# General Measures of Health for use in Health Interview Surveys and Censuses: the UK experience

Professor Howard Meltzer
Social Survey Division
Office for National Statistics
London, SW1V 2QQ
United Kingdom

# 1 Introduction

In the late 1990s, the Department of Health in England commissioned a team of social researchers from the Office for National Statistics (ONS), the National Centre for Social Research (NSCR) and the London School of Economics (LSE) to make a comprehensive review of general measures of heath used in population surveys and the Census in the UK. (Sturgis et al., 2001)  Their findings, *Comparative Review and Assessment of Key Health State Measures of the General Population*, were published on the web (www.doh.gov.uk/pdfs/healthreport.pdf) in September 2001.

The aim of this review was to provide information that would assist in the design of surveys and the interpretation of survey results on the *general health* of the population resident in private households in the UK.

The project was empirically driven, being based around results from large population surveys in the UK. Several of these had been commissioned by governmental bodies over the previous decades. These included the General Household Survey (GHS), the Health Survey for England (HSE), the Health Education Monitoring Survey (HEMS) and a number of others.

Evidence from health surveys conducted outside central government was also considered and special attention was paid to 'calibration' exercises designed to throw light on the way questions about respondents' *general health* are interpreted and answered by members of the public.

Whereas the review included many general health measures, this paper has extracted those sections concerned with measures of health covered by one or two questions which are generally applicable, cover a range of dimensions of health of importance to the public, and are simple to understand and use in sample surveys of the general population.

## 1.1  Use of *general health* measures

At a population level, general health measures can be used to produce prevalence estimates and thus provide a method of monitoring the population's health and of assessing the likely demand for health care and services.

When used to produce health measures at a national level, they may be collected either via a *census* or a *sample surve*y. The use of the decennial census in the UK to collect health information has the advantage of near complete coverage of the whole population and, thus, the ability to provide estimates for small areas of the country.

The disadvantage of using a census, apart from the cost, is that space on the census form is limited, and the forms are normally completed by one member of the household, without the presence of an interviewer to probe for a full answer. In the UK, census forms are put through letter boxes by enumerators and returned by post. In addition, there are long periods between censuses, which means that information collected on them will become outdated.

Because of these disadvantages, the monitoring of the nation's health is normally

carried out by the inclusion of health measures in continuous or frequently repeated surveys of the general population, using face-to-face interviewing. Therefore the main focus of this paper is the measurement of health in the face-to-face sample survey context. However, the national Censuses for 1991 and 2001 are included here because they include general health questions and therefore have an important place in health monitoring strategy.

## 2. Conceptual issues

### 2.1 Why are questions about *'general health'* included in surveys?

Single questions or brief sets of questions about *general health* are frequently included both in specialised health surveys and in general surveys of the population. Three main needs seem to underlie this popularity.

The first need is to control both the burden on respondents and the cost and complexity of surveys by minimising the number of questions on any one topic that has to be included in a questionnaire. A single question providing an indicator of general health is cheap and may appear straightforward to interpret. Simplicity is an important advantage, particularly in the case of large surveys that present and compare results for many sub-samples and over time. If survey respondents are willing and able to answer these simple-seeming questions, why incur more expense and complication by asking more?

The second need, which is implicit in all quantitative survey work, is to derive a simple indicator (or small set of indicators) to subsume the detail which emerges when a person is questioned in depth about something as complex as his or her state of health

A third reason for including general health measures in surveys may be as a relatively straightforward way of estimating the 'burden of ill health' in the population. Here there may be a subtext which defines 'ill health' as 'that which requires input from the health services'. Without such an indicator there is no simple way of using a continuous or repeated health survey to answer to the question 'How well are we doing in our efforts to improve health in the population?' Of course, the fact that such a measure would be very useful does not, of itself, imply that it should not be subject to serious methodological scrutiny.

### 2.2 The concept of 'general health'

The concept of *'general health'* may at first sight appear rather straightforward and commonsensical. In everyday conversation we often address to a friend or acquaintance such questions as 'How are you these days?'. Sometimes this is mere politeness, but sometimes we actually expect.and appear to receive more or less informative answers bearing on the person's *general health*.

However, in asking and answering such questions we seldom give conscious attention to such issues as:

- whether we mean the individual to take account of stable long-term conditions or disabilities, or only of recent or acute episodes of ill health;

- whether they are giving appropriate and consistent weightings to different aspects of health, mental and physical;

- whether or not we expect them to 'discount'  health problems associated with advancing age;

In short, we seldom ask ourselves whether or not the person we are talking to is drawing the same conceptual boundaries around the idea of *'general health'* as we do ourselves, or as other persons to whom we might address the same question. Interview respondents probably take formal interview questions more seriously than casual conversational enquiries, but the evidence suggests that terms within the conceptual domain of *'health'* are unlikely to be interpreted very consistently either across different individuals or within the same individual over time.

Even if respondents appear to understand consistently what they are being asked to do in providing an assessment of their *general health*, there is no way, without special cognitive studies, that we can assess whether the response given is based on careful and comprehensive thought and the application of 'reasonable' standards of judgement, or not. The exception perhaps occurs in cases of gross discrepancy, such as where persons who on objective evidence appear to be very ill give responses suggesting that they have no health problems

In other words, there can be no ultimate "gold standard" that can be applied to questions inserted in health surveys to distinguish "correct" from "wrong or misleading" responses to questions asking for a self-assessment of health.

## 2.3 Cognitive tasks required by *general health* questions

To arrive at a single, summary answer to a question about his or her *general health* the respondent must, in theory, strike a weighted average of how they feel they stand on different dimensions of health (the weights representing the importance to them, personally, of the different dimensions). In trying to understand what really goes on in respondents' minds as they take in a question about their *general health*, decide what is required by way of an answer and apply standards and judgement to their personal experience in order to produce a response, our main source of information has been to the relatively small amount of "cognitive" question testing work that has been done on the way *general health* questions are answered.

Indicators of different aspects of health may, perfectly legitimately, move in different directions over time. For example, a person's mobility may improve while their sight (or digestion, or depression, or migraines) get worse; and differential movement of indicators will often be observed at the population as well as at the individual level.

## 3. Methodological issues

In order to assess the quality of data derived from the responses to survey questions, a well-established list of criteria is available to social researchers.

## 3.1 Validity

Validity is the degree to which a measure captures the concepts it is intended to measure and is not *systematically* affected by other, irrelevant variables. Also, the *same* concept needs to be measured in the same way, and using the same standards, for all respondents. There are several different ways of assessing validity.

*Face validity* appeals to semantic or observational judgements of whether the measure being evaluated appears to capture what it is intended to capture. For example the response category 'Yes' to the question 'Do you suffer from any longstanding illness or disability?' has face interpretability.

*Criterion or external validity* makes comparisons with other sources. Criterion validity looks for appropriate correlation between the measure being evaluated and some other independent and trusted measure or classification of the same concept. For example, correlation with clinical diagnosis of severe arthritis might be used to validate a questionnaire item on long standing health problems. However, it would be very surprising if a measure with face validity as a measure of longstanding health problem did *not* produce *some* degree of positive correlation with diagnosed arthritis. To be convincing as validation of the measure a *very strong positive relationship* would need to be shown.

*Construct validity* is assessed by testing theory-based predictions of the pattern of statistical relationships between the measure being evaluated and other, conceptually-related measures. For example, a variable said to measure "social isolation" might be predicted to correlate more highly with "living alone" and "mobility problems" than with "digestive problems". Again, it is not enough for the predicted pattern of correlation to be present andstatistically significant on large samples. To be convincing the observed differentials need to be quite large.

*Predictive validity* is assessed by testing theory-based predictions of how health-related outcomes (for example, hospitalisation or death) should vary for cases having different scores on the measure. Once again, almost any measure claiming to detect serious ill-health should be associated with a higher-than-average chance of early death. To provide convincing proof of the validity of the measure the prediction achieved needs to be striking, *even after controlling for other variables such as age.*

## 3.2 Freedom from overall bias

The idea of freedom from overall bias is linked to, but not the same as, the idea of 'validity' and also to that of 'sensitivity' Clinical examination of a representative sample of the population would probably show that 'perfect health', like 'very poor health' was relatively rare, though the health defects suffered by many of the population would no doubt be relatively trivial or latent (such as, for example, unfitness and obesity due to lack of exercise or poor diet, which is known to be a predictor of serious diseases in middle age). Therefore it could be argued that a

questionnaire measure of general health that suggested that the majority of the population had no health defects is either biased in an 'optimistic' direction, or, alternatively, that it is insensitive to real differences in health within that part of the population which is free of *major* health problems.

## 3.3 Sensitivity

Measures should be sufficiently sensitive to differences in health states. A measure needs to able to detect changes over time, or mean differences between groups, in the aspect of health that it is intended to measure. This sensitivity should ideally be standard over the whole range of the underlying health variable, so that there are neither "ceiling effects" (loss of sensitivity in distinguishing "very good" from "good" health), nor "floor effects" (loss of sensitivity in distinguishing "very poor" from "poor" health).

## 3.4 Freedom from bias between sub-groups

If our aim is to monitor the health of all sections of the general household population, it is important that the criteria above apply equally to all subgroups of the population, so that the health of *particular subgroups* is not *spuriously* represented as being better or worse than that of other subgroups. In other words, measures must be *equivalent in their meaning and interpretation* for all members of the population. A degree of *random* variability in how individuals interpret and answer survey questions is tolerable, but *systematic relative bias* in the way questions are interpreted and answered (say) by men versus women, or by younger people versus older people, undermines the aims of monitoring (with a view to determining health policy priorities), unless it can be corrected for in some way.

## 3.5 Reliability

The term *'reliability'* is here used in the technical sense which distinguishes it from *validity*. Ideally it is assessed by special *test-retest* studies. A reliable measure is one that is not subject to excessive random variability in the results it obtains at the individual level. The presence of such *measurement variance* has the same effect on survey estimates as a reduction in sample size.

## 3.6 Portability

Measurement instruments used in monitoring must not be prone to relative bias in their application. For flexibility in developing a health monitoring strategy it may be desired to mount a measure on different survey vehicles and to vary the precise questioning context or use of proxy responses.

It is therefore very desirable that a measure should be *portable between surveys* in the sense that it will produce the same results, irrespective of whether it is included on a dedicated health survey or a multi-purpose survey *(freedom from context/order bias).*

The ideal measure should also be independent of *mode of administration* (whether the measure is administered by telephone or face-to-face, for example) and use of

*proxy response*s. Given the fact that health means in the population tend to change rather slowly and that small changes are therefore of interest, lack of portability in measures may have serious consequences in causing statistical artifacts that may be mistaken for true changes or differences in *general healt*h.

### 3.7 Practicality

While the criteria already discussed are of prime importance in scientific and methodological terms, *a criterion which in practice tends to outweigh them* is practicality. In survey contexts this concerns:

a) the length of time it takes to administer and complete the survey questionnaire module concerned and hence the associated operational and opportunity costs;

b) whether the results will slot readily into an existing time series and offer scope for useful comparisons.

c) whether survey respondents seem able and willing to provide answers to the questions involved without any untoward reaction (acceptability);

d) the cost and complication of processing the resulting data;

e) the suitability of the results for presentation in descriptive survey reports.

### 3.8 Stability of measurement performance over time

To fulfil the purpose of monitoring health over time, it is important that the format and wording of the measure used and *the way in which they relate* to what we intend to measure be *invariant over tim*e. Then one can be confident of interpreting a change over time in the measure as indicating a real change in the population's health, not a change in expectations or in the relative weight of different components of the measure.

The ideal measure would be one that also provided continuity with available time series, thus extending the current monitoring data rather than having to start a new series. However, some doubts arise over whether the health standards applied by respondents do remain stable over time. For example, the aims of monitoring over time are undermined if what respondents in one year understand and intend (on average) by the response "My health is good" is different from what respondents understood and intended by the same response five years earlier.

### 3  The self-assessed general health question.

### 3.1 Use of the question in the UK

Surveys  in the UK which have included a question on self-reported general health include:

The Health Survey for England;

The General Household Survey;
The Health and Lifestyles Survey
The Health Education Monitoring Survey
The ONS Psychiatric Morbidity Surveys
The National Child Development Survey
The 1970 British Cohort Survey
The Allied Dunbar National Fitness Survey
The Health Education Authority's Today's Young Adults Survey.


## 3.2 Range of questions used in the UK.

The Health Survey for England uses the following question, which is recommended by the WHO Regional Office for Europe, as an instrument for collecting internationally comparable data for measuring progress towards achieving WHO-Europe Health for All targets.

Use of this question therefore provides a basis for international comparisons of self-assessed health, although respondents' understanding of what constitutes 'good' or 'bad' health will be influenced by cultural and historical contexts

*[*] Now I would like to ask you some questions about your health. How is your health in general? Would you say it was..*
*RUNNING PROMPT*
*1 very good*
*2 good*
*3 fair*
*4 bad*
*5 or very bad?*

The General Household Survey (GHS) has included a single-item question since 1976 and therefore offers twenty five years of annual estimates and the Health Survey for England has included a question since its inception in 1991.

The General Household Survey uses the following question which, unlike that used by the HSE, specifies a time period.

*[*] Over the last 12 months would you say your health has on the whole been...*
*1 good*
*2 fairly good*
*3 or not good?*

Questions on other surveys ask respondents to compare themselves with others; the question used by the Health and Lifestyles Survey, for example, asked respondents to say how good their health was 'for someone of your age'.

Although self-assessed health is often measured by a single item, there is widespread evidence that this question nevertheless covers several dimensions of health, and that people implicitly go through a process of considering and weighing these dimensions when answering the question.

### 3.3 Cognitive testing of the question

Respondents to the 1984 Health and Lifestyles Survey, for example, were asked what they understood by the term 'health': among the aspects which they mentioned were absence of disease, functional ability, and fitness (both physical fitness and psychological well-being). Also identified were a 'moral' dimension, whereby health depended on will-power, self-discipline and self-control; health as healthy behaviour (being a non-smoker or non-drinker, taking exercise); and health as a 'reserve' which could be diminished by neglect and accumulated by good behaviour (Blaxter, 1990).

Cognitive work carried out for the pilot phase of the 1997 Health Education Monitoring Survey (HEMS) identified very similar themes. Respondents interpreted 'health in general' as absence of ill-health, the ability (or not) to lead a normal life, a state of mind, and physical fitness   Participants in the 2001 Census question-testing programme also referred to frequency of doctor consultations, whether or not people were absent from school or work because of ill-health, and whether or not they were taking medication.

### 3.4 Socio-economic differences

Many questions on self-assessed health were specifically designed for inclusion in surveys of the general population. As single items, they take very little time in an interview or when a respondent is self-completing a questionnaire. There is evidence (Calnan, 1987) that those with higher levels of education are able to produce more elaborated definitions of health; there may be therefore systematic differences between social groups in their understanding of questions and hence in the meaning of their answers.

Blaxter (1990) believes that this distinction does not hold when people are encouraged to elaborate on their ideas in an in-depth interview, but warns that respondents do not have the time to do this in most surveys. It may be that less well-educated respondents are more likely to draw on narrower concepts of health in the survey setting.

### 3.5 Validity

Self-assessed general health has been shown by studies in several countries to be a good predictor of mortality. In the UK, a follow-up study to the Health and Lifestyles Survey (HALS2) showed that, after the existence of a serious disease, self-reported poor health was *one of the most powerful predictors of mortality*. Among those who said in their 1984 interview that they had no serious disease, men at all ages who assessed their general health as 'fair' or 'poor' were twice as likely as those who rated it as 'excellent' or 'good' to die in the seven years between the initial and the follow-up study. For women, self-assessment was a good predictor only for those aged 55 or over (Blaxter & Prevost, 1993).

Similar studies in Sweden (Sundquist and Johansson, 1997), the USA (Berkham and Syme, 1979; Idler et al. 1990) and France (Grand et al. 1990) have shown similar results. The Swedish study had a very large sample of almost 40,000 respondents. It

found that poor self-reported health status was a significant risk for men and women of all ages, when the effects of age, marital status and low socio-economic status (measured by educational level and tenure status) were controlled for.

The validity of questions on self-reported health has also been tested by comparing them with other measures of health. In an analysis of the 1984 Health and Lifestyle Survey results, Blaxter (1990) constructed a health index based on four dimensions: the presence or absence of disease, the presence or absence of illness (as measured by reported symptoms), fitness and unfitness, and a measure of perceived well-being. The presence or absence of disease was partially validated by nurse assessments and by details of medication reported by respondents. The fitness/unfitness dimension was based on physiological measurements such as Body Index, blood pressure and respiratory function. Blaxter found a high level of agreement between self-reported general health and the index at the two 'extremes'; that is, those whose measured health was best and worst (as measured on the four dimensions) were most likely to give an appropriate self-assessment.

Self-assessed health has also been shown to be associated with doctor consultation rates, with the mean rates of consultation increasing as self-perceptions of health deteriorate. However, Blaxter (1985) found that, once social class was taken into account, self-assessments and consultation rates were clearly associated only for those belonging to the manual social classes; she suggests that not consulting is part of the definition of being in good health for these groups.

Evidence suggests that there is an overall tendency for respondents to give positive rather than negative assessments of their health, but as with other measures discussed here, there are systematic variations between the assessments given by people in different social groups. Evidence from a number of surveys suggests that older people have lower expectations of health, and are more likely to make a positive assessment of their health than a younger person with similar illnesses or symptoms might; they consider themselves healthy despite the difficulties associated with ageing. Similarly, people with a disability can give assessments of their health as good, 'despite the disability'. Those in families where the head of household is defined as belonging to the manual social classes are more likely to make a more pessimistic assessment than objective measures suggest is appropriate (Blaxter, 1990).

People in different social groups also emphasise different dimensions in their definitions of health; functional ability is more likely to be mentioned by older people, and fitness by younger people. Psychosocial well-being is stressed more by people in the middle years, by women and by more highly educated respondents.


### 3.6 Reliability

Data from the 1997 HEMS (Bridgwood et al. 1998) indicate that individual changes in self-rated health are associated with objective changes in health. The 1997 survey was a follow-up, in which respondents who were first interviewed in 1996 were interviewed for a second time in 1997. As well as being asked about their health, they were also asked whether they had experienced one or more of a series of

events in the last year. Those who reported a serious illness, injury or operation since their first interview were three to four times as likely as other respondents to give an assessment of their health in 1997 which was more than one category 'poorer' than in 1996.

Blaxter (1990) warns that people are often inconsistent in their assessments of their own general health. One of the reasons for this may lie in the answer categories available. The cognitive work carried out for the 2001 Census and the 1997 HEMS pilot explored respondents' understanding of the different answer categories. The 'fairly good' category in the GHS question and the 'fair' category in the HSE question were least easy to define; 'fairly good' was considered to be a vague term, while 'fair' was seen as an average of good and bad days. Those who described their health as 'fair' in the 1996 HEMS were most likely to have changed their assessment; less than half used the same description in 1997. Similarly, about one in six of those who described their health as 'good' and more than a quarter of men and more than two fifths of women who said it was 'bad' in 1996, opted for 'fair' in 1997. If the term 'fair' is difficult to define clearly, than it is perhaps not surprising that some respondents change their assessments over time. Similarly, some respondents had difficulty distinguishing between the 'very good' and 'good' categories in the HSE question; some movement between these two categories is therefore perhaps to be expected.

## 3.7 Ease of interpretation

Responses to questions on self-reported general health offer a simple summary measure with an intuitively comprehensible meaning, which can be used to compare different social and health status groups. They give an overall summary assessment of health, although it is difficult to know whether any differences in reported health for a given population over time are real differences or a difference in the relative weight attached to the component dimensions of health, particularly as these dimensions are implicit rather than explicit. When analysing differences between social groups, it should be borne in mind that there are systematic differences in the dimensions which respondents have in mind when making an assessment of their own health, and in the extent to which these assessments correlate with more objective measures

## 4. Long-standing and limiting long-standing illness questions

## 4.1 Use of the question in the UK

Surveys in the UK which have included a question on longstanding illness include:

The Health Survey for England;
The General Household Survey;
The Health and Lifestyles Survey
The Health Education Monitoring Survey
The ONS Psychiatric Morbidity Surveys
The National Child Development Survey
The Survey of the Physical Health of Prisoners

The National Survey of Sexual Attitudes and Lifestyles
The 1991 and 2001 Censuses
The National Child Development Survey

Questions on long-standing illness or disability have been included in the General Household Survey since 1971 (with a separate question on limiting long-standing illness since 1974), with a break in 1977 and 1978, which provides time series data spanning a period of over 25 years. A question on limiting long-standing illness was included in the Census for the first time in 1991, and repeated in 2001 in part to obtain an improved indicator of the likely need for health services for small areas than could be produced from survey data.

## 4.2 Range of questions used in the UK.

The Health Survey for England, the General Household Survey, and many other surveys, use the following question:

*[*] Do you have any long-standing illness, disability or infirmity? By long-standing I mean anything that has troubled you over a period of time or that is likely to affect you over a period of time?*
*1 Yes*
*2 No*

The GHS also asks whether the condition is a limiting one:

*[*] Does this illness or disability (Do any of these illnesses or disabilities) limit your activities in any way?*
*1 Yes*
*2 No*

The 1991 Census used the following question:

*Do you have any long-term illness, health problem or handicap which limits your daily activities or the work you can do? (Include problems which are due to old age)*
*1 Yes*
*2 No*

The 2001 Census used a slightly different question:

*Do you have any long-term illness, heath problem or disability which limits your daily activities or the work you can do? (Include problems which are due to old age)*
*1 Yes*
*2 No*

These core questions are sometimes supplemented with further questions on Activities of Daily Living (OPCS, 1994) or by a checklist of symptoms (Health Promotion Trust, 1987).or asking "What is the matter with you?) (General Household Survey)

The question asking for details of illness is sometimes asked only as a courtesy with no intention of analysing the responses, as in most years of the GHS; at other times, interviewers are asked to probe the nature of the self-reported illness or disability fully. This was done in 1988, 1989, 1994 and 1996 for the GHS, for all years of the Health Survey for England, for the first Health and Lifestyles Survey and for the Survey of the Physical Health of Prisoners.

The dimensions of health covered by the questions are not explicit, but there is some evidence that they measure physical morbidity more successfully than psychiatric morbidity

Answers to these questions are used to produce estimates for the prevalence of self-reported long-standing and limiting long-standing morbidity among people living in private households. Long time series, such as those produced by the GHS, provide a point of comparison for local, ad hoc or irregular surveys. International comparisons are possible, as other countries use similar questions, although prevalence estimates will be influenced by cultural understandings of illness, disability and normal activities. The data have also been used to produce estimates of Healthy Life Expectancy (Bone et al. 1995b) and combined with other measures, including more objective measures such as blood pressure and lung function, to produce a summary scale of health (Blaxter, 1987).

## 4.3  Use in surveys of the general population

Questions on long-standing illness and disability are short and easy to administer and therefore take little interview time. They can, however, be sensitive to changes in question wording and to mode of administration. For example, the overall prevalence of limiting long-term illness as measured by the 1991 Census among those resident in private households was 12%, significantly lower than the estimate of 18% from the 1991 GHS. The authors of the 1992 GHS report argue that differences in methodology accounted for some of the difference; the census information was collected by self-completion, usually by one member of the household and related to one night in April, while for the GHS all adult members of the household are interviewed individually by a trained interviewer and fieldwork goes on throughout the year (Thomas et al. 1994). The change in wording to include reference to 'the work you can do' may also have contributed to the discrepancy'.

A comparison of responses to the Census question and to an identical question on the 1991 Census Validation Survey (CVS) found a 'gross error', that is the proportion of times the answers to the two studies were different, of 4.9% (Heady et al. 1996). Higher estimates of prevalence were obtained in the Census Validation interview than from completed Census forms. The authors of the CVS report point out, however, that the differences may reflect genuine changes in health between the Census and  the survey, or lack of knowledge on the Census form-filler's part about the health of other members of the household. The comparison between the Census and GHS questions, together with several other studies, also show that quite small differences in survey design, question wording and possibly in question order also appear to influence response (OPCS, 1994). In this regard some of the most prominent effects are:

- Surveys which attempt to measure both limiting and non-limiting chronic illness with one question tend to produce lower overall estimates of prevalence than those which ask two separate questions.

- Asking whether respondents 'have' a long-standing illness produces higher estimates than asking whether they 'suffer' from an illness; some people may answer 'no' to the latter on the grounds that they are not actually suffering (Goddard, 1990).

- Asking whether an illness limits activities compared with 'people of your age' produces lower estimates than asking whether it limits them 'in any way'; it is believed that elderly people in particular would say no because most of their contemporaries were as limited in their activities as they were (OPCS, 1975).

- Using a checklist of symptoms stimulates reporting (Blaxter, 1987). One advantage of a checklist is that it provides all informants with a common frame of reference; it is possible, however, that it might produce overestimates of prevalence as informants who are not sure whether they have a condition might include themselves (Goddard, 1990). A checklist cannot be used to produce accurate prevalence estimates for more serious diseases as sufferers are more likely than others to be in hospital or unavailable for interview (Blaxter, 1987).

- Analysis of GHS data suggests that asking informants for full details of their illness before they are asked whether the illness limits their activities might result in lower estimates of *limiting* long-standing illness or disability. The authors of the 1988 report suggest that some informants may be reluctant to say that an illness limits them when interviewers know what it is; they also note, however, that unexplained fluctuations in the levels of self-reported limiting illness were a feature of GHS data throughout the 1980s (Foster et al. 1990).

- Asking interviewers to use directed probes, rather than generalised ones, can result in marginally more codeable conditions being reported. The cognitive question-testing carried out for the 1997 HEMS pilot found that respondents were able to define the terms 'illness' and 'disability' without difficulty, but that some had difficulty in understanding 'infirmity'. For some respondents, infirmity was synonymous with old age.

### 4.4 Validity

Assessments of the validity of questions on long-standing illness or disability have been based on comparisons with standardised mortality ratios (SMRs), the results of clinical examination and doctors' reports. They show a high level of agreement for overall prevalence, although the level of agreement varies for specific conditions and for different social groups. Commentators note that discrepancies do not necessarily indicate that data from self-reported sources is inaccurate; informants may not have brought a condition to the attention of a doctor, medical records could be inaccurate, doctors may not have informed patients of their diagnosis, and lay descriptions may differ from those given by doctors (White, 1995).

A comparison of age-standardised ratios for overall prevalence of self-reported

chronic sickness and standardised mortality ratios carried out for the first GHS in 1971, showed that for males, with the exception of Scotland, regions where SMRs were higher than expected also had higher than expected age-standardised ratios of long-term illness. This was also true for limiting long-standing illness and disability. There was less apparent correspondence between the two measures for females (OPCS, 1975). A similar comparison carried out at local authority level on 1987 Census test data showed correlations of 0.80 for men and 0.82 for women between all-cause mortality (as measured by standardised mortality ratios) and limiting long-standing illness (Charlton et al. 1994).

Interview data from the 1984 Health and Lifestyles Survey yielded an estimate of 30% overall prevalence of self-reported long-standing illness; information collected from respondents during a subsequent nurse session, which included recording details of medication, increased this estimate by only two percentage points (Blaxter, 1987).

Evidence from several sources indicates that these questions underestimate the prevalence of long-standing illness and disability among the elderly; for example, a proportion of informants who reported difficulties with Activities of Daily Living nevertheless say they had no chronic illness or disability. Even when there is no reference to 'people of your age', it appears that elderly people regard limitations in their daily activities, particularly difficulties with eyesight and hearing, as a normal part of growing old, not as evidence of illness or disability (Martin et al. 1988).

However, when the data from the 1991 Census Validation Survey were analysed, it was found that the proportion of those with a disability who reported a long-standing condition actually increased with age; the overall underestimation of chronic conditions among the elderly arose because the number who are disabled is much greater among the elderly than other age-groups, so that a slightly smaller proportionate under-recording produces a much larger absolute effect (Heady et al. 1996).

Supplementing the questions on long-standing illness with questions on Activities of Daily Living and on eyesight and hearing, as is done periodically on the GHS, is one way of improving estimates of prevalence for the elderly, as the estimates from the two different measures could be cross-referenced at the case level.

When comparing self-reported morbidity among different groups in the population, it must also be remembered that some people are more troubled by a certain kind of symptom than others, and that the need to limit activities will depend on what people usually do (Bennett et al. 1996). Informants may also vary in the amount of information they choose to give or in their knowledge of the extent and nature of their ill-health (Blaxter, 1990).

Comparisons have been made for estimates for specific conditions, as well as for overall prevalence. Blaxter (1990) found an 80% agreement between self-reported data and clinical assessments on the presence or absence of specific chronic conditions. The majority of the serious conditions which were reported were treated (and therefore presumed to be medically diagnosed); conditions which were most likely to be untreated were conditions such as varicose veins, migraine,

haemorrhoids and 'back trouble'. Those belonging to a non-manual social class were more ready to declare a chronic condition, even if it was not functionally troublesome or accompanied by symptoms. Informants in manual social classes, particularly men, were likely to say they had a named disease only if it was actually troublesome; this was particularly true for mental disorders. Very few of those with a severe condition said it did not affect their lives (Blaxter, 1990). Analysis of the 1987 Census Test results showed the highest correlations at Local Authority level between named conditions and standardised mortality ratios were for circulatory diseases (Charlton et al. 1994).

## 4.5 Reliability

There is little or no data on how well the questions on self-reported health problems or disability perform using a test-retest methodology. There is some evidence on reliability, however, from the 1997 HEMS; respondents who reported a serious illness, injury or operation in the life events section of the interview were twice as likely as others to give an assessment of self-reported morbidity which was poorer in 1997 than in 1996 (Bridgwood et al. 1998).

## 4.6 Ease of interpretation

Data from the GHS enable trends over time to be measured; these show year-to-year fluctuations, but the overall trend for both long-standing and for limiting long-standing illness and disability is upwards. Caution needs to be exercised, however, when interpreting changes in the prevalence of self-reported morbidity as changes over time may reflect changes in people's expectations of health as well as the prevalence or duration of sickness (Bennett et al. 1996).

## 5. Empirical comparisons

## 5.1 Introduction

Because we are dealing with surveys of the general population, we rarely have objective data on the health status of individuals in the sample. Thus, we must rely on self-reported measures of health status to evaluate other self-reported measures of health status, a circularity which it is hard to avoid when using general population survey data.

## 5.2 Context effects

There are a number of reasons why questions which aim to measure the same concept produce different estimates for the same population; even a relatively small difference in the wording of the question or of the response categories, as on the self-assessed health questions on the HSE and GHS, can have a significant effect. Consistency of results across surveys cannot, however, be guaranteed, even if identical questions are used, because of the context in which the questions are asked. There is a substantial body of methodological and survey literature demonstrating such context effects for a wide range of different types of questions. Secondary analysis of the HSE, GHS and other surveys provides evidence of the

scale of context effects for three of the general health measures under consideration: self-assessed general health, long-standing illness and limiting long-standing illness.

Thus, it can be seen that identical questions do not produce identical estimates – although any differences tend to be small. Differences could emerge for a number of reasons; if, for example, the surveys had differing approaches to the taking of proxy information, or if they were affected by different types of non-response bias. On the three surveys analysed, however, questions on health would not be answered by proxy as they are opinion questions, and, in general, all three have similar characteristics of non-response (younger adults tend to be under-represented). It is quite possible, therefore, that the observed variation may occur because of the context in which the questions are asked. It might be expected that there would be a difference between answers to questions asked on a general  survey, and those asked on a specific health survey, but there was also a difference between the two general surveys, the GHS and the Omnibus. Despite both of these surveys covering several different substantive topics, they are quite different in their actual content. The GHS carries relatively long question modules on major aspects of a person's life, such as housing tenure, education and employment, while the Omnibus carries a selection of much shorter modules that could be on a wide variety of topics. It may be that the latter survey does not encourage as much consideration of health issues before the answer is given, but this can only be speculation.

## 5.3 Service use

One way of validating health measures is to examine how they relate to use of health services. In this section results from the 1996 GHS are used to show the relationship between the health measures included in that survey, and whether or not a doctor had been consulted in the two weeks prior to interview. There is, of course, no reason to expect that all those reporting a  health problem will have consulted a doctor recently, particularly if the health problem is of a long-standing nature, but the proportions of those who have consulted do give some indication of the validity of the measure..

 22% of men and 30% of women with a long-standing illness or disability had consulted a doctor in the two weeks before interview, while slightly higher proportions of those with a limiting long-standing illness had done so. A better predictor of doctor's consultation (though not necessarily ill health) appears to be the question on self-assessed general health. Around a third (35%) of men and two-fifths (42%) of women who said that their health in the last 12 months had not been good, had consulted a doctor in the previous two weeks. A fifth (19%) of men and a quarter (24%) of women who said that their health had been fairly good had consulted a doctor, while only 9% of men and 14% of women who reported good health had done so.

Thus, for these three general health measures (all asked on the GHS in 1996), the expected relationship between poor health and doctor consultations was observed. However, of all three measures, the presence of a long-standing illness or disability showed the weakest association.

The ability of health measures to predict use of health services is clearly important

from a policy perspective. An instrument which discriminates well between those likely and those unlikely to use health services would clearly be of benefit for planning for future demand. It should be borne in mind, however, that the associations discussed here are not really predictive relationships in this sense. The use of services reported here refers to GP consultations *prior* to completion of the general health measures. It is equally likely that consulting a doctor affects how one subsequently rates one's general health rather than causality running in the opposite direction. In order to assess the ability of general health measures to *predict* future service use, a longitudinal design would be necessary.

## 5.4 Distributions self-reported general health by age and sex

A direct comparison between the prevalence of self-reported good health, as measured by the HSE on the one hand, and the GHS on the other, is not possible because of these differences in response scale format. 'Good' health is normally derived on the HSE by combining the categories 'very good' and 'good'; this category almost certainly includes some of those who would rate their health as 'fairly good' in response to the GHS. More than three-quarters of respondents to the 1993 and 1996 HSE, for example, rated their health as 'very good' or 'good', compared with between half and three-fifths of GHS respondents who chose the 'good' option. This in itself shows that how people rate their health depends crucially on how the question is framed. In addition, the GHS question specifies a time period, 'in the last 12 months', while the HSE does not.

All surveys, however, show a similar pattern of association between self-reported general health, age and sex. Men were consistently more likely than women to say that their health was good, although the differences were not significant on the two Health Surveys for England. Similarly, all the surveys showed a strong relationship between self-reported health and age, with the proportion of respondents who reported being in good health declining with age.

Differences between the proportions of men and women who said they had 'bad' or 'very bad' health on the HSE, or 'not good' health on the GHS, were small and not always statistically significant. Not surprisingly, however, the likelihood of reporting poor health increased with age on all surveys.

## 5.5 Self-reported long-standing illness, disability or infirmity by age and sex

All surveys showed a clear association between the prevalence of long-standing illness, disability or infirmity and age. Below the age of 55, between a fifth and two-fifths of respondents reported a chronic condition; among those aged 55 and over, between a half and two-thirds said that they had such an illness, disability or infirmity. The prevalence of long-standing illness as estimated by the HSE was higher than for the GHS; authors of previous HSE reports have suggested that respondents to a health survey may be more likely than those participating in a general survey to report an illness due to the subject matter of the questionnaire stimulating them to think more closely about all aspects of their health. On those surveys which included a question on limiting long-standing illness, between a tenth and a half of respondents said they had such a condition, the proportion increasing quite steeply with age. On the 1994 GHS, for example, 10% of men and women

aged 16-24 said they had a limiting illness, compared with 44% of men and 48% of women aged 75 and over.

## 5.6 Association between self-reported health and long-standing illness

All five surveys under consideration included questions on self-reported general health and self-reported long-standing illness although, as noted earlier, the wording of the question on general health and the response categories used varied across surveys. All surveys show an association between the two measures, with respondents reporting good health much less likely than those whose health is not good to report a long-standing illness or disability. Thus, for example, only 19% of respondents to the 1994 GHS with 'good' health said they had a chronic illness or disability, compared with 86% of those with 'not good' health. Similarly, 97% of respondents to the 1996 HSE with 'bad' or 'very bad' health reported a long-standing complaint, compared with 28% of those whose health was 'very good' or good'.

While this represents a high degree of congruity between these two instruments, it should be noted that a significant minority of respondents whose self-reported health was 'good', nevertheless said they had a chronic illness or disability, suggesting that the two questions are measuring somewhat different aspects of health. The key to this difference probably lies in the fact that the self-rated general health question contains an implicit valuation component while the long standing illness question does not. Therefore, while someone may report having a long standing illness the same person may nevertheless report their general health as being very good, because they may see the long standing illness as minor or unproblematic (e.g. minor skin complaints or correctable visual problems

## 6. Implications of the UK review on the Minimum European Health Module

The first question in the Minimum European Health Module is:

*How is your health in general?*
*Very good*
*Good*
*Fair*
*Bad*
*Very bad*

This question is very similar to that used in the national health survey in England. Therefore all the comments made above in terms of how it is administered, how it is understood and how it is answered are relevant.

The second question in the Minimum European Health Module is:

*Do you have any long-standing illness or health problem?*
*Yes*
*No*

There is no comparable question to this in the UK surveys. All the UK surveys add

the words "disability" or "infirmity" or give a reference period. The expression "health problem only occurs in the census questions, not in the large population sample surveys.

The third question in the Minimum European Health Module is:

*For at least, the past six months, have you been limited in activities people usually do because of a health problem?*
*Yes*
*No*

This question asks respondents to concentrate first on the last six months, then whether they have had a limitation in activity during this time, then whether it is an activity people usually do, and finally whether it is a result of a health problem. In the UK, questions on limiting long standing illness, in population surveys at least, tend to ask this as two questions: first to establish whether there is a problem and then to establish its consequences in terms of limitations in activity. In these questions and in the census question, which does put the concepts together, the focus is on the limitation in the respondents' own activities including in the census, the work that they can do, and not compared to what people usually do.

## 7. Conclusions

Subjecting any survey question to rigorous conceptual and methodological scrutiny is bound to throw up inconsistencies in interpretation and response. This is especially apparent when the task relates to asking people to rate or evaluate their health.

All the evidence from the UK experience suggests that at the most fundamental level it is important to have the same question wording across surveys and that help should be given in telling the person answering the question what we mean by health, perhaps by means of a preamble. This would help get over the problems of differential response by age, sex, and level of education. Also we should also be aware that when we compare data collected in different contexts, by subject and by proxy and with different modes of administration, that these have an effect on responses.

**References**

Bennett N. et al (1995) *Health Survey for England* 1993, London: HMSO

Bennett N, Jarvis L, Rowlands O, Singleton N & Haseldon, L. (1996) *Living in Britain: results from the 1994 General Household Survey,* London: HMSO

Blaxter M. (1987) *'Self-reported health' in The Health and Lifestyles Survey* London: Health Promotion Research Trust

Blaxter M. (1990) *Health and Lifestyle*s. (London: Routledge)

Bone M. (1995) *Trends in dependency among older people in England*, London: HMSO)

Bone M, Bebbington AC, Jagger C, Morgan K & Nicolaas G. (1995) *Health expectancy and its uses*. London: HMSO

Bowling,A. (1991/1997) *Measuring Health: a review of quality of life measurement scales.* Milton Keynes: Open University Press

Breeze E. et al (1994) *Health Survey for England 1992* London: HMSO

Bridgwood, A. (1993) *Baseline '93: health status and performance*

Bridgwood A & Malbon G. (1995) *Survey of the Physical Health of Prisoners 1994*. London: HMSO

Bush JW, Chen MM, Patrick,DL (1972) Social Indicators for Health Based on Function Status and Prognosis. *Proceedings of the American Statistical Association Social Statistics Section*: 71.

Cadman D, Boyle MH, Offord DR, Szatmari P, Rae-Grant NI, Crawford J, Byles J (1986) Chronic illness and functional limitation in Ontario children: findings of the Ontario Child Health Study *CMAJ* **135(7):**761-7

Crosnick J (1999) Survey Research *Annual Review of Psychology* **50** 537-567.

Department of Health (1992) *The Health of the Nation: a strategy for health in England* London: HMSO

Donovan, JL, Frankel SJ and Eyles JD (1993) Assessing the need for health status measures, *Journal of Epidemiology and Community Health*, **47**,158-162.

Foster K, Wilmot A & Dobbs, J. (1990) *General Household Survey 1988* London: HMSO

Franks P, Gold MR and Clancy CM (1996) *Use of care and subsequent mortality: the importance of gender*. Health Serv Res **Aug;31(3):**347-63.

Goddard E. (1990) *Measuring morbidity and some of the factors associated with it', in Health and Lifestyle surveys: towards a common approach*: report of a workshop held on 7 November 1989 organised by the HEA and OPCS. (London: HEA and OPCS)

Goddard E & Savage D.(1994*) General Household Survey: People aged 65 and over*: GHS No. 22 Supplement A. London: HMSO

Grand A, Grosclaude P, Bocquet H, Pous J, Albarede (1990) Disability, psychosocial factors and mortality among the elderly in a rural French population *Journal of Clinical Epidemiology* **43(8**):773-82.

Kind, P (1995) Measuring the reliability of individual assessments of the life quality associated with health states. *Survey Methods Centre Newsletter* Vol. 15 No 2.

Lawton MP & Brody EM (1969) Assessment of older people: self-maintaining and instrumental activities of daily living, *Gerontologist*, **9**, 179-186.

Long A (1993) *General Health Measures - an introduction to multidimensional profile*s, Paper prepared for the sub-group of the Chief.125 Medical Officers' Health of the Nation Survey.

Sundquist J & Johansson SE (1997) Indicators of socio-economic position and their relation to mortality in Sweden. *Social Science and Medicine* **45(12)**, 1757-66

*The Health and Lifestyle Survey* (1987) London: Health Promotion Research Trust

Thomas M, Goddard E, Hickman M and Hunter P (1994) *General Household Survey 1992* London: HMSO

Thomas R & Purdon S (1994) *Survey Methods Centre Newsletter*, **14( 2)** National Centre for Social Research 1994

White A (1995) *Measuring subjective health statu*s. (Unpublished paper: Social Survey Division).

White A et al (1993) *Health Survey for England 1991* (London: HMSO)