Assessing the need for health status measures

Jenny L Donovan, Stephen J Frankel, John D Eyles

The current reorganisation of the NHS obliges health authorities to purchase a pattern of health care provision which accords with the health needs of the local population.1 This seemingly unexceptional expectation has revealed considerable uncertainty as to how needs assessment should be carried out in practice.

The broad characteristics of the population are reflected in a range of demographic, socioeconomic and mortality data. While these sources of data offer broad insights into patterns of need, they cannot support more intimate decisions. In particular, they are fundamentally incapable of generating valid measures of morbidity.2 Aspects of morbidity are reflected in health services utilisation data, but their interpretation suffers from the necessary deficiency that patterns of demand and supply are inseparable from patterns of need. Routine data sources are therefore inadequate for determining those attributes of each population that should inform purchasing decisions. Adequate needs assessment must depend upon sketching some relief from this unacceptable void in the portrayal of the population’s experience of health and illness.

One option which is advocated is the administration of generic measures of health status to samples of the population in the expectation that the resultant profiles and indices may offer a representative view of the perceived need for health care. In this paper we consider the relevance of generic health status measures to those responsible for advising on the purchase of what may constitute an appropriate mix of services.

Structured measures of health status

The distinguishing purpose of structured health measures has been to offer insights into subjective aspects of health that cannot be ascertained from laboratory tests or population statistics. They therefore aim to record and represent, in a systematic way, aspects of health that cannot be determined from ‘hard’ data or morbidity statistics, but must instead rely on the judgements of individuals, whether patients or clinicians. Data are obtained usually from self-completed questionnaires or from the observations of clinicians or therapists. Most researchers and clinicians are very familiar with the methods of natural science, and so have been relatively willing to incorporate structured, quantitative measures of health into clinical and survey work. It is not the purpose here to review in detail the very many structured social survey instruments developed to measure health status (see, for example).3-5 Health status instruments fall into seven basic categories.

Types of health status measures

General health measures

Measures such as the Nottingham Health Profile (NHP),6-7 Sickness Impact Profile,8,9 and the Medical Outcomes Study Instrument/SF3610-11 aim to provide global profiles of health, including well being, function, and social and emotional health. There is also the single item global health measure which asks if the respondent rates her or his health as excellent, good, fair, or poor (see 12-14).

Measures of physical function

These aim to determine levels of disability, impairment, and physical function within general populations (such as the Lambeth Disability Screening Questionnaire13) or for specific groups such as the elderly in residential accommodation (such as the Pulses Profile16). These measures focus on activities of daily living such as dressing and going up steps.

Pain measures

Instruments such as the McGill Pain Questionnaire17 and visual analogue scale18,19 seek to represent the intensity of pain.

Social health measures

Social health measures such as the Social Health Battery20 aim to assess the strength of people’s social support networks.

Psychological measures

Measures such as the General Health Questionnaire21 identify people with psychological or psychiatric morbidity.

Quality of life measures

Examples such as the Four Single Items of Well-Being22; and the Quality of Life Index,23 seek to measure the overall satisfaction and well being of individuals.

Specific disease measures

The Arthritis Impact Measurement Scale24 and the Oswestry low back pain questionnaire25 are examples. These concentrate on issues of particular importance to patients with specific diagnoses.

Typically, these instruments consist of established sets of closed questions which are asked in the same order and ways of every individual. Questions tend to have limited numbers of responses and respondents (or clinicians) have to tick those closest to their own view. Subjective reports of health are not inherently quantitative, and so some form of rating method is required to
Assessing the need for health status measures

translate statements such as 'unbearable pain' into a form suitable for statistical analysis. Rating methods and the weighting of scales have allowed health measures to take on the style of quantitative research. The extent to which the adoption of this idiom has distracted from critical scrutiny of a number of fundamental characteristics of these measures is discussed below.

Validity and reliability of health status measures

VALIDITY

The validity of an instrument refers to the degree to which it measures what it purports to measure. Structured health status instruments claim to measure aspects of health, but health is an abstract and variable concept which is notoriously difficult to define.27 Ethnographic studies have shown that health is also difficult to define for lay individuals.27 These problems of definition have been acknowledged by only a small number of authors (see 28 29). In order to simplify the measurement of health, structured instruments tend to avoid these issues by selecting indicators that are assumed to be related to health but are, with few exceptions, states of ill health.

The tests of the validity of measures of health status fall into three groups: criterion, construct, and content validity. Criterion validity is determined by comparing a measure with an external 'gold standard'. Criterion validity has little relevance, however, where there is no agreed operational gold standard. Construct validity examines the relationship between underlying theoretical concerns and the measure. It relies on a proper theoretical basis for the particular measurement, but again this is of doubtful relevance where the theoretical base is so under-developed.

Content validity concerns the inferences that can be drawn from the results of the measure. It depends usually upon the subjective assessment of an expert, and so potentially has greater pertinence. There are, however, no clear 'experts' in the field of subjective health status, although it could be argued that lay respondents could fulfill this role. There has been some work assessing lay perceptions and experiences of health (see 27 30 31), but as it has been conducted in the qualitative tradition, there has been little constructive cross-fertilisation. Some measures, such as the NHP, have attempted to draw upon statements made by the general population, but the methods of selecting and refining statements into the final profile have not been made explicit.

Qualitative research offers a pertinent method for studying the validity of structured instruments from the perspective of the respondent. The findings of this qualitative research allow insight into the extent that individuals’ scores truly represent their views concerning their own health. A study of regional variations in lay perceptions of health and health care can be used to illustrate this point.27 32 33 Here respondents were asked to fill in part of the General Household Survey concerning health and social attributes; two measures of health and health care developed by the Rand Corporation; the NHP; and then to participate in a semi-structured interview about their health care and any issues raised by the standardised instruments. All respondents agreed that their interview should be tape-recorded and all the material was fully transcribed. We concentrate here on the measure that has become widely used to determine health status, the NHP, a self-completed profile of aspects of perceived physical, social, and emotional health. Although we have concentrated on the NHP, however, the issues raised are not exclusive to this one instrument but apply to the genre.

A number of aspects of the NHP created uncertainty and difficulty for respondents. The NHP is answered according to the simple 'yes'/no' dichotomy. This leads to straightforward analysis, but does not allow people to express what they really feel. The tape recorder was able to capture this: (In the following excerpts from tape transcriptions with different respondents, the NHP statements are in italics).

(Things are getting me down) 'I won't let them if I can. Can I put sometimes?'
(I have pain at night) 'Not so much pain as discomfort, so what do I put?'
(I'm tired all the time) 'I'm tired, but not all the time."
(I have unbearable pain) 'It's hard to answer some of these. I don't have unbearable pain except when I have the backache. Is that yes or no?'
(I take tablets to help me sleep) 'I take tablets at night for the cramp and they help me sleep. What do I put there?'

The statements made people think about aspects of their health, but then constrained their responses. Many found this frustrating, and large amounts of useful information would be lost when a tape recorder was not used. The apparent simplicity of the 'yes'–'no' dichotomy was far from simple in practice. The forcing of responses into predefined categories negated people's desires to negotiate the meanings of health and illness.

Another interesting finding from these tape recorded interviews was that respondents often discussed their possible answers out loud, but then filled in the schedule seemingly arbitrarily, sometimes clearly contradicting what they had thought. In part, this was because the meanings of the statements seemed unclear and unspecified, but it also resulted from people's desire to give socially acceptable responses and to conform to peer pressures. In most cases, they did not want to be categorised as being ill or disabled in some way:
(I find it hard to bend) 'I do find it hard to bend, but I'm not ticking yes there.'
(Worry is keeping me awake at night) 'Well yes, but it's only stupid things. I lie awake thinking. I'll put no because I'm just being stupid.'

In some cases, people misunderstood the purpose of statements and responded according to problems unassociated with their health:
(I find it hard to reach for things) 'I do find it hard to reach for things, yes, because I am short.'
(I'm waking up in the early hours of the morning) 'It's what I've always done, got up early in the morning. I find that quite easy, yes.'

Several other people also ticked the statement about waking up early in the morning because they had young children who always woke them up early. Respondents were ticking 'yes' even though they did not perceive any morbidity. Such res-
Reliability and responsiveness

The reliability of a measure is the extent to which the measuring procedure yields the same results on independently repeated trials under the same conditions. Reliability is typically measured by coefficients, defined as the proportion of the score that contains information about the concept of interest as distinct from random error. Conventional methods aim to decrease random error and to achieve scores as close to 1 as possible (representing 0% random error). There are several types of coefficient, each referring to the different sources of measurement error: internal consistency, interobserver reliability, intraserver reliability and test-retest reliability.

The levels of reliability coefficients deemed to be acceptable seem somewhat arbitrary: 0.6 for the internal consistency of the AIMS24 compared with 0.97 for the 235 item version of the SIP, for example. There is also some variability within instruments, with global measures reporting high test-retest and internal consistency scores, but low values for particular items. The timing of retesting can also be crucial to the levels of reported reliability, yet the choice of time interval again seems somewhat arbitrary, ranging from one hour to one year.

In addition to the aspects of reliability that lie within the conventional idiom of epidemiological method, more fundamental questions can be raised concerning variations of response to health measures. If we are interested in the question of health status, we must accommodate the variation that is one of its key characteristics. The values of many of the variables associated with health fluctuate. The ability to climb stairs, for example, may vary during the day for someone with rheumatoid arthritis, or it may be possible with a stick but impossible without. Health status measures tend to measure these abilities in static and rigid terms—whether or not they can be done, rather than the degree of difficulty—thus failing to incorporate their variability. As we have seen, the act of forcing complex experience, including variation, into a rigid scheme can result in statistics whose meaning is difficult to interpret. Qualitative work, however, aims to incorporate this rich variety of experience and response within its analysis.

Structured health measures, because they concentrate on indicators of disease, tend to focus on the most severely ill, missing out large proportions of the population who may have some symptoms, but are identified in this way as ‘healthy’. It has been found, for example, that approximately one third of all respondents to the NHP score zero. These zero scores imply that (or at least the absence of any ill health), and yet they will mask a range of people with discomfort and illness.

There are many problems with the scoring of scales, particularly in the adding of scores when measurements have not been at the interval or ratio scales. Particular scores can be achieved in a number of ways that may not be comparable. Many instruments also have biases, including end aversion (the avoiding of extreme values); straight line (ticking responses one under another); and acquiescent response set (the tendency to agree with statements). The act of weighting responses can also build in and accentuate biases.

A more general problem that follows from the demand for consistency that is characteristic of the scientific method is the potential incompatibility between methodological development directed towards high reliability and the supposed subject matter of interest. The pursuit of consistency may lead to the removal of those very questions that explore the qualities of health itself, particularly its complexity and variety. The search for high reliability may result in the evacuation from the instrument of those very questions which may capture the quality of subjective health.

Structured health status instruments largely reflect the values of their originators. Researchers can impose their values, consciously and unconsciously, at almost every stage in the development of instruments. Reliability statistics give insight into the appearance of being value-free, but in fact, researchers can further impose their own values by determining the levels of tests that are acceptable and by setting the conditions under which tests are taken. Instruments can be ‘made’ more reliable by the exclusion of items that skew test results. The values of researchers are also involved in the decisions about scoring techniques and weightings of instruments.

The researcher faced with a choice between health status instruments will want to be able to make a judgement according to the instruments’ reported levels of reliability, validity and responsiveness. Unfortunately, these are not always readily available. As McDowell and Newell report in their review of very many measures: ‘from our experience in writing these reviews, it has become clear that many methods lack information on relevant issues such as precisely for whom the measurement is intended, how valid it is and
Assessing the need for health status measures

Perceived health status and needs assessment
In spite of their acknowledged failings, structured health status measures continue to be used because the health status of the population is patently a legitimate interest and because these measures purport to reflect this quality. We cannot, however, evade the question of what it is that these instruments actually measure. In attempting to become an accepted part of the scientific paradigm, many instruments have become overly concerned with their abilities to be reliable and valid as measured by statistical tests. Subjective health then becomes a single number or set of numbers that are taken out of their context and used in policy making alongside so-called 'hard' data. It remains to be proved that aspects of perceptual health can be reduced to such simplistic numbers. These numbers are, in any case, often derived somewhat arbitrarily. The instruments and their scores are grounded in the values of the originators, obfuscated by apparent scientific reliability and validity.

Instead of addressing these prior concerns, the debate about the measurement of health seems somewhat stuck at the level of which instrument to choose. Researchers are encouraged to choose a measure that is reported to reflect this quality. We cannot, however, evade the question of what it is that these instruments actually measure. In attempting to become an accepted part of the scientific paradigm, many instruments have become overly concerned with their abilities to be reliable and valid as measured by statistical tests. Subjective health then becomes a single number or set of numbers that are taken out of their context and used in policy making alongside so-called 'hard' data. It remains to be proved that aspects of perceptual health can be reduced to such simplistic numbers. These numbers are, in any case, often derived somewhat arbitrarily. The instruments and their scores are grounded in the values of the originators, obfuscated by apparent scientific reliability and validity.

The SF-36 is now, in turn, replacing the NHP in epidemiological studies and outcome assessments. It is reported to be valid and reliable, more so than the NHP, although the issues of face validity explored above are no better resolved for this instrument than for its predecessors.

Health is an extremely complex concept. It is closely related to other complex concepts such as well being and quality of life. All of these have difficult and variable definitions which have been tackled rarely in the published reports. Health status instruments seem to offer an easy solution: the measurement of health by relatively simple instruments, thoroughly tested for reliability and validity. They are also attractive to those familiar with the methods of natural science because they produce quantitative results. The use of statistics and statistical tests have given many instruments such a sense of validity and reliability that individual researchers make judgements based on these criteria rather than on what they actually want to measure in their studies.

Generic health status measures have a clear role in complementing condition-specific studies, particularly in the measurement of the outcome of treatment. In this application it is important to complement these measures with the sort of qualitative analysis already described. This commentary upon measures of health status is, however, directed particularly towards their potential relevance in health needs assessment. A more basic problem arises that the attributes that these measures focus upon are not especially relevant to the sort of health needs that must be assessed by those purchasing care on behalf of particular populations.

Health needs assessment is a broad term that encompasses two quite distinctive purposes. On the one hand there is the interest in those broad characteristics of the population that may reflect their general state of health. The research interest in such health needs has been focused particularly upon questions of inequality and health. In policy terms such measures of health needs may inform broad patterns of resource allocation.

The more intimate decisions facing those responsible for a local pattern of health care are quite different. Here health needs are those services that are necessary for the alleviation of particular forms of morbidity. The focus is therefore upon the services that are necessary, rather than the neediness of the population. The term need is unfortunate in collapsing both meanings. As discussed elsewhere, there are advantages in using a quite different term for each form of need. Health needs can be reserved for the broader environment of individual health, encompassing questions of deprivation and inequality. The term health care requirements rather than health needs would be a preferable term for the needs for particular services. The global view of health needs can then be complemented by a focus upon the requirements for particular interventions.

The potential contribution of generic health status measures can be considered in terms of this distinction. The assessment of health needs, the broader environment of health, can draw upon the range of demographic, socioeconomic, and morbidity data that is available. While some assessment of subjective health might in principle be desirable, it is not apparent how such measures would in practice modify the interpretation of conventional measures of health need.

The assessment of health care requirements can draw upon current patterns of utilisation, but this source is necessarily unsatisfactory as demand is known to reflect supply. Empirical assessments of health care requirements are clearly desirable, but in this case the level of generality characteristic of health status measures makes their interpretation difficult. It is not possible to map general measures of pain, immobility or distress onto the requirements for specific services. Measures mask multiple disabilities, and although they can identify some states of ill health, they cannot...
identify the cause and/or nature of the dysfunction and thus cannot inform interventions. 39

Generic health status measures seem therefore to fall between the two stools of reflecting health needs and reflecting specific health care requirements. It is not clear what would be gained from the health status questionnaire material that might not be found more economically from routinely available statistics. Where particular health care requirements are to be determined, the generic nature of the information retrieved by many of these instruments would, even if it were useful, be inadequate for supporting detailed purchasing decisions.

Aspects of this work were funded by the Economic and Social Research Council and the Department of Health.


27 Eyles JD, Donovan JI. The social effects of health policy. Aldershot: Avebury; 1990.

28 Bergner M. Quality of life, health status and clinical research. Medical Care 1989; 27(3) suppl: S148-S156.

29 Mulley AG. Assessing patients' utilities: can the ends justify the means? Medical Care 1989; 27(3) suppl: S209-S281.


