

Chapter 3

Measuring subjective well-being

Introduction

This chapter aims to present best practice in measuring subjective well-being. It covers both the range of concepts to be measured and the best approaches for measuring them. This includes considering issues of sample design, survey design, data processing and coding and questionnaire design. In particular, the chapter presents a single primary measure intended to be collected consistently across countries, as well as a small group of core measures that it is desirable for data producers to collect where possible. Beyond this core suite of measures, the chapter provides more general advice to support data producers interested in identifying and measuring aspects of subjective well-being that will meet their particular research or policy needs, as well as a range of question modules relating to different aspects of subjective well-being.

The chapter has four substantive sections. The first section focuses on issues associated with planning the measurement of subjective well-being. This includes addressing the relationship between the intended policy or research use of the data and the appropriate measurement objectives. A crucial element of deciding what to measure is thinking about the relevant co-variates to be collected alongside the measures of subjective well-being to support analysis and interpretation. Section 2 of the chapter addresses survey and sample design issues. These include the choice of survey vehicle, sample design, target population, collection period and survey frequency. The third section of the chapter looks at questionnaire design, which includes both issues of question order and questionnaire structure, as well as the precise question wording. A key element of this section is the inclusion of model questions on the different aspects of subjective well-being. The final section focuses on survey implementation. This includes brief guidelines on interviewer training as well as data processing. The chapter does not, however, cover issues relating to the use and analysis of subjective well-being data in detail. These are addressed in Chapter 4 (*Output and analysis of subjective well-being measures*). A recurring issue throughout the chapter is the lack of standards in the methods used to gather supporting information for subjective well-being analyses, such as information about child well-being, attitudes, personality, etc. These issues are deemed to be beyond the scope of this chapter, but remain an important gap.

Core measures of subjective well-being

Core measures of subjective well-being are those for which international comparability is the highest priority. These are measures for which there is the most evidence of validity and relevance, for which the results are best understood, and for which policy uses are most developed. Although the guidelines are intended to support producers of measures of subjective well-being rather than being overly prescriptive, the core measures proposed here are quite specific in content and collection method.

The core measures outlined in this chapter consist of five questions. The first is a primary measure and is intended to be collected consistently across countries. The primary measure should be regarded as the highest priority for national statistical agencies and should be the first question included in surveys where the measurement of subjective well-being is considered. The additional three affect questions in the core are also important and should be collected where possible. However, it is recognised that not all national statistical offices will be able to collect these measures in their core surveys. Finally, an experimental eudaimonic question is attached, picking up the element of eudaimonia for which there is the most evidence of relevance.

Beyond articulating a suite of core measures, the main goal of this chapter is to provide general advice to data providers. In particular, the chapter is intended to support national statistical agencies and other data providers in the process of deciding what to measure and how to implement the measurement process most effectively. While models are provided for specific questions, the chapter aims to provide options and advice rather than directions.

1. What to measure? Planning the measurement of subjective well-being

This section looks at the planning stage of a measurement project. It is concerned with what concepts to measure and how these concepts affect decisions about the final output and analysis. In doing so, the chapter touches on the issues that are the main focus of Chapter 4 (*Output and analysis of subjective well-being measures*). However, where Chapter 4 focuses on how to analyse, interpret and present subjective well-being data, the discussion here is limited to how user needs determine what information to collect.

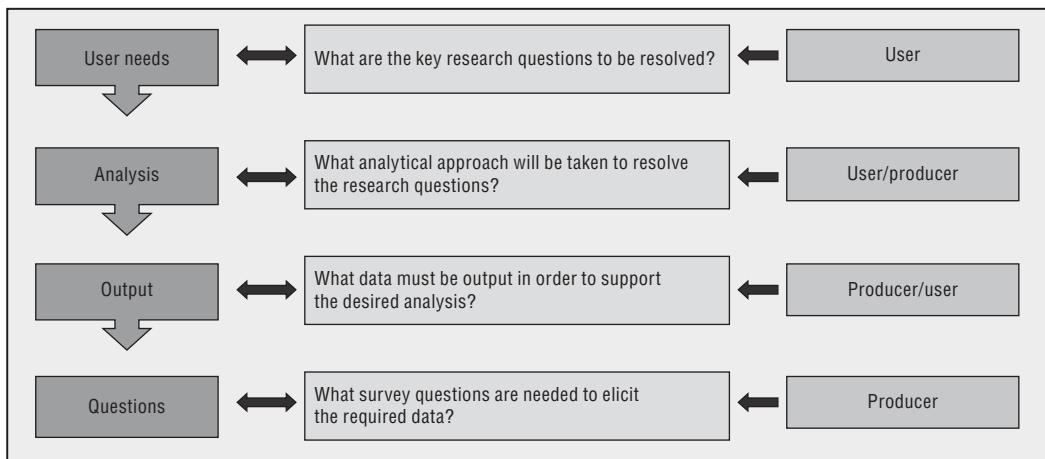
The initial planning stage of a project to measure subjective well-being – or indeed any statistical programme – is critically important. All subsequent decisions will be heavily influenced by choices made early on about the research objectives of the project. Clarity about objectives is thus crucial.

Decisions about what to measure should always be grounded in a clear understanding of user needs. Only if the needs for the data are clearly understood is it possible to make informed decisions about the information that should be collected to meet these needs. Understanding user needs is not, however, straight-forward. A relatively simple research question can be approached in a range of different ways using different methodologies. For example, one can understand what motivates behaviour both by asking people directly what they would do in a given set of circumstances or by collecting information on the course of action people take and on the circumstances they face.¹ Each methodological approach has its own strengths and weaknesses, and will have different implications for measurement. Having an analytical model can assist in thinking in a structured way about how user needs relate to specific decisions about what data to collect.

Figure 3.1 presents a simple model relating user needs to the specific survey questions used to collect information. The model is intended to provide a framework for thinking about the various stages involved in moving from a user's information needs to specific questions that can be included in a survey.

The first column of Figure 3.1 identifies the four stages involved in going from user needs to specific survey content. Conceptually, these stages involve working back through the process of collecting the data and using them in decision-making in reverse order. Column 2 articulates the key issues to be addressed in each stage of the project in order to make well-informed decisions about the most appropriate measures. Finally, the third

Figure 3.1. The planning process: From user needs to survey questions



column indicates which party has the lead role in making decisions. Although the process of going from user needs to survey content is fundamentally collaborative in nature, there are stages in the process when users can be expected to play a more important role than data providers, as well as cases where the reverse is true.

In practice, the process of working through these four stages is likely to be less clearly defined than Figure 3.1 suggests. In some cases, where the level of analysis required is relatively simple, the analysis and output stages of the process can merge into each other. Users will sometimes have clear views about the best measures to support the analysis that they would like to undertake, and it would be foolish to ignore these views in instances where a sophisticated user has a better understanding of the issue at hand than a data provider with little experience of measuring subjective well-being. Similarly, data producers may suggest possibilities that will result in changes in user needs or in the analytical approach taken to address them.

User needs

Understanding user needs involves understanding the key policy and research questions that the user is trying to address. While it is not possible in this chapter to give a full discussion of all possible user needs for subjective well-being data, some general questions can be articulated:

- Are the user needs related to one of the general policy uses for subjective well-being data described in Chapter 1?
- What are the policy questions?
- Is the subjective wellbeing content being proposed appropriate to respond to the policy questions? Is the content proposed sensitive to monitoring changes over time or between population groups?
- What population groups are of interest to the user? For example, is the focus on international comparisons (making countries the key unit of analysis), the same population at different points in time (for time series analysis), or different sub-groups of the same population (such as age, sex, location or ethnicity)? This will have implications both for sampling and for the types of measure that are most appropriate.

In the case of cross-country comparisons, measures with good cross-cultural reliability will be most important, while for analysis of groups within a country low respondent burden may be a more important consideration in order to allow a larger sample size.

- Does the user's interest lie in comparing outcomes of different groups or in understanding the relationship between different aspects of subjective well-being? In the first case, a relatively narrow range of subjective well-being measures may suffice, while in the latter case more detail on a range of co-variates is likely to be necessary.
- Is the user's primary interest in overall subjective well-being (captured by summary measures of life evaluation, affect or eudaimonic well-being) or in a specific dimension of subjective well-being (such as satisfaction with income or satisfaction with work/life balance)? Are other measures of well-being more appropriate?
- What are the frequency requirements of the users to monitor changes over time?
- What within-country comparisons are required, such as geographic level?

A thorough understanding of user needs should allow the identification of one or more clear research questions that the project should address.

Analysis

Understanding the overall research question is not sufficient to make meaningful decisions about the type of output or the most appropriate measures to use. A given research question may be addressed in more than one valid way. It is therefore essential to understand how the specific research question can be answered:

- Will the analytical approach be primarily descriptive, or will it require more sophisticated statistical techniques (e.g. regression, factor analysis, etc.)?
- What contextual and other variables are required to answer the research question? If the research question simply involves identifying differences between specific population groups in terms of a small set of key outcomes, the range of relevant co-variates may be relatively limited. However, if the research question is focused on understanding what drives group differences in subjective well-being or on examining the joint distribution between subjective well-being and other dimensions of well-being, the range of co-variates is likely to be significantly broader.
- What level of accuracy is required to produce meaningful results from the proposed analysis? This will have implications for sample size and sampling strategy. For example, if obtaining precise estimates for small population sub-groups is a priority, then oversampling of these groups may be necessary.

After considering the proposed analytical strategy, it should be possible to articulate how the research questions can be answered in quite specific terms. This will form the basis for evaluating what data needs to be output to support the desired analysis.

Output

Output refers to the statistical measures released by a national statistical agency or by another data producer. These can take the form of tables of aggregate data such as average results by group, micro-data files, interactive data cubes or other forms. The key distinction between output and analysis is that output does not, in itself, answer a research question. Instead, it provides the base information that is analysed in order to produce the answer.

In some cases, the answer may be directly evident from the output, requiring only limited interpretation, comment and caveats, while in other cases extensive statistical analysis may be required.

Because output forms the basis for all subsequent analysis, it provides the key link between specific survey questions and the use of the data in analysis. The required output must therefore be clearly specified before appropriate questions can be designed. Some key issues to consider when specifying the desired output for information on subjective well-being include:

- Will the analysis require tabular output of averages or proportions, or is micro-data needed? Simple comparisons of how different population groups compare with each other can be accomplished via tabular output, but understanding the drivers of such differences will require a much finer level of detail.
- Will the analytic techniques used treat the data as ordinal or cardinal? This makes little difference if micro-data is required (since users can decide for themselves), but will influence how summary measures of central tendency and distribution are presented in tabular form. Information on a cardinal variable² can be presented via techniques that add and average scores (e.g. mean, standard deviation), while ordinal data will need to be reported by category.
- How important is it to present measures of the central tendency of the data (e.g. mean, median, mode) as opposed to the dispersal (e.g. standard deviation) or full distribution of the data (e.g. proportion responding by category)?

In planning a measurement exercise, the aim should be to clearly specify the desired output, and the data items required to produce this, before considering question design. This will involve, at a minimum, defining the measures to be used and the break-downs and cross-classifications required. In many cases, particularly if multivariate analysis is proposed, more detailed information may be required.

Questionnaire design

Once a clear set of outputs has been identified based on the analysis required to meet user needs, it will be possible to make specific decisions about survey design, including the most appropriate survey vehicle, collection period, units of measurement and questionnaire design. These decisions should flow logically from the process of working down from user needs through analysis and output. The remainder of this chapter sets out a strategy for the measurement of subjective well-being. This includes both specific proposals for how a national statistical agency might approach the measurement of subjective well-being and more general information that can be used in a wider range of circumstances.

What other information should be collected: Co-variates and analytical variables

All potential uses of subjective well-being data require some understanding of how subjective well-being varies with respect to other variables. This applies whether the goal is understanding the drivers of subjective well-being – which requires understanding the causes of change – or where the main purpose is monitoring well-being over time and across countries – which requires understanding changes in demographics in order to understand a given change is due to changes in average levels or in the ratios of different population groups in society. It is therefore imperative to consider not only how best to measure subjective well-being *per se*, but also what other measures should be collected alongside measures of subjective well-being for analytical purposes.

A need for additional information to aid in interpreting and analysing results is not unique to subjective well-being. Most statistical measures are collected alongside, at the least, basic demographic data. Demographics matter to subjective well-being measures just as much as they do to labour market statistics. There are pronounced differences in average levels of subjective well-being across a range of different demographic groups, including based on gender, age and migration status (Dolan, Peasgood and White, 2008). For example, one of the best-known features of life satisfaction data is the “U-shaped” relationship between age and average life satisfaction (Blanchflower and Oswald, 2008). Similarly, there are differences between men and women in life satisfaction and affect measures that are not fully accounted for, even when controlling for income and education (Boarini, Comola et al., 2012).

Beyond demographics, subjective well-being affects and is affected by a wide range of different factors. Material conditions (e.g. income, consumption, wealth) affect subjective well-being (Dolan, Peasgood and White, 2008), but so do factors relating to quality of life. Health status, unemployment, social contact and safety all impact on life satisfaction in important ways (Boarini, Comola et al., 2012). In the context of affect data collected through time-use diaries, it is possible to collect information on an additional range of variables, such as the activity associated with a particular affective state.

Finally, there is a strong case for collecting some additional psychological variables alongside measures of subjective well-being. These include measures of personality type, expectations about the future and views about past experiences. Such measures may help to disentangle fixed effects at the personal level when it is not possible to collect panel data.

The precise range of co-variates to collect alongside measures of subjective well-being will vary with the specific aspect of subjective well-being that is of interest and with the research question being examined. Despite this, it is possible to present some general guidelines on the most important information that should be collected alongside measures of subjective well-being.

Most of the co-variates described below are regularly collected by national statistical agencies, and international standards for their collection do exist. No attempt is made here to specify the details of how these variables should be collected, and it is assumed that existing standards apply. This is not true for a few measures, such as those related to personality, trust and belonging. In these cases some general guidelines are provided. However, as many of these issues (such as the measurement of personality traits) are complex topics in their own right, this chapter does not provide detailed recommendations.

Demographics

Demographic variables cover the basic concepts used to describe the population being measured and to allow the analysis of how outcomes vary by population sub-group. As such, including a range of demographic measures in any attempt to measure subjective well-being is of utmost importance, in particular the following measures:

- **Age.** The age of the respondent, in single years if possible. Age bands, while allowing for some cross-classification, are less desirable both because they allow less flexibility with respect to the groups examined, and because they do not facilitate analysis of age as a continuous variable.
- **Sex or gender.** The sex or gender of the respondent.

- **Marital status.** *The legal marital status* of the respondent, including whether the respondent is widowed, divorced or separated and the **social marital status** of the respondent, including whether the respondent is living as married even if not legally married.
- **Family type.** Family type refers to a classification of the respondent's family unit, including whether they are single or living with a partner and whether children are present.
- **Children.** The number and age of children in the respondent's family unit, along with relationship to the respondent.
- **Household size.** The number of people living in the respondent's household. Household size is a distinct concept from family size, as more than one family unit can live in a dwelling. Household size is essential to allow an understanding of the impact of household income on subjective well-being.
- **Geographic information.** While privacy concerns may prevent the release of detailed geographical information relating to the respondent, estimates can be disaggregated by some broad level geographic regions such as urban and rural, capital city, states/provinces, etc. Geo-coding allows for merging with other datasets also containing geo-codes, such as environmental data.

In addition to the demographic measures identified above, which can be considered essential, a number of additional demographic variables may also be desirable to include. The precise relevance of these may, however, vary depending on national circumstances and the research priorities being considered:

- **Migration status/country of birth/year of arrival.** Migration status, such as permanent residence, citizenship, etc., and/or country of birth of the respondent.
- **Ethnic identification.** The ethnic identity or identities of the respondent may be of high policy importance in ethnically diverse societies.
- **Language.** The primary language of the respondent. It may also be desirable, in some circumstances, to collect information on other languages spoken. Proficiency in the main language of the country in which the survey is taking place may also be important for some purposes.
- **Urbanisation.** The classification of the area in which the respondent lives in terms of degree of urbanisation.

Material conditions

The term "material conditions" is used here to cover income, wealth and consumption, as well as other aspects of the material living circumstances of the respondent. Much of the interest in measures of subjective well-being has been focused on the relationship between the material conditions of the respondent and their level of subjective well-being. Traditionally, income has been a major focus. The so-called "Easterlin paradox", described by Richard Easterlin (1974), notes that a rise in household income leads to higher subjective well-being for individuals in the household, but that a rise in average incomes for a country appears not to give rise to a corresponding increase in the country's average subjective well-being. Understanding this apparent paradox is important given the degree to which much of policy is focused on economic growth. There are a number of possible explanations for the paradox, but one is lack of high-quality data linking measures of subjective well-being to the household income of the respondent.

Including income measures in surveys of subjective well-being is essential, and should be considered as important as basic demographic variables. It is important that the income measures used are of high quality, and ideally relate to a relatively long period of time (such as a year). Measures of subjective well-being are likely to be more sensitive to changes in long-term income levels than to short-term fluctuations in weekly or monthly income. The relationship between income (at both aggregate average and individual level) and subjective well-being is log linear (i.e. shows diminishing returns) for measures of life evaluation (Sacks, Stevenson and Wolfers, 2010), and the relationship between income and positive affect in a US sample flattens off entirely (Kahneman and Deaton, 2010). This suggests that, if income is collected in bands rather than as a continuous variable, the bands will need to be narrower in currency terms (but constant in proportionate terms) at lower levels of income than at higher levels:

- **Income.** Household income is of greater importance than individual income, since it is household income that drives living standards and consumption possibilities (Stiglitz, Sen and Fitoussi, 2009). However, where it is possible, capturing information on individual and household income is also of interest. In both cases, it is desirable to have information on net (post-tax and transfers) income as well as gross income, and equivalised household income should also be available (to take account of household size and composition).³ Space permitting, information on the source of the income (wages and salary, capital and investment earnings, government transfers) may also be of interest.
- **Expenditure and consumption.** Income flows are a relatively limited measure for the actual level of consumption that a household can support. People may draw on previously accumulated assets or run up debt to smooth consumption over time. Thus, for exploring the relationship between consumption and subjective well-being it is desirable to have measures of expenditure and/or access to specific goods and services. Such measures may perhaps allow for separating living standards (consumption) from status and rank effects (income). Questions on financial stress or the ability to access a given amount of money in an emergency may also be valuable for analytical purposes.
- **Deprivation.** Because of the difficulty and costs associated with collecting high-quality information on expenditure, surveys are often under pressure with respect to space available for additional questions. Measures of material deprivation provide an alternative to detailed expenditure data as a way of assessing adequacy of consumption.⁴ Because such measures impose a much smaller respondent burden than collecting detailed expenditure data, material deprivation can be a useful way to collect information on consumption so as to inform analysis of the relationship between low material living standards and subjective well-being.
- **Housing quality.** Housing quality is an important element of the material conditions in which people live, and there is evidence that housing conditions affect subjective well-being (Oswald et al., 2003). Where the impact of material conditions on subjective well-being is a significant part of the research question, collecting information on housing quality will be important. Key dimensions of housing quality to be collected might include number of rooms, housing costs and specific aspects of quality such as dampness or noise.⁵ Data on the number of rooms can be used alongside household composition information to assess overcrowding, while housing costs may be used to measure income net of housing costs.

Measures of subjective well-being and household economic statistics are complements rather than substitutes. Understanding the relationship between subjective well-being and economic variables such as income and expenditure is an important rationale for collecting subjective measures in the first place. Collecting data on the economic and social determinants of subjective well-being can then permit their relative importance to be established. Careful thought should be given to facilitating this analysis, not only by including measures of income and wealth in surveys focused on well-being, but also – space and cost permitting – by placing measures of subjective well-being in household income and expenditure surveys and in employment-related surveys.

Quality of life

Quality of life is a broad term covering those aspects of overall well-being that are not captured only by material conditions. The Sen/Stiglitz/Fitoussi commission described quality of life as comprising “the full range of factors that influences what we value in living, reaching beyond its material side” (Stiglitz, Sen, Fitoussi, 2009). Information on these factors is important when measuring subjective well-being, because they are strongly correlated with subjective well-being even after controlling for income and demographic factors (Helliwell, 2008; Dolan, Peasgood and White, 2008; Boarini et al., 2012). In fact, it is likely that much of the simple correlation between individual income and subjective well-being occurs only because income is itself correlated with some measures of quality of life. Evidence of this can be found in the fact that the size of the coefficient on income decreases sharply when quality-of-life measures are included in a regression model (Boarini et al., 2012). This suggests either that higher income is one of the channels through which the quality of life has been improved or that there are other factors that improve both incomes and quality-of-life measures.⁶

Measurement of some aspects of quality of life is less developed than in the case for income, and it is therefore not possible to point to internationally accepted standards for some areas of quality of life that could be collected alongside measures of subjective well-being. In addition, the range of concepts covered by the notion of “quality of life” is so broad that an attempt to be comprehensive in identifying potential co-variates of subjective well-being would be prohibitively large. Nonetheless, it is possible to identify some of the key concepts for which measures would be desirable:

- **Employment status** – employment status is known to have a large influence on subjective well-being, with unemployment in particular associated with a strong negative impact on measures of life satisfaction (Winkelmann and Winkelmann, 1998) and affect (Boarini et al., 2012). There is also good evidence that measures of satisfaction with work predict subsequent labour market behaviour (Clark, Georgellis and Sanfrey, 1998; Card et al., 2010).
- **Health status** – both physical and mental health are correlated with measures of subjective well-being (Dolan, Peasgood and White, 2008), and there is evidence that changes in disability status cause changes in life satisfaction (Lucas, 2007). Although health status is complex to measure in household surveys, there is a large pool of well-developed measures available, such as the health state descriptions from the World Health Survey (WHO, 2012), or more specialised question modules, such as the GHQ-12 for mental health (Goldberg et al., 1978). Since 2004, a joint work programme of the UNECE, WHO and Eurostat has been engaged in developing common core measures of health status for inclusion in surveys (the Budapest Initiative). When these measures are finalised and become commonly available, they will form a suitable basis for monitoring health outcomes in general population surveys (UNECE, 2009).

- **Work/life balance** – there is significant evidence that aspects of work/life balance impact on subjective well-being, in particular commuting (Frey and Stutzer, 2008; Kahneman and Kruger, 2006), and time spent caring for others (Kahneman and Krueger, 2006). Relevant measures include hours worked (paid and unpaid), leisure time, perceived time crunch as well as information on how time is used.
- **Education and skills** – education and skills have obvious interest both as variables for cross-classification and because there is good evidence that education is associated with subjective well-being at a bivariate level (Blanchflower and Oswald, 2011; Helliwell, 2008). In analyses that control for additional factors, such as income and social trust, the correlation falls, suggesting that education may affect subjective well-being partly through its impact on other intermediate variables. The highest qualification attained and years of schooling may be used to measure education and skills. There may also be some value in collecting information on current engagement with education.
- **Social connections** – social contact is one of the most important drivers of subjective well-being, as it has a large impact both on life evaluations and on affect (Helliwell and Wang, 2011b; Kahneman and Krueger, 2006; Boarini et al., 2012). Although only some elements can be measured well in the context of general household surveys, measures of human contact, such as frequency of contact with friends and family, volunteering activity, and experience of loneliness, should also be collected where possible.
- **Civic engagement and governance** – generalised trust in others as well as more domain-specific measures of neighbourhood and workplace trust are crucial factors when accounting for variation in subjective well-being (Helliwell and Wang, 2011b) and should be collected. More generally, corruption and democratic participation have been shown to affect life evaluations (Frey and Stutzer, 2000), and measures of these concepts are of interest.
- **Environmental quality** – environmental quality is inherently a geographic phenomenon, and integrating datasets on environmental quality with household level data on life satisfaction is costly. Nonetheless, there is some evidence that noise pollution (Weinhold, 2008) and air pollution (Dolan, Peasgood and White, 2008) have a significant negative impact on life satisfaction. Silva, De Keulenaer and Johnstone (2012) also show that subjective satisfaction with air pollution is correlated with actual air pollution. To understand the impact of environmental quality on subjective well-being, it will be important to link actual environmental conditions to reported subjective well-being via geo-coding. Particular issues of concern include air quality and the extent of local green space.
- **Personal security** – security is important to subjective well-being. This is reflected in correlations between experience of victimisation and subjective well-being at the individual level (Boarini et al., 2012), as well as by subjective perceptions of safety. For example, living in an unsafe or deprived area is associated with a lower level of life satisfaction, after controlling for one's own income (Dolan, Peasgood and White, 2008; Balestra and Sultan, 2012). Measures of experience of victimisation and perceived safety should both be collected, as is already done in standard victimisation surveys, because subjective well-being appears to be more strongly affected by perceived crime rates than by actual rates (Helliwell and Wang, 2011b).

Psychological measures

Personality type has a significant impact on how people respond to questions on subjective well-being (Diener, Oishi and Lucas, 2003; Gutiérrez et al., 2005). While this will not normally bias results if personality is uncorrelated with the main variables used in the analysis of subjective well-being, it is desirable to control for it if possible. In panel surveys, personality type can be controlled for, to some extent, using individual fixed effects. In cross-sectional household surveys this is not possible. One approach is to incorporate measures of personality type such as the standard instrument for the Five Factor Model (Costa and McCrae, 1992) in surveys focusing on subjective well-being.⁷ Although such measures are rarely used in official statistics, this is an area that may warrant further investigation.

Aspirations and expectations, which form part of the *frame of reference*⁸ that individuals use when evaluating their lives or reporting their feelings, are also of interest when analysing data on subjective well-being. There is good evidence that life evaluations are affected by aspirations (Kahneman, in Kahneman, Diener and Schwarz, 1999), and it has been suggested that differing aspirations may account for some cultural differences in life evaluations (Diener, Oishi and Lucas, 2003). There is less evidence with respect to how aspirations impact on measures of affect or eudaimonia. Nonetheless, information on people's aspirations and expectations would be useful for investigating this relationship. There are no standard approaches to measuring aspirations and expectations, so it is not possible to be specific as to best practice in approaching measures of this sort. However, this area is one where further research would be of high value.

Time-use diaries

Although all of the measures identified as relevant to household surveys remain equally relevant to time-use surveys, the use of time diaries opens the way to collect additional co-variates not possible in standard household surveys. This is particularly the case where information on aspects of subjective well-being, such as affect, is collected in the diary itself. Several implications of collecting subjective well-being measures via time-use diaries are worth noting specifically:

- **Activity classification** – the standard activity classifications (Eurostat, 2004) are central to time-use diaries, and are of primary importance in interpreting information on subjective well-being.
- **With whom** – there is evidence that whether an activity is performed alone or with others, and the respondent's relationship to the others, are important to subjective well-being (Kahneman and Krueger, 2006). This reinforces the value of collecting information on "with whom" an activity took place where subjective well-being measures are collected.
- **Location** – the location of the activity in question and the impact that this has on subjective well-being is little researched. However, such information potentially brings useful context to analysis, and should be collected alongside subjective well-being and activity classification where possible.
- **For whom** – permits disaggregation of activity data by the *purpose* of the activity. This allows for analysing activities done for voluntary organisations, persons with a disability, family and non-family members, which may all have useful analytical possibilities for subjective well-being.

2. Survey and sample design

One important distinction between measures of subjective well-being and many of the measures typically included in official statistics is that subjective well-being measures will almost invariably need to be collected through sample surveys. In contrast to many economic or population statistics, there is generally no administrative database that would produce subjective information without, in effect, incorporating survey questions in an administrative process.⁹ Thus, issues relating to survey and sample design are fundamental to producing trustworthy and reliable measures of subjective well-being.

It is not the role of this chapter to provide detailed guidelines on sample frames and sample design. These are specialist areas in their own right, and excellent guides exist for data producers who are seeking advice on these technical aspects of data collection (UN, 1986). However, in survey design, as in other aspects of design, form should follow function. The fact that subjective well-being is the goal of measurement has implications for survey design. This section discusses some of the most significant considerations for the measurement of subjective well-being with respect to the target population, to when and how frequently the data should be collected, to what collection mode should be used, and to identifying the most appropriate survey vehicle.

Target population

The target population for a survey describes the complete set of units to be studied. A sample survey will generally attempt to achieve a representative sample of the target population. However, the target population may be more detailed than the total population from which the sample is drawn. It may also specify sub-populations that the survey describes. For example, the total population might be all persons aged 15 and over living in private dwellings in a specified area. However, the target population might also specify males and females as sub-populations of interest, requiring the sampling frame to accommodate distinct analysis of these two groups. More generally, sub-groups are often defined by such characteristics as age, gender, ethnicity, employment status or migrant status.

Some surveys with the household as the unit of measure rely on a single respondent (such as the head of household) to provide responses for the household as a whole. This cannot be used for measures of subjective well-being, since the cognitive process of evaluating and responding with respect to one's own subjective well-being is very different to that of providing an estimate of another householder's state of mind. Responses to questions on subjective well-being are inherently personal, and consequently the unit of measure for subjective well-being must be the individual. This implies that the sampling frame must produce a representative sample of individuals or households as if all individuals are personally interviewed. While this will typically not be an issue for surveys where the individual is the primary unit of analysis, some household surveys may require an additional set of individual weights to derive individual estimates. Surveys where the response is on the basis of "any responsible adult" will in particular be problematic in this regard.

The target age group for measures of subjective well-being will vary with respect to the goals of the research programme. For example, in the context of research on retirement income policies, it may be appropriate to limit the target population to persons aged 65 or older. In general, however, measures of subjective well-being would usually be collected for all the adult population (aged 15 years and older).

Children

Child well-being is a significant policy issue, both because child well-being has an important impact on later adult outcomes (OECD, 2009), and because the well-being of children is important in its own right. In the analysis of child outcomes, parental or household responses are often used as proxies for the situation of children. In many cases this is a reasonable assumption. Household income, for example, is obviously much more relevant to the living circumstances of a child than would be the income earned by the child themselves. However, this is not true for measures of subjective well-being. If there is a policy interest in the subjective well-being of children, providing data on this will necessarily involve obtaining responses from children. While parents may be able to provide a second-person estimate of the well-being of their child, this is a conceptually different construct from the child's own subjective well-being.

National statistical agencies rarely collect information from respondents younger than 15 or 18 years-old. This reflects both legal issues and concern about the acceptability of such practices to respondents. These are real and significant issues, and must be treated accordingly. However, it is important to note that the available evidence suggests that children are capable of responding effectively to subjective well-being questions from as young as age 11 with respect to measures of life evaluation and affective state (UNICEF, 2007).

While there may be ethical issues with interviewing young children, it is also important to consider the implications of not including the voices of children when measuring subjective well-being. As the focus of this chapter is on general population surveys, questions focused specifically at young children are therefore not provided. However, this remains a significant gap that future work should address.

People not living in private households

One population group that may be of high policy interest, but which is not typically covered in household surveys, is people not living in private households. This group includes people living in institutions, including prisons, hospitals or residential care facilities, as well as people with no fixed residence, such as the homeless. These groups raise two issues with respect to the measurement of subjective well-being. The first problem is common to all attempts to collect statistical information on such groups – that such population groups tend to be excluded from standard household survey sample frames. This means that, at a minimum, specific data collection efforts will be required based on a sample frame designed to cover the relevant institutions. In some cases, such as for the homeless, it may be difficult to develop any statistically representative sampling approach at all.

A more significant challenge faced in the measurement of subjective well-being is that many of the people in the relevant groups may not be able to respond on their own behalf. This is particularly the case for people institutionalised for health-related reasons that affect mental functioning (including people with some mental illnesses, or with physical illnesses limiting the ability to communicate, and the very old). In these cases it is not possible to collect information on a person's subjective well-being. Proxy responses, which might be appropriate for some types of data (income, marital status, age), are not valid with respect to subjective well-being.

Frequency and duration of enumeration

The frequency with which data is collected typically involves a trade-off between survey goals and available resources. All other things being equal, more frequent collection of data will improve the timeliness of estimates available to analysts and policy-makers, and will make it easier to discern trends in the data over time. More frequent enumeration, however, is more costly both in terms of the resources involved in conducting the data collection and in terms of the burden placed upon respondents. It is therefore important that decisions around the frequency of data collection are made with a clear view to the relationship between the timeliness and frequency of the data produced and the goals of the data collection exercise.

It is not possible to provide specific guidelines for how frequently measures of subjective well-being should be collected covering every contingency, since the range of possible data uses is large and the frequency at which data are needed will vary depending on the intended use and on the type of measure in question. However, some general advice can be provided. Aggregate measures of subjective well-being generally tend to change only slowly over time. This reflects the relatively slow movements in most of the social outcomes that affect subjective well-being and the fact that many changes only impact on a small proportion of the population. For example, unemployment – which is associated with a change of between 0.7 and 1 on a 0 to 10 scale (Winkelmann and Winkelmann, 1998; Lucas et al., 2004) – typically affects between three and 10% of the adult population. Thus, even a large shift in the unemployment rate – say, an increase of 5 percentage points – will translate only into a small change in measures of subjective well-being (Deaton, 2011).

The relatively slow rate of change in measures of subjective well-being might appear to suggest that such measures do not need to be collected frequently. However, the small absolute size of changes in subjective well-being also means that standard errors tend to be large relative to observed changes. A number of observations are therefore needed to distinguish between a trend over time and noise in the data. Box 3.1 illustrates this point. For this reason, despite (or indeed, because of) the relatively slow rate of change in subjective well-being data, it is desirable that measures are collected on a regular and timely basis. For the most important measure used in monitoring well-being, an annual time series should be regarded as the essential minimum in terms of frequency of enumeration. More frequent monthly or weekly data is, however, likely to be of lower value (Deaton, 2011). (It should be pointed out that frequent, or rolling sample, surveys increase the possibilities for identifying the causal impacts of other factors whose dates can be identified. It was only the daily frequency of observations that made it so easy to discover and eliminate the question-order effects in Deaton (2011)).

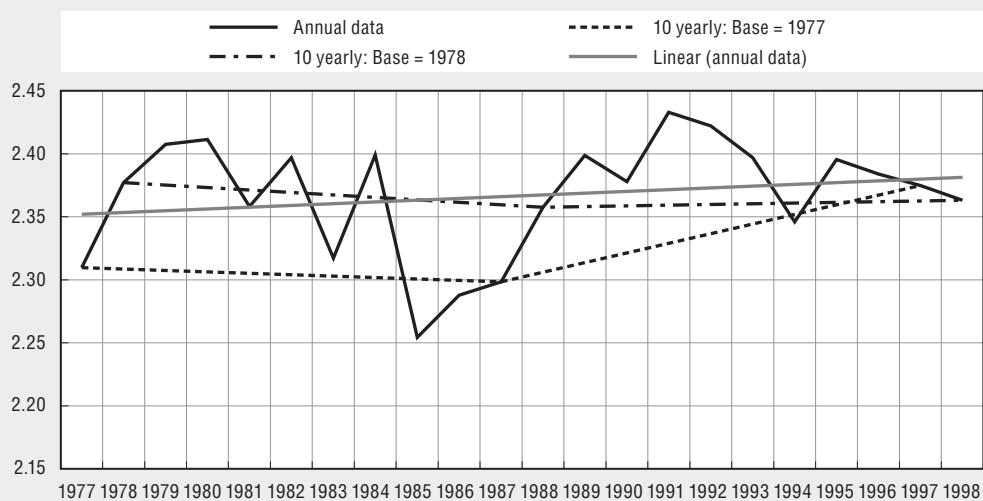
Duration of enumeration

The duration of the enumeration period (i.e. the period of time over which information is collected) is immensely important for measures of subjective well-being. Unlike measures of educational attainment or marital status, for which it does not usually matter at what point during the year the data are collected, the precise timing of the collection period can have a significant impact on measured subjective well-being (Deaton, 2011). For example, measures of positive affect are higher on weekends and holidays than on week days (Helliwell and Wang, 2011a; Deaton, 2011).

**Box 3.1. Identifying trends in time series of subjective well-being:
Implications for frequency of measurement**

It might seem logical that, if measures of subjective well-being change only slowly over time, they will need to be measured only infrequently. Figure 3.2 illustrates why this assumption does not necessarily hold, by plotting the time series changes of life satisfaction in the Netherlands using data collected in the Eurobarometer. Because the Eurobarometer has a relatively small sample size, the error is large relative to the size of changes over time. Between 1977 and 1997, life satisfaction in the Netherlands was largely static. Despite large fluctuations in the annual time series, the trend line for this 20-year period is almost flat.

Figure 3.2. Life satisfaction in the Netherlands 1977-97: Eurobarometer



Source: Eurobarometer.

Two additional lines included in Figure 3.1 illustrate the picture that would emerge if information had been collected only every ten years. If the base year were 1977, then a 10-yearly data collection would show a slight decline in life satisfaction between 1977 and 1987, followed by a substantial increase from 1987 to 1997. The net effect over twenty years is seen to be a significant increase in life satisfaction. Changing the base year to 1978 produces a different pattern, with life satisfaction declining from 1978 to 1988, before flattening out from 1988 to 1998. The overall effect over 20 years this time would be a slight decrease in life satisfaction. None of these 10-yearly patterns reproduces the pattern highlighted by annual data.

The fact of being sensitive to the point in time at which they are collected is not unique to measures of subjective well-being. Many core labour market statistics, for example, have a pronounced seasonality, and published statistics usually adjust for this. However, such adjustments require collecting data over the course of a whole year in order to produce the information required for seasonal adjustments.

The fact that subjective well-being does vary over the year suggests strongly that a long enumeration period is desirable. Ideally, enumeration would take place over a full year, and would include all days of the week, including holidays. This would ensure that

measures of subjective well-being provide an accurate picture of subjective well-being across the whole year. Where a year-long enumeration period is not possible, enumeration should be spread proportionately over all the days of the week as far as is possible. All days of the week need to be covered because day of the week can impact on the level of subjective well-being reported (Helliwell and Wang, 2011a). Any attempt to measure the “typical” level of subjective well-being for a group would need to account for regular variations over time, and it may be necessary to develop a specific set of weights to ensure that responses from all days contribute equally to the final estimate.

Holidays (and to some degree the incidence of annual leave) are more problematic in that they tend to be distributed unevenly over the course of the year. Thus, if enumeration cannot be spread over a whole year, there is a risk that an incidence of holidays during the enumeration period that is greater or lesser than normal might bias the survey results. For this reason, it is essential in surveys collected with relatively short enumeration periods that the impact of the inclusion of data collected during any holidays is checked. While it may not be necessary to omit data collected during holidays from output if the impact is negligible or weak, the available evidence on the magnitude of some holidays suggests that testing for potential bias from this source is important. What constitutes a holiday will need to be considered with respect to the context in which the survey is collected. However, it is worth noting that Deaton (2011) finds a large effect for Valentine’s Day in the United States, despite the day not being a public or bank holiday.

Sample size

Large samples are highly desirable in any survey, as they reduce the standard error of estimates and allow both a more precise estimate of subjective well-being as well as a greater degree of freedom with respect to producing cross-tabulations and analysis of results for population sub-groups. With measures of subjective well-being, sample size is particularly important because of the relatively small changes in subjective well-being associated with many areas of analytical interest. Deaton (2011), for example, notes that the expected decline in life satisfaction due to the changes in household incomes and unemployment associated with the 2008 financial crisis is less than the standard error on a sample of 1 000 respondents, and only three times larger than the standard error on a sample of 30 000 respondents. Thus, large samples are highly desirable for measures of subjective well-being.

Although it is impossible to give precise guidelines for what is an appropriate sample size, some general criteria can be noted. Most of the factors that should be taken into account in the planning of any survey also apply when collecting information on subjective well-being. Available resources, respondent burden, sample design (a stratified sample will have a different sample size to a random sample with the same objectives, all other things equal), anticipated response rate and the required output will all influence the desirable sample size. The need for sub-national estimates, in particular, will play an important role in determining the minimum required sample.

Over and above this, some features specific to measures of subjective well-being will influence the desired sample size. On a 0-10 scale, an effect size of 1 implies a very large effect when analysing the determinants of subjective well-being (Boarini et al., 2012). Changes over time are even smaller. The analysis of subjective well-being data therefore requires a relatively large sample size in order to achieve the statistical precision required.¹⁰

Mode

Surveys can be carried out in a number of different modes. Because the mode of collection influences survey costs and respondent burden and can induce mode effects in responses, the choice of mode is an important decision when collecting data. The two modes most commonly used to collect information on subjective well-being are Computer-Assisted Telephone Interviewing (CATI), conducted by an interview over the telephone, and Computer-Assisted Personal Interviewing (CAPI), where the interviewer is personally present when recording the data. Computer-Assisted Self-Interview (CASI) surveys can occur in the presence of an interviewer, when the interviewer is on hand but the respondent enters their own data into a computer questionnaire, or without an interviewer present, such as when the respondent completes an Internet survey. For some purposes traditional chapter-based self-complete surveys are still likely to be relevant. Most time-use diaries, for example, are self-completed chapter diaries filled in by the respondent.

As outlined in Chapter 2, there is good evidence that the collection mode has a significant impact on responses to subjective well-being questions. In general, the use of CASI as a mode tends to produce lower positive self-reports than the use of CAPI, and this is assumed to be because interviewer-led approaches are more likely to prompt more socially desirable responding. CATI is viewed as the least reliable way to collect consistent subjective well-being data, because in these conditions the interviewer is unaware of whether the respondent is answering in a private setting or not, and it is more challenging for interviewers to build rapport with respondents.

As with other features of survey design, the choice of the survey mode will be influenced by a variety of factors, including resource constraints. However, the balance of evidence suggests that, where resources permit, CAPI is likely to produce the highest data quality. This is probably due in part to the rapport that interviewers can build in face-to-face situations. However, CAPI also provides the opportunity to use show cards, which CATI lacks. Show cards that include verbal labels for the scale end-points are particularly valuable in collecting information on subjective well-being where the meaning of the scale end-points changes between questions, as this can impose a significant cognitive burden on respondents (ONS, 2012).

In terms of data quality, CAPI with show cards should be considered best practice for collecting subjective well-being data. Where other modes are used it is important that data producers collect information to enable the impact of mode effects to be estimated. National statistical agencies, in particular, should consider experimentally testing the impact of the mode on responses to the core measures of subjective well-being and publishing the results along with any results from CATI or CASI¹¹ surveys.

Survey vehicles

Questions on subjective well-being should not typically be the subject of a specific survey. As discussed earlier in this chapter, analytical interest in measures of subjective well-being is commonly focused on the interaction between measures of subjective well-being and measures of objective outcomes, including income, aspects of quality of life and time use. It should also be considered that, in most cases, subjective well-being measures are relatively simple and easy to collect. For example, the UK Office for National Statistics found that the four subjective well-being questions used in the Integrated Household Survey take approximately 30 seconds to complete (ONS, 2011). Even a

relatively comprehensive approach to measuring subjective well-being is likely to be more on the scale of a module that could be added to existing surveys rather than requiring a whole survey questionnaire in itself. A key question to consider then is which survey vehicles are most appropriate to the task of measuring subjective well-being.

It is impossible to provide definitive guidance on this issue, because the range of household surveys collected – even among national statistical agencies – varies significantly from country to country. However, it is possible to identify the roles that different survey vehicles can play in collecting subjective well-being data. Seven classes of survey vehicle are relevant to subjective well-being and meet slightly different needs. These are:

- Integrated household surveys.
- General social surveys.
- Time-use surveys.
- Victimization surveys.
- Health surveys.
- Special topic surveys.
- Panel surveys.

Integrated household surveys

Integrated household surveys include the primary surveys used by national statistical agencies to collect information on issues such as income, expenditure and labour market status. In some countries information such as this is collected through separate surveys, such as a labour force survey, while other countries, such as the United Kingdom, rely on an integrated household survey with sub-samples focused on particular topics. Another similar example is the EU-SILC, which consists of a core survey focused on income and living conditions alongside a range of special topic modules. The 2013 EU-SILC module is focused explicitly on well-being. Such surveys are generally not appropriate to be the sole source of information on subjective well-being, as they have a clearly-defined focus that may not align well with an extensive module of subjective well-being and space in these surveys is at a premium. However, such surveys may be more appropriate as a vehicle for a limited set of core questions or a primary measure of subjective well-being intended for monitoring purposes. These questions take up relatively little space in a survey and demand both large sample sizes and regular collection in order to support the effective monitoring of outcomes. Further, subjective measures of this sort complement the economic focus of many integrated household surveys by capturing information on the impact of non-economic factors in a relatively compact form.

General social surveys

Not all national statistical agencies run general social surveys, and among those that do, the content and focus vary considerably. Some national statistical agencies, such as the Australian Bureau of Statistics, focus their general social survey primarily on measures of social capital and social inclusion, while others rotate modules on different topics between survey waves (Statistics Canada) or are explicitly multi-dimensional (Statistics New Zealand). The latter two approaches are particularly appropriate vehicles for collecting information on subjective well-being (and indeed, both Statistics Canada and Statistics

New Zealand collect information on subjective well-being in their general social surveys). Surveys with rotating content, such as the Canadian General Social Survey, offer the opportunity for a subjective well-being module that can collect information in some depth if this is determined to be a priority. Surveys with a wider focus, such as the New Zealand General Social Survey, are particularly valuable in that they allow for the analysis of the joint distribution of subjective well-being and of a wide variety of other topics, including material conditions and objective aspects of quality of life. Regardless of whether a specific subjective well-being module is collected as part of a general social survey, it is very desirable that at least the core module be collected in all general social surveys.

Time-use surveys

Time-use surveys typically involve respondents completing a time-use diary alongside a questionnaire on demographic and other information. The inclusion of a time-use diary offers a unique opportunity to gather information that relates activities to particular subjective states and to collect information on the amount of time spent in different subjective states. In particular, time-use surveys have been used to collect data on affect at varying levels of detail. The American Time Use Survey 2011 included an implementation of the Day Reconstruction Method (Kahneman and Krueger, 2006), which collected detailed information on the affective states associated with a representative sample of episodes drawn from the diaries. This allows analysis of how different affective states vary depending on activity type and calculation of the aggregate amount of time spent in different affective states. The *Enquête Emploi du temps* 2010, run by the French statistical agency, the INSEE, uses an alternative approach to collecting information on subjective well-being in time diaries. Rather than collecting detailed information on a sample of episodes, the INSEE selected a sub-sample of respondents to self-complete a simple seven point scale (-3 to +3), rating each activity from *très désagréable* to *très agréable*. This gives less information on each activity for which information is collected but gathers information on all recorded diary time, therefore providing a larger effective sample of diary entries with subjective well-being information attached.

Victimisation surveys

Victimisation surveys collect information on the level and distribution of criminal victimisation in a society. They are intended to answer questions such as how much crime takes place, what are its characteristics, who are its victims, whether the level of crime is changing over time, who is at risk of becoming a victim, and how do perceptions of safety relate to the actual risk of victimisation (UNECE, 2010). The interaction between victimisation, perceptions of safety and subjective well-being is of high interest, both from the perspective of understanding how victimisation affects well-being, and in order to better understand the impact on the victim of different types of victimisation. Subjective well-being questions are thus of high value to such surveys.

Health surveys

Health surveys already have a considerable tradition of the inclusion of measures of subjective well-being as part of overall and mental health modules such as the widely used GHQ-12 and SF-36 modules. These include questions relating to all three aspects of subjective well-being. However, these modules are calibrated for a specific purpose – measuring overall health status or pre-screening for mental health issues – and in many ways do not conform

with best practice in measuring subjective well-being as outlined here. Because of the importance of health status to subjective well-being, there is considerable value in adding a small number of specific subjective well-being measures to such surveys where possible.

Special topic surveys

Many national statistical agencies run one-off or periodic special topic surveys that are intended to explore a topic in greater detail than would be possible through a question module in a regular survey. Because the content of such a survey can be tailored to the topic in question, such surveys are excellent vehicles for exploring aspects of subjective well-being in more depth. Issues relating to the relationship between different aspects of subjective well-being (i.e. life evaluation, affect, eudaimonic well-being), and between single-item and multiple-item measures of subjective well-being can be examined with such data. However, because of the “one-off” nature of such surveys (or the long periodicity associated with such surveys when they do repeat), special topic surveys are less appropriate for monitoring well-being over time.

Panel surveys

Panel surveys follow the same individuals over time, re-interviewing them in each wave of the survey. Because of this, panel surveys are able to examine questions of causality in a way that is not possible with cross-sectional surveys. Both the German Socio-Economic Panel (GSOEP) and Understanding Society (formerly the British Household Panel Survey) have included questions on subjective well-being for some time, and much of the evidence on the nature of the relationship between life evaluations and their determinants derives from these surveys.

3. Questionnaire design

Questionnaire design is an iterative process involving questionnaire designers, those responsible for determining survey content, and data users. A questionnaire designer must balance the cognitive burden on the respondent, a limited time budget for the survey, and the need to have a questionnaire that is clear, comprehensible and flows well, with different (and often competing) data needs. It is neither possible nor desirable for this chapter to provide a single questionnaire on subjective well-being for users to implement. Instead, the intent of this section is to provide a set of tools to support the development of surveys containing questions on subjective well-being rather than to prescribe a single approach to its measurement.

Some general guidance on issues affecting the inclusion of measures of subjective well-being into a survey is provided below. In particular, the issues of question placement and translation are discussed on their own. This is accompanied by a set of prototype question modules that questionnaire designers should adapt to the specific conditions under which they are working. This section also describes the rationale behind the question modules and an explanation of the template used to describe them. The question modules are attached to these guidelines as Annex B (A to F).

Question placement

Question order and the context in which a question is asked can have a significant impact on responses to subjective questions (see Chapter 2). Although measures of subjective well-being are not uniquely susceptible to such effects – question order and

context will impact on all survey responses to some extent – the effect is relatively large in the case of subjective well-being. Several well-known examples suggest that such effects do need to be taken into account when incorporating questions on subjective well-being into a survey.

In general, question order effects appear to occur, not because the question was early or late in the questionnaire *per se*, but because of the contextual impact of the immediately preceding questions. Thus, the key issue is to identify the most effective way to isolate questions on subjective well-being from the contextual impact of preceding questions. The most direct way of managing contextual effects of this sort is to put subjective questions as early in the survey as possible. Ideally, such questions should come immediately after the screening questions and household demographics that establish respondent eligibility to participate in the survey. This practice almost eliminates the impact of contextual effects and ensures that those that cannot be eliminated in this way are consistent from survey to survey.

However, this cannot be a general response to the issue of dealing with contextual effects for two reasons. First, there will be instances when questions on subjective well-being are added to well-established surveys. In these conditions, changing the flow of the questionnaire would impose significant costs in both resources and data quality. Introducing questions on subjective well-being early in such a survey might ensure that contextual effects do not impact the subjective questions, but this would come at the expense of creating significant contextual effects for the following questions. Second, in cases where there are several such questions in the survey, they cannot all be first.

With these factors in mind, four key recommendations emerge with regard to the placement of subjective well-being questions in surveys. These are as follows:

- *Place important subjective well-being questions near the start of the survey.* Although, as noted above, placing questions early in a survey does not eliminate all of the problems associated with context effects, it is the best strategy available and should be pursued where possible. In particular, for the core measures of subjective well-being, for which international or time series comparisons are an important consideration, it is desirable to place the questions directly after the initial screening questions that result in a respondent's inclusion in the survey. The core measures module included as an annex to this chapter is intended to be placed at the start of a survey in this way.
- *Avoid placing the subjective well-being questions immediately after questions likely to elicit a strong emotional response or that respondents might use as a heuristic for determining their response to the subjective well-being question.* This would include questions on income, social contact, labour force status, victimisation, political beliefs or any questions suggesting social ranking. The best questions to precede subjective questions might be relatively neutral factual demographic questions.
- *Make use of transition questions to refocus respondent attention.* One technique that has been used to address contextual effects resulting from a preceding question on a subjective well-being question is using a transition question designed to focus the respondent's attention on their personal life. Deaton (2011) reports that the introduction of such a question in the Gallup Healthways Well-being Index in 2009 eliminated over 80% of the impact from a preceding question on politics on the subsequent life evaluation measure.¹² However, it is important to consider the risk that transition questions might introduce their own context effects. For example, drawing attention to a respondent's

personal life may lead them to focus on personal relationships or family when answering subsequent questions about life overall. Development of effective transition questions should be a priority for future work.

- Use of introductory text to distinguish between question topics. Well-worded text that precedes each question or topic can serve as a buffer between measures of subjective well-being and sensitive questions. However, there is little hard evidence on the degree of effectiveness or optimal phrasing of such introductory text. A standard introductory text has been included in each of the prototype question modules included as an annex to this chapter. This text is based on what is believed to be best practice. Consistent use of it should help reduce context effects (and will eliminate bias caused by inconsistent introductory text). Further cognitive testing or experimental analysis of the impact of different types of introductory text would, however, be of high value.

Question order within and between subjective well-being modules

Questions on subjective well-being can be affected by previous subjective well-being questions just as easily as by questions on other topics. This has implications for the structure of subjective well-being question modules (particularly where more than one aspect of subjective well-being is addressed), as well as for the presentation of questions within modules and whether it is advisable to include several questions that address very similar topics (see Chapter 2).

In terms of ordering question modules themselves, overall the evidence suggests that moving from the general to the specific may be the best approach. This implies that overall life evaluations should be assessed first, followed by eudaimonic well-being, with more specific questions about recent affective experiences asked next and domain-specific questions last. This is because domain-specific measures in particular risk focusing respondent attention on those domains included in the questions, rather than thinking about their lives and experiences more broadly.

Question order within a battery of questions can also be important – particularly where a group of questions include both positive and negative constructs (such as in the case of affect and some measures of eudaimonia). Although full randomisation of such questions may be optimal, in practice switching between positive and negative items may prove confusing for respondents, who may deal more easily with clusters of questions of the same valence. As discussed in Chapter 2, more evidence is needed to resolve this trade-off, but in the meantime, consistency in the presentation approach (whether randomised or clustered) across all surveys will be important, particularly in terms of whether positive or negative constructs are measured first. In the question modules attached to these guidelines, a clustered approach has been adopted.

Finally, asking two questions about a very similar construct can be confusing for respondents, leading them to provide different answers because they anticipate different answers must be required of them. This means that including several very similar questions about life evaluations, for example, could mean respondents react differently to these questions than when each question is presented in isolation. Thus it is important to have consistency in the number of measures used to assess a given construct, and the order in which those measures are used.

Translation

The exact question wording used in collecting information on subjective well-being matters a lot for responses. As discussed in Chapter 2, a standardised approach to question wording is important for comparisons over time or between groups. This is relatively straight-forward where all surveys are in a single language. However, international comparisons or studies in multi-lingual countries raise the issue of translation. This is a non-trivial matter. Translating survey questionnaires to work in different languages is challenging for any survey, and the potential sensitivity of subjective well-being questions to differences in wording only reinforces this issue.

Potential issues arising from translation cannot be entirely eliminated, but they can be managed through an effective translation process. An example of good practice in the translation of survey questionnaires is provided by the *Guidelines for the development and criteria for the adoption of Health Survey Instruments* (Eurostat, 2005). Although focused on health survey instruments, the framework for translation presented there has broader applicability, and is highly relevant to the measurement of subjective well-being. The health survey guidelines identify four main steps to the translation procedure:

- *Initial or forward translation* of the questionnaire from the source document to the target language.
- *Independent review* of the translated survey instrument.
- *Adjudication* of the translated survey instrument by a committee to produce a final version of the translated survey instrument.
- *Back translation* of the final version of the translated survey instrument into the source language.

Most of the best-practice recommendations identified by Eurostat for health surveys also apply with respect to the measurement of subjective well-being. It is desirable that the initial translation be carried out by at least two independent translators who have the destination language as their mother tongue and who are fluent in the source language. Translators should be informed about the goal of the study and be familiar with the background, origin and technical details of the source questionnaire as well as with the nature of the target population. The reviewer at stage 2 should be independent from the translators, but will ideally need a very similar skill set. Both the reviewer and the translators should be on the adjudication panel, along with an adjudicator whose main area of expertise is the study content and objective. As with any survey design, cognitive interviewing and field testing should be undertaken and the results of this reviewed before the full survey goes into the field.

Back translation is somewhat controversial in the literature on survey translation, with some experts recommending it and others not (Eurostat, 2005). The effect of back translation is generally to shift the focus onto literal translation issues rather than the conceptual equivalent of the original instrument. In the case of the measurement of subjective well-being, back translation is strongly advised. This reflects the sensitivity to question wording of subjective well-being measures (see Chapter 2).

Choice of questions

The choice of which questions to use is of critical importance for measuring subjective well-being. Different questions capture different dimensions of subjective well-being and,

as discussed in Chapter 2, the precise question wording can have a non-trivial impact on results. In selecting questions to incorporate into existing survey vehicles, statistical agencies face trade-offs between the time taken to ask any new questions, the potential impact of new questions on responses to existing questions, and the added information gained from the new questions. These trade-offs will come under particularly severe scrutiny if the survey in question refers to an important and well-established concept (e.g. household income or unemployment).

In selecting subjective well-being questions themselves, there is also a trade-off to manage between using existing questions from the literature that will enable reasonable comparability with previous work, and modifying questions or response formats in light of what has been learned about good practice – including the evidence described in Chapter 2. The approach adopted in this chapter is to recommend tried-and-tested questions from the literature first and foremost. Where a variety of approaches have been used in the past, the rationale for selecting between these is explained. Finally, where there is a case for making small alterations to the question wording based on the evidence in Chapter 2, some modifications are proposed.

For statistical agencies already using subjective well-being measures in their surveys, a crucial question will be whether the potential benefit of using improved measures, and/or more internationally comparable measures, outweighs the potential cost of disrupting an established time series. This is a choice for individual statistical agencies, and will depend on a number of factors, including what the current and future intended use of the data set is, how drastic the change may be, and how long the time series has been established for. It is recommended that any changes to existing questions are phased in using parallel samples, so that the impact of the change can be fully documented and examined. This will enable insights into the systematic impact of changes in methodology and provide agencies with a potential method for adjusting previous data sets (e.g. Deaton, 2011).

In recognition of the different user needs and resources available to statistics producers, this chapter does not present a single approach to gathering information on subjective well-being. Instead, six question modules are attached to the guidelines as Annex B (A to F). Each question module focuses on a distinct aspect of subjective well-being. Question Module A contains the core measures for which international comparability is the highest priority. These are measures for which the evidence on their validity and relevance is greatest, the results are best understood, and the policy uses are the most developed. Of all the six question modules, Module A is unique in that it contains both life evaluation and affect measures, and because all national statistical agencies are encouraged to implement it in its entirety. When this is not possible, the primary measure outlined in the module should be used at the minimum. Modules B through to E are focused on specific aspects of subjective well-being. These modules are not intended to be used in their entirety or unaltered, but provide a resource for national statistical agencies that are developing their own questionnaires.

The six modules are listed below, and those which it is recommended that national statistical offices implement are highlighted as *recommended* in order to distinguish them from those modules intended as a resource for data producers of all types that are developing more detailed questionnaires.

Recommended:

- A. Core measures.

Resource:

- B. Life evaluation.
- C. Affect.
- D. Eudaimonic well-being.
- E. Domain evaluation.

Recommended for time-use surveys:

- F. Experienced well-being.

A. Core measures

The core measures are intended to be used by data producers as the common reference point for the measurement of subjective well-being. Although limited to a few questions, the core measures provide the foundation for comparisons of the level and distribution of life evaluations and affect between countries, over time and between population groups.

Data producers are encouraged to use the core measures in their entirety. The whole module should take less than 2 minutes to complete in most instances. It includes a basic measure of overall life evaluation and three short affect questions. A single experimental eudaimonic measure is also included.

There are two elements to the core measures module. The first is a primary measure of life evaluation. This represents the absolute minimum required to measure subjective well-being, and it is recommended that all national statistical agencies include this measure in one of their annual household surveys.

The second element consists of a short series of affect questions and an experimental eudaimonic question. The inclusion of these measures complements the primary evaluative measure both because they capture different aspects of subjective well-being (with a different set of drivers) and because the difference in the nature of the measures means that they are affected in different ways by cultural and other sources of measurement error. While it is highly desirable that these questions are collected along with the primary measure as part of the core, these questions should be considered a lower priority than the primary measure. In particular, the inclusion of the eudaimonic measure in the core should be considered experimental.

There are essentially two candidate questions for the primary measure. These are the Self-Anchoring Striving Scale (the Cantril Ladder) and a version of the commonly-used question on satisfaction with life. Both have been widely used and have an extensive literature attesting to their validity and reliability. Both questions focus on the evaluative aspect of subjective well-being and have been used in large-scale surveys across many different nations and cultures. The choice between the two measures comes down to a balancing of the strengths and weaknesses of each measure.

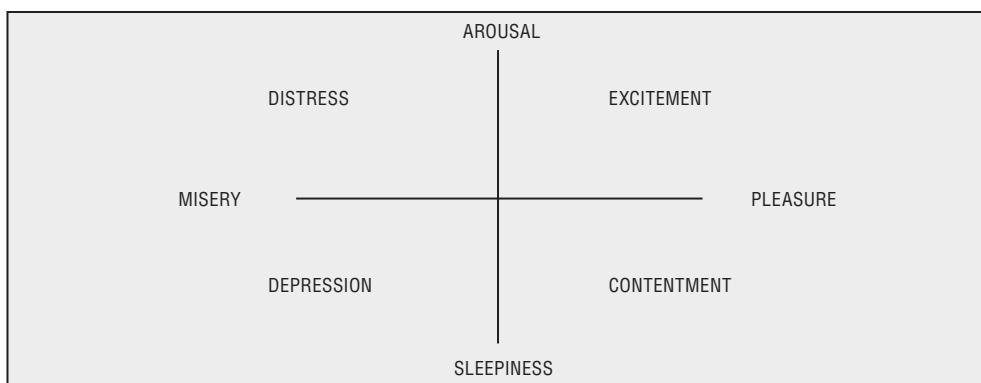
The Cantril Ladder is designed to be “self-anchoring”, and is therefore thought to be less vulnerable to interpersonal differences in how people use the measurement scale. In addition, the anchoring element of the scale is explicitly framed relative to the respondent’s aspirations. This has led some authors to suggest that it may be more rather than less vulnerable to issues of cross-country comparability (Bjørnskov, 2010). Also, the Cantril Ladder tends to produce a marginally wider distribution of responses than does satisfaction with life. However, the Cantril Ladder is a relatively lengthy question, requiring some explanation of the “ladder” concept involved.

By way of contrast, the satisfaction with life question is simple and relatively intuitive. Compared to the Cantril Ladder, the satisfaction with life question has been the subject of much more analysis, reflecting its inclusion not just in the World Values Survey, but also in crucial panel datasets such as the German Socio-Economic Panel and the British Household Panel Survey.

The Cantril Ladder and the satisfaction with life question are relatively similar in terms of their technical suitability for use as an over-arching measure, particularly if both use the same 11-point (0 to 10) scale.¹³ Given this situation, the primary measure included in the core module is a variant of the satisfaction with life question using a 0-to-10 scale. The decisive factor in favour of this choice is the relative simplicity of the question, which will make it easier to incorporate in large-scale household surveys where respondent burden is a significant issue.

Several affect questions are included in the core module. This is because affect is inherently multi-dimensional and no single question can capture overall affect. The various dimensions of affect can be classified in two ways. One of these relates to positive versus negative emotions, while the other relates to level of “arousal”. This gives four affect quadrants and is known as the Circumplex model (Larson and Fredrickson, 1999).¹⁴ Figure 3.3 illustrates the Circumplex model. The quadrants are: positive low arousal (e.g. contentment); positive high arousal (e.g. joy); negative low arousal (e.g. sadness); and negative high arousal (e.g. anger, stress). A good measure of affect might attempt to cover all four quadrants.

Figure 3.3. The circumplex model of affect



Source: Derived from Russell (1980).

Unlike overall life satisfaction, there is not an obvious choice of a simple affect measure that is suitable for inclusion in general household surveys. Most affect scales have been developed in the context either of the measurement of mental health or of more general psychological research. In the former case, many of the existing scales focus excessively on negative affect, while in the latter the scales may be too long for practical use in a household survey. One model for collecting affect measures in a household survey is provided by the Gallup World Poll, which contains a range of questions on affect covering enjoyment, worry, anger, stress and depression, as well as some physical indicators such as smiling or experiencing pain. These questions now have a significant history of use and analysis behind them (Kahneman and Deaton, 2010). A very similar set of questions (on positive affect only) was proposed by Davern, Cummins and Stokes (2007).

The affect questions contained in the proposed prototype module are based on those in the Gallup World Poll and proposed by Davern, but reduced to a list of two questions covering both the negative quadrants of the Circumplex model of affect and a single positive affect question. Only a single positive question is used because the different aspects of positive affect are, in practice, relatively closely correlated. The moods proposed for measurement are happy, worried and depressed. In each case, a 0-to-10 frequency scale is used for responses (ranging from “not at all”, to “all of the time”, similar to the scale anchors used in the European Social Survey).

The eudaimonic question is based on a question trialled by the ONS: “to what extent do you feel the things you do in your life are worthwhile?” There is good evidence from the ONS data that this question captures information not covered by life evaluation and affect measures (NEF, 2012). In addition, a similar question was included in the American Time Use Survey well-being module (Krueger and Mueller, 2012). The question proposed here is similar to that used by the ONS. However, because there is as yet no over-arching theory linking individual questions such as the one proposed to “eudaimonia” as a broad concept, the question should be regarded as experimental.

B. Life evaluation

The life evaluation module is not intended to be used in its entirety. To some degree, the measures it contains should be considered as substitutes for each other rather than as complements. Nonetheless, all of the measures included in the module add something over and above the basic satisfaction with life question contained in Module A. Broadly speaking, there are three groups of question contained in Module B.

The first two questions (the Cantril self-anchoring striving scale and the overall happiness question) are alternative measures of the same underlying concept as satisfaction with life. Although there is some debate as to whether the measures do indeed capture exactly the same concept (Helliwell, Layard and Sachs, 2012) or whether the Cantril scale is a more “pure” measure of life evaluation and overall happiness somewhat more influenced by affect (Diener, Kahneman, Tov and Arora, in Diener, Helliwell and Kahneman, 2010), there is no doubt that the measures are all predominantly evaluative. As discussed above, the Cantril scale is somewhat more awkwardly worded than the satisfaction with life question, but tends to produce a slightly wider distribution of responses (ONS, 2011) and has been thought to have a stronger association with income (Helliwell, 2008; Diener, Kahneman, Tov and Arora, in Diener, Helliwell and Kahneman, 2010). However, when the Cantril Ladder and life satisfaction questions are asked of the same respondents, they show essentially identical responses to income and other variables (Chapter 10 of Diener, Helliwell and Kahneman, 2010), so much so that an average of life satisfaction and the Cantril Ladder performs better than either on its own. Some authors have noted that the word “happiness” can be challenging to translate effectively (Bjørnskov, 2010; Veenhoven, 2008), and Bjørnskov further argues that life satisfaction is easier to translate more precisely. However, happiness may be easier to communicate to the public than the more “technical” satisfaction measures. Helliwell, Layard and Sachs (2012) note, based on analysis of the European Social Survey, that averages of overall happiness and life satisfaction perform better than either does alone in terms of the proportion of variance that can be explained by a common set of explanatory variables.

Some of these questions (B3 and B4) capture information on the respondent's perceptions of prior life satisfaction and their anticipated future life satisfaction. This potentially provides some information about how optimistic or pessimistic the respondent feels, but it can also add information on the respondent's overall life evaluation, as a person's expectations of the future are part of how they evaluate their life. This view is reflected in the methodology for life evaluation used by the Gallup Healthways Well-being Index, which is calculated as the average of the Cantril scale and the anticipated Cantril scale 5 years in the future (Gallup, 2012).

Finally, the module includes the five questions (B5 to B9) that together define the Satisfaction With Life Scale (SWLS) developed by Ed Diener and William Pavot. The SWLS is one of the best-tested and most reliable multi-item scales of life evaluation. Since its development in 1985, the SWLS has accumulated a large body of evidence on its performance and has been tested in a number of different languages (Pavot and Diener, 1993). Because the SWLS is a multi-item measure, it has a higher reliability than single-item measures and is more robust to inter-personal differences in scale interpretation than they are. The SWLS adds value to the primary life evaluation measure in contexts where more space is available in a questionnaire, and where a more reliable measure of life evaluation would help interpret and calibrate the results from the primary measure.

C. Affect

Best practice for collecting data on directly-experienced affect involves either sampling people throughout the course of the day and recording their affective state (experience sampling method or ESM) or a detailed reconstruction of daily activity and of the associated affective states (the day reconstruction method or DRM). The former approach (ESM) is not discussed here in detail as it typically involves the use of electronic pagers or similar devices more suited to experimental research design than to official statistics. The DRM approach can be implemented in large-scale surveys containing a time-use diary and forms the basis of the experienced well-being module presented in this chapter. However, time-use diaries are expensive to collect and code, and there are times when it may be desirable to collect affect data from a general household survey. This module provides an approach to collecting affect data in such a survey and expands on the more limited range of affect questions contained in the core questions module.

There are several approaches to measuring affect in household surveys. The European Quality of Life Survey (Eurofound, 2007), for example, asks five questions about how people felt during the previous two weeks. These questions ask respondents to rate how much of the previous two weeks they experienced each feeling on a 6-point scale. Similarly, the European Social Survey has 15 questions on the respondent's affective state over the past week, with responses on a 4-point scale. The SF-36 health measurement tool contains a set of nine items relating directly to the respondent's affective state over the previous four weeks, also using a 6-point scale (Ware and Gandek, 1998). Five of these nine items have been included in the EU-SILC 2013 well-being module to capture affect. However, the length of the reference period in both the EQLS and SF-36 questions is potentially problematic, as errors are likely to increase with the length of the reference period.

While the four-week period used in the SF-36 is well suited for its intended purpose – assessing mental health – recall of affective states is likely to be better when the recall period is short and the question refers to a specific day (see Chapter 2). The affect questions presented in Module C, therefore, are similar in structure to those contained in the

Gallup World Poll (although also drawing on the ESS questionnaire). These questions focus specifically on the affective state of the individual on the previous day. In addition, they ask for a 0 to 10 frequency judgement, rather than requiring judgements relating to the intensity of the feeling.

The affect module includes 10 questions, largely drawn from those used in the Gallup World Poll and those used by the ONS. Four of the questions are related to positive affect and six to negative affect, reflecting the apparent potential for multi-dimensionality in negative affect in particular.

D. Eudaimonic well-being

Eudaimonic well-being encompasses a range of concepts, many with clear policy relevance. Some aspects of eudaimonic well-being – such as meaning or a sense of purpose in life, and a sense of belonging – capture elements of subjective well-being not reflected in life evaluation or affect, but with high intuitive relevance. Other aspects of eudaimonic well-being – such as “agency” and locus of control – are more tenuously related to subjective well-being conceived of as an outcome, but are powerful explanatory variables for other behaviour. However, this level of potential relevance is not matched by an equally good understanding of what eudaimonic well-being actually “is”, and more specifically, how it should be measured. In particular, as discussed in Chapter 1, it is not clear whether eudaimonic well-being captures a single underlying construct like life evaluation or is rather an intrinsically multi-dimensional concept like affect. For this reason, the proposed measures of eudaimonic well-being should be considered experimental.

The eudaimonic well-being module proposed here is based on elements of the European Social Survey well-being module and the Flourishing Scale proposed by Diener et al. (2010). This provides a starting point for data producers who wish to collect measures of eudaimonic well-being. The proposed questions are consistent with what is known about best practice in collecting such information (see Chapter 2), but cannot be considered definitive in the absence of a coherent body of international and large, representative-sample research comparable to that existing for life evaluation and affect.

E. Domain evaluations

In addition to evaluating life as a whole, it is also possible to collect information evaluating specific life “domains” such as health or standard of living. Such information has a wide range of potential uses (Dolan and White, 2007; Ravallion, 2012) and may be better adapted to some policy and research questions than over-arching evaluations relating to life as a whole. A challenge to providing advice on measuring domain evaluations is the sheer range of possible life domains that could be measured. Some areas, such as job satisfaction, have substantial literatures in their own right, while others do not. The goal of the domain evaluation module is not to provide an exhaustive approach to measuring subjective evaluations of all policy-relevant life domains. Instead, the focus is on the more limited objective of detailing a limited set of domain evaluation measures that can be used in a general social survey focused on measuring well-being across multiple domains or as the basis for an analysis of the relationship between overall life satisfaction and domain evaluations (e.g. Van Praag, Frijters and Ferrer-i-Carbonell, 2003). Each individual question can, of course, be used in its own right in analysis and monitoring of the particular outcome that it reflects.

Ideally, the questions comprising the domain satisfaction block would meet two key criteria. First, they would be independently meaningful as measures of satisfaction with a particular aspect of life; and second they would collectively cover all significant life domains. A major practical challenge to this sort of approach, however, is that there is no generally agreed framework for identifying how to divide well-being as a whole into different life domains. Different authors have taken different approaches. For example, the Stiglitz/Sen/Fitoussi commission identified eight domains of quality of life alongside economic resources, while the ONS proposal for measuring national well-being identified a slightly different set of nine domains of well-being, including “individual well-being” as a distinct domain capturing overall life evaluations. The OECD’s Better Life Initiative uses eleven life domains, including a distinct domain on subjective well-being, while the New Zealand General Social Survey uses a slightly different set of ten domains. Another approach is that adopted for constructing the Personal Wellbeing Index (PWI; International Wellbeing Group, 2006). This consists of eight primary items that are meaningful on their own but which can also be used to calculate an overall index of subjective well-being. The domains that are included in the PWI have been subject to considerable testing and reflect the results of extensive factor analysis. Table 3.1 compares these different approaches.

Table 3.1. A comparison of life domains

SSF	BLI	ONS	NZGSS	PWI
Material conditions	Income and wealth	Personal finance	Economic standard of living	Standard of living
Economic insecurity		The economy		Future security
Housing				
	Jobs and earnings	What we do	Paid work	
Personal activities	Work and life balance		Leisure and recreation	
Health	Health status	Health (physical and mental)	Health	Personal health
Education	Education and skills	Education and skills	Knowledge and skills	
Social connections	Social connections	Our relationships	Social connectedness	Personal relationships Community connectedness
Political voice and governance	Civic engagement and governance	Governance	Civil and political rights	
Personal insecurity	Personal security	Where we live	Safety	Personal safety
Environmental conditions	Environmental quality	The environment	The environment Cultural identity	
			Achieving in life	
	Subjective well-being	Individual well-being	Life satisfaction	

Note: The SSF domains have been widely used by other agencies as the basis of their own analysis of well-being. The European Union Sponsorship Group on Measuring Progress, Well-being and Sustainable Development, the UNECD Sustainable Development Task Force, and the ISTAT, for example, have all adopted variants of the SSF approach for various purposes.

Source: The acronyms used in Table 3.1 are: *Commission on the Measurement of Economic Performance and Social Progress – Sen, Stiglitz, Fitoussi (SSF); Your Better Life Index – OECD (BLI); Office for National Statistics (ONS); New Zealand General Social Survey (NZGSS); and Personal Wellbeing Index (PWI)*.

There is a high degree of overlap in the different approaches outlined in Table 3.1. The proposed domain evaluations module draws on these to identify ten questions pertaining to ten specific life domains. These domains include the constituent elements of the PWI as a subset, but also three additional domains (time to do what you like doing, the quality of the environment, and your job) that are of potential policy relevance in and of themselves. The nine domains are:

- Standard of living.

- Health status.
- Achievement in life.
- Personal relationships.
- Personal safety.
- Feeling part of a community.
- Future security.
- Time to do what you like doing.
- Quality of the environment.
- Your job (for the employed).

The ten proposed questions cover all of the main domains of well-being identified in Table 3.1 except one: governance. The range of concepts covered by political voice, governance and civil and political rights is very broad, and there is no model question or set of questions that could be used as the basis for inclusion in these guidelines. Similarly, there would be little value in developing a question from scratch without testing to see how the question performs. However, governance is undeniably an important dimension of well-being. The issue of how best to collect information on satisfaction with governance, political voice and civil and political rights therefore remains a key area for future research.

F. Experienced well-being

As noted in the section describing Module C (affect), the gold standard for measuring affect is via the experience sampling method. When this is not possible, the day reconstruction method (DRM) provides a well-tested methodology that produces results consistent with the experience sampling method (Kahneman and Krueger, 2006). Although it is not possible to implement the DRM in general household surveys, it is possible in time-use surveys. This module provides approaches to implementing the measurement of affect in time-use diaries. Because of the value that time-use diary information on subjective well-being adds, and because information on affect yesterday from general household surveys is not a good substitute¹⁵ for measures like those collected through the DRM, it is strongly recommended that information on experienced well-being be collected in time-use surveys whenever possible.

The experienced well-being module presents two approaches to measuring subjective well-being in time-use diaries. The first is essentially the implementation of the DRM used in the 2011 American Time Use Survey (ATUS). This provides aggregate information similar to the full DRM, but restricts the information collected to only three diary episodes per respondent. This helps reduce the respondent burden and the amount of interviewer time required per respondent, which is otherwise relatively high with the full DRM. The data collected using this method is exceptionally rich, as it involves collecting information on a number of different moods and feelings. As with the affect questions in Module C, it uses a 0-10 scale. This is longer than the 0-6 scale currently used in the ATUS, but is preferred for reasons of consistency with other scales used in these guidelines and because the (relatively limited) literature on the subject tends to support the choice of the longer scale (Kroh, 2006; Cummins and Gullone, 2000).

An alternative to the DRM is also included in the experienced well-being module. This is based on the “unpleasant/pleasant” (“très désagréable/très agréable”) approach used by the INSEE in the *Enquête Emploi du temps* 2010. Although the INSEE approach captures less

information than the DRM – the measure used to collect information on affective state is uni-dimensional – it does have two significant advantages. First, it is a self-complete question that can be included on the diary form. This significantly reduces interviewer time and the associated costs, and does not add much to the time required for respondents to fill in the diary (INSEE, 2010). Because of this, information can be collected on the respondent's affective state during all diary episodes, allowing more comprehensive analysis. The self-completed nature of the question also makes it potentially suitable for inclusion in “light” time-use surveys that rely more heavily on respondents to self-complete their diary. The second point in favour of the INSEE approach is that analysis of the available data suggests that the results are broadly comparable with results from the DRM when these are reduced to a uni-dimensional measure such as the “U-index” or affect balance.

There is currently relatively little basis to assess which method is preferable overall. The DRM is better grounded in the research literature, with good evidence of its validity, and provides a more detailed view of the different moods people experience. On the other hand, the INSEE approach appears to manage adequate data quality combined with significantly lower respondent and interviewer burden, as well as detail on a complete sample of episodes. Resolving the issue of which approach is to be preferred will require further analysis, drawing on data derived from both methodologies. For this reason, both approaches are detailed in the experienced well-being module.

Question templates

The six question modules are attached to these guidelines as Annex B. Each question module is presented in the same format, containing a common set of headings that outline the objectives of the module (what kind of information it is trying to gather), a description of the contents of the module, the origin of the questions in the module, how the data from the module should be presented, background information for interviewers, and the detailed question wording. These headings are described in more detail below.

Objectives

The objective succinctly outlines the purpose of the block, including both the type of information it is designed to elicit and the rationale behind the scope of the question block.

Description

A description of the contents of each question module is provided, outlining the role of each of the questions in the module with respect to the module's objectives. The description is intended to assist users to identify which questions they wish to use in the event that they choose to implement only part of the module.

Origin

Questions included in each module are drawn from existing sources and remain unchanged wherever possible to maximise comparability with previous work. However, some items have been modified to a greater or lesser extent where a variety of question versions exist in the literature, and/or where there are clear grounds for small changes in item wording or response scales, for example based on the evidence in Chapter 2. The origin section indicates the source of the questions and notes whether they have been altered.

Completion time

This gives an estimate of the time required to run the entire module.

Output

The output section contains basic information on the production of standard tables and measures from the question block. This information is not exhaustive, but is intended to provide some basic guidelines for data producers. Such guidelines are important, both in order to assist producers in presenting the data appropriately, but also to provide context for why the questions are framed in the way that they are.

A number of the question blocks are intended to produce multi-item measures of subjective well-being derived from the survey questions. The output section provides details on the construction of these multi-item measures, and how they should be reported.

Guidelines

The quality of any survey data is heavily influenced by the attitude of the respondents to the questions they are being asked. Although the evidence is overwhelming that measures of subjective well-being are not regarded as particularly challenging or awkward by respondents (particularly when compared to questions on some other commonly-asked topics, such as income), better-quality information is likely to result if interviewers understand what information is being collected and how it will be used, and they are able to communicate this clearly to respondents. This enables interviewers to answer queries from respondents on why the information is important or on what concept the question is trying to elicit from them.

The guidelines for interviewers contained in this module are not intended as a substitute for the more extensive notes and/or training that would normally be provided to interviewers in the process of preparing to conduct a household survey. However, they do provide a basis from which users of the module can develop their own more substantive guidelines.

4. Survey implementation

How a survey is implemented is crucial to its effectiveness. A carelessly-implemented survey will result in low-quality and unreliable data regardless of the quality of the underlying questionnaire. In general, the features relevant to the effective implementation of any household survey also hold for those collecting information on subjective well-being. These guidelines make no attempt to provide a detailed discussion of best practice in survey implementation, for which high-quality standards and guidelines already exist (UN, 1984). However, there are several points where the specific nature of measures of subjective well-being raises survey implementation issues that are worth noting.

Interviewer training

Interviewer training is crucial to the quality of responses in any survey. However, the measurement of subjective well-being raises additional issues because the subject matter may be unfamiliar to interviewers. This is, ironically, particularly so for national statistical agencies with a permanent force of field interviewers. Although a body of trained interviewers will generally contribute to higher response rates and better responses, interviewers may struggle with questions if they cannot explain adequately to respondents why collecting such information is important and how it will be used. Anecdotal evidence

and feedback from cognitive testing shows that this can be an issue with some subjective measures, particularly measures of affect (ONS, 2012). In some cases, respondents may find it difficult to understand why government might want to collect this information and that the concept that the survey intends to collect is their recently-experienced affective state rather than their normal affective state.

To manage risks around respondent attitudes to questions on subjective well-being, it is imperative that interviewers are well-briefed, not just on what concepts the questions are trying to measure, but also on how the information collected will be used. This is essential for interviewers to build a rapport with respondents and can be expected to improve compliance by respondents and the quality of responses. While the notes on interviewer guidelines contained in the question modules provide some crucial information specific to each set of questions, a more comprehensive approach should draw on information on the validity and use of measures of subjective well-being (Chapter 1) and the analysis of subjective well-being data (Chapter 4).

Ethical issues

Evidence suggests that measures of subjective well-being are relatively non-problematic for respondents to answer. Rates of refusal to respond are low, both for life evaluations and for measures of affect (Smith, 2013). In general, item-specific non-response rates for subjective well-being measures are similar to those for marital status, education and labour market status, and much lower than for measures of income (Smith, 2013). This suggests that, in general, such questions are not perceived as problematic by respondents.

Cognitive testing of measures of subjective well-being supports the conclusions reached from an examination of item-specific non-response rates (ONS, 2012), with some notable exceptions. In particular, the ONS found that eudaimonic questions relating to whether respondents felt that what they did in life were worthwhile and the experience of loneliness caused visible distress in some respondents, particularly among disabled and unemployed respondents.

Best practice suggests that statistical providers should consider how to manage the risks associated with questions that are distressing to respondents. Although it is important not to overstate the risks – they apply mainly to eudaimonic questions, and to a small proportion of respondents – such issues should be dealt with effectively. A complicating factor is that it might not be evident at the time of the interview whether a respondent has been affected by the questioning. One approach to managing this proposed by the ONS (2012) is to distribute a leaflet at the time of the interview giving respondents information on the purpose of the survey and reiterating the confidentiality of the data collected. The leaflet would also contain information for distressed respondents about where to seek help.

Coding and data processing

The coding of information on subjective well-being is generally straight-forward. In general, numerical scales should be coded as numbers, even if the scale bounds have labels. Much analysis of subjective well-being data is likely to be quantitative and will involve manipulating the data as if it were cardinal. Even for fully-labelled response scales (such as the “yes/no” responses that apply to many questions), it is good practice to code the data numerically as well as in a labelled format in order to facilitate use of the micro-data to

produce summary measures of affect balance or similar indices. “Don’t know” and “refused to answer” responses should be coded separately from each other as the differences between them are of methodological interest.

Normal data-cleaning procedures include looking for obvious errors such as data coders transposing numbers, duplicate records, loss of records, incomplete responses, out-of-range responses or failure to follow correct skip patterns. Some issues are of particular relevance to subjective data. In particular, where a module comprising several questions with the same scale is used, data cleaning should also involve checking for response sets (see Chapter 2). Response sets occur when a respondent provides identical ratings to a series of different items. For example, a respondent may answer “0” to all ten domain evaluation questions from Module E. This typically suggests that the respondent is not, in fact, responding meaningfully to the question and is simply moving through the questionnaire as rapidly as possible. Such responses should be treated as a non-response and discarded. In addition, interviewer comments provide an opportunity to identify whether the respondent was responding correctly, and a robust survey process will make provision for allowing such responses to be flagged without wiping the data record.

Finally, it is important to emphasise that much of the value from collecting measures of subjective well-being comes from micro-data analysis. In particular, analysis of the joint distribution of subjective well-being and other outcomes and use of subjective well-being measures in cost-benefit analysis cannot usually be accomplished through secondary use of tables of aggregate data. Because of this, a clear and comprehensive data dictionary should be regarded as an essential output in any project focusing on subjective well-being. This data dictionary should have information on survey methodology, sampling frame and correct application of survey weights, as well as a description of each variable (covering the variable name, the question used to collect it and how the data is coded). If a variable is collected from only part of the survey sample due to question routing, this should also be clearly noted in the data dictionary.

Notes

1. See the section of Chapter 4 (*Output and analysis of subjective well-being measures*) relating to cost-benefit analysis for an example of this distinction when comparing stated preference approaches to estimating non-market values as opposed to using subjective well-being measures.
2. The distinction between cardinal and ordinal measures is important to measuring subjective well-being. With ordinal measures the responses are assumed to show the rank order of different states, but not the magnitude. For example, with ordinal data a 5 is considered higher than a 4 and an 8 is considered higher than a 7. However, nothing can be said about the relative size of the differences implied by different responses. For cardinal data it is assumed that the absolute magnitude of the response is meaningful, and that each scale step represents the same amount. Thus, a person with a life satisfaction of 5 would be more satisfied than someone reporting a 4 by the same amount as someone reporting an 8 compared to a 7. Most subjective well-being measures are technically ordinal, but the evidence suggests that treating them as cardinal does not generally bias the results obtained (Ferrer-i-Carbonell and Frijters, 2004).
3. In many cases it may be possible to collect one income measure (say, gross household income) and impute net household income on the basis of household size and composition and the relevant tax rates and transfer eligibility rules.
4. One example of these sorts of measure is provided by the material deprivation questions contained in the EU-SILC.
5. See for example the Eurostat social inclusion and living conditions database, http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/data/database.

6. These other factors potentially include both the confounding effects of shared method variance if the quality-of-life measures in question are subjective, or more substantive factors such as events earlier in the life course that may impact both income and quality of life.
7. The Five Factor Model is a psychological framework for analysing personality type. It identifies five main factors that relate to personality: Neuroticism; Extraversion; Openness to experience; Agreeableness; and Conscientiousness. The scale is widely used in psychological research and is well suited to inclusion in survey questionnaires.
8. “Frame of reference” refers to the situation or group on which respondent’s base comparisons when formulating a judgement about their lives or feelings. The respondent’s knowledge of how others live and their own prior experiences can influence the basis on which judgements are reached about the respondent’s current status.
9. This is not, in fact, beyond the realm of possibility. Many government agencies may have an interest in collecting measures of client satisfaction. However, the case for collecting general measures of subjective well-being as a standard part of interactions with government service delivery agencies is beyond the scope of this paper.
10. The need for a relatively large sample size is one reason to prefer a simple measure of subjective well-being with a low respondent burden in place of a technically more reliable multi-item measure with a higher respondent burden. The quality gains from a more detailed measure need to be assessed carefully against the quality losses associated with any reduction in sample size associated with a longer measure.
11. Internet surveys are, from this perspective, a way of implementing CASI.
12. In this case the precise transition question used was: “Now thinking about your personal life, are you satisfied with your personal life today”, and the subjective well-being measure that followed was the Cantril self-anchoring ladder of life measure. It does not follow that the same transition question will work in other contexts, and transition questions should be tested empirically before being relied on.
13. Some versions of the satisfaction with life question use different response scales, such as a 5-point labelled Likert scale or a 1-10 scale. Based on the conclusions from Chapter 2, the core module uses a 0-10 end-labelled scale.
14. Technically the Circumplex model implies that positive and negative affect are ends of a single dimension rather than a way of grouping several independent types of feeling. Here the Circumplex model is used as an organising framework to help impose some structure on the range of different affective states, without assuming continuity on the positive/negative axis.
15. Information on affect yesterday from general household surveys, while of interest in its own right, does not allow analysis of how different activities, locations and the people with whom the respondent is with impact on subjective well-being.

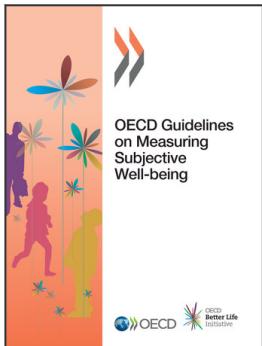
Bibliography

- Balestra, C. and J. Sultan (2013), “Home Sweet Home: The Determinants of Residential Satisfaction and Its Relation with Well-Being”, *OECD Statistics Directorate Working Papers* (forthcoming), OECD, Paris.
- Bjørnskov, C. (2010), “How Comparable are the Gallup World Poll Life Satisfaction Data?”, *Journal of Happiness Studies*, Vol. 11, pp. 41-60.
- Blanchflower, D. and A. Oswald (2011), “International happiness”, *NBER Working Paper*, No. 16668, National Bureau of Economic Research.
- Blanchflower, D. and A. Oswald (2008), “Is well-being U-shaped over the life cycle?”, *Social Science and Medicine*, Vol. 66(8).
- Boarini, R., M. Comola, C. Smith, R. Manchin and F. De Keulenaer (2012), *What Makes for a Better Life? The determinants of subjective well-being in OECD countries: Evidence from the Gallup World Poll*, STD/DOC(2012)3, OECD.
- Card, D., A. Mas, E. Moretti and E. Saez (2010), “Inequality at work: the effect of peer salaries on job satisfaction”, *NBER Working Paper*, No. 16396, National Bureau of Economic Research.
- Clark, A.E., Y. Georgellis and P. Sanfey (1998), “Job Satisfaction, Wage Changes and Quits: Evidence From Germany”, *Research in Labor Economics*, Vol. 17.

- Costa, P.T. Jr. and R.R. McCrae (1992), *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual*, Odessa, FL, Psychological Assessment Resources.
- Cummins, R.A., R. Eckerslet, J. Pallant, J. Vugt and R. Misajon (2003), "Developing a National Index of Subjective Wellbeing: The Australian Unity Wellbeing Index", *Social Indicators Research*, No. 64, pp. 159-190.
- Cummins, R.A. and E. Gullone (2000), "Why we should not use 5-point Likert scales: The case for subjective quality of life measurement", *Proceedings, Second International Conference on Quality of Life in Cities*, National University of Singapore, pp. 74-93.
- Davern, M., R. Cummins and M. Stokes (2007), "Subjective Wellbeing as an Affective-Cognitive Construct", *Journal of Happiness Studies*, No. 8, pp. 429-449.
- Deaton, A. (2011), "The financial crisis and the well-being of Americans", *Oxford Economic Papers*, No. 64, pp. 1-26.
- Diener, E., J. Helliwell and D. Kahneman (eds.) (2010), *International Differences in Well-Being*, Oxford University Press.
- Diener, E., S. Oishi and R. Lucas (2003), "Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life", *Annual Review of Psychology*, No. 54, pp. 403-425.
- Dolan, P., T. Peasgood and M. White (2008), "Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being", *Journal of Economic Psychology*, Vol. 29, pp. 94-122.
- Dolan, P. and M. White (2007), "How can measures of subjective well-being be used to inform policy?", *Perspectives on Psychological Science*, Vol. 2(1), pp. 71-85.
- Easterlin, R. (1974), "Does Economic Growth Improve the Human Lot? Some Empirical Evidence", in David, P.A. and M. W. Reder, *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz*, New York, Academic Press Inc, pp. 89-125.
- Eurofound (2007), *2007 European Quality of Life Survey Questionnaire*.
- European Social Survey (2007), *Final Source Questionnaire Amendment 03*.
- Eurostat (2005), *Guidelines for the development and criteria for the adoption of Health Survey instruments*, Eurostat, Luxembourg.
- Eurostat (2004), *Guidelines on harmonised European Time Use Surveys*.
- Ferrer-i-Carbonell, A. and P. Frijters (2004), "How important is methodology for the estimates of the determinants of happiness?", *The Economic Journal*, No. 114, pp. 641-659.
- Frey, B.S. and A. Stutzer (2000), "Happiness, Economy and Institutions", *The Economic Journal*, Vol. 110(466), pp. 918-938.
- Frey, B.S. and A. Stutzer (2008), "Stress that Doesn't Pay: The Commuting Paradox", *Scandinavian Journal of Economics*, Vol. 110(2), pp. 339-366.
- Gallup Organisation (2012), *Indexes and Questions*.
- Goldberg, D.P. et al. (1978), *Manual of the General Health Questionnaire*, Windsor, England, NFER Publishing.
- Gutiérrez, J., B. Jiménez, E. Hernández and C. Puente (2005), "Personality and subjective well-being: big five correlates and demographic variables", *Personality and Individual Differences*, No. 38, pp. 1561-1569.
- Harkness, J., B.E. Pennell and A. Schoua-Glusberg (2004), "Survey Questionnaire Translation and Assessment", in S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin and E. Singer (eds.), *Methods for Testing and Evaluating Survey Questionnaires*, John Wiley and Sons, Inc, Hoboken, NJ, USA.
- Helliwell, J.F. (2008), "Life Satisfaction and the Quality of Development", *NBER Working Paper*, No. 14507, National Bureau of Economic Research.
- Helliwell, J.F., R. Layard and J. Sachs (eds.) (2012), *World Happiness Report*, The Earth Institute, Columbia University.
- Helliwell, J.F. and S. Wang (2011a), "Weekends and Subjective Well-being", *NBER Working Paper*, No. 17180, National Bureau of Economic Research.
- Helliwell, J.F. and S. Wang (2011b), "Trust and Well-being", *International Journal of Wellbeing*, available online at: www.internationaljournalofwellbeing.org/index.php/ijow/issue/current.

- Huppert, F and T. So (2008), "Deriving an objective definition of well-being", *Working Paper, Well-being Institute, University of Cambridge*. See also J. Michaelson, S. Abdallah, N. Steur, S. Thompson and N. Marks, *National Accounts of Well-being: Bringing real wealth onto the balance sheet*, New Economics Foundation.
- INSEE (2010), *Enquête Emploi du temps*.
- International Wellbeing Group (2006), *Personal Wellbeing Index*, 4th edition, Melbourne, Australian Centre on Quality of Life, Deakin University, available online at: www.deakin.edu.au/research/acqol/instruments/wellbeing_index.htm.
- Kahneman, D. and A. Deaton (2010), "High income improves life evaluation but not emotional well-being", *Proceedings of the National Academy of Sciences*, Vol. 107(38), pp. 16489-16493.
- Kahneman, D., E. Diener and N. Schwarz (1999), *Well-being. The Foundations of Hedonic Psychology*, Russel Sage Foundation, New York.
- Kahneman, D. and A.B. Krueger (2006), "Developments in the Measurement of Subjective Well-Being", *Journal of Economic Perspectives*, Vol. 20(1), pp. 19-20.
- Kroh, M. (2006), "An experimental evaluation of popular well-being measures", *DIW Berlin Working Paper*, No. 546.
- Krueger, A.B. and A.I. Mueller (2012), "Time Use, Emotional Well-Being, and Unemployment: Evidence from Longitudinal Data", *The American Economic Review*, Vol. 102(3), pp. 594-599.
- Larsen, R.J. and B.L. Fredrickson (1999), "Measurement issues in emotion research. Well-being", *The Foundations of Hedonic Psychology*, pp. 40-60.
- Lucas, R.E. (2007). "Long-term disability is associated with lasting changes in subjective well-being: evidence from two nationally representative longitudinal studies", *Journal of Personality and Social Psychology*, Vol. 92(4), p. 717.
- Lucas, R.E., A. Clark, Y. Georgellis and E. Diener (2004), "Unemployment alters the set point for life satisfaction", *Psychological Science*, No. 15, pp. 8-13.
- NEF (2012), *Well-being patterns uncovered: An analysis of UK data*, United Kingdom.
- OECD (2009), *Doing Better for Children*, OECD Publishing, Paris.
- ONS (2012), *Subjective Well-being: A qualitative investigation of subjective well-being questions*, ONS, United Kingdom.
- ONS (2011), *Initial investigations into Subjective Well-being from the Opinions Survey*, ONS, United Kingdom.
- Oswald, F., H. Wahl, H. Mollenkopf and O. Schilling (2003), "Housing and Life Satisfaction of Older Adults in Two Rural Regions in Germany", *Research on Ageing*, Vol. 25(2), pp. 122-143.
- Pavot, W. and E. Diener (1993), "Review of the Satisfaction With Life Scale", *Psychological Assessment*, Vol. 5(2), pp. 164-172.
- Ravallion, M. (2012), "Poor, or just feeling poor? On using subjective data in measuring poverty", *World Bank Policy Research Working Paper*, No. 5968, World Bank Development Research Group, Washington, DC.
- Russell, J. (1980), "A Circumplex Model of Affect", *Journal of Personality and Social Psychology*, Vol. 39(6), pp. 1161-1178.
- Sacks, W.D., B. Stevenson and J. Wolfers (2010), "Subjective Well-being, Income, Economic Development and Growth", *NBER Working Paper*, No. 16441.
- Silva, J., F. De Keulenaer and N. Johnstone (2012), "Individual and Contextual Determinants of Satisfaction with Air Quality and Subjective Well-Being: Evidence based on Micro-Data", *OECD Environment Directorate Working Paper*, OECD Publishing, Paris.
- Smith, C. (2013), "Making Happiness Count: Four Myths about Subjective Measures of Well-Being", *OECD Paper prepared for the ISI 2011: Special Topic Session 26*.
- Stiglitz, J.E., A. Sen and J.P. Fitoussi (2009), *Report by the Commission on the Measurement of Economic Performance and Social Progress*.
- Tinkler, L. and S. Hicks (2011), *Measuring Subjective Well-being*, ONS, United Kingdom.
- UNECE Secretariat (2009), *Revised terms of reference of UNECE/WHO/Eurostat steering group and task force on measuring health status*, UNECE.

- UNECE (2010), *Manual on Victimization Surveys*, United Nations.
- UNICEF (2007), *Child poverty in perspective: An overview of child well-being in rich countries*, Innocenti Report Card 7.
- United Nations Statistical Division (1986), *National Household Survey Capability Programme, Sampling Frames and Sample Designs for Integrated Survey Programmes. Preliminary version*, United Nations, New York.
- United Nations Statistical Division (1984), *Handbook of Household Surveys*, United Nations, New York.
- Van Praag, B.M.S., P. Frijters and A. Ferrer-i-Carbonell (2003), "The anatomy of subjective well-being", *Journal of Economic Behaviour and Organisation*, No. 51, pp. 29-49.
- Veenhoven, R. (2008), "The International Scale Interval Study: Improving the Comparability of Responses to Survey Questions about Happiness", in V. Moller and D. Huschka (eds.), *Quality of Life and the Millennium Challenge: Advances in Quality-of-Life Studies, Theory and Research*, Social Indicators Research Series, Vol. 35, Springer, pp. 45-58.
- Ware, J. and B. Gandek (1998), "Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project", *Journal of Clinical Epidemiology*, Vol. 51(11), pp. 903-912.
- Weinhold, D. (2008), "How big a problem is noise pollution? A brief happiness analysis by a perturbable economist", *MPRA Working Paper*, No. 10660.
- Winkelmann, L. and R. Winkelmann (1998), "Why Are The Unemployed So Unhappy? Evidence From Panel Data?", *Economica*, Vol. 65, pp. 1-15.
- World Health Organisation (2012), *World Health Survey Instruments and Related Documents*.



From:

OECD Guidelines on Measuring Subjective Well-being

Access the complete publication at:

<https://doi.org/10.1787/9789264191655-en>

Please cite this chapter as:

OECD (2013), "Measuring subjective well-being", in *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264191655-7-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.