

Chapter 4

Output and analysis of subjective well-being measures

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Note by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

Introduction

This chapter provides guidance regarding the release and use of subjective well-being data. It briefly re-caps the policy and public interest in the data (outlined in Chapter 1, *Concept and validity*), before covering how information can be reported and analysed. This includes the statistical outputs that may be released; basic information about the methods of analysis that may be adopted; and a discussion of key interpretive issues, placing particular emphasis on the extent to which levels of subjective well-being can be expected to vary in different circumstances.

The chapter is divided into three main sections, summarised in Table 4.1. The first and largest section focuses on the use of subjective well-being data to complement existing measures of well-being. This includes examination of trends in subjective well-being over time, the distribution of subjective well-being across different groups within society, and its distribution across different countries. The first part of this section outlines approaches to measuring well-being and the value added that subjective well-being brings relative to other measures. The second part of the section then describes how subjective well-being can be reported, including the summary statistics that may be of interest. Finally, issues in the analysis and interpretation of descriptive statistics on subjective well-being are explored. These include the size of change over time, or difference between groups, that can be expected, as well as the risk of cultural “bias” in cross-country comparisons.

The remaining two sections of the chapter deal with more detailed analyses of subjective well-being, which might be conducted by government analysts and others on the basis of micro-data released by statistical agencies. Section 2 addresses analyses of the *drivers* of subjective well-being. This includes the relationship between subjective well-being and other important well-being outcomes, such as income and health, as well as the use of subjective well-being data to inform the appraisal, design and evaluation of policy options. Section 3 addresses subjective well-being data as an input for other analyses. First, it considers the use of subjective well-being as an explanatory variable for other outcomes, and then focuses on the potential use of subjective well-being data in cost-benefit analysis.

Section 1 will be of most direct interest to large-scale data producers, such as national statistical agencies, as it concerns the kinds of outputs and analyses that they are most likely to report for a wide range of audiences. Sections 2 and 3 provide a sense of the broader uses of subjective well-being data – which are essential to consider when planning its measurement (as set out in Chapter 3, an approach to *Measuring subjective well-being*). Analyses of drivers, for example, require consideration of the co-variates to be collected alongside subjective well-being data, and ideally call for data from which causal inferences can be drawn. The potential risk of measurement error, and the various biases that may be present in the data, are also major themes throughout the chapter. However, as the relevance of measurement errors depends on the *intended usage* of the data (Frey and Stutzer, 2002), the chapter is organised around data uses, rather than around these sources of error. Key interpretive issues for each type of analysis are summarised in Table 4.1.

Table 4.1. Summarising possible uses of subjective well-being data

Data use	What	Why	Who	Key interpretive issues
1) Complementing existing measures of well-being	Core measures/headline indicators used to examine: <i>i)</i> National trends over time. <i>ii)</i> Distribution of outcomes across different groups within society. <i>iii)</i> Distribution of outcomes across countries. Includes indicators of central tendency or “level”, as well as distribution, and the relative rate of rise or decline over time.	To know if the changes affecting society have an impact on subjective well-being. To identify vulnerable groups and areas of suffering – highlighting where key drivers of subjective well-being may lie – and where there may be opportunities for policy interventions. To conduct international benchmarking, assist in the interpretation of national data, and identify where countries may be able to learn from others’ experiences.	Governments (central, regional, local). Wider public. Public, private and third sector organisations. Researchers interested in country-level drivers of national well-being. Individuals and organisations – e.g. making decisions about where to live and work.	<i>i)</i> What size of difference between groups or over time can be expected? <i>ii)</i> What alternative explanations should be considered for observed differences? <i>iii)</i> What is the role of culture and cultural bias in cross-country comparisons?
2) Better understanding the drivers of subjective well-being	Analyses based on national and international micro-data, with subjective well-being used as the dependent variable, to: <i>i)</i> Examine the relationship between subjective well-being and other important life circumstances, such as income and health. <i>ii)</i> Inform policy options appraisal, design and evaluation. <i>iii)</i> Inform policy trade-offs.	To improve our understanding of well-being overall, by examining the relationship between subjective well-being, life circumstances, and other important well-being outcomes. To highlight areas of policy with the greatest potential to improve subjective well-being, and the life events/circumstances most likely to put subjective well-being at risk. To assist in government decision-making processes, including the allocation of resources and the design elements of policies. To inform the public and employers about the likely drivers of individual subjective well-being, providing better information for individual and organisational decision-making.	Governments. Researchers. Individuals wanting better information to support decision-making. Employers wanting to understand and improve employee well-being.	<i>i)</i> What size of impact can be expected? <i>ii)</i> How can the impacts of different drivers be compared?
3) Subjective well-being as an input for other analyses, particularly cost-benefit analysis	Micro-data on subjective well-being, used as an input for other analyses, including: <i>i)</i> As an explanatory variable for other elements of well-being or behaviour. <i>ii)</i> Used to estimate the value of non-market goods and services, for the purposes of cost-benefit analyses.	To better understand how subjective well-being can contribute to other well-being outcomes and shed light on human decision-making processes, including the various biases that may be present. To provide an alternative to traditional economic approaches to estimating the value of non-market goods, supporting government (and other organisations) in making decisions about complex social choices.	Researchers. Governments. Individuals wanting better information to support decision-making. Employers wanting to understand and improve employee well-being.	<i>i)</i> The sensitivity of subjective well-being data to non-market goods. <i>ii)</i> Measurement error and its impact on valuations. <i>iii)</i> Co-variables to include in regression models. <i>iv)</i> Time horizons for study.

1. Using subjective well-being to complement other outcome measures

Introduction

Subjective well-being is an essential element of a broader and multi-dimensional concept of human well-being and can be used in the context of monitoring reports on the living conditions of different countries or sub-national units. Indicators of interest will include the overall level of subjective well-being, its rate of change over time and its distribution across different groups within society. This section is organised in three parts. The first part addresses what is meant by “measuring well-being” and briefly discusses how subjective well-being can contribute in this area. This includes outlining what subjective well-being data can add to more conventional measures and why subjective well-being might be considered an important outcome in its own right. The second part focuses on reporting measures of subjective well-being. This examines the relative merits of a number of different approaches to summarising and describing subjective well-being

data. The section concludes with discussion of the issues arising in analyses that aim to compare different levels of subjective well-being. This includes consideration of how to interpret observed differences between groups, over time and between different countries – including when there may be a risk of cultural “bias” in the data.

What does “measuring well-being?” mean, and why do it?

Measuring human well-being involves identifying the key components of a good life and then selecting a set of indicators that provide information about the progress of society with respect to these outcomes. There are three key elements of well-being that it will be important to measure: i) trends over time; ii) the distribution of outcomes across different members of society; and iii) the distribution of outcomes across countries.

Measures of well-being are important to governments and to the general public. Although many societal outcomes are often not under direct government control, governments still seek to have a positive influence on well-being and are often called to intervene to address poor outcomes and/or declines in well-being over time. Governments will generally have an interest in all three elements: trends over time; distributions across society; and international benchmarking.

Businesses and voluntary sector organisations may also have an interest in monitoring well-being. National trends influence the business environment. Businesses can play a role in meeting the needs of vulnerable groups and bridging inequalities in society, and they may look to international measures when considering export, expansion and/or relocation opportunities. Voluntary sector organisations may have a strong interest in the distribution of outcomes across society – including what this tells us about vulnerable groups and the support that they may need. Voluntary sector organisations may also look overseas for examples of different practices, and some voluntary organisations will be engaged in international work that seeks to address global inequalities in well-being outcomes.

Approaches to measuring well-being

There are a number of different approaches to monitoring well-being through dedicated reports. GDP per capita is commonly used as a proxy measure for the overall well-being of countries. Other commonly-cited indicators of national progress include poverty, unemployment levels, infant mortality, life expectancy, educational attainment, crime figures and air quality. These provide information on outcomes that may not be accurately captured by GDP per capita, but which are important to well-being.

Nonetheless, it can be difficult to compile a coherent overall picture of well-being from a disparate range of measures. One approach is to develop composite indices, such as in the UN’s Human Development Index (HDI),¹ which combines information on life expectancy at birth, mean years of schooling, expected years of schooling and gross national income per capita, to produce a single overall figure. Alternatively, a range of indicators can be presented in a “dashboard”, such as that adopted in *How’s Life?* (OECD, 2011a) or the various sets of sustainable development indicators available, such as those in the EU sustainable development strategy (Eurostat, 2009), or *Measuring New Zealand’s Progress Using a Sustainable Development Approach* (Statistics New Zealand, 2008).

The role of subjective well-being in measuring well-being

What people feel about their lives matters. A nation of materially wealthy, healthy, but miserable citizens is not the kind of place where most people would want to live. The available evidence suggests that the general public, at least in affluent countries, do regard subjective well-being as an important component of national well-being overall. For example, Dolan and Metcalfe (2011) report an initial survey asking UK respondents to rank seven ways of measuring progress, in which “people’s happiness” was ranked behind the “state of the economy” and “peoples’ health”, but above “crime rates”, “education levels”, “the environment” and “depression rates”.

A recent public consultation by the UK Office for National Statistics (ONS, 2011a) found that 79% of 6 870 respondents endorsed “life satisfaction” as a measure of “national well-being and how life in the UK is changing over time” – second only to “health statistics” (80%), with measures such as “income distributions” endorsed by 62%, and “economic measures such as GDP” endorsed by just 30% of respondents. The OECD’s web-based interactive tool *Your Better Life Index* offers individuals the opportunity to create their own international well-being index, rating the importance of 11 different dimensions of well-being on a 1-5 scale. Ratings shared by around 4 000 users of the website (OECD, 2011b) indicate that life satisfaction is the domain most often ranked the highest (with over 10% of users identifying it as the most important domain), closely followed by health, education, the environment and work-life balance.²

One benefit of using subjective well-being to complement existing measures of national progress is that it emphasises the views of individuals. It thus presents an overall picture of well-being that is grounded in people’s preferences, rather than in *a priori* judgements about what *should* be the most important aspects of well-being. Subjective well-being measures reflect the unique mix of factors that influence an individual’s feelings and assessments. This is not to say that subjective well-being should *replace* other important economic, social and environmental indicators, but it does provide a useful and easy-to-understand complement to existing measures, because it can indicate the *combined impact* of life circumstances on subjective perceptions and emotions.

Subjective well-being measures may also capture some aspects of well-being that are difficult to otherwise observe or quantify through more traditional measures. An example of this, cited in Chapter 1 (Box 1.2), is the marked decline in evaluative measures of subjective well-being in Egypt and Tunisia in the years preceding the 2011 “Arab Spring”. Conventional indicators of progress, such as economic growth, and the UN’s Human Development Index, continued to rise during this period – thus failing to detect an important social trend.

The public policy applications of subjective well-being measures (described in Chapter 1) are wide-ranging. Extensive reviews on this topic have been published recently by Diener, Lucas, Schimmack and Helliwell (2009), Bok (2010), the European Commission (Chapple et al., 2010), and the New Economics Foundation (Stoll, Michaelson and Seaford, 2012). These reviews build on the earlier conceptual work of Kahneman et al. (2004), Layard (2005), Dolan and White (2007) and Krueger (2009), to name just a few. Specific examples from the field include using life satisfaction and eudaimonic indicators alongside a wide variety of outcome measures to evaluate public projects to enhance well-being, such as the UK Big Lottery Fund well-being evaluation (CLES Consulting and NEF, 2011); and the evaluation of the Community Employment Innovation Project in Canada (Gyarmati et al., 2008), as well as for cost-benefit analyses of psychological therapy (Layard et al., 2007),

estimating the well-being impact of various policy-relevant daily activities, such as commuting (Kahneman and Krueger, 2006; Stutzer and Frey, 2008), as well as to explore policy trade-offs, such as those between inflation and unemployment (Di Tella, MacCulloch and Oswald, 2001) or income and airport noise (Van Praag and Baarsma, 2005). Research linking subjective well-being, and particularly positive affect, to health outcomes (Pressman and Cohen, 2005; Danner, Snowdon and Friesen, 2001; Cohen et al., 2003; Kiecolt-Glaser et al., 2002; and Steptoe, Wardle and Marmot, 2005), as well as income, employment outcomes and productivity (Diener et al., 2002; Wright and Staw, 1999; Keyes 2006; Clark and Oswald, 2002) also suggests a public interest in monitoring such measures.

Like many other measures of well-being, however, subjective well-being data do come with some notable caveats and trade-offs, specifically around data comparability and the risk of measurement error (Ravillion, 2012; see Chapter 2 for a summary). Some of these risks are common to other self-report measures, including the risk of various response biases, and the impact that both question wording and response formats can have on how people answer questions. Frame-of-reference effects³ and adaptation⁴ to life circumstances over time can also potentially influence the levels of subjective well-being observed among different populations and population sub-groups, as well as the nature of the relationship between subjective well-being and its determinants. These issues mean that subjective well-being data, like most self-reported data, need to be interpreted with care and should be used to *complement* rather than replace other indicators of well-being. Interpretive issues are described at length in the sections that follow.

Reporting subjective well-being data

Using subjective well-being data to complement other measures of well-being requires producers of statistical information to regularly collect and release high-quality nationwide data from large and representative samples. Key audiences include policy-makers, public service providers, private businesses and voluntary sector organisations, researchers and the wider public – all of whom may have an interest in whether, where and when conditions in society are improving. For monitoring exercises in particular, it is important that the figures released *mean something* to the general public, as well as to more specialist audiences (New Economics Foundation, 2009).

Many of these audiences will not read statistical releases directly, but rather will rely on how these are reported in a variety of media. It is therefore important to consider how to package the data in a concise yet precise manner to ensure that the necessary information can be easily located and conveyed with accuracy by other parties.

The language used to describe measures is also important. The term “happiness” is often used as convenient shorthand for subjective well-being, in both popular media and parts of the academic literature – not least because *happiness* may be more attention-grabbing and intuitively appealing. The key risk surrounding the term “happiness” is conceptual confusion: whilst the experience of positive emotion (or positive affect) is an important part of subjective well-being, it represents only part of the over-arching concept, and the term “happiness” underplays the evaluative and eudaimonic aspects of subjective well-being as well as the experience of negative affect (pain, sadness, anxiety, etc.), all of which may be of interest to policy-makers.⁵ We therefore recommend against describing results only in terms of “happiness”, particularly for data releases from national statistics agencies.

Several authors have also shown a tendency to drop the term “subjective” from their reporting, simply describing results in terms of “well-being”. This is also a potential source of confusion. Whilst subjective measures of well-being offer an important insight into respondents’ views about their own well-being, the OECD regards subjective measures as only one of several measures required to develop a balanced view of well-being overall (OECD 2011a; Stiglitz, Sen and Fitoussi, 2009). This concurs with the outcome of the UK ONS’s recent public consultation on what matters for measuring national well-being (ONS, 2011a). For both the ONS and OECD, measuring well-being requires a mix of subjective and objective indicators, and measures across a variety of other dimensions (e.g. education, health, income and wealth, social connections and the environment, to name just a few) are viewed as an essential part of the overall well-being picture.

These considerations mean it will be important, especially when reporting the results of national surveys, to provide a full description of the indicators used – including the underlying constructs of interest, and what they might reflect in addition to “happiness”. This could be accompanied by a brief explanation of the rationale for measuring subjective aspects of well-being and their role in complementing (rather than *replacing*) other well-being indicators. Chapter 1 also discusses these issues.

For the purposes of high-level communication about subjective well-being results, particularly with non-specialist audiences, it is desirable to identify a small set of key measures and figures. These guidelines recommend that this set should include one primary measure of life evaluation and its dispersion, as well as a limited number of affect measures if possible (see Chapter 3). Eudaimonia and domain-specific life evaluations may also be of interest, although, as multi-dimensional constructs, they can be more challenging to convey in single headline figures. There are several different ways in which current levels of subjective well-being data can be presented for the purposes of monitoring progress – and the choice of method should ultimately be driven by user need and demand. Recent examples are available from France’s National Institute of Statistics and Economic Studies (INSEE – Godefroy, 2011) and the UK’s Office for National Statistics (ONS, 2012). Chapter 3 provides recommendations for the basic output associated with the different question modules proposed as part of these guidelines.

Because of the range of possible approaches to presenting and reporting on subjective well-being data, it is useful to consider the issue within some sort of organising framework. At the most general level, the question of how to report subjective well-being data for the purposes of monitoring progress has four elements:

- How to report *central tendency and level*.
- How to report *distribution*.
- Whether and how to *aggregate* responses.
- How to report *change over time* and *differences between groups*.

Reporting central tendency and level

The most fundamental information to report with respect to subjective well-being is the level of the outcome. This can be thought of as addressing the issue of “how high or low is the level of subjective well-being in the population under consideration?”. There are three main approaches to describing the level of either single-item or summed multi-item aggregate measures. First, the frequency of responses can be described by category: this involves presenting the proportion of the population that select each response category of

the subjective well-being scale used. Second, the data can be summarised in relation to one or more thresholds. This involves reporting the proportion of the population with a level of subjective well-being above or below a particular threshold level. Finally, the data can be summarised via some measure of central tendency, such as the mean, median or mode. Each of these three approaches has its own strengths and weaknesses.

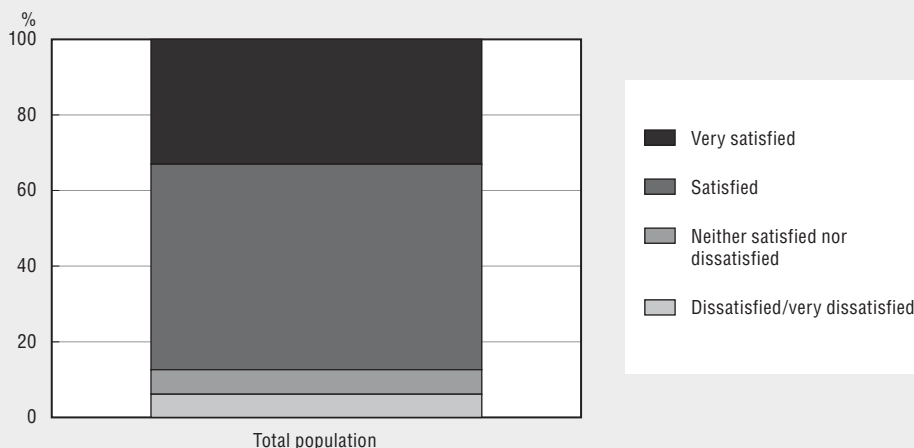
Reporting the proportion of respondents selecting each response category is the method that requires the data producer to make the fewest decisions about presentation. Such an approach has some significant strengths with respect to information on subjective well-being. Because the entire distribution is described, no information is lost. Also, a presentation by category respects the ordinal nature of subjective well-being data⁶ and requires no assumptions about the differences among ordinal categories (i.e. there is no assumption that the difference between a 3 and a 4 is the same as that between a 7 and an 8).

However, presenting the whole distribution of responses for each measure also has significant draw-backs. In particular, for a non-specialist audience it is difficult to directly compare two distributions of this sort and reach judgements about which represents a higher or lower “level” of well-being – although non-parametric statistical tests are available for these purposes. While reporting the whole distribution may be a viable strategy where the number of response categories is relatively limited (e.g. example shown in Box 4.1), as the number of categories increase it becomes more difficult to reach overall judgements from purely descriptive data.

Box 4.1. Reporting on the proportion of respondents by response category

Statistics New Zealand publishes a number of measures of subjective well-being in the statistical releases for the biannual New Zealand General Social Survey. These include overall life satisfaction and satisfaction with particular aspects of life, namely financial satisfaction and a subjective assessment of health status. In all cases a five-point labelled Likert scale is used for responding to the questions. Although such a measure is sub-optimal in many respects, it lends itself well to being presented as a proportion of respondents by response category (Figure 4.1).

Figure 4.1. Reporting the proportion of respondents selecting each response category



Source: Statistics New Zealand, *New Zealand General Social Survey*.

One way to manage a large number of scale responses is to report on the *proportion of responses falling above or below a given threshold, or set of thresholds*. For example, responses can be reported as the percentage of the sample falling above or below a certain cut-off point, or banded into “high”, “medium” and “low” categories (Box 4.2). Threshold descriptions of the data can be grasped quickly – providing an anchor for interpretation, and offering a way of

Box 4.2. Output presentation examples – threshold-based measures

The Gallup-Healthways Life Evaluation Index classifies respondents as “thriving”, “struggling”, or “suffering”, according to how they rate their current and future lives (five years from now) on the Cantril Ladder scale with steps numbered from 0 to 10, where “0” represents the worst possible life and “10” represents the best possible life. “Thriving” respondents are those who evaluate their current state as a “7” or higher and their future state as “8” or higher, while “suffering” respondents provide a “4” or lower to both evaluations. All other respondents are classified as “struggling”. Table 4.2 shows thriving struggling and suffering in the EU.

Table 4.2. **Gallup data on thriving, struggling and suffering in the EU (sorted by percentage suffering)**

Column 1	% thriving	% struggling	% suffering	% thriving minus % suffering (pct. pts)
Bulgaria	5	50	45	-40
Romania	18	54	28	-10
Hungary	15	57	28	-13
Greece	16	60	25	-9
Latvia	16	61	23	-7
Portugal	14	65	22	-8
Estonia	24	60	17	7
Poland	23	60	17	6
Lithuania	23	57	16	7
Slovenia	32	53	14	18
Germany	42	52	6	36
Czech Republic	34	53	13	21
Slovak Republic	27	61	12	15
Malta	34	55	11	23
Spain	39	54	7	32
Cyprus	44	49	7	37
Italy	23	71	6	17
United Kingdom	52	44	6	46
Ireland	54	43	4	50
France	46	50	4	42
Austria	59	38	3	56
Finland	64	34	3	61
Denmark	74	24	2	72
Luxembourg	45	54	1	44
Netherlands	66	33	1	65

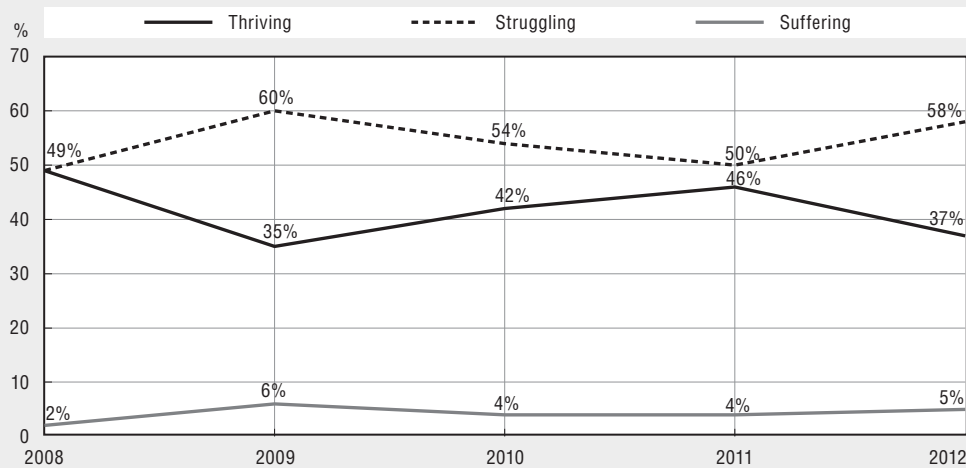
Note: Data collected between March and June 2011. Data unavailable for Sweden and Belgium at time of publishing.

Source: Gallup World web article by Anna Manchin, 14 December 2011, “More suffering than thriving in some EU countries”, www.gallup.com/poll/151544/Suffering-Thriving-Countries.aspx.

Box 4.2. **Output presentation examples – threshold-based measures (cont.)**

Change in subjective well-being over time can also be presented relative to a given threshold (Figure 4.2).

Figure 4.2. **Share of the French population classified as “thriving”, “struggling” and “suffering”**



Source: Gallup World web article by Anna Manchin, 4 May 2012, “French Adults’ Life Ratings sink in 2012”, www.gallup.com/poll/154487/French-Adults-Life-Ratings-Sink-2012.aspx.

communicating something about the distribution of the data with a single figure. The use of thresholds is also consistent with the ordinal nature of much subjective well-being data, as it requires no assumptions about the cardinality of scale responses.

The downsides of threshold measures include losing some of the richness of the data,⁷ and the risk of encouraging a distorted emphasis on shifting people from just below to just above a threshold. This is a particular risk if only one threshold (e.g. “6 and above”) is used, because it may be important for policy-makers in particular to understand what characterises communities at both high and low ends of the subjective well-being spectrum. Although thresholds have the potential to be more sensitive to change when carefully selected around the area of greatest movement on the scale, there is a considerable risk that a threshold positioned in the *wrong* part of the scale could mask important changes in the distribution of the data. For example, if the risk of clinically-significant mental health problems is greatest for individuals scoring 5 or less on a 0-10 life evaluation measure, setting a threshold around 7 could lead to a failure to identify changes that could have significant consequences for policy. In addition, reporting based on thresholds runs the risk of presenting two very similar distributions as quite different, or vice versa. For example, for some countries the distribution of subjective well-being is bi-modal, while for others there is a single mode. Depending on where a threshold is set, two such distributions might be presented as very different, or essentially the same. The central difficulty, therefore, lies in identifying *meaningful* threshold points that have real-world validity.

Thresholds can be set through examining the underlying distribution of the data and identifying obvious tipping points, but this data-driven approach limits both meaningful interpretation (what is the real-world meaning of a data cliff?) and comparability among

groups with different data distributions, whether within or between countries. A more systematic approach may be to adopt something similar to a “relative poverty line”, whereby individuals falling, for example, below half of the median value on a scale are classified as faring badly. This capitalises on thresholds’ ability to convey distributional characteristics, but has the downside of conveying relatively little about the average level, which is essential for both group and international comparisons.⁸ It also remains an essentially arbitrary method for identifying a threshold. The final option would be to select an absolute scale value below which individuals demonstrate a variety of negative outcomes (and an upper bound associated with particularly positive outcomes), based on the available empirical evidence. This would at least give the threshold some real-world meaning.

Blanton and Jaccard (2006) make a strong case for linking psychological metrics to meaningful real-world events, highlighting the risk of assigning individuals to “high”, “medium” and “low” categories without justifying or evidencing what these categories mean in practice. In particular, they note the conceptual and practical problems associated with the intuitively appealing practice of “norming”, i.e. setting threshold values based on the proportion of the sample falling above or below that threshold. For example, in an obesity reduction programme, if an individual’s weight loss result was described as “high” because relative to others in the group they lost more weight, the clinical significance of the finding remains obscured: it is possible that *everyone* in the group lost a clinically significant amount of weight, or *no-one* in the group lost a clinically significant amount. In both of these scenarios, what matters is not how the individual fares relative to the rest of the sample, but how their weight loss is likely to relate to other health outcomes. There is a clear analogy here with both relative poverty lines and international comparisons of subjective well-being: what would be categorised as “high” life satisfaction by normative standards in Denmark will be quite different to “high” life satisfaction according to normative standards in Togo – making these two categorisations impossible to compare. This emphasises the challenges associated with setting suitable thresholds and suggests *against* emphasising threshold-based measures too strongly in data releases. Given the wide range of potential uses of the data, a wide range of thresholds may be also relevant to policy-makers and others.⁹

Summary statistics of central tendency provide a useful way of presenting and comparing the level of subjective well-being in a single number. The most commonly-used measures of central tendency are the mean, the mode and the median. However, due to the limited number of scale categories (typically no more than 0-10), the median and modal values may lack sensitivity to changes in subjective well-being over time or to differences between groups. The mean is therefore generally more useful as a summary statistic of the level of subjective well-being.

Although the mean provides a good summary measure of the level of subjective well-being, it has shortcomings. First, the use of the mean requires treating the data from which it is calculated as cardinal. Although most subjective measures of well-being are assumed to be ordinal, rather than cardinal,¹⁰ evidence suggests that treating them as if they were cardinal in subsequent correlation-based analysis does not lead to significant biases: the practice is indeed common in the analysis of subjective well-being data, and there appear to be few differences between the conclusions of research based on parametric and nonparametric analyses (Ferrer-i-Carbonell and Frijters, 2004; Frey and Stutzer, 2000; Diener and Tov, 2012). That said, Diener and Tov also note that when it comes to simpler analyses, such as comparisons of mean scores, ordinal scales that have been

adjusted for interval scaling using Item Response Theory can produce different results to unadjusted measures (p. 145). Second, the mean can be strongly affected by outliers and provides no information on the distribution of outcomes. Both of these issues therefore highlight the importance of complementing the mean with information on the distribution of data.

Distribution

It is also important to present information on the distribution of responses across the different response categories. If the primary way of presenting the data is by reporting the proportion of responses falling in each response category, the need for separate measures of distribution is less important. If, however, reporting is based on thresholds or summary statistics of central tendency, specific measures of distribution are important. The choice of distributional measure will depend partly on whether the data is treated as ordinal or cardinal.

When cardinality is assumed, it is possible to use summary statistics of distribution such as the Gini coefficient. Both the Gini coefficient and the standard deviation are based on calculations that are unlikely to hold much meaning for the general public, and may therefore be less effective as a tool for public communication. The Gini in particular also perhaps has less *intuitive meaning* for subjective well-being than it does for its more traditional applications to income and wealth.¹¹ This means that other measures of dispersion, such as the interquartile range (i.e. the difference between individuals at the 25th percentile and individuals at the 75th percentile of the distribution), or the point difference between the 90th and the 10th percentile (Box 4.3), may be preferred in simple data releases. Where space allows, graphical illustrations of distribution are likely to be the most intuitive way to represent distributions for non-specialist audiences, although such graphs can be difficult to compare in the absence of accompanying summary statistics.

Aggregation of multi-item measures

Where a survey includes more than one question about subjective well-being, a key reporting decision for data producers will be whether to report responses to each question separately, or alternatively to aggregate some questions into broader multi-item measures. Single-item life evaluation questions are most often reported as stand-alone headline measures.¹² However, in addition to the single-item life evaluation primary indicator, the suite of question modules proposed in Chapter 3 also includes several multi-item measures intended to capture evaluative, affective (or hedonic), eudaimonic and domain-specific aspects of subjective well-being.

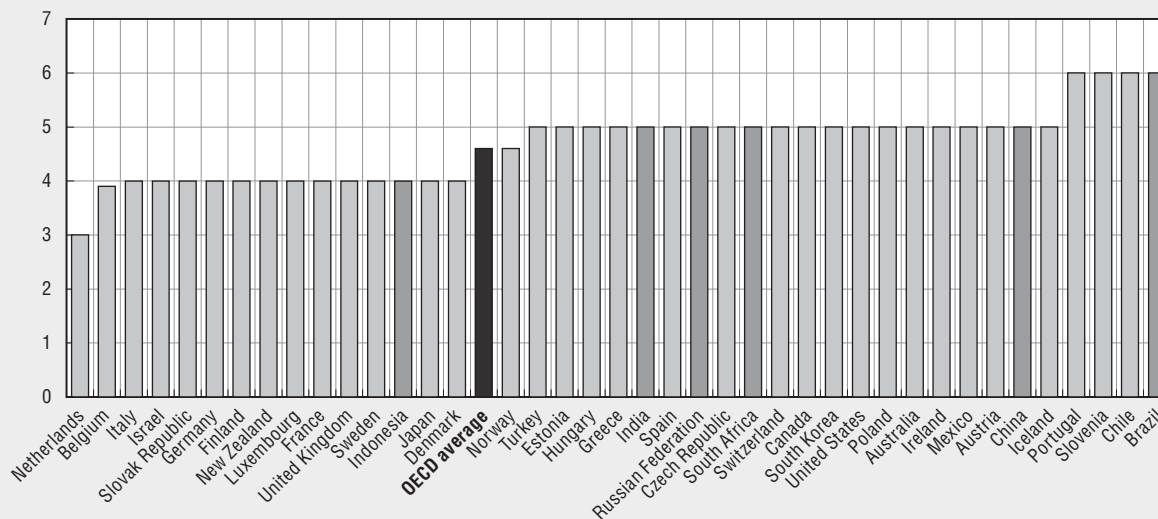
Although there may be value in looking at responses to individual questions or scale items in more detailed analyses, it is desirable to summarise longer multi-item measures, particularly for the purposes of reporting outcomes to the general public. Furthermore, summing responses across multiple items should generally produce more reliable estimates of subjective phenomena, reducing some of the impact of random measurement error on mean scores – such as may result from problems with question wording, comprehension and interpretation or bias associated with a single item. However, whilst summing responses across different life evaluation questions should pose relatively few problems, affect and eudaimonia are by nature more multidimensional constructs, and thus there is a greater risk of information loss when data are aggregated.

Box 4.3. Distribution of subjective well-being among OECD and emerging countries (OECD, 2011a)

In *How's Life?*, the OECD used the gap between the 10th and 90th percentiles as a measure of distribution (the “90/10 gap”). Conceptually similar to the interquartile range, the 90/10 gap was used because the clustered nature of life satisfaction responses meant that the interquartile range provided little to distinguish between countries.

Figure 4.3. **Inequality in life satisfaction in OECD and emerging economies, 2010**

Point difference between the 90th percentile and the 10th percentile



Source: Gallup World Poll data, reported in *How's Life?* (OECD, 2011a).

Options for aggregation, specific to each scale, include:

- *Positive and negative affect*: Where several items are used to examine experienced affect, most scales are designed such that one can calculate positive and negative affect subtotals for each respondent, summarising across items of similar valence. For example, in the core affect measure proposed in Module A of Chapter 3, positive affect is calculated as the average score (excluding missing values) for questions on “enjoyment” and “calm”, and negative affect is calculated as the average score for questions on “worry” and “sadness”. As with any summary measure, this risks some degree of data loss, particularly where affect dimensions can be factored into one or more sub-dimensions – for example, the high-arousal/low-arousal dimensions identified in the Circumplex model of mood (Russell, 1980; Russell, Lewicka and Niit, 1989; Larsen and Fredrickson, 1999). However, for the purposes of high-level monitoring of affect, examining summary measures will be more feasible than looking at each affect item individually, and the increased reliability of multi-item scales will be advantageous.
- *Affect balance*: Positive and negative affect measures can be further summarised into a single “affect balance” score for each respondent by subtracting the mean average negative affect score from the mean average positive affect score. This can then in turn be reported as either a mean score (positive minus negative affect) or as a proportion of the population with net positive affect overall.

- Where information is available on the frequency of positive and negative affect experiences throughout the day, such as that provided by time-use studies, it is also possible to calculate the proportion of time that people spend in a state where negative affect dominates over positive affect. This is described as the “U-index” (Kahneman and Krueger, 2006), and again this can also be reported at the aggregate population level. Time-use data also enable the mean affect balance associated with different activities to be described (Table 4.3).

Table 4.3. **Mean net affect balance by activity, from Kahneman et al. (2004)**

Activity	Percentage of sample	Time spent (hours)	Net affect ¹
Intimate relations	11	0.21	4.74
Socialising after work	49	1.15	4.12
Dinner	65	0.78	3.96
Relaxing	77	2.16	3.91
Lunch	57	0.52	3.91
Exercising	16	0.22	3.82
Praying	23	0.45	3.76
Socialising at work	41	1.12	3.75
Watching TV	75	2.18	3.62
Phone at home	43	0.93	3.49
Napping	43	0.89	3.27
Cooking	62	1.14	3.24
Shopping	30	0.41	3.21
Computer at home	23	0.46	3.14
Housework	49	1.11	2.96
Childcare	36	1.09	2.95
Evening commute	62	0.62	2.78
Working	100	6.88	2.65
Morning commute	61	0.43	2.03

1. Net affect is the average of three positive adjectives (enjoyment, warm, happy) less the average of five negative adjectives (frustrated, depressed, angry, hassled, criticised). All the adjectives are reported on a 0-6 scale, ranging from “not at all” to “very much”. The “time spent” column is not conditional on engaging in the activity. The sample consists of 909 employed women in Texas.

Source: Kahneman, Krueger, Schkade, Schwartz and Stone (2004), Figure 2, p. 432.

Both affect balance and the U-index are similar to threshold-based measures, but ones that have both clear meaning and the considerable advantage of reducing affect data to a single variable. However, there is some risk of data loss in adopting these aggregation approaches, particularly when exploring group differences. For example, the ONS subjective well-being data release (ONS, 2012) found that, for most age groups, on average women reported slightly higher happiness yesterday than men, but they also reported higher anxiety yesterday. If aggregated as an affect balance measure, these differences may not be detectable.

Ultimately, the judgement of the most appropriate measure should be driven by the primary data use. For overall monitoring, the benefits of reporting affect balance are likely to outweigh the drawbacks – but when attempting to understand, for example, the links between affect and health outcomes, it may be more important to examine dimensions of affect separately (Cohen and Pressman, 2006).

Eudaimonia: Most of the literature regards eudaimonia as a multidimensional construct (e.g. Huppert and So, 2011; Ryff, 1989; Ryan and Deci, 2001), and therefore summarising across all items on a multi-item scale again risks some data loss. For detailed analysis, it may be important to examine each sub-component of eudaimonia separately, at least initially. Nonetheless, for the purposes of monitoring well-being, if positive correlations are found between each of the sub-dimensions, it may be appropriate to sum across items.¹³ The first option is to take the mean average value of all responses, omitting missing values. Alternatively, a threshold-based approach has been proposed by Huppert and So (2011; see Box 4.3), which categorises respondents according to whether they meet the criteria for “flourishing”. The “flourishing” construct may offer a powerful communicative device. However, partly because it is based on groups of items with different numbers and different response categories, Huppert and So’s operational definition of “flourishing” ends up being quite complex (with different thresholds being applied to differentially distributed data, and different items grouped according to various subscales assumed to be present). As noted earlier, the present difficulty with threshold-based measures is that there is little consensus on where the meaningful cut-off points lie. Further research is therefore needed before this approach can be regarded as preferable to reporting mean average scores.

Domain satisfaction: Questions about satisfaction with individual domains of life can be meaningful as stand-alone measures, and may be particularly useful for policy-makers seeking specific information on the effects of a particular policy intervention. However, some sets of domain-specific questions have been designed with a view to creating a composite measure of life evaluation overall, by summing responses across each of the domains (e.g. the Australian *Personal Wellbeing Index* – International Wellbeing Group, 2006, in Module E, Chapter 3). This overall approach requires making strong assumptions about the weights to apply to each life domain (as well as the universality with which those weights apply across the population) along with some judgements about which domains of life are relevant to subjective well-being overall. In the case of the *Personal Wellbeing Index*, domains have been selected as the most parsimonious list for capturing “satisfaction with life as a whole”, and equal weights are adopted for each domain, in recognition of the fact that empirically-derived weights may not generalise across data sets. These assumptions notwithstanding, composite measures of domain satisfaction may offer a more rounded and potentially more reliable picture of life “as a whole”, as respondents are encouraged to consider a variety of different aspects of life when forming their answers.

Aggregating several subjective well-being indicators into an overall index

Although the various subcomponents of subjective well-being (e.g. life evaluation, eudaimonia and affect) will convey most information when measured and reported separately, there may be demand for aggregating these into a single over-arching index of subjective well-being, particularly for the purposes of high-level communication and monitoring¹⁴ (see Stiglitz, Sen and Fitoussi, 2009, for a detailed discussion of aggregation issues in relation to well-being indicators). Where there is pressure to report just one overall headline measure, selecting only one element of subjective well-being (such as life evaluations) may neglect other important components, making aggregation across life evaluations, eudaimonia and affect an attractive prospect to those who wish to see all three components of subjective well-being reflected in headline measures.

The communication advantages in reducing different measures of subjective well-being to one number must, however, be set against a number of methodological objections. Most fundamentally, the different aspects of subjective well-being (life evaluation, affect, eudaimonia) represent distinct constructs, and it is not clear that it is possible to provide a coherent account of what an aggregate index of overall subjective well-being actually represents. Similarly, there is no clear basis for determining the relative weights to assign to different dimensions or sub-dimensions of subjective well-being. This problem is analogous to those encountered when trying to develop composite measures for other well-being outcomes, such as health or skills. Until further consideration has been given to how composites could be created, the most sensible approach may be for data producers to provide disaggregated measures – enabling users to experiment and create their own composite indices as necessary. In the meantime, where single headline figures are to be reported, life evaluations are likely to remain the focus, because they are currently the most established of the three measures in terms of their use to complement existing measures of well-being (see Chapter 1).

Reporting change over time and differences between groups

National levels of subjective well-being are difficult to interpret when examined in isolation. External reference points are essential in order to understand whether a mean life satisfaction score of 7.2 is “good”, or not. In order to interpret current observations of subjective well-being, two broad comparisons are likely to be of interest to data users: 1) comparisons between current and previous levels of subjective well-being; and 2) comparisons between different countries, particularly those regarded as peers in terms of their overall levels of development.

A third type of comparison involves examining group differences *within* a country. Identifying groups of individuals who report lower or higher subjective well-being, or whose well-being is changing at a faster or slower rate over time, is an essential use of national statistics. Defining reference groups for such comparisons is also important – and by providing information about the level of subjective well-being across the whole population, national statistics provide a baseline against which population sub-groups can be compared. Further breakdowns in national statistics (such as by age, gender, education, region, occupation, socio-economic and employment status, health status, etc.) can also enhance their usefulness. Understanding what characterises communities at both high and low ends of the subjective well-being spectrum will be important for policy users seeking both to reduce extreme suffering and to better understand how high levels of subjective well-being can be achieved.

Examining whether gaps in subjective well-being between groups within society are growing or shrinking is also important. Central and local governments, the wider public sector, researchers and voluntary organisations may be particularly interested in inequalities in subjective well-being in order to assist the identification of vulnerable groups who may benefit from specific interventions.

Comparisons over time and between groups, both within and across countries, can also signal where to look in terms of the potential *drivers* of subjective well-being. For example, if regional differences in subjective well-being are identified, looking at other variables which differ across regions can have implications for better understanding what matters for subjective well-being. This will be of interest to government and researchers, but also to members of the public and the organisations that they work for.

Methods for reporting change over time and differences between groups

There are also several ways in which comparisons over time and between groups can be reported. The first step involves basic descriptive statistics. These include tracking mean changes in time series, calculating changes in the mean score between time points, examining absolute or percentage differences between groups, and looking at group differences over time or relative to a given threshold (see Figures 4.2 and 4.3, Box 4.2). Changes in the overall distribution of subjective well-being over time are also of interest, as they can indicate whether society as a whole is becoming more or less equal in terms of people's experiences of subjective well-being. Finally, differences in the rate and direction of change both between groups within societies and between countries more broadly may also be important.

Data users will find summary statistics easier to understand if there is some degree of consistency between the methods used for reporting current levels of subjective well-being and those used for reporting change over time or comparisons between groups. Thus, if current levels are described using the mean, ideally change over time should also be reported on this basis. Once again, threshold-based estimates offer both advantages and disadvantages. Ease of communication and sensitivity to changes around the threshold level come at the cost of failing to detect changes or differences elsewhere in the scale.¹⁵ Although the overall information loss can in theory be managed through careful selection of the threshold value (and potentially through multiple thresholds), it is not obvious where that threshold should be drawn. Selecting cut-offs according to the distribution of the data could result in setting different thresholds for different population groups and/or different countries, making comparisons impossible.

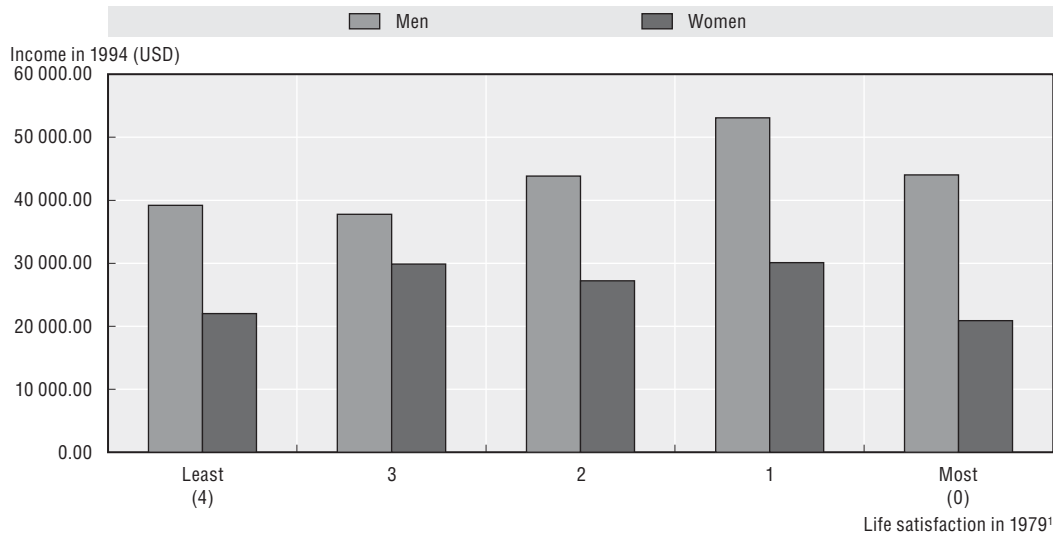
A number of factors can make comparisons of basic descriptive statistics challenging to interpret: for example, differences in sample sizes, or the variability of the data, can make simple comparisons between summary statistics misleading. Thus, both the sample size and standard errors (i.e. the standard deviation of the sampling distribution of a statistic) should be considered when comparing two or more different observations – whether over time or between groups. Robust estimates of standard errors require large and representative samples: when sample sizes are small, standard errors can be larger and the risks of false inferences greater.

One approach is to ensure that whenever group means are reported, both the group sample size and the standard deviations are reported alongside. To assist in the interpretation of standard errors, it may be preferable to display these graphically, for example through the use of box-plots or of error bars added to charts comparing mean levels (Figure 4.4), or simply by providing bar charts to show any differences in the distribution of the data in each group. Finally, statistical inference testing offers a way to examine the likelihood that the observed difference between two values would occur by chance – taking both sample size and standard errors into account.

Analysing and interpreting descriptive subjective well-being data

Almost all analysis associated with monitoring progress will be concerned with examining differences between observations. Whilst statistical analyses can provide a sense of the statistical significance of an observed difference, they cannot indicate the practical significance of a finding – both in terms of its overall size (*is this difference big enough*

Figure 4.4. **Australian longitudinal study: Life satisfaction and income**
(Oishi et al., 2007)



Note: Data drawn from the Australian Longitudinal Study. Data are from men and women surveyed in 1979 (for life satisfaction) and again in 1994 (for income). Along the x-axis, respondents (men and women) are grouped according to the level of life satisfaction reported on a 5-point scale, from most satisfied (0) to least satisfied (4). The y-axis shows mean average reported income in 1994 for each of the respondent groups. Error bars represent standard error. 1. 0 = most satisfied; 4 = least satisfied.

Source: Oishi et al., *Perspectives on Psychological Science*, 2007, No. 2, pp. 346-360.

to matter?) and the possible source of the difference observed (*how can we know if the difference is genuine?*). The section that follows examines three central issues when seeking to interpret patterns of subjective well-being over time or among groups:

- What size of difference can we expect to see?
- What alternative explanations should be considered for the observed differences?
- What is the role of culture in international comparisons, and can data be “corrected” for “cultural bias”?

For subjective well-being to be a useful complement to other measures of well-being, it needs to reflect changes in the things that matter to people. While there is clear evidence that life circumstances have a significant impact on subjective well-being levels, the average measures for countries generally appear to change very slowly over time, and sometimes only by small amounts in response to quite substantial events. In contrast, differences between countries can sometimes appear large relative to what is known about how those countries differ solely on economic measures of well-being.

Even when making very simple comparisons over time, among groups, or among countries, it is important to consider the possible drivers and alternative explanations for observed differences over time or among groups. One particular source of concern is the potential for cultural “bias” to influence cross-country comparisons.

What size differences can be expected?

Despite the wide variety of factors that can limit the size of differences observed in subjective well-being data (considered below), evidence clearly shows that measures can and do change in response to life circumstances (Veenhoven, 1994; Lucas, 2007a, 2007b; Lucas, Clark, Georgellis and Diener, 2003; Diener, Lucas and Napa Scollon, 2006). The expected size of differences in subjective well-being measures, however, varies depending on the context of the analysis. What might be considered a “medium-sized” difference between two population sub-groups within a country would constitute a “large” change in the average level within a country over time, but only a “small” difference between countries.

In addition to considering differences in the mean levels of subjective well-being, it can be valuable to consider the standard errors associated with estimates for different groups, for different observations over time and for different countries. This is currently overlooked in a number of reports, but is important both to understand the likely robustness of mean differences and to better indicate the distribution of the data. The inequality of subjective well-being within groups and across society can be an important indicator, and evidence also suggests that individuals’ subjective well-being can vary considerably in response to certain life events, such as disability (Diener, Lucas and Napa Scollon, 2006; Schulz and Decker, 1985). This makes the standard errors of mean estimates particularly relevant.

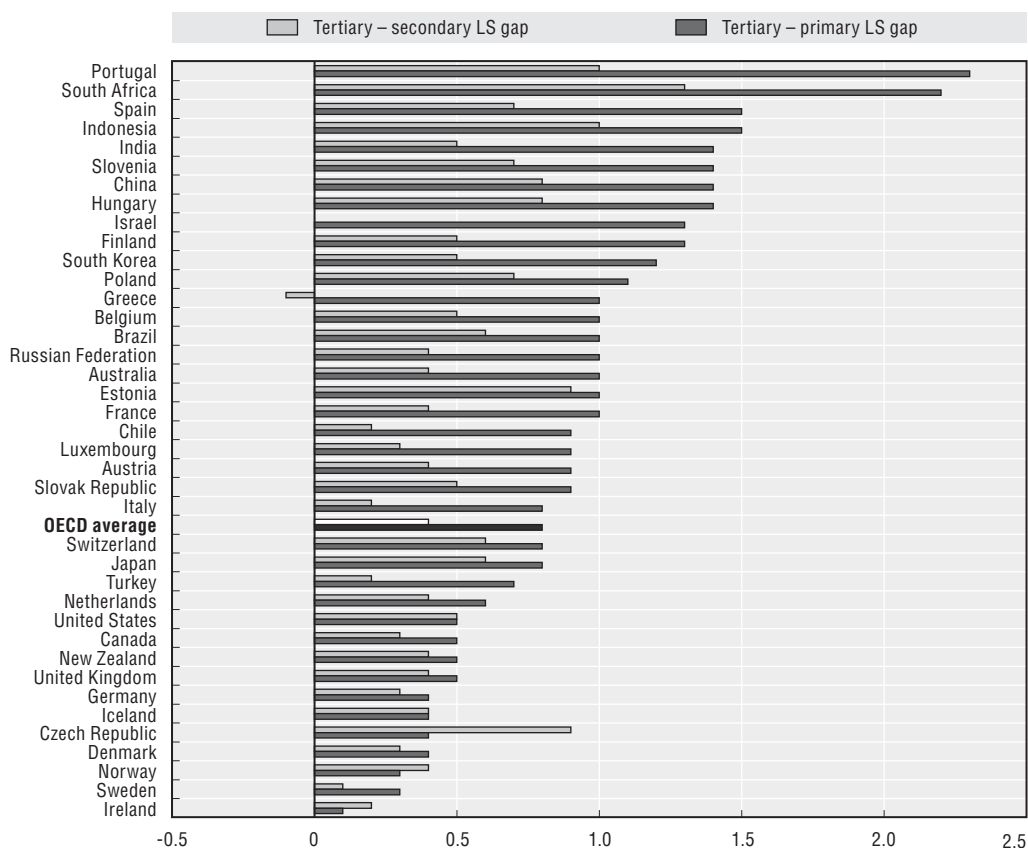
Differences among groups. Within affluent countries, simple mean differences in the range of around 0.5 to 2 scale points on a 0-10 scale (a 5-20% difference) have been detected among different population sub-groups on life evaluation, eudaimonia and affect measures. For example, analysis of experimental subjective well-being data collected from a nationally-representative sample of 80 000 United Kingdom adults in 2011 (ONS, 2012) found differences between employed and unemployed respondents of around 1 scale point in response to life evaluation and eudaimonia questions, and around half a scale point in response to “happy yesterday” and “anxious yesterday” questions (all measured on a 0-10 scale). Similar magnitude differences were observed between those married or in a civil partnership, and those who were divorced or separated – although mean affect differences were closer to 1 scale point between these groups.

Health is another important component of well-being overall, and reductions in subjective well-being have also been observed for groups experiencing health problems. The ONS (2012) reported mean life satisfaction, eudaimonia, and “happy yesterday” responses between 1.7 and 2.0 scale points lower, and “anxious yesterday” responses 1.7 scale points higher, among those out of work due to long-term sickness, in comparison to total population means (all measured on a 0-10 scale). Lucas (2007a) reports data from two very large-scale nationally-representative panel studies several years *before and after* the onset of a disability. In this work, disability was associated with moderate to large drops in happiness (with effect sizes ranging from 0.40 to 1.27 standard deviations), and little adaptation over time.

The OECD (2011a) also reports gender differences in life evaluations and affect balance. While in the United States, Japan, Finland and China women report higher average levels of both life evaluation and affect balance than men (with the ratio of men’s scores to women’s in the range of 0.90-0.99), in Eastern and Southern Europe, Latin America and the Russian Federation men are more likely to report positive affect balance and higher life evaluations. This difference is most marked in the cases of Hungary, Slovenia and Italy, where the ratio of men’s scores to women’s is 1.05 or above for life

evaluations, and 1.15 or above in the case of positive affect balance. Looking at education gaps, low levels of education are associated with lower levels of life evaluations overall, but in Portugal, Spain, Slovenia, Hungary and many of the non-OECD countries for whom data is provided there are particularly large differences in life evaluations between people with and without tertiary education (Figure 4.5).

Figure 4.5. **Gap in life satisfaction by level of education for OECD and selected countries, 2010**



Note: The gap is defined as the difference between the mean life satisfaction of people with tertiary attainment and the mean life satisfaction of people with primary (secondary) education.

Source: OECD's calculations on Cantril Ladder (0-10 scale) life evaluations Gallup World Poll, reported in *How's Life?* (2011a).

Between-country comparisons. In analyses of subjective well-being among different countries, life evaluations appear to vary by up to 5 scale points (on a 0-10 scale) globally, and affect balance by up to 0.5 on a 0-1 scale, although it is important to note that the available evidence on global differences is currently based on small and in some cases unrepresentative samples. The *World Happiness Report* (Helliwell, Layard and Sachs, 2012, Figure 2.3) describes mean average life evaluations averaging 7.6 (on a 0-10 scale) for the top four countries (Denmark, Finland, Norway and the Netherlands), whereas mean reported levels fall well below 4 in the bottom four countries (Togo, Benin, Central African Republic and Sierra Leone). Differences of up to 0.5 are reported (on a 0-1 scale) for positive affect balance, with respondents in Iceland, Laos and Ireland reporting an average positive affect balance of around 0.7, and those in the Palestinian Territories, Armenia and Iraq reporting an average below 0.2.

Even among relatively affluent societies, differences in levels of subjective well-being are non-trivial. Among OECD countries, the mean average life evaluations (on a 0-10 scale) reported in 2010 ranged from over 7.5 in the cases of Denmark, Canada and Norway to between 5.0 and 6.0 in the cases of Portugal, Estonia, Turkey, Greece and Poland (OECD, 2011a). In terms of the proportions of the population experiencing a positive affect balance,¹⁶ Denmark, Iceland, Japan, the Netherlands, Norway and Sweden each score more than 85%, whilst Turkey, Italy, Israel, Portugal and Greece score around 70%.

Both the OECD (2011a) and Helliwell, Layard and Sachs (2012) also report substantial differences between countries in the distribution of subjective well-being. The standard deviations reported for life evaluations on the 0-10 Cantril Ladder reported by Helliwell et al. were around or below 1.5 for the Netherlands, Denmark, Finland and Belgium, indicating quite consistently high levels of subjective well-being. However, similar standard deviations were also observed for Namibia, Tajikistan, Senegal, Madagascar, Côte d'Ivoire, Niger, Cambodia, Burkina Faso, Chad, Comoros, Burundi and the Central African Republic, which among these countries indicates quite consistently low levels of subjective well-being. While the mean life evaluation scores for Puerto Rico and Colombia may seem quite high, considering their levels of economic development, variability among these scores is also high, with standard deviations of around 2.6 and 2.5 respectively. Marked variations in life evaluations are also evident for countries such as Honduras, Pakistan, Nicaragua, Lebanon and the Dominican Republic. These relatively high standard deviations invite more research into their sources, permanence and consequences.

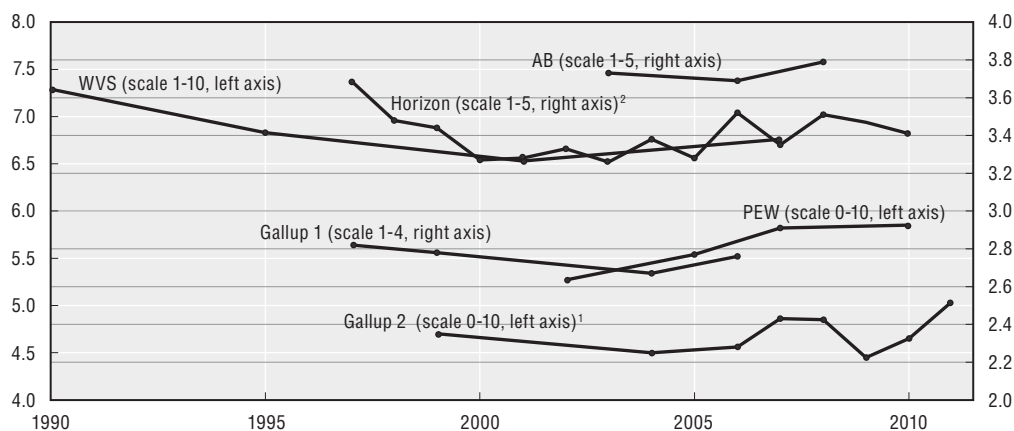
Changes over time. When looking at changes in average country scores over time, the level of change to expect will be highly dependent on the extent of social, political and economic change taking place – and the combined effect of several different changes needs to be considered collectively when interpreting the overall pattern of subjective well-being. Evidence suggests that at the national aggregate level, a *long-term*¹⁷ mean average shift of 0.3 or 0.5 scale points on a 0-10 life evaluation scale may represent a very sizeable change, occurring only in response to major societal shifts. If one bears in mind that little more than 4 scale points separate the top and bottom national life evaluations among the 150 countries in the Gallup World Poll, a shift of 0.4 points would change a country's international ranking by ten to twenty places.

Several authors have examined the question of whether increases in income over time are associated with increases in subjective well-being, and particularly increases in life evaluations (e.g. Easterlin, 1974, 1995, 2005; Hagerty and Veenhoven, 2003; Sacks, Stevenson and Wolfers, 2010). Frijters, Haisken-DeNew and Shields (2004) looked at the effect of the large increase in real household income in East Germany on life satisfaction following the fall of the Berlin Wall in 1989. In the 10-year period between 1991 and 2001, the authors estimated that around 35-40% of the observed increase in average life satisfaction was attributable to the large (over 40%) increase in real household incomes during this time period, with a one-unit increase in log income corresponding to around a 0.5 unit increase in life satisfaction for both men and women.

However, it is important to consider other social and political changes that may be co-occurring when examining the effects of increasing income – such as the potential impact of growing inequalities in society. For example, Easterlin et al. (2012) chart the changing levels of overall life evaluations in China between 1990 and 2010, a period during which China's GDP per capita increased fourfold. According to World Values Survey data

analysed by these authors, there was a *decline* in life satisfaction of 0.53 scale points on a 1-10 scale from 1990 to 2007, although an upturn begins to emerge from early 2000 onwards. Although the early World Values Survey data points in this analysis may be biased upwards, the same general “U-shape” with a turning point between 2000 and 2005 is visible in other data (Figure 4.6).

Figure 4.6. **Easterlin et al. 2012: China’s life satisfaction, estimated from six time series data sets, 1990-2010**



1. Scale of 1-10 for 1999 and 2004 data.

2. For 1997, 1998, 1999 and 2001: Scale of 1-4 and mean computed from 5, 4, 2, 1 coding.

Source: Easterlin, Morgan, Switek and Wang (2012), p. 2.

In conclusion, what constitutes a “big” difference in subjective well-being partly depends on the nature of the difference under consideration. Although results from individual studies need to be interpreted with caution, current evidence suggests that within countries group differences in mean scores of around 10% may be considered large, whereas between countries much greater differences can be observed because the cumulative impacts of different life circumstances can stack on top of one another. Looking at change within countries over time, it is important to consider a wide variety of potential drivers when interpreting the results – because a smaller change than might be expected, or an unexpected direction of change, may be due to the combined and possibly interacting effects of a number of variables. It is also very valuable to examine differences in the *distribution* of subjective well-being data, both over time and between groups. This important consideration is often neglected, but deserves closer attention.

What influences effect sizes?

Several factors potentially limit the size of the difference between groups or change over time that one can expect to see in subjective well-being data. These include the boundedness of the scale and the overall distribution of responses; the sample size in each group and the proportion of the sample affected by a given societal change; changes or differences in other important determinants of well-being; the influence of frame-of-reference effects; the possibility that subjective well-being may be both a cause and an effect of group differences and societal changes; and the time frame over which differences or changes are examined. When interpreting the magnitude of changes in subjective well-being over time, or differences among population sub-groups, it is important to consider how these factors are influencing estimates, particularly when these deviate from the expected pattern.

Bounded scales, response categories and the distribution of responses. One practical issue that potentially limits movements over time is that subjective well-being data are collected using bounded scales with a limited number of response categories. Unlike some indicators of well-being where the scale is unbounded (e.g. income; life expectancy), the average subjective well-being response can never move beyond the top response category. Furthermore, where the choice of response categories is very limited (such as the three and four response categories in the happiness questions often used to investigate the Easterlin paradox), a quite massive shift in life circumstances may be required to move individuals up or down by even just one scale point. Bradburn, Sudman and Wansink (2004) suggest that, on a three-item scale with extremes at either end (for example, “best ever” “worst ever” and “somewhere in between”), most people will tend to select the middle category. This scale sensitivity issue is a key part of the rationale behind preferring longer response scales with a greater number of categories (discussed in depth in Chapter 2). Among highly-developed countries, life evaluations and affect data also tend to be skewed so that the large majority of responses are in the upper range. Whilst at present even the highest-scoring countries still report mean scores several points below the scale ceiling on 0-10 measures, it is rare for respondents to declare their lives or affective states as being absolutely perfect. This implies that “quick wins” in terms of improving overall subjective well-being may be relatively few and far between, and that those high-ranking countries seeking to further improve overall well-being may best focus their energies on the lower tail of the distribution. However, “quick losses” in mean scores of subjective well-being also still remain possible, as highlighted in the experience of Egypt and Tunisia in the years prior to the Arab Spring.

Despite skewed distributions in life evaluation and affect in affluent countries, in emerging and developing economies there is considerable scope for subjective well-being to improve (as documented in the recently published *World Happiness Report*, Helliwell, Layard and Sachs, 2012¹⁸). In the case of eudaimonia, Huppert and So (2011) examined data from 23 European countries and found that even in the best-performing country (Denmark) only 41% of the population were considered to be “flourishing”. The next-best performer was Switzerland at 31%, whereas in Portugal fewer than 10% of the population met the criteria. Thus, whilst the nature of the subjective well-being response scales place theoretical limits on how high subjective well-being can ultimately go, the vast majority of countries do not currently appear to be anywhere near those limits.

Proportion of the sample affected. When interpreting national data, it is also important to consider that life circumstances or conditions that only affect a small percentage of the overall sample (or groups being compared) may have large effects at the individual level whilst having a relatively small effect on the aggregate level. For example, unemployment is known to have a powerful negative impact on the life evaluations of those individuals affected, and this is evident when comparing average scores for the unemployed and the employed. However, even relatively large increases in the unemployment rate (for example, from 5% to 10%) may lead to only small decreases in mean life evaluations for the country as a whole (e.g. Deaton, 2012), in part because they only affect a small proportion of the population. When combined with the high signal-to-noise ratio of subjective well-being indicators, this means that large samples are often needed to detect meaningful long-term shifts in subjective well-being over time at the national level.

The wide variety of subjective well-being determinants. The very large number of determinants of subjective well-being also means that changes over time in any one variable, or in any one difference between two groups of individuals, may have only a small impact on mean scores. The range of variables showing significant associations with subjective well-being includes health, income and material wealth, employment status, migrant status, education, marital status, social relationships, trust in others, volunteering, confidence in institutions, governance, freedom, air quality, personal safety and crime – to name just a few (Boarini et al., 2012). Thus, when examining either changes in country-level mean scores over time, or mean score differences between groups, it is important to consider *other variables* that may also be changing over time or differing between those groups, which could serve to reduce or obscure the effects of the variable in question.

Frame-of-reference effects and adaptation. A person's responses to questions about subjective well-being will be informed by the limits of his or her own experience. "Frame-of-reference effects" refer to differences in the way respondents formulate their answers to survey questions, based on their own life experiences as well as their knowledge about the experiences of others, including both those they consider as within their "comparison group" and those outside it (Sen, 2002; Ubel et al., 2005; Beegle, Himelein and Ravallion, 2012). This knowledge and experience sets the frame of reference, relative to which a respondent's own current circumstances and feelings are felt and evaluated.

Frames of reference produce real differences in how people genuinely feel, rather than simply differences in how people report those feelings. Thus, frame-of-reference effects do not bring into question the validity of subjective well-being measures as *measures of subjective constructs*, but rather they are concerned with the relationship between objective and subjective experiences. Framing effects matter when using subjective well-being as a complement to other measures of well-being, because they concern the extent to which subjective well-being is a *relative* construct, rather than something reflecting *absolute* achievements in society. However, the available evidence suggests that, while framing effects may influence the size of group and country differences observed in subjective well-being data, they are not sufficiently large to prevent the impact of life circumstances from being detected (e.g. Boarini et al., 2012; Fleche, Smith and Sorsa, 2011; Helliwell and Barrington-Leigh, 2010).

Adaptive psychological processes can also either restore or partially repair subjective well-being, and particularly affective experiences, in the face of some types of adversity (e.g. Cummins et al., 2003; Diener, Lucas and Napa Scollon, 2006; Clark et al., 2008; Riis et al., 2005). Adaptation to positive life events such as marriage or winning the lottery has also been observed (Clark et al., 2008; Brickman, Coates and Janoff-Bulman, 1978).

The possibility of shifting reference frames and psychological adaptation again mean that differences over time, between groups and between countries might be smaller than one might expect based on objective changes or differences in life circumstances. However, there is strong evidence that adaptation does not occur (or is incomplete) for a range of policy-relevant life circumstances, such as chronic pain from arthritis or caring for a severely-disabled family member (Cummins et al., 2003), disability (Oswald and Powdthavee, 2008a; Lucas 2007a; Brickman, Coates and Janoff-Bulman, 1978) and unemployment (Lucas et al., 2003). Focusing on instances of incomplete adaptation could help policy-makers and public service providers to focus on areas where intervention may be most valuable.

Reverse and two-way causality and time frames for analyses. The possibility of reverse or two-way causality¹⁹ among subjective well-being and its determinants may also limit the size of difference likely to be observed in subjective well-being data. For example, at the cross-sectional level, there is evidence that being married is associated with higher levels of both life satisfaction and positive affect, and with lower levels of negative affect/anxiety (e.g. Boarini et al., 2012; ONS, 2012). However, there is also some evidence that happier people are more likely to get married (Luhmann et al., 2012), and that after the initial boost in subjective well-being observed around the time of marriage, subjective well-being reduces back to its pre-marriage levels in the years after the event (Clark et al., 2008). Thus, an increase in the proportion of the population getting married in year n may not necessarily produce a large increase in the average levels of subjective well-being reported in year $n + 5$. Conversely, an increase in average levels of subjective well-being may actually precede an increase in marriage rates.

Both reverse/two-way causality and adaptation raise the issue of the appropriate *time frame* to consider when examining changes in subjective well-being. Whilst some determinants of subjective well-being might be expected to have an immediate impact on feelings of well-being (such as the sudden onset of disability or the death of a family member), others may take longer to unfold because their effects are indirect (for example, the influence of education on subjective well-being, or of having sufficient income to enable investment in healthcare insurance or a pension for retirement). Thus, as with any attempt to evaluate impact, for sizeable differences to be detected the correct time-frame for analysis needs to be adopted, based on what is known about the variables in question and the causal pathways through which they take effect.

Although longer time-frames might be required to detect significant changes in subjective well-being data, it is also true that these measures can be relatively bumpy over short time periods. Deaton (2012), for example, raises the possibility that long-term trends in country-level subjective well-being risk being swamped by “cognitive bubbles”, i.e. by the temporary impact of short-term reactivity to national events that affect everyone (such as public holidays or major news events). If time-series data on subjective well-being are examined over only short time periods, these bubbles can potentially drown out the more meaningful changes associated with important societal shifts in well-being (such as rising unemployment rates), particularly if these affect only a small proportion of the population during the time-frame examined. On the one hand, short-term measures may act as a useful barometer for public mood – and the short-term worry and stress that accompanied the immediate impacts of the 2008 financial crisis “is surely real enough, and worth measuring and taking into account in policy” (Deaton, 2008, p.23). On the other hand, it is important to view short-term fluctuations in subjective well-being in the context of much broader long-term trends in order to capture wider changes in what most people might regard as societal progress.

What alternative explanations should be considered for observed differences in subjective well-being?

Because subjective well-being is affected by so many different life circumstances, several factors need to be taken into account when interpreting the magnitude of a difference between groups or a change in subjective well-being data over time. A number of background characteristics, such as age, gender and marital status, can influence mean

levels of subjective well-being. When making group comparisons (for example, between the employed and unemployed, or between different regions within a country), it is important to consider the impact of differences in these background characteristics.

Regarding age differences, evidence suggests that among affluent OECD and especially English-speaking countries, there is a U-shaped relationship between age and life satisfaction (with average levels lowest between the ages of around 35 and 55), and this persists even after controlling for other age-related factors such as income and health status (OECD, 2011a; Blanchflower and Oswald, 2008; Deaton, 2010). However, among lower- and middle-income countries, there is evidence to suggest that life satisfaction decreases with age, and age-related decreases appear to be particularly marked in transition countries in Eastern Europe and the former Soviet Union (Deaton, 2010). Gender also has a small but significant impact on reported subjective well-being in several countries, as detailed in the preceding section. Finally, there may be systematic differences in patterns of responses associated with different cultures, discussed in the section that follows.

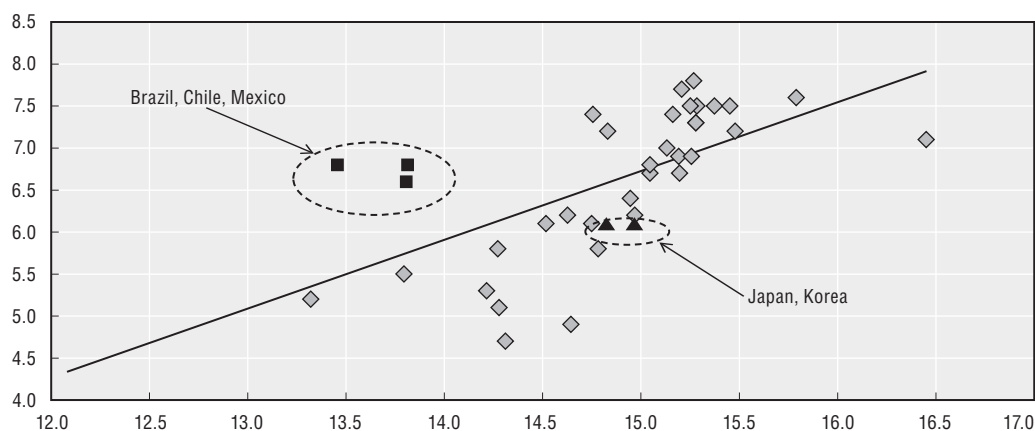
When making group comparisons in subjective well-being, it is therefore important that the gender, age and cultural composition of the groups in question are taken into account. This is also important when examining changes in national subjective well-being over time, which could be linked to demographic shifts among the population. Without taking these factors into account, spurious interpretations of the data are possible. For example, whilst it is evident that the onset of widowhood is associated with a very substantial decrease in life evaluations at the individual level (Clark and Oswald, 2002), the probability of widowhood increases with old age, and in affluent countries life evaluations also tend to increase after the age of 55. This means that in cross-sectional national data, the life evaluations of widowed individuals may not be as low as one might expect, due to the interaction between widowhood risk and old age. In a recent small-scale study in New Zealand (UMR, 2012), for example, widows appeared to report *higher* happiness than married respondents, a pattern that is almost certainly due to widows being, on average, older than married, divorced or single respondents.

What is the role of culture in international comparisons, and can data be “corrected” for “cultural bias”?

The fact that the global distribution of both life satisfaction and affect is wide and varied suggests that differences in country-level life circumstances are likely to produce differences in country-level subjective well-being, and this has been confirmed empirically in a wide variety of studies based on large international datasets (e.g. Helliwell and Barrington-Leigh, 2010; Helliwell et al., 2010; Helliwell, Layard and Sachs, 2012; Deaton, 2010; Boarini et al., 2012; Fleche, Smith and Sorsa, 2011). For example, in the Cantril Ladder life evaluations data set examined in the *World Happiness Report* (Helliwell et al., 2012), the countries with the top four rankings are reported to have average incomes 40 times higher than those countries with the bottom four rankings.

On the other hand, countries with relatively similar levels of economic development can sometimes report quite different mean levels of subjective well-being.²⁰ Inglehart, Foa, Peterson and Welzel (2008) illustrate how international measures of subjective well-being data can diverge from the pattern that might be predicted based solely on their level of economic development (see also Figure 4.7). This indicates that Latin American countries

Figure 4.7. **Subjective well-being (SWB) and per capita gross domestic product (GDP)**
Mean life satisfaction¹ versus log of GDP per capita – OECD and selected countries, 2010



1. Mean Cantril Ladder score.

Source: OECD (2011), *How's Life?*, OECD Publishing.

in particular tend to report higher levels of subjective well-being than might be expected based only on their GDP per capita, whilst ex-Communist countries appear to report lower subjective well-being than one might expect.

While income may be an important determinant of country differences in subjective well-being, a number of other non-economic factors are also important, many of which relate to measurable differences in life circumstances (such as health and social context, see Helliwell et al., 2012 and the discussion below). As with interpersonal comparisons of subjective well-being, there may also be some differences *between countries* in terms of the frames of reference used by individuals to report on their own well-being, as well as differences in how questions are understood and interpreted, and how response formats are used. Chapter 2 described the evidence around hypothesised cultural response styles and the methodological steps that can be taken to reduce the risk of differences in how scales are understood by respondents, and Chapter 3 covered the issue of scale translation in more detail. The purpose of the present discussion is therefore to focus on the interpretation of observed international differences in average levels of subjective well-being, possible sources of those differences, and whether data can and should be “corrected” for linguistic or cultural bias *after* it has been collected.

Cultural impacts versus cultural “bias”. Before attributing differences in average subjective well-being between countries at similar levels of economic development to “cultural bias”, it is important to remember that these differences may have many sources. A helpful distinction can be made between *cultural impact*, which refers to valid sources of variance between cultures, and *cultural bias*, which refers to inter-cultural differences that result from measurement artefacts²¹ (Van de Vijver and Poortinga, 1997).

If we assume that an identical measurement approach has been adopted across all countries, and therefore that observed differences cannot be attributed to the methodological differences described in Chapter 2, differences in average levels of subjective well-being between countries may have at least four different sources:

- *Life circumstances*

In addition to income and other economic variables, there may be other important differences between countries in terms of social context and other life circumstances. As noted above, levels of economic development are just one group of potential drivers of subjective well-being, but a very wide variety of others exist, often playing a more substantial role than income (e.g. health, social relationships, unemployment rates, freedom of choice and control). These drivers can include valid country differences in subjective well-being connected to levels of democracy, tolerance of outgroups, strength of religiosity, or trust in others (Inglehart et al., 2008; Bjørnskov, 2010), and perceived freedom, corruption and the quality of social relationships (Helliwell, 2008; Helliwell et al., 2010; Helliwell et al., 2012). The socio-demographic structure of countries may also contribute to mean differences observed between countries. Because of the very wide range of factors that impact on average levels of subjective well-being, comparing countries on the basis of income alone is insufficient.

- *Differences in how people feel about their life circumstances*

There may be differences between countries in how people *feel* about their current life circumstances. Many factors may potentially influence how life circumstances are appraised, including an individual's reference group (i.e. frame-of-reference effects, discussed just above), past life experiences, the past or present political and economic situation, the policy environment and the country's religious, cultural and historical roots. These differences may contribute to *appraisal styles* that influence the connection between objective life circumstances and subjective feelings – for example, the degree of optimism or pessimism individuals feel about the future. Rather than representing cultural “bias”, these should arguably be regarded as valid sources of difference between countries – because they influence the level of subjective well-being actually experienced by individuals, even if this does not mirror exactly the measures of their objective life circumstances.²²

- *Language differences that influence scale use*

Systematic differences between countries may also arise as a result of imperfect translatability of subjective well-being constructs. For example, Veenhoven (2008) has shown differences between the (0-10) numerical ratings that respondents assigned to English and Dutch translations of verbal response categories (*very happy, quite happy, not very happy* and *not at all happy*). In this instance, linguistic differences would produce biases in how people respond to a verbally-labelled scale that bear no relation to how individuals actually feel about their lives – and thus it would be desirable to remove this bias from the data.

- *Cultural response styles or biases*

There may be country-specific differences in how individuals *report* their feelings, regardless of their actual experiences. For example, a “modesty” or moderate-responding bias might have a downward influence on self-reports, without having a negative impact on private feelings of subjective well-being. Similarly, tendencies towards “extreme responding” (i.e. using scale end-points) or more socially desirable responding could imply

differences in modes of cultural expression, rather than substantive differences in the subjective well-being actually experienced. These effects could be described as group differences in scale use, or cultural response styles. If differences in scale use do not represent differences in how people feel about their lives, they can be regarded as a source of bias that it would be desirable to remove.

It is important to distinguish between these four potential sources of country differences, because they have different implications for the *validity* of between-country comparisons, and for the actions one might take to address country-specific differences in subjective well-being. In the case of unmeasured life circumstances, there is a country-specific effect that may or may not be related to culture. In the second case, differences between countries can reflect cultural impact – i.e. differences in how respondents *genuinely feel*, and which would add to the predictive validity of the overall subjective well-being measure (e.g. in its association with future behaviour or health states). One would not necessarily want to *correct* subjective well-being scores for either the first or the second of these country-specific differences. Linguistic differences or cultural response styles, on the other hand, can be expected to add *bias* to the data, reducing its overall validity and predictive ability. In these instances, it would be desirable to find a way to either minimise the problem at source through survey design (Chapter 2) and translation (Chapter 3), or to adjust the data *ex post* to remove the bias and enhance the overall usefulness of the measures.

Methods for examining and “correcting” cultural bias

Counterfactuals. The “counterfactual” approach attempts to isolate the problem of unmeasured life circumstances by examining the extent to which country differences can be explained by a variety of objective outcomes. It involves using information about objective life circumstances to adjust subjective reports so that only the variance in subjective reports that can be explained by objective life circumstances is retained. For example, Jürges (2007) used detailed information about a wide range of objective medical complaints to adjust self-reported health data from 10 European countries so that only variance that could be explained with reference to the objective health indicators was included.

Two fundamental assumptions in the counterfactual approach are that: a) all relevant objective variables have been included in analyses; and that b) there is no valid variation among respondents in how objective states are perceived and experienced. Using such a procedure to “correct” for country-specific effects in practice eliminates *any* country differences in the relationship between objective outcomes and subjective experiences. In the work of Jürges (2007), this means that any country-specific differences in, for example, the care received by patients, or the support received from friends and family, are eliminated from the adjusted data. Unfortunately, these country-specific differences are precisely the kinds of differences that are likely to be the focus of interest of governments and public service professionals. Thus, in removing the country-specific influences that might “bias” self-reported data, the study design also potentially removes any *substantive* differences between countries in how health conditions affect perceptions of health. However, such an approach may have value in helping to understand the nature and composition of differences between countries (Fleche, Smith and Sorsa, 2011).

Fixed effects models. One could argue that if counterfactuals are to be constructed to better understand subjective well-being differences between countries, they will need to take a wide variety of life circumstances into account – including social context variables that often rely on self-report measures. The difficulty is, however, that cultural differences in scale use are likely to affect many self-report measures, not just subjective well-being. Thus adding self-reported measures into a counterfactual model could serve to mask cultural response styles, rather than teasing them out. An alternative approach involves examining country and regional fixed effects in the data.

Helliwell, Barrington-Leigh, Harris and Huang (2010) reported that among a global sample of between 50 000 and 140 000 respondents in 125 countries, a regression equation including demographics, income, being unable to afford food, having friends to count on, perceived freedom, perceived corruption, charitable donations of money and time, helping strangers, and religion explained between 30 and 45% of the individual-level variance in life evaluations. Adding dummy variables for regional fixed effects added between 0.3 to 1.1% to the variance explained, with a significant and positive coefficient for the South and Central America region (indicating more positive evaluations of life than might be predicted from other variables in the model).²³ Adding individual dummies for every country surveyed added between 2.5 and 4% to the overall variance explained. Importantly, however, neither set of dummies substantially reduced the coefficients for the other predictors in the model – indicating that the strong relationships between social indicators and subjective well-being was not due to country- or region-specific fixed effects. Helliwell et al. (2010) also report that when separate regression models are created for each of the different regions, and each of the different countries, the coefficients are markedly similar to those obtained in the single global model.

The work of Helliwell et al. (2010) suggests that even if some variation in average scores between countries may be due to unexplained factors, there do not appear to be big country or regional differences in the *structure* of the relationship between subjective well-being and some of its key known determinants. Although they did not explicitly set out to measure cultural bias, their results also suggest that when a much larger predictor set is examined, relatively little of the unexplained variance in life evaluations is due to region- and country-specific fixed effects. This does not *preclude* the presence of cultural biases in the data, but if these biases operate at the regional- or country-level, they appear to explain only a small amount of variation in individual-level responding, above and beyond life circumstances.

Vignettes. The “vignette” approach (e.g. King et al., 2004) attempts to measure the different ways in which individuals and/or cultures may understand, interpret, benchmark or respond to the same survey question, issues that are collectively known as “differential item functioning” (DIF). DIF can result from either scale translation problems or cultural response styles.

Vignettes are short descriptions of hypothetical scenarios that respondents are asked to rate, using the same scale format used to obtain self-reports. The vignette method works on the assumption of “vignette equivalence”, i.e. that, because respondents are each evaluating *the same* vignette, they should in principle assign identical ratings to that vignette. Thus, any differences between individuals (or groups of individuals) in the ratings assigned to vignettes are attributed to differential item functioning or response styles²⁴ (often interpreted as “cultural bias” in cross-country studies).

The vignette approach has been used in recent cross-country studies to identify cultural effects in subjective data (Angelini et al., 2011; Kapteyn, Smith and van Soest, 2009; Kristensen and Johansson, 2008 – Box 4.4). For example, Angelini et al. (2011) utilised data from respondents in ten European countries ($N = 5\,606$), who were asked to provide life satisfaction ratings for both themselves and for two fictional elderly characters described in two separate vignettes, detailing the characters' age, family and social relations, income and health circumstances. In Angelini et al.'s study, the (unadjusted) self-assessments of life satisfaction showed Danes to be the most satisfied with their lives and Italians the least satisfied. However, the vignette method indicated differences in the scale thresholds used by Danish and Italian respondents to define the response categories (on a 5-point verbally-labelled scale, ranging from "very satisfied" to "very dissatisfied"). In simulations estimating life satisfaction for other countries using Danish scale thresholds, more than 95% of respondents in all countries would rate themselves as satisfied or very satisfied with their own life. In simulations using Italian thresholds, this reduces to between 80 and 95% in most cases, but around 60% for Poland and the Czech Republic.

**Box 4.4. Use of vignettes to investigate job satisfaction
– Kristensen and Johansson, 2008**

Kristensen and Johansson (2008) examined subjective assessments of job satisfaction across seven EU countries ($N = 5\,988$), using 19 different sets of five vignettes. Respondents evaluated both their own jobs, and fictional jobs described in vignettes, on a 0-10 job satisfaction measure.

As can be seen from Table 4.4, the relative country rankings in terms of job satisfaction shift: a) when objective job characteristics are controlled; and b) when satisfaction rankings are adjusted on the basis of vignette responses.

Table 4.4. Differences in country rankings of job satisfaction, 2008

Rank	"Positivity" of vignette scores ¹	Average unadjusted self-report job satisfaction	a) Job satisfaction rankings, controlling for objective job characteristics	b) Vignette-adjusted job satisfaction ranking ²
1	Finland	Denmark (7.5)	Finland	Netherlands
2	Spain	Finland (7.4)	Denmark	Greece
3	Greece	Netherlands (7.3)	Greece	Denmark
4	Netherlands	Greece (6.9)	Netherlands	Finland
5	United Kingdom	France (6.6)	Spain	France
6	Denmark	Spain (6.5)	France	United Kingdom
7	France	United Kingdom (6.4)	United Kingdom	Spain

1. Calculated from the information provided by Kristensen and Johansson in Table 3, p. 104.

2. Also with objective job characteristics controlled.

Kapteyn, Smith and van Soest (2009) used vignette ratings as a means of examining response scale differences between Dutch and US respondents, and found that Dutch life evaluations become more spread out when re-scaled using US threshold parameters – such that there are increases in the percentage of respondents who are very satisfied (from 21.5% to 26.7%) or very dissatisfied (from 1.5% to 3.1%). However, this re-scaling has little impact on the total scale mean: if one re-calculates mean scores by assigning 1-5 scale

values (from very dissatisfied to very satisfied respectively) the original Dutch ratings average 4.06, and when re-scaled according to US thresholds the mean average is 4.07. From either perspective, then, differential item functioning appears to produce only small scale shifts among this sample.

Although vignette-based studies have revealed some differences between countries in the subjective well-being ratings assigned to vignettes, the *source* of these country differences requires further investigation. The vignette method assumes that any country-specific differences in the scores assigned to vignettes must be brought about by response styles, or cultural bias. Another likely possibility is that this method picks up problems in scale translation between languages, particularly where short and verbally-labelled response categories (typical of vignette studies) are used. However, it is also possible that differences between vignette ratings reflect meaningful differences between countries (such as differences in the work, social and policy environments) that have real consequences for how individuals perceive and respond to the narrow set of life circumstances described in vignettes.²⁵ For example, estimates of the likely subjective well-being impact of unemployment in a vignette may be affected by country-specific differences in the macroeconomic and jobs climate, the social safety nets available, social norms around work (Stutzer and Lalive, 2004), and total levels of unemployment (Clark, 2003).

Vignettes also require that individuals are able to accurately forecast how they would feel in different circumstances and that people respond to vignettes in the same way that they do to actual questions about their own subjective well-being. Both of these assumptions can be questioned. It is important, therefore, to *empirically demonstrate* that the vignette-adjusted ratings are more accurate reflections of what respondents are subjectively feeling, or more accurate predictors of future behaviour, than are unadjusted mean scores.

Migrant data. Senik (2011) proposes a technique in which the effects of culture can be investigated by comparing the subjective well-being of native and migrant respondents within a country. When a variety of background and economic variables are controlled for, if migrants systematically differ from natives in their subjective well-being assessments (despite exposure to similar within-country conditions), one might infer cultural causes for these differences – although it will be important to eliminate a variety of other ways in which migrants and natives differ in their experiences. It is also possible to compare the subjective well-being of individuals living in their native country with that of their fellow countrymen who have emigrated overseas (e.g. French people living in France, versus French people living in other countries).

In examining the “French unhappiness puzzle”, Senik found that across a sample of 13 European countries in total, living in France reduced average happiness by 0.23 points (on a 0-10 scale) and reduced the probability of reporting 7 or more by 19% relative to the rest of the sample. By contrast, people living in Denmark were 50% more likely to score 7 or more. Furthermore, although the general trend across the sample was to find that natives are happier than immigrants, among the French sample this pattern was reversed. The actual reported level of happiness among French natives was 7.22, but when simulated using regression parameters obtained from immigrants living in France, this was predicted to rise to 7.36. Conversely, the average reported happiness of immigrants (7.25 in the original self-reports) was predicted to decrease to 7.15 when simulated using regression

parameters based on responses from French natives. Finally, when estimated using the parameters obtained for natives among the other 12 European countries, the happiness of French natives was predicted to rise to 7.54.

This method thus implies something unique to French natives that has a tendency to depress happiness reports by between one-eighth and one-quarter of a scale point, on a 0-10 scale, relative to immigrants from a wide range of countries, and natives among other European countries.²⁶ What is much less clear from this work is the precise source of this effect. Senik is clear that international differences identified through her methods should not be interpreted as “meaningless anchoring biases and measurement errors, but as identity and cultural traits” (p. 7). This approach thus views culture and mental attitudes not as something that should be statistically controlled in cross-country comparisons, but rather as genuine cultural impacts which can point to possibilities for policy interventions, particularly in school and childhood experiences, that can support the development of more positive subjective well-being overall. Senik also notes that the observed “French unhappiness” is mirrored by low levels of trust in the market and in other people.

Comparison of life evaluations and affect balance. Other insights into response biases can potentially be obtained through comparing life evaluations and affect balance measures. Recent experiences of affect are thought to be less susceptible to retrospective recall biases than life evaluations in particular (Kahneman and Riis, 2005; Diener and Tov, 2012; Oishi, 2002; Tsai, Knutson and Fung, 2006; Blanchflower, 2009). In terms of response styles or biases that have been linked with culture, such as more extreme and more moderate responding, it is also possible that *affect balance* measures, by subtracting mean negative affect from mean positive affect, reduce the impact of such biases in the final data set.²⁷ This requires further examination.

Krueger et al. (2009) reported that, based on results from studies with US and French samples, the French (on average) report spending more time in a more positive mood, and spend more of their time in activities that are rated as more enjoyable. This contrasts with responses to life satisfaction questions, which typically find US samples reporting higher life satisfaction than the French, and suggests affect data can add something new to the overall well-being picture.

In an analysis of subjective well-being among 40 OECD and emerging countries, there are some marked differences between affect balance and life satisfaction measures in terms of countries’ relative rankings (OECD, 2011a). For example, Japan falls below OECD-average levels on life satisfaction, but Japan’s relative position on affect balance almost reverses, such that it is ranked third-highest overall. All Asian countries considered in this data set ranked higher using affect balance relative to life satisfaction, and this was particularly striking for China, Indonesia and Japan, who move from near the bottom of the rankings to near the top. In contrast, Israel, Italy, Finland, Switzerland and Canada ranked significantly lower on affect balance relative to life satisfaction.

However, the fact that the two measures perform differently doesn’t in any way show that one is less “biased” than the other – and there are several alternative explanations for observed differences in the patterns among affect and life evaluations. For example, it is known that these measures are affected in different ways by objective life circumstances, such as income, which has a stronger impact on life satisfaction than on affect balance (Boarini et al., 2012;

Kahneman and Deaton, 2010). As a result, affect data could not be used to “correct” life evaluations, but comparisons of the two measures are nonetheless interesting in the context of other investigations of cultural impact – including vignettes and migrant data.

Conclusions: What is the role of culture in international comparisons, and can data be “corrected” for “cultural bias”? The gap between standard economic variables and subjective well-being arguably reflects where using subjective well-being can add value to existing measures of progress. However, this gap contains noise as well as signal. Separating the signal from the noise is a particularly vexed issue when it comes to comparisons of average subjective well-being levels between countries.

While counterfactuals, vignettes, migrant data and comparisons between life evaluations and affect offer interesting insights into the impact of country (and, by extension, culture²⁸) on subjective well-being, none of these approaches has yet been able to convincingly distinguish between substantive cultural *impacts* and cultural *bias*. The relatively small number of countries sampled in the existing research also makes it difficult to extrapolate more widely on the basis of this work – meaning that there is little that can be said even about the expected *magnitude* of cultural effects, particularly at a global level. For example, although the “cultural differences” between US and Dutch respondents (Kapteyn et al., 2009) or French and other European respondents (Senik, 2011) appeared to be reasonably small, more disparate cultural groups, such as Latin American and former Soviet countries, may reveal larger differences. The findings of Helliwell et al. (2010, cited earlier) suggest that when a wide range of predictors are taken into account in a global sample, both region- and country-specific fixed effects (which are likely to reflect both unmeasured variables *and* sources of systematic biases if these vary between the regions and countries identified) explain a relatively small percentage of the overall variance in subjective well-being (between 0.3 and 4%). Access to further high-quality data on subjective well-being from large and nationally-representative samples will help to shed light on the issue of *what proportion* of average-level differences between countries can be attributed to cultural biases in determining whether the benefits of data adjustments outweigh the costs of lost information.

A further practical limitation in using vignettes and migrant data to “correct” country averages of subjective well-being data is that the impact of culture in the data cannot be quantified in simple absolute terms – rather, it is defined relative to other countries in the sample. This provides a further challenge if the goal is to adjust national-level data to provide “culture-free” estimates, and implies that only a large and representative global sample could really be used as a basis for such adjustments.

Given the current state of the evidence available, these guidelines do not recommend using the methods described in this section for “correcting” mean average country-level subjective well-being data for cultural influences. It is not yet clear, for example, whether adjusting subjective well-being according to these methods actually *adds* to the validity of the data or to its usefulness in terms of predicting future behaviour and other well-being outcomes. Correcting data for *all* country-specific influences on how objective circumstances are perceived would risk removing the influence that all unmeasured country differences (including the influence of a country’s policy environment, social safety nets, and a wide range of valid cultural differences) have on how subjective assessments and feelings are formed.

Thus, until further research and analyses become available, the risk of cultural bias is best managed through adopting survey design and translation principles (described in Chapters 2 and 3 of these guidelines) that seek to minimise differences in scale interpretation and use among respondents. Inclusion of covariates that could help to explain cultural *impacts* may also be valuable in international comparisons. Finally, supplementing life evaluation data with affect balance measures may provide a more rounded picture of country differences overall.

2. Better understanding the drivers of subjective well-being

Introduction

The present section is concerned with analyses examining the drivers of subjective well-being. If identifying vulnerable groups and international benchmarking are core elements of monitoring well-being, better understanding the drivers of subjective well-being can help to *explain* some of the differences observed over time or between groups – both within and among nations. This analysis might then suggest areas where policy interventions and individual life choices might raise levels of subjective well-being overall.

This section is organised in three parts. It begins with an overview of what better understanding the drivers of well-being means in practice, and the types of drivers typically examined in such analyses – including other high-level well-being outcomes, life events and more specific policy interventions. The use of subjective well-being data to inform the appraisal, design and evaluation of different policy options, as well as to examine policy trade-offs, is also described.

Key methods involved in the analysis of subjective well-being drivers are then covered in the second part of this section. This includes discussion of data requirements and the types of survey design that facilitate causal interpretations, as well as brief consideration of the types of statistical analysis involved in these investigations. Finally, the third part of this section discusses the challenges associated with interpreting analyses of the drivers of subjective well-being. The fundamental questions addressed are *what size of impact can be expected?* and *how can the impacts of different drivers be compared?* Key issues considered include interpretation of regression coefficients, the generalisability of results, the risk of error in the measurement of both drivers and outcomes, and the time frames under investigation.

What does “better understanding the drivers” mean, and why does it matter?

Understanding the drivers of subjective well-being means identifying variables that appear to have causal relationships with subjective well-being and examining some of the mechanisms through which drivers take their effects. Drivers of subjective well-being can include high-level well-being outcomes, such as income and health conditions, as well as specific life events and circumstances such as unemployment or the onset of disability, or specific patterns of behaviours and time use, such as commuting, watching TV, or interacting with friends and family.

Governments and researchers may be interested in the drivers of subjective well-being for a number of reasons, which are described in more detail in the sections that follow. Organisations and individuals may also have an interest in both the life circumstances and the daily events that influence subjective well-being in order to help inform decision-making and increase the well-being of workers and their families (Box 4.5).

Box 4.5. Wider public interest in the drivers of subjective well-being

Drivers of subjective well-being that could interest the general public include:

- Other high-level well-being outcomes (such as income, social connections and health) and the trade-offs that may exist between them.
- The impact of certain life events, and the factors associated with positive adaptation to life events over time.
- How time use plays a role in both short-term mood states and longer-term well-being.

For example, Layard (2005) discusses how geographic labour mobility might bring positive economic benefits, but could potentially lead to an overall decrease in well-being, including subjective well-being, through losses in both work and social connections and the weakening of local community ties. The trade-off between economic benefits and the “hidden costs of mobility” (Dolan and White, 2007) can be explored through examining subjective well-being data, which can also be used to illuminate the factors associated with successful adaptation to relocation. This information may be useful for both the individuals making those trade-offs as well as organisations seeking to support the well-being of staff that have been relocated.

Data obtained through a combination of time-use and survey methods may also prove interesting for individual decision-making. Loewenstein and Ubel (2008) view informing the public about the likely consequences of particular actions as the main way in which affect data should be used. Kahneman and Riis (2005) meanwhile suggest that paying more attention to the allocation of time is one of the more practical ways to improve experienced well-being.

Several studies have provided insights into both subjective well-being gained from individual activities and the net impact that time allocation has on national well-being. For example, Kahneman et al. (2004) conducted an investigation of affect among nearly 1 000 working women in Texas using the Day Reconstruction Method. Among this sample, the three work-related activities (the morning commute, time spent at work, and the evening commute) were associated with the lowest average levels of positive affect balance, whilst intimate relations, socialising after work and eating dinner were associated with the highest average levels of positive affect balance. These authors propose calculating *national accounts of well-being*, based on the proportion of time individuals report being engaged in different activities, and the net affective experience reported during each of those activities (e.g. Krueger et al., 2009).

Individual activities that have been investigated in detail for both their short- and long-term influence on subjective well-being include TV watching, Internet use and commuting. Frey, Benesch and Stutzer (2007) for example reported that people watch more TV than they consider optimal for themselves, and that heavy TV viewers – particularly those with significant opportunity costs of time – report lower life satisfaction. Gross, Juvonen and Gable (2002) examined Internet use among adolescents and found that, whilst the overall duration of time spent online was not associated with evaluative subjective well-being or daily affect, the emotional closeness of instant message communication partners was associated with daily social anxiety and loneliness in school. Finally, Stutzer and Frey (2008) found that people with longer commuting times reported systematically lower satisfaction with life overall – consistent with the finding from Kahneman et al. (2004, above) that commuting is associated with low levels of positive affect balance.

Because of the various challenges associated with interpreting analyses of drivers, however, it may be misleading to place too much emphasis on comparing the relative effect sizes of the different drivers (see below). In communicating with the wider public, then, results of these types of analyses should not be presented as a *recipe* for subjective well-being, but rather more as a list of ingredients, with a broad indication of their impacts – and allowing flexibility to adapt the recipe according to taste.

Life events and life circumstances as drivers of subjective well-being

Examining the relationship between subjective well-being and other key well-being outcomes (e.g. income, jobs, health, social relationships, work-life balance, personal security, education, civic engagement, governance, housing, environmental quality) is then the first step for better understanding differences in subjective well-being observed between different groups or over time. As well as enhancing understanding of well-being as an over-arching construct, analysing the drivers of subjective well-being data offers a way to test empirically whether the outcomes currently used to measure and describe societal progress align with the outcomes that determine people's perceptions of their own well-being. These analyses can also assist in identifying potential opportunities for new policy approaches, improving the design of existing policies or highlighting areas where policies or regulations can be withdrawn in order to improve subjective well-being outcomes. In particular, identifying the factors that influence how people react to adversity, such as the onset of disability or unemployment, as well as how successfully they adapt to these events over time, may be particularly relevant to policy-makers.

Although not all of the life circumstances that are important to subjective well-being will be amenable to policy interventions, subjective well-being evidence can have particular implications for government approaches to issues such as: mental health and resilience; employment, training and labour market flexibility; child welfare, and family and community policies; and taxation approaches to products (e.g. addictive substances) and activities known to have an impact on subjective well-being (Layard 2005; 2011). The next step may then be to examine the impacts that particular interventions are likely to have on subjective well-being outcomes, and how consideration of subjective well-being impacts can be used to inform certain policy design features.

Using subjective well-being data to inform the options appraisal, design and evaluation of policies

When making funding allocation decisions, it is important for governments to have information about the efficiency with which resources can be used to achieve policy objectives. Estimating the efficiency of expenditure, often described as *value for money*, delivered by a project, programme or policy intervention involves quantifying the various impacts it might have on outcomes of interest – including economic outcomes (e.g. does the intervention boost jobs or decrease regulatory burdens on business?), social outcomes (e.g. does the intervention improve educational attainment or health outcomes?) and environmental outcomes (e.g. does the intervention contribute to carbon reduction or increase peoples' access to green space in the local area?). These questions are relevant in the process of initially appraising policy options, but may also be asked as part of ongoing refinements to policy design and implementation, as well as when examining the potential impacts of *stopping* a particular policy intervention or regulation.²⁹

Options appraisal takes place before a policy is implemented, whereas policy evaluation involves the assessment of policy during its implementation – and might include both specifically commissioned research, as well as less formal evaluations based on existing evidence. Formal programme evaluations may involve an experimental or quasi-experimental design for investigations and include measures both before and after a policy has been introduced, enabling causal inferences to be drawn about the impact of

the policy. Policy design considerations are relevant both before and during policy implementation, as well as in interpreting the evaluation of policy impacts and what can be done to enhance them.

Alongside some of the typical economic, social and environmental outcomes, subjective well-being data can provide policy-makers with an additional perspective on the potential impact of a policy. As noted earlier, subjective well-being data may offer unique insights into the effects of a given action, taking into account a variety of objective well-being outcomes and how they combine to produce an overall *perception* of well-being. It should be noted, however, that because subjective well-being has so many drivers, the impact of any one policy, particularly one that affects only a small number of people, may prove difficult to detect. This has implications for the sample sizes required, and the study design adopted in, for example, formal policy evaluations – issues that will be discussed in the section that follows.

In assessing the likely impacts of a policy intervention on subjective well-being, analysts are likely to draw on prior literature, including academic sources, and international examples. More comparable data will enhance the quality of these sources of information and provide better baseline information about the levels of subjective well-being to expect among different population sub-groups. This baseline data can provide essential information about the “do nothing” policy option – i.e. what to expect in the absence of intervention.

Diener and Tov (2012) list a wide variety of policy considerations where it may be valuable to consult subjective well-being data. These include issues such as deciding how to support day care for elderly Alzheimer patients, examining the moods and emotions of caregivers when the patient is in day care or at home and the life satisfaction of caregivers when respite care is provided; or examining the well-being benefits of parks and recreation, testing whether parks are more crucial to well-being in areas where dwellings have no outdoor space and whether life satisfaction is higher in cities with plentiful parks than in cities where parks are rare.

In terms of applied examples, Gruber and Mallainathan (2002) have used subjective well-being data to examine optimal cigarette taxation across a range of areas in the United States and Canada. Boarini et al. (2012) show how data from OECD member countries can be used to explore the impact of health co-payments and unemployment replacement rates on national levels of subjective well-being, as well as well-being among certain population subgroups, such as people working versus those outside the labour market. Using a quasi-experimental design, Dolan and Metcalfe (2008) examined the subjective well-being impact of an urban regeneration project in Wales.

Using subjective well-being data to inform policy trade-offs

A key part of governing involves making decisions not only between different policy options but also between different policy objectives. When confronting policy *trade-offs*, one of the perennial challenges lies in comparing the relative value of different economic, social or environmental policy outcomes. This is like comparing apples with oranges: is an increase in educational attainment any “better” or “more necessary” than an increase in health outcomes? Governments may use various methods to assist in making these decisions, such as international benchmarking and national targets, but policy trade-offs, and particularly those involving different departments with different objectives, remain thorny issues to resolve.

Although it is difficult to imagine ever solving this problem through evidence alone, subjective well-being data offer one way of looking at the *societal preferences* (Loewenstein and Ubel, 2008) for different trade-offs. In the terms of Bjørnskov (2012), governments face a *massive information problem* – i.e. in a large society it becomes impossible to know enough about the detailed preferences and needs of the population to direct policies according to those preferences and needs. By providing information about what is likely to increase the subjective well-being of the population at large, subjective well-being data offers an alternative to listening to the arguments of relatively narrow interest and lobby groups. Subjective well-being can also offer a standard unit of comparison, thereby facilitating more “joined-up government” where departments are better able to consider the spill-overs from their interventions onto a wider range of domains (Dolan and White, 2007).

Practical examples of trade-offs that have been examined in the literature include the trade-off between inflation and unemployment (Di Tella, MacCulloch and Oswald, 2001) and that between income and airport noise (Van Praag and Baarsma, 2005). Fitoussi and Stiglitz (2011) highlight the value in further investigating the well-being impacts of moving towards greater flexibility in the labour market: whilst labour market flexibility is assumed to deliver strong economic benefits, it could also negatively affect two key determinants of well-being, i.e. the quality of jobs and economic security. Although Shackleton (2012) warns about the risks of adopting a “one-size-fits-all” approach to well-being at work, particularly across different countries, having better data with which to examine these trade-offs is surely important.

As noted earlier, attempts to directly compare the different drivers of subjective well-being face a number of challenges, and there will be limits to the extent that these analyses contain *answers* to policy trade-offs *per se*. Some of the interpretive challenges associated with analyses of the drivers of subjective well-being – and particularly the issue of comparing different drivers – are discussed in more detail in the interpretation section that follows.

Methods: How can drivers of subjective well-being be analysed?

Data requirements, survey design principles and causality

Research aimed at better understanding the drivers of subjective well-being requires the inclusion of a wide range of co-variables in the analyses, including a number of standard demographic and control variables (described in Chapter 3), as well as measures of the drivers of interest, and their potential co-variables. Examples of key variables of interest are outlined in Diener, Diener and Diener (1995), Dolan, Peasgood and White (2007), Fleche, Smith and Sorsa (2011), and Boarini et al. (2012). Analyses of drivers require access to micro-level data, and may be undertaken by government analysts, researchers and organisations or institutes with an interest in informing government policy, academic enquiry and public discourse, as well as in organisational well-being (including business approaches to employee well-being).

Ideally, research into the drivers of subjective well-being should utilise data that enable some inferences about the causality of relationships between variables. Although the gold standard for determining causality involves experimental manipulations of hypothesised drivers under controlled conditions, this is close to impossible for most of the policy-relevant determinants of life evaluations and eudaimonia in particular. The model scenario for determining causality in real-life settings therefore tends to be *randomised controlled trials* (RCTs), which involve the random allocation of individuals into groups, each

of which are assessed before and after receiving a different treatment (e.g. one group receives intervention A, one group intervention B, and a third group acts as the control, receiving no intervention). In practice, for many of the potential policy drivers of subjective well-being, RCTs are also rare, particularly in terms of ensuring perfect randomisation and double-blind or single-blind conditions.³⁰

Quasi-experimental designs refer to “natural experiments” where a group of respondents exposed to a particular intervention can be matched with and compared to a similar group of respondents not exposed to that intervention. However, investigators tend to have little control over either the level of variation in the determinant of interest, or in the allocation of treatment groups, which is rarely completely random. It may also be difficult to obtain pre-intervention outcome measures – thus inferences may have to be based on measures collected only after the event has occurred. However, quasi-experimental designs do offer some advantages over RCTs, particularly where it would be unethical to randomise treatments, and/or where experimental designs can be challenged in terms of their real-world applicability.³¹ Quasi-experimental designs can also enable researchers to draw on much larger and more representative data sets, including national-level data.

Quasi-experimental designs typically require panel data (longitudinal surveys collecting repeated measures of individuals over time) or the collection of pre- and post-data for the populations of interest. These data offer the opportunity to explore whether a change in the level of a given determinant is associated with a subsequent change in subjective well-being over time.³² While panel data do not enable the researcher to experimentally manipulate the main variables of interest, and panels can suffer from attrition, this approach has the benefit of being able to utilise data sets from large and high-quality samples such as those obtained by national statistical agencies – thus enhancing the representativeness of the sample, and the generalisability of the findings.

Large sample sizes are particularly important for detecting the impact of minor drivers and/or drivers that typically affect only a small proportion of the overall population. In comparison to more experimental methods (such as RCTs), observational data also carry less risk of experimental demand characteristics (e.g. Hawthorne or placebo effects), where a respondent’s knowledge that he or she is part of a special treatment group may influence subjective well-being outcomes and/or how they are reported. The same is true of international comparisons, which offer another form of natural experiment, where a particular intervention has been applied in one country but not in another. However, it is very difficult to infer causality from international comparisons of cross-sectional (rather than longitudinal) data, given the variety of uncontrolled differences between countries in terms of both sample characteristics and other variables of interest.

At present, the majority of studies investigating the drivers of subjective well-being tend to rely on cross-sectional data, simply because these are the most widely-available data sets. Strictly speaking, these analyses are concerned with *co-variates* rather than *drivers*, although the term drivers will be used in the sections that follow to denote the underlying intention of the analyses described. Cross-sectional data do not enable causal inferences to be made directly, but can be interpreted alongside evidence about the direction of causality from other sources.

Methods for analysis: Tests of association

The most appropriate method for analysis depends largely on the type of data collected, the method of collection and the nature of the research question. The simplest test for the strength of a relationship between two variables is a bivariate correlation. The Pearson or product-moment coefficient can be calculated when the data are assumed to be normally distributed and the expected relationship between them is linear; Spearman's Rank and other non-parametric tests are available for ordinal data and non-linear relationships. Partial correlation enables examination of the relationship between two variables while removing the effect of one or two other variables. Correlations indicate the possible existence of a predictive relationship between two variables, but they do not imply causation.

For more thorough examination of the drivers of subjective well-being in cross-sectional, international and longitudinal studies, regression analysis is widely adopted. Regression is a correlation-based statistical technique that examines how well a set of explanatory or *independent variables* can predict a given *dependent variable*, i.e. the chosen subjective well-being measure. Regression is particularly suited to complex real-life problems because it allows the impact of several independent variables to be assessed simultaneously in one model, and it can tolerate independent variables being correlated with one another. However, the "best" regression solution (in terms of variance explained per independent variable) is produced when each independent variable is strongly correlated with the outcome variable, but uncorrelated with other variables, whether these other variables are included or excluded from the model. If two correlated independent variables are both included in the same regression model, their relationship with the dependent variable may be obscured (Tabachnick and Fidell, 2001). However, if an independent variable is correlated with some other excluded variable with causal claims, then the included variable will falsely be given credit for explanatory power really due to the excluded variable (a difficulty commonly described as the "omitted variable problem").

Given the ordinal nature of subjective well-being measures, linear regression models (based on ordinary least squares estimates) are theoretically inefficient when compared to methods designed to analyse ordinal outcomes (e.g. Probit). However, Ferrer-i-Carbonell and Frijters (2004) have examined both methods in relation to subjective well-being drivers, and concluded that in practice there are few differences between estimates based on ordinary least squares estimates and Probit methods. Similar results were reported by Frey and Stutzer (2000), and reviewing the literature overall, Diener and Tov (2012) reach a similar conclusion. As the interpretation of ordinary least squares outputs is more straightforward, these are often the results reported. However, it remains advisable to examine both Probit and ordinary least squares approaches in the course of analyses to test for (and report on) any major differences between the results observed.

Where curvilinear relationships are expected, such as in the case of both income and age in predicting subjective well-being, squared values (in the case of age, where the expected relationship is U-shaped) and log values (in the case of income, where the expected relationship is asymptotic) are typically used in regression models.

Other analytical options that may be used to investigate the drivers of subjective well-being include structural equation modelling, also known as causal modelling or analysis, analysis of co-variance structures or path analysis. Like regression, structural equation modelling involves examining a set of relationships between one or more independent variables and a dependent variable (or sometimes several dependent variables); but rather

than using raw measured variables as they are observed, structural equation modelling combines factor analysis³³ with regression – and involves multiple regression analysis of factors. The key advantage of this approach is that it enables complex pathways to be tested simultaneously, and by focusing on relationships among underlying factors (rather than measured variables), estimated relationships are thought of as “free” of measurement error.³⁴ Detailed discussion of structural equation modelling is beyond the scope of this chapter, but because it also involves association- and regression-based techniques, some of the issues raised below remain relevant.

What constitutes a significant association?

Correlation coefficients (here denoted as r) range from -1 to +1, with -1 signifying a perfect negative linear association, and +1 signifying a perfect positive linear association. The square of the coefficient (or r square) denotes the per cent of the variation in one variable that is related to the variation in the other. Thus, an r of 0.60 (r squared = 0.36) means that 36% of the variance in the dependent variable is related to the variance in the independent variable. The statistical significance of a given correlation coefficient indicates the likelihood that the coefficient would be found in a sample by chance when no significant association actually exists between the variables.

In regression-based analyses, the overall model “fit” with the observed data is described in terms of the proportion of variance in the dependent variable that is explained by the variance in the independent variables (the overall *multiple-correlation coefficient*, or R^2 value). Statistical significance is used to indicate whether the overall model provides better-than-chance prediction of the dependent variable.

In order to further understand how each independent variable contributes to the prediction of the dependent variable, one examines the set of *regression coefficients* for the independent variables. The size (and sign) of the coefficient for each independent variable indicates how much the dependent variable is expected to increase (if positive) or decrease (if negative) when the independent variable increases by one unit, while holding all the other independent variables constant.

Interpreting drivers of subjective well-being

Two key questions for interpreting analyses of the drivers of subjective well-being are *what size of impact can be expected?* and *how can the impacts of different drivers be compared?* The first question needs to be considered with reference to the overall sensitivity of subjective well-being measures, as well as the time frames for the analysis. The second question raises issues about the interpretation of regression coefficients – including problems of correlations among independent variables, the effects of unmeasured or omitted variables, the possibility of shared measurement error (or method variance) between variables and outcomes, the possible presence of reverse or two-way causality between variables and outcomes, and the generalisability of results.

What size of impact can be expected?

There are *a priori* reasons not to expect large movements in subjective well-being data as a result of single drivers. Many of the interpretive issues described in the first section of this chapter (looking at basic patterns of change over time and differences between groups) also apply to analyses involving the drivers of subjective well-being.³⁵ Factors such as the initial distribution of responses, the proportion of the sample affected, the number of

significant drivers in the model, frame-of-reference effects and adaptation can all limit the impact on subjective well-being that one might expect as a result of any one individual driver, as well as the size of sample required in order to detect that impact.

Senik (2011) notes that the typical model R^2 value of an ordinary least squares estimate of life evaluation varies between 3% and 15%, depending on the control variables and drivers included in the model and the sample size. Drawing on two waves of Gallup World Poll data (2009 and 2010) from all 34 OECD countries and including measures for a large number of key known drivers of subjective well-being, Boarini et al. (2012) obtained R^2 values of 0.35 in the case of life evaluations, and 0.19 in the case of affect balance. Fleche, Smith and Sorsa (2011) report cross-country comparisons of key life satisfaction drivers over two to three waves of data (from the World Values Survey 1994-2008) for 32 different countries and find R^2 squared values ranging from 0.40 in New Zealand to 0.14 in Turkey, with an OECD average of 0.22. In Helliwell, Barrington-Leigh, Harris and Huang's (2010) analysis of data from a global sample of between 50 000 and 140 000 respondents in 125 countries, income plus a range of social and cultural variables explained between 30 and 45% of the individual-level variance in life evaluations.

Given the very large number of potential drivers of subjective well-being, these R^2 values suggest that the proportion of variance explained by any one driver is likely to be small. Furthermore, if the initial amount of variability in a given driver is itself limited (for example, because only a small proportion of the sample is affected by it), then the proportion of variability in the subjective well-being outcome it can explain will also be limited. The key statistic of interest in the analysis of drivers, however, is not the overall model R^2 , but the size and significance of the individual regression coefficients associated with each driver – which (in the absence of correlations among independent variables) indicate how much the dependent variable is expected to increase or decrease when the independent variable increases by one unit.

As noted in Section 1, a mean change of 0.3 or 0.5 scale points on a 0-10 life evaluation scale may represent a very sizeable result that one might expect only in response to major life events at the individual level. Between countries, differences may be larger due to the cumulative impact of differences across a wide range of subjective well-being drivers.

Using an appropriate time horizon for analysis is another key consideration when interpreting effect sizes. As with any attempt to evaluate the impact of a driver, it will be important to consider the mechanisms through which that driver is assumed to operate – and how long it may take for these effects to emerge. Due to psychological resilience and adaptation over time, the immediate impact on subjective well-being for some life events and interventions may be greater than the impact several years down the line. As noted previously, the variables that influence the rate and extent of adaptation to life events over time may be of key interest to policy-makers, and thus adaptation should be considered a feature rather than a “bug” in subjective well-being data. Nonetheless, the process of psychological adaptation means that close attention to time horizons is warranted, in particular to avoid misinterpreting the effects of exogenous events.

How can the impacts of different drivers be compared?

Particularly in the case of investigating policy trade-offs, or when deciding between two different courses of action, there may be times when it is useful to focus on the *relative size* of different drivers of subjective well-being. Directly *comparing* the regression coefficients

associated with different drivers requires caution, however. Challenges include the need to consider units of measurement as well as the potential for correlation among the drivers, and problems potentially arising due to shared method variance or self-report biases. The generalisability of findings obtained through regression analysis is also of crucial importance for understanding policy implications.

Although the issues associated with comparison of regression coefficients apply to all regression-based analyses, they are particularly relevant here because there are often strong inter-correlations among the drivers of subjective well-being, and because a growing literature suggests some degree of reverse-causality between life evaluations and its drivers in particular. The problems of shared method variance and the possibility of self-report “bias” are also discussed. Finally, a key consideration for policy evaluations in particular are the time horizons over which well-being drivers might be expected to take effect.

Regression coefficients and correlations among independent variables. As noted earlier, when there are no interrelationships between independent variables, the size of the regression coefficient gives an indication of how a one unit change in the independent variable can be expected to influence the dependent variable. However, because the high-level drivers of subjective well-being (such as income, health status and social connections) are likely to be so strongly interrelated, interpretation of their individual contributions must proceed with caution, because there may be *mediation*, *confounding* and *suppression* effects in the data.

Although conceptually distinct, mediation, confounding and suppression each describe scenarios when relationships between an independent variable and a dependent variable are affected by the presence of a third related independent variable (the *mediator*, *confound* or *suppressor*). When the third variable is actually measured, these effects can be detected by a substantive change in the regression coefficient for the independent variable when the third variable is included in the model (relative to a model that excludes the third variable). In the case of mediation, the third variable is described as “transmitting” the effect of the independent variable to the dependent variable. In the case of confounding effects, the third variable is described as a “nuisance” variable, producing spurious correlations between the independent variable and the dependent variable.³⁶ Conversely, when suppression effects are present, relationships between the independent and the dependent variable become *stronger* when the third variable is included in the model. In the event that the third variable remains unmeasured, the coefficient observed for the independent variable can be misleading (see Dolan, Peasgood and White, 2007).

Boarini et al. (2012) also raise the possibility of “over-measurement” of individual drivers – where, if several measures of the same driver are included, correlations among the measures can crowd one another out, such that some relevant variables fail to reach significance in the overall model. This means that a significant driver could be overlooked if there are too many measures of it in the model, because the overall effect will be distributed among too many independent variables.

The fact that the regression coefficient for an independent variable is often dependent on the other variables in the regression equation means that selecting the variables to include in an analysis is a crucially important task. A clear theoretical structure and an understanding of the hypothesised causal pathways must underpin these decisions. While the use of hierarchical (or sequential) regression and structural equation modeling can

provide an analytical strategy for examining causal pathways among variables, techniques of this sort cannot provide a definitive solution with regard to the relative “importance” of interrelated independent variables, in terms of absolute size of impact.

The omitted variable problem. In addition to the mediating, confounding and suppression that can occur as a result of measured variables, causal inferences about relationships between variables can be severely hampered by *unmeasured* or “omitted” variables. Specifically, a significant *statistical* relationship can be observed between two variables not because there is a causal relationship between them, but because both variables are causally related to a third unobserved variable that has been omitted from the analyses.³⁷ Omitted variables can also suppress causal relationships between observed variables, causing results to fail to reach statistical significance due to unmeasured factors.³⁸ This is a problem right across econometric analyses (and all tests of association) and is in no way limited to examination of the drivers of subjective well-being.

Among subjective well-being data sets, because so many drivers help to explain final subjective well-being outcomes, several counter-intuitive findings (such as repeated failures to find relationships between income growth and subjective well-being, despite strong cross-sectional relationships between income and well-being) could potentially be explained with reference to variables that have been omitted from analyses (such as changes in *relative* income, or patterns of decline in other important determinants of subjective well-being, such as health, social connections, perceived freedom, corruption, etc.). The effects of relative income, and aspirations about income, have in particular been studied by several authors (for reviews, see Dolan, Peasgood and White, 2007, and Clark, Frijters and Shields, 2008) and reflect the *frame-of-reference effects* discussed in Section 1 of this chapter.

Another set of omitted variables often discussed in relation to subjective well-being involve personality- and temperament-based measures. Individual fixed effects do appear to account for a sizeable proportion of the variance in subjective well-being measures (see Diener, Inglehart and Tay, 2012), and some of these fixed effects may in turn reflect dispositional tendencies. For example, Lucas and Donnellan (2011) reported that 34-38% of variance in life satisfaction was due to stable trait-like differences – although this study did not include measures of the objective life circumstances that might impact on stable trait-like components. The issue of whether these stable differences reflect a true causal impact of personality and temperament on experienced subjective well-being, or simply a response style that affects self-reported measures (including subjective well-being, but also health and exposure to stress) is discussed in relation to shared method variance, below.

Self-report measures and shared method variance. A final factor to consider in the interpretation of subjective well-being drivers is shared method variance,³⁹ also known as common method variance, which can inflate the estimated impact of self-reported drivers relative to those measured through other means (such as objective observations). In particular, due to a combination of social desirability biases, response sets, differences in scale interpretation or use, and similarities between the questions themselves, one might expect that subjective well-being and other self-report measures such as self-rated health, self-reported mental health, self-reported social connections, and/or personality and dispositional variables might have correlated errors. Indeed, response formats to such questions are often very similar (e.g. 0-10 labelled scales). Furthermore, several items on

current measures of eudaimonia and affect bear a strong resemblance to some of the questions used to measure personality and mental health. Concepts such as self-efficacy, often included in constructions of eudaimonia, are also often considered to be an aspect of personality or dispositional tendency.

When comparing the effects of different drivers of subjective well-being, particularly in cross-sectional data, it is therefore important to consider *how* each of those determinants was measured. The possibility of shared method variance has led some authors to suggest that dispositional measures such as personality or negative affectivity⁴⁰ should be included as control variables in analyses of self-report data, and particularly when analyses are cross-sectional (Brief et al., 1988; Burke, Brief and George, 1993; McCrae, 1990; Schaubroeck, Ganster and Fox, 1992), in order to remove any bias associated with subjective self-report processes more generally. The risk in doing so, however, is that this could potentially swamp the effects of other important determinants, and remove variance in subjective well-being data that is likely to be of policy interest. For example, if exposure to childhood poverty or long-term health problems influences responses to both personality and subjective well-being questions, controlling for personality in the analyses could mask the true impact of childhood poverty and/or long-term health problems on the outcomes of interest. Personality, and negative affectivity in particular, may also play a substantive role in the overall development of subjective well-being (Moyle, 1995; Bolger and Zuckerman, 1995; Spector et al., 2000).

An alternative approach to investigating the issue of shared method variance and self-report bias is to use longitudinal panel data, in which individual fixed effects can be controlled. In such models, the ability of self-reported drivers to predict changes in subjective well-being over time can be investigated, and this is a much stronger test of causality. These types of analyses enable the effects of more objective indicators to rise to the forefront, whilst problems associated with shared method variance recede into the background.

Reverse and two-way causality. Understanding the direction of causality when examining drivers of subjective well-being is crucial to establishing their policy-relevance. As noted in the *data requirements* section above, an analyst's ability to make causal inferences is strongest where experimental data, or data from randomised controlled trials, is available. Quasi-experimental designs and longitudinal panel data can also offer insights into likely causal relationships, because analyses can be restricted to factors that temporally precede changes in subjective well-being over time. In cross-sectional data, the ability to make causal inferences is severely limited – and thus results need to be interpreted alongside evidence about the direction of causality from other sources.

In regression-based analyses, one method for exploring issues of reverse-causality is to include an *instrumental variable*.⁴¹ Instrumental variables are sometimes used when there are problems of endogeneity in regression models – i.e. when the independent variable of interest is correlated with the model error term. Two-way or reverse causality can be a key source of endogeneity, as can omitted variables (described above). Dolan and Metcalfe (2008) and Powdthavee (2010) report using instrumental variables to obtain better estimates of the exogenous effect of income on life evaluations. This typically increases the estimate of the income coefficient (Fujiwara and Campbell, 2011). In practice, however, it is very difficult to identify appropriate instrumental variables for income, as most of the key variables strongly associated with income also tend to be associated with life satisfaction.

Generalisability of results. Analyses of drivers are strongly affected by both the variables included in the model and the population sampled – which in turn both influence the extent to which results can be generalised. The importance of different drivers of subjective well-being may vary systematically according to certain group characteristics, because different groups within and across societies may be characterised by very distinct initial resource endowments. For example, Boarini et al. (2012) examined the determinants of life satisfaction among different population sub-groups (i.e. by gender, age and participation in the labour market) across 34 OECD countries. While the overall pattern of coefficients was quite similar, there were a number of non-trivial differences in the subjective well-being functions⁴² observed in the different groups.

This evidence suggests that, for different population sub-groups, the relative impact of the determinants of subjective well-being may differ. Heterogeneity in the relative size and significance of the drivers of subjective well-being has implications for how we might inform the public about the relative importance of the different drivers. Policies aimed at increasing subjective well-being may also need to consider the distribution of well-being resource endowments among different population sub-groups. Regression analyses generate results for the *average* individual – and in practice, there may be wide individual differences in the specifics of the well-being function. Different people may find happiness in different ways.

Although several studies have highlighted strong consistencies among affluent countries in terms of the direction and significance of some of the high-level determinants of subjective well-being (Helliwell and Barrington-Leigh, 2010; Fleche, Smith and Sorsa, 2011), one might also expect to see some differences in subjective well-being functions between countries, because countries also vary in terms of both their initial resource endowments and how those resources are distributed. For example, Inglehart et al. (2008) found that among less economically-developed countries, there were stronger associations between happiness and in-group solidarity, religiosity and national pride, whereas at higher levels of economic security, free choice becomes a more important predictor. Drawing on data from the Gallup World Poll, Bjørnskov (2010) reports that Cantril Ladder life evaluations showed a strong relationship with levels of GDP per capita among countries with lower relative incomes, whereas social trust became a strong and significant determinant only among countries with higher relative incomes. In the same vein, Helliwell and Barrington-Leigh (2010) show that coefficients on a number of social variables are higher in OECD than in non-OECD countries, while the coefficients on log income were identical in the two parts of the global sample.

3. Subjective well-being as an input to cost-benefit analysis

Introduction

The first two sections of this chapter have largely been concerned with analyses in which subjective well-being is the ultimate outcome of interest. But in addition to the intrinsic value of knowing more about subjective well-being, subjective well-being data can play an important role as an input for other analyses – offering insights into human behaviour and decision-making, as well as on how other well-being outcomes develop (Box 4.6). Thanks, in part, to these kinds of insights, subjective well-being data has also been suggested as an alternative means for estimating the monetary value of non-market factors (i.e. goods and services that do not have market prices) for the purposes of cost-benefit analysis.

Box 4.6. Subjective well-being insights into health, human behaviour and decision-making

Beyond the intrinsic value of subjective well-being, evidence suggests it is also important to other aspects of human functioning. Fujiwara and Campbell (2011) summarise a broad range of evidence suggesting that individuals with higher levels of subjective well-being are more likely to get married, earn more money and be healthier. Positive affect, negative affect and measures of life evaluations are associated with better long-term health and greater longevity (Danner, Snowdon and Friesen, 2001; Ostir et al., 2001), as well as shorter-term cardiovascular and immune system functioning (Cohen et al., 2003; Kiecolt-Glaser et al., 2002; Steptoe, Wardle and Marmot, 2005) which may mediate longer-term relationships between emotions and health. For example, Pressman and Cohen (2005) report that people with high positive affect have been shown to be less likely to become ill when exposed to a cold virus, and more likely to recover quickly.

In addition to income and health, subjective well-being may have other implications for economic performance and overall well-being. Research has found prospective links between positive emotions and workplace performance ratings and productivity (Diener et al., 2002; Estrada, Isen and Young, 1997; Wright and Staw, 1999). Keyes (2006) also reports evidence that mentally healthy individuals missed fewer days of work, were more productive at work, and had fewer limitations in daily activities. Summarising existing evidence, Clark and Oswald (2002) report that measures of subjective well-being have been shown to predict the likelihood of job quits, absenteeism and non-productive work, as well as the duration of unemployment. Bertrand and Mullainathan (2001) also found that job satisfaction was a strong predictor of the probability of changing jobs in the future.

Finally, subjective well-being data can offer insights into people's (in)ability to estimate the well-being impacts of both market and non-market factors as well as of life events – enabling us to compare estimates of what makes us happy against the level of happiness actually attained as a result. This work suggests that our ability to predict future well-being gains or losses (or our *affective forecasting*) is subject to various biases, such as irrelevant cues* (Sugden, 2005), lack of sensitivity to the size of the good or service valued (Kahneman and Tversky, 2000; Sugden, 2005), and focusing illusions, whereby “nothing that you focus on will make as much difference as you think” (Schkade and Kahneman, 1998; Wilson and Gilbert, 2005). These biases can produce marked discrepancies between the degree to which we think we want something, and the degree to which evidence suggests it will actually make us happy (described by Gilbert and Wilson, 2000, as “miswanting”). These findings have practical consequences for current methods of cost-benefit analysis (see below), as well as more general relevance, providing individuals with better information about the correlates of subjective well-being, so that they can make more informed choices in their own pursuit of happiness.

* An example is the starting point bias, whereby questions that begin: “Would you pay \$x for...?” can heavily bias responses towards the valuation of x used in the first question (Sugden, 2005).

This section focuses on how subjective well-being data can complement existing approaches to the valuation of non-market factors. It begins with a brief description of cost-benefit analysis and of why it is useful to place a monetary value on non-market factors. It then describes methods currently used for estimating the monetary value of non-market factors, and the ways in which subjective data may be able to complement these methods. Finally, the section briefly discusses some interpretive challenges and the caveats that need to be applied to valuations obtained through the use of subjective well-being data.

What is cost-benefit analysis, and how can subjective well-being data help?

Cost-benefit analysis (CBA) is one of the tools that governments and organisations often use to inform decision-making about complex social choices that include a variety of different well-being outcomes, including economic, social and environmental outcomes. CBA involves quantifying a number of the foreseen costs and benefits associated with a particular project or policy intervention. The information from this analysis can then be used as part of a wider decision-making process, which may involve a variety of other information sources, such as the results of public consultations and/or the net costs and benefits associated with alternate policy or programme options competing for the same funding. The key measurement challenge in CBA concerns finding methods to adequately value all potential costs and benefits, based on a common metric. As used by economists, the common denominator of choice is usually a monetary value, and the costs and benefits investigated tend to focus on those with established market values.

Approaches to valuation

Where costs (and/or benefits) have explicit economic values observable in the marketplace, these values can be used in the estimation process – although as market prices often fail in some ways, adjustments may be necessary. However, for non-market factors, alternative valuation methods are required to estimate monetary value. *Revealed preference* techniques involve calculating shadow prices, inferred from observed behaviour. *Stated preference* techniques on the other hand involve surveying respondents about their “willingness to pay” in order to gain or avoid a certain outcome, and/or their “willingness to accept” compensation to give up a good or put up with something undesirable.

Both preference-based techniques make the assumption that people make choices on the basis of what will maximise their future well-being, and this will be directly revealed by their patterns of expenditure. Despite numerous challenges to these assumptions, preference-based methods have become standard practice for public policy appraisal in the United States and the United Kingdom (Dolan and Metcalfe, 2008). Sugden (2005) also highlights that public policy-making requires *some way* of accounting for the impact of non-market factors on well-being if policies options are to be rationally compared. This is essential whether or not preference-based approaches, or even CBA itself, are deemed to be the correct methods by which to proceed.

The role of subjective well-being data in valuation

All preference-based approaches rely on people’s ability to make rational and accurate judgements about how something will make them feel in future. This is also true of market prices. However, evidence from psychology and behavioural economics suggests that people’s rationality may be bounded at best and “coherently arbitrary” (Ariely, Loewenstein and Prelec, 2003) at worst. In particular, various biases have been identified that distort estimates of well-being gained from various experiences (Sugden, 2005; Kahneman and Tversky, 2000; Sugden, 2005; Schkade and Kahneman, 1998; Wilson and Gilbert, 2005; see Box 4.6). Fujiwara and Campbell (2011), Sugden (2005) and Frey, Luechinger and Stutzer (2004) review how these biases challenge stated preference-based approaches to the valuation of non-market factors.

An alternative approach to valuation involves using life evaluations (usually life satisfaction) data to directly estimate the impact of a particular outcome on how people feel after the event – thus replacing a hypothetical judgement with *ex post* calculations of impact based on the level of subjective well-being actually achieved. Relative to stated preference approaches, this should remove problems associated with, for example, focusing illusions, because respondents are not prompted to think about the *source* of their well-being. There is also less risk of strategic responding on the part of individuals, i.e. intentionally over- or under-estimating the value of a good due to personal interests in the outcome of a valuation process.

Methods: How can subjective well-being data be used to value non-market factors?

Clark and Oswald (2002) suggest a method by which the life satisfaction gained or lost from experiencing certain life events can be converted into a monetary figure. This is done by estimating the life satisfaction gain or loss achieved, controlling for relevant background characteristics including income, in a regression analysis. The coefficients from this calculation are then used to estimate the amount of income that would be required to hold life satisfaction constant after the occurrence of a particular life event. Ideally, one would perform this analysis with longitudinal panel data, so that *transitions* from one state to another can be explored, lending more confidence to the interpretation of causality and giving insight into the typical duration of subjective well-being reactions.

Using this technique, Clark and Oswald (2002) calculated that (at 1992 prices) getting married produced the same impact as an additional GBP 6 000 per month, whilst widowhood was estimated to be equivalent to losing GBP 14 000 per month. The same authors found that the impact of becoming unemployed was far greater than simply the loss of income incurred – with a monthly payment of GBP 23 000 required to offset the negative effects of unemployment. Finally, the impact of moving from “excellent” self-rated health to “fair” self-rated health was estimated as being equal to a loss of GBP 41 000 per month. To put these figures into context, the average monthly household income over the whole sample (7 500 individuals) was just under GBP 2 000. Frey, Luechinger and Stutzer (2004) have used life satisfaction data to estimate the monetary value that would be necessary to “compensate” for subjective well-being loss caused by terrorist activities in the most terrorism-prone regions of France, the United Kingdom and Ireland, as compared to the least terrorism-prone regions. Their findings indicated that between 1975 and 1998, a resident of Northern Ireland (compared to residents of Great Britain and the Republic of Ireland) experienced losses valued at 41% of total income.

Most of the work on valuation of non-market outcomes using measures of subjective well-being has focused on measures of life evaluation. However, Deaton, Fortson and Totoro (2009) looked at the value of life in sub-Saharan Africa using both life evaluation measures and affect measures. They found that affect measures produced a higher value of life than that obtained using life evaluation measures.

How do the results of life satisfaction-based valuations compare to preference-based approaches? Dolan and Metcalfe (2008) note the lack of empirical research directly comparing the two, but present results from a quasi-experiment looking at the impact of an urban regeneration project in Wales. They found that whilst urban regeneration had no impact on house prices, individuals from the control group (in an adjacent area without urban regeneration) would be willing to pay on average GBP 230-245 per year for the next

three years for the public and private benefits of such a project. Meanwhile, the life satisfaction valuation method placed the impact at between GBP 6 400 and GBP 19 000 in total (over an indefinite time span).

Based on both Dolan and Metcalfe's house price data and the willingness-to-pay estimates, the GBP 10 million scheme would not look like an efficient use of resources: benefits are estimated at zero on the basis of house price differences, and at only GBP 240 000 across all households in the willingness-to-pay example. Based on the life satisfaction estimates, however, the scheme brought between GBP 6.1 million and 18.1 million in overall benefits. Such a wide range of estimates suggests a combination of approaches may be most effective in estimating value.⁴³

An extension of this approach can be used in other valuation problems, including those where monetary values are not required. For example, an increasingly-used statistic in the public health field is quality-adjusted life years (QALYS). These can be used to help make decisions about which conditions and treatments to prioritise when allocating healthcare resources, particularly when investing in new healthcare technologies. According to this approach, health states are assigned a value (e.g. 0 = death; 1 = full health), and then multiplied by how long that state lasts. Dolan et al. (2009) suggest that subjective well-being data from patients could be used as a more direct way to estimate the quality-of-life improvements that result from specific health conditions and treatments. Specifically, they propose that subjective well-being could be assessed, alongside health and other important subjective well-being determinants, before and during various stages in a treatment.

Subjective well-being data can add value over other preference-based health valuation methods⁴⁴ – which involve asking either the public or patients to imagine hypothetical health-related scenarios – because both public and patient preferences can be at odds with the level of suffering actually reported by patients with those conditions (Dolan and Kahneman, 2008). For example, Dolan et al. report that in hypothetical time trade-off scenarios, individuals drawn from the general population estimate that the impact of moderate pain would be worse than the impact of moderate depression. However, when one examines both affect and life evaluation data from real patients suffering from each of these conditions, this ordering is reversed. When two subjective statements conflict, it is difficult to know which one is the “right” one; however, it seems that subjective well-being data from patients might make a useful contribution to the overall evidence base on which valuations are made.

Direction of change: Loss aversion

Evidence also suggests that particular attention should be paid to the direction of change in the variable being valued. In stated preference methods, respondents often exhibit *loss aversion* – where the negative psychological impact of a loss is expected to be greater than the corresponding positive impact of a gain in the same good (Sugden, 2005; Guria et al., 2005). Valuations based on subjective well-being data can offer critical insights into whether loss aversion reflects a *true* imbalance in the well-being derived from losses versus gains, or whether it is a result of some of the biases in decision-making noted earlier (Box 4.6). For the time being, it may be helpful to estimate valuations for losses and gains separately where possible, because the well-being impact of, for example, withdrawing a particular policy initiative may not be equivalent to the initial well-being impact of introducing it in the first place.

Using subjective well-being to adjust preference data

Where stated and revealed preference data are used in CBA, Layard, Mayraz and Nickell (2008) and Powdthavee (2010) propose that subjective well-being data can be used to resolve the problem of how to value costs and benefits accruing to people with different incomes. Direct use of prices in CBA assumes that \$1 is equally valuable to all parties concerned. However, valuation surveys tend to produce very different valuation estimates for people at different points on the income distribution – because those with higher incomes are typically willing to pay more for the non-market factor in question. This can distort results because it weighs the views of higher-income individuals more heavily than the views of lower-income individuals. Layard et al. (2008) suggest using subjective well-being data (evaluative happiness, life satisfaction, or a combination of the two) to estimate the marginal utility of income. Whilst this assumes the position that subjective well-being can be used as a measure of utility, a position on which not everyone agrees, it nonetheless provides a way of managing the impact of the marginal utility of income in an evidence-based manner.

Challenges in the interpretation of subjective well-being valuations

The valuation of non-market factors using subjective well-being data is still in its infancy. As the analytical methods on which valuations are based are very similar to those used to investigate the drivers of subjective well-being, all of the interpretive challenges discussed in the previous section also apply here. Rather than repeating the previous section, however, the focus here will be on the implications these issues have for how monetary valuations should be conducted and interpreted – including data requirements and co-variables to include in analyses. Four factors, in particular, bear on valuations based on subjective well-being:

- Sensitivity of life evaluations – and what can and cannot be valued through this approach.
- Measurement error in estimating regression coefficients.
- Correlations among independent variables and the co-variables to include in regression models.
- Time horizons over which analyses are conducted.

Sensitivity of life evaluations

As noted earlier in this chapter, life evaluation data are sensitive to major life events and show strong associations with a variety of other well-being outcomes. However, there are *a priori* grounds not to expect large movements in life evaluations as a result of relatively small-scale policy initiatives or non-market factors. If the life satisfaction valuation technique is used to assess drivers that operate on a much smaller scale, this can risk under-valuing non-market factors that nevertheless people do regard as important. Fujiwara and Campbell (2011) also note that life evaluations may not be sensitive to the non-use value of items such as cultural monuments – so again, scale is important, and it may be more realistic to look at the collective impact of these goods.

One subjective well-being indicator that might be more sensitive to immediate surroundings and activities – for example, small changes in environmental quality, or the availability of green space – is affect. Kahneman and Sugden (2005) propose an approach to valuation based on “experienced utility”. This is estimated from short-term affect data collected through the day reconstruction method, which includes information about both activities and locations, as well as the affective states accompanying those activities. The

effects of different activities or locations on positive and negative affect can then be reconstructed. Kahneman and Sugden do not propose a specific method for linking experienced utility and money – and further work is needed to address this – but they note that the life satisfaction valuation approach could be adapted for affect-based valuations.

One further limitation is that the valuation technique based on subjective well-being is retrospective, i.e. it cannot be used to project the *potential* impact of something that does not yet exist – in contrast to the hypothetical scenarios on which stated preferences are based. As policy-makers using cost-benefit analysis are frequently interested in assessing the potential effects of a policy that has not yet been put in place, analyses will often need to draw on examples of policy initiatives in other communities – where the generalisability of results to the population of interest may come with caveats.⁴⁵

Measurement error in estimating regression coefficients

Monetary valuations obtained using subjective well-being data are typically based on regression coefficients, and thus require a high degree of precision in estimating those coefficients. Measurement error among the set of drivers (independent variables) examined in the course of valuations can be especially problematic. Of particular concern is the measurement error in self-reported income – which risks reducing the income coefficient, leading to higher valuations of non-market factors. For example, Powdthavee (2009) found an increased income coefficient (producing lower valuations for non-market factors) where objective income information was obtained by interviewers through examination of payslips.

Dolan and Metcalfe (2008) and Powdthavee (2010) also report using *instrumental variables*⁴⁶ in subjective well-being valuations to obtain better estimates of the exogenous effect of income on life evaluations. Fujiwara and Campbell (2011) note that instrumenting for income typically increases the estimate of the income coefficient, thus producing lower overall valuation estimates for non-market factors. For example, in Dolan and Metcalfe's analysis, this correction brings estimates of the value of urban regeneration down from GBP 19 000 to around GBP 7 000. Powdthavee reports that this technique lowers the valuations of marriage from around GBP 200 000 to GBP 3 500 per annum. In practice, however, it is very difficult to identify appropriate instrumental variables for the purposes of valuations, as most of the key variables strongly associated with income also tend to be associated with life satisfaction.

Measurement error in non-market factors could also reduce coefficients attached to the variable in question, leading to under-valuation. Conversely, if measurement error in non-market factors is positively correlated with measurement error in the life evaluations (for example, due to shared method variance or response biases), this could inflate rather than depress their coefficients, leading to over-valuation, unless the income variable was similarly affected. Again, instrumental variables could be of particular use in separating out causal effects from correlated errors.

The large impact that measurement error in independent variables can have on valuations means that, particularly when small or non-representative samples are used in regressions, it will be essential to check the coefficients obtained for income and other variables in the model (and especially for the non-market factor in question) to ensure that they fall within the range that might be expected, based on larger and more representative samples – and preferably those utilising high-quality panel data (Box 4.7). Further work on potential instrumental variables for use in valuations will be important for future development of the technique.

Box 4.7. The range of income estimates observed in the life satisfaction literature

The method used to estimate the value of non-market factors on the basis of life satisfaction data is very sensitive to the coefficient estimated for income. When interpreting the results of such valuations it is therefore helpful to consider the range of coefficients for income identified in the wider literature.

A very large number of authors have examined the role of income in life evaluations and the issue of whether increases in a country's average income over time are associated with increases in a country's subjective well-being (e.g. Easterlin, 1974, 1995; 2005; Hagerty and Veenhoven, 2003; Sacks, Stevenson and Wolfers, 2010). In practice, the effects of income are highly complex, and can vary both between countries and within different population sub-groups. Some authors report a consistent finding that income plays a more important role in developing and transition countries, and a less important role in more affluent societies (Bjørnskov, 2010; Clark, Frijters and Shields, 2008; Sacks, Stevenson and Wolfers, 2010), whereas others report a similar magnitude effect for income across all countries (Deaton, 2008; Helliwell, 2008; Helliwell and Barrington-Leigh, 2010). Estimates for income coefficients are also critically sensitive to other variables included in the regression model. Clark, Frijters and Shields (2008), Sacks, Stevenson and Wolfers (2010), and Fujiwara and Campbell (2011) provide a more extensive overview of several other important issues – including those associated with reverse causation, individual effects and the importance of relative income (i.e. an individual's income in comparison to a given reference group). A final issue is the problem of income non-response rates, which are rarely reported but could also affect coefficients estimated for income.*

Although the results from any one study should be interpreted with caution, research described below illustrates how estimates for the effect of income can vary when different background characteristics and life circumstances are controlled. In all cases, a 0-10 life evaluation measure and log-transformed income data are used.

- Sacks, Stevenson and Wolfers (2010) report results from several large life evaluation data sets, together spanning 140 countries. In cross-sectional data (pooled across all countries) and controlling only for country fixed effects, regression coefficients for log household income on Cantril Ladder life evaluations range from 0.22 to 0.28. Results remain similar when controlling for age and sex, while adjusting for the effects of permanent income or instrumenting income increases coefficients to between 0.26 and 0.5. The authors conclude that at the within-country level, the coefficient for the permanent effect of income lies somewhere between 0.3 and 0.5. They also suggest similar magnitude effects at the between-country level and for changes in income over time.
- Boarini et al. (2012) use two waves of Gallup World Poll data (2009 and 2010) to examine the determinants of Cantril Ladder life evaluations among 34 OECD countries. Pooled across countries, the coefficient for log household income is estimated at 0.18 when only key background characteristics are controlled. Controlling for a variety of other individual-level well-being outcomes (health problems, social connections, environmental quality, personal security, having enough money for food), the coefficient reduces to 0.13. When regressions for different population sub-groups are examined, the coefficient for log income is very similar for men and women (around 0.15), but much larger for those of working age (around 0.18) in comparison to the youth and the elderly (around 0.10). Drawing on a much larger number of countries involved in the Gallup World Poll (125 in total), Helliwell, Barrington-Leigh, Harris and Huang (2010) found coefficients of around 0.4 for log household income. Their analyses controlled for a wide range of variables, including demographics, social connections, religion, perceived corruption, charitable giving (time and money) and food inadequacy (not enough money for food), as well as GDP per capita and a national-level measure of food inadequacy. The food inadequacy measure was defined net of its strong and significant correlation with household income – and this raised the estimated coefficient on household income.

Box 4.7. The range of income estimates observed in the life satisfaction literature (cont.)

- Frijters, Haisken-DeNew and Shields (2004) looked at the effect of the large increase in real household income in East Germany on life satisfaction following the fall of the Berlin Wall in 1989. In the 10-year period between 1991 and 2001, the authors estimated that around 35-40% of the observed increase in average life satisfaction was attributable to the large (over 40%) increase in real household incomes during this time period, with a one-unit increase in log income corresponding to around a 0.5 unit increase in life satisfaction for both men and women.
 - Dolan and Metcalfe (2008) examined the life satisfaction impact of an urban regeneration project in a quasi-experimental design, which involved comparing two different communities in Wales. Background variables such as gender, age, relationship status and employment status were controlled in analyses. Whole sample analyses failed to find a significant effect of log household income on life satisfaction. With analyses restricted to individuals of working age, coefficients for household income were observed in the range of 0.65 to 0.93. These authors also note that in additional analyses, controlling for measures of social capital reduced the income coefficient by a non-trivial amount.
- * Although hard data is rarely reported, it appears that a relatively high proportion of individuals refuse to answer questions about their income in non-official and telephone-based surveys. For example, Smith (2013) estimated that income non-response rates ranged from under 5% to just over 35% for countries involved in the 2008 European Values Survey, with a similar variation in the World Values Survey. Gasparini and Gluzmann (2009) found non-response rates in the 2006 Gallup World Poll among Latin American and Caribbean countries to range from 2% (in Ecuador) to 39% (in Trinidad and Tobago). The potentially non-random nature of income non-responses (Riphahn and Serfling, 2005; Gasparini and Gluzmann, 2009; Zweimüller, 1992) and the various techniques deployed to manage them (which range from dropping all observations from the analysis, to imputation methods) can potentially impact on estimated coefficients. Further research and reporting on this phenomenon is needed.

Co-variates to include in the regression model

As noted earlier, any attempts to compare coefficients obtained through regression analyses need to consider the possible impact of correlations among independent variables. When using life satisfaction data to value non-market factors, Fujiwara and Campbell (2011) recommend that measures for all known determinants of well-being should be included in the model. Although they note the lack of consensus in this area, they list key determinants from the literature as: income, age, gender, marital status, educational status, employment status, health status, social relations, religious affiliation, housing, environmental conditions, local crime levels, number of children and other dependents (including caring duties), geographic region, personality traits (such as extroversion) and the non-market factor being valued.

The same authors also note that for policy purposes, there may be some indirect effects that need to be controlled in valuation regressions to fully estimate the impact that a *marginal change* in a non-market factor may have on subjective well-being. They take the example of pollution, noting that although pollution is expected to have negative effects on subjective well-being, individuals may be partially compensated for those effects through lower house prices and reduced commuting times. These offset the overall impact of pollution on subjective well-being and may cause the *true value* of a marginal reduction in pollution to be underestimated. Frey, Luechinger and Stutzer (2004) note the same difficulty in estimating the impact of living in terrorism-prone areas, where higher wages and lower rents potentially compensate individuals – and these authors conclude that all potential channels of compensation need to be controlled for.

The difficulty in attempting to control for all possible drivers and indirect effects is that this may crowd out the variables of interest. For example, including controls that also co-vary with income in the regression equation may shrink the coefficient for income, thus shrinking the increase in life satisfaction brought about by each additional per cent increase in income. Underestimating the impact of income can therefore risk over-valuing the impact of non-market factors. However, the same under-valuation risk is also present for the non-market factors. For example, if the effects of air pollution on life evaluations are mediated by respiratory health conditions, the coefficient for a measure of air pollution is likely to be substantially reduced if a measure of respiratory health conditions is included in the model. This would lead to a lower valuation of air pollution than if the health variable were excluded from the model. The choice to include or exclude other variables in the regression therefore depends on the assumed causal pathways – and these must be clearly described when conducting valuations. Mediation analyses therefore need to play an essential role in preparing models, to better understand how predictors interact with one another. In further developing this valuation technique, it is also important to establish a better overall consensus regarding which explanatory variables should be included in (and excluded from) a valuation regression, and under what circumstances. As noted previously, it may be helpful to report results as a range of values, derived from various different models with and without the presence of certain control variables.

Time horizons

Whereas more traditional approaches to valuation enable time horizons to be specified if necessary (i.e. a period of time over which respondents would be willing to pay, or willing to accept, a particular non-market good), the subjective well-being valuation approach does not come with a fixed time-frame. Dolan and Metcalfe (2008) note that current valuation approaches fail to address the issue of how long changes in subjective well-being are thought to last as a result of the effects of a particular non-market factor.⁴⁷

Time horizons are also important for being able to adequately detect and value the impact of events, interventions or non-market factors. Consideration of the mechanisms through which subjective well-being impacts are realised, and the time frames over which those mechanisms operate, is thus important in selecting data sets for valuation purposes and in interpreting the findings (Frey, Luechinger and Stutzer, 2004). The potential for partial adaptation to life events over time (see Section 2) also has implications for valuations. For example, the subjective well-being impact of widowhood ten years after the event is likely to be different to the subjective well-being impact just one year after the event. Thus, for the valuation of exogenous events, it is important to specify, *a priori*, the time frame(s) of interest and to collect, examine and interpret the data accordingly. In particular, scope exists for measuring effects at given points in time following a change, as well as for establishing longer-term averages or cumulants.

Combining non-market and market prices in cost-benefit analysis

The ultimate goal of assigning monetary values to non-market goods and services is to enable them to be examined in cost-benefit analyses alongside goods and services with market prices. Because the subjective well-being valuation technique is still in its infancy, and tends to produce wide-ranging estimates of value, the UK government Treasury's current position (as articulated in the 2011 update to *The Green Book: Appraisal and Evaluation in Central Government*) is that whilst valuations based on life satisfaction might be a useful

way to quantify the *relative* value of two or more non-market goods, they are not yet robust enough to allow comparisons with market prices. An alternative to this view would be to generate and test a series of cost-benefit analysis models – starting with a benchmark model based on market prices and other estimations widely regarded as robust, and extending this with valuations based on preference-based approaches on the one hand, and subjective well-being on the other.

One example of this multi-stage approach is the work of Gyarmati et al. (2008), who undertook a comprehensive quasi-experimental evaluation of the Community Employment Innovation Project in Canada in order to estimate the overall costs and benefits of the project to individuals, communities and governments. The project was designed to evaluate a long-term active re-employment strategy for unemployed individuals who volunteered to work on locally-developed community projects in return for wages (as an alternative to receiving state-funded income transfers). A benchmark cost-benefit analysis model was based on administrative costs, participant earnings and the market prices of fringe benefits, taxes, transfer payments and premiums, as well as market-based estimations of the value of volunteering. An extended model was then developed that included a valuation of foregone leisure (based on 20% of earnings), as well as a valuation based on subjective well-being (based on Helliwell and Huang, 2005) of the social networks that respondents built as a result of their work placements, and the reduction in perceived hardship experienced. In the benchmark model, each dollar in net cost to government was estimated to produce between \$1.02 and \$1.39 in net benefits to society.⁴⁸ In the extended model, each dollar in net cost to government was estimated to bring between \$1.21 and \$1.61 of net benefits. Gyarmati et al. point out that one dollar direct cash transfer (one alternative to the employment project) has meanwhile been estimated to deliver only \$0.85 in net benefits to the intended recipient.

Conclusions

Fujiwara and Campbell (2011) summarise the advantages of the life satisfaction approach to valuation as follows: i) the cost and time-effectiveness of the data collection; ii) the ability to use statistics drawn from very large and representative samples (in contrast to stated preference techniques, which require a separate data collection exercise); iii) the possible application to a whole variety of life events and circumstances; iv) the presence of fewer biases and less strategic behaviour on the part of respondents; and v) the fact that the data do not rest on assumptions about market structure. Conversely, some of the disadvantages highlighted by these authors include: i) difficulties in estimating the marginal utility of income, including the effects of relative as compared to absolute income, as well as the indirect effects of income and variables that operate counter to the effects of income; and ii) difficulties in estimating the marginal utility of the non-market factor, including indirect effects and the consumption of complementary goods alongside the non-market factor. All of the main approaches to monetary valuation of non-market factors (revealed preference, stated preference and subjective well-being-based estimates) are associated with methodological shortcomings, but the nature of these shortcomings is different in each case. Using several different methods provides more information than relying on a single approach, and using subjective well-being data offers a relatively low-cost option that avoids some of the biases connected to preference-based approaches. However, as the method based on subjective well-being is still in its infancy, significant methodological and interpretive questions remain, and it should therefore be regarded as a complement to rather than a replacement for existing methods.

Notes

1. <http://hdr.undp.org/en/statistics/hdi/>.
2. It is important to note that these findings are based on worldwide visitors to the OECD's *Your Better Life Index* website, <http://oecdbetterlifeindex.org/>, a sample of individuals known to be non-representative and non-random, and as such they should be interpreted with care.
3. *Frame-of-reference effects* refer to differences in the way respondents formulate their answers to survey questions, based on their own life experiences, as well as their knowledge about the experiences of others – including both those they consider as within their “comparison group” and those outside it.
4. *Adaptation* refers to psychological processes that may either restore or partially repair subjective well-being, and particularly affective experiences, in the face of some types of adversity. People may also show adaptation to positive life events over time (whereby the initial subjective well-being boost delivered by a positive change in life circumstances, such as marriage, reduces over time).
5. Much of the critique surrounding the use of subjective well-being for public policy centres around a view that increasing positive emotions is not an appropriate goal for governments (e.g. Booth et al., 2012; McCloskey, 2012). Although this view potentially underestimates the health and well-being implications of emotional experiences (described above), and fails to distinguish between the usefulness of *monitoring* positive emotions, versus making them *primary objectives* of government policy, it is nonetheless further grounds to avoid describing subjective well-being data solely in terms of “happiness”.
6. Ordinal data are those measured on scales where the intervals between scale points are not assumed to be equal, but there is an underlying sequence or rank order. For example, we assume that a 5 is lower than a 6 and a 6 is lower than a 7, but we do not assume that the distance between 5 and 6 is equivalent to the distance between 6 and 7. Linear regression relies on continuous variables, where cardinality is assumed, i.e. where the size of the number on a scale is expected to have a direct linear relationship with the amount of the variable in question. Tabachnick and Fidell (2001), however, note that in the social sciences, it is common practice to treat ordinal variables as continuous, particularly where the number of categories is large – e.g. seven or more – and the data meet other assumptions of the analysis.
7. Data users are likely to want to know, for example, *how* good or bad a person's experience was, not just on which side of a cut-off they fall. This is less of a problem for the reporting of headline national aggregate figures, but becomes particularly relevant when comparing responses between groups. It can be ameliorated to some extent by banding responses into several categories rather than selecting just one cut-off point.
8. For example, a country with universally low levels of subjective well-being would have few individuals falling below the relative poverty line, thus masking the extent of difficulties faced.
9. For example, the thresholds associated with clinically-significant mental health outcomes may be very different to the thresholds associated with different educational or income levels.
10. There are some who disagree with this, arguing that cardinal interpretations of subjective well-being are possible – e.g. Ng (1997).
11. It is unclear, for example, what it really means to say that the bottom 10% of the population achieves only 1% of the total subjective well-being. This can be contrasted with income, where it is easier to understand the practical implications of the bottom 10% earning just 1% of the total income across a population.
12. Although Helliwell, Barrington-Leigh, Harris and Huang (2010) took the simple mean average of the Cantril Ladder and a single-item life satisfaction measure and found this was more closely correlated with predictors of subjective well-being (such as demographics, income and a set of social indicators) than either measure on its own.
13. The UK's ONS have also proposed a single-item eudaimonia question, for high-level monitoring purposes: “Overall, to what extent do you feel the things you do in your life are worthwhile?” (ONS, 2011b).
14. In the case of more detailed analyses, such as group comparisons or investigation of the drivers of subjective well-being, separate estimates of the different sub-components of subjective well-being will be preferred due to the risk of information loss when summing across sub-components.

15. For example, if the threshold on a 0-10 scale is set at 7, movements across that threshold will be very salient, but large-scale movements from 8 to 9, or from 2 to 5, may go undetected.
16. Represented as a percentage of the population reporting higher positive than negative affect; OECD calculations based on figures from the 2010 Gallup World Poll.
17. I.e. a change over a one, five or ten-year period. As discussed earlier, short-term fluctuations of this magnitude can also be detected, but may not represent meaningful societal shifts in overall levels of well-being (e.g. Deaton, 2012).
18. For example, the *World Happiness Report* (Figure 2.3) lists 63 countries where the mean average life evaluation between 2005 and 2011 (measured on a 0-10 Cantril Ladder scale) is lower than the scale midpoint, 5. These include India, China, Iraq, Afghanistan and Syria; and particularly low-scoring countries, with scores below 4.0, include Congo, Tanzania, Haiti, Comoros, Burundi, Sierra Leone, the Central African Republic, Benin and Togo.
19. *Reverse causality* in this context refers to when subjective well-being drives the independent variable, rather than vice versa. For example, in a cross-sectional analysis, a significant association could be observed between income (the independent variable) and subjective well-being (the dependent variable), but this could be because subjective well-being drives income (rather than vice versa). *Two-way causality* is where there are reciprocal and causal relationships between two variables in both directions – i.e. income drives subjective well-being, but subjective well-being also drives income. *Endogeneity* refers to a situation where there is a correlation between an independent variable and the error term in a regression model. This can be due to measurement error, omitted variables, sample selection errors, and/or reverse or two-way causality.
20. To quote from Kahneman and Riis (2005): "... consider the Americans and the French. The distributions of life satisfaction in the US and France differ by about half a standard deviation. For comparison, this is also the difference of life satisfaction between the employed and the unemployed in the US, and it is almost as large as the difference between US respondents whose household income exceeds USD 75 000 and others whose household income is between USD 10 000 and USD 20 000 (in 1995)... Is it possible to infer from the large differences in evaluated well-being that experienced well-being is also much lower in France than in the USA? We doubt it, because the sheer size of the difference seems implausible" (p. 297).
21. Such as tendencies to use either extreme or more "moderate" scale response categories, as well as the likelihood of socially desirable responding.
22. There are times when people's subjective perceptions matter, even when they don't reflect objective reality: "cultural differences may in some cases be relevant to policy and in some cases irrelevant. For example, people's satisfaction with leisure opportunities might be relevant to policy deliberations, regardless of the objective conditions" (Diener, Inglehart, and Tay, 2012, p. 20). Another classic example of this is perceptions of regulatory burden, which can influence important business decisions and behaviour regardless of their accuracy (OECD, 2012). If perceptions of regulatory burden are misplaced, activity should be focused on better informing businesses and the wider public. Few would argue that the correct response would be to simply adjust the perceptions of regulatory burden so that they fit the pattern observed among more objective measures.
23. This Latin American effect has been explored in depth by Graham and Lora (2009).
24. See Chapter 2 for a full account of response styles.
25. As Senik (2011) puts it, "If the French evaluate the happiness of some hypothetical person in a less positive manner than the Danes, perhaps it is because they would actually feel less happy in the situation of that hypothetical person" (p. 8).
26. In practice, however, substituting even the highest adjusted figure for French natives (7.54) would only cause a very minor adjustment in country rankings overall, causing French natives (7.22) to exchange places with natives of Great Britain (7.38) only. Overall, mean average happiness ratings among natives in the 13 countries sampled range from 6.74 in the case of Portugal to 8.34 in the case of Denmark.
27. Suppose that two individuals both have the same levels of underlying positive and negative affect. But imagine that the first has a tendency towards extreme responding – so that on a 0-10 scale this individual reports 9 on positive affect and 7 on negative affect. The second individual has a tendency towards more moderate responding, thus reporting a 7 on positive affect and a 5 on negative affect. The net affect balance for both individuals will be +2. This of course assumes that response biases operate in a similar way for both positive and negative affects, which requires further examination.

28. A substantial part of the literature in this field uses *country* as a proxy for culture, inferring cultural differences on the basis of country differences. Whilst this is potentially problematic, the words *country* and *culture* are often used interchangeably in many studies on the subject.
29. Although the implications of the subjective well-being literature are often interpreted in terms of opportunities for policy *interventions*, subjective well-being has just as much potential to identify areas where existing government interventions can be redesigned or stopped altogether.
30. Double-blind conditions refer to scenarios where neither the respondent nor those implementing the intervention are aware of which treatment group a given respondent has been assigned to. Single-blind is where those implementing the intervention know which treatment group has been assigned to which respondent, but respondents are unaware.
31. Wider applicability can be challenged where there are concerns about the extent to which a given result might generalise to other situations, beyond the experimental or quasi-experimental setting.
32. For some research questions investigating international differences in subjective well-being, where the driver in question is hypothesised to operate at an aggregate country level, pooled cross-sectional time series data (i.e. international data containing repeated study waves and representative, but different, samples in each wave) may also enable some causal inferences.
33. Factor analysis is a statistical procedure that is conducted to identify distinct (relatively independent) clusters or groups of related items or variables (called factors). It is based on patterns of correlations among items or variables.
34. Because error is estimated and removed in the process of extracting the underlying factors. A “factor loading” is calculated for each item or variable, which reflects the variance it shares with the underlying factor – and all other variance is assumed to be error. When factors are used in the analysis (instead of measured variables), only this common variance is analysed, and thus measurement error is, in theory, purged from the data.
35. One interesting exception, however, is cultural biases. Although there is currently some evidence of cultural bias in direct country comparisons of mean average levels, there is currently little to suggest that cultural biases exert a problematic influence on cross-country analyses of the drivers of subjective well-being – and drivers tend to be reasonably consistent across countries (e.g. Fleche, Smith and Sorsa, 2011; Helliwell and Barrington-Leigh, 2010). The issue of the extent to which regression solutions are replicable and can be generalised from one sample or country to another is, however, an important consideration that is discussed below.
36. However, statistically, mediation and confounding are identical: both are indicated when the inclusion of the third variable in the model reduces the relationship between the independent variable and the dependent variable by a non-trivial amount.
37. An example would be a significant statistical relationship between the time I spend talking to the plant in my office and how much it grows per year – both of which are related to an (unmeasured) causal variable: how often I remember to water the plant.
38. An example here might be that a causal relationship between how often I remember to water my office plant and how much it grows is obscured by a third (unmeasured) variable: how much plant food my colleague gives it.
39. Shared method variance refers to variance that is attributed to the measurement method, rather than the constructs of interest. In the case of subjective well-being, the main concern is that if drivers are also measured through subjective self-report data, self-report biases (including retrospective recall biases, response styles, cultural bias, etc.) could inflate observed relationships between those drivers and the subjective well-being outcomes of interest.
40. I.e. a dispositional tendency towards experiencing negative affect.
41. An instrumental variable is one that has a direct association with the independent variable in question (e.g. income), but not with the outcome of interest (e.g. life evaluations).
42. The term *function* is used here to describe the overall pattern of relationships between the independent variables and the dependent variable, including the size and significance of coefficients.
43. Dolan and Metcalfe conclude that “we need much more research into the extent and the sources of the differences between these valuation methods” (p. 25), particularly given that the valuation through subjective well-being approach is still in its infancy and “literally thirty years behind that of generating monetary values from revealed and stated preferences”.
44. These include the “standard gamble” and “time trade-off” methods (see Dolan and Kahneman, 2008).

45. See the previous section on interpreting the drivers of subjective well-being for a more detailed treatment of the generalisability of results.
46. An instrumental variable is one that has a direct association with the independent variable in question (e.g. income), but that is associated with the outcome of interest (e.g. life evaluations) only via the independent variable in question.
47. Based on the current state of knowledge, these authors suggest that “there would seem to be good grounds for viewing the ICs (income compensation – i.e. valuations) as a total value over a finite horizon. Clearly, the actual assumption made on how life satisfaction incorporates future expectations is crucial to the methodology of the value of the non-market good by experiences, and merits further investigation” (p. 23).
48. Two different estimates were produced because different models were estimated for participants, based on which type of welfare payments they received from the government prior to their participation in the programme.

Bibliography

- Abelson, R.P. (1985), “A variance explanation paradox: When a little is a lot”, *Psychological Bulletin*, Vol. 97(1), pp. 129-133.
- Angelini, V., D. Cavapozzi, L. Corazzini and O. Paccagnell (2011), “Do Danes and Italians rate life satisfaction in the same way? Using vignettes to correct for individual-specific scale biases”, *Health, Econometrics and Data Group Working Paper*, No. 11/20, University of York, available online at: www.york.ac.uk/res/herc/hedgwp.
- Ariely, D., G. Loewenstein and D. Prelec (2003), “Coherent arbitrariness: Stable demand curves without stable preferences”, *The Quarterly Journal of Economics*, Vol. 118, pp. 73-105.
- Beegle, K., K. Himelein and M. Ravallion (2012), “Frame-of-reference bias in subjective welfare”, *Journal of Economic Behaviour and Organization*, Vol. 81, pp. 556-570.
- Bertrand, M. and S. Mullainathan (2001), “Do people mean what they say? Implications for subjective survey data”, *The American Economic Review*, Vol. 92(2), pp. 67-72.
- Bjørnskov, C. (2012), “Wellbeing and the size of government”, in P. Booth (ed.), ... *and the Pursuit of Happiness: Wellbeing and the role of Government*, London: Institute of Economic Affairs, in association with Profile Books.
- Bjørnskov, C. (2010), “How Comparable are the Gallup World Poll Life Satisfaction Data?”, *Journal of Happiness Studies*, Vol. 11, pp. 41-60.
- Blanchflower, D.G. (2009), “International evidence on well-being”, in *Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being*, pp. 155-226, University of Chicago Press.
- Blanchflower, D.G. and A.J. Oswald (2008), “Is well-being U-shaped over the life cycle?”, *Social Science and Medicine*, Vol. 66(8), pp. 1733-1749.
- Blanton, H. and J. Jaccard (2006), “Arbitrary metrics in psychology”, *American Psychologist*, Vol. 61, pp. 27-41.
- 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”, *OECD Statistics Working Papers*, No. 2012/03, OECD Publishing, DOI: <http://dx.doi.org/10.1787/5k9b9ltjm937-en>.
- Bok, D. (2010), *The Politics of Happiness: What government can learn from the new research on well-being*, Princeton and Oxford: Princeton University Press.
- Bolger, N. and A. Zuckerman (1995), “A framework for studying personality in the stress process”, *Journal of Personality and Social Psychology*, Vol. 69(5), pp. 890-902.
- Booth, P. (ed.) (2012), ... *and the Pursuit of Happiness: Wellbeing and the role of government*, London: The Institute of Economic Affairs, available online at: www.iea.org.uk/publications/research/and-the-pursuit-of-happiness, last accessed 10 August 2012.
- Bradburn, N., S. Sudman and B. Wansink (2004), *Asking Questions: The Definitive Guide to Questionnaire Design – from Market Research, Political Polls, and Social and Health Questionnaires*, San Francisco: Jossey-Bass.
- Brickman, P., D. Coates and R. Janoff-Bulman (1978), “Lottery winners and accident victims: Is happiness relative?”, *Journal of Personality and Social Psychology*, Vol. 36(8), pp. 917-927.

- Brief, A.P., M.J. Burke, J.M. George, B.S. Robinson and J. Webster (1988), "Should negative affectivity remain an unmeasured variable in the study of job stress?", *Journal of Applied Psychology*, Vol. 73(2), pp. 193-198.
- Burke, M.J., A.P. Brief and J.M. George (1993), "The Role of Negative Affectivity in Understanding Relations between Self-Reports of Stressors and Strains: A Comment on the Applied Psychology Literature", *Journal of Applied Psychology*, Vol. 78, No. 3, pp. 402-412.
- Carnevale, P.J.D. and A.M. Isen (1986), "The influence of positive affect and visual access on the discovery of integrative solutions in bilateral negotiations", *Organizational Behavior and Human Decision Processes*, Vol. 37(1), pp. 1-13.
- Chapple et al. (2010), "Subjective well-being and social policy", *European Commission Working Paper*, published in the *Social Europe* series, under the Directorate-General for Employment, Social Affairs and Inclusion, available online at: <http://ec.europa.eu/social/main.jsp?langId=en&catId=22>, last accessed 10 August 2012.
- Clark, A.E. (2003), "Unemployment as a social norm: Psychological evidence from panel data", *Journal of Labor Economics*, Vol. 21, No. 2, pp. 323-351.
- Clark, A.E., E. Diener, Y. Georgellis and R.E. Lucas (2008), "Lags and leads in life satisfaction: A test of the baseline hypothesis", *The Economic Journal*, Vol. 118, pp. 22-243.
- Clark, A.E., P. Frijters and M. Shields (2008), "Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles", *Journal of Economic Literature*, Vol. 46(1), pp. 95-144.
- Clark, A.E. and A.J. Oswald (2002), "A simple statistical method for measuring how life events affect happiness", *International Journal of Epidemiology*, Vol. 31, pp. 1139-1144.
- CLES Consulting and the New Economics Foundation (2011), "Big Lottery Fund national well-being evaluation", progress report and interim findings, prepared by CLES Consulting and presented to the Big Lottery Fund, February, available online at: www.biglotteryfund.org.uk/evaluation_well-being.htm?fromsearch=-uk, last accessed 9 August 2012.
- Cohen, S., W.J. Doyle, R.B. Turner, C.M. Alper and D.P. Skoner (2003), "Emotional style and susceptibility to the common cold", *Psychosomatic Medicine*, Vol. 65(4), pp. 652-657.
- Cohen, S. and S.D. Pressman (2006), "Positive affect and health", *Current Directions in Psychological Science*, Vol. 15, pp. 122-125.
- Cummins, R.A. (2003a), "Normative Life Satisfaction: Measurement Issues and a Homeostatic Model", *Social Indicators Research*, Vol. 64, pp. 225-256.
- Cummins, R.A. (2003b), "Subjective well-being from rich and poor", *Social Indicators Research Series*, Vol. 15, Part 3, pp. 137-156.
- Cummins, R.A. (2001), "The subjective well-being of people caring for a family member with a severe disability at home: a review", *Journal of Intellectual and Developmental Disability*, Vol. 26(1), pp. 83-100.
- Cummins, R.A., R. Eckersley, J. Pallant, J. van Vugt and R. Misajon (2003), *Developing a national index of subjective wellbeing: The Australian Unity Wellbeing Index*, *Social Indicators Research*, Vol. 64, pp. 159-190.
- Danner, D., D. Snowdon and W. Friesen (2001), "Positive emotions in early life and longevity: Findings from the Nun Study", *Journal of Personality and Social Psychology*, Vol. 80, pp. 804-813.
- Deaton, A. (2012), "The financial crisis and the well-being of Americans; 2011 OEP Hicks Lecture", *Oxford Economic Papers*, Vol. 64, pp. 1-26.
- Deaton, A. (2010), "Income, aging, health and well-being around the world: Evidence from the Gallup World Poll", in D.A. Wise (ed.), *Research Findings in the Economics of Aging*, Chicago: University of Chicago Press, pp. 235-263.
- Deaton, A. (2008), "Income, aging, health and well-being around the world: Evidence from the Gallup World Poll", *NBER Working Paper*, No. 13317, National Bureau of Economic Research.
- Deaton, A., J. Fortson and R. Titora (2009), "Life (evaluation), HIV/AIDS, and death in Africa", *NBER Working Paper*, No. 14637, National Bureau of Economic Research.
- Di Tella, R. and R. MacCulloch (2010), "Happiness adaption to income beyond "basic needs"", in E. Diener, J. Helliwell and D. Kahneman (eds.), *International Differences in Well-being*, New York: Oxford University Press.
- Di Tella, R. and R. MacCulloch (2008), "Gross national happiness as an answer to the Easterlin Paradox?", *Journal of Development Economics*, Vol. 86, pp. 22-42.

- Di Tella, R., R. MacCulloch and A.J. Oswald (2003), "The Macroeconomics of Happiness", *The Review of Economics and Statistics*, Vol. 85(4), pp. 809-827.
- Di Tella, R., R. MacCulloch and A.J. Oswald (2001), "Preferences over inflation and unemployment: Evidence from surveys of happiness", *The American Economic Review*, Vol. 91(1), pp. 335-341.
- Diener, E., M. Diener and C. Diener (1995), "Factors predicting the subjective well-being of nations", *Journal of Personality and Social Psychology*, Vol. 69, pp. 851-864.
- Diener, E., R. Inglehart and L. Tay (2012), "Theory and Validity of Life Satisfaction Scales", *Social Indicators Research*, published in an online first edition, 13 May.
- Diener, E., D. Kahneman, R. Arora, J. Harter and W. Tov (2009), "Income's differential influence on judgements of life versus affective well-being", in E. Diener (ed.), *Assessing Well-Being: The collected works of Ed Diener*, Social Indicators Research Series, Vol. 39, Dordrecht: Springer, pp. 247-266.
- Diener, E., R.E. Lucas and C. Napa Scollon (2006), "Beyond the hedonic treadmill: Revising the adaptation theory of well-being", *American Psychologist*, Vol. 61(4), pp. 305-314.
- Diener, E., R.E. Lucas, U. Schimmack and J. Helliwell (2009), *Well-being for public policy*, New York: Oxford University Press.
- Diener, E., C. Nickerson, R.E. Lucas and E. Sandvik (2002), "Dispositional affect and job outcomes", *Social Indicators Research*, Vol. 59(3), pp. 229-259.
- Diener, E., C.K.N. Scollon, S. Oishi, V. Dzokoto and E.M. Suh (2000), "Positivity and the construction of life satisfaction judgments: global happiness is not the sum of its parts", *Journal of Happiness Studies*, Vol. 1, pp. 159-176.
- Diener, E. and M.E.P. Seligman (2004), "Beyond money: Towards an economy of well-being", *Psychological Science in the Public Interest*, Vol. 5(1), pp. 1-31.
- Diener, E. and W. Tov (2012), "National Accounts of Well-Being", in K.C. Land, A.C. Michalos and M.J. Sirgy (eds.), *Handbook of Social Indicators and Quality of Life Research*, Springer: Dordrecht.
- Dolan, P. (2009), "NICE should value real experiences over hypothetical opinions", letter to *Nature*, published Vol. 462(5), p. 35.
- Dolan, P. and D. Kahneman (2008), "Interpretations of utility and their implications for the valuation of health", *The Economic Journal*, Vol. 118, pp. 215-234.
- Dolan, P., H. Lee, D. King and R. Metcalfe (2009), "How does NICE value health?", *British Medical Journal*, Vol. 229, pp. 371-373.
- Dolan, P. and R. Metcalfe (2011), "Comparing measures of subjective well-being and views about the role they should play in policy", *UK Office for National Statistics Paper*, July.
- Dolan, P. and R. Metcalfe (2008), "Comparing willingness-to-pay and subjective well-being in the context of non-market goods", *CEP Discussion Paper*, No. 890, London: Centre for Economic Performance, London School of Economics.
- Dolan, P., T. Peasgood and M. White (2007), "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.P. 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.A. (2005), "Feeding the illusion of growth and happiness: A reply to Hagerty and Veenhoven", *Social Indicators Research*, Vol. 74(3), pp. 429-433.
- Easterlin, R.A. (1995), "Will raising the incomes of all increase the happiness of all?", *Journal of Economic Behaviour and Organisation*, Vol. 27(1), pp. 35-48.
- Easterlin, R.A. (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.
- Easterlin, R.A., R. Morgan, M. Switek and F. Wang (2012), "China's life satisfaction, 1990-2010", *Proceedings of the National Academy of Sciences*, early edition 2012 (contributed 06 April 2012).
- Estrada, C.A., A.M. Isen and M.J. Young (1997), "Positive affect facilitates integration of information and decreases anchoring among Physicians", *Organizational Behavior and Human Decision Processes*, Vol. 72(1), pp. 117-135.

- Eurostat (2009), *Sustainable Development in the European Union: 2009 monitoring report of the EU sustainable development strategy*, Eurostat Statistical Books, available online at: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-78-09-865/EN/KS-78-09-865-EN.PDF.
- Ferrer-i-Carbonell, A. and P. Frijters (2004), "How important is methodology for estimates of the determinants of happiness?", *The Economic Journal*, Vol. 114, pp. 641-659.
- Fitoussi, J.P. and J. Stiglitz (2011), "On the measurement of social progress and well being: Some further thoughts", paper presented at the 16th World Congress of the International Economic Association, Beijing, 4-8 July.
- Fleche, S., C. Smith and P. Sorsa (2011), "Exploring Determinants of Subjective Wellbeing in OECD Countries: Evidence from the World Value Survey", *OECD Economics Department Working Papers*, No. 921, OECD Publishing, DOI: <http://dx.doi.org/10.1787/5kg0k6zlc5k-en>.
- Frey, B., C. Benesch and A. Stutzer (2007), "Does watching TV make us happy?", *Journal of Economic Psychology*, Vol. 28, pp. 283-313.
- Frey, B.S., S. Luechinger and A. Stutzer (2004), "Valuing public goods: The life satisfaction approach", *CESifo Working Paper*, No. 1158, Zurich: University of Zurich Institute for Empirical Research in Economics.
- Frey, B.S. and A. Stutzer (2002), "What can economists learn from happiness research?", *Journal of Economic Literature*, Vol. 40(2), pp. 402-435.
- Frey, B.S. and A. Stutzer (2000), "Happiness, economy and institutions", *The Economic Journal*, Vol. 110(446), pp. 918-938.
- Frijters, P., J.P. Haisken-DeNew and M.A. Shields (2004), "Money does matter! Evidence from increasing real income and life satisfaction in East Germany following reunification", *American Economic Review*, Vol. 94, pp. 730-40.
- Fujiwara, D. and R. Campbell (2011), *Valuation techniques for social cost-benefit analysis: Stated preference, revealed preference and subjective well-being approaches, a discussion of the current issues*, United Kingdom: HM Treasury and Department for Work and Pensions, July.
- Gasparini, L. and P. Gluzmann (2009), "Estimating income poverty and inequality from the Gallup World Poll: The case of Latin America and the Caribbean", *ECINEQ Working Paper*, 2009-151, published December 2009, Palma de Mallorca, ECINEQ Society for the Study of Economic Inequality, available online at: www.ecineq.org/milano/wp/ecineq2009-151.pdf, last accessed 14 August 2012.
- Gilbert, D.T., M.D. Lieberman, C.K. Morewedge and T.D. Wilson (2004), "The peculiar longevity of things not so bad", *Psychological Science*, Vol. 15(4), pp. 14-19.
- Gilbert, D.T. and T.D. Wilson (2000), "Miswanting: Some problems in the forecasting of future affective states", in J.P. Forgas (ed.) (2000), *Feeling and thinking: The role of affect in social cognition*, Studies in emotion and social interaction, second series, pp. 178-197, New York, NY, US: Cambridge University Press.
- Godefroy, P. (2011), "Life-satisfaction: French people give themselves an average score of 7 out of 10", *INSEE, France, portrait social – édition 2011*, available online at: www.insee.fr/en/publications-et-services/dossiers_web/stiglitz/VE4-Anglais.pdf.
- Graham, C. and E. Lora (eds.) (2009), *Paradox and Perception: Measuring Quality of Life in Latin America*, Washington, DC, The Brookings Institution Press.
- Gross, E.F., J. Juvonen and S.L. Gable (2002), "Internet use and well-being in adolescence", *Journal of Social Issues*, Vol. 58(1), pp. 75-90.
- Gruber, J. and S. Mullainathan (2002), "Do cigarette taxes make smokers happier?", *NBER Working Paper*, No. 8872, Cambridge, Mass.: National Bureau of Economic Research.
- Guria, J., J. Leung, J. Jones-Lee and G. Loomes (2005), "The willingness to accept value of statistical life relative to the willingness-to-pay value: Evidence and policy implications", *Environmental and Resource Economics*, Vol. 32, pp. 113-127.
- Gyarmati, D., S. de Radd, B. Palameta, C. Nicholson and T. Shek-Wai Hui (2008), "Encouraging work and supporting communities: Final results of the Community Employment Innovation Project", Ottawa: Social Research and Demonstration Corporation, available online at: www.srdc.org/uploads/CEIP_finalrpt_ENG.pdf, last accessed 10 August 2012.
- Hagerty, M.R. and R. Veenvhoven (2003), "Wealth and happiness revisited: Growing wealth of nations does go with greater happiness", *Social Indicators Research*, Vol. 64, pp. 1-27.
- Halpern, D. (2010), *The Hidden Wealth of Nations*, Cambridge: Polity Press.

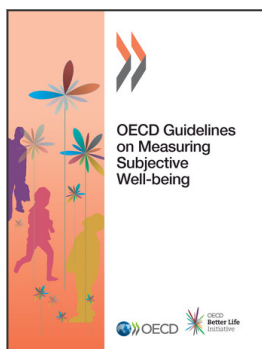
- Helliwell, J.F. (2008), "Life satisfaction and quality of development", *NBER Working Papers*, No. 14507, National Bureau of Economic Research.
- Helliwell, J.F. and C.P. Barrington-Leigh (2010), "Measuring and understanding subjective well-being", *NBER Working Paper*, No. 15887, National Bureau of Economic Research.
- Helliwell, J.F., C.P. Barrington-Leigh, A. Harris and H. Huang (2010), "International evidence on the social context of well-being", in E. Diener, J.F. Helliwell and D. Kahneman (eds.), *International Differences in Well-Being*, New York: Oxford University Press.
- Helliwell, J.F. and H. Huang (2005), "How's the job? Well-being and social capital in the workplace", Cambridge: National Bureau of Economic Research, available online at: <http://faculty.arts.ubc.ca/jhelliwell/papers/NBERw11759.pdf>, last accessed 10 August 2012.
- Helliwell, J.F., R. Layard and J. Sachs (eds.) (2012), *World Happiness Report*, Earth Institute, New York, Columbia University, available online at: www.earthinstitute.columbia.edu/sitefiles/file/Sachs%20Writing%202012/World%20Happiness%20Report.pdf.
- HM Treasury (2011), *The Green Book: Appraisal and Evaluation in Central Government*, United Kingdom.
- Huppert, F.A. and T.T.C. So (2011), "Flourishing across Europe: Application of a new conceptual framework for defining well-being", *Social Indicators Research*, published Online First.
- Huxley, A. (1932; reprinted 2004), *Brave New World*, London: Vintage Classics.
- Inglehart, R., R. Foa, C. Peterson and C. Welzel (2008), "Development, freedom and rising happiness: A global perspective 1981-2007", *Perspectives on Psychological Science*, Vol. 3, pp. 264-285.
- 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.
- ISTAT (2011), *The Importance of Well-Being Dimensions for Citizens*, web article published 4 November, available online at: www.istat.it/en/archive/44236, accessed 09 August 2012.
- Jürges, H. (2007), "True health vs response styles: Exploring cross-country differences in self-reported health", *Health Economics*, Vol. 16, pp. 163-178.
- Kahneman, D. (2010), "The riddle of experience versus memory", lecture delivered February 2010 as part of the TED Talks series, available online at: www.ted.com/talks/daniel_kahneman_the_riddle_of_experience_vs_memory.html.
- Kahneman, D. (1999), "Objective happiness", in D. Kahneman, E. Diener and N. Schwarz (eds.), *Well-being: Foundations of hedonic psychology*, pp. 3-25, New York: Russell Sage Foundation Press.
- Kahneman, D. and A. Deaton (2010), "High income improves life evaluation but not emotional well-being", *Proceedings of the National Academy of Sciences*, Vol. 207(38), pp. 16489-16493.
- Kahneman, D. and A.B. Krueger (2006), "Developments in the measurement of subjective well-being", *The Journal of Economic Perspectives*, Vol. 20(1), pp. 3-24.
- Kahneman, D., A.B. Krueger, D. Schkade, N. Schwartz and A. Stone (2004), "Towards National Well-Being Accounts", *The American Economic Review*, Vol. 94(2), pp. 429-434.
- Kahneman, D. and J. Riis (2005), "Living and thinking about it: Two perspectives on life", in F.A. Huppert, B. Kaverne and N. Baylis (eds.), *The Science of Well-Being*, London: Oxford University Press, pp. 285-304.
- Kahneman, D. and R. Sugden (2005), "Experienced utility as a standard of policy evaluation", *Environmental and Resource Economics*, Vol. 32, pp. 161-181.
- Kahneman, D. and A. Tversky (2000), *Choices, values and frames*, Cambridge: Russell Sage Foundation and Cambridge University Press.
- Kahneman, D., P.P. Wakker and R. Sarin (1997), "Back to Bentham? Explorations of experienced utility", *Quarterly Journal of Economics*, Vol. 112(2), pp. 375-405.
- Kapteyn, A., J.P. Smith and A. van Soest (2009), "Comparing Life Satisfaction", *RAND Labor and Population Working Paper Series*, RAND and IZA.
- Keyes, C.M. (2006), "Subjective well-being in mental health and human development research worldwide: An introduction", *Social Indicators Research*, Vol. 77, pp. 1-10.

- Kiecolt-Glaser, J.K., L. McGuire, T.F. Robles and R. Glaser (2002), "Psychoneuroimmunology: Psychological influences on immune function and health", *Journal of Consulting and Clinical Psychology*, Vol. 70(3), pp. 537-547.
- King, G., C.J.L. Murray, J.A. Salomon and A. Tandon (2004), "Enhancing the validity and cross-cultural comparability of measurement in survey research", *American Political Science Review*, Vol. 98(1), pp. 191-207.
- Kristensen, N. and E. Johansson (2008), "New evidence on cross-country differences in job satisfaction using anchoring vignettes", *Labour Economics*, Vol. 15, pp. 96-117.
- Krueger, A.B. (2009), *Measuring the subjective well-being of nations: National accounts of time use and well-being*, Chicago: University of Chicago Press.
- Krueger, A.B., D. Kahneman, C. Fischler, D. Schkade, N. Schwarz and A.A. Stone (2009), "Time use and subjective well-being in France and the US", *Social Indicators Research*, Vol. 93, pp. 7-18.
- Lacey, H.P., D.M. Smith and P.A. Ubel (2006), "Hope I die before I get old: Mispredicting happiness across the adult lifespan", *Journal of Happiness Studies*, Vol. 7, pp. 167-182.
- Larsen, R.J. and B.L. Fredrickson (1999), "Measurement issues in emotion research", in D. Kahneman, E. Diener and N. Schwarz (eds.), *Well-being. The Foundations of Hedonic Psychology*, Russel Sage Foundation, New York, pp. 40-60.
- Layard, R. (2011), "Wellbeing and public policy", *CentrePiece Winter 2011/12*, CEP's Big Ideas Series, London School of Economics, Centre for Economic Performance.
- Layard, R. (2005), *Happiness: Lessons from a New Science*, London: Penguin Books.
- Layard, R., D. Clark, M. Knapp and G. Mayraz (2007), "Cost-benefit analysis of psychological therapy", *National Institute Economic Review*, Vol. 202, pp. 90-98.
- Layard, R., G. Mayraz and S. Nickell (2008), "The marginal utility of income", *Journal of Public Economics*, Vol. 92(8-9), pp. 1846-1857.
- Loewenstein, G. and P.A. Ubel (2008), "Hedonic adaptation and the role of decision and experience utility in public policy", *Journal of Public Economics*, Vol. 92, pp. 1975-1810.
- Lucas, R.E. (2007a), "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), pp. 717-730.
- Lucas, R.E. (2007b), "Adaptation and the set-point model of subjective well-being", *Current Directions in Psychological Science*, Vol. 16(2), pp. 75-79.
- Lucas, R.E. and M. Donnellan (2011), "Estimating the Reliability of Single-Item Life Satisfaction Measures: Results from Four National Panel Studies", *Social Indicators Research*, forthcoming.
- Lucas, R.E., A.E. Clark, Y. Georgellis and E. Diener (2003), "Re-examining adaptation and the set point model of happiness: Reactions to changes in marital status", *Journal of Personality and Social Psychology*, Vol. 84, pp. 527-539.
- Luhmann, M., R.E. Lucas, M. Eid and E. Diener (2012), "The prospective effect of life satisfaction on life events", *Social Psychological and Personality Science*, published online before print, 20 March.
- Lyubomirsky, S., L. King and E. Diener (2005), "The benefits of frequent positive affect: Does happiness lead to success?", *Psychological Bulletin*, Vol. 131(6), pp. 803-855.
- McCloskey, D. (2012), "Happyism: The creepy new economics of pleasure", *The New Republic*, 8 June, available online at: www.tnr.com/article/politics/magazine/103952/happyism-deirdre-mccloskey-economics-happiness, last accessed 10 August 2012.
- McCrae, R.R. (1990), "Controlling neuroticism in the measurement of stress", *Stress Medicine*, Vol. 6, pp. 237-241.
- Metcalfe, R., N. Powdthavee and P. Dolan (2011), "Destruction and distress: Using a quasi-experiment to show the effects of the September 11 attacks on mental well-being in the United Kingdom", *The Economic Journal*, Vol. 121(550), pp. 81-103.
- Moyle, P. (1995), "The role of negative affectivity in the stress process: Tests of alternative models", *Journal of Organizational Behavior*, Vol. 16(6), pp. 647-668.
- Napier, J.L. and J.T. Jost (2008), "Why are Conservatives happier than Liberals?", *Psychological Science*, Vol. 19(6), pp. 565-572.

- New Economics Foundation (2009), contributing authors: J. Michaelson, S. Abdallah, N. Steuer, S. Thompson and N. Marks, "National Accounts of Well-Being: Bringing real wealth onto the balance sheet", *New Economics Foundation*, January, London, available online at: www.national-accountsofwellbeing.org/learn/download-report.html.
- Ng, Y.K. (1997), "A case for happiness, cardinalism and interpersonal comparability", *The Economic Journal*, Vol. 107, pp. 1848-1858.
- Ng, Y.K. (1996), "Happiness surveys: Some comparability issues and an exploratory survey based on just perceivable increments", *Social Indicators Research*, Vol. 38, pp. 1-27.
- Nussbaum, M.C. (2008), "Who is the happy warrior? Philosophy poses questions to Psychology", *The Journal of Legal Studies*, Vol. 37(2), pp. S81-S113.
- OECD (2012), *Measuring regulatory performance – A practitioners guide to perception surveys*, Paris, OECD Publishing.
- OECD (2011a), *How's Life? Measuring Well-Being*, Paris, OECD Publishing.
- OECD (2011b), internal analysis of data shared by around 4 000 users of *Your Better Life Index*, as of September, available online at: <http://oecdbetterlifeindex.org/>.
- Oishi, S. (2002), "The experiencing and remembering of well-being: A cross-cultural analysis", *Personality and Social Psychology Bulletin*, Vol. 28, pp. 1398-1406.
- Oishi, S., E. Diener and R.E. Lucas (2007), "The optimum level of well-being: Can people be too happy?", *Perspectives on Psychological Science*, Vol. 2(4), pp. 346-360.
- Olejnik, S. and J. Algina (2000), "Measures of effect size for comparative studies: Applications, interpretations and limitations", *Contemporary Educational Psychology*, Vol. 25, pp. 241-286.
- ONS UK (2012), "Analysis of experimental subjective well-being data from the Annual Population Survey April to September 2011", *Statistical release*, 28 February, available online at: www.ons.gov.uk/ons/rel/wellbeing/measuring-subjective-wellbeing-in-the-uk/analysis-of-experimental-subjective-well-being-data-from-the-annual-population-survey-april-september-2011/report-april-to-september-2011.html.
- ONS UK (2011a), "Findings from the National Well-Being Debate", *UK Office for National Statistics Paper*, July 2011.
- ONS UK (2011b), "Initial Investigation into Subjective Well-Being from the Opinions Survey", *Working Paper* released 1 December, ONS, Newport, available online at: www.ons.gov.uk/ons/rel/wellbeing/measuring-subjective-wellbeing-in-the-uk/investigation-of-subjective-well-being-data-from-the-ons-opinions-survey/initial-investigation-into-subjective-well-being-from-the-opinions-survey.html.
- Ostir, G.V., K.S. Markides, M.K. Peek and J.S. Goodwin (2001), "The association between emotional well-being and the incidence of stroke in older adults", *Psychosomatic Medicine*, Vol. 63, pp. 210-215.
- Oswald, A.J. and N. Powdthavee (2008a), "Does happiness adapt? A longitudinal study of disability with implications for economists and judges", *Journal of Public Economics*, Vol. 92, pp. 1061-1077.
- Oswald, A.J. and N. Powdthavee (2008b) "Death, happiness, and the calculation of compensatory damages", *The Journal of Legal Studies*, Vol. 37(S2), pp. S217-S251.
- Pavot, W. and E. Diener (1993), "Review of the Satisfaction with Life Scale", *Psychological Assessment*, Vol. 5(2), pp. 164-172.
- Posner, E.A. and C.R. Sunstein (eds.) (2010), *Law and Happiness*, Chicago: University of Chicago Press.
- Powdthavee, N. (2010), *The Happiness Equation*, London: Icon Books.
- Powdthavee, N. (2009), "I can't smile without you: Spousal correlation in life satisfaction", *Journal of Economic Psychology*, Vol. 30, pp. 675-689.
- Prentice, D.A. and D.T. Miller (1992), "When small effects are impressive", *Psychological Bulletin*, Vol. 112(1), pp. 160-164.
- Pressman, S.D. and S. Cohen (2005), "Does positive affect influence health?", *Psychological Bulletin*, Vol. 131(6), pp. 925-971.
- 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.
- Riis, J., G. Loewenstein, J. Baron, C. Jepsen, A. Fagerlin and P.A. Ubel (2005), "Ignorance to hedonic adaptation to hemodialysis; A study using Ecological Momentary Assessment", *Journal of Experimental Psychology: General*, Vol. 134, pp. 3-9.

- Riphahn, R. and O. Serfling (2005), "Item non-response on income and wealth questions", *Empirical Economics*, Vol. 30, pp. 521-538.
- Russell, J.A. (1980) "A Circumplex Model of Affect", *Journal of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161-1178.
- Russell, J.A., M. Lewicka and T. Niit (1989), "A Cross-Cultural Study of a Circumplex Model of Affect", *Journal of Personality and Social Psychology*, Vol. 57(5), pp. 848-856.
- Ryan, R.M. and E.L. Deci (2001), "On happiness and human potentials: A review of research on hedonic and eudaimonic well-being", *Annual Review of Psychology*, Vol. 52, pp. 141-166.
- Ryff, C.D. (1989), "Happiness is everything, or is it? Explorations on the meaning of psychological well-being", *Journal of Personality and Social Psychology*, Vol. 57(6), pp. 1069-1081.
- Sacks, D.W., B. Stevenson and J. Wolfers (2010), "Subjective well-being, income, economic development and growth", *NBER Working Paper*, No. 16441, Cambridge, Mass.: National Bureau of Economic Research.
- Schaubroeck, J., D.C. Ganster and M.L. Fox (1992), "Dispositional affect and work-related stress", *Journal of Applied Psychology*, Vol. 77, pp. 332-335.
- Schkade, D.A. and D. Kahneman (1998), "Does living in California make people happy? A focusing illusion on judgments of life satisfaction", *Psychological Science*, Vol. 9(5), pp. 340-346.
- Schulz, R. and S. Decker (1985), "Long-term adjustment to physical disability: The role of social support, perceived control, and self-blame", *Journal of Personality and Social Psychology*, Vol. 48(5), pp. 1162-1172.
- Schwartz, P. (2012), "Happiness is not within the government's remit: The philosophical flaw in Happiness Economics", in P. Booth (ed.), ... *and the Pursuit of Happiness: Wellbeing and the role of Government*, London: Institute of Economic Affairs, in association with Profile Books.
- Scollon, C.N., E. Diener, S. Oishi and R. Biswas-Diener (2004), "Emotions across cultures and methods", *Journal of Cross-Cultural Psychology*, Vol. 35(3), pp. 304-326.
- Sen, A. (2002), "Health: perception versus observation", *British Medical Journal*, Vol. 324, pp. 860-861.
- Senik, C. (2011), "The French unhappiness puzzle: The cultural dimension of happiness", *Paris School of Economics Working Paper*, No. 2011-34, Paris-Jourdan Sciences Économiques.
- Shackleton, J.R. (2012), "Wellbeing at work: Any lessons?", in P. Booth (ed.), ... *and the Pursuit of Happiness: Wellbeing and the role of Government*, London: Institute of Economic Affairs, in association with Profile Books.
- Smith, D., K.M. Langa, M. Kabeto and P.A. Ubel (2005), "Health, wealth and happiness: Financial resources buffer subjective well-being after the onset of disability", *Psychological Science*, Vol. 16, pp. 663-666.
- Spector, P.E., D. Zapf, P.Y. Chen and M. Frese (2000), "Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water", *Journal of Organizational Behavior*, Vol. 21(1), pp. 79-95.
- Statistics New Zealand (2008), available online at: www.stats.govt.nz/browse_for_stats/environment/sustainable_development/sustainable-development.aspx.
- Steptoe, A., J. Wardle and M. Marmot (2005), "Positive affect and health-related neuroendocrine, cardiovascular, and inflammatory processes", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102(18), pp. 6508-6512.
- Stevenson, B. and J. Wolfers (2008), "Economic Growth and Subjective Wellbeing: Reassessing the Easterlin Paradox", *NBER Working Paper*, No. 14282, Cambridge, Mass.: National Bureau of Economic Research.
- Stiglitz, J.E., A. Sen, J.P. Fitoussi (2008), *Report by the Commission on the Measurement of Economic Performance and Social Progress*, available online at: www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.
- Stoll, L., J. Michaelson and C. Seaford (2012), "Well-being evidence for policy: A review", London: New Economics Foundation, available online at: www.neweconomics.org/publications/well-being-evidence-for-policy-a-review, last accessed 10 August 2012.
- Stutzer, A. and B.S. Frey (2008) "Stress that doesn't pay: The commuting paradox", *The Scandinavian Journal of Economics*, Vol. 110(2), pp. 339-366.
- Stutzer, A. and R. Lalive (2004), "The role of social work norms in job searching and subjective well-being", *Journal of the European Economic Organisation*, Vol. 2(4), pp. 696-719.

- Sugden, R. (2005), "Anomalies and stated preference techniques: A framework for a discussion of coping strategies", *Environmental and Resource Economics*, Vol. 32, pp. 1-12.
- Tabachnick, B.G. and L.S. Fidell (2001), *Using multivariate statistics*, fourth edition, Needham Heights, MA: Allyn and Bacon.
- Thaler, R.H. and C.S. Sunstein (2008), *Nudge: Improving decisions about health, wealth and happiness*, New Haven, CT: Yale University Press.
- Tsai, J.L., B. Knutson and H.H. Fung (2006), "Cultural variation in affect valuation", *Journal of Personality and Social Psychology*, Vol. 90(2), pp. 288-307.
- Ubel, P.A., A. Jankovic, D. Smith, K. Langa and A. Fagerlin (2005), "What is perfect health to an 85-year-old? Evidence for scale recalibration in subjective health ratings", *Medical Care*, Vol. 43(10), pp. 1054-1057.
- UMR Research (2012), "UMR Omnibus Results: Happiness of New Zealand", UMR research summary, reported by Scoop, available online at: http://img.scoop.co.nz/media/pdfs/1202/HappinessOfNZUpdate_Jan12.pdf.
- van de Vijver, F.J.R. and Y.H. Poortinga (1997), "Towards an integrated analysis of bias in cross-cultural assessment", *European Journal of Psychological Assessment*, Vol. 13(1), pp. 29-37.
- Van Praag, B.M.S. and B.E. Baarsma (2005), "Using Happiness Surveys to Value Intangibles: The Case of Airport Noise", *The Economic Journal*, Vol. 115, pp. 224-246.
- Veenhoven, R. (1994), "Is happiness a trait? Tests of the theory that a better society does not make people any happier", *Social Indicators Research*, Vol. 32(2), pp. 101-160.
- 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.
- Waterman, Alan S. (2007), "On the importance of distinguishing hedonia and eudaimonia when contemplating the hedonic treadmill", *American Psychologist*, Vol. 62(6), pp. 612-613.
- Wilson, T.D. and D.T. Gilbert (2005), "Affective forecasting: Knowing what to want", *Current Directions in Psychological Science*, Vol. 14(3), pp. 131-134.
- White, M.P. and P. Dolan (2009), "Accounting for the richness of daily activities", *Psychological Science*, Vol. 20(8), pp. 1000-1008.
- Wright, T.A. and B.M. Staw (1999), "Affect and favourable work outcomes: Two longitudinal tests of the happy-productive worker thesis", *Journal of Organizational Behavior*, Vol. 20, pp. 1-23.
- Yeaton, W.H. and L. Sechrest (1981), "Meaningful measures of effect", *Journal of Consulting and Clinical Psychology*, Vol. 49(5), pp. 766-767.
- Zweimüller, J. (1992), "Survey non-response and biases in wage regressions", *Economics Letters*, Vol. 39, pp. 105-109.



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), "Output and analysis of subjective well-being measures", in *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264191655-8-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.