT SUMMARIES

Child Health and Illness Profile/CHIP

Description

The CHIP-AE is a child-report measure for children aged 11-18, which the authors comment may be particularly useful for evaluating community and school health service programmes. It aims to document the state of health in adolescent populations, identify differences in the health of sub-populations, assess the impact of health service interventions on health, and provide an initial assessment of adolescent health for screening services.

It consists of six domains (and 20 sub-domains): satisfaction with health (self-perceptions of overall health and self-esteem), discomfort (physical and emotional symptoms, and limitations of activity), achievement (including academic performance), risks (threats to subsequent health), resilience (characteristics protecting future health), and disorders (conditions). It also contains socio-demographic questions and questions about health service utilisation. The final instrument has 126 items, plus 46 disease-specific or injury-specific questions.

The following recall periods are built into the instrument: 28 days for items on discomfort, the family involvement sub-domain of the resilience domain, and threats to achievement; one year for conditions in the disorder domain, current status for satisfaction, items in the problem-solving, home safety and health sub-domains of the resilience domain, items in the peer influence sub-domains of the risks domain, and last reported experience in the individual risks sub-domain; four weeks or two years (depending on the item) in the school achievement sub-domain; and four weeks for the work items.

Items are scored from a minimum of 0 to a maximum of 4. A score for each domain is derived by averaging a person's responses to items in that domain, where 70% of items are answered. In the current (revised) scoring, higher scores indicate better health. Responses to the CHIP are reported to cover the full range of options provided.

The CHIP-AE was modified by others (Chen and Chen, 1999), omitting seven sub-domains of CHIP-AE not directly related to school health-programme planning, in order to reduce respondent burden. It contains six domains and 13 sub-domains from the original instrument.

A taxonomy of health-profile types describing adolescents' patterns of health as self-reported on the health status questionnaire has also been developed (Riley et al., 1998 a & b). Individuals are assigned to mutually exclusive and exhaustive groups, characterising the important aspects of their health and need for health services. Four domains of health (satisfaction, discomfort, risks, and resilience) were used to group individuals into 13 distinct profile-types, describing distinct patterns of health and health service requirements. They identify sub-groups having distinct needs for health services, with potential utility for health policy and planning. The profiles are designed to characterise individuals according to their functioning across all domains. Individuals are assigned to one profile only.

A version for children aged 6-11 (CHIP-CE), and a parallel parent version, have recently been developed. Each item in this younger child version is illustrated. This child-completed instrument contains 45 items assessing five domains (satisfaction, comfort, resilience, risk avoidance, and achievement) and 4 demographic questions. The parent version includes the child items (without the illustrations) and some additional optional items, including a domain

of disorders and medical conditions (children do not report on their disorders). Both versions report on symptoms and signs of illness and well-being, health-related behaviour, problem behaviour, school performance, and involvement with family and peers. Most items assess frequency or degree, typically over the previous four weeks.

Preliminary psychometric data, unpublished but available from the authors, is promising: in a US sample of 1708 children (mixed sex and ethnicity), factor analysis was generally supportive of domains, and internal consistency reliability alphas ranged from 0.64 to 0.85 across age and sex. Test-retest correlations ranged from 0.64 to 0.78 across domains. No validity data has yet been presented. The ability of children in the middle of the elementary school age-range to complete this instrument was supported in extensive cognitive interviewing studies (Rebok, Riley, Forrest, et al., in press).

Both the comprehensive and the standard versions of the parent form for young children have been evaluated (again, unpublished) with 583 parents. Both versions were generally supported by factor analysis, and internal consistency reliability estimates ranged from 0.63 to 0.89 (across versions, sex, and age). Test-retest correlations ranged from 0.63 to 0.86 (excluding one very low correlation of 0.36). No validity data has yet been presented. The validation testing has been done and validation manuscripts on the CHIP-CE are being prepared; these are expected to be available by autumn 2001 (personal communication).

Item generation
The items in the CHIP-AE were generated from literature reviews, focus groups (children, adolescents, and parents), health professionals, and expert panels. Nine healthy adolescents provided comments on the language and content of the instrument. Item-total correlations were examined in the development of the instruments, which led to the removal and rewording of items, although 20 items with poor item-total correlations were retained for conceptual reasons (Starfield, 1995).

The validation of item placement in domains and sub-domains was conducted by ten health experts; it is also reported that factor analysis demonstrated the integrity of the sub-domains (Starfield, 1993) whilst second-order factor analysis led to the reorganisation of some domains (Starfield, 1995 op.cit.). As a result of validity, reliability and factor structure testing, revisions were made to the original CHIP. Efforts to reduce respondent burden and improve reliability of items led to the simplification of response formats and reduction of the number of response options (ibid.).

Acceptability
Although the earliest version of the CHIP-AE took an average of 45 minutes to complete (Starfield, 1993 op.cit.), the current CHIP-AE can be completed in 30 minutes (Starfield, 1995 op.cit.). Both the CHIP taxonomy and the Modified CHIP-AE took 20 minutes to complete.

Response and completion rates
Failure to complete the CHIP-AE was due either to absence on the day of testing or to refusals. Response rates by location ranged from 62% to 92% with absences and refusals generally equally divided. Observed completion rates by sub-domains ranged from 1.1% (physical discomfort sub-domain of discomfort scale in one area) to 46.1% (threats to achievement sub-domain of risks domain). In general, completion rates were poorer the later a sub-domain appeared in the questionnaire (ibid.).

Child Health Questionnaire/CHQ

Description

The Child Health Questionnaire is intended as a instrument to measure and compare the health of general and specific groups of children, and to evaluate interventions. There are various versions of the Child Health Questionnaire. It has been published as a parent/proxy form for children aged 5-13 (CHQ-PF98) and as a child self-completion form for children aged 10-18 (CHQ-CF87). In response to demand, a shortened parent version was constructed (CHQ-PF50) using regression techniques and item-scaling analysis. For larger population studies, an even shorter parent version was devised: CHQ-PF28. Although parent and child versions of the CHQ are available, they are not parallel. A shortened child-completed form is under development. The CHQ is currently being anglicised by Eiser and colleagues in Swansea (Eiser & Morse, op.cit.).

Each dimension is measured along three parameters: status, disability, and personal evaluation, with each version yielding a health profile comprising 12 or 13 concepts and two summary component scores. Scales are scored so that higher scores equal better health. To generate a scale score, at least half the items in a scale must be completed. The raw scale score is summed, transformed into a mean and a point on a continuum of 0-100. The change in health item is scored on a 0-5 continuum. The multi-item scales (except the family activities scale) are used to calculate the psychosocial and physical summary scores.

All items are based on a recall of health status over the previous four weeks, except change in health (which assesses health over the past year), general health perceptions, and family cohesion (there is no recall period for the last two). General health perceptions, behaviour, and family cohesion all include a stand-alone global item. The various versions of the CHQ contain the following concepts: physical functioning, role/social-physical, general health perceptions, bodily pain, parental time impact, parental emotional impact, role/social-emotional/behavioural, self-esteem, mental health, general behaviour, family activities, family cohesion, and change in health. Evidence suggests these are essential components of children's health-related quality of life. Scales use Likert-type categories with between four and six response options.

Item generation

Items were generated from multiple sources, viz. comprehensive literature reviews (including the adult quality of life literature), interviews, and focus groups with parents and children. Items were constructed to be relevant for girls and boys of varying ethnicity and socio-economic background. Consistently low item-total correlations were observed for five items: three general health items, one behavioural scale item, and one mental health item, but these were retained for theoretical reasons. The structure of the CHQ has been supported by factor analysis (Waters et al., 2000), although the two summary scores were not supported for use with general populations (ibid.).

Data quality

i. Parental form

Scaling success rates were very high in the Australian evaluation of the CHQ-PF50 (ibid.). Multi-trait analysis was used in the Australian sample, which reported item internal consistency (percentage of items with Pearson ≥0.40) for the parent forms of 90% (ages 12-18) and 96% (ages 5-11), whilst corresponding figures for discriminative validity were 98% and 99% (Waters et al., 1999). For the parent version, 85% of the UK sample and 87% of the

US sample met item internal consistency criteria (Landgraf et al., 1998). Item discriminant validity success rates ranged from 96% to 100% for the UK sample, and from 83% to 100% for the US sample (ibid.).

Negligible floor effects were observed, although ceiling effects exceeded 50% in three UK cases and six US cases (ibid.), whilst ceiling effects exceeded 50% in six cases from the Australian sample (Waters et al., 1999 op.cit.). In a separate Australian evaluation, only six items had item-scale internal consistency values lower than the 0.4 criterion, whilst perfect item discriminant validity success rates were observed for eight of 11 multi-item scales (Waters et al., 2000 op.cit.).

For the CHQ-PF28, scaling test results (item internal consistency and discriminant validity) were good (Landgraf and Abetz, 1996 op.cit.).

ii. Child form

Floor effects were again minimal whilst ceiling effects exceeded 50% for four scales (Waters et al., 1999 op.cit.). Multi-trait scaling techniques employed with the CHQ-CF87 showed that perfect success rates in terms of item internal consistency were observed among 6/10 CHQ scales, although low item correlations were found for the General Health, Mental Health and Behaviour scales, consistent with findings in the parent-completed version (Landgraf and Abetz, 1997). Scaling success rates for the child version for item discriminant validity exceeded 93%; consistent responses were observed for 70% of the child-completed version (ibid.). Item internal consistency was 84% and discriminant validity 98% in the Australian evaluations (Waters et al., 1999 op.cit.). Floor effects were negligible whilst ceiling effects exceeded 50% in four cases. The Australian evaluation of the CHQ-CF87 recommended the reduction in the number of items to 80, given the poor performance of several items (ibid.).

Acceptability

In one study response rates were fairly low, with just over 50% of parents and children returning questionnaires (ibid.). 231 parents (92.5% response rate) replied to a feedback questionnaire; 90% reported no problem completing the questionnaire. 83 students replied to the feedback questionnaire (48% response rate); 65% found the questionnaire very easy to understand, and 34% found it hard to understand or confusing. 81% stated they felt fine about their parent filling out a similar questionnaire, and the same percentage stated there were no questions they did not like or felt bad about reporting. 17% reported there were questions they did not like or felt bad about reporting (ibid.). For the CHQ-PF50, 72% of the parents in Australia responded, compared with 68% of the original US sample (Waters et al., 2000 op.cit.).

Another study reported that 63% of questionnaires were fully completed (Landgraf and Abetz, 1997 op.cit.). Completion rates tended to be lower for children aged 10-12 (53-60%) than children aged 13 and over (72-74%) (ibid.). 1% were excluded from the psychometric evaluation of the Australian sample since over 50% of the data for that scale were missing (Waters et al., 2000 op.cit.). 9/100 UK cases were excluded due to missing data (Landgraf et al., 1998). Consistent responses were observed for 70% of the school sample (Landgraf and Abetz, 1997 op.cit.).

Although the CHQ-CF87 is designed for children aged 10-18, in the Australian evaluation it was limited to children aged 12-18, since earlier data had reportedly shown that children aged 10-12 took 45-60 minutes and needed help to complete it, compared with 15 minutes without help for older children.

Child's Health Self-Concept Scale/CHSCS

The CHSCS is a child-report instrument designed to measure a child's perception of his or her health-related behaviours, for use in nursing research and practice. It is based on a health continuum, with positive health perceptions at one end and negative health perceptions at the other. It is potentially useful for nursing research and practice, and knowledge of an individual's health self-concept can be useful in the planning and evaluation of interventions, especially in the area of health promotion.

There are five items for ten sub-scales, six items for one sub-scale (emotion) and two items for one sub-scale (health). Children are presented with a bipolar structure, with one pole representing a positive health perception and the other a negative health perception. Score 1 is given to the negative pole 'really true', score 2 to the negative pole 'sort of true', score 3 to the positive pole 'sort of true', and score 4 to the positive pole 'really true'. The highest possible score on the CHSCS is 232 and the lowest 58. High scores indicate a positive health self-concept and low scores, a negative health self-concept.

Items were generated by asking a convenience sample of 225 children aged 6-13 what they think a healthy and an unhealthy child is like. 12 categories were identified: nutrition, physical health, sleep, dental health, friends, healthiness, family, play, activity and exercise, personal grooming, emotional, and non-specific. Expert review of the draft instrument was conducted by nursing and education professionals, and a group of 40 children aged 5-13. The instrument was empirically tested with two other groups of children, on the basis of which several items were deleted (too few responses to one pole or low factor loadings). The sub-scale of play was completely eliminated through item deletion.

The final instrument consist of 34 items in five sub-scales: psychosocial (13 items), physical health (8 items), healthiness (3 items), values (5 items), energy (5 items). Five factors were generated in the factor analysis, but this solution was unstable. Evidence suggests that only one factor is being measured since item-total correlations exceeded item/sub-scale correlations.

The longer draft version of the instrument (with 41 items) took approximately 25 minutes to complete.

Children's Health Ratings Scale

The Children's Health Rating Scale is a 17-item child-report scale designed as a report of general health in children for group comparison. It uses a five-point scale defined as 'true', 'mostly true', 'don't know', 'mostly false', and 'false'. Higher scores reflect more favourable ratings of health. It assesses current health and illness, resistance to illness, and health outlook.

The original instrument consisted of 22 items adapted from the General Health Ratings Index, which was constructed by factor analysing adult responses to the general health perception items of the Rand Corporation's Health Insurance Experiment. A teacher was consulted to make the items readable for fourth-graders. It was administered to a sample of 25 second- to sixth-grade subjects who provided information on readability and ease of administration. Pilot-testing followed with a sample of 137 fourth- to sixth-grade students, and five items exhibiting low inter-item correlation were deleted.

Most item means were found to be slightly above the midway score. The distribution tended to be negatively skewed and flatter than a normal distribution. The observed range of values (n=58) approached the possible range of 68.

Principal factor analysis was employed and five factors were identified with eigenvalues above 1.0, which accounted for 55.8% of the total variance in responses. The factors were labeled 'Current Health Quality', 'Current Illness State', 'Current Comparative Health', 'Resistance to Illness', and 'Health Outlook'. However, no prior hypotheses were made concerning the factor structure.

Small, statistically significant correlations with grade, sex and social desirability scores were not considered a threat to validity.

Child Health Status Questionnaire

This measure of health status was developed for use in the Health Insurance Study, which was designed to test the effects of different health care financing arrangements on health status. It was designed as a parent-report measure, although in one study (Diaz et al., 1986) it is reported that children also completed the measure. The Child Health Status Questionnaire aims to measure physical, mental, and social components of health, and general health.

Physical health is defined in terms of functional performance and capacity with regard to specific categories, including self-care (e.g. bathing), physical (e.g. walking), mobility (e.g. confinement indoors), and role activities (e.g. schoolwork). Mental health focusses on psychological states, such as mood and feelings, and assessed both positive and negative states. Social relations encompasses interpersonal interactions (home, school, and neighbourhood). General health ratings are defined with respect to time (current and prior health) and resistance/susceptibility to illness.

Questionnaires are specific to two age ranges: 0-4 and 5-13, the division marking the start of schooling. Items for those aged 0 to 4 years relate to functional limitations, satisfaction with development, and general health perceptions, whilst for those aged 5-13 the instrument contains items relating to functional limitations, mental health, social health, and general health perceptions.

Scores for scales are computed using the simple algebraic sum of scores for items, after reversing where necessary. In addition, 12 mental health items are combined to construct a Mental Health Index, and seven general health items are combined to construct a General Health Rating Index. Response categories vary according to the item, with between four and six options.

In terms of item generation, categories were selected from those found in the children's and adults' physical health literature. Questionnaire items representing these categories were similar to those used in previous studies of children and adults in general populations. Categories and items were reviewed by physician consultants to assess face validity and age appropriateness. To reduce the influence of other non-health items, almost all items contained a phrase focussing on the health-relatedness of limitations.

The number of children showing any physical health problems was small. For children aged 0-4 years, 96% were free of limitations; for 5-13 year-olds, the percentage was 94%. For the mental, social, general health, and satisfaction with development scores, distributions were skewed, with mean values consistently on the favourable side.

It is reported that, in general, the pattern of rotated factor loadings strongly supported the hypothesised item groupings. Item-scale correlations exceeded 0.30 (the criterion value) for all except one item. Missing responses ranged between 0.3% and 6.2% (Eisen et al., 1979). Item-scale correlations exceeded 0.30 (the criterion value) except in one instance in the younger population. The scaling errors observed appeared to be site-related.

Comprehensive Quality of Life Scale/ComQol

The ComQol was originally developed and evaluated for use with adults, as an assessment tool covering subjective and objective domains of life for research and applied purposes. The child-report version assesses subjective and objective quality of life in seven domains: material well-being, health, productivity, intimacy, safety, place in the community, and emotional well-being. Two changes were made to the instrument to enhance its relevance to adolescents: adolescents were asked to nominate parents' occupation rather than income, and the number of response options for the satisfaction items was reduced from seven to five.

For the subjective dimension, the participant rates each item twice, once for importance and once for satisfaction. Scores range from one (terrible/not at all important) to five (delighted/could not be more important). There are seven satisfaction items and seven importance items, one for each domain. These are then combined by weighting satisfaction scores by importance scores. Each subjective score can range from +20 to –20. The ComQol yields scores on several parameters: total scores (sum across all domains) for the objective, satisfaction, importance, and subjective (satisfaction x importance) domains. The objective scale comprises three items for each domain.

No ceiling or floor effects were reported. Parents of students from randomly selected classes were given consent forms and, with the exception of those absent from school on the day of testing, there was 100% participation.

Dartmouth COOP Functional Health Assessment Charts

The COOP child-report charts have been developed as a survey instrument to evaluate treatment outcomes and as a tool for the detection of important health problems. It consists of six charts addressing physical fitness, emotional feelings, schoolwork, social support, family communications, and health habits. Respondents answer using a five-point scale with a score of five indicating the worst possible scores. Items relate to the previous month.

Items for the COOP chart were generated by means of a literature review of available measures, from which 17 potential categories were defined; picture-and-word charts corresponding to these categories were designed. Focus groups, consisting of 51 primary care physicians and 31 adolescents, rated the importance of the 17 picture-and-word charts. Following focus groups, the number of picture-and-word charts was reduced to 14. Six charts were dropped on the grounds of poor validity and reliability results.

Over 50% of the 658 respondents reported chart scores of one or two (the categories signifying best health) on four of six scales, and the distribution of scores was reported to be influenced by age and sex. In a sub-sample of 360 teenagers from New England, those who completed charts in the physician's offices generally had better scores than their peers in schools, so responses may be affected by mode of administration.

The 188 teenagers who completed the charts were asked to compare the relative ease of answering and the honesty of their responses for the two assessment methods (questionnaires versus picture-and-word charts). 27% considered the charts to be easier to understand and 7% claimed the questions were easier. 7% felt the charts might induce dishonest responses, as opposed to 21% who thought questionnaires might do this.

Exeter quality of life measure/Exqol

The Exqol aims to be a generic child-report measure of quality of life for children aged 6-12, based on the authors' experience with chronically ill children. It is computer-delivered and consists of 12 sex-specific pictures, each of which is rated twice: first in terms of 'like me' and second in terms of 'as I would like to be'. The theoretical model is based on an assumption that poorer quality of life is the result of discrepancies between an individual's actual and ideal self, which is based on observations of children with chronic illnesses. Most items are framed in the social context, as this seemed to be an important factor in young children's lives.

The instrument is completed under supervision and items are read aloud twice, to eliminate reliance on reading ability. The use of computers with gender-specific picture stimuli, where children indicate responses by clicking an on-screen visual analogue scale, is designed to make it more fun. For each of the 12 items, two ratings are recorded by the computer: the actual self score and the ideal self score; the difference between the two is then calculated for each item. The mean absolute discrepancy is calculated for the 12 items, yielding a possible range from 0 to 100. Discrepancy scores are calculated so that a high score represents a poor quality of life.

Items were generated on the basis of literature reviews and clinical experience with children. The Exqol takes approximately 20 minutes to complete and children were reported to have had no problems with using the mouse. A response rate of 57% was observed (mainly due to parents not returning consent forms). There were no significant effects for age or sex on discrepancy scores.

Functional Status II(R)

The Functional Status II(R) is designed to measure children's health status across a wide age-range and is especially intended for children with chronic physical conditions, although it has also been used with a general population. The parent-completed FS II(R) has a long version consisting of 43 items, and a short version consisting of 14 items. It is a revised version of the Functional Status Measure FS I which was developed to measure individual child health status and to characterise populations, and is modelled on the Sickness Impact Profile (Eiser & Morse, 2001 op.cit.).

The parent-completed measure considers behavioural manifestations of illness that interfere with an individual's performance of age-appropriate activities. The elements in the conceptual framework are communication, mobility, mood, energy, play, sleep, eating, and toilet patterns as they interfere with normal social role performance in three sites (home, neighbourhood, and school) during leisure, work, and rest activities. Responses are indicated on two three-point scales, indicating whether the item occurs 'never or rarely', 'some of the time' or 'almost always', whether it is due 'fully', 'partly' or 'not at all' to a health problem.

Items for the original FS I were generated from literature reviews, interviews with mothers and health care professionals, and clinical experience. Items have pairs of questions: one regarding the child's behaviour and the other, whether this was related to illness. Behavioural questions relate to specific ages: infants (0-9 months), toddlers (9-23 months), pre-school children (2-5 years), and school-age children (over 5 years). A small number of items overlap all age ranges.

The score is the percentage of possible points for that scale and age. Items on each scale are summed, the score is then divided by the total possible score for that scale and multiplied by 100. The internal structure of both versions is supported by factor analysis.

Generic Children's Quality of Life Measure

The Generic Children's Quality of Life Measure was designed to be (a) suitable for use with chronically ill children as well as children in the general population, (b) based on children's reports rather than adults' perceptions of quality of life, (c) child-friendly, and (d) able to consider the degree to which things matter to each individual child.

Children are asked to complete the questionnaire by relating to the responses of children in a story: first, by answering questions relating to the child they feel they are most like and second, by answering questions relating to the child they would most like to be. Each of the questions score 1-5, with 1 for never and 5 for always, and scores reversed on ten items. To determine quality of life, the discrepancy between the perceived and desired scores are calculated. The discrepancy scores are then transformed so that higher scores indicate a higher quality of life.

The final version of the instrument has 25 items, and the authors suggest the GCQ may be appropriate for use with nearly all linguistically able children. In one study, children aged six completed the questionnaire under close supervision (two children and the researcher), children aged 7-10 were supervised in groups of four, and children aged 11-14 completed the measure in class groups (Collier et al., 2000). The authors of this study conclude that the instrument is suitable for use over a wide age-range (6-14) without the scores being confounded by age, sex, geographic location, or social deprivation. After evaluation with the two populations listed in Table III and a chronically ill group (not detailed here), the instrument was amended to produce a final version with 25 items.

Items were generated by 80 children aged 6, 11 and 13 who were approached in schools and asked to identify what made their lives good or bad, i.e. both positive and negative influences. Following item generation, the draft questionnaire comprised 22 questions covering general affect (happiness, worrying), peer relationships (friends, bullying), attainments (sport, academic), relationships with parents (like their parents, told off by parents), and one general satisfaction question (how much of the time they feel happy with their lives).

Both the perceived self and the QOL scores were normally distributed and there was no evidence that younger children were failing to discriminate across the range of responses (ibid.). The possible score range for the perceived self is 24-120 and observed was 51-112; corresponding values for the quality of life scores were 0-100 and 27-199, respectively. Factor analysis is reported to have revealed eight sub-scales (ibid.).

Where schools provided the information, there was a non-return rate of consent-forms of 34.2%; 4.3% of parents refused, as did 0.2% of pupils; 2.5% were absent on the day of testing. Therefore, 58.6% of children and their parents consented to participation, were present on the day of the study, and were actually tested. 93.5% of children completed all 25 self-perceived items and 91.3% completed all preferred-self items. It is reported that children found the GCQ easy to use (over 90% of children said they had fun completing it) and there were few administrative difficulties (ibid.).

Instrument for monitoring adolescent health issues

This child-report instrument is designed to monitor health status and health-related behaviours in secondary school students; it aims to determine the prevalence of a range of health issues and health behaviours, in order to identify clusters of negative health outcomes. It was created in the context of a national policy framework for children and young people in Australia, and contains domains relating to tobacco use, alcohol use, other substance use, sun exposure, leisure, dietary habits, exercise and fitness, sexual health, mental health, violence, safety, and injury. It consists of questions for both senior and junior students (aged 12-18). The questionnaire for junior students does not include items relating to sexual health, contraception, and pregnancy.

Regarding item generation: first, the literature was examined, from which a list of possible items and scales were compiled; possible questions were then circulated among experts. Second, workshops with health professionals were held to identify and finalise criteria and discuss items. On the basis of these workshops, draft questionnaires were developed for different ages and forwarded to workshop participants for comments. Third, focus groups were held with students to discuss difficulties with understanding items, and how participants felt about the items. A draft version was developed, to which some amendments were made after pilot-testing (mainly the response category options and the layout).

The instrument takes around 35 minutes to complete, but those with lower levels of literacy skill are reported to need more time to complete it. Feedback from school staff was said to be supportive and favourable with regard to the choice of issues.

Juvenile Wellness and Health Survey/JWHS-76

The JWHS-76 aims to be a comprehensive child-report screening instrument, assessing both mental and general health of adolescents in the school context, from a clinical child psychiatrist's perspective. By developing service profiles and individual risk profiles, it seeks to aid in the planning of school-based and school-linked services. The requirements of the screening instrument were: simplicity, administration in one class period, assessment of multiple domains of mental and physical functioning, and a balanced representation of mental and general health, as well as specific risk-behaviours.

The JWHS-76 contains 104 questions covering general health, mental health, risk-taking behaviour, socio-demographic information, and health-care habits. Of the 104 questions, 76 are lead questions answered by everyone; the remainder are follow-up questions. Scales range from one to five, with five indicating poorer outcomes.

An unpublished instrument applied in another school was used as the basis for item generation. Relevant health and school professionals were consulted regarding content areas, and the language was simplified to that of a fifth-grader. Three focus groups of high school students then completed the instrument, and discussed its problems and inadequacies.

All lead questions were skewed and a principal components analysis was performed on the 76 lead questions, generating a five-factor solution. 14 lead questions were excluded because of poor factor loadings.

Students were given one class period to complete the instrument. The response rate among the students was 99% (parents had to opt out if they did not want their child to participate). 1769 questionnaires were collected; however, after removing those with fictitious responses, 1755 questionnaires remained. 32% of questionnaires were complete. Responses could be computed for others, giving a figure of 79%.

Pediatric HealthQuiz

The Pediatric HealthQuiz is based on reports by parents; it is designed to screen for a range of paediatric health issues and provide a comprehensive health database for paediatric patients. The questionnaire contains 375 items and is divided into two modules: Medical Peds covering biomedical issues, and Prevent Peds covering prevention, psychosocial, educational, and safety topics. More specifically, Medical Peds covers pregnancy, perinatal health, child development, past illnesses, operations and accidents, symptoms, and family history. Prevent Peds covers family relationships, nutrition, preventive health care, and psychosocial issues such as mental illness, behavioural, and educational problems. Three response options are available: ‘yes’, ‘no’, or ‘not sure’.

It is administered via an Internet application using a touch screen. The computerised questionnaire is structured using a decision-tree, and is designed so that only questions appropriate to age and sex are asked. Two reports are generated: a physician’s report summarising potential health issues, and a patient’s report suggesting steps that could be taken - such as in injury prevention.

The items were developed by the author, who submitted a draft to four expert paediatric reviewers; interviews to assess the questionnaires were then conducted with 132 parents at paediatric ambulatory clinics in the US. As a result of these steps, the total number of items was reduced from 478 to 375.

The mean time to complete the Prevent Peds module was 12 minutes (range 6-30 minutes) and for the Medical Peds module, 19 minutes (range 11-31 minutes). 40/50 parents felt the HealthQuiz was not too long, and the majority are said to have described it as interesting or enjoyable, and as providing important information for the doctor. 47/50 parents were not upset by any of the questions asked.

Pictorial Scale of Perceived Competence and Social Acceptance for Young Children

The Pictorial Scale of Perceived Competence and Social Acceptance for Young Children may be useful for determining behaviour and motivation, and for assessing children under stress. It is a child-report instrument consisting of 24 items divided into two main constructs and four sub-scales: general competence (cognitive and physical), and social acceptance (peer acceptance and maternal acceptance). There are six items per sub-scale. The existence of two main constructs was confirmed by factor analysis.

It is a self-completion measure designed for children aged four to seven. There are two versions of the instrument: the first is for pre-school children and those in kindergarten (four and five year-olds), the second for first- and second-graders (six and seven year-olds). This was because the specific skills reflecting competence and social acceptance change between these ages; the younger child's version also excludes the self-worth scale. Both versions use pictorial formats rather than a written questionnaire, and there are gender-specific sets of pictures for both age-groups.

The instrument is reported to have undergone numerous revisions in terms of scale structure, item content and question format, and was based on extensive piloting with large numbers of subjects. The child-respondent is faced with two pictures: a girl or boy who is good at an activity, and one who is not. The child then indicates which of the two he or she is most like, and whether he or she is a lot or a little like that child. The pictures are given to the child and the item is read by an examiner.

Each item is scored on a four-point scale, whereby four indicates the most competent or accepted and one, the least competent or accepted. Item scores are averaged across the six items for a given sub-scale, and the four means provide the child's profile of perceived competence and social acceptance. A teacher's rating scale parallels the child's instrument, though the maternal acceptance sub-scale is excluded. Interviews with children as to the reasons why they answered a particular way showed that they could provide definite reasons for their alleged competencies.

The version for pre-school/kindergarten children has been applied with predominantly African-American children of low socio-economic status, participating in an urban Head Start programme. However, this evaluation did not find internal validity (factor analysis) for the instrument; it also found that children did not understand the concepts of quantity, and were unable to identify pictures based on verbal descriptions (Fantuzzo et al., 1996).

Quality of Life Profile - Adolescent Version/QOLPAV

The QOLPAV measures child-reported health from a broad quality-of-life perspective; the authors suggest it could be used to assess current states of coping and functioning, identify adolescents' service needs, develop health-enhancing environments, and assess the effects of illness and treatments. These concepts have previously been operationalised with the elderly and people with developmental disabilities. It consists of 54 items, each of which is rated for importance and enjoyment/satisfaction on a five-point scale. It covers three main aspects of adolescent functioning in nine sub-domains: being (physical, psychological, and spiritual), belonging (physical, social, and community), and becoming (practical, leisure, and growth). There are six items in each of nine domains. About 50% of items are specific to adolescents.

The scales range from one (not at all important/no satisfaction at all) to five (extremely important/extremely satisfied). Importance scores serve as a weight for converting satisfaction scores into quality of life scores. Quality of life scores can range from –3.33 (extremely important and no satisfaction) to +3.33 (extremely important and extremely satisfied). In addition, single items address the amount of control and opportunities the adolescent perceives in each of the nine sub-domains. These items are not part of the quality of life score computation but provide contextual information. Control scores can range from one (almost no control) to five (almost total control), as do opportunity scores (from 'almost none' to 'a great many'). Quality of life scores were found to be normally distributed with virtually no skewing.

Items were generated using group meetings mainly of high-school students across a range of grades (9-13), although separate meetings with guidance counsellors were also held. Responses were collected and developed into items. The adolescent development and health literature was also drawn upon to generate items. The draft instrument was pilot-tested with a class of 20 adolescents, and modifications were made. Scores were normally distributed.

Factor analysis was conducted for the nine sub-domain scores, which generated three factors. It was not conducted for individual items, since there were too few respondents to meet the standards for factor analysis.

Administration of the instrument took 40 minutes.

Warwick Child Health and Morbidity Profile/WCHMP

The Warwick Child Health and Morbidity Profile is a parent-report measure of health and morbidity in infancy and childhood. It is suitable for research and planning purposes, and capable of measuring both cross-sectional and longitudinal health and morbidity experience in a child population. It provides a parent's perception of the child's health, illness, functional health status, and health-related quality of life.

The instrument consists of ten domains: general health status, acute minor illness status, behavioural status, accident status, acute significant illness status, hospital admission status, immunization status, chronic illness status, functional health status, and health-related quality of life. Each consists of a single item with four response-categories. Details of acute minor illness, behaviour problems, accidents, hospital admission, acute significant illness, and chronic illness are obtained using second-tier questions. Domains are not weighted or scored; it is a profile of a child's health and illness experience.

In the first phase of testing, the global questions were tested and parents were invited to explore the meaning of concepts like health, using a series of open questions. Modifications were made in the domain questions to improve comprehensibility and acceptability.

The WCHMP takes a maximum of ten minutes to complete. Inter-observer variation between the researcher and the family health visitor was found to be low.

Chapter 5: CONCLUSIONS

The present review has focussed on generic, multi-dimensional instruments evaluated for use in general populations of children. The review was based on a comprehensive search strategy including several electronic databases and hand-searching of key literature. The reference-lists of all published articles retrieved were checked for relevant papers.

The search strategy produced 16 instruments that were included in the review. Only four of these instruments had undergone evaluation in a UK population. There was a surprisingly high proportion of child self-completion measures, although these tended to be more common for older age-groups.

There was some consensus as to important domains, but considerably less agreement concerning specific items within domains. The constructs of quality of life, health status, and functioning were not clearly separated. However, there was more variety in the mode of administration than is usually found in instruments for adults.

As regards recommending individual instruments, it is essential to assess an instrument according to the purpose of its application. Once basic psychometric criteria are fulfilled, the main issue is whether the instrument provides relevant information. When measuring health-related quality of life in children, the chief concerns are whether child or proxy reports are more acceptable, given the different data they provide, and whether the instrument is appropriate for use with the particular age-group being evaluated. Most of the instruments have not yet been fully evaluated in a UK setting, or provide only preliminary data. It is, therefore, important that any application of these instruments be accompanied by an evaluation of their use.

The measures developed in the UK are the Exqol, Generic Children's Quality of Life Measure, and the Warwick Child Health and Morbidity Profile. The first two measures are child-completed and suitable for young children (aged six and over). Both cover similar domains, including school performance and family relationships; the Exqol also covers symptoms. They are designed to be easy and fun to complete: the Generic Children's Quality of Life Measure takes a storybook format, whilst children respond to the Exqol via a computer. Neither addresses risk-taking behaviour, which population-level interventions may seek to change, and neither includes an accompanying parent version. If used, both would need further concurrent testing of test-retest reliability and validity. The Warwick Child Health and Morbidity Profile is a different type of measure. It elicits information from the parent concerning health service contacts and health status for infants and young children. It would need to undergo internal consistency reliability testing.

The Child Health Questionnaire and the CHIP have been the most extensively evaluated from a psychometric standpoint; results suggest these instruments are reliable and valid. Both could be recommended for use as self-completion measures with children aged 11 and over, although they are rather long (shorter versions of the CHQ-CF87 are being developed). Most reviews conducted to date concur that the CHQ has much to recommend it. Data relating to UK populations is lacking for both measures at the time of writing, but a UK evaluation of the CHQ is currently underway. Versions of both instruments have been developed for younger children, although the younger child version of the CHQ is parent-completed.

The CHIP-CE is potentially the most interesting instrument for younger children. This child-completed measure is designed for use with children as young as six, and includes risk-taking

behaviour and school-functioning. Importantly, there is an accompanying parent version which includes items on specific childhood disorders, and would allow for comparison between child and proxy responses. Preliminary data suggest this is a reliable instrument in a US setting, and validity data are soon to be available. Application of this instrument in the UK would require concurrent reliability testing and validity testing.

Although evaluations of non-English language instruments were excluded from this review, one such instrument, the KINDL, warrants mention. This child-reported questionnaire is suitable for ill and healthy children aged eight to 16 years, and assesses psychological well-being, physical state, social relationships, and functional capacity. Preliminary reliability and validity results are promising (Salek, 1998 op.cit.) although, again, testing in a large English-speaking population would be required.

REFERENCES

Apajasalo, M., Sintonen, H., Holmberg, C., Sinkkonen, J., Aalberg, V., Pihko, H., Siimes, M. A., Kaitila, I., Makela, A., Rantakari, K., Anttila, R., and Rautonen, J. (1996a). Quality of life in early adolescence: A sixteen-dimensional health-related measure (16D). *Quality of Life Research* **5**, 205-211.

Apajasalo, M., Rautonen, J., Holmberg, C., Sinkkonen, J., Aalberg, V., Pihko, H., Siimes, M. A., Kaitila, I., Makela, A., Erkkila, K., and Sintonen, H. (1996b). Quality of life in pre-adolescence: A 17-dimensional health-related measure (17D). *Quality of Life Research* **5**, 532-538.

Barr, R. D., Pai, M. K. R., Weitzman, S., Feeny, D., Furlong, W., Rosenbaum, P., and Torrance, G. W. (1994). A multi-attribute approach to health status measurement and clinical management - Illustrated by an application to brain tumors in childhood. *International Journal of Oncology* **4**, 639-648.

Boyle, M. H., Furlong, W., Feeny, D., Torrance, G. W., and Hatcher, J. (1995). Reliability of the Health Utilities Index-Mark III used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. *Quality of Life Research* **4**, 249-257.

Bradlyn, A. S., Harris, C. V., Warner, J. E., Ritchey, A. K., and Zaboy, K. (1993). An investigation of the validity of the quality of Well-Being Scale with pediatric oncology patients. *Health Psychology* **12**, 246-250.

Bullinger, M. and Ravens-Sieberer, U. (1995). Health related QOL assessment in children: A review of the literature. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée* **45**, 245-254.

Chen, S. P. and Chen, E. H. (1999). Application of modified CHIP-AE in a vocational high school. *ABNF Journal* **10**, 104-110.

Collier, J. (1997). Developing a generic child quality of life questionnaire. *The British Psychological Society, Health Psychology Update* **28**, 12-16.

Collier, J., MacKinlay, D., and Phillips, D. (2000). Norm values for the Generic Children's Quality of Life Measure (GCQ) from a large school-based sample. *Quality of Life Research* **9**, 617-623.

Colver, A. and Jessen, C. (2000). Measurement of health status and quality of life in neonatal follow-up studies. *Seminars in Neonatology* **5**, 149-157.

Connolly, M. A. and Johnson, J. A. (1999). Measuring quality of life in paediatric patients. *PharmacoEconomics* **16**, 605-625.

Diaz, C., Starfield, B., Holtzman, N., Mellitis, E., Hankin, J., Smalky, K., and Benson, P. (1986). Community-Based Assessment of Morbidity in Children. *Medical Care* **24**, 848-856.

Dossetor, D. R., Liddle, J. L. M., and Mellis, C. M. (1996). Measuring health outcome in paediatrics: Development of the RAHC Measure of Function. *Journal of Paediatrics and Child Health* **32**, 519-524.

D. Drotar, ed., (1998). "Measuring health-related quality of life in children and adolescents: Implications for research and practice." Mahwah, NJ, USA: Lawrence Erlbaum Associates, Inc., Publishers. (1998). xiii, 372 pp..

Eisen, M., Ware, J. E. J., Donald, C. A., and Brook, R. H. (1979). Measuring components of children's health status. *Medical Care* **17**, 902-921.

Eisen, M., Donald, C. A., Ware, J., and Brook, R. H. (1980). Conceptualization and measurement of health for children in the Health Insurance Study. R-2313-HEW, 314pp..

Eiser, C., Havermans, T., Craft, A., and Kernahan, J. (1995). Development of a measure to assess the perceived illness experience after treatment for cancer. *Archives of Disease in Childhood* **72**, 302-307.

Eiser, C., Vance, Y., and Seamark.D (2000). The development of a theoretically driven generic measure of quality of life for children aged 6-12 years: a preliminary report. *Child: Care, Health and Development* **26**, 445-456.

Eiser, C. and Morse, R. (2001a). A review of measures of quality of life for children with chronic illness. *Archives of Disease in Childhood* **84**, 205-211.

Eiser, C. and Morse, R. (2001b). Quality-of-life measures in chronic diseases of childhood. *Health Techology Assessment* **5**(4).

Erling, A. (1999). Methodological considerations in the assessment of health-related quality of life in children. *Acta Paediatrica, International Journal of Paediatrics, Supplement* **88**, 106-107.

Fantuzzo, J., McDermott, P., Holliday Manz, P., Hampton, V., and Burdick, N. (1996). The pictorial scale of perceived competence and social acceptance: does it work with low-income urban children? *Child Development* **67**, 1071-1084.

Fayers, P. M. and Machin, D. (2000). "Quality of Life: Assessment, Analysis and Interpretation." John Wiley and Sons, Ltd., Chichester, UK.

Feeny, D., Furlong, W., Barr, R. D., Torrance, G. W., Rosenbaum, P., and Weitzman, S. (1992). A comprehensive multi-attribute system for classifying the health status of survivors of childhood cancer. *Journal of Clinical Oncology* **10**, 923-928.

Finkelstein, J. W. (1998). Methods, models, and measures of health-related quality of life for children and adolescents. *In* "Measuring health-related quality of life in children and adolescents: Implications for research and practice" (D. Drotar, ed. op. cit.), pp. 39-52.

Fitzpatrick R., Davey, C., Buxton, M. J., Jones, D. R. (1998). Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* **2**(14).

Gaudin, J. M. Jr., Polansky, N. A., and Kilpatrick, A. C. (1992). The Child Well-Being Scales: a field trial. *Child Welfare* **71**, 319-328.

Glaser, A. W., Furlong, W., Walker, D. A., Fielding, K., Davies, K., Feeny, D. H., and Barr, R. D. (1999). Applicability of the health utilities index to a population of childhood survivors of central nervous system tumours in the U.K. *European Journal of Cancer* **35**, 256-261.

Goldbloom, R. B., Kim, R. K., Hodder, M. C., Mingay, D. J., Summerell, D., Lee, J., Randel, P., and Roizen, M. F. (1999). Design and reliability of pediatric HealthQuiz: preliminary report of a comprehensive, computerized, self-administered child health assessment. *Clinical Pediatrics, Philadelphia* **38**, 645-654.

Graham, P., Stevenson, J., and Flynn, D. (1997). A new measure of health-related quality of life for children: Preliminary findings. *Psychology and Health* **12**, 655-665.

Gullone, E. and Cummins, R. A. (1999). The comprehensive quality of life scale: A psychometric evaluation with an adolescent sample. *Behavioural Change* **16**, 127-139.

Harter, S. and Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development* **55**, 1969-1982.

Hester, N. (1984). Child's health self-concept scale: its development and psychometric properties. *Advanced Nursing Science* **7**, 45-55.

Jordan, T. (1983). Developing an international index of quality of life for children: the NICQL Index. *Journal of Research in Social Health* **103**, 127-130.

Kaplan, R. M. and Anderson, J. P. (1988). A general health policy model: Update and applications. *Health Services Research* **23**, 203-235.

Kozinetz, C. A., Warren, R. W., Berseth, C. L., Aday, L. A., Sachdeva, R., and Kirkland, R. T. (1999). Health status of children with special health care needs: Measurement issues and instruments. *Clinical Pediatrics* **38**, 525-533.

Landgraf, J. and Abetz, L. (1996). Measuring Health Outcomes in Pediatric Populations: Issues in Psychometrics and Application. *In* "Quality of Life and Pharmacoeconomics in Clinical Trials. 2nd edition" (B. Spilker, ed.), pp. 793-802. Lippincott-Raven Publishers, Philadelphia, USA.

Landgraf, J. M. and Abetz, L. N. (1997). Functional status and well-being of children representing three cultural groups: Initial self-reports using the CHQ-CF87. *Psychology and Health* **12**, 839-854.

Landgraf, J. M., Maunsell, E., Speechley, K. N., Bullinger, M., Campbell, S., Abetz, L., and Ware, J. E. (1998). Canadian-French, German and UK versions of the child health questionnaire: Methodology and preliminary item scaling results. *Quality of Life Research* **7**, 433-445.

Landgraf, J. M. and Abetz, L. (1998). Influences of sociodemographic characteristics on parental reports of children's physical and psychosocial well-being: Early experiences with the Child Health Questionnaire. *In* "Measuring health-related quality of life in children and adolescents: Implications for research and practice" (D. Drotar, ed. op. cit.), pp. 105-126.

Lansky, S. B., List, M. A., Lansky, L. L., Ritter, S. C., and Miller, D. R. (1987). The measurement of performance in childhood cancer patients. *Cancer (Philadelphia)* **60**, 1651-1656.

Levi, R. and Drotar, D. (1998). Critical issues and needs in health-related quality of life assessment of children and adolescents with chronic health conditions. *In* "Measuring health-

related quality of life in children and adolescents: Implications for research and practice" (D. Drotar, ed. op. cit.), pp. 3-24.

Lewis, C. C., Pantell, R. H., and Kieckhefer, G. M. (1989). Assessment of children's health status. Field test of new approaches. *Medical Care* **27**, S54-S65.

Lindstrom, B. and Eriksson, B. (1993). Quality of life among children in the Nordic countries. *Quality of Life Research* **2**, 23-32.

Marra, C., Levine, M., McKerrow, R., and Carleton, B. (1996). Overview of health-related quality-of-life measures for pediatric patients: application in the assessment of pharmacotherapeutic and pharmacoeconomic outcomes. *Pharmacotherapy* **16**, 879-888.

Maylath, N. S. (1990). Development of the Children's Health Ratings Scale. *Health Education Quarterly* **17**, 89-97.

Munzenberger, P. J., Van Wagnen, C. A., Abdulhamid, I., and Walker, P. C. (1999). Quality of life as a treatment outcome in patients with cystic fibrosis. *Pharmacotherapy* **19**, 393-398.

Neff, E. J. and Dale, J. C. (1990). Assessment of quality of life in school-aged children: a method - phase I. *Maternal Child Nursing Journal* **19**, 313-320.

Pal, D. K. (1996). Quality of life assessment in children: A review of conceptual and methodological issues in multi-dimensional health status measures. *Journal of Epidemiology and Community Health* **50**, 391-396.

Pantell, R. H. and Lewis, C. C. (1987). Measuring the impact of medical care on children. *Journal of Chronic Diseases* **40**, 99S-108S.

Raphael, D., Rukholm, E., Brown, I., Hill-Bailey, P., and Donato, E. (1996). The quality of life profile - Adolescent version: Background, description, and initial validation. *Journal of Adolescent Health* **19**, 366-375.

Ravens-Sieberer, U. and Bullinger, M. (1998). Assessing health-related quality of life in chronically ill children with the German KINDL: First psychometric and content analytical results. *Quality of Life Research* **7**, 399-407.

Rebok, G., Riley, A. W., Forrest, C., Starfield, B., Green, B. F., Robertson, J. A., and Tambor, E. (2001). Development of a child health status questionnaire using cognitive interviewing methods (in press). *Quality of Life Research*.

Riley, A. W., Forrest, C. B., Starfield, B., Green, B., Kang, M., and Ensminger, M. (1998a). Reliability and validity of the adolescent health profile-types. *Medical Care* **36**, 1237-1248.

Riley, A. W., Green, B. F., Forrest, C. B., Starfield, B., Kang, M., and Ensminger, M. E. (1998b). A taxonomy of adolescent health: development of the adolescent health profile-types. *Medical Care* **36**, 1228-1236.

Salek, S. (1998). Compendium of Quality of Life Instruments. New York: Wiley.

Seaberg, J. R. (1988). Child Well-Being Scales: A critique. *Social Work Research and Abstracts* **24**, 9-15.

Schor, E. L. (1998). Children's Health and the Assessment of Health-Related Quality of Life. *In* "Measuring health-related quality of life in children and adolescents: Implications for research and practice" (D. Drotar, ed. op. cit.) pp. 25-37.

Spencer, N. J. and Coe, C. (1996). The development and validation of a measure of parent-reported child health and morbidity: the Warwick Child Health and Morbidity Profile. *Child: Care, Health and Development* **22**, 367-379.

Spieth, L. E. and Harris, C. V. (1996). Assessment of health-related quality of life in children and adolescents: An integrative review. *Journal of Pediatric Psychology* **21**, 175-193.

Stanton, W. R., Willis, M., and Balanda, K. P. (2000). Development of an instrument for monitoring adolescent health issues. *Health Education Research* **15**, 181-190.

Starfield, B., Bergner, M., Ensminger, M., Riley, A., Ryan, S., Green, B., McGauhey, P., Skinner, A., and Kim, S. (1993). Adolescent health status measurement: Development of the child health and illness profile. *Pediatrics* **91**, 430-435.

Starfield, B., Riley, A. W., Green, B. F., Ensminger, M. E., Ryan, S. A., Kelleher, K., Kim, H. S., Johnston, D., and Vogel, K. (1995). The adolescent child health and illness profile. A population-based measure of health. *Medical Care* **33**, 553-566.

Starfield, B., Forrest, C. B., Ryan, S. A., Riley, A. W., Ensminger, M. E., and Green, B. F. (1996). Health status of well vs ill adolescents. *Archives of Pediatrics and Adolescent Medicine* **150**, 1249-1256.

Starfield, B. and Riley, A. (1998). Profiling health and illness in children and adolescents. *In* "Measuring health-related quality of life in children and adolescents: Implications for research and practice" (D. Drotar, ed., op. cit.), pp. 85-104.

Stein, R. E. and Jessop, D. J. (1990). Functional status II(R). A measure of child health status. *Medical Care* **28**, 1041-1055.

Stein, R. E. and Jessop, D. J. (1991). "Functional status II(R): A measure of child health status": Erratum. *Medical Care* **29**, 490-491.

Stein, R. E. and Jessop, D. J. (1991). Manual For The Functional Status II(R) Measure.

Steiner, H., Pavelski, R., Pitts, T., and McQuivey, R. (1998). The juvenile wellness and health survey (JWHS-76): A school-based screening instrument for general and mental health in high school students. *Child Psychiatry and Human Development* **29**, 141-155.

Theunissen, N. C., Vogels, T. G., Koopman, H. M., Verrips, G. H., Zwinderman, K. A., Verloove-Vanhorick, S. P., and Wit, J. M. (1998). The proxy problem: child report versus parent report in health-related quality of life research. *Quality of Life Research* **7**, 387-397.

Varni, J. W., Seid, M., and Rode, C. A. (1999). The PedsQL: measurement model for the pediatric quality of life inventory. *Medical Care* **37**, 126-139.

Verrips, G. H., Vogels, A. G. C., Verloove-Vanhorick, S. P., Fekkes, M., Koopman, H. M., Kamphuis, R. P., Theunissen, N. C. M., and Wit, J. M. (1997). Health-related quality of life measure for children-the TACQOL(+). *Journal of Applied Therapeutics* **1**, 357-360.

Vogels, T., Verrips, G. H. W., Verloove-Vanhorick, S. P., Fekkes, M., Kamphuis, R. P., Koopman, H. M., Theunissen, N. C. M., and Wit, J. M. (1998). Measuring health-related quality of life in children: The development of the TACQOL parent form. *Quality of Life Research* **7**, 457-465.

Walker, L. S. and Greene, J. W. (1991). The functional disability inventory: Measuring a neglected dimension of child health status. *Journal of Pediatric Psychology* **16**, 39-58.

Wasson, J. H., Kairys, S. W., Nelson, E. C., Kalishman, N., and Baribeau, P. (1994). A short survey for assessing health and social problems of adolescents. *Journal of Family Practice* **38**, 489-494.

Waters, E., Wright, M., Wake, M., Landgraf, J., and Salmon, L. (1999). Measuring the health and well-being of children and adolescents: A preliminary comparative evaluation of the Child Health Questionnaire in Australia (includes commentary by Stein REK). *Ambulatory Child Health* **5**, 131-141.

Waters, E., Salmon, L., and Wake, M. (2000). The parent-form Child Health Questionnaire in Australia: comparison of reliability, validity, structure and norms. *Journal of Pediatric Psychology* **25**, 381-391.

Appendix 1: PHIG database search strategy

((acceptability or appropriateness or (component* analysis) or comprehensibility or (effect size*) or (factor analys*) or (factor loading*) or (focus group*) or (item selection) or interpretability or (item response theory) or (latent trait theory) or (measurement propert*) or methodol* or (multi attribute) or multiattribute or precision or preference* or proxy or psychometric* or qualitative or (rasch analysis) or reliabilit* or replicability or repeatability or reproducibility or responsiveness or scaling or sensitivity or (standard gamble) or (summary score*) or (time trade off) or usefulness* or (utility estimate) or valid* or valuation or weighting*) and ((COOP or (functional status) or (health index) or (health profile) or (health status) or HRQL or HRQoL or QALY* or QL or QoL or (qualit* of life) or (quality adjusted life year*) or SF-12 or SF-20 or SF?36 or SF-6) or ((disability or function or subjective or utilit* or (well?being)) near2 (index or indices or instrument or instruments or measure or measures or questionnaire* or profile* or scale* or score* or status or survey*)))) or ((bibliograph* or interview* or overview or review) near5 ((COOP or (functional status) or (health index) or (health profile) or (health status) or HRQL or HRQoL or QALY* or QL or QoL or (qualit* of life) or (quality adjusted life year*) or SF-12 or SF-20 or SF?36 or SF-6) or ((disability or function or subjective or utilit* or (well?being)) near2 (index or indices or instrument or instruments or measure or measures or questionnaire* or profile* or scale* or score* or status or survey*))))

Appendix 2: Non-English language measures excluded from the review

Instrument name	Reference	Aim	Population	Age
KINDL	Ravens-Sieberer & Bullinger, 1998	Generic psychometrically based practical self-report measure for children	Chronically ill and healthy children Germany	10-16
Nordic Quality of Life Questionnaire for Children	Lindstrom & Eriksson, 1993	Quality of life structure that can be used in studies of populations as well as individuals	General population Nordic countries	Parent-completed with input from children 2-18 years
TACQOL	Theunissen et al. 1998; Verrips et al.. 1997; Vogels et al. 1998	Generic measure intended for assessment of health-related quality of life in medical research and clinical trials	Representative sample Netherlands	Proxy (parent) or self (child) completion 8-11 Parent-completed for pre-school children
16D	Apajasalo et al., 1996a	Generic measure of HRQL in early adolescent children	Schoolchildren and children with illnesses Finland	12-15
17D	Apajasalo et al., 1996b	Generic measure of perceived HRQL	Schoolchildren and children with illnesses Finland	Self-rating of health status, with parents providing the dimension importance weights 8-11

Appendix 3: Other measures excluded from the review

Instrument name	References	Reason for exclusion
Child Quality of Life Questionnaire	Graham et al., 1997	Disease-specific populations
Functional Disability Inventory/FDI	Walker & Greene, 1991	Disease-specific populations
Health Utilities Index/HUI	Feeny et al., 1992; Glaser et al., 1999	Disease-specific populations
Global multi-attribute health status utility scores/MAHS	Barr et al., 1994	Disease-specific populations
PedsQL	Varni et al., 1999	Disease-specific populations
Perceived Illness Experience/PIE	Eiser et al., 1995	Disease-specific populations
Play performance scale for children	Lansky et al., 1987	Disease-specific populations
Quality of Well-being Scale	Munzenberger et al., 1999; Kaplan & Anderson, 1988; Bradlyn et al., 1993	Disease-specific populations, or no child/adolescent-specific evaluation
RAHC Measure of Function/MOF (modified from the Child Global Assessment Scale)	Dossetor et al., 1996	Disease-specific populations
RAND FSQ	Lewis et al., 1989	Disease-specific populations
Batelle Developmental Inventory	Identified by Bullinger & Ravens-Sieberer, 1995	Clinical screening method
Child Well-being Scales	Gaudin et al., 1992; Seaberg, 1988	Method of evaluating family functioning in social welfare programmes, USA
Health Utilities Index (HUI) mark 3	Boyle et al., 1995	General population survey (not child/adolescent-specific)
National Index of Children's Quality of Life Scores/NICQL	Jordan, 1983	A quantitative socio-economic indicator for international comparisons
NIE Functional Status Index	Identified by Bullinger & Ravens-Sieberer (op.cit.)	General population survey (not child/adolescent-specific)
Ontario Child Health Study Scales	Identified by Bullinger & Ravens-Sieberer (ibid.)	Not multi-dimensional; considers only emotional/behavioural problems
Ten-question screen for disability	Identified by Marra et al., 1996	Disability screening method for developing countries
Semi-structured interviews (obtains information regarding physical, psychosocial, and social development, while focussing on the child's activities, family life, and home environment)	Neff & Dale, 1990	No psychometric evaluation
Vineland Adaptive Behavior Scales	Identified by Bullinger & Ravens-Sieberer (op.cit.)	Clinical scale