2 Conceptual framework for measuring HRQoL in asthma

The development of valid and standardised methods for measuring HRQoL is challenging because of the uniqueness inherent in an individual's perception of their quality of life. Nonetheless, it is widely appreciated that measuring HRQoL as an outcome of diseases such as asthma is essential to understanding their impact (Guyatt et al. 1993; Schipper 1983). It is for this reason that standardised methods of assessment of HRQoL have been developed and validated so that comparisons can be made between populations and various groups (Jones et al. 1994).

In this chapter, we describe a conceptual framework for measuring HRQoL for the purpose of population monitoring in relation to asthma. This encompasses what is being measured, why it is being measured and how it is measured. Included is a review of how asthma impacts on HRQoL, what types of measures are available to assess HRQoL, and what characteristics indicate a good measure (attributes, breadth and depth). The implementation of HRQoL measures in Australian health surveys to date is reviewed in light of the conceptual issues raised. At the end of this chapter, the conceptual framework is used to provide principles that can be used to guide the selection of HRQoL measures for different purposes in population monitoring. The strengths and weaknesses of specific HRQoL measurement instruments are reviewed in Chapter 3.

2.1 How does asthma affect HRQoL?

Most people who identify asthma as their main disabling condition report some restriction in their core activities and also report poorer health status than people without asthma. Table 2.1 summarises the impacts of asthma on the domains of HRQoL. In the 1995 National Health Survey, 12% of people with asthma reported taking days off from work or school in the preceding two weeks due to asthma (ABS 1995). There is also evidence that asthma is associated with a predisposition to anxiety and depression in adults, although the subject remains controversial (Harrison 1989; Osman 2002; Rand & Butz 2000). People with asthma experience sleep disturbances and often feel tired and frustrated because of their asthma (Sawyer & Fardy 2003). In the United States, people with asthma report more physically unhealthy days (6.5 days vs 2.9 days), mentally unhealthy days (5.2 days vs 3.0 days) and days with activity limitation (3.7 days vs 1.6 days) in the previous month than respondents who did not have asthma (Ford et al. 2003).

Children with asthma may also identify specific issues that impact on their HRQoL, such as feeling angry, frustrated and socially isolated (Juniper 2001). In the Living With Asthma study, one in five children with asthma did not ride a bike, play at school or play with animals and one in three did not participate in organised sports (Sawyer & Fardy 2003). Parents of children with asthma were more anxious than parents of children who did not have asthma. In another Australian study conducted among school children (Sawyer et al. 2001), the physical health, mental health and role and social functioning dimensions of HRQoL were significantly worse among children with asthma than among those without asthma. Children with more severe asthma had the poorest HRQoL outcomes.

Table 2.1: Impact of asthma on HRQoL for the individual and family

Core domains of HRQoL	Impact on individual	Impact on family
GLOBAL		
Overall assessment of wellbeing	Influenced by disease severity and level of disability as well as underlying emotional and social factors that can impact on the outcomes of the disease as well as on the ability to manage and control symptoms and risk factors	Members of the family may take on a carer role and provide support and assistance in daily/core activities. In adults, there may be the presence of comorbidities impacting on overall health, or asthma may have been present over a longer duration with adaptation of the family to limitations on lifestyle.
PHYSICAL		
Symptoms	Coughing, wheezing, loss of sleep	Sleep disrupted
Physical functioning	Walking up stairs, playing sport, exercise and other physical activity Sleep disrupted	Dependence on family members for assistance with activities such as shopping and housework
Disability	Restriction in ability to perform normal actions Limited in ability to complete activities of daily living	
PSYCHOLOGICAL		
Mental and emotional health	Fear of lack of control and anxiety about an asthma attack Embarrassment in taking medication Stress in remembering to take medication Increased risk of depression (especially if other chronic diseases are present) Children and adolescents often have lower self-esteem and see themselves as different.	Anger, frustration, depression by burden asthma places on family Parents anxious, worried about child's asthma, fear of an attack, lack of control, risk of their child's death Stress on family members due to difficulties in negotiating medication compliance and communication between family, carers and clinicians
Behaviour	Asthma can be a contributing factor in causing behavioural problems in children.	
SOCIAL		
Daily role	Restricted in usual activities Restricted in study activities Increased sick days and missed school days Restriction in participation in community social activities	Family life disrupted (e.g. night disturbances, visits to health services) Family restricted in social activities, holidays and keeping pets
Work	Restricted in work activities Increased sick days Long-term limitations in employment, and possibly lower educational attainment	Can contribute to restriction in employment for family members either in choice of occupation or in hours able to work Carer burden for parents if child sick, with lower productivity
Personal relationships	Impaired contact with friends, relatives and reduced participation in social events and increased isolation In children and adolescents, asthma can inhibit relationships with peers and modify social circles.	Contact with relatives and friends can be restricted.

Other studies have also found that children and adolescents with asthma have more behavioural problems (Bussing et al. 1995), lower self-perceived health status (Forrest et al. 1997), and lower self-esteem, self-pity and sometimes embarrassment in taking medication (Donnelly 1994). In a United Kingdom study of 773 children aged between 5–17 years who had current asthma, children reported that asthma restricted their participation in everyday activities and caused frequent school absences and night disturbances (Lenney et al. 1994). Substantial proportions stated that there were times when they could not complete a sports lesson (up to 50%), when school work productivity was reduced due to being sleepy in lessons and having attention deficit problems (>50%) or when they were sometimes not able to go to school following a disturbed night (41%).

Asthma also has impacts on HRQoL for the family. Having a child with asthma has an impact on the parent or caregiver's time, other siblings and family-related activities (Halfon & Newacheck 2000). Families may be confronted with decisions about holidays, keeping pets, installation of special furnishings, and extra cleaning to control the environment (Warner & Warner 1991). There may be an added burden from the costs of medications and health care (Toelle et al. 1995). A parent or caregiver of a child with asthma may have to take time off from work or from daily activities to care for their child (Halfon & Newacheck 2000). The extra demand on time and responsibility adds to the family's emotional and financial burden and can increase stress and put pressure on relationships (Rand & Butz 2000). These findings highlight the impact of asthma on the emotional and social dimensions, as well as on the physical dimension, of HRQoL.

2.2 Purposes of measuring HRQoL

HRQoL can be used to describe health outcomes, guide clinical management, predict health outcomes, formulate clinical policy and direct the allocation of resources. The main functions for which HRQoL measures are used may be classified as discrimination, evaluation and prediction (Kirshner & Guyatt 1985).

2.2.1 Discrimination

One of the purposes of population monitoring in asthma is to discern subgroups of the population who have greater or lesser impacts attributable to asthma (Feeny et al. 1999). This requires an instrument that can discriminate between groups with a higher burden of disease. High burden subgroups identified in this way may then be targeted for specific interventions or further investigation into the causes (e.g. environmental, economic or cultural) of the observed disparities.

2.2.2 Evaluation

Perhaps the most common context for health research is evaluating the effect of an intervention. In clinical trials the intervention may be a drug or some other form of treatment, which is usually evaluated in a randomised controlled trial. In the population setting, it is common to evaluate the impact of new programs or management guidelines, either using a cluster randomised design or, more simply, by tracking change in outcomes over time. Evaluative measures of HRQoL are required for this purpose. Many HRQoL measurement instruments have been designed for these settings, particularly asthma-specific HRQoL measures. The key attributes of these measurement instruments is that they are valid measures of change in HRQoL and that they are responsive to within-subject change in the HRQoL attributes (Kirshner & Guyatt 1985).

2.2.3 Prediction

Predictive instruments are used in HRQoL measurement either to predict the result in another measure or to forecast an outcome at a future time (Feeny et al. 1999). These can be useful for assisting in decision making processes, classifying individuals entering a study or identifying those who are likely to develop a particular outcome (Kirshner & Guyatt 1985). Predictive HRQoL measures might be used to predict future health needs and economic impacts. For example, Eisner et al. (2002) conducted a prospective cohort study aiming to determine the effectiveness of HRQoL measures for identifying those at risk of adverse health outcomes. This study measured HRQoL using the Short-Form 12 questions (SF-12) and the Integrated Therapeutics Group Asthma Short Form (ITG-ASF) battery measurement instruments to test HRQoL as a predictor of future health care utilisation based upon the subjects' current asthma status and known risk factors for health care utilisation. It found that people with better baseline asthma-specific HRQoL scores had a significantly lower risk of all cause hospitalisation.

2.3 Types of HRQoL measures

2.3.1 Generic and specific HRQoL measures

The focus of the content within an HRQoL instrument may be on impacts that are relevant to a specific disease or, alternatively, on impacts that are relevant to a broad range of health conditions. Both generic and disease-specific instruments have a role in the assessment of HRQoL. Generic questionnaires aim to assess the impact of any and all adverse health states on HRQoL, without reference to the impacts of any specific disease. Disease-specific HRQoL instruments measure the specific impacts of the target disease.

Generic HRQoL measurement instruments can be used to assess overall HRQoL in all individuals in the study population. The strength of these instruments is that all members of the population, including those with no illness and those with a range of different illnesses, are measured on the same scale. It therefore allows comparison of HRQoL outcomes between population groups with different diseases.

Reference values, based on the scores in healthy individuals, have been derived for some generic HRQoL questionnaires (Mishra & Schofield 1998). This facilitates the assessment of the HRQoL of subgroups, such as those with asthma, relative to other members of the population or relative to reference values (Ware & Gandek 1998). The limitation of these questionnaires is that they may not adequately focus on those aspects of HRQoL that are particularly relevant to the people with a particular disease and, hence, may lack sensitivity in relation to the impacts of a specific disease.

Specific measurement instruments are designed for specific diagnostic or population groups, such as people diagnosed with asthma. The rationale for these questionnaires is that they will be more relevant and more sensitive to differences between population subgroups and responsive to changes over time (Patrick & Deyo 1989). Disease-specific profiles or health indexes are widely recognised as useful tools for assessing the impact of asthma, and particularly for evaluating the impact of interventions to ameliorate the condition.

In population-based monitoring the important limitation of disease-specific instruments is that they are only applicable to people with that condition in the population and, unlike generic instruments, cannot be used to compare HRQoL with the general population or with other diseases or population groups. However, in order to achieve a time series that can be used to monitor changes in disease outcomes over time and allow comparison between subgroups or populations with a particular condition, there is value in using disease-specific measures. These are more sensitive to the specific HRQoL issues of concern in the subpopulation with the disease of interest.

Another possible limitation of some disease-specific measures is that they may not be accurate in attributing impacts to the specific disease in question. This is not an issue when the impact is unique to a specific disease (e.g. wheeze, or embarrassment about inhaler use, for people with asthma) but may be a problem when the adverse outcome could have many possible causes (such as tiredness or time away from work or school). Respondents may inadvertently underestimate or overestimate the importance of a specific cause for these non-specific adverse outcomes.

2.3.2 Utility scales

Utility-based measures of HRQoL differ from all other types of HRQoL measures in one fundamental way; they value health as well as describing it. The HRQoL instruments described in other sections of this chapter are designed to quantify a respondent's perception of his or her own current health state, in terms of a set of standardised questions and responses. These instruments are often explicitly multi-dimensional, with a separate summary score for each dimension, and although various dimensions of health are described, their relative value is not captured. Health states in utility instruments are also described in terms of a number of dimensions, but the value of each health state is summarised as a single index. This utility index incorporates the relative value of the component dimensions and levels of health, and reflects respondents' preferences for different health states. However, the value that is linked with a particular health state is not necessarily the value of a particular individual, nor do respondents necessarily value their own health state.

The theories and methodologies underlying utility-based measures are rooted in economic theories of decision making, and the measurement methods are conceptually and operationally complex. Consistent with the conceptual framework used in this report, utility-based measures are summarised here in terms of what is being measured, why it is being measured and how it is measured.

Utility measures include a defined set of health states, covering a wide range from worst to best possible health. The values associated with a particular health state are called health state preference scores or utility weights. Under a set of strong assumptions, utility is a cardinal scale, with an absolute zero (death). Full health is given a value of one, and states worse than death are possible. However, interval scale properties have not been proven empirically (Cook et al. 2001).

Measurement in the utility-based approach has two parts: one describes the relevant health states and the other ascribes utility values to those health states. Multi-attribute utility indices (MAUI) describe health states systematically in terms of a series of domains (or 'attributes') and levels, similar to a HRQoL profile. The number of health states defined by a MAUI is a function of the number of items and response options. For example, the generic utility instrument EQ-5D (formerly known as EuroQoL), describes health states in terms of five domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), each of which has three levels (e.g. no pain, moderate pain, extreme pain) (Rabin & de Charro 2001). Thus, the EQ-5D describes a total of 243 health states, representing all possible 3^5 combinations of those domains and levels. MAUIs can be used like HRQoL profiles to allow individual patients to describe their own current health state in terms of the domains and levels in the MAUI. The health states described by MAUIs may not be suitable for a particular research study. In this case, health states may be described in a series of vignettes specific to the particular research context.

The second component in the utility-based approach ascribes utility weights to health states. Three methods commonly used for valuing health states are the standard gamble (SG), time trade off (TTO), and the visual analogue (VAS) (see Glossary). SG and TTO are cognitively complex and must be administered by a trained interviewer. Determining utility weights is, therefore, labour-intensive and expensive, which may explain why Australian weights are available for only one MAUI, the Assessment Quality of Life Instrument (AQoL). Some MAUIs define an enormous number of health states, and it is not always feasible to value all of them. Instead, their value is interpolated from the values of a subsample of health states, using an algorithm that combines the utility associated with each dimension into an overall utility index, either algebraically or by statistically modelling. Thus, the utility weight associated with a particular health state in a MAUI represents a very complex synthesis of a sample of respondents' valuations.

A key question in the valuation exercise is: 'Whose preferences and values matter?' Decisions about the allocation of health budgets require a societal perspective and may warrant values from a general population, while decisions about best treatment may be better informed by people who have experienced the health condition, whether personally or vicariously via a friend or relative. People who have experienced a poor health state tend to value it more than do people without such experience. Arguably, only people who have experienced a health state can value it truly, but on the other hand they may over-value it. This conundrum cannot be resolved, and is perhaps a conceptual limitation of the utility approach. A pragmatic solution may be to recognise that values from different perspectives may differ, and to choose the appropriate perspective and sample from which to determine utility weights for a particular decision context.

Most of the widely used MAUIs have published general population-based utility weights. However, the validity of the MAUI within a specific population depends, in part, on the extent to which the weights are applicable to that population. Most sets of weights have been derived in British or North American populations. The AQoL is the only MAUI with utility weights from an Australian sample.

2.4 Attributes of HRQoL measures

Attributes of HRQoL measurement instruments that are important for population health monitoring include validity, reliability, responsiveness, sensitivity and interpretability. In addition, practical issues such as cost and the suitability for use in special populations need to be considered when evaluating available HRQoL measures. Table 2.2 summarises the attributes of HRQoL measures as they relate to the purposes of measuring HRQoL.

2.4.1 Validity

Since HRQoL cannot be directly observed, it cannot be directly quantified. Validation is a process of establishing the extent to which an instrument measures what it is intended to measure (in this case, HRQoL) (Fayers & Machin 2000; Streiner & Norman 2001). The ability of HRQoL instruments to measure HRQoL accurately can be addressed through assessment of content validity, criterion validity and construct validity.

Content validity refers to whether an instrument adequately covers the topic being measured (Streiner & Norman 2001). The method used to derive the content of the questionnaire is relevant to its content validity. For instance, the use of psychometric techniques to sample content adequately from the HRQoL domains of interest contributes evidence of content validity (Kaplan et al. 1976). Face validity is related to content validity and assesses the extent to which the items within the instrument appear, to the person interpreting the data, to both encompass, and be limited to, the range of topics relevant to impacts on HRQoL.

Criterion validity refers to the degree of agreement of the measure with a gold standard (or 'criterion'). This is not possible in relation to HRQoL measurement instruments, as there is no gold standard. In quality of life research, comparisons of test instruments with longer in-depth interviews exploring the domain the instrument purports to measure are sometimes used as assessments of criterion validity (Fayers & Machin 2000).

Construct validity refers to whether the measurement instrument produces findings that are consistent with expectations based on the hypothetical model (or construct) that underpins the instrument (Kaplan et al. 1976; Kirshner & Guyatt 1985). Determining construct validity is an ongoing process whereby the larger the body of supporting evidence confirming expectations for a construct, the stronger the construct validity. In HRQoL measurement for asthma, correlations between HRQoL measurement instruments and markers of severity have been used to support the construct validity of some measurement instruments (Marks et al. 1992, 1993).

2.4.2 Reliability

The assessment of reliability examines the extent to which a measurement instrument has reproducible and consistent results, and encompasses internal consistency and repeatability (Fayers & Machin 2000). Internal consistency refers to the degree to which items within a measurement instrument are interrelated and measure the same thing. The correlation between items within the instrument can be statistically assessed, with the most widely used statistic for assessing internal consistency being Cronbach's α (Cronbach 1951). Internal consistency is an important attribute of all scales that are scored, as it is a prerequisite for valid interpretation of the overall score.

Repeatability refers to the level of agreement between repeated administrations under the same conditions (test–retest reliability), usually over a short time interval. It is quantified for each item and for the overall questionnaire using the kappa statistic, for binary and categorical outcomes, and the intraclass correlation coefficient, for continuous measures (Fleiss & Cohen 1973). Repeatability is a major consideration in the population monitoring context as surveys are almost always periodically repeated.

2.4.3 Responsiveness and sensitivity

Responsiveness is the ability of an instrument to detect change within individuals over time, and sensitivity is the ability of the instrument to detect differences between groups (Fayers & Machin 2000). Instruments in which a large proportion of respondents select the highest or the lowest response categories ('ceiling' and 'floor' effects) and those in which there is a large gap between the available levels, so that most respondents are clustered on either side of this gap, lack responsiveness and sensitivity. The importance of responsiveness and sensitivity depends on the purpose of the HRQoL measurement. Responsiveness is particularly important in evaluative instruments, which are commonly used in the clinical setting but not in population health surveys. Sensitivity is important in discriminative instruments.

In a population health survey, sensitivity is a key issue for detecting differences between groups in the population such as people with and without asthma. Sensitivity is also an important attribute of questionnaires used in repeated cross-sectional surveys to measure change, over time, in a population because the individuals responding each time will differ. Therefore, sensitivity is generally more important than responsiveness in population health.

Table 2.2: Summary of attributes needed for the purposes of HRQoL measurements

Purpose of measurement	Validity	Reliability	Responsiveness / sensitivity	Example in people with asthma
Discriminative	Cross-sectional construct validity – relationship between the measure and external measures at a point in time	Internal consistency and test–retest repeatability	Ability to detect differences between subjects (sensitivity)	Health surveys to compare HRQoL in people with and without asthma or with severe and mild asthma
Evaluative	Longitudinal construct validity – relationship between changes in measure and external measures over time	Internal consistency is relevant to interpretation. Should be repeatable in subjects known to be stable but responsive in those who have changed.	Ability to detect within-subject changes over time (responsiveness)	Evaluation of an asthma self-management intervention Assessment of an asthma control program for school children Clinical trial for new asthma medication or treatment regimen
Predictive	Predictive validity – predictions based on the measures are proven correct	As for discriminative instruments	Not applicable	Classification of subjects into categories according to a criterion/gold standard measure Prediction of demand for health care services for asthma

Sources: Feeny et al. 1999; Guyatt et al. 1992; Kirshner & Guyatt 1985.

2.4.4 Interpretability

Interpretability has been defined as 'the degree to which one can assign qualitative meaning—that is, clinical or commonly understood connotations—to a quantitative score' (Lohr & Aaronson 1996). It is an essential attribute of any HRQoL instrument. Much as for validity, determining interpretability is an evolving process through accumulation of a body of evidence with repeated experience in a variety of contexts (Ware & Keller 1996).

The interpretation of HRQoL scores poses a number of difficulties. HRQoL means different things to different people at different times and in different contexts. A person's perception of his/her health state may change over time. Furthermore, the numeric values of HRQoL measurement scales are arbitrary and there are many different HRQoL instruments with their own scales, meaning it is difficult to standardise across measures (Gonin et al. 1996).

It is important to point out that statistical significance testing does not necessarily assist in interpreting the findings. A statistically significant result (for example, $p < 0.05$) indicates that the observed difference is unlikely to have occurred by chance. However, it does not convey any information about the size or meaning of the observed difference.

One approach to the interpretation of population data on HRQoL is to compare the observed levels to population normative values (see Figure 2.4), or alternatively, to the values seen in other diseases or other population groups. This gives a reference point or points, which the reader can use in interpreting the data for the disease and population under study (Osoba & King 2004).

2.4.5 Feasibility and practical issues

Population surveys are commonly administered by telephone, face-to-face interview or self-completion. Inclusion of HRQoL instruments within a survey necessitates that the instrument be compatible with the survey design. For example, the use of telephone interviews precludes the administration of visual analogue scales. Furthermore, the mode of administration may influence the outcome of the HRQoL measurements. Participants may respond differently in the anonymous setting of a self-completed questionnaire compared with a face-to-face interview.

A critical issue relating to survey design is respondent burden, that is the demand placed on respondents to participate in the survey. The number and complexity of survey questions largely determine the time required to complete the survey and, hence, the respondent burden. In telephone or interviewer-administered surveys, the time required to complete the survey also affects the cost of conducting the survey. In large health surveys, it is likely that HRQoL measures will be competing for survey space with a range of other measures, such as questions about service utilisation and disease management. For this reason there are limitations on the amount of time available for HRQoL questions in population health surveys. These limitations and costs need to be considered when selecting HRQoL measures for this purpose.

The time period over which participants are asked to recall events is also a major consideration in population surveys, particularly when comparing results between surveys. In relation to asthma, it is important that the time period be long enough to encompass some of the short-term variability that is inherent in the disease. However, as for all disease states, it is important that it not be so long that recall error is likely to occur.

2.4.6 Applicability to special populations

In addition to the general performance criteria described above, population monitoring measures used in Australia must be suitable for use in a culturally and linguistically diverse society. Methods for iterative forwards and backwards translation of questionnaires to obtain valid data in languages other than the original language have been described (Chwalow et al. 1992) and many of the widely used questionnaires have been translated into other European languages. However, translations into languages common within the Australian community are less widely available. Furthermore, simple linguistic translation may not be adequate. It seems likely that cultural differences in attitudes, values and beliefs would influence the content of domains of HRQoL that are appropriate to measure. Under some circumstances it may be advantageous to develop questionnaires that are specifically appropriate to cultural groups.

Adult Indigenous Australians report diagnoses of asthma more commonly and have higher rates of hospitalisation for asthma than non-Indigenous adults (ACAM 2003). It is likely that assessing the quality of life impact of asthma and other diseases among Indigenous Australians poses some specific challenges in developing measures that are linguistically and culturally sensitive and appropriate. In a study of urban Indigenous Australians, family and spiritual beliefs were important determinants of perceptions of health (King et al. 1999). Other issues are similar to those seen in non-Indigenous communities in Australia (Freidoon Khavarpour confirmed this by email on 11 November 2003). Therefore, the inclusion of the spiritual domain in a measurement instrument may be a consideration when measuring HRQoL in this population.

A similar issue arises in relation to differing age groups: the content of quality of life domains differs through the phases of life. This has been recognised, to a limited extent, with the development of child-specific HRQoL questionnaires and some adolescent questionnaires. However, in general, issues of the elderly have not been specifically addressed in asthma-related quality of life questionnaires.

2.5 Breadth and depth of HRQoL measures

Within the broad types of HRQoL measurement instruments exist instruments of differing levels of breadth (coverage) and depth (precision), ranging from single item (single question) and very brief questionnaires to comprehensive, multi-item, multi-dimensional HRQoL profiles. These are described in the following sections and summarised in Figure 2.1

Coverage of an instrument can be evaluated in terms of its content validity (Section 2.4.1), while precision (or reliability) is related to responsiveness and sensitivity as well as internal consistency (Sections 2.4.2 and 2.4.3). In population monitoring, sufficient precision is needed to discriminate subgroups.

Dimensions covered (breadth) →

Number of items (depth) ↓

	Single dimension		Multi-dimensional	
	Single item, single dimension measures		**Global measures**	
Single item	**Description** Single question to measure one dimension of HRQoL	**Examples** Days off school or work due to illness (generic) Sick days due to asthma (disease-specific) Symptom-free days (disease-specific)	**Description** Single question to give overall measure of the global domain of HRQoL	**Examples** In general, how would you rate your health? ('SF-1', generic) 'Delighted to Terrible' scale (generic)
	Multi-element, single dimension measures		**HRQoL profiles**	
Multi-item	**Description** Multiple questions to measure one dimension of HRQoL	**Examples** Kessler-10 Scale of Psychological Distress (generic)	**Description** Multiple questions to measure several dimensions of HRQoL	**Examples** SF-36 Questionnaire (generic profile) EQ-5D (generic utility scale) Asthma Quality of Life Questionnaire (disease-specific profile)

Figure 2.1: Classification of HRQoL instruments by breadth and depth

2.5.1 Single item and brief measures

The broadest and simplest class of HRQoL measures are those that endeavour to summarise the domains and dimensions of HRQoL simultaneously in a single question (sometimes referred to as global domain measures). A widely used example is the question 'In general, would you say your health is excellent, very good, good, fair or poor?' sometimes referred to as the 'SF-1'.

Brief global measures have the advantage of being simple to use with low respondent burden (the effort and time required for a respondent to answer) and this can be particularly attractive in large-scale population surveys where there are many questions competing for space in the survey and each question adds substantially to the costs. Furthermore, global measures of self-perceived health status have been shown to be predictive of mortality (Heidrich et al. 2002; Idler & Benyamini 1997; Miilunpalo et al. 1997). This supports the construct validity of these measures.

The main disadvantage of single item or very brief instruments is that the content, although it may be broad ranging or global in intent, does not adequately sample from a comprehensive range of HRQoL dimensions and may not adequately reflect all the relevant domains for all individuals. Using one question is vulnerable to influence by the respondents' individual interpretations of the question, and is also unable to provide detail about the dimensions of HRQoL that may have influenced the response. These measures do not provide information about the relative impact on the individual physical, psychological and social domains of health (Sloan et al. 2002), and this limits their usefulness in terms of planning an appropriate response. These limitations relate to content validity (Section 2.4.1). A further disadvantage is that since they usually have only a small number of possible response options, the measurement range is coarse in relation to the underlying latent continuum of real health states in the population. The limited response options in single item measures reduces the instrument's precision and, hence, its sensitivity or ability to discriminate differences in HRQoL between population groups. Hence, due to problems with content validity, sensitivity and reliability, studies using these single item or very brief global instruments as the sole tool for assessing HRQoL should be interpreted with some caution (Bradley 2001; Jones et al. 1994).

Some single item measurement instruments only focus on a single HRQoL domain rather than HRQoL globally. Sick days due to asthma—that is, the number of days away from work or school or the number of reduced activity days due to asthma—and symptom-free days—that is, the number of days in which the subject does not experience asthma symptoms—are both examples of this form of disease-specific, single domain, single item measures for the impact of asthma (CDC 2000). These single item, single dimension measures may be more valid and sensitive for their intended purpose than the single item global measures, as long as their interpretation does not extend beyond the single domain or dimension that has been measured. As asthma is an episodic disease, it can be difficult to capture adequately the time-variable impacts in a single measure. Some of the single item, single dimension measures referred to above, such as sick days, unhealthy days or healthy days, represent a useful way to address this issue of time variability. However, they should *not* be interpreted as global measures of HRQoL impacts.

2.5.2 Multi-item and multi-dimensional HRQoL profiles

In contrast to single item or very brief HRQoL measures, HRQoL profiles that contain multiple items to measure multiple dimensions are able to assess the physical, psychological and social domains of HRQoL more comprehensively (Testa & Simonson 1996). By measuring several dimensions (issues, or areas of interest) within each domain, such questionnaires may more relevant to the disease or intervention that is being investigated (Table 2.3). By including multiple items relevant to a domain, these questionnaires achieve greater precision in measuring that domain. In other words, multi-item, multi-dimension instruments generally measure HRQoL with greater content validity and precision than the single item or very brief questionnaires referred to above.

There are some circumstances when the purpose of monitoring may relate particularly to one domain of HRQoL. For example, in evaluating the impact of an intervention designed to reduce school absences due to asthma, it would be most appropriate to choose a measure with maximal validity, reliability and sensitivity in this dimension. Indeed, this may not be an asthma-specific questionnaire but rather a measure of overall absence from school. Similarly, an intervention addressing the psychological consequences of asthma might best be evaluated by using a psychological questionnaire. In other circumstances, the physical domain may be the focus of attention and one of the questionnaires which focuses on physical function would be most appropriate. The important issue is that investigators should be aware of the domains that are encompassed by the measures they use and, where possible, should select measures that target the domains that are relevant to their monitoring purpose.

Table 2.3: Summary of key HRQoL elements for assessing the impact of asthma

Core domains	Dimensions	Elements of HRQoL in people with asthma		
Physical	Symptoms, impairment in physical functioning, disability	• Tiredness • Restricted physical activity	• Impairment of physical functioning • Exercise limitations	• Symptom free days • Days limited in core activities
Psychological	Positive and negative affect, behaviour	• Distress • Anxiety • Depression • Fear	• Frustration • Coping with an attack • Dependence on sprays/medication	• Expression of being bothered by asthma • Embarrassment at taking medication
Social	Role performance, personal relationships	• Restriction in work and usual activities	• Sick days • Missed school days	• Contact with friends, relatives • Participation in social events

There are several approaches to scoring or summarising the information contained within multi-item (or multi-element) instruments. The psychometric approach is to extract meaning about dimensions and domains from a number of items or elements using a variety of statistical tools. A number of specific strategies are employed to select relevant items, group them in a meaningful way and combine information from responses to individual items to generate summary information (Juniper et al. 1997). This may yield an overall summary score or a profile of scores for specific dimensions, or both. These scores can be used to summarise the impact of having asthma on the core domains of HRQoL and make comparisons between different population groups. Psychometric measures provide quantitative information but can be used only to compare with data collected using the same scale.

There is no absolute reference or anchor point for psychometric scales and, hence, the meaning of any given scale score is unique to that scale. An alternative scoring approach is to quantify information about health status on a scale between perfect health and death. This approach is based on utility theory and is discussed in Section 2.3.2.

The main disadvantage of HRQoL profiles is that they are longer and, therefore, more expensive to implement. They also involve a greater respondent burden. Generally, longer measurement instruments are more precise. However, for population monitoring purposes, in which surveys are administered to large populations, the precision of multi-item profiles may be greater than that needed to distinguish population subgroups adequately or to detect clinically relevant change over time. Under these circumstances, shorter instruments may be adequate, as long as they have sufficient content validity; that is, they sample from all HRQoL domains. Consideration should be given to the balance between level of precision required and efficiency when selecting instruments for population monitoring.

2.5.3 Dynamic health assessment

Most of the multi-item instruments developed to date have been developed with classical psychometric theory. In this approach, a large pool of relevant items is developed, then various procedures and criteria are used to select a subset of the best items for inclusion in the instrument. The same items are then administered to every person every time the instrument is used. In this sense, these instruments are fixed or 'static'. As noted above, practical considerations dictate that relatively few items are used in many health applications.

Brief, static instruments have three important limitations. First, if the items represent a broad range of health, they are spread sparsely along the underlying latent continuum of real health states, producing a coarse, imprecise scale prone to measurement error. Poor precision in the measurement of each individual's health is not relevant when the purpose is to estimate the mean health status of a population; precise estimates of the mean are achieved by surveying very large samples. However, population surveys may also be used to investigate relationships among various factors, such as determinations of health. In this case, greater precision in the health measurement scale increases the power of subgroup analyses and regression.

Second, if the items are targeted at a limited range of health, representing only a portion of the underlying continuum, the resulting scale will suffer from ceiling or floor effects when used in subgroups whose true health lies outside the measured range. As noted above, ceiling and floor effects compromise the sensitivity of a scale to differences among patients and its responsiveness to change.

The third consideration is the integration of evidence across levels of health care, from population health monitoring through clinical research to individual patient management. These levels require different precision: instruments used to screen and monitor individual patients must be very precise to minimise classification errors and to detect individual changes reliably, while imprecise instruments are suitable for population health monitoring when errors at the individual level do not matter. The precision required for clinical trials and health services research falls somewhere between these two extremes. Instruments developed for one level are often not appropriate for another; they are either too long or too imprecise or they target the wrong part of the health range. For example, the SF-36 (with 36 items and eight domains) is suitable for clinical research, but it is not precise enough for use in individual patient management (McHorney & Tarlov 1995). Different instruments are often used at different levels, making it difficult to translate knowledge derived at one level to another level, and to link populations and policy to patients and practice.

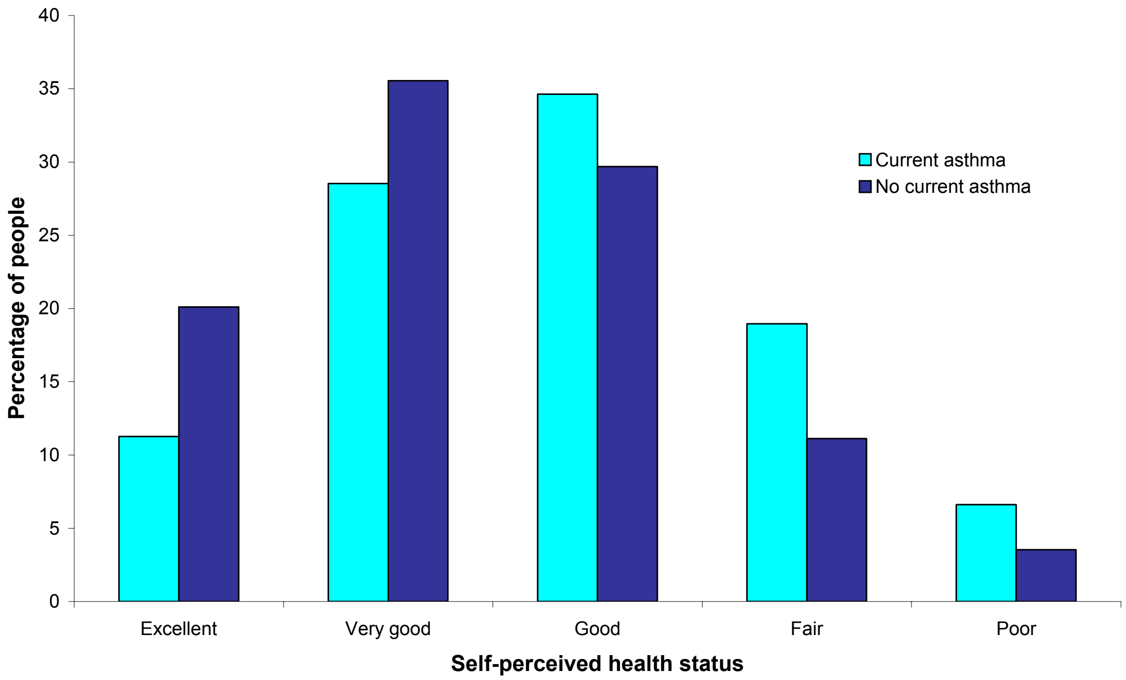
Ideally, we would measure health on a common metric with a range of instruments that could be cross-calibrated and whose precision and content could be suited to the context and needs of the application. New research suggests this ideal may be achievable (Hays et al. 2000). There is growing appreciation of another psychometric approach, 'modern psychometrics', because of its potential to address the limitations of static instruments. This approach has the same starting point as does the classical psychometrics approach: it begins with a large pool of relevant items. This pool of items is then administered to a heterogeneous sample, representing the full spectrum of possible health states. Item response theory is then used to characterise each item in terms of where it sits along the latent health continuum and how sharply it discriminates among people in different states of health. The corresponding item response statistics calibrate items relative to the latent variable. A response to a single item, or any combination of items, can then be given a score which locates the respondent on a common metric. The more items that are asked, the more precisely the respondent is located on the latent continuum.

In this measurement approach, the only question common to every respondent at every assessment time is the first question. The second question is determined by the answer to the first, the third question is determined by the answer to the second, and so on. Thus, each respondent is asked questions that are relevant to their current state of health; people in good health are not asked questions about poor health and vice versa. This is in contrast to static instruments, where everyone is asked the same questions, including some that may not be at all relevant to some people. The number of questions asked depends on the precision required. Since the number and content of questions varies each time a subject's health is assessed, this approach is called 'dynamic health assessment'. The iterative, logical process that determines which and how many items are used is suited to computer administration. Initially developed for educational applications, this was called computer adaptive testing; now it is being applied to health assessment it is called dynamic health assessment (Bayliss et al. 2000).

This new dynamic approach overcomes a number of the limitations of traditional, static health assessment. First, it matches precision to the assessment context, allowing the same (albeit dynamic) instrument to be used for monitoring patients and populations, resolving the problem of interpretation across the three levels of health care described above. Second, it optimises the number of questions asked with respect to the information needs and purpose of the assessment, resolving past tension between respondent burden and precision. Third, it ensures the content is relevant to the respondent, facilitating compliance with questionnaire completion. Fourth, it allows existing static instruments to be calibrated to a common metric, resolving the problem of interpretation across different instruments.

The implications for population health are that dynamic assessment will allow the most efficient allocation of a quota of questions to the competing topics of interest in a survey, and will maximise interpretability and, hence, usefulness of the ensuing data.

One aspect of dynamic health assessment is currently identified as a potential concern and limitation: the assumption of unidimensionality in the underlying item response theory. This means the pool of items that the dynamic instrument draws from must pertain to a single aspect of health or HRQoL, the notional latent variable or underlying continuum. HRQoL is multi-dimensional; the challenge is to identify a complete set of distinct dimensions and to operationalise them in a way that is meaningful for people in different states of health and with different disease conditions or disabilities. While the potential and limitations of dynamic health assessment are not yet fully realised or understood, it is definitely worthy of further investigation (Cella & Chang 2000; Hambleton 2000).



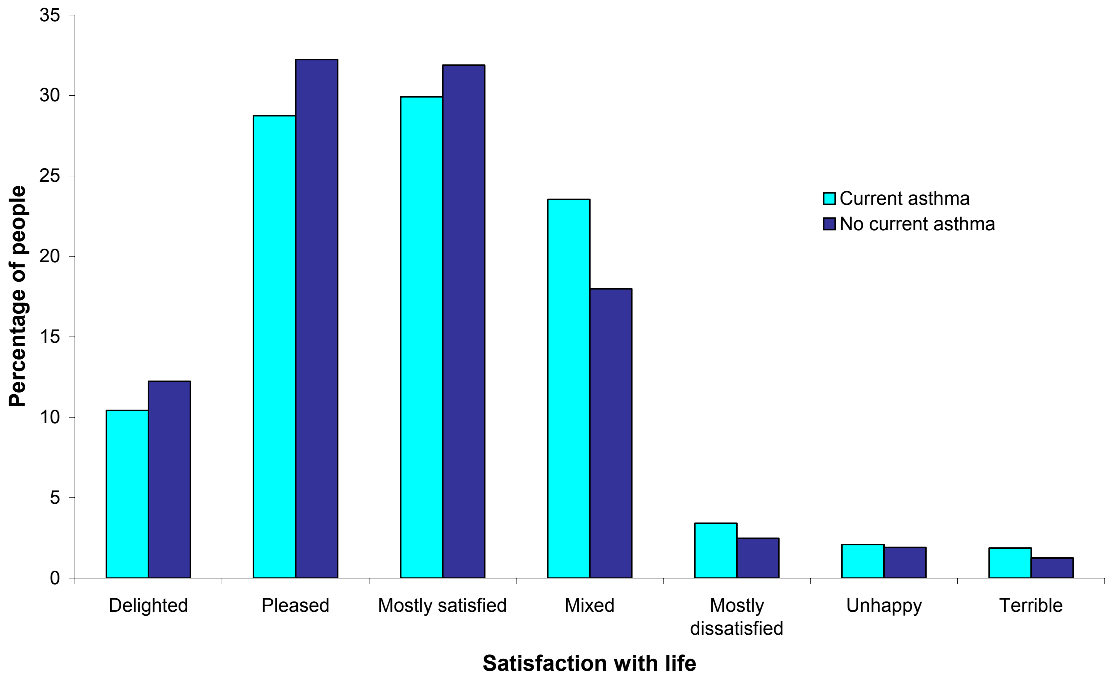
Source: ABS National Health Survey, 2001.

Figure 2.2: Self-reported health status by asthma status, age 18 years and over, Australia 2001

2.6 Examples of population monitoring of HRQoL: two Australian health surveys

Population health monitoring is usually accomplished through repeated cross-sectional surveys on selected health issues in a representative sample of the population or a subset of the population. These surveys afford the opportunity to compare HRQoL and other outcomes for different diseases with the general population norms for a broad range of population health data. The selection of items for inclusion can be based on identified health concerns, such as the National Health Priority Areas (AIHW & DHFS 1997), and behavioural factors, such as physical activity and diet, that are known to influence health. This section presents data collected in two population health surveys in Australia to demonstrate the use of a range of HRQoL measures. The findings are discussed in light of the strengths and weaknesses of the measures used.

In the 2001 National Health Survey (NHS), measures that related to HRQoL were the SF1 self-rated health status measure (five response options), and a question to rate life satisfaction (seven response options). These are examples of single item global measures, which are often used in large population surveys because of the minimal cost and time to implement such measures. Compared with people without current asthma, people with asthma were less likely to select the most positive response options and more likely to select negative response options for both of these questions (Figures 2.2 and 2.3).



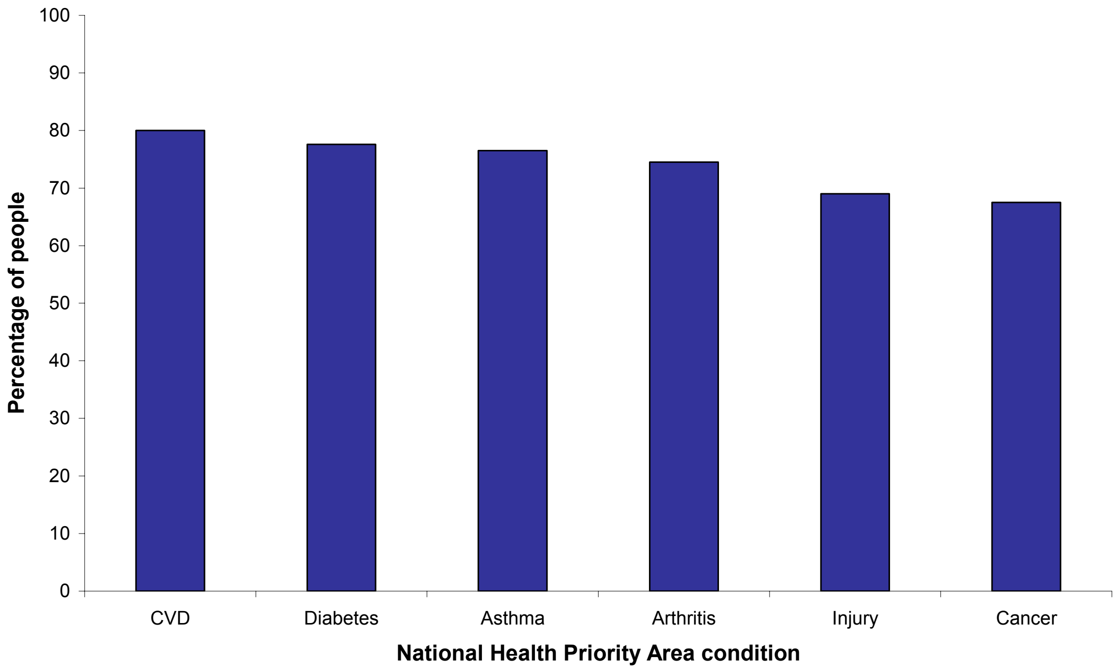
Source: ABS National Health Survey, 2001.

Figure 2.3: Satisfaction with life by asthma status, age 18 years and over, Australia 2001

The 2001 NHS also included single item, single dimension HRQoL questions:

- 'In the last 2 weeks, have you stayed away from your (work/school/place of study) for more than half the day because of any illness or injury you had?'
- 'In the last two weeks, have you had any other days of reduced activity?'

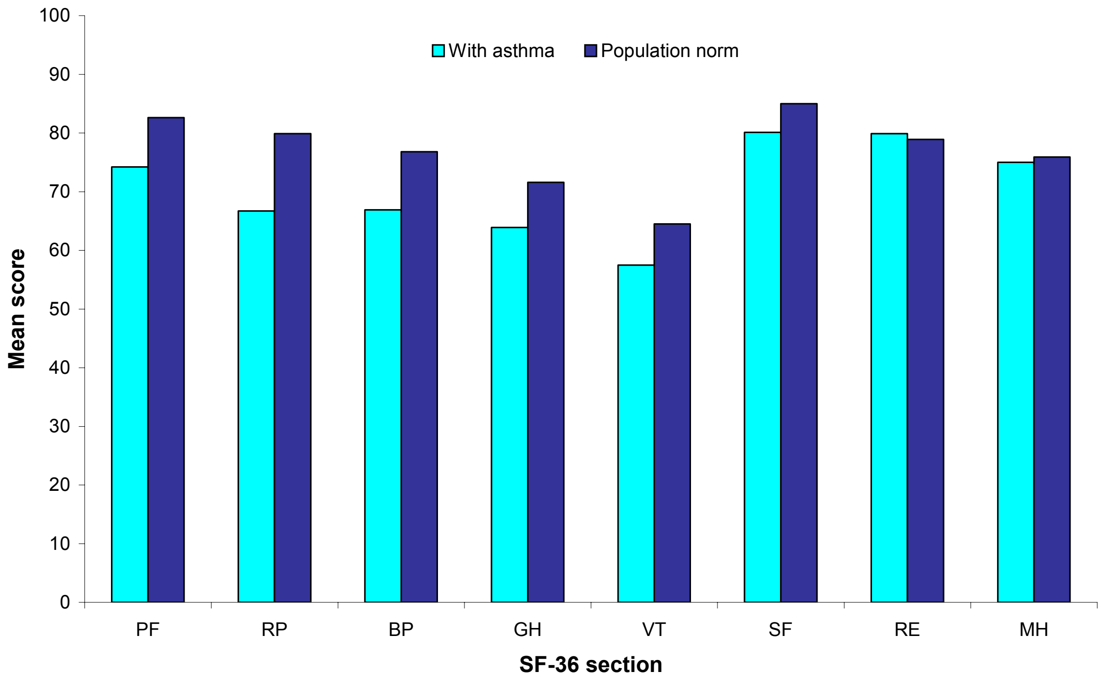
In Figure 2.4, these two questions were combined to create 'Any reduced activity days' and used to make comparisons among diseases that were the subject of National Health Priority Areas at that time. More than two-thirds of people who currently had the selected conditions reported reduced activity days in the last two weeks. The highest prevalence was in those with cardiovascular disease (CVD) and diabetes. People with asthma were more likely to report reduced activity days than those with arthritis, injuries or cancer. Unlike the global measures, this has a narrower focus on elements within HRQoL domains (Table 2.3).



Source: ABS National Health Survey, 2001.

Figure 2.4: Percentage of people with each National Health Priority Area condition reporting any reduced activity days, age 18 years and over, Australia 2001

The South Australian Health Omnibus Survey, conducted in 1998 among 3,010 adults sampled from the general population, incorporated the SF-36 questionnaire (Ware & Sherbourne 1992) to assess HRQoL (Wilson et al. 2002). This is an example of a multi-item, multi-dimensional HRQoL profile in a population survey. This instrument provides a profile of scores on eight HRQoL or health status dimensions (Ware & Sherbourne 1992). The information provided from this measure is able to indicate the relative impacts of asthma on the different HRQoL dimensions. Figure 2.5 shows that having current asthma reduces scores in most dimensions of the SF-36 compared with the population norm. However, the greatest impact was on physical dimensions, with little impact on emotional and mental health.



Note: PF—physical functioning, RP—role: physical, BP—bodily pain, GH—general health, VT—vitality, SF—social functioning, RE—role: emotional, MH—mental health. There was a significant difference between the mean scores of people in metropolitan areas with and without asthma for all domains ($p<0.001$ except for RE, where $p=0.003$).

Source: Wilson et al. 2002.

Figure 2.5: SF-36 scores in people with asthma and the population norm, age 15 years and over, South Australia, 1998

2.7 Selecting HRQoL measures for population monitoring

The conceptual framework developed in this chapter will be used in this section to derive principles that can guide the selection of HRQoL measures. In selecting instruments for measuring HRQoL in populations, it is important to identify those that are suitable for the intended monitoring purpose and context. The three commonly described purposes for measuring HRQoL are discrimination, evaluation and prediction (Section 2.2). These correspond to three key purposes of population monitoring which are discussed here: (1) Comparison of the impact of different diseases, (2) monitoring of changes over time and (3) economic evaluation. In this section, we provide guidelines to assist in the selection of measures for each of these purposes, focusing on monitoring the impact of asthma.

2.7.1 Comparisons of the impact of different diseases or health states

An advantage of population surveys is that they can collect information about many diseases and health states across a representative sample of the general population. Therefore, measuring HRQoL in these surveys can be used to make comparisons between different diseases and health states. This has value for understanding the relative burden that different conditions have in the population and enables policy makers to determine how priorities should be set in the health care system. It also supports the development of interventions that will target those conditions that have the greatest impact in the population.

The measure used for this purpose should be discriminative, so that it is optimised for comparisons between groups in the population with different disease and health states. As it is also necessary to measure HRQoL without reference to specific diseases or specific disease manifestations, a generic HRQoL measure is likely to be most appropriate. The content of the generic questionnaire should not only be interpretable to people with all states of ill-health but also encompass a comprehensive range of impacts, so that the specific effects of various diseases can be measured.

2.7.2 Monitoring changes over time

Another important reason for population health monitoring is to monitor changes in health outcomes over time in repeated cross-sectional surveys. This is used to examine the impact of changes in the physical, social and economic environment, and in disease management practices, and health and other policy.

The specific choice of an evaluative instrument (with high responsiveness) or a discriminative instrument (with high reliability and sensitivity) depends on the study design. In a cohort study, where the same subjects are being monitored over time, an evaluative instrument is required. However, in a repeated cross-sectional study design, in which different subjects are surveyed at each time point, a sensitive, discriminative instrument is required.

There is value in using disease-specific measures in order to achieve a time series that can be used to monitor changes in a disease outcome over time and allow comparison between subgroups or populations with a particular condition. It is also important that the scope of content of the selected instrument is well matched to the expected effects of the interventions or exposures it is required to evaluate or monitor. For example, where the purpose is to monitor the impact of an asthma policy intervention, a disease-specific questionnaire that focuses on asthma will be more responsive than a generic questionnaire, in which scores will be heavily influenced by impacts that are not relevant to the asthma policy intervention (Marks et al. 1993; Rutten-van Molken et al. 1995).

2.7.3 Resource allocation

A third purpose of monitoring HRQoL in population surveys is to generate information that can be used to guide decision making processes by forecasting an outcome at a future time, such as future health needs and economic impacts (Feeny et al. 1999), or by identifying those who are likely to develop a particular outcome (Kirshner & Guyatt 1985). For this purpose, the measure should be suitable for predictive functions and should be measured on a scale that can be incorporated into economic analysis.

In economic evaluation, the consequences of health care programs or treatments are compared with their costs (Drummond et al. 1997). Health outcomes are key components of such analyses, where the aim is to determine which programs or treatments are worth funding, given the alternative uses of resources. Utility-based approaches were developed for use in economic evaluations, and are generally used in this way, but are sometimes also used as outcome measures in their own right. Cost–utility analysis (CUA) requires that health outcomes are adjusted by utility weights, yielding units such as quality adjusted life years (QALYs). In CUA, utilities provide a common metric, allowing comparison across diverse health conditions such as asthma, cancer and heart disease. Cost-effectiveness analysis (CEA) requires only that the outcomes are measured in the same units in the programs or treatments being compared. When HRQoL is the health outcome of interest, utilities may be an appropriate unit and are suitable for CEA because they integrate domains of HRQoL into a single index.