3 Evaluation of HRQoL measures used in asthma

Over the last 20 to 30 years there has been rapid development of HRQoL measurement instruments and this field continues to evolve. As described in the preceding Chapter, there are many options for HRQoL measurement, with strengths and weaknesses applying for different purposes. The challenge is to identify the instruments with attributes that are suited to the specific population health monitoring task.

There is an increasing appreciation of the benefits of using formally evaluated and well validated measures to assess HRQoL. Although a few surveys have used multi-item, multi-dimensional instruments such as the SF-36 (e.g. Wilson et al. 2002), most general health surveys have used single item measures, both global and single dimensional, for measuring HRQoL or health status. Some brief disease-specific measures (e.g. sick days due to asthma) have also been used. In most cases there has been little or no formal evaluation of the attributes of these brief or single item instruments. They have the benefit of low cost when used in large monitoring activities. However, in Chapter 2, the limitations of these instruments, including problems with sensitivity and content validity, were identified.

In this chapter, we present the findings of a systematic review of the attributes of instruments that have been used in population studies to assess the HRQoL impact of asthma.

3.1 Review inclusion criteria

The aim was to systematically review the attributes of HRQoL measurement instruments to assess their suitability for population health monitoring tasks. Studies evaluating the reliability and validity of generic and asthma-specific HRQoL measurement instruments were identified using Medline, World Wide Web and expert input. The HRQoL measurement instruments included in the review were:

- those used to investigate populations with asthma between 1991 and June 2004;
- those used in population studies (applied to generic measures only); and
- those with formal evaluation of attributes, including validity and reliability.

In addition, we included only asthma-specific measures that had been used by multiple research groups.

It is acknowledged that there are a number of important measures that did not meet these inclusion criteria. This is because this evaluation focused on measures that had been used in population-based studies in which asthma had been one of the focuses of investigation. This was necessary for identifying evidence relevant to asthma monitoring. However, these selection criteria resulted in the inclusion of a wide range of multi-dimensional measures. A list of measures that were considered but not included in the evaluation has been compiled in Appendix B with reasons for exclusion.

3.2 Framework for assessment of HRQoL measures

A systematic approach was developed to evaluate the HRQoL measurement instruments included in this review. The purpose was to identify measures that would be sensitive to differences between populations, subgroups and changes over time; include content that was relevant to HRQoL concerns of people with asthma and, hence, be valid as measures of HRQoL impact of asthma; and also be meaningful and useful in populations with and without asthma. The framework for describing, assessing and making recommendations relating to the suitability of these instruments for population monitoring is described in Table 3.1. This framework included a rating out of six stars (see Table 3.2.).

Table 3.1: Framework for assessing HRQoL measurement instruments

Type of instrument The type of HRQoL measurement instrument: global, profile or utility measure
HRQoL domains The domains included in instrument: global, physical, psychological and social
Content areas A description of the dimensions included in each instrument
Mode of administration How the instrument was administered (e.g. self-administered, interview, computer assisted telephone survey)
Respondent burden Time effort and other demands placed on those completing the instrument
Time recall The time period over which respondents were asked to recall events
Settings used The setting(s) in which the study using the instrument was conducted
Reliability ▪ Internal consistency: the extent to which elements of the questionnaire are measuring the same domain (quantified with Cronbach's α) ▪ Test–retest repeatability: the extent to which the repeated administration of the instrument under the same conditions results in similar scores (quantified with the interclass correlation coefficient—ICC)
Validity The degree to which an instrument measures what it is supposed to measure **Content validity** The extent to which the material covered by the instruments encompasses, and is limited to, the intended purpose of the questionnaire. Provides an evaluation of the processes used to derive the content of the instrument. This includes: ▪ Source of items: source from which items for the instrument were identified, such as from focus groups (qualitative methods) or previous questionnaires; and ▪ Method of selection of items: process used to select items for inclusion in the final instrument (e.g. psychometric methods such as factor analysis). **Construct validity** The extent to which the correlation with or difference from other measures, such as markers of disease severity, accords with theoretical expectations. **Criterion validity** Describes comparisons with a gold standard. This method of assessment is not applicable to the evaluation of HRQoL measures.
Responsiveness Describes evidence of the ability of an instrument to detect changes in individuals over time

(continued)

Table 3.1 (continued): Framework for assessing HRQoL measurement instruments

Sensitivity Describes evidence of the ability of an instrument to detect differences between populations / subgroups / repeated surveys
Australian data Identifies studies implementing the instrument in Australia
Other comments Any further information that informs the overall evaluation of the instrument
Usefulness for population monitoring A star rating system used to rate the usefulness of a measure for population monitoring based on six key questionnaire attributes (see Table 3.2)

A star rating system was adopted to summarise six attributes that were selected for their relevance for population health monitoring (Table 3.2). For respondent burden, HRQoL domains, construct validity and sensitivity, the ratings categories were based on the conceptual framework described in Chapter 2. For the reliability measures (test–retest and internal consistency) cut-offs for statistical values were used that were applicable to a population monitoring context (Streiner & Norman 2001). Good ratings were assigned a black star; moderate ratings, a white star; and poor ratings (or no data), no star. An overall rating was derived by adding all the stars, whereby two white stars were equated to one black star (see Tables 3.3, 3.4, 3.6 and 3.7).

In interpreting this information, it is important to consider the relevance of specific attributes to the population monitoring tasks (as discussed in Chapter 2). The rating used in this evaluation gave all attributes equal weighting; however, some users might choose to apply weights that reflect their own resources and priorities. For example, it is acknowledged in this report that respondent burden is a particularly important issue in a population monitoring context. However, we have chosen not to give this greater weight in our evaluation because, as suggested in Chapter 2, it needs to be balanced with other attributes. These and many of the issues that need to be considered in evaluating measures are discussed in the following sections.

Table 3.2: Evaluation rating system for HRQoL instruments

Attribute	★	☆	No star
Respondent burden (RB)	<3 minutes to complete, or approximately 1–5 items	3–9 minutes to complete, or approximately 6–20 items	10+ minutes to complete, or >20 items
HRQoL domains (D)	Samples from physical, psychological and social domains	Global domain sampled	Samples one or two of physical, psychological and social domains
Construct validity (CV)	Extensive evidence (consistent with several other measures)	Some evidence	No evidence
Test–retest repeatability (T–R)	ICC>0.7	ICC 0.4–0.7	ICC<0.4
Internal consistency (IC)	Cronbach's α >0.7	Cronbach's α 0.4–0.7	Cronbach's α <0.4
Sensitivity (S)	Extensive evidence (several studies)	Some evidence	No evidence

Note: Where there was a range of values for an attribute for a questionnaire, the least favourable value was used as the basis for the rating.

3.3 Evaluation of measures in relation to monitoring tasks

The details of the review of the 30 evaluated HRQoL measures are contained in Appendix A. The star rating summary is reported in Table 3.3 (adult generic measures), Table 3.4 (adult disease-specific measures), Table 3.6 (childhood generic measures) and Table 3.7 (childhood asthma-specific measures). A more detailed interpretation of the evaluation is contained in subsequent sections.

3.3.1 Generic measures

The selection among generic measures of HRQoL represents a compromise between feasibility, on the one hand, and validity, reliability and sensitivity or discriminative capacity on the other. Single item measures are by far the most widely used generic measures of HRQoL in Australian population surveys. However, any single item measure is limited in content validity, reliability and sensitivity.

The SF-36 is a multi-item, multi-dimensional measure that has 36 questions, measures eight HRQoL dimensions and takes five to ten minutes to complete (Bousquet et al. 1994; McHorney 1993). The Sickness Impact Profile (SIP) has 136 questions, measures 12 dimensions and can take up to 30 minutes to complete (Bergner et al. 1981; Rutten-van Molken et al. 1995). Long, detailed HRQoL measurement instruments can be unattractive for use in large population health surveys because of respondent burden. This is a major limitation of the Sickness Impact Profile and is reflected in its infrequent use compared with the SF-36 in population-based studies.

More recently, shortened versions of the SF-36 have been developed such as the SF-12, which has 12 items (Ware et al. 1996). The SF-12 has been used in population studies and in people with asthma (Garratt et al. 2000) and rated relatively well in our evaluation (Table 3.3). These instruments reduce respondent burden and cost. However, the compromise is that they measure HRQoL with less precision than the longer version (Ware et al. 1996). This is more a limitation for individual monitoring, while for population monitoring they have the advantage of increased efficiency.

Healthy Days is another relatively short multi-dimensional HRQoL measure that has been used for several years in the United States Behavioural Risk Factor Surveillance System. It has four questions taking only one minute to complete. It also has a 14-question version (not included in evaluation, see Appendix B) (Hennessy et al. 1994). This measure has low respondent burden. However, its scope is restricted to the physical and psychological domains of HRQoL: 'focusing on the quality and functional impact of perceived physical and mental health during the immediate past.' (Hennessy et al. 1994:569).

Measures used to assess the impact of asthma should have a period of recall that is sufficiently long to capture intermittent symptom or exacerbation episodes but not so long that recall is unreliable. Although there is no clear evidence about appropriate recall period, clinical observation would suggest that two to four weeks may be optimal. The SF-36 and SF-12 have been evaluated for recall over the last four weeks and last week (acute). Similarly, Healthy Days measures health impacts over the last 30 days. The SIP focuses on 'today', making it less suitable for asthma monitoring based on this criterion.

Table 3.3: Ratings of usefulness for population monitoring: generic adult measures

Instrument	Respondent burden	HRQoL domains	Construct validity	Test–retest	Internal consistency	Sensitivity	Total (2☆=★)
EuroQol-5D (EQ-5D)	★	★	☆			☆	★★★
Healthy Days (CDC-HRQoL 4)	★	★	☆	★		☆	★★★★
Health Utilities Index Mark III (HUI)			☆			☆	★
Medical Outcomes Study, short form 36 (SF-36)		★	★	☆	★	☆	★★★★
Medical Outcomes Study, short form 12 (SF-12)	☆	★	★	★	★	☆	★★★★★
Nottingham Heath Profile (NHP)		★	☆		☆	☆	★★☆
Sickness Impact Profile (SIP)		★	☆	★	★		★★★☆

In summary, HRQoL profiles are not commonly used in population surveys due to respondent burden and cost. However, shorter profiles such as the SF-12 are more efficient for measuring all domains of HRQoL with acceptable validity, reliability and sensitivity and these may be used more widely in population health monitoring. An added advantage of the SF-12 is that it includes the single item health status measure often referred to as the SF-1 (Section 2.5.1), which has been used in many population surveys. Therefore, adoption of the SF-12 for population monitoring will not compromise time series based on the SF-1.

3.3.2 Disease-specific measures

In order to monitor changes in disease outcomes over time, there is value in using disease-specific measures, as these are more sensitive to the specific HRQoL issues of concern in the subpopulation with the disease of interest. The disease-specific measures for asthma that have been used in population surveys are mainly single item, single dimension measures such as 'sick days due to asthma' and 'nights woken due to asthma'. However, as noted in Chapter 2, these cannot be considered holistic measures of asthma-related quality of life. This can best be accomplished by including multi-item, multi-dimensional measures in asthma monitoring surveys. The questionnaires in Table 3.4 are potentially suitable for this task. Three of these have been extensively evaluated for use in adults with asthma: the St George's Respiratory Questionnaire (SGRQ) (Jones 1991), the McMaster Asthma Quality of Life Questionnaire (AQLQ-McMaster) (Juniper et al. 1992), and the Sydney Asthma Quality of Life Questionnaire (AQLQ-Sydney) (Marks et al. 1992). These measures were given relatively high ratings in our evaluation (Table 3.4). The original AQLQ-McMaster includes five items that are individually tailored to respondents. This design feature increases the instrument's responsiveness in longitudinal study designs, such as clinical trials. However, it makes it unsuitable for use in cross-sectional studies because the actual content of the questionnaire is not the same for all respondents. The Standardised AQLQ-McMaster (AQLQ(S)-McMaster) was developed to overcome this problem. It replaces the five variable items with five standardised items and this questionnaire is suitable for use in cross-sectional studies. However, this questionnaire has only recently been developed and has not been evaluated or used extensively at this point in time. Hence, Table 3.4 shows that the AQLQ(S)-McMaster did not rate as highly as the questionnaires referred to above.

In relation to respondent burden, the SGRQ contains more items (76) than theAQLQ-McMaster and the AQLQ-Sydney, and takes approximately 10 minutes to complete. The AQLQ-McMaster contains 32 items and takes 10–15 minutes to complete while the AQLQ-Sydney contains 20 items and takes around five minutes to complete. Therefore, the AQLQ-Sydney has the lowest respondent burden, which is an advantage when including the instrument as a component in a broader population health survey, and is reflected in its higher rating than the other measures. Briefer versions of both the AQLQ-McMaster (the Mini AQLQ-McMaster) (Juniper et al. 1999b) and the SGRQ (Paul Jones, personal communication) may make them more acceptable for use in large surveys. However, the Mini AQLQ-McMaster retains five non-standardised items, which makes it unsuitable for use in cross-sectional surveys.

The SGRQ was designed for use in people with both asthma and chronic obstructive pulmonary disease (COPD) whereas the other questionnaires are designed for use only in adults with asthma. This broader range of the SGRQ comes at the cost of less disease specificity and, hence, potentially less sensitivity and responsiveness (Sanjuas et al. 2002). The SGRQ, AQLQ-McMaster and AQLQ-Sydney have been mainly used in clinical populations of patients with asthma. However, some have been used in population-based samples of patients with asthma (Marks et al. 1997; Premaratne et al. 1999).

All three questionnaires have been shown to have good test–retest reliability: AQLQ-McMaster (intraclass correlation coefficient, ICC>0.9), SGRQ (ICC>0.9), and AQLQ-Sydney (ICC=0.8) (Appendix A: 49, 52, 57).

Of the disease-specific multi-item, multi-dimensional HRQoL questionnaires, the AQLQ-Sydney, which is the only one of these developed and tested in Australia, may be the most suitable for population monitoring purposes.

Table 3.4: Ratings of usefulness for population monitoring: disease-specific adult measures

Instrument	Respondent burden	HRQoL domains	Construct validity	Test–retest	Internal consistency	Sensitivity	Overall (2☆=★)
Asthma Quality of Life Questionnaire (McMaster) (AQLQ-McMaster)		★	★	★	★	★	★★★★★
Mini Asthma Quality of Life Questionnaire (McMaster) (Mini AQLQ-McMaster)	☆	★	☆	★	★		★★★★
Standardised Asthma Quality of Life Questionnaire (McMaster) (AQLQ(S)-McMaster)		★	☆	★	★	☆	★★★★
Sydney Asthma Quality of Life Questionnaire (AQLQ-Sydney)	☆	★	★	★	★	★	★★★★★☆
Asthma Symptom Utility Index (ASUI)	☆		☆	☆			★☆
Integrated Therapeutics Group Asthma Short Form (ITG-ASF)	☆	★	☆		★		★★★
Living with Asthma Questionnaire (Hyland) (LWAQ)		★	★	★	★		★★★★
Quality of Life for Respiratory Illness questionnaire (QoLRIQ)		★	☆	★	★		★★★☆
St George's Respiratory Questionnaire (SGRQ)		★	★	★	★	☆	★★★★☆

3.3.3 Utility scales

Utility measures were developed for use in economic evaluations. There are a number of generic multi-attribute utility indices (MAUIs), including the EQ-5D, the Health Utilities Index (HUI) (Furlong et al. 2001), the Assessment of Quality of Life (AQoL) (Hawthorne et al. 2001), and the SF-6D (Brazier et al. 1998) (see Table 3.5).

Of these, the EQ-5D is by far the most widely used with over 200 published papers relating to this instrument (reviewed in Brazier et al. 1998; Garratt et al. 2002; Hawthorne & Richardson 2001). The EQ-5D has been widely evaluated in the population context. The construct validity of this instrument as a measure of HRQoL is supported by comparison with the SF-12 and the SF-36 (Essink-Bot et al. 1997; Jenkinson et al. 1997; Johnson & Coons 1998; Johnson & Pickard 2000). Respondents who reported a problem on the EQ-5D scale also had lower mean scores in the corresponding dimensions of the SF-12 and SF-36. A major limitation identified in these studies was that the EQ-5D was prone to ceiling effects; that is, a high proportion of respondents had the highest possible score, which occurred when respondents reported no problem in all five dimensions. As a consequence, this instrument is relatively insensitive for discriminating differences in the general population where the majority of individuals do not have chronic illnesses (Guyatt et al. 1997). This represents a major limitation on the usefulness of the EQ-5D for population monitoring purposes, particularly in relation to asthma. The SF-6D is a relatively new instrument, but its derivation from the widely used SF-36 assures its wider use in the future. Disease-specific MAUIs have been developed to provide more sensitive measures for specific contexts. For example, the Asthma Symptom Utility Index (ASUI) was developed for clinical trials and cost-effectiveness studies in which reduction in symptom frequency and intensity is the primary clinical outcome (Revicki et al. 1998).

Table 3.5: Generic multi-attribute utility indices

	HUI Mark 3	EQ-5D	AQoL	SF-6D
Country of origin	Canada	United Kingdom	Australia	United Kingdom
Dimensions	8: hearing, speech, ambulation, dexterity, emotion, cognition, pain	5: self-care, usual activities, pain/discomfort, anxiety/depression	5: independent living, social relationships, physical senses, psychological wellbeing	6: role limitation, social function, bodily pain, mental health, vitality
No. of items	12	5	15	14
No. of response levels	4–6	3	4	2–6
No. of health states	972,000	243	1,073,741,824	9000
Sample for utility weights	General population	General population	General population	General population
Weights for Australia	No	No	Yes	No
Utility elicitation method	VAS/SG	TTO/VAS	TTO	VAS/SG
Utility algorithm form	Multiplicative	Regression/ Additive	Multiplicative	Additive
Range of utility weights	–0.36 to 1.00	–0.59 to 1.00	–0.04 to 1.00	+0.46 to 1.00

As noted previously, the validity of the MAUI within a specific population depends, in part, on the extent to which the weights are applicable to that population. The AQoL is the only MAUI with utility weights from an Australian sample. Thus, if any of the other MAUIs are used for Australian applications, subsequent decisions would be based on the utility weights of British, Canadian or American population samples and may not reflect the values of multicultural Australia. At this time, further work is required to develop a utility measure for use in people with asthma in Australian population monitoring.

3.3.4 Measuring HRQoL in children

Designing HRQoL indicators for children presents additional methodological challenges. A child's perspective on his or her wellbeing and functional status is dependent on the child's developmental stage and can differ greatly from the parents', carer's, or health professional's perspective (Jenney & Campbell 1997). Overall, the generic multi-item, multi-dimensional HRQoL scales that we reviewed (Table 3.6) were relatively long and, hence, had a substantial respondent burden, making them unsuitable for use in population monitoring surveys. They also tended to lack evidence for construct validity and test–retest reliability.

As for adults, there are circumstances in which it is important to measure HRQoL impacts that are specific to asthma. Several questionnaires that have been developed for this purpose are reviewed in Table 3.7. Probably the greatest challenge in measuring child and adolescent HRQoL is not only to capture the individual perspective, but also to accommodate the physical, emotional, and social changes that occur as the child develops and understands the concepts that are being addressed (Christie et al. 1993). The Childhood Asthma Questionnaires (French et al. 1998) are divided into three age groups: 4–7 years, 8–11 years and 12–16 years. This approach acknowledges that the issues relating to asthma and HRQoL are different in different stages of childhood. These measures rated moderately well in relation to other childhood measures for asthma. However, there may be insufficient power to detect differences for items that are relevant to a small age range in a sample from the general population, and none of the questionnaires rated well on the respondent burden criterion. Furthermore, the inclusion of self-completed and visual components in the administration of these surveys could be incompatible with some population health survey designs such as those administered by telephone. The particular advantages of this measurement instrument are that part of it can be administered to children without asthma, for comparison, and that it has been adapted for use in the Australian context (French 1996).

Table 3.6: Ratings of usefulness for population monitoring: generic childhood measures

Instrument	Respondent burden	HRQoL domains	Construct validity	Test–retest	Internal consistency	Sensitivity	Total (2☆=★)
Child Health and Illness Profile—Adolescent Edition (CHIP-AE)		★		☆	★	☆	★★★
Child Health Questionnaire Parent Form 50 (CHQ-PF50)		★		☆	☆	☆	★★☆
Child Health Questionnaire Parent Form 28 (CHQ-PF28)		★	☆			☆	★★
Pediatric Quality of Life Inventory (PedsQL)		★	☆		★	☆	★★★

Another example of an asthma-specific HRQoL instrument for use in children is the Pediatric Asthma Quality of Life Questionnaire (PAQLQ) (Juniper et al. 1996). This contains 23 items and takes approximately 10 minutes to complete, which, while rating low on the respondent burden criterion, is shorter than most childhood measures. It also has the advantage in population monitoring of being designed for children with asthma across a wide age range (7–17 years) and addresses the physical, psychological and social domains of health with scores for HRQoL dimensions in symptoms, activity limitations and emotional function. The child can self-complete the questionnaire (providing he or she has appropriate reading skills) or it can be administered via interview with the child.

The Adolescent Asthma Quality of Life Questionnaire (AAQLQ) (Rutishauser et al. 2001) also rates relatively highly, is designed for the 12–17 year age range, and has 32 items taking 5–7 minutes to complete. The instrument with lowest respondent burden in the evaluation of children's measures is the Integrated Therapeutics Group Child Asthma Short Form (ITG-CASF) (Bayliss et al. 2000) with only eight items. However, this instrument rates poorly in other criteria, including that the content is restricted to the physical and social domains.

The PAQLQ may be a preferable choice for population monitoring because, despite moderate respondent burden, it is designed for use across a wide age range. The AAQLQ may also be suitable for studies limited to the adolescent age range.

Table 3.7: Ratings of usefulness for population monitoring: asthma-specific childhood measures

Instrument	Respondent burden	HRQoL domains	Construct validity	Test–retest	Internal consistency	Sensitivity	Total (2☆=★)
About My Asthma		★		☆	★		★★☆
Adolescent Asthma Quality of Life Questionnaire (AAQLQ)		★	★	★	★		★★★★
Childhood Asthma Questionnaire A (CAQ-A)			★	☆	☆	☆	★★☆
Childhood Asthma Questionnaire B (CAQ-B)		★	★	★	☆	★	★★★★☆
Childhood Asthma Questionnaire C (CAQ-C)		★	☆	★	☆	★	★★★★
Children's Health Survey for Asthma (CHSA)		★	☆	☆	★		★★★
How Are You? (HAY)		★	☆	☆	★	☆	★★★☆
Integrated Therapeutics Group Child Asthma Short Form (ITG-CASF)	☆		☆		★	★	★★★
Paediatric Asthma Quality of Life Questionnaire (PAQLQ)		★	★	☆	★	☆	★★★★
Pediatric Quality of Life Asthma Module (PedsQL-Asthma Module)		★	☆		★		★★☆