EEPRU Policy Research Unit in Economic Evaluation of Health and Social Care Interventions

Research Report

A conceptual comparison of well-being measures used in the UK

Tessa Peasgood, John Brazier, Clara Mukuria, Donna Rowen

Correspondence to: John Brazier, ScHARR, University of Sheffield, j.e.brazier@sheffield.ac.uk

RR0026

September 2014

The University Of Sheffield.

UNIVERSITY of York

The Policy Research Unit in Economic Evaluation of Health and Care interventions is funded by the Department of Health Policy Research Programme. It is collaboration between researchers from the University of Sheffield and the University of York.

The Department of Health's Policy Research Unit in Economic Evaluation of Health and Care Interventions is a 5 year programme of work that started in January 2011. The unit is led by Professor John Brazier (Director, University of Sheffield) and Professor Mark Sculpher (Deputy Director, University of York) with the aim of assisting policy makers in the Department of Health to improve the allocation of resources in health and social care.

Executive Summary

Motivation and background

There is significant political interest in the UK in measuring subjective well-being (SWB) and the possibility of incorporating such measures into policy, including health policy [All-Party Parliamentary Group on Well-being Economics, 2014]. A number of different, yet related, measures of well-being and health are used across government departments. This includes four summary subjective (personal) well-being questions which ask about life satisfaction, happiness yesterday, anxiety yesterday and worthwhileness adopted by the Office of National Statistics (ONS) under its Measuring National Well-being Programme (referred to here as the ONS-4). They have also adopted the use of the short Warwick Edinburgh Mental Well-being Scale (S-WEMWBS) and the General Health Questionnaire (GHQ-12) which is a mental health screening measure that has been used in well-being measurement. In addition to the measures used within the ONS framework, the National Institute for Health and Care Excellence (NICE) currently rely upon the EQ-5D, a measure of health-related quality of life (HRQoL), in the assessment of medical technologies and public health interventions,[NICE 2012; 2013a] while social care guidance includes measures of capability and need, ICECAP-A and Adult Social Care Outcomes Toolkit (ASCOT).[NICE 2013b]

There is limited evidence on how the GHQ-12, ONS-4, S-WEMWBS, EQ-5D, ASCOT and ICECAP-A relate to each other, which causes difficulty in the comparison of results across datasets and evaluations containing different measures and for informing decisions across sectors. Given that these measures are used to inform policy making throughout Government, it is important to better understand how these measures compare. For example, it is unclear whether the GHQ-12 is redundant if the S-WEMWBS is included alongside the ONS-4 within the ONS well-being indicator set.

In order to assess how these measures are related, it is important to establish what construct they seem to measure. Well-being has no agreed upon meaning. There are five main theories that offer a possible explanation of what it is for a life to go well:

1. Objective list theories: Life goes well if an individual has certain goods or characteristics, regardless of whether they want or enjoy them.

2. Preference satisfaction theories: Life goes well if an individual gets what she wants. This has been the dominant paradigm within economics.
3. Hedonist theories: Life goes well if an individual experiences pleasure and not pain; therefore the key focus is feelings.
4. Evaluative theories: Life goes well if an individual evaluates her life positively.
5. Flourishing theories: Life goes well if an individual has ideal qualities for a human being.

There are a number of other approaches that focus on what they take to be policy relevant outcomes, including HRQoL, capabilities and the fulfillment of basic needs. Though the theories are distinct, there is considerable overlap in the constructs that they cover; consequently, questions asked across measures assessing the different theories may be very similar.

The instruments

The 6 main measures used in the UK were designed for different purposes and have one or more theory of well-being as the main focus.

1. GHQ-12 contains 6 positive and 6 negative worded items that cover both feelings and psychological functioning. It was designed as a screening tool for mild mental health problems, but has been used (mostly for convenience) as a measure of SWB. The response options differ between the positive and negative items and in both cases are referenced to 'usual', making it problematic to use in order to identify stationary health or SWB states. There has been intense disagreement on whether the 12 items in the GHQ act as a single domain.
2. WEMWBS contains 14 positively worded items designed to measure positive mental health and tap into both feelings and psychological flourishing. A shortened 7-item version (S-WEMWBS) was developed using rasch analysis and has been found to have a single domain.
3. ONS-4 contains 4 (experimental) questions, each aiming to represent a different aspect of SWB, including feelings (happiness and anxiety), life evaluation (life satisfaction) and flourishing (worthwhile). However, with only single items it does not provide a complete picture of these concepts.

4. ICECAP-A aims to measure capability through prefixing its 5 items with the words ‘able to’ or ‘can’, for example, ‘can have a lot of enjoyment and pleasure’. The dimensions cover aspects of psychological well-being or flourishing along with a more hedonic item (enjoyment). The content was developed from in-depth interviews with adults to define attributes of quality of life and the capabilities that are important to people in their lives.[Al-Janabi et al, 2012]
5. ASCOT is a measure of social-care related quality of life that is designed to assess the extent to which an individual’s social-care needs and wants are being met. It also draws upon Sen’s theory of capabilities, with the level of met needs in each of the 8 domains being disaggregated into having needs met and an ‘ideal’ state in which the level of functioning in the domain is consistent with individual preferences. The instrument can be anchored to the QALY scale.[Netten et al, 2012] Its content overlaps with some aspects of psychological well-being and health, but it is not concerned with the extent that an individual can meet their needs or wishes on their own but with informal and formal care.
6. EQ-5D is a measure of HRQoL rather than well-being or full quality of life. It has two components: five questions describing health and a visual analogue rating scale. The five item descriptive system generates a utility score based on a valuation exercise to extract preferences of a general population sample.

Comparison of measures

When we look at the currently used measurement instruments, most do not fit neatly into any particular theoretical concept of well-being. Measures of health and well-being may draw from different approaches yet have similar items within their measure. Despite this potential over-lap, it is clearly useful to understand the theoretical foundations behind the well-being measurement tools and their conceptual differences. Our six instruments draw from a number of competing conceptual theories and measure different overall constructs: mental health and psychological well-being (GHQ-12, WEMWBS), social-care needs (ASCOT), capabilities (ICECAP-A), evaluative, affect and flourishing (ONS-4), and health-related quality of life (EQ-5D). Despite drawing on specific conceptual theories, there is also considerable overlap in questions asked across the measures. GHQ-12, WEMWBS, ONS-4, ICECAP-A and ASCOT all have questions covering constructs that are important for flourishing; GHQ-12, WEMWBS and EQ-5D

also include questions on feelings. One of the distinguishing features between the measures is the inclusion/exclusion of positive and negative items; GHQ-12 has a mix while WEMWBS and ICECAP-A have only positively worded questions, which has implications for potential ceiling and floor effects of the instrument. Where items draw upon more than one theory of well-being, the question of whether this aims to represent a very broad view of a single construct of well-being or whether they are to be treated as separate domains is not clear. Empirical evidence suggests that GHQ-12 may have more than one domain, though the evidence is inconclusive. There is some (limited) evidence that the full WEMWBS may have more than one domain. ICECAP-A, ASCOT and EQ-5D have questions which are intended to measure different domains but they have relative weights for each domain to explicitly combine them. EQ-5D and ASCOT indices also can be combined with survival to provide information that can be used in economic evaluation. There are a number of other differences between the measures including issues to do with the top and bottom anchors of the measures as well as the recall period which may have an impact on what is measured and hence their comparability.

Summary and conclusion

The conceptual assessment of four of the measures (GHQ-12, WEMWBS, ONS-4 and ICECAP) used in the UK indicate that whilst the instruments derive from different backgrounds and a number of other differences (such as recall periods), there may still be considerable empirical overlap between the instruments and possible redundancy when using more than one of these instruments to measure well-being. The EQ-5D and ASCOT have a narrower focus and overlap rather less. Empirical evidence comparing these measures is limited and the next stage will be to compare the measures using existing datasets to examine the degree of over-lap.

Table of contents

1 Motivation and background

The promotion of well-being is an important objective of economic and social policy; the components of what makes a 'good life' have been the subject of much research and debate in the social sciences. Government initiatives to provide education, social housing, law and order, income (for the unemployed and those who cannot work), health and other social amenities are geared towards the provision of a 'better life' for their citizens. Governments have finite resources and therefore need to ensure that these are being used in the most appropriate way. Neo-classical economists have relied on the preferences of individuals, translated into monetary terms, as the basis of informing decision-makers on what improves well-being. In recent years there has been increasing interest in going beyond outcomes assessed in monetary terms to the measurement of well-being directly, with a number of different measures of health-related quality of life (HRQoL) and well-being used across different UK government departments.

The Office for National Statistics (ONS) was commissioned to undertake work on measuring national well-being in November 2010, with the aim of publishing national statistics to support the understanding and monitoring national well-being. The work was guided by an extensive public debate on 'what matters' to people. The stance adopted has been that subjective well-being should matter ***in addition to*** objective life circumstances, and that it is "possible to collect meaningful and reliable data on subjective well-being".[Stiglitz et al 2009: p216]

Subjective well-being (SWB) is used as an umbrella term for how people think and feel about their lives. It is best thought of as describing subjective perceptions of life rather than offering a coherent theory of well-being. SWB could be any aspect of well-being that is subjectively perceived by the individual themselves.

As a result of the work undertaken by ONS "Measuring National Well-being Programme", a set of ten domains in which to measure well-being have been developed: personal well-being (this domain was initially labelled subjective well-being but was changed to personal well-being for ease of understanding), relationships, health, what we do, where we live, personal finance, the economy, education and skills, governance and the natural environment; each domain has between 3 and 5 indicators, giving 41 headline measures. The measures include both objective data (for example, number of crimes against the person per 1,000 adults) and subjective data (for example, percentage of people who felt safe walking alone after dark). Subjective aspects of well-being are included to supplement rather than replace objective indicators.

The two domains that are of most interest for our report are the personal (subjective) well-being and health domains. The ONS personal (subjective) well-being domain originally contained four indicators: individual questions on life satisfaction, happiness, anxiety and worthwhileness (which in this report is referred to as the ONS-4). Since April 2011 these have been collected from respondents to the Annual Population Survey, amongst others. In 2013 the short Warwick Edinburgh Mental Well-being Scale (S-WEMWBS) (with data taken from the panel survey Understanding Society) was added to this domain at request of the Department of Health.

The ONS health domain contains four indicators: healthy life expectancy at birth (based on life expectancies and self-reported health), self-reports of a long-term illness or disability (with data taken from the Labour Force Survey), satisfaction with health (with data taken from Understanding Society), and the GHQ-12 (with data also taken from Understanding Society).

In addition to the measures used within the ONS framework, the National Institute for Health and Care Excellence (NICE) currently rely upon the EQ-5D measure of HRQoL in the assessment of medical technologies and public health interventions,[NICE 2012; 2013a] while social care guidance includes measures of capability and need, ICECAP-A and Adult Social Care Outcomes Toolkit (ASCOT).[NICE 2013b] The NHS outcome framework also uses EQ-5D as a Patient Reported Outcome Measure (PROMs) across a number of health care conditions to monitor the performance of providers.[Department of Health, 2012]]. The EQ-5D along with GHQ-12 and more recently the WEMWBS has been used in the health survey for England.

There is limited evidence on how the GHQ-12, ONS-4, S-WEMWBS, EQ-5D, ASCOT and ICECAP-A relate to each other, which causes difficulty in the comparison of results across datasets and evaluations containing different measures and for informing decisions across sectors. Given that these different measures are used to inform policy making throughout Government, it is important to better understand how they compare to each other. For example, it is unclear whether the GHQ-12 is redundant if the S-WEMWBS is included alongside the ONS-4 within the ONS well-being indicator set. It is also not clear what would be the implications for health policy of moving from health to more well-being measures of outcome.

The Department of Health has asked the Policy Research Unit in Economic Evaluation of Health and Care Interventions (EEPRU) to undertake a conceptual and empirical comparison of these six commonly used measures of health and well-being: WEMWBS, GHQ-12, ONS-4, ICECAP-A, ASCOT and EQ-5D. This report examines the first part of this task by providing a conceptual comparison of these measures. It presents an overview of the way well-being has been conceptualised in the

literature and the key issues in its measurement. It then describes the six commonly used measures and reviews them against these concepts and issues.

2 What is well-being?

Well-being, broadly conceived, is how well an individual's life is going. Distinct schools of thought have been developed to explain what it is that makes a life go well and the label 'well-being' can be taken to mean very different things. The traditional tripartite classification[Parfit, 1984] distinguishes between:

1. objective list theories (well-being increases when an individual achieves certain specified goods or characteristics)
2. preference satisfaction theories (well-being increases when an individual's desires are met)
3. hedonism (well-being increases when an individual experiences more pleasure and/or less pain).

This classification has been extended by Haybron[2008] to include:

4. flourishing theories (well-being increases when an individual more closely fulfils their nature as a human being, or 'flourishes')
5. life evaluation[1] or life satisfaction (well-being increases when an individual assesses her life more positively).

We shall set out in a little more detail the five main concepts of well-being listed above before looking at what concepts the well-being and health measures currently used in the UK assess.

2.1 **Objective list theories**

Objective list theories identify a set of goods or characteristics, such as being able to see, or read, or having a good friend, which are taken as non-instrumentally good (i.e. good for their own sake and not because they enable other states to be attained). The attributes and characteristics which are important for well-being may include factors such as basic needs (such as housing and food) but can extend beyond 'objective' indicators to more subjective attributes such as feelings of safety or happiness. But how do we decide what should go on the list? Or which items on the list are most

[1] Haybron refers here to the account developed by Sumner[1996] of 'Authentic Happiness', which conceives of well-being as an individual's cognitive and emotional response to their life from an informed and autonomous perspective.

important? The list of items will vary depending on the theorists' ideas of what is important (not what the individual thinks is important to their life), although some factors may be easily agreed upon, for example health.

2.2 **Preference satisfaction theories**

According to preference satisfaction, or desire theories, an individual's well-being increases when they get what they want; though the majority of desire theories set restrictions upon which desires should count towards well-being (for example only those based on full information should count). Neo-classical welfare economics is derived from this approach to well-being. Individuals' choices in the market place for goods and services reveal their preferences by their willingness to pay. As income increases, individuals can choose more of what they want which makes income a useful proxy for utility or well-being.

One problem with using observed decisions is that markets are not well-defined for many of the benefits from Government programmes and even where markets do exist it cannot be assumed that individual decision makers are acting rationally or with full information. Economists have developed another application of the preference satisfaction account that involves asking individuals to state their strength of preference for different states of the world. One of these stated preference methods is to ask individuals how much they are willing to pay (WTP) for any given state of the world, such as one with a better environment or access to health care. This approach has been widely used in UK Government departments, including transport and environment, though less so in health economics. In health economics there has been a concern with the distributional implications of WTP and the evidence of insensitivity to the scale and scope of the health care benefit being valued.[Kahneman and Knetsch, 1992] Policy makers have also been unwilling to monetise intangible health benefits. For this reason health economists have developed the Quality Adjusted Life Years (QALYs) rather than money as the preferred unit of benefit for economic evaluation.

QALYs combine length of life with quality of life by assigning a valuing to each health state on a scale anchored at 0 (for a state equivalent to death) and 1 (for full health). A common way to estimate preference weights is to ask the individual to give the length of time in full health that they would consider equivalent to a given length of time in the health state in question (time trade-off (TTO) method).[Brazier et al, 2007] More recently, economists have been using ordinal methods to value health states, including pairwise comparison discrete choice experiments.[Bansback et al, 2012]

Whilst on an abstract level, methods like TTO can be argued to capture the impact that health states

have upon overall well-being there are a number of reasons why this may not be the case. The first is priming. These methods focus the respondent's attention on health and they are typically not encouraged to think more broadly about their life (such as relationships, job satisfaction, income and so forth). The valuation techniques also assume that individuals are able to predict the likely impact of the health state being described on their lives in the future; however, this has been shown not to be the case in health and other contexts.[Dolan and Kahneman, 2008] For example, general population respondents in valuation surveys imagining health states tend not to take into account the extent of any adaptation they will make over time. This is one of the reasons why some economists advocate the use of a more direct measurement of well-being in those experiencing the health states through measures of subjective well-being.[Dolan and Kahneman, 2008]

2.3 **Hedonic theories**

Hedonism is based on the theory that pleasure and pain are the only things that are good or bad for anyone. Hedonic theories conceive of well-being as the total of all hedonic moments, or the product of the intensity of pleasure (absence of pain) and its duration.[Crisp, 2006; Kahneman et al, 1997] They require that pleasure and pain share a certain common property or quality which enables them to be compared. This may be a shared mental state or sensation common in all pleasure or pain, or the extent of the desirability of the feeling. From this perspective, how well someone's life is going depends upon how they are feeling and the emotions they experience in any given moment. The notion that well-being should be viewed as simply the sum of hedonic happiness has been seen by many psychologists and philosophers as too simplistic. Arguably, positive and negative affect are not the only things that people care about (or have desires about).

There has been a lot of research effort on measuring hedonic well-being, particularly by psychologists, but this is not usually based on 'hedonism' as a theory (in the sense of aiming for a complete measure of well-being), rather it has been in the context of measuring mental health, particularly poor mental health. Typically, hedonic well-being has been assessed by asking people to rate their current level of a range of positive feelings (or affect) (such as enjoyment, joy, happiness, excitement) and negative feelings (such as worry, anger, stress, sadness). Psychologists have written extensively on whether these various feelings can be treated as a single dimension, or whether they form two or even more dimensions.[Watson, 1988] The 'Circumplex model'[Larsen and Fredrickson, 1999] distinguishes between both valence (positive/negative) and the degree of arousal, placing affect into four categories: positive high arousal (e.g. excitement), positive low arousal (e.g. contentment), negative high arousal (e.g. stress), and negative low arousal (e.g. sorrow).

The number of dimensions within affect is important as it influences what can be done with any information collected on affect. If affect forms a single dimension then it is reasonable to combine the data to construct a single (affect balance) index. Furthermore, asking about one or two feelings may be a reasonable representation of a more complete picture of the individual's feelings. However, this issue remains unresolved. There is a reasonable consensus of opinion that positive affect forms a single dimension (incorporating high and low arousal affects), but mixed opinion on whether negative affect is best presented as two dimensions, and/or whether it can be combined with positive affect. Findings depend upon the dataset, the instrument, the time-scale, and the treatment of measurement error.[Green et al, 1993; Russell and Carroll, 1999]

Considerable attention has also been devoted to assessing the concern that individuals may not be able to accurately recall their past feelings.[Kahneman et al, 1997] To overcome problems with memory recall, methods have been developed which report affect as it happens: the Experience Sampling Method (ESM)[Csikszentmihalyi and Hunter, 2003] or soon after, the Day Reconstruction Method,[Kahneman et al, 2004] and measures related to time-use. These are quite sophisticated methods of data collection that are not employed by the instruments being reviewed in our report that ask people about today or to recall a period from one day to a few weeks.

2.4 **Evaluative theories**

Evaluative theories focus on how individuals feel about their life and their cognitive appraisal of their life. In this approach, an individual has high well-being if they have a positive attitude towards their life which includes both thinking life is good and feeling "satisfied or fulfilled by it".[Sumner, 1996: 146] An individual's assessment may include positive and negative feelings, whether they are happy with their lives more generally, or whether they feel they are meeting their idea of human potential. It is left up to the individual to choose what to take into consideration when assessing their life.

This approach to well-being underlies the use of life satisfaction questionnaires as measures of well-being, such as the Satisfaction with Life Scale (SWLS)[Diener et al, 1985] and single evaluative style questions which focus on life satisfaction or overall happiness. The term 'happiness' in this context is distinct from happiness as an emotion; hence, surveys asking about whether the respondent is happy with their life overall (or an aspect of it) assume respondents are able to distinguish this from *feeling happy* at a particular moment in time.

A number of concerns have been raised about interpreting life satisfaction or evaluative questions as overall well-being, particularly whether or not individuals can evaluate their lives using these types

of simple questions. There is also concern that questions may unduly reflect current mood,[Schwarz and Clore, 2003] context factors or the individual's expectations and adaptation. Despite these concerns, there is reasonable evidence on the validity of life satisfaction responses.[Diener et al, 2013][2]

2.5 **Flourishing theories**

Flourishing or perfectionist theories (often referred to as eudaimonic theories) identify levels of well-being by reference to human nature: what is the best way of 'being' for a human. Most perfectionist accounts of well-being draw from Aristotle's understanding of eudaimonia, which he took to be the highest good (the only good that individuals desire for its own sake) and to arise from activities based on virtue and practical reason. However, there is no agreement on what it means to perfect human nature.

From a theoretical perspective, aspects of 'flourishing' are strictly non-instrumental. If knowledge, achievement and friendship are good it is because they realise aspects of human nature, independently of any pleasure they may bring.[Hurka, 1993] There is no theoretical mechanism that would enable one aspect of flourishing to be traded against another.

Psychologists have drawn from this body of theory but moved away from ideas of species perfection towards notions of individuals reaching their own individual true potential in life,[Ryff, 1995] and possessing the psychological resources necessary to flourish. They have been interested in psychological functioning, with less concern as to whether the aspects of psychological functioning are taken as instrumental (will cause some other valuable state such as happiness) or non-instrumental (of value in themselves). This generates some potential for confusion; different theories share similar content on their understanding of what they call 'psychological health' or 'psychological well-being',[McDowell, 2006] yet they have a different understanding of how this relates to well-being overall. Here, we focus on those theories that can be considered to draw directly on flourishing accounts but will discuss what may be considered needs-based theories in the summary section.

2.5.1 Psychological well-being as a complete measure of well-being

Psychological well-being can be equated to well-being. One of the best known theories which does

[2] We use the term validity here to mean that the self-report life satisfaction questions are a reasonably accurate way of getting at an individual's judgement of their life, and they converge with other types of measures such as informant reports and biometric measures. It is hard to say that something is a valid measure of well-being as we do not have a gold standard, or agreement, on what we conceive well-being to be.

this is Ryff's theory and measure of Psychological Well-being (PWB)[Ryff, 1995] which presents psychological well-being as the true potential of an individual. Ryff's theory draws from convergences between theories of positive mental health (such as Jahoda's theories of ideal mental health), those from developmental psychologists and those from clinical psychologists (such as Maslow's ideas of self-actualisation, and Jung's ideas of self-realisation). Ryff takes a purely theory driven approach and sees individual well-being as equating with psychological well-being, which itself is comprised of six distinct components: autonomy, personal growth, self-acceptance, purpose in life, environmental mastery and positive relations.[Ryff, 1995]

2.5.2 Human flourishing scales as a complete measure of well-being

Some psychologists have added positive emotions and evaluations to measures of psychological well-being to generate a broader measure of well-being. Seligman [2011] has a model of flourishing that contains: Positive emotions, Engagement, Relationships, Meaning, and Achievement (PERMA). Seligman claims that these are aspects of life that individuals would choose for their own sake. These different aspects can be measured through a profiler or 'dashboard' (PERMA-P) which generates a score for each of the separate dimensions.

Keyes [2002] takes the basis of flourishing (or the top end of the mental health continuum) as being the mirror opposite of mental health problems. He represents flourishing as: psychological well-being (based on the 6 domains from Ryff's PWB), emotional well-being (positive affect and life satisfaction), and social well-being (social acceptance, social actualisation, social contribution, social coherence and social integration). The inclusion of life satisfaction here does not imply that this is meant to represent an overall assessment of the individual's *well-being*, rather that the individual's ability to assess their life positively is one of a number of attributes which would be expected to be present in a flourishing life.

Huppert and So [2013] also adopt Keyes' idea of identifying flourishing as the mirror opposite of identifying mental health problems, and use data from the European Social Survey to frame flourishing as both positive feeling and positive functioning: competence, emotional stability, engagement, meaning, optimism, positive emotion, positive relationship, resilience, self-esteem, and vitality. Huppert and So [2013] note that their model does not include "psychodynamic constructs such as 'personal growth' or 'self-acceptance' …or the construct of autonomy or self-determination", which is a central concept for other models such as Ryff [1995] and Deci and Ryan.[2002]

2.6 **Other approaches**

2.6.1 The capability account

The capability account [Sen, 1999] shares aspects of a simple list theory but moves away from the idea that well-being can be identified by the attainment of items on the list or key functionings (things people can be and do), towards the notion that the relevant criteria for public policy is whether the individual had the opportunity to achieve those functionings. An individual's capability represents the alternative combinations of functionings they can achieve, from which they can choose one combination.[Sen, 1993] This re-introduces a role for individual preferences; if the individual 'could' have had a close friendship but chose not to, their well-being is not considered lower because of the absence of a close friend.

Sen does not prescribe a particular set of capabilities or way to operationalise the concept, arguing that societies may choose their own list of important functionings with priority given to the functionings that we have 'reason to value', and which arise from democratic public deliberation.[Sen, 2004] There have been attempts by others, the most notable being Nussbaum's list of 10 central human functionings.[Nussbaum, 2000] Alkire [2002] reviewed 15 different lists of human development and quality of life, finding considerable overlap in the domains identified by different researchers and different approaches. Although favouring dialogue and deliberate process in the implementation of the capabilities approach, Alkire [2008] identifies a core list of capabilities for the opening of that process: health and security, understanding, achievement, participation, relationships, satisfaction and self-integration, and harmony. Robeyns and van der Veen [2007] identify 13 key capabilities which include shelter, the living-environment, labour and non-discrimination. Whilst these share some similarities with the typical domains within psychological health or psychological well-being, the 'capabilities' lists usually include more objective items, such as physical health and safety, democratic participation, understanding and knowledge and the living-environment.

As with other 'flourishing' accounts, there is no unit to use in order to compare achievement in one capability against another, and both Sen and Nussbaum have stressed the incommensurability of important capabilities. It is also generally accepted that a full capabilities set cannot be observed.[Alkire 2002: 181] Questionnaires will necessarily be limited to self-reported perceptions of 'ability' rather than objective measures. Furthermore, both the presence of the ability to attain a functioning, and the individual's preference towards that functioning may change over time. However, with a heavy dose of pragmatism, instruments have been developed within health economics that attempt to extract 'capabilities' through survey questionnaires and use techniques

to generate overall scores from those reported capabilities (e.g. ICECAP).

2.6.2 Health and Health-Related Quality of Life

The long-standing definition of health from the World Health Organisation (WHO) [WHO, 1958: 459] defined health as "a state of complete physical, mental and social well-being, and not merely the absence of disease and infirmity". However, a far narrower focus on health as an 'absence of illness, disease and injury' dominated medical decision making until the 1970s. Around this time, when interest in social progress and quality of life indicators was burgeoning, interest also developed in using 'quality of life' as a means of evaluating treatments in which lifesaving was not the only objective.

In health services research and health economics 'health-related quality of life' (HRQoL) was taken to be the most appropriate concept to use rather than quality of life *per se*. HRQoL is only those aspects of quality of life that can be influenced by health or health care treatment. A commonly used definition describes HRQoL as a 'multidimensional concept that encompasses the physical, emotional, and social components associated with an illness or treatment'.[Revicki, 1989] However, not all health economists were in agreement over the relevant domains for HRQoL. Torrance, for example, argued that the WHO criteria of fulfilling a social role and achieving social well-being were "outside the scope of the health domain and, therefore, are outside the scope of health-related quality of life".[Torrance, 1987: 594] He also excluded economic, political, cultural, environmental, aesthetic, and spiritual aspects of an individual's existence from the health domain; his definition of HRQoL being physiological and emotional functioning only.[Torrance, 1987: 593]

Whilst HRQoL may be in its purest sense include all aspects of well-being which could theoretically be impacted by health or health care, for many the term is used to identify the sub-set of the important or most common ways in which health or health care impact upon well-being. The rationale for taking a narrower HRQoL perspective rather than full quality of life or well-being was to reflect the interests of healthcare policy makers and doctors. Narrower measures of HRQoL are likely to be more responsive to health care interventions.

There are literally thousands of instruments which are designed to measure health and HRQoL, whether that is the presence of symptoms, or achievement/impairment of key functionings either related to specific health conditions or to general (generic) health and HRQoL. The purpose of these instruments varies from being used as diagnostic tools, to support patient-level treatment decisions, for measuring patient reported outcomes (PROMs), or for group level assessment of change in research studies.

Instruments have aimed to locate health states throughout a spectrum of physical and mental health states, from very poor to very good health. In physical health, positive or very good health is generally seen as the absence of problems in physical functioning, with a subsequent lack of sensitivity at the top end of the spectrum. High levels of energy, fitness, resistance to disease, high life expectancy may be aspects found in the top of the spectrum of physical health, but these are not generally considered.

Instruments for measuring outcomes in mental health have focused on either the poor mental health (distress) end of the spectrum, on only the positive mental health end (experiencing positive affect and emotions), or both, with considerable disagreement as to whether positive mental health and negative mental health lie on the same spectrum or are independent constructs. GHQ is an example of the latter, and WEMWBS focused exclusively on positive items. Interestingly, many of the domains in the theory driven measures of psychological health (such as Ryff's PWB) overlap with domains (CHIME: connectedness, hope, identity, meaning and empowerment) identified from reviewing the literature on mental health recovery [Leamy et al, 2011] and from interviews with mental health service users.[Connell et al, 2012]

2.6.3 Needs based theories

Needs accounts assert that the only policy relevant aspect of well-being is whether essential human needs for survival and flourishing are met. Therefore, needs theories do not provide an outcome measure that represents how well an individual's life is going overall, rather they aim to give an indication of whether attributes which are instrumental to well-being are attained. Need-based theories such as Maslow,[1943] Rawls,[1971] Doyal and Gough,[1991] Max-Neef et al,[1992] and Deci and Ryan[2002] combine both physical resources and psychological resources as necessary for well-being. Psychological health can be taken as an input into current or future well-being. For example, Ryan and Deci's self-determination theory includes the need for autonomy, relatedness and competence. The individual cannot attain well-being (however defined) unless these essential psychological needs are met. This view was influential in the development of ASCOT.[Netten et al, 2012]

2.7 **Summary of theoretical concepts of well-being**

Figure 1 provides a summary of the main theoretical concepts of well-being. These different theoretical approaches provide very different ways of thinking about what it means for a life to go well. Some theories focus on explaining *why* something can be thought of as good for an individual,

others set out a list of *what* constitutes well-being but may not give a reason as to why. The latter lends itself more to measurement. The traditional split between objective and subjective well-being theories is not very useful here as measurement will ultimately rely upon self-report and even within 'objective' lists we may wish to include attributes which can really only be assessed subjectively, such as friendship. A more useful distinction is between those theories which rely upon judgements and preferences of the individual and those in which what is important in life is not determined by the individual. Within perfectionist (or flourishing) theories the individual has not decided that these particular attributes (say autonomy or self-development) are things that matter to them. Similarly with hedonism, the individual may not value happiness. Where what people care about differs across society we may expect to see differences arising between measures of well-being drawn from these different approaches. The importance of certain aspects of life (say purpose and meaning) may relate to cultural and socio-demographic characteristics of the individual, which would lead to differences between well-being measures being related to individual characteristics.

Some of the approaches describe attributes which are necessary, but not sufficient, for well-being to arise. Psychological functioning and mental health have been viewed this way - as an essential need. We would therefore expect to see greater correlation between aspects of 'need' and low levels of well-being, than high levels of well-being.

The term SWB has morphed from referring to a combination of life evaluation and affect (both positive and negative) to more recently including aspects of functioning and flourishing. It therefore draws upon three competing approaches to understanding how well a life is going.

Research on datasets containing multiple SWB questions has tended to identify separate factors within the data: evaluative/life satisfaction, positive affect and negative affect.[e.g. Arthaud-Day et al, 2005] A growing body of evidence finds that evaluative/life satisfaction measures are more closely associated with life circumstances, particularly income, than measures of affect.[Kapteyn et al, 2013; Kahneman and Deaton, 2010; Helliwell et al, 2013; Helliwell and Wang, 2012] Unemployment has also been found to have a stronger impact on evaluative measures of SWB than on measures of affect.[Knabe et al, 2010; Schimmack et al, 2008] Overall, this evidence suggests that the well-being theories are separate constructs, but there may be overlap particularly where there are distinctions between instrumental and non-instrumental constructs that are important in separate well-being theories.

The 6 measures will now be considered to assess their conceptual background. If the measures are designed to measure different well-being theories, then this provides a conceptual basis for

expecting these measures to be associated in different ways to policy relevant aspects such as poor health and unemployment.

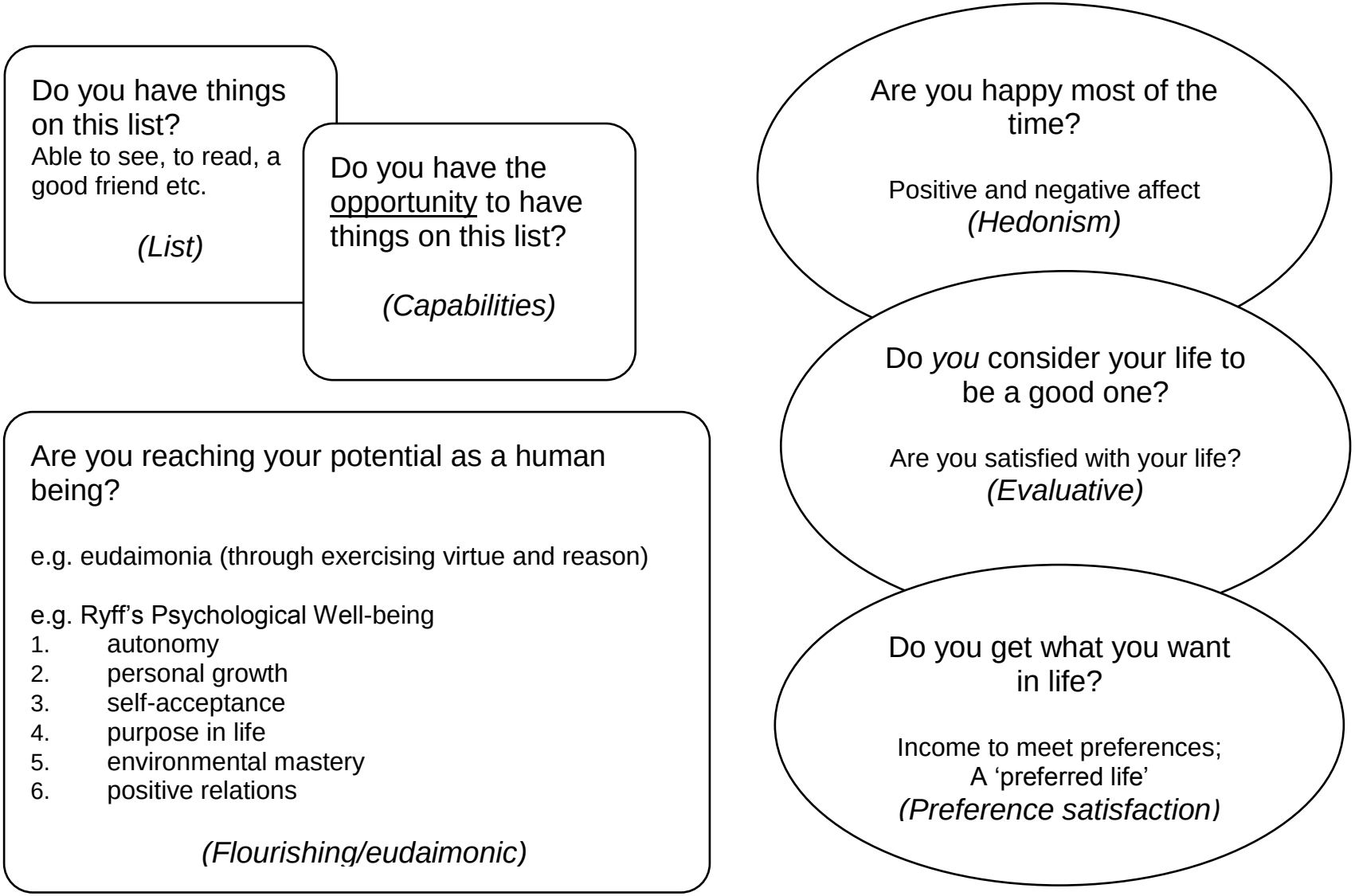

Figure 1: Concepts of well-being

3 The Measures

This section describes the 6 most widely used measures of health and well-being in the UK. These include the GHQ-12, WEMWBS (and its shortened version the S-WEMWBS) and the ONS-4 which measure mental health and well-being, the ICECAP capability measure, the ASCOT which measures social care related quality of life, and the most widely used HRQoL measure, the EQ-5D. See Appendix 1 for the measures.

3.1 **GHQ-12**

The GHQ was developed as a first-stage screening tool to measure mild somatic and psychological symptoms in a non-clinical environment and identify those in need of psychiatric care.[Goldberg and Williams, 1988] The questionnaire focuses on two major areas: the "inability to carry out one's normal 'healthy' functions, and the appearance of new phenomena of a distressing nature",[Goldberg and Hillier, 1979: p139] the aim being to identify individuals who are disturbed or altered from their usual self.

The original scale comprised of 60 items but 30, 28, 20 and 12-item versions have been developed. The GHQ-12 includes six positive and six negative questions and a choice of four options for each in which the presence or intensity of the state over the last few weeks is related to its usual frequency or intensity (see Appendix 1). Negative items have response options of "not at all, no more than usual, rather more than usual, much more than usual", and positive items have response options of "more so than usual, same as usual, less so than usual, much less than usual". Scoring can adopt a number of different forms. The most common scoring methods are Likert scoring (0,1,2,3) or binary (0,0,1,1), the latter representing number of the 12 symptoms that are present. This is often referred to as the Caseness score with a chosen cut off (usually 4) indicating probable psychological disturbance or mental ill health. An alternative scoring, known as the Corrected GHQ (CGHQ), treats the response "no more than usual" for negative items as still indicating the presence of a problem, hence uses a (0,1,1,1) binary score for negative items, and the response "no more than usual" for positive items as indicating no problem, hence uses the conventional (0,0,1,1) for positive items.[Goodchild and Duncan Jones, 1985] This corrected scoring aims to enable the instrument to be used to identify those with a chronic mental health disorder, rather than just those experiencing short-term psychiatric disorders.

The GHQ-12 is a well validated and widely used measure to detect poor mental health, and was not designed as a measure of positive mental health, mental well-being or subjective well-being. That

said, the GHQ has been used in the SWB literature as a measure of overall psychological well-being (mostly due to its availability within large secondary datasets). The instrument has been interpreted as a proxy for individuals' overall assessment of their life. Both the positive and negative questions include aspects of feeling or affect, functionings (such as, 'lost much sleep over worry') and evaluation (such as, thinking of oneself as worthless). The question is whether or not these are measuring different aspects of well-being.

The factor structure of the GHQ-12 has been hotly debated amongst psychologists. Some studies have found that responses from the GHQ-12 separate into two (positive/negative) or three (anxiety/social dysfunction/loss of confidence) factors.[Graetz, 1991] However, other studies have found that responses fit a single factor model reasonably well once measurement error is accounted for.[Hankins, 2008; Smith et al, 2013]

3.2 **WEMWBS and S-WEMWBS**

The WEMWBS was developed from the Affectometer,[Kammann and Flett, 1983] (an existing measure of positive and negative feelings or general happiness) by the Universities of Warwick and Edinburgh. The scale aimed to be able to identify levels of positive mental health in the general population and drew from a number of different conceptions of well-being, including hedonic (feelings) and flourishing accounts (psychological functioning and self-realisation).[Tennant et al, 2007]

The full version asks for time spent in 14 positive states over the last two weeks with five response categories ranging from 'all of the time' to 'none of the time' (see Appendix 1). The minimum score is 14 and the maximum is 70. All questions are asked in a positive manner. The questions ask about 'feelings' and 'thoughts', hence the distinction between positive 'hedonic' type items and those of psychological functioning is not always clear; for example, 'feelings' of optimism could be interpreted as an aspect of psychological health or a positive 'pleasurable' mental state.

The full WEMWBS was found to have good internal consistency and reliability, and no ceiling or floor effects in a population sample.[Tennant et al, 2007] Tests of the dimensionality of the full WEMWBS suggest it may not be uni-dimensional, though strictly only one factor was significant.[Tennant et al, 2006] The full WEMWBS has a reasonable correlation with measures of positive affect and psychological flourishing, but a lower correlation with negative affect. Despite the fact that there are no items within the scale that ask about life evaluations, the WEMWBS correlated quite highly with life satisfaction.[Tennant et al, 2007]

A shortened 7-item version has been derived using Rasch.[Stewart-Brown et al, 2009] This shortened version is uni-dimensional since the poorly fitting items were dropped. The 7-item measure asks about optimism, usefulness, feeling relaxed, thinking clearly, dealing with problems, feeling close to others, and being able to make up one's own mind. The exclusion of 'cheerfulness', which was the closest feeling to 'pleasurableness', means it does not directly measure the hedonic conception of well-being. Alternative psychometric work applying Mokken scale analysis on the full WEMWBS preferred a 10 item scale[Stochl et al, 2012] which re-includes "I've had energy to spare", "I've been interested in things" and "I've been feeling loved".

The full version has been used in a number of national surveys and the shortened version has been adopted by the Pricing and Currency initiative as an outcome measure in secondary mental health care during a pilot phase.[Department of Health, 2013] Results are awaited from the recent application in mental health to see whether the absence of negative items has implications for its psychometric performance. Evidence on the psychometric properties of these measures across other populations is now only starting to emerge.

3.3 **ONS-4**

From April 2011 the ONS included four Subjective Well-being questions on the Annual Population Survey (APS), Living Cost and Food Survey (LCF), Wealth and Assets Survey and the Opinions & Lifestyle Survey (OPN). The term 'subjective' has since been altered to 'personal' since focus groups with responders found the term 'personal' easier to understand.

The following four questions touch upon three different theoretical approaches to individual well-being: evaluative or cognitive assessment of life overall ("Overall, how satisfied are you with your life nowadays?"), eudemonic or flourishing ("Overall, to what extent do you feel the things you do in your life are worthwhile?") and affect or day-to-day experience ("Overall, how happy did you feel yesterday? Overall, how anxious did you feel yesterday?"). Respondents indicate their answer on scales from 0 not at all, to 10 completely. The questions were not intended to provide an aggregate score but rather a profile of personal (subjective) well-being.

The four questions were drawn from recommendations from Dolan, Layard, and Metcalf[2011] and advice from Felicia Huppert. The restriction to four questions enables the questions to be included in a broad range of surveys; however, it necessitated single questions to be used as representative of a larger number of items. 'Anxiety' is chosen as representative of possible negative feelings, and 'worthwhileness' is chosen as representative of psychological flourishing. The use of single items

raises potential issues of measurement error, but may be additionally problematic if items within negative affect or psychological 'flourishing' are not uni-dimensional, i.e. does an item such as feelings of relatedness (which appears on most concepts of psychological flourishing) track the same pattern as feelings of worthwhileness? The use of the reference period 'yesterday' for the questions on feeling anxious and happy draw from the Day Reconstruction Method designed by Kahneman.[Kahneman et al, 2004]

These four questions were initially set out to be experimental in nature and open to future revision in the light of user engagement, further testing and international demands for harmonised questions. Qualitative work on understanding peoples' responses to the ONS-4 questions raised issues, with some respondents interpreting 'satisfied' as a negative or neutral state.[ONS, 2012]

The RAND American Life Panel included the ONS-4 questions [Kapteyn et al, 2014] on a panel with an average of 27 days apart, allowing test re-rest reliability of these questions to be estimated. They found ratios of 0.74 for the life satisfaction question, 0.65 for the worthwhile question, 0.57 for the happy yesterday question, and 0.45 for the anxious yesterday question. Whilst we might expect the affect measures to show greater change over time, the reliability for the anxious question is low. Interestingly, the recommendations from OECD on core questions for measuring SWB replicate the first three ONS questions but use the negative affects feeling 'worried' and 'depressed' rather than anxiety.[OECD, 2013]

There is (as yet) no means by which different aspects of SWB can be combined. One option available is to rely upon life satisfaction as a summary of SWB. If aspects of psychological functioning or positive and negative affect and meaning are important to individuals, would they not be incorporated into life satisfaction? A reluctance to rely fully upon cognitive appraisals of life overall may arise because they are not considered to be a complete measure of well-being, or because they are subject to excessive measurement error, or because individuals cannot be considered sufficiently informed. Life evaluations are taken here, not as the individual offering a global assessment of their life overall, rather as one characteristic (among many) which indicates that an individual has a high subjective well-being. Analysis on the ONS-4 questions in the American Life Panel [Kapteyn et al, 2014] found that the happiness question, the worthwhile question and the life satisfaction question loaded onto a life evaluation factor, whist the anxiety question loaded to a negative affect factor but with an unintuitive sign. This suggests a need for greater testing of this item.

3.4 **ICECAP-A**

The ICECAP-A (or Investigating Choice Experiments Capability Measure for Adults) is a capability

well-being measure for adults which draws upon Sen's capability theory. An earlier version (ICECAP-O) was developed for use in older populations.[Grewal et al, 2006] The ICECAP-A has 5 items that can take one of four levels: stability (being able to feel settled and secure), attachment (being able to have love, friendship and support), autonomy (being able to be independent), achievement (being able to achieve and progress) and enjoyment (being able to have enjoyment and pleasure) (see Appendix 1).

The authors attempt to measure capability by prefixing items with the words 'able to' or 'can', for example 'can have a lot/quite a bit/a little/ or cannot have any'. The dimensions cover aspects of flourishing or psychological well-being along with a more hedonic item (enjoyment). The content was developed from in-depth interviews with a sample of the general adult population to define conceptual attributes of quality of life and the capabilities that are important to people in their lives.[Al-Janabi et al, 2012] This approach is arguably closer to Sen's recommendations of public reasoning and democratic decision making [Sen, 2004] than capability lists based on theoretical reasoning, [Nussbaum, 2000] or a list of important functionings drawn from psychological theories.[Ryff, 1995]

The instrument has been scored from a survey of the general population using best-worst scaling (BWS), a technique rooted in a preference-based paradigm called random utility theory [Flynn et al, 2013]. The values have not been anchored on the QALY scale, where zero is equivalent to being dead, but on a 0-1 scale where zero is the worst (i.e. no capability on any dimension) and one the best (full capability on all dimensions) state defined by the measure. Anchoring zero at no-capabilities is a normative judgment, and the authors of the scale accept that "although death implies no capability, the reverse is not necessarily true" [Flynn et al, 2013].

There is published evidence of its construct validity in the general population [Al-Janabi et al, 2012] and it has been adopted by NICE as one of two measures for social care. It is being used in clinical trials and more may be learnt about its properties in due course. It is comparatively untested and the wording used to design the items assesses capabilities such as "I can have a lot of love, friendship and support", "I can have a lot of enjoyment and pleasure" makes it seem clumsy. A 'think aloud' study on the ICECAP-A identified that some individuals have problems with interpreting the wording of the instrument, and may use a different timescale to that specified.[Al Janabi et al, 2010]

Another concern is the exclusive use of positive items and so, like WEMWBS measures, it may not be appropriate for more severely ill populations.

3.5 **ASCOT**

The Adult Social Care Outcomes Toolkit (ASCOT) is a measure of social-care related quality of life that is designed to assess the extent to which an individual's social-care needs and wants are being met. It also draws upon Sen's theory of capabilities, with the level of met need in a domain being disaggregated into having needs met and an 'ideal' state in which the level of functioning in the domain is consistent with individual preferences. This suggests that whilst capabilities matter, the individual with low levels of functioning can be judged by society as having an unacceptable level of need regardless of whether they recognize this to be the case.[Netten et al, 2011]

The descriptive system was developed using qualitative work with social care service users, carers and staff. It has eight dimensions: 5 reflecting basic social-care related needs (accommodation cleanliness/comfort, safety, food and drink, personal care, being treated with dignity), and three reflecting higher order concerns (control over daily life, social participation, and involvement/occupation). For each item the level of met need is assessed across 4 levels: ideal, no unmet needs, some unmet needs and high unmet needs (for example 'My home is: as clean and comfortable as I want, is adequately clean and comfortable, not quite clean or comfortable enough, or not at all clean or comfortable) (see Appendix 1).

There are two methods of scoring the instrument. One is to use scores developed from a general population survey using BWS; another has anchored these BWS scores onto the QALY scale, where zero is for states equivalent to being dead, using TTO values for a sample of states.[Netten et al, 2012]

It can be used in the community or in nursing/residential homes. It has been used in the local authority Adult Social Care Survey and has been adopted by NICE in its methods for economic evaluation of social care. Evidence for its construct validity is starting to emerge.[Malley et al, 2012]

3.6 **EQ-5D**

The EQ-5D [EuroQol Group, 1990] is a measure of health-related quality of life rather than well-being or full quality of life. It has two components: five questions describing health and a visual analogue rating scale.

3.6.1 5-item questionnaire

The first component of the EQ-5D is a self-report questionnaire comprising five questions dealing with various aspects of current (today) physical and mental health status (mobility, self-care, usual activities, pain/discomfort, anxiety/depression). The response to each question is one of three

possible degrees of impairment (no problems, some problems or extreme problems). This gives an EQ-5D profile described by the levels on each item. For example, 21112 represents a health state with some problems with mobility, no problems with self-care, no problems with usual activities, no problems with pain/discomfort and some problems with anxiety/depression. The five dimensions aim to cover important aspects of health-related quality of life (HRQoL).

Dimensions for the instrument were selected after a review of the descriptive content of several generic health status questionnaires. The instrument was designed with a deliberate intention of being used in valuation exercises, which require a limited number of domains; hence there was a concentration upon only the 'core' aspects of HRQoL. Symptoms (such as pain, and mobility) which are clearly health but closer to inputs into quality of life than outputs or quality of life itself, are included in the HRQoL measure in part because they provide potentially clinically useful information. The instrument was designed to have only one item per dimension, though concepts may be interrelated. Pain, mobility and usual activities may be expected to be related but this may depend on the data, for example, it may apply in arthritis patients but not in patients with headaches.

Data from the EQ-5D questionnaire (which can define 243 health states) can be converted to an EQ-5D utility index using one of several scoring algorithms. Utility weights for the UK are usually those produced from a survey of the UK general public.[Dolan, 1997] Respondents used the TTO method to value hypothetical EQ-5D states. Respondents indicate the number of years in full health they consider to be indifferent to a certain number of years in a particular chronic health state. For example, if they are indifferent between living for 6 years in full health or 10 years in state 21112, this implies a utility value for state 21112 of 0.6 for that individual. The tariff for EQ-5D states were then derived using a regression model with an additive function of each attribute and level, with an additional interaction term for health states that experience 'severe' on any level (known as the N3 term). The tariff allows all EQ-5D profiles to be converted to health state utilities, on a zero to one scale, where zero represents the health state considered to be as bad as being dead and one represents full health.

The EQ-5D has been widely used in clinical trials and economic evaluations of health care across a large range of conditions. There is evidence for the validity of the measure in many patient populations, but concerns have emerged about the appropriateness in populations with sensory impairment, cognitive problems and more severe or complex mental health problems.[Longworth et al, 2014; Brazier et al, 2014] There has also been a concern that three levels are insufficient to measure smaller changes. A new version called the EQ-5D-5L has been developed which has five levels of severity (no problems, slight problems, moderate problems, severe problems, unable

to/extreme problems).[Herdman et al, 2011] Health state values for the EQ-5D-5L in the UK are still in development. In the interim a "crosswalk" value set that maps the EQ-5D-5L states to the utility scores for the EQ-5D-3L is available. This 5L version has been shown to offer some improvements in terms of greater sensitivity to differences between populations and a reduced ceiling effect.[Janssen et al, 2013]

The EQ-5D utility score does not clearly fit under any conception of well-being. The aim is to give a value to all health states; as such it draws on the notion of health-related quality of life, being based primarily on functioning and symptoms. However, the number of dimensions has been kept deliberately small, hence not all aspects of quality of life which could be impacted by health are included.

The use of population preferences to derive the 'value' of each health state implies the use of a social consensus on how good the state is, rather than a value that reflects the extent to which individual preferences or desires are satisfied. Therefore, the EQ-5D utility score is not a measure of preference satisfaction.

3.6.2 EQ-VAS

The second component of the EQ-5D is a visual analogue scale (VAS), which is a vertical scale for recording an individual's rating for their current health. This is anchored at the bottom at 0 (worst imaginable health state) and at the top at 100 (best imaginable health state). The EQ-VAS differs from the EQ-5D questionnaire and tariff score in a number of ways. The EQ-VAS may capture aspects of health that are not covered in the EQ-5D items (such as fatigue, cognition, vision). The bottom anchor is not anchored at dead unless individuals also mark the health state 'dead' onto the scale while the top anchor is best imaginable health rather than full health-related quality of life. Finally, the EQ-VAS relies only on judgment from the patient/respondent and does not draw upon values elicited from the general population. Indeed, the EQ-5D profiles have been found to only partly explain the EQ-VAS scores.[Feng et al, 2012]

VAS has been used both to value a respondents 'own' health state and to value one or more hypothetical health states. Critics of VAS have noted that they are subject to a number of biases, including spreading, context and end point bias.[Brazier et al, 2007] It has been argued that they provide only an ordinal ranking, and so do not provide information on strength of preference. As this method does not explicitly involve a choice, it has been argued that the VAS score does not represent utility, though this criticism is not universally accepted.[Parkin and Devlin, 2006]

The self-report EQ-VAS is only identifying a measure of 'health'. However, individuals may use as broad an interpretation of health as they like. They may include future anticipation of health states, such as prognosis or life expectancy, aspects of well-being, or adopt a very narrow view of health based on the presence of symptoms.

4 Comparison of measures

The measures will now be compared in terms of the concepts covered by the measures and the content in terms of the items, the response choices, recall period and scoring.

4.1 **Concepts and content of measures**

4.1.1 Measuring the different aspects of well-being

A summary of the concepts covered by each measure is presented below in Table 1.

Measuring affect: None of the six measures rely upon affect only. However, they each incorporate pleasant and unpleasant feelings to some extent. The ONS-4 asks about positive affect (happy) and negative affect (anxiety) experienced yesterday. The full WEMWBS aims to cover positive affect. It asks about 'cheerfulness' and 7 of its other items use the language of 'feeling', other questions are also indirectly about feelings (such as having energy, and feelings of optimism and feeling relaxed). The EQ-5D only taps into 'feelings' through the question on anxiety and depression. ASCOT asks respondents about being able to spend their time doing things they value or enjoy, which would be expected to tap into positive feelings. The GHQ-12 asks whether the respondent has been feeling "unhappy or depressed?", and "reasonably happy, all things considered". ICECAP-A asks if individuals "can have a lot of enjoyment and pleasure". So they all contain some content on affect, but none of the six measures here give solely an aggregate of positive and/or negative affect.

Measuring evaluation: The only measure that contains a direct question on life satisfaction is the ONS-4 and here it is taken as a partial measure of the broader concept of SWB. Although it is less direct, the WEMWBS also taps into how the respondent is feeling about their life through a number of different questions, including "feeling optimistic about the future". Within the GHQ-12, the question: have you "been feeling reasonably happy, all things considered", may be interpreted as an overall evaluation of one's life.

Measuring psychological well-being: The WEMWBS was designed as a measure of positive mental well-being. The authors describe the process of identifying items for the scale: "With reference to

current academic literature describing psychological and subjective well-being, the expert panel agreed key concepts of mental well-being to be covered by the new scale: positive affect and psychological functioning (autonomy, competence, self-acceptance, personal growth) and interpersonal relationships".[Tennant et al, 2007] The authors note that concepts of meaning of life and spirituality were not included because they may not have received "endorsement from the general UK population as related to mental well-being".[Tennant et al, 2007] The items on the GHQ-12 share many similarities with measures of psychological health; it contains affect (positive and negative) and psychological functioning (feelings of competence and usefulness). The ICECAP-A also covers a number of typical psychological flourishing domains (such as positive relationships, independence and achievements), though drawn from the public consensus on key functionings rather than based on psychological theory, and phrased in the form of opportunity or capability.

Measuring HRQoL: The EQ-5D measures an individual's physical functioning (mobility and self-care), activities (usual activities), and two widespread symptoms of many medical conditions: pain or discomfort, and anxiety or depression. None of the other measures contain direct measurement of physical functioning. EQ-VAS asks about health in general though respondents appear to be mainly thinking about physical health when they complete the rating task. ASCOT measures HRQoL indirectly through the extent to which care needs are met but not whether a person is able to meet those needs on their own (i.e. self-care). GHQ-12, WEMWBS and EQ-5D also aim to assess mental health, though the questions asked may fall into hedonic and/or flourishing accounts of well-being.

Measuring need: The social-care related quality of life, ASCOT, draws upon the idea of universal needs focusing on the sub-set that can be met by social care. The ASCOT stands out from the other well-being measures since it is concerned with need fulfillment, but by having the best level as an ideal it does cover how recipients feel about the fulfillment of these needs. It is not an objective account of needs fulfillment. Many of the areas covered differ from the other measures (for example accommodation, food and drink, personal care) and it is nearer to Maslow's idea of a hierarchy of need. However, the areas of safety, control, social participation and occupation are common with other well-being measures (such as WEMWBS), even though they are designed for the social care context.

Measuring capabilities: The ICECAP-A was designed to assess capabilities (or opportunity), but the content of these capabilities also overlaps with the concepts of psychological well-being (achievement, independence, relationships, security). The idea of capabilities is operationalised by the use of the terms 'able to' and 'can', rather than 'do' or 'feel'.

Positive and negative aspects of well-being: The different use of negatively and positively phrased

questions is a clear distinguishing feature between the different measures. The EQ-5D only identifies the presence of problems across physical and mental health domains. The highest achievable state is 'no problems' in any of the five domains; hence it suffers from ceiling effects in the general population where health problems are uncommon. The ASCOT, ICECAP-A, and WEMWBS use only positively phrased questions, whereas the GHQ-12 uses half positive, half negative. The range which may be covered by the worst response state in a positively phrased question may be quite broad; for example, 'feeling useful' 'none of the time' (WEMWBS) may differ considerably from 'thinking of yourself as a worthless person' 'more than usual' (GHQ-12), suggesting potential floor effects for purely positively phrased questionnaires. Some populations may experience severe negative feelings that may not be picked up by WEMWBS or ICECAP-A, such as those diagnosed with a severe or complex mental illness.

Three of the ONS-4 questions are positively phrased, with only 'anxiety' being negative. Again, it may be that the extreme response to absence of a valued functioning, such as responding 'not at all' to the question 'Overall, to what extent do you feel the things you do in your life are worthwhile?', may not be as negative a state as feeling that life is completely pointless. Given that only about 5% of respondents in UK population level surveys respond in the bottom 5 response options to this question, the impact of floor effects at the population level may be minimal. The positive phrasing may still mean it is insufficiently sensitive to identify important differences in psychological well-being in groups with severe mental health problems.

Where there are differences in the use of positive and negatively worded items between instruments, not only may the sensitivity differ at different levels of well-being, but measures may also be subject to different response errors. Negatively worded items may be more prone to socially desirable responding, or respondents feeling that questions are intrusive and overly personal.

There are a number of issues going on here. The first is the on-going debate on whether positive and negative affect are bipolar ends of the same dimension. Whilst they may be at any individual moment in time, there is an additional question of whether they can be treated as such over a time scale relevant for policy evaluation. This question of bipolarity applies both to affects such as happiness and sadness, as well as to broader experiences such as feelings of purposefulness or pointlessness.

The second issue is whether knowing only about the absence of problems or aspects which lower well-being will have ceiling effects and equivalently, whether knowing only about the presence of aspects which raise well-being will have floor effects.

4.1.2 Summary of concepts and content

Table 1 summarises the concepts that are covered by the measures. It reflects the considerable overlap between the scales, but no one measure covers all domains. The extent to which they measure the same things also depends on the precise wording of items and how they compare empirically.

Table 1: Comparison concepts covered by measures

Concepts	GHQ-12	WEMWBS	S-WEMWBS	ONS-4	ICECAP-A	ASCOT	EQ-5D	EQ-VAS
Preferences					*	*	*	
Affect (Hedonic)								
Positive	*	*		*				
Negative	*			*			*	
Evaluative				*				
Flourishing accounts:								
Psychological well-being	*	*	*	*	*	*		
Capabilities					*	*		
Health								
Physical							*	*
Mental	*	*	*	*			*	*
Needs based						*		

4.2 **Administration**

4.2.1 Response choices

The WEMWBS uses the response options of frequency (how often an emotion is felt), whereas ONS-4, ASCOT, ICECAP-A and EQ-5D opt for intensity/severity. Frequency, duration and intensity may be highly correlated.[Diener, 1994; 2000] Where the time scale is very short (now, yesterday) then intensity may be more appropriate. The GHQ-12 has a more problematic wording for the response choices since it uses frequency but in relation to 'usual', which may reflect short term changes but will be less sensitive to long term conditions that impact upon the respondent's perception of usual. This can, in part, be addressed by using the corrected GHQ scoring.

4.2.2 Recall time period

The duration of recall used in the instruments ranges from the moment (ICECAP-A), today (EQ-5D), yesterday (ONS-4 happy and anxious), last 2 weeks (WEMWBS), last few weeks (GHQ-12) and 'nowadays' (ONS-4 life satisfaction). The length of recall period of used in HRQoL (7 days versus today) has been found to make a difference to the utility values assigned to health states (Bansback et al, 2008].

Longer time frames are more likely to suffer from problems with memory recall and individuals may focus on particularly memorable experiences rather than their day-to-day experiences; on the other hand, short periods of recall may result in a very unrepresentative picture of someone's well-being. Differences in time-scale may also result in people interpreting response options differently.[Winkielman et al, 1998] The ONS technical advisory group report that some respondents were reluctant to report affects experienced yesterday when they were unusual, preferring to give a more typical response.[ONS, 2012] This was overcome by specific text asking for yesterday's affect even if atypical.

The time scale set out in the instrument instructions may not be the one that respondents apply when actually answering questions, which may be influenced by the questions being asked. For example, respondents to the ICECAP-A may draw upon their whole life experience when asked about 'achievements' rather than their life at the moment.[Al Janabi et al, 2010] Differences in time-scale (as it is actually used, not just as it is set out in the pre-amble) may also result in differences in the impact of personality factors and current mood upon the responses.

4.2.3 Social desirability, ordering and mode of administration

Responses to health and well-being questionnaires are known to be influenced by social desirability responding, question ordering effects, mode of administration and how respondents feel at the time of completing the interview.

A positive correlation was found between WEMWBS and the Impression Management and Self-Deception scores from the Balanced Inventory of Desirable Response, yet it was also found to compare favourably to other instruments they tested.[Tennant et al, 2007] ASCOT may also be vulnerable to social desirability responding as responders may seek not to offend care providers or relatives. This may well be a problem for the other instruments.

The physical layout of the instruments may introduce different susceptibility to response bias, particularly where responses are set out in a grid and respondents may wish to be consistent with

earlier responses. The early questions (for example optimism in the WEMWBS) may generate a potential framing effect for later responses (which would lead to an inflated internal consistency between responses). For this reason it has been argued that the ordering of items should be randomised between respondents, though this cannot fully overcome the framing effect.

The initial preamble to the questions could also set a particular framing. Both the EQ-5D and GHQ-12 include the term 'health' in the preamble, ICECAP-A asks about 'quality of life' and WEMWBS asks about 'experience'. This would suggest that 'health' may play a greater role in GHQ-12 response than in ICECAP-A or WEMWBS responses.

4.2.4 Scoring

The ONS-4 does not sum responses across the four items and so only provides profile information.

The WEMWBS and GHQ-12 measures can be scored by simple summative methods, whereby the responses are summed assuming equal weighting across the items. This assumes the instruments are uni-dimensional and are tapping into a single 'latent' construct. Having a number of separate questions helps reduce measurement error and if the instrument is designed to measure a single construct it should not matter if a respondent has a one point improvement on item A or a one point improvement on item B. There are a few important issues here; firstly, whether these measures are really picking up a single dimension – some evidence suggests this may not be the case particularly in relation to the GHQ. Secondly, even if the data shows a single dimension, this could mean that people with high levels on one item 'usually' do have high levels on the other item, but that they are separate constructs, and may come apart for certain subgroups. People may have preferences towards the importance of the different items. For example, an individual may consider an improvement in feeling close to others as more important than feeling relaxed.

The EQ-5D, ICECAP-A and ASCOT use relative weights between items and domains that have been derived from surveys of preferences of the general population. This allows the trade-offs between different levels on different domains to be based on the judgments of the general population. For example, surveys can reveal how much of a change in the other ICECAP-A dimensions they would consider equivalent to a move from level 3 to level 4 in independence. The EQ-5D generated preferences through Time Trade-Off (TTO), ICECAP-A and ASCOT both used Best Worst Scaling, and ASCOT also used TTO. These choice-based methods assume independence of the domains. Rather than assuming a single latent construct, these methods require individuals to make trade-offs between separate domains. These instruments avoid the concerns with lack of uni-dimensionality

since these instruments explicitly weight the different dimensions to derive an index. This also means they could be interpreted as providing a level of well-being which reflects society's preferences towards different types of lives, hence they are more compatible with a preference satisfaction account of well-being. However, some of the trade-offs between domains may be unrealistic given the level of correlation between them.

The EQ-5D and ASCOT have the additional advantage that they are able to be used in economic evaluation where the outcomes cover well-being and survival. GHQ-12, WEMWBS, ICECAP-A and ONS-4 cannot be combined with survival data and so have limited application in health economics. In principle it may be possible to recalibrate the well-being scales onto a full well-being-death scale (or 'WELBY') and this is the topic of another Department of Health project being undertaken by EEPRU.

In summary, the scores between the 6 measures are not comparable for a variety of reasons. They are often scored in different ways, with some using preferences and others using different psychometric ones (e.g. summative scoring or rasch analysis). Furthermore, each of the six measures use different top and bottom anchors (Table 2) which has implications for what scale respondents are using. This may even limit comparability between preference-based measures where measures are on a 0 to 1 scale, since one will mean different things between measures and ICECAP does not use dead for 0.

For this reason, the empirical comparisons presented in the next report uses standardized effect sizes to aid comparability.

Table 2: Top and bottom anchors of measures

Measure	Top anchor	Bottom anchor
EQ-5D	No problems on any of the five items.	Equivalent of dead anchored at 0. The worst state possible is severe problems with mobility, self-care, usual activities, pain/discomfort, and depression and anxiety. In the UK tariff this is valued at -0.59.
EQ-VAS	Best imaginable health.	Worst imaginable health. If 'dead' is valued onto the VAS, health states may be recalibrated such as that 0 is equivalent to dead.
ICECAP-A	The ability to achieve each of the 5 key functionings.	No ability to achieve any of the 5 key functionings. Set at 0.
ASCOT	No unmet social-care related needs/wants. And the opportunity for chosen levels of control over daily life, social contact and doing enjoyable and valuable things.	Anchored at 0 as equivalent to dead. The worst state possible is responding: • My home is not at all clean or comfortable • I don't feel at all safe • I don't always get adequate or timely food and drink, and I think there is a risk to my health • I don't feel at all clean or presentable • I have no control over my daily life • I have little social contact with people and feel socially isolated • The way I'm helped and treated completely undermines the way I think and feel about myself • I don't do anything I value or enjoy with my time
S-WEMWBS	Experiencing each of the 7 positive states 'all of the time'.	Experiencing each of the 7 positive states 'none of the time'.
GHQ-12	Experiencing the 6 positive states 'more so than usual' (or 'same as usual' depending on the scoring), and the 6 negative states 'not at all'.	Experiencing the 6 positive states 'much less than usual' and the 6 negative states 'much more than usual'.

5 Discussion and Conclusion

Different theories of well-being offer a number of competing conceptions of well-being. We have described five different theories of well-being, which present competing views on what it means for a life to go well (see Figure 1 for an overview).

We can usefully distinguish in theory between:

i. Complete theories of well-being that identify what it is to lead a good life such as preference fulfilment, hedonism and life satisfaction.
ii. Partial theories of well-being which state that something is an important aspect of well-being but do not aim to give a complete explanation of well-being.
iii. Descriptions of inputs or basic functionings (either physical or psychological) which are not setting out to describe well-being *per se,* but rather to establish whether conditions exist that would enable well-being (however defined) to occur. Achieving these basic functionings would be a necessary but not sufficient condition for high well-being. The justification for their importance may be based on theory (as in the case of Ryan and Deci's Self-Determination theory), social agreement, or on a statistical relationship.

A clear distinction between outcomes (non-instrumental) and determinants (instrumental) is not always easy in practice. The same attribute may be taken as important for different reasons. For example, 'relatedness' may be seen as part of what it is for life to go well (as it is in Ryff's Psychological Well-Being theory), a universal psychological need required for health and well-being (as it is in Ryan and Deci's Self-Determination theory), or a characteristic which, on average, predicts higher life satisfaction (as is often found to be the case in empirical data).[Dolan et al, 2008] As a consequence, measurement instruments drawn from different theoretical backgrounds may share similar content.

From a policy perspective, and particularly from the perspective of conducting economic evaluation, it clearly does matter if the well-being measure is to be taken as a complete measure of well-being, part of well-being or an important input to well-being. Whilst inputs to well-being are important for policy, for evaluation purposes, considering both inputs and outputs runs the risk of double counting potential benefits.

Our six instruments draw from a number of competing conceptual theories somewhat 'pick and mix' style, taking some aspects from a hedonist theory, some from evaluative theory, some from preference satisfaction. If a number of possible ways of thinking about well-being are attractive, but

there is no commitment to a single theory which explains what well-being is, then it becomes even more difficult to conceive of a way of constructing a single index of overall well-being.

We have considered a number of ways in which our six instruments differ:

- They are measuring different overall constructs: health-related quality of life (EQ-5D), mental health (GHQ-12, WEMWBS), social-care needs (ASCOT), capabilities (ICECAP-A), SWB including evaluative, affect and flourishing (ONS-4).
- Those that measure mental health or psychological well-being break down the construct into different dimensions (GHQ has positive and negative items, WEMWBS has only positive items).
- They have a different potential for accuracy of measuring dimensions due to different number of items (for example EQ-5D and ONS-4 have one item on negative affect, GHQ-12 has 6).
- They are likely to have a different degree of sensitivity at different levels of the construct (for example the WEMWBS may be sensitive to a change in a good mental health state but not in a very poor mental health state), hence subject to potential ceiling or floor effects.

With these potential theoretical and measurement differences in mind, the key question is whether or not these differences are also evident empirically. The measures are used to judge whether or not a good life is being lived. It is relatively easy to think of examples in which judgment would collide (a happy paraplegic, for example). However, a more relevant question is whether well-being measures drawn broadly from these different approaches present levels of well-being for individuals or groups which substantively differ empirically in ways which are policy relevant. Furthermore, it may be the case that some measures/questions are redundant given other measures/questions already included. For example, there may be redundancy across the measures included within the ONS well-being battery, particularly in relation to the WEMWBS and GHQ-12. Is it necessary to include feeling 'relaxed' (WEMWBS), and feeling 'anxious' (ONS-4), and feeling 'constantly under strain' (GHQ-12)? The next step will be to compare the measures empirically using existing datasets.

References

Al-Janabi H, Flynn T, Coast J. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research*. 2012; *21*(1), 167-176.

Alkire S. Dimensions of human development. *World Development*. 2002; *30*(2), 181-205.

Alkire S. The capability approach to the quality of life. Working paper prepared for the Working Group "Quality of Life" The Measurement of Economic Performance and Social Progress. 2008.

All-Party Parliamentary Group on Wellbeing Economics, Wellbeing in four policy areas, September 2014. Downloaded from http://parliamentarywellbeinggroup.org.uk/ on 5-10-2014.

Arthaud-Day ML, Rode JC, Mooney CH, Near JP. The subjective well-being construct: A test of its convergent, discriminant, and factorial validity. *Social Indicators Research*. 2005; 74, 445–476.

Bansback N, Sun H, Guh DP, Li X, Nosyk B, Griffin S, Barnett PG, Anis AH, OPTIMA team. Impact of the recall period on measuring health utilities for acute events. *Health economics*. 2008; *17*(12), 1413-1419.

Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *Journal of Health Economics*. 2012; 31(1), 306-318.

Baumeister R, Vohs K, Aaker J, Garbinsky E. Some Key Differences between a Happy Life and a Meaningful Life. *Journal of Positive Psychology.* 2013; 8, 505-516.

Bradburn NM. The structure of psychological well-being. Chicago: Aldine. 1969.

Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford University Press. 2007.

Connell J, Brazier JE, O'Cathain A, Lloyd-Jones M, Paisley S. Quality of life of people with mental health problems: a synthesis of qualitative research. *Health and Quality of Life Outcomes.* 2012; 10:138.

Crisp R. Hedonism reconsidered. *Philosophy and Phenomenological Research*. 2006; 73(3), 619-645.

Csikszentmihalyi M, Hunter. Happiness in everyday life: The uses of experience sampling. *Journal of Happiness Research*. 2003; 4(2), 185-199.

De Wit GA, Busschbach JJ, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? *Health Economics*. 2000; 9(2), 109-126.

Deci EL, Ryan RM. Overview of self-determination theory: An organismic dialectical perspective. *Handbook of self-determination research.* 2002; 3-33.

Department of Health. The NHS Outcomes Framework 2013–14. London: November 2012.

Department of Health, 2013.

Diener E. Assessing subjective well-being: Progress and opportunities. *Social Indicators Research*. 1994; 31, 103-157.

Diener E, Emmons RA. The independence of positive and negative affect. *Journal of Personality and*

Social Psychology. 1984; 47(5), 1105-1117.

Diener E, Emmons RA, Larson RJ, Griffin S. The satisfaction with life scale. *Journal of Personality Assessment*. 1985; 49, 71-75.

Diener E, Seligman ME. Beyond money toward an economy of well-being. *Psychological science in the public interest*. 2004; 5(1), 1-31.

Diener E, Suh E. Measuring quality of life: Economic, social, and subjective indicators. *Social indicators research*. 1997; 40(1-2), 189-216.

Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, Biswas-Diener R. New measures of well-being: Flourishing and positive and negative feelings. *Social Indicators Research*. 2009; 39, 247-266.

Diener E, Inglehart R, Tay L. Theory and validity of life satisfaction scales. *Social Indicators Research*. 2013; 112(3), 497-527.

Dolan P, Kahneman D. Interpretations of Utility and Their Implications for the Valuation of Health. *The Economic Journal*. 2008; 118(525), 215-234.

Dolan P, Peasgood T, White M. 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.* 2008*;* 29(1), 94-122.

Dolan P. Modeling valuations for EuroQol health states. *Medical Care*. 1997; 35(11), 1095-1108.

Dolan P, Metcalfe R. Measuring subjective wellbeing: recommendations on measures for use by national governments. *Journal of Social Policy*. 2012; 41(02), 409-427.

Doyal L, Gough I. *A theory of human need*. Palgrave Macmillan. 1991.

EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. *Health policy.* 1990; 16(3), 199.

Feng Y, Parkin D, Devlin NJ. Assessing the performance of the EQ-VAS in the NHS PROMs programme. *Quality of Life Research. 2014; 23*(3), 977-989.

Flynn TN, Huynh E, Peters TJ, Al-Janabi H, Moody A, Clemens S, Coast J. Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. Health Economics. 2013; Online first: doi: 10.1002/hec.3014.

Fordyce MW. A review of research on The Happiness Measures: A sixty second index of happiness and mental health. *Social Indicators Research*. 1988; 20, 63-89.

Goldberg DP, Hillier VF. A scaled version of the General Health Questionnaire. *Psychol Med*. 1979; 9:139–145.

Goldberg DP, Williams P. A user's guide to the GHQ. *Windsor, NFER Nelson*. 1988.

Goodchild ME, Duncan-Jones P. Chronicity and the General Health Questionnaire. *Br J Psychiatry*. 1985; 146:55–61.

Graetz B. Multidimensional properties of the general health questionnaire. *Social psychiatry and*

psychiatric epidemiology. 1991; *26*(3), 132-138.

Green DP, Goldman SL, Salovey P. Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*. 1993; 64(6), 1029.

Grewal I, Lewis J, Flynn TN, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Social Science & Medicine.* 2006; 62: 1891-1901.

Hankins M. The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC Public Health*. 2008; *8*(1), 355.

Haybron DM. *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford University Press. 2008.

Helliwell JF, Wang S. The state of world happiness. *World happiness report 2012*, 10-57. 2012.

Helliwell JF, Layard R, Sachs J. (Eds.) *World happiness report 2013*. Sustainable Development Solutions Network. 2013.

Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G and Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011; 20: 1727–36.

Huppert FA, So TT. Flourishing across Europe: Application of a new conceptual framework for defining well-being. *Social Indicators Research.* 2013; 110(3), 837-861.

Huppert FA and Whittington JE. Evidence for the independence of positive and negative well-being: implications for quality of life assessment. *Br. J. Health Psychol*. 2003; 8, 107–122.

Hurka T. *Perfectionism*. Oxford: Oxford University Press. 1993.

Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Swinburn P, Busschbach J. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research*. 2013; 22(7), 1717-1727.

Kahneman D, Knetsch JL. Valuing public goods: the purchase of moral satisfaction. *Journal of environmental economics and management*. 1992; 22(1), 57-70.

Kahneman D, Wakker PP, Sarin R. Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*. 1997; 112(2), 375-406.

Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: The day reconstruction method. *Science*. 2004; 306(5702), 1776-1780.

Kahneman D, Krueger AB. Developments in the measurement of subjective well-being. *The journal of economic perspectives*. 2006; 20(1), 3-24.

Kahneman D, Deaton A. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Science.* 2010; 107, 16489–16493.

Kammann R, Flett R. Affectometer 2: a scale to measure current level of general happiness. *Australian Journal of Psychology*. 1983; 82, 1007-1022.

Kapteyn *A*, Lee J, Tassot C, Vonkova H, Zamarro G. Dimensions of Subjective Well Being. CESR working paper series, paper no. 2013005. 2013.

Keyes CL. The mental health continuum: From languishing to flourishing in life. *Journal of health and social behavior*. 2002; 207-222.

Keyes CLM. Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology*. 2005; 73: 539–548.

Kinghorn PB, Robinson A. Smith R Developing the capability approach to assess quality of life in patients with chronic pain. In *iHEA 2007 6th World Congress: Explorations in Health Economics Paper*. 2007.

Knabe A, Rätzel S, Schöb R, Weimann J. Dissatisfied with Life but Having a Good Day: Time-use and Well-being of the Unemployed. *Economic Journal.* 2010; 120, 867–889.

Larsen RJ, Fredrickson BL. Measurement issues in emotion research. In Kahneman D, Diener Ed, Schwarz N (Eds) *Well-being: The foundations of hedonic psychology.* New York, NY, US: Russell Sage Foundation . 1999.

Leamy M, Bird V, Le Boutillier C, Wiliams J, Slade M. A conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *British Journal of Psychiatry*. 2011; 199, 445-52.

Longworth L, Yang Y, Young T, Hernandez Alva M, Mukuria C, Rowen D, Tosh J, Tsuchiya A, Evans P, Keetharuth A, Brazier J. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: systematic review, statistical modeling and survey. *Health Technology Assessment*. 2014; *18*, 9.

Malley J, Towers A, Netten A, Brazier J, Forder J, Flynn, T. An assessment of the construct validity of the ASCOT measure of social care-related quality of life with older people. *Health and Quality of Life Outcomes*. 2012; 10:21.

Maslow AH. A theory of human motivation. *Psychological Review*. 1943; *50*(4), 370.

Max-Neef M. Elizalde A, Hopenhayn M. Development and human needs. *Real-Life Economics: understanding wealth creation. Edited by: Ekins P, Max-Neef M. Routledge, London*, 1992; 197-213.

McDowell I. *Measuring health: a guide to rating scales and questionnaires*. Oxford University Press. 2006.

The National Research Council. *Subjective well-being: measuring happiness, suffering and other dimensions of experience*. Panel on Measuring Subjective Well-being in a Policy Relevant Framework; Committee on National Statistics, Division on Behavioral and Social Sciences and Education. Stone AA and Mackie C (Eds), National Academies Press (US), 2013. Dec.

Netten AP, Beadle-Brown J, Caiels J, Forder JE, Malley J, Smith NJ, Windle K *et al*. ASCOT Adult Social Care Outcomes Toolkit: Main Guidance v2. 1 PSSRU Discussion Paper 2716/3. University of Kent. 2011.

Netten A, Burge P, Malley J, Potoglou D, Towers AM, Brazier J, Forder J. *et al*. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technology Assessment*. 2012; *16*(16), 1-166.

National Institute for Health and Care Excellence (NICE). Methods for the development of NICE public health guidance (3rd Edition). 2012.

National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal. 2013a.

National Institute for Health and Care Excellence (NICE). The Social Care Guidance Manual. 2013b.

Nussbaum MC. *Women and human development: The capabilities approach*. Cambridge University Press. OECD, 2000.

OECD. OECD Guidelines on Measuring Subjective Well-being. Paris: OECD Publishing, 2013.
Office for National Statistics, Subjective well-being: A qualitative investigation of subjective well-being questions, results of cognitive testing carried out during development of the ONS's experimental subjective well-being questions. December 2012.

Parfit D. *Reasons and Persons.* Oxford University Press. 1984.

Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics*. 2006; 15(7), 653-664.

Pudney S. *An experimental analysis of the impact of survey design on measures and models of subjective wellbeing* (No. 2010-20). ISER Working Paper Series. 2010.

Rawls J. *A Theory of Justice*. Cambridge. Mass.: Harvard University. 1971.

Revicki DA. Health related quality of life in the evaluation of medical therapy for chronic illness. *The Journal of Family Practice.* 1989; 29, 377-380.

Robeyns I, van der Veen R J, Petersen AC. (2007). Sustainable quality of life. *Conceptual analysis for a policyrelevant empirical specification*.

Russell JA, Carroll JM. On the bipolarity of positive and negative affect. *Psychological bulletin*. 1999; *125*(1), 3.

Deci EL, Ryan RM. Overview of self-determination theory: An organismic dialectical perspective. *Handbook of self-determination research.* 2002; 3-33.

Ryff CD. Psychological well-being in adult life. *Current Directions in Psychological Science*. 1995; 4(4), 99-104.

Ryff CD, Singer B. The contours of positive human health. *Psychological Inquiry*. 1998; 9, 1-28.

Salgueiro MF, Malta J. Latent growth curve modeling of psychological well-being trajectories. *American Journal of Theoretical and Applied Statistics*. 2013; *2*(3), 61-66.

Schwarz N, Clore GL. Mood as information: 20 years later. *Psychological Inquiry.* 2004; 14(3-4), 296-303.

Seligman ME. *Flourish: A visionary new understanding of happiness and well-being*. Simon and Schuster. New York. 2011.

Sen A. *Commodities and capabilities.* Oxford University Press. 1999.

Sen A. Capabilities, Lists, and Public Reason: Continuing the conversation, *Feminist Economics. 2004;* 10(3), 77-80.

Shevlin M, Adamson G. Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychological Assessment*. 2005; *17*(2), 231.

Smith AB, Oluboyede Y, West R, Hewison J, House AO. The factor structure of the GHQ-12: the interaction between item phrasing, variance and levels of distress. *Quality of Life Research*. 2013; *22*(1), 145-152.

Stiglitz J, Sen A, Fitoussi JP. *Report by the Commission on the Measurement of Economic Performance and Social Progress.* 2009.

Stochl J, Jones PB, Croudace TJ. Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers. *BMC Medical Research Methodology*. 2012; 12(1), 74.

Strack F, Martin LL, Schwarz N. Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology* 1988; 18, 429-42.

Sumner LW. Welfare, Happiness, and Ethics. Oxford University Press. 1996.

Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, Stewart-Brown S, *et al*. The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health and Quality of life Outcomes.* 2007*; 5*(1), 63.

Torrance GW. Utility approach to measuring health-related quality of life. *Journal of chronic diseases*, 1987; *40*(6), 593-600.

Watson D. Intra-individual and inter-individual analyses of positive and negative affect: their relation to health complaints, perceived stress, and daily activities. *Journal of Personality and Social Psychology*. 1988; 54(6), 1020-1030.

Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scale. *Journal of Personality and Social Psychology.* 1988; 54, 1063-1070.

Winkielman P, Knäuper B, Schwarz N. Looking back at anger: Reference periods change the interpretation of emotion frequency questions. *Journal of Personality and Social Psychology.* 1998; 75.3: 719.

Woodard C. Classifying theories of welfare. *Philosophical Studies*. 2013; *165*(3), 787-803.

World Health Organization. The first ten years of the World Health Organization. Geneva: World Health Organization, 1958.

APPENDIX 1: Measures of individual well-being

General Health Questionnaire 12 (QHQ12)

We want to know how your health has been in general over the last few weeks. Please read the questions below and each of the four possible answers. Circle the response that best applies to you. Thank you for answering all the questions.

Have you recently:

1. been able to concentrate on what you're doing?
better than usual, same as usual, less than usual, much less than usual

2. lost much sleep over worry?
not at all, no more than usual, rather more than usual, much more than usual

3. felt that you are playing a useful part in things?
more so than usual, same as usual, less so than usual, much less than usual

4. felt capable of making decisions about things?
more so than usual, same as usual, less so than usual, much less than usual

5. felt constantly under strain?
not at all, no more than usual, rather more than usual, much more than usual

6. felt you couldn't overcome your difficulties?
not at all, no more than usual, rather more than usual, much more than usual

7. been able to enjoy your normal day to day activities?
more so than usual, same as usual, less so than usual, much less than usual

8. been able to face up to your problems?
more so than usual, same as usual, less so than usual, much less than usual

9. been feeling unhappy or depressed?
not at all, no more than usual, rather more than usual, much more than usual

10. been losing confidence in yourself?
not at all, no more than usual, rather more than usual, much more than usual

11. been thinking of yourself as a worthless person?
not at all, no more than usual, rather more than usual, much more than usual

12. been feeling reasonably happy, all things considered?
not at all, no more than usual, rather more than usual, much more than usual

NB: Measures are copyrighted. Contact developers to register use.

The Full Warwick Edinburgh Mental Well-Being Scale (WEMWBS)

Below are some statements about feelings and thoughts. Please tick the box that best describes your experience of each over the last 2 weeks

STATEMENTS	None of the time	Rarely	Some of the time	Often	All of the time
I've been feeling optimistic about the future	1	2	3	4	5
I've been feeling useful	1	2	3	4	5
I've been feeling relaxed	1	2	3	4	5
I've been feeling interested in other people	1	2	3	4	5
I've had energy to spare	1	2	3	4	5
I've been dealing with problems well	1	2	3	4	5
I've been thinking clearly	1	2	3	4	5
I've been feeling good about myself	1	2	3	4	5
I've been feeling close to other people	1	2	3	4	5
I've been feeling confident	1	2	3	4	5
I've been able to make up my own mind about things	1	2	3	4	5
I've been feeling loved	1	2	3	4	5
I've been interested in new things	1	2	3	4	5
I've been feeling cheerful	1	2	3	4	5

The Short Warwick-Edinburgh Mental Well-being Scale (S-WEMWBS)

Below are some statements about feelings and thoughts. Please tick the box that best describes your experience of each over the last 2 weeks

STATEMENTS	Noneof thetime	Rarely	Some of the time	Often	All ofthetime
I've been feeling optimistic about the future	1	2	3	4	5
I've been feeling useful	1	2	3	4	5
I've been feeling relaxed	1	2	3	4	5
I've been dealing with problems well	1	2	3	4	5
I've been thinking clearly	1	2	3	4	5
I've been feeling close to other people	1	2	3	4	5
I've been able to make up my own mind about things	1	2	3	4	5

"Short Warwick Edinburgh Mental Well-Being Scale (SWEMWBS)

ONS subjective well-being overview questions (ONS-4)

On a scale of 0 to 10 where ‘0 ’means ‘not at all’ and ‘10’ is ‘completely’:

Overall, how satisfied are you with your life nowadays?

Overall, how happy did you feel yesterday?

Overall, how anxious did you feel yesterday?

Overall, to what extent do you feel the things you do in your life are worthwhile?

ICECAP-A

ABOUT YOUR OVERALL QUALITY OF LIFE

Please indicate which statements best describe your overall quality of life at the moment by placing a tick in ONE box for each of the five groups below.

1. Feeling settled and secure
    - I am able to feel settled and secure in all areas of my life
    - I am able to feel settled and secure in many areas of my life
    - I am able to feel settled and secure in a few areas of my life
    - I am unable to feel settled and secure in any areas of my life

2. Love, friendship and support
    - I can have a lot of love, friendship and support
    - I can have quite a lot of love, friendship and support
    - I can have a little love, friendship and support
    - I cannot have any love, friendship and support

3. Being independent
    - I am able to be completely independent
    - I am able to be independent in many things
    - I am able to be independent in a few things
    - I am unable to be at all independent

4. Achievement and progress
    - I can achieve and progress in all aspects of my life
    - I can achieve and progress in many aspects of my life
    - I can achieve and progress in a few aspects of my life
    - I cannot achieve and progress in any aspects of my life

5. Enjoyment and pleasure
    - I can have a lot of enjoyment and pleasure 4
    - I can have quite a lot of enjoyment and pleasure 3
    - I can have a little enjoyment and pleasure 2
    - I cannot have any enjoyment and pleasure 1

The Adult Social Care Outcome Toolkit (ASCOT)

Domain level
Accommodation cleanliness and comfort
1. My home is as clean and comfortable as I want
2. My home is adequately clean and comfortable
3. My home is not quite clean or comfortable enough
4. My home is not at all clean or comfortable
Safety
1. I feel as safe as I want
2. Generally I feel adequately safe, but not as safe as I would like
3. I feel less than adequately safe
4. I don't feel at all safe
Food and drink
1. I get all the food and drink I like when I want
2. I get adequate food and drink at OK times
3. I don't always get adequate or timely food and drink
4. I don't always get adequate or timely food and drink, and I think there is a risk to my health
Personal care
1. I feel clean and am able to present myself the way I like
2. I feel adequately clean and presentable
3. I feel less than adequately clean or presentable
4. I don't feel at all clean or presentable
Control over daily life
1. I have as much control over my daily life as I want
2. I have adequate control over my daily life
3. I have some control over my daily life, but not enough
4. I have no control over my daily life
Social participation and involvement
1. I have as much social contact as I want with people I like
2. I have adequate social contact with people
3. I have some social contact with people, but not enough
4. I have little social contact with people and feel socially isolated

Domain level
Dignity
1. The way I'm helped and treated makes me think and feel better about myself
2. The way I'm helped and treated does not affect the way I think or feel about myself
3. The way I'm helped and treated sometimes undermines the way I think and feel about myself
4. The way I'm helped and treated completely undermines the way I think and feel about myself
Occupation and employment
1. I'm able to spend my time as I want, doing things I value or enjoy
2. I'm able do enough of the things I value or enjoy with my time
3. I do some of the things I value or enjoy with my time, but not enough
4. I don't do anything I value or enjoy with my time

NB: Measures are copyrighted. Contact developers to register use.

EQ-5D Health Questionnaire

Here are some simple questions about your health in general. By ticking one answer in each group below, please indicate which statements best describe your own health state TODAY.

Please tick one

1. Mobility

- I have no problems in walking about ☐
- I have some problems in walking about ☐
- I am confined to bed ☐

2. Self-care

- I have no problems with self-care ☐
- I have some problems washing or dressing myself ☐
- I am unable to wash or dress myself ☐

3. Usual Activities

- I have no problems with performing my usual activities (e.g. work, study, housework, family or leisure activities) ☐
- I have some problems with performing my usual activities ☐
- I am unable to perform my usual activities ☐

4. Pain/Discomfort

- I have no pain or discomfort ☐
- I have moderate pain or discomfort ☐
- I have extreme pain or discomfort ☐

5. Anxiety/Depression

- I am not anxious or depressed ☐
- I am moderately anxious or depressed ☐
- I am extremely anxious or depressed ☐

To help people say how good or bad their health is, we have drawn a scale (rather like a thermometer) on which the best state you can imagine is marked by 100 and the worst state you can imagine is marked by 0.

We would like you to indicate on this scale how good or bad your own health is today, in your opinion. Please do this by drawing a line from the box below to whichever point on the scale indicates how good or bad your current health state is.

Your own health state today

Best imaginable health state

Worst imaginable health state