

**U.S. National Institute on Aging
U.K. Economic and Social Research Council
U.S. National Academy of Sciences**

**Workshop on the Role of Well-Being Measures in Public Policy
The Keck Center of the National Academies
Washington, DC**

November 8-9, 2010

WORKSHOP SUMMARY¹

Rev. March 18, 2011

Subjective well-being can be conceptualized as the combination of experienced affect (emotions ranging from joy to misery) and evaluations of life satisfaction/dissatisfaction. In recent decades, interest has grown rapidly in research on broader measures of well-being that go beyond material standard of living and market-based economic concepts. The United Kingdom (U.K.), the Gallup World Poll, the Organisation for Economic Co-operation and Development (OECD), and the Commission on the Measurement of Economic Performance and Social Progress² have argued that governments and population surveys should measure people's well-being as a way of assessing societal progress.

In an effort to engage research and policy personnel in dialogue on the potential policy applications of well-being science, the National Institute on Aging (NIA) and the U.K. Economic and Social Research Council (ESRC) jointly commissioned a National Academies workshop on November 8-9, 2010, to gather leading academic and policy personnel from the United States and the United Kingdom to examine the potential for subjective well-being measures to inform the design and evaluation of specific social and economic programs. Establishing the policy relevance of well-being data and measures creates a compelling reason to continue supporting their development. In the United States (U.S.), the challenge is not necessarily in creating the measures, but in convincing agencies, legislators, and the general public that subjective well-being data would be cost effective to collect and publish, and useful to inform policy where appropriate.

In her welcoming remarks, workshop chairperson Katharine Abraham (University of Maryland)

¹ This workshop summary was prepared by Chandra Keller-Allen and Rose Maria Li, Rose Li and Associates, Inc. under contract to the National Institute on Aging (HHSN263200700991P) and the U.K. Science and Innovation Network. The statements, conclusions, and recommendations contained in this document reflect opinions of the workshop participants and are not intended to represent the official position of the meeting sponsors, the National Institutes of Health, the U.S. Department of Health and Human Services, or the U.K. Economic and Social Research Council. We gratefully acknowledge those who reviewed earlier drafts of the report: Paul Allin, Susan Cartwright, Somnath Chatterji, Paul Dolan, Carol Graham, Robert Groves, Brian Harris-Kojetin, Felicia Huppert, Steve Landefeld, Richard Layard, Richard E. Lucas, Christine McGuire, Lis Nielsen, Jacqui Smith, Andrew Steptoe, Richard Suzman, Joy Todd, and Katherine K. Wallman.

² Established by French President Nicolas Sarkozy, see: <http://www.stiglitz-sen-fitoussi.fr/en/index.htm>.

traced the origin of interest on subjective well-being to a paper in 1980 written by Richard Layard (London School of Economics).³

The NIA's interest in subjective well-being has focused on supporting the development and refinement of measures and concepts on the full range of well-being from misery to happiness, from experienced to evaluative, and understanding the causal relationships. Richard Suzman (NIA Division of Behavioral and Social Research [BSR]) observed that the measures of well-being used in aging research have focused almost exclusively on life satisfaction, which is not sufficient. Until shown otherwise, measures of experienced well-being and affect should be considered equally valuable. They likely tap somewhat different domains of psychological functioning, and empirical research is needed to discover how the experience of the two different types of well-being differentially impacts health and other outcomes. Experiential measures should be assessed alongside measures of life satisfaction in longitudinal studies.

This workshop is part of a larger effort of the NIA to support the integration of measurement, policy, and life-course themes in the science of well-being, to augment survival statistics, and to enhance the policy relevance of this field. In the past, the NIA has funded a center at Princeton University that expanded on work to develop measures of momentary experienced well-being. More recent efforts to advance this area include a request for applications (RFA) to improve measurement of well-being and understanding of how different measures affect people,⁴ funding to the Bureau of Labor Statistics (BLS) to add subjective well-being measures to the American Time Use Survey, and commissioning a National Academies panel chaired by Joseph Newhouse (Harvard University) on the design of national health accounts.⁵ The NIA is especially interested in longitudinal panel data that will allow for the examination of well-being throughout the life course and its impact on health. The NIA has encouraged grantees with ongoing longitudinal panel studies to add measures of well-being; taken together, these surveys represent roughly 65 percent of the world's population over the age of 50. Lis Nielsen (NIA BSR) underscored the program's interest in bridging approaches from psychology and economics as this field progresses.

The considerable recent interest in subjective well-being measures in the United Kingdom has been in part stimulated by an increasing desire among policy makers for evidence on what works to promote well-being. The ESRC has a role in funding high quality empirical and methodological research with high potential for impact on policy and practice whilst maintaining its independence from Government agendas. The ESRC takes an interdisciplinary lifecourse perspective to issues around ageing and well-being. Joy Todd (ESRC) described the ESRC's approach to working with government departments to develop research agendas (e.g., Department for Work and Pensions, Department of Health, Department for Education, and the Department for Communities and Local Government). A key factor in discussions with policy makers in the area of subjective well-being is the need to bridge the gaps between researchers

³ Layard, R. (1980). Human satisfactions and public policy. *The Economic Journal*, 90, 737-750.

⁴ Subjective Well-Being: Advances in Measurement and Applications to Aging (R01, RFA-AG-11-003) <http://grants.nih.gov/grants/guide/rfa-files/RFA-AG-11-003.html>.

⁵ Panel to Advance a Research Program on the Design of National Health Accounts; National Research Council. 2010. Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement. Washington, DC: The National Academies Press. Prepublication available: http://www.nap.edu/catalog.php?record_id=12938.

and policy makers. The ESRC has supported a range of work including a professorial fellowship, international work with Andrew Steptoe on positive well-being, research at the Centre for Economic Performance at the London School of Economics, and the development of well-being measures. The current political climate in the United Kingdom is very supportive of integrating measures of well-being into various aspects of the policy process. The Office of National Statistics (ONS) has received funding to develop measures of well-being, and it has published an article outlining how the ONS will engage stakeholders and address recommendations from the Report by the Commission on the Measurement of Economic Performance and Social Progress.⁶

Connie Citro (Committee on National Statistics [CNSTAT], The National Academies) reported that CNSTAT has undertaken several studies and has worked in cooperation with the NIA, ESRC, and the National Academies Division of Behavioral and Social Sciences and Education (DBASSE). After this workshop, CNSTAT will review what has been learned and possibly work further with the NIA to move forward with a study. CNSTAT also is working with the OECD to develop a handbook on measures of well-being. Economic statistics provide a great deal of information, but capturing a full range of the well-being of individuals and families is important.

Several broad themes emerged from the workshop presentations, invited remarks, and discussion sessions:

- There was general agreement on the importance of this field and that subjective well-being measures could be useful and appropriate for targeted populations and targeted policies.
- There was less agreement about the use of well-being measures for national monitoring purposes and whether a national indicator of well-being should be one that is responsive and changes over time.
- There was disagreement on the issue of timing, particularly for a national progress monitoring function. Some advocated launching well-being measures now with what knowledge we have and revise as needed. Others felt that much more research is needed before doing so.
- Most agreed and were encouraged by the possibility of focusing on changes in well-being in smaller geographic areas and the possibility that policy interventions, and the use of well-being measures to evaluate policy interventions, may have the greatest impact at a more targeted level.
- The current political climate in the United Kingdom is favorable to use of well-being measures in national surveys to inform policy. Many saw this as a time to boldly move forward.

⁶ Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). *Report of the Commission on the Measurement of Economic Performance and Social Progress*. Retrieved March 16, 2011, from http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.

- U.S. agencies have plans to improve elements of the gross domestic product (GDP), continue support of in-depth research on well-being measures in the academic sector, and pilot the use of some well-being measures in some national surveys under certain circumstances.
- It is necessary to be clear about the purposes to which the well-being measures are being put (i.e., national progress monitoring, informing policy design, policy evaluation), which domains of well-being are appropriate for a given purpose, and the types of measures needed for specific domains of well-being (evaluated, experienced, eudaimonic).
- There is an important role for researchers in the international community to learn from each other as measures are developed and implemented. The suggestion of a separate role for a body of the international community to coordinate and enable comparisons was more contentious.
- There was little consensus about the proper role of the public in accepting, understanding, and legitimizing the use of well-being measures.
- There was disagreement and debate about whether an understanding of the causality of well-being is needed to inform cost-benefit analyses and policy decisions. Some argued strongly for more research on causal modeling of well-being, and others emphasized field experiments as deserving of focus.

A summary of the workshop proceedings follows. The meeting agenda and list of participants are included as appendices 1 and 2. Biographical sketches of presenters and their responses to questions circulated in advance of the meeting are available from the Committee on National Statistics.

Overview

Daniel Kahneman

Eugene Higgins Professor of Psychology, emeritus, Professor of Psychology and Public Affairs, emeritus, and Senior Scholar, Princeton University

There is a distinction between experiencing life as it is lived and remembering life—the experiencing self and the remembering self. The term “well-being” has traditionally focused on the remembering self. Ways to measure well-being of the experiencing self were being developed (i.e., experience sampling) 10 years ago, but they did not appear to be effective methods.

Several years ago, Kahneman convened a team of psychologists and economists, most notably Alan Krueger (Princeton University), with the goal of developing measures of well-being that economists would take seriously. The group developed and Krueger has extended the day reconstruction method, which may become a part of national statistics. The Gallup Poll also has made efforts to measure well-being worldwide. The Gallup-Healthways Well-Being Index samples 1,000 Americans per day with measures that include both affect and life satisfaction

aspects of well-being. A future need may be for high-frequency panels to look at well-being dynamics and the relationship between affect, life events, and health more often than once per year. This type of measurement will be the core of work on well-being in the future.

There are significant constraints in the ability of policies to contribute to individual well-being, and in some sense it is almost statistically impossible. There are huge individual differences in affect and life satisfaction. When the effects of an intervention are assessed using well-being measures, the individual differences are also measured. The changes that an intervention can accomplish are minute in comparison to the variance of individual differences in well-being. This does not mean that such interventions may not be worth doing, but the implications of these constraints must be considered and accounted for. One implication is that extremely large datasets will be needed in order to detect such small effects of an intervention on well-being relative to huge individual differences.

There are three classes of policy interventions with respect to attempts to improve well-being: policies that seek to improve the circumstances of people; policies that directly attempt to influence individual behavior to improve life satisfaction or affect; and policies that induce individual behavior that might indirectly improve well-being.

The terminology used to describe well-being is an issue because of its complexity and multidimensionality. Happiness does not equal well-being. People tend to emphasize a particular dimension of well-being. For example, there is a tendency to focus on increasing individuals' happiness or well-being, but perhaps policy (or particular policies) should be focused on reducing misery instead. Another example is the notion of meaning. When a person is asked what is important in life, he or she rarely mentions meaning. However, when meaning is mentioned and then a person is asked if it is important, it becomes the most important thing in the person's life. The research on meaningfulness is mostly correlational. A person who says life is meaningless is very likely clinically depressed. Focusing on meaningfulness will point us in very different directions in terms of improving life. The point is that the objectives of policy should not be slaves of language, and attention must be paid to what particular aspects are being focused on, measured, and factored into public policy.

There is a confounding of intensity and dimensionality in terms of emotional well-being. Measures of positive and negative affect are correlated -0.6 . Some people report a great deal of emotion, positive and negative, and others report very little emotion. This confounding of intensity and dimensionality has implications for international comparisons. For example, people discuss the "mystery" of the high levels of happiness found among South Americans. In reality, measures indicate that they are not only happier than others, but also more miserable than others too; they report *more* emotion in general than other people. This is an example of the pitfall of focusing only on one end of a dimension of well-being.

Positive and negative are not the only dimensions of affect; another dimension is activation and lethargy. The activation dimension includes different aspects of the low and high end: tense/stressed versus depressed/blue (low end) and joyous/thrilled versus serene/content (high end). Our current measures do not represent all of these dimensions. Sleep is critical to well-being and is on the activation dimension. Paying attention to sleep is important because it is

along that dimension that cultures differ most in terms of well-being. Attending to activation and evaluation and the interaction of the two is clearly one of the important tasks in this research. The main point is that we need to be clear about both the aspect and the pole within that aspect of well-being on which we want to focus.

Related to sleep is the notion of stress. Ed Diener (University of Illinois) and his collaborators have found a correlation of 0.4 between GDP and the level of stress reported across multiple countries. The United States has the fourth highest level of stress in the world according to Gallup Poll data. Across individuals, stress tends to be correlated with negative attributes. When negative affect is controlled, stress is also correlated with some positive attributes, such as income, education, being married, and having children. According to Gallup data, the best predictors of stress are having a headache or being alone the previous day, being a caregiver, having children, being a smoker, and being a college graduate. The distinction between good stress and bad stress must be defined if we are to fully understand well-being.

Different measures of life satisfaction will have different implications for the use of well-being data to inform policy. We have learned in recent years that if questions are asked about happiness, the answers tend to be modestly correlated with real-time measures of affect. At the other extreme there is the ladder of life question, which the Gallup poll uses for evaluating life satisfaction; it is the purest measure in terms of life evaluation. When Hadley Cantril formulated the ladder of life measure in 1965, he called it a self-anchoring scale.⁷ The wording on the scale indicates that the top rung is the best possible life and the bottom rung is the worst possible life for the respondent, relative to what is perceived as able to be accomplished. However, this is not how respondents necessarily understand the question. People have a concept of a “good life,” but it is not relative. The average response on the ladder question is 8 in Denmark and 3.2 in Togo, which is an enormous difference on a 10-point scale (someone with a 3.2 in Denmark might be considered clinically depressed). The concept of a “good life” appears to be very similar in Denmark and in Togo. The correlation between log GDP and the average ladder of life in a country is .94. This measure is an almost objective assessment of where people perceive they are in life compared to a social conventional definition of a “good life.” This is not what we normally think of as life satisfaction.

Life satisfaction is very much like the ladder, but it is slightly contaminated by affect. The ladder and affect have different predictors. The distinctive predictors of the ladder of life question are income and education. Education contributes nothing to affective well-being in the United States. In some senses, the ladder measure is redundant for the purposes of policy. The ladder measure will be optimized by improving things that can be more easily measured in different ways (e.g., income, professional status, level of education). According to Gallup data, the predictors of blue affect (to be worried and depressed) are being lonely, having a headache, being a caregiver, having a health condition, or having children at home. If we are going to use two families of measures, then we should have the ladder question, and measures of affect. The two types of measures we use should be as different as possible, and these two types of measures are expected to have different predictors.

⁷ Cantril, H. (1965). *The Pattern of Human Concern*. New Brunswick: Rutgers University Press.

Analysis of Gallup data of more than 450,000 respondents shows life evaluation rises steadily when plotted against log income.⁸ When log income is plotted against three measures of affect (positive affect, stress, and negative affect), there is no further improvement of affect beyond an income of \$75,000. The reason for this deserves further research. It is clear that people with high income can buy more pleasures, but a recent study in *Psychological Science* suggested that people with high income have a reduced ability to enjoy small pleasures and that might be the factor that evens things out.

We need to understand the essential distinction between the two types of well-being and address the policy question of which aspect of well-being should be the focus and which pole of that dimension (positive or negative) should policies target.

Paul Dolan

Professor, London School of Economics and Political Science

Dolan discussed three topic areas in an effort to set the conceptual landscape: a) concepts of well-being and contexts of policy; b) a comparison of subjective well-being measures; and c) concerns about measures and conclusions.

Defining concepts and contexts is extremely important in this discussion. We must place the measurement of subjective well-being in its appropriate policy context and compare it to other routinely used measures. There are three broad kinds of relevant measures: measures of objective circumstances (e.g., health status, literacy), preference satisfaction (e.g., income), and mental states, which include measures of subjective well-being. The policy context is very important; often we discuss measures of well-being without being clear about the purpose to which those measures will be put. We will want different measures for different purposes with different levels of sensitivity and specificity. The three policy contexts are: monitoring national progress, informing policy design, and policy appraisal.

The Human Development Index, which crudely measures life expectancy, health, and education, is an example of a measure of objective circumstances of well-being for the purpose of monitoring national progress (e.g., the fraction of the population that can read and write). GDP is used to monitor national income. It might be satisfactory to have some high-level measures such as ladder of life and/or life satisfaction to measure subjective well-being for the purpose of national monitoring. For the purposes of informing policy design and the evaluation of policy interventions (e.g., determining the impact of spending more money on one health service versus another) we need measures that are more detailed and that allow the expression of benefits in terms of monetary values so that they can be used in cost-benefit models.

There are three broad types of measures of subjective well-being: evaluation, experience, and eudemonic. Evaluation measures include the ladder of life, life satisfaction, domain satisfaction, and thoughts about domains. Experience measures include day affect, day reconstruction method affect, and thoughts about an experience. Eudemonic measures include overall purpose of life,

⁸ Kahneman, D. & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academies of Sciences*, n/a. doi 10.1073/pnas.1011492107. Retrieved March 16, 2011, from <http://www.pnas.org/content/early/2010/08/27/1011492107>.

day reward, and day reconstruction method reward. A “eudemonic” well-being measure is an individual’s assessment of how much meaning or purpose of engagement he/she has in his/her life overall or with certain activities throughout the day.

Dolan presented analyses of German data.⁹ The data show the pleasurable nature of particular activities relative to other activities. For example, going to work brings slightly less pleasure than housework. Participants also were asked the extent to which these activities contributed toward the well-being of others or toward important goals, and the extent to which the activities were worthwhile (i.e., they were asked about the activities in a more eudemonic sense), which resulted in some interesting differences. For example, time spent with children is more meaningful or rewarding relative to other activities, despite it sometimes not being very pleasurable (e.g., changing diapers). These data are correlational and no claims are made about causality or how individuals maximize utility over the course of the day or their lives. These findings are relevant to some of the conversations happening in the United Kingdom about measuring the value of the benefits associated with a Big Society.¹⁰ It may be that we have to think about items that measure a more eudemonic sense of welfare in order to capture societal benefits.

Another aspect of well-being that has not been measured very well to date is thoughts about domains. These are not just the experience associated with activities or global assessments of life satisfaction in a generic overall sense, but the thoughts that arise throughout the day. Data were collected on thoughts about domains from 1,200 Americans in a study the NIA helped to start. Participants were asked about their thoughts during the previous day. Results indicated that 27 percent of people said they thought about their finances many times or continually. It is not obvious that existing measures of well-being capture this information very well. Thoughts about finances are more strongly associated with income than life satisfaction or than affect data are associated with income.

The study also found that 16 percent of people thought about their health many times or continually. Dolan looked at individuals’ willingness to trade life years in order to improve their current health conditions.¹¹ Individuals’ thoughts about health were stronger predictors of their willingness to give up life years than the actual health circumstances. We need additional research on this.

Dolan presented findings from data gathered from a large U.S. sample using the ladder of life, day affect, and willingness to trade life years for improved current health plotted against five health circumstances taken from the EQ-5D.¹² The health circumstances were: some problems walking about, some problems washing or dressing, some problems performing usual activities, moderate pain or discomfort, and moderate anxiety or depression. The coefficients are the relative decrement in the well-being measures associated with each of those dimensions where 1

⁹ White, M. P. & Dolan, P. (2009). Accounting for the richness of daily activities. *Psychological Science*, 5, 327-334. Retrieved March 16, 2011, from <http://pss.sagepub.com/content/20/8/1000.abstract>.

¹⁰ Prime Minister David Cameron launched a plan for a Big Society Programme in May 2010 for public service reform. More information can be found at <http://www.cabinetoffice.gov.uk/news/big-society-heart-public-sector-reform> (accessed March 16, 2011).

¹¹ Dolan, P. (2010). Thinking about it: Thoughts about health and valuing QALYs. *Health Economics*, n/a. doi 10.1002/hec. 1679. Retrieved March 16, 2011, from <http://onlinelibrary.wiley.com/doi/10.1002/hec.1679/abstract>.

¹² <http://www.euroqol.org/>.

represents the best possible health and 0 represents the worst. Analyses indicate that the decrement associated with “some problems walking about” using the ladder of life is slightly under .05 whereas the public’s perception of some problems walking about is 3 times larger. Indeed, the data indicate that the physical functioning conditions are worse in our imaginations than they are in our experiences. Moderate pain or discomfort has a smaller effect on well-being measures than we might believe it to have on ladder of life and day affect measures. A question that requires further research is whether or not this represents the real effect of pain or if it is being picked up by other dimensions.

The most interesting dimension is moderate anxiety or depression. The ladder of life loss associated with anxiety or depression is .25. We would have predicted that moderate anxiety or depression would have had more of an effect on day affect than on ladder of life (note that the sample in this study is not nearly as large as that of the Gallup poll).

In general, most of the time, things are much worse in our imaginations than in our experiences. We forecast things to be terrible and stay terrible most of the time, yet things are typically not that bad, and most of the time they get better. The anxiety and depression dimension does not appear to improve with the passing of time, relative to physical dimensions. This is an important finding for policy purposes. Our policy decisions and much of our economic cost-benefit models are based on people’s preferences expressed in willingness to pay in contingent valuation studies or in willingness to give up life years, and, in a nutshell, we get it wrong in terms of the actual impact on our real experiences.

There are several concerns to consider about the measures themselves. Salience is about where the attention is focused at the time of assessment; issues include context (i.e., question ordering), priming (i.e., the messenger), attention direction, and time frame. Where the attention is focused at the time of the assessment, whether the person is aware of how their attention is being focused, or being primed to think about things in certain ways (question ordering, context) will all have an impact on how the measures perform. We know from a number of studies that context and question ordering matters. Subjective well-being questions should ideally come at the beginning of a survey rather than be primed by other questions, but the ordering within the subjective well-being questions is a matter for further discussion. The messenger of the survey (i.e., academic versus government) also may have an effect on answers; there is some evidence in some contexts that if the messenger is a health survey organization then health satisfaction loads more heavily on life satisfaction responses.

Scaling effects include adaptation (i.e., shifting domain weights) and endpoints (i.e., change over time). With adaptation we have a theoretical model and some data to populate it. For example, when a person gains weight, especially if it happens quickly, he or she can expend effort to lose weight or expend effort to shift the importance of different domains of life so that the ones associated with weight become less important in the person’s overall life satisfaction. There is some evidence now to support this shift in domain weights. We need further work on adaptation and domain weight shifting

There are two main concerns about using subjective well-being measures for policy appraisal purposes: selection effects and spillover effects. We do not know very much about those who are

not included in the survey. It is possible that the type of person who participates in the survey somehow systematically differs from those who do not participate, which is not the same concept as the representativeness of the sample. People select into certain circumstances and into surveys, possibly based on some observable characteristic or trait, and we need to know more about this potential selection bias. With regard to spillover effects, we need to better understand network effects and the effect policy has on people who are not necessarily the target of the policy. If we are interested in interventions that are meant to impact the Big Society, then we want to intervene in ways that will engage the population as effectively as possible while having network and spillover effects.

Any measure used to inform public policy needs to be theoretically rigorous. We will not be able to resolve the question of which measures are the best ones; we ought to use all of them to assess how different populations fare. Opportunities for policy interventions arise when we observe people performing differently on these different measures. In addition to being theoretically rigorous and empirically robust, the measures must be policy relevant. To be policy relevant we need to find out much more than we now know about the expectations and preferences of stakeholders, citizens, policy makers, and consumers with respect to using subjective well-being measures to inform policy. The Green Book, the U.K. government's primary guidance document for economic appraisal in the public sector, is not used as often as it should be.¹³ Where it is used, we can do a better job by augmenting conventional measures with more evidence on direct experience of interventions on people's well-being.

There are different measures for different purposes, and they need to be selected with that in mind. Policy appraisal, as opposed to monitoring national progress, requires different levels of sensitivity, specificity, more frequent assessment in panel datasets, and good field experimentation. There are opportunities right now with numerous policy interventions in the United Kingdom that are not currently being evaluated but are amenable to experimental design.

Discussants

Jacqui Smith

Professor of Psychology and Research Professor, University of Michigan, Ann Arbor

There are different life phases that we might consider from a policy perspective. Old age is an interesting phase including multiple segments: older workers, people who were retired but going back to work, people who are retired, and those who are frail and moving toward the end of life.

Measures of global or evaluated well-being have been included in the Health and Retirement Study (HRS)¹⁴ for quite some time. A recent analyses of HRS data on the over-50 population determined that survey measures of global evaluated well-being (life satisfaction, positive affect, and negative affect) are reliable, show valid relationships to the central outcomes for this population (health, disability, stress of older workers, retiree lifestyle, financial difficulties, burden of caregiving), and have associations with age cohorts and social gradient. These measures also predict survival and are sensitive to change over time. However, response biases

¹³ http://www.hm-treasury.gov.uk/data_greenbook_index.htm.

¹⁴ <http://hrsonline.isr.umich.edu/>.

complicate subgroup and cross-national comparisons. People have different connotations of what it means to be satisfied, they disclose satisfaction information differently, or they require different standards to be satisfied. This is not a problem for assessing intra-individual change, but it makes difficult comparisons with other panels or nations. Variations in measure metrics, wording, and item context are also challenges (e.g., 10-point versus 7-point scale, labeling of items).

HRS data from 2008 show that people troubled by pain or in poor health are less likely to report being satisfied with life. There are age cohort trends in life satisfaction as measured on the Diener scale of five items. Those in their 50s score lower compared to those in their 70s. We do not yet know if those in their 50s now will score higher in life satisfaction once reaching their 70s or if they will continue to score lower in life satisfaction. In the 80+ population, there is much more variation within each age group and within different levels of frailty.

In 2009, the HRS and RAND Corporation conducted an Internet study of the effects of the economic crunch using a single item to measure life satisfaction to model intra-individual change over time. There is a large dip in life satisfaction in 2009 and a rebound in 2010 but not to the same level, demonstrating a period effect. This effect was true regardless of age and education, with the exception that people in the highest quintile of income did not show a dip. This data will be finalized in 2010, and more in depth analyses will be conducted.

A comparison of the English Longitudinal Study of Ageing (ELSA) and the HRS, also using the Diener scale, demonstrates that in general, the English are more satisfied with life in all age groups. However, the country differences actually disappear when response bias (based on item response analysis), item order effects, the probability of choosing each answer, and age differences are all statistically controlled for.

There are challenges to adding measures of experienced well-being to longitudinal panel surveys. The day reconstruction method questions are very long (45 to 75 minutes), and we do not yet know the essential features of a shorter measure. We need to determine which positive and negative affective experiences are important. The benefits of experienced happiness may differ depending on the activity (e.g., drinking beer and watching television versus volunteering). These distinctions will matter for determining the benefits of well-being for the purpose of policy. Additional questions that remain include: a) When is experienced well-being distinct theoretically from evaluated well-being? b) What is the best metric of experienced well-being? and c) Is non-hedonic experience more important than hedonic? Pilot survey measures of experienced well-being are currently being collected in the HRS, the Panel Study of Income Dynamics,¹⁵ the American Time Use Survey,¹⁶ and the American Life Panel.¹⁷

¹⁵ <http://psidonline.isr.umich.edu/>.

¹⁶ <http://www.bls.gov/tus/>.

¹⁷ http://www.rand.org/labor/roybalfd/american_life.html.

Amanda Rowlatt

Chief Analyst and Director for Child Poverty and Analysis, U.K. Department for Work and Pensions

There are two main items of interest from a policy perspective. First, the public debate needs to move from a focus on money and GDP to a focus on well-being. We need aggregate measures, informative break downs, more of the academic work summarized and in the press, and a real public debate about what matters. Second, policy has to be based on proper cost-benefit analysis, and there is not enough in the Green Book on the social impact on well-being in order to factor well-being into cost-benefit analysis models.

There is currently very strong political and institutional support for measuring well-being in the United Kingdom. In particular, the budget statement charged the ONS with developing national statistics measuring well-being and stimulating a public debate. The debate will begin with a high profile conference with the Prime Minister on November 25, 2010. There is an institutional framework for this work in addition to current political support. A Cabinet committee on social justice is chaired by the Secretary of State for Work and Pensions, Iain Duncan Smith. Smith's main passion is to significantly improve the assessment of policy to ensure the greatest social impact, or as he terms it, social return on investment. His goal is to use the committee to drive politicians to justify choices on the basis of cost-benefit analysis. There is also a social impact task force with high-level analysts from all the main government departments, chaired jointly by Rowlatt and the chief analyst from the Department for Environment, Food and Rural Affairs, to develop additions to the Green Book that will better guide evaluation of social impact. There is pressure to complete this work quickly, and the task force will need to partner with the academic sector and policy makers in the United States to implement these changes.

Three areas specifically linked to the Department for Work and Pensions (DWP) should be explored. First, we need to determine the value for the well-being impact of work over and above the financial impact. Figures must be robust and credible with politicians and useful in cost-benefit analysis. Second, we should investigate the well-being impact on children of having a parent at work. Third, we need to think about the well-being impact of financial smoothing over life through pensions or the role of working age welfare payments. Major gaps include the value to individuals of being a part of a community and the well-being value of trust. There are many questions, but also a driving desire for action now, because there is a sense of urgency given the favorable political atmosphere in the United Kingdom.

Open Discussion

Budget Constraints

The consequences of the budget cuts in the United Kingdom are still being assessed and discussed. David Halpern (Institute for Government, London, United Kingdom) and Rowlatt noted that well-being measures are being added into existing surveys. In the future, it would be ideal if well-being measures could help to inform decisions about budget cuts. With the current political focus on Big Society and the Prime Minister's specific interest in well-being, the

atmosphere in the United Kingdom is quite ambitious despite the economic climate. The ONS has the needed resources to move forward with this.

In the United States the President's budget proposes a significant increase for the Bureau of Economic Analysis (BEA) to begin some of the work to include distribution of income information and some notions of sustainability in GDP estimates. BEA is starting with financial sustainability and investment sustainability to build a platform of better measures of sustainability (e.g., energy data and its relationship to emissions and environmental sustainability). Other opportunities include: a) pairing the American Time Use Survey with work on the new BLS module to quantitatively value leisure and time use; and b) work at the BEA supported by the NIA to develop satellite health accounts. These are examples where common ground exists in terms of measures of well-being and the extensions of measures to subjective well-being.

Measurement Issues

There was discussion about the adequacy and purpose of a short version of a measure to assess affect. Gallup asks a few questions about how the respondent felt yesterday; the result is that basically the same conclusions have emerged from more detailed study and less detailed study in terms of the basic contrast between life evaluation and affect, indicating that we can recover a lot of information from simple interrogation of huge samples. The best single indicator of positive affect is a response to one question: how much time did you spend with people you like yesterday? The NIA has provided funding to Arthur Stone (Stony Brook University) for the creation of a 3-minute version of a measure that can be used in surveys. However, if we are trying to maximize people's engagement in rewarding experiences in order to promote some kind of life-course profile of meaning maximization, then we need to consider if a 3-minute measure will be satisfactory to assess this dimension of people's experiences from a measurement perspective. The two different measures (experience and evaluation) are correlated, but not substitutable. Initial measures may require decades of methodological refinement.

Steven Landefeld (BEA) and Robert Groves (U.S. Bureau of the Census) expressed concern about efforts to standardize measures too early in the process. Monitoring national progress is one purpose of well-being measures, which is attractive. However, in this particular measurement domain it is attractive only with auxiliary ongoing research that is based on rich, frequent longitudinal surveys and field experiments done in the academic sector. The problem for the governmental statistical agencies in the United States is how to mount an experimental measure that will change over time by design and keep it from becoming ossified while the academic sector is learning more about the measurement properties. It would be risky at this point to intervene with public policy actions using measures that are not fully understood. Results of such an intervention might be considered unexpected when the situation is really that the outcome measures are not understood. A preferable structure might be monitoring with multiple government surveys using a smaller module in partnership with academic research that is feeding new measures refined from use in a longitudinal survey.

Purpose

Most agreed that different policy purposes will require different measures. The Gallup poll measures are quite effective. The sophistication of the measure will vary by policy domains depending on the purpose. Andrew Steptoe (University College London) offered that perhaps a graded system of measures is needed across domains in order to fully capture different purposes. Kahneman contended that government policy will likely have very little effect on people's measureable well-being on a global scale. Well-being measures can be very useful for targeted populations and targeted interventions (e.g., retirement policy impact on retired population). It may be true that the government does want to satisfy people, but that does not mean that the best measure of how it satisfies people is to increase their life satisfaction. The most important characteristic of a measure is that it captures changes in a credible way; if that type of subjective well-being measure does not exist for the population as a whole, then we should change our objectives. Dolan agreed that it is crucial to measure changes in well-being in a credible way. However, there may be cases where the well-being measures are not moving in response to an intervention not because the measure is inadequate, but because the intervention is not effective. Sometimes well-being measures will give counterintuitive findings.

From a policy perspective, it is important to look at the entire distribution of well-being. The epidemiological evidence shows that if we treat people with disorders, then we may help those individuals, but we will never reduce the incidence of the disorder. Therefore, we need to think in terms of a population shift intervention. In order to have a population shift, we need to understand what we are shifting it toward, in other words, the characteristics of the top end of the distribution. Any common condition (e.g., mental health, substance abuse, poverty) can be addressed by small population shifts toward the higher end of the distribution. Although it is important to consider the entire distribution, all people are not necessarily accorded the same weight at all points of the distribution. It might very well be appropriate to focus on reducing the misery of those at the low end rather than on increasing the happiness of others.

Focus on Small Areas

Participants discussed the importance of focusing on smaller geographic areas rather than only the national level. At least for certain measures, people's assessment of their well-being is relative. In large national surveys, such as the ones in which we are talking about adding well-being measures, we do not know a lot about the respondents' reference groups. The National Science Foundation is interested in setting up social observatories using smaller areas to collect rich data in order to better understand some of the influencing factors that may be missing from larger surveys. People's networks may be geographically based, and that may play into a person's comparison reference.

Locally focused studies that might be coordinated into sets of studies in different contexts are an important strategy to consider, as context is important for the way the indicators are understood. It might be incorrect to assume when developing a policy that things will work similarly with different populations in different locations. We need to consider whether a nationally representative sample will reflect local differences. Also, the idea of sustainability may

necessitate some broader indicators of well-being in communities as a whole, beyond individual measures of well-being.

Public Acceptability

Some participants contended that public acceptability needs to be a very important criterion of a well-being measure. It needs to be a measure that the public understands and that corresponds to commonsense wording, rather than a complex index based on several questions. Ideally, we want well-being to be like temperature if we want it to be taken seriously. People more routinely think about reward and meaning and the purpose of activities rather than a generic global assessment of meaning. We need to be cautious about well-being measures and their usefulness in every single context. There will be trade-offs between the things that policy makers and the social welfare function cares about. Sustainability issues may not always go hand in hand with individual well-being.

Well-Being, Cost-Benefit Analysis, and the Life Course

Richard Layard

Emeritus Professor of Economics, London School of Economics and Political Science

For well-being data to be useful to public policy, we need to not only measure well-being but also have some idea of what is causing it and how it can be changed by interventions. The goal is to be able to use well-being data in cost-benefit analysis. There needs to be a great deal more basic research on the causality of well-being (i.e., the coefficients and model that explain well-being) before we can adequately model it as an outcome measure for cost-benefit analysis.

Well-being can be framed in terms of public economics. The maximand is the sum over future time of individual i 's well-being, aggregated and weighted across individuals (i.e., more weight on lifting well-being when it is very low as opposed to already high). The government's role is to use policies to try to maximize this maximand, subject to all constraints. The function that determines the well-being of individual i at time t is what we need to know a lot more about (i.e., what is in that function that determines well-being). The main variables in the function are circumstances at time t , but also things that have happened before, genes, and other parts of the life course up to time t .

Let W_{it} be well-being of person i at time t

Let maximand be $\sum (\sum W_{it})$

Government chooses policies to maximize this, subject to all constraints, including

$W_{it} = f(\text{Circumstances}_{it}, \text{Genes}_i, \text{Background}_i, \text{Life-course}_i)$

We need to define well-being, determine the contents of the function, understand the role of mental health over the life course, and conduct cost-benefit analysis with well-being as the outcome. For policy purposes the definition of well-being must be acceptable to the public, well-studied, reasonably responsive to policy variables, easily described, and close to the quality of experience. We need to understand the performance capabilities of particular measures of well-being: life satisfaction, degree of happiness, affect yesterday, flourishing, and the ladder of life.

The function of well-being should include current circumstances (i.e., income, health, employment) and background (i.e., genes, earlier experiences). Current circumstances are important because this is where policy can intervene. Background helps to determine what the person is like (i.e., how happy a person is in any given set of circumstances). A huge amount of variation among individuals is probably not due to circumstances but to the ability people have to enjoy their circumstances. Also, the background is important in helping to explain how people have come to have their circumstances. We seem to have a better handle on the role of contemporaneous circumstances but have not adequately considered how background circumstances, previous experiences, and states affect well-being later in the life course.

Mental health often has been neglected, but it is very important to include in the function. Sample data from the British Cohort Study¹⁸ demonstrate the importance of mental health in the function of well-being. Standardized regression coefficients (i.e., partial correlations) demonstrate that emotional disorders at younger ages predict life satisfaction at age 34, in some cases, more so than income. For example, emotional disorder at age 26 explained four times more of the variation of life satisfaction at 34 than income; emotional disorder at age 16 explains as much of life satisfaction at 34 as income. These data demonstrate that more attention should be accorded to mental health as a background variable.

To look at the effects of all the main contributory variables, we need a proper path model of the life course including variables that measure state: emotional well-being (outcome variable), conduct/criminality, physical health, and cognitive/educational status. When a person becomes an adult, we are interested also in employment, income, welfare dependence, and his/her own parenting activity. In principle, any one of these state variables at time t could be explained by all the state variables in previous periods, the background variables, and the person's genes. We should be able to add events to the path model as well.

Layard presented four different regression analyses using data from the British Cohort Study with life satisfaction, malaise, earnings, and criminality as the dependent variables and many childhood and background controls. The life satisfaction equation shows that there is an effect when adult life satisfaction is regressed on previous childhood mental health problems. Emotional problems at age 5 reduced life satisfaction at age 34 by 0.05 standard deviations, and conduct problems at age 5 reduced life satisfaction at age 34 by 0.08 standard deviations. The criminality equation shows that conduct problems at age 5 influence earnings by 0.06 standard deviations and criminality by 0.20 standard deviations.

The coefficients attributed to emotional or conduct problems at age 5 can then be used in the evaluation of policies. For example, we can use these coefficients to determine the implied effects of a child mental health treatment program called the Incredible Years Programme for 6-year-olds.¹⁹ The program provides training to parents of children with conduct disorders at a cost

¹⁸ <http://www.cls.ioe.ac.uk/studies.asp?section=000100020002>.

¹⁹ Scott, S., Sylva, K., Doolan, M., Price, J., Jacobs, B., Crook, C. et al. (2009). Randomised controlled trial of parent groups for child antisocial behaviour targeting multiple risk factors: The SPOKES project. *Journal of Child Psychology and Psychiatry*, n/a. doi:10.1111/j.1469-7610.2009.02127.x. Retrieved March 16, 2011, from <http://www.incredibleyears.com/library/items/SPOKES-JCPP-2010.pdf>.

of £2,500 per child. The program has been shown to reduce parent-rated conduct disorder by 0.5 standard deviations. The implied effect for the 6-year-old's future adult life satisfaction using the coefficients from the previous regression analyses would be 0.5 (0.08) standard deviations for adult mental health (i.e., benefit to well-being) and 0.5 (0.20) standard deviations for adult criminality. From this we can calculate mental health gain per unit of net cost.

The fundamental problem in cohort studies is unmeasured personal factors and the persistence of problems throughout a person's life, such as conduct. It is unclear if there is a causal relationship between conduct problems and criminality or if the persistence of conduct problems and criminality are both simply reflecting something about the person that was there all the time. Twin data would help in addressing the causal problem and explain subsequent persistence. Koning et al. has done regression analyses before and after including the fixed effect of the twin relationship and showed that the effect was halved. Layard and his colleagues are hoping to do that kind of work using Robert Plomin's (King's College London) data from the Twins Early Development Study²⁰ and the Norwegian and Swedish panels.

We need a method for cost-benefit analysis with a function of well-being as the outcome to measure savings through a path model. Valuing benefits by willingness to pay does not make sense for many policy outcomes (e.g., health, parenting, crime, community cohesion, local environment). We need a parallel Green Book with well-being effects as the outcome. However, there are some pitfalls to this approach. There are fallacies of composition when starting from individual data and in analyzing income outcomes. Nevertheless, a key aim of current social science should be the painstaking attempt to estimate these causal models.

Discussant

Amanda Sacker

Professor in Quantitative Social Science, Institute for Social and Economic Research, University of Essex, United Kingdom

Layard's presentation put the focus on a different dimension of well-being—the life course. In the field of social epidemiology, models for thinking about the life course deal with identifying critical periods, pathways to endpoints, and the accumulation of risk. It is difficult to disentangle these three elements. We can demonstrate linkages through time, but a person's life course is also an accumulation of effects in all different spheres in different domains; we need to be aware of how domains are correlated in one point of time and over time to ensure the right model. A simple pathways model might not capture all of life's experiences.

Public policies may involve intervention (e.g., treatment for children with conduct disorder) or prevention (e.g., encouraging breastfeeding because breastfed babies have resilience to stress later on in childhood). There could be policies targeted at an early stage of the life course that are designed to prevent later stages of poor well-being, such as Layard described in his examples.

²⁰ <http://www.iop.kcl.ac.uk/departments/?locator=336>.

Pathways can be different for people in different social circumstances, and the relationship between health and well-being will depend on those circumstances. For example, people in more advantaged circumstances with health problems will show a dip in well-being initially but can bounce back; the ability to bounce back is not seen among people in disadvantaged circumstances. Data from Understanding Society,²¹ a new large U.K. panel study that has measures on well-being, psychological functioning, and life satisfaction, will enable us to look at many of these issues. An innovation panel component of Understanding Society also will serve as a platform for developing new measures of well-being. There are difficulties with cross-national comparisons because people from different cultures may evaluate their well-being differently. Likewise, people of different ages and circumstances may evaluate their well-being differently. We need more research on all of these things before we can make concrete assumptions about a path model.

Open Discussion

Life Course Issues

Many of the findings concerning pathways from longitudinal studies need to be checked through experimental manipulation of some kind. The NIA has an interest in combining longitudinal panel studies and experiments, when possible, either through natural (e.g., Medicare Part D) or randomized experiments (e.g., Oregon health study lottery). Also, one longitudinal panel funded by the NIA is now in the process of genotyping; perhaps this information will help answer the question about the extent to which some of these things are thermostatically set at birth or through some form of expression later on in the life course.

Insofar as people's current well-being depends on things that have happened in the past, experiments would need to have a very long follow-up, which is impractical for all the interventions we might consider. This constraint provides an argument for using causal models from longitudinal panels instead to enable us to estimate the long-term effect of short-term changes. In some sense, the major issue is the rate of fading of effects. Another piece of evidence would be to compare the fading of effects in experimental studies where there is a long-term follow-up period to the rate of fading in estimates from longitudinal studies.

We must not neglect the enormous impact of circumstances and contexts on individual well-being and individual behavior when thinking about the life course. The day reconstruction method has helped us to think about analyzing data differently. It is not only about people but also about activities and the circumstances people find themselves in.

Stability of Well-Being Measure

Participants returned to the topic of having a national indicator of well-being that changes. Panel studies have enabled researchers to trace the decline of stability over time in the life satisfaction measures; the stability declines more quickly than with personality measures, indicating that they are more changeable than personality measures. The stability of national levels is important, as

²¹ <http://www.understandingsociety.org.uk/>.

are differences within nations which merit greater attention. There are some data that suggest real differences between counties, regions, and states; there seems to be a disconnect between what happens at the national level where we do not see change over time and what happens in smaller geographical areas. These differences may be linked to geographic characteristics, which gives us hope that something is happening differently in smaller regions that might be amenable to policy interventions.

Paul Allin (ONS) discussed the current plans for using the U.K.'s Integrated Household Survey of 450,000 people randomly selected per year in 450 local authority areas (i.e., 1,000 people per local authority area). Well-being measures will be included in the survey from April 2011 and will allow for analysis at the national and subnational level. Headline results also will be available at the local level. There is a tension between going firm on questions as soon as possible while also wanting to evolve the questions and ensure that they meet our needs; this is an unresolved conflict. It will be some time before a standard harmonized approach is achieved.

Policy Relevance

Gallup poll data show that people who do not have health insurance and have a health problem worry more than those without health insurance but do not have a current health problem. Kahneman advocated for a focus on interactions such as this because that is where policy can have an impact. The implication methodologically is that we need very large samples in order to look at the interactions.

Halpern discussed the relevance of social policy evaluations in general and the difficulty in conducting them. There needs to be routine evaluation of social policies and decisions based on the evaluation results. Including well-being in those evaluations is just an additional component. The current political atmosphere in the United Kingdom will raise the sense of urgency to bring subjective well-being measures to scale. Public debate about how to assemble the measures in terms of trade-offs across the life span and identifying key policy priorities can be ongoing.

Rowlatt reflected on the implications of Layard's presentation on valuing future mental health gains and negative societal impacts (e.g., criminality). Translating this type of evaluation to policy decisions would mean reallocating funding from justice programs to mental health programs. These potential policy decisions need a range of cost-benefit analysis and debate.

Richard Frank (HHS) discussed the distinction between well-being and the outcomes of focus when mental health treatments were developed (i.e., reduction of symptoms). Treatments are not necessarily designed to improve overall well-being; rather treatments are designed to reduce symptoms of mental health conditions. There are implications for measuring the success of a mental health treatment by measuring well-being if improving well-being was not the purpose of the treatment. For example, studies of depression consistently show strong and large effects on symptoms for evidence-based treatment, but much smaller effects on work. When evaluating treatments of mental health, it is important to look at the production function of both clinical outcomes and well-being outcomes because they are not the same thing.

Layard commented that the argument for having measures taken with great frequency when they are not longitudinal is public relations. Having unemployment figures every month has a huge effect on the importance of unemployment as a policy target. Having high-frequency data will impact the policy salience of the issue. Layard speculated that since GDP reporting began, the rate of change in the GDP has become a lot higher, which might indicate an effect of reporting GDP.

Well-Being: Health and Medical Care

Andrew Steptoe

Professor of Psychology and Department Head, Epidemiology and Public Health, University College London

Two relevant issues are the importance of well-being in the maintenance of health and prevention of illness and an evaluation of well-being among people with health problems. There has been a focus in recent years on the development of physical illness and the role that negative states play (e.g., depression as a precursor to illness) and various measures of positive states including affective and evaluative measures. There is a long literature on depression as a precursor to illness, but there is less research on positive states and resilience.

Observational epidemiology is the basic method used in health research to study these topics. This approach involves measuring health outcomes among a large cohort of healthy people, documenting exposure to the factor of interest, tracking people prospectively for mortality and morbidity, and conducting multivariate analysis of predictors to determine if exposure is related to outcome. Using data on positive affect and incidence of coronary heart disease from a cohort in Nova Scotia of 1,200 initially healthy adults, analyses show that a positive affect at baseline is associated with reduced coronary heart disease risk over 10 years.²² Depressive symptoms are associated with an increased risk of coronary heart disease. Factors such as body mass index, history of diabetes, cholesterol, and smoking status were controlled for. These results support the argument that affective states are independent predictors of coronary heart disease. These are simultaneous models; positive affect is protective in the same model that depressive symptoms are increasing risk, indicating that these are two independent effects.

There are problem with this observational approach. There could be confounding (i.e., an unaccounted for factor may be responsible for these associations) despite attempts to control for other variables. Other problems include reverse causality (i.e., illness changing well-being as opposed to well-being contributing to illness) and the possibility that negative affect is independent of positive well-being. A meta-analysis of 20 large-scale prospective observational studies showed overall that positive affect in various measures is associated with reduced mortality when negative affect is controlled for and when a number of covariates are included.²³ There is a certain amount of evidence of an association.

²² Davidson, K. W., Mostofsky, E., & Whang, W. (2010). Don't worry, be happy: Positive affect and reduced 10-year incident coronary heart disease: The Canadian Nova Scotia Health Survey. *European Heart Journal*, *31*, 1065-1070.

²³ Chida, Y. & Steptoe, S. (2008). Positive psychological well-being and mortality: A quantitative review of prospective observational studies. *Psychosomatic Medicine*, *70*, 741-756.

Another approach to looking at the influence of well-being on illness is to look at the biological correlates of well-being. Biology is an important element because it provides mechanistic evidence for processes linking well-being with health outcomes, and biological data are often regarded as more convincing than self-report data. The other perspective is to look at the extent to which these biological markers could be used as actual measures of well-being. The problem is that there are many other factors that influence these biomarkers, which make it difficult to use them as objective measures to prove or disprove self-report differences in well-being.

There are several examples of biomarkers. First, the best established biological marker associated with positive well-being is cortisol. It is difficult to measure because it varies over the course of the day. Potentially damaging effects of high levels of cortisol include increased low-density lipoproteins (LDL) cholesterol in the blood, suppression of the immune function, decalcification of bones, increased abdominal fat, damage to the hippocampus, muscle wasting, and impaired reproductive function. A study that measured salivary cortisol throughout the day demonstrated that higher levels of positive affect, measured by ecological momentary assessments, are associated with lower levels of cortisol when controlling for several covariates of both variables.²⁴ A number of observational studies have shown this pattern. Using measures developed by Cohen to establish a baseline of emotional style prior to attending a laboratory session, another study demonstrated that people with higher positive emotional style do not have elevated cortisol in response to a stressor when compared to people who are less happy.²⁵

The second biological correlate of well-being is cardiovascular function, although the findings are mixed due to measurement problems. Blood pressure tends to be lower in happier people, but there have been findings of the reverse as well. Blood pressure represents a measurement problem because there is huge variation in blood pressure throughout the day, which makes it very difficult to use blood pressure as an indicator related to affective state. Another measure of cardiovascular function of interest is heart rate variability. It is an interesting marker because it reflects balance between two parts of an autonomic or vegetative nervous system, one part stimulating an acceleration of heart rate and another part stimulating a slowing of heart rate. People with higher levels of heart rate variability have greater control of the parasympathetic nervous system. There is now growing evidence that positive affect is associated with higher levels of heart rate variability.²⁶ Positive well-being is associated with more rapid post-stress recovery and healthier heart rate variability, but there is mixed evidence on blood pressure.

The third biological correlate of well-being is inflammation. Markers of inflammation include C-reactive protein, interleukin 6, fibrinogen, and tumor necrosis factor α . Studies of these markers are somewhat limited because they need to be measured in the blood. In an example using ecological momentary assessment measures with Whitehall 2 cohort data, analyses showed greater inflammation (as measured by interleukin 6) among less happy women when controlling

²⁴ Steptoe, A., O'Donnell, K., Badrick, E., Kumari, M., & Marmot, M. (2008). Neuroendocrine and inflammatory factors associated with positive affect in healthy men and women. *American Journal of Epidemiology*, 167, 96-102.

²⁵ Bostock, S., Hamer, M., Wawrzyniak, A.J., Mitchell, E.S., and Steptoe, A. (in press). Positive emotional style and subjective, cardiovascular and cortisol responses to acute laboratory stress. *Psychoneuroendocrinology*,

²⁶ Bhattacharyya, M. R., Whitehead, D. L., Rakhit, R., & Steptoe, A. (2008). Depressed mood, positive affect, and heart rate variability in patients with suspected coronary artery disease. *Psychosomatic Medicine*, 70, 1020-1027.

for depression and other factors. An analysis of fibrinogen and an evaluative measure of well-being (CASP measure: control, autonomy, self-realization, pleasure) in ELSA data showed women with higher levels of well-being had lower levels of fibrinogen. Overall, there is mixed evidence for inflammation.

There is a great deal of literature on health-related quality of life with both generic measures and more disease-specific measures. Measuring the functional effects of therapy on a person as perceived by the person is not the same as measuring general well-being that includes other aspects of life or positive affect. There is a question of whether or not general well-being and affect also are impaired in people with health problems. Cross-sectional ELSA data showed that people with different specific health conditions have a wide variation in impairment of general quality of life (as measured by CASP), but it is not the same sort of variation you might expect from health-related quality of life measures (e.g., those with cancer scored higher on CASP than those with diabetes). Measures of happiness also show variation by specific health conditions but not in the same pattern as for general quality of life.

In order to explore this finding further, longitudinal data were used to compare general quality of life between pre- and post-illness in the case of specific diseases. Those with the disease had lower quality of life than the healthy controls. Also, people who developed stroke or coronary heart disease had greater initial impaired general quality of life compared to the controls. Another analysis looked at the happiness among people who experience coronary heart disease at different points after the event, which can be used to look at the extent to which happiness is an independent factor from conditions like depression. Even if a person is not depressed, there is a wide range of happiness. An analysis of how happiness levels change throughout the day among people who had acute coronary syndrome 3 weeks prior reveals that the people who are not depressed have more stable ratings of happiness. There is more variation among those who are depressed; however, the depressed people do have moments of happiness indicating the possibly that particular activities are important factors.

General well-being and positive affect are impaired in people with chronic health problems, but there is surprisingly little data showing that. The absence of depression and anxiety does not necessarily imply well-being. Concurrent activities are critically relevant to experienced well-being and need to be taken into account in understanding the links between well-being and health outcomes.

Discussants

Susan Cartwright

*Professor of Organizational Psychology and Well-Being, School of Health and Medicine,
Lancaster University*

It is clear that the absence of illness or disease is not the same thing as positive well-being or happiness. The impact of increased frequency and duration of work on health and on well-being is different with different antecedents. In terms of work psychology, we know a lot about the antecedents of what makes people stressed at work and have developed many sophisticated measures; however, the same information might have been gained by simply asking how stressed

the person is. Sometimes single items do capture what we want to know. This point goes back to what has already been discussed: what do we want to capture and what do we want to do with the information?

In the area of work, questions of interest include the kinds of interventions that will not only prevent people from becoming ill from their work, but also improve well-being at work. People most often indicate that their boss is their biggest source of stress as well as their biggest source of support at work; this is a complex issue. Positive impacts of work often have less to do with the status of the work, but the meaning and value people associate with their work. For example, in an examination of hospital cleaning staff, those who reported enjoying work and having work satisfaction were those who perceived that their work involved fighting infection and had an important purpose for the hospital environment, not as performing menial tasks. This is a positive finding because it indicates that people can have positive well-being at work that is not connected to status or income. By making work more meaningful, managers can improve workers' well-being as well as work performance.

There is a lot of focus on unemployment, and we know that being unemployed is more detrimental to health than being employed. But we do not really focus on the aspects of work that are healthy and positive. Government tends not to intervene in work organizations in Western developed economies. Work is very important in the discussion of well-being but it is an area where policy may have little ability to intervene. More than any other domain of people's activity, work is the one they have the least control over.

Richard Frank

Deputy Assistant Secretary for Disability, Aging, and Long-Term Care Policy, U.S. Department of Health and Human Services

At the U.S. Department of Health and Human Services (HHS) there are concerns about trying to measure performance in terms of the well-being of people with illnesses. We need to determine the policy areas in health care where well-being measurement might be uniquely useful and address the challenges in using these measures for tracking and accountability from a regulator's perspective.

It is not surprising that the measurement of well-being has the advantage of tracking outcomes we actually care about in public policy beyond a particular set of signs and symptoms. We can use well-being for the purposes of evaluation for organizations, agents, or policies to allow for a broader perspective than a narrow clinical one. There is currently a booming industry for measuring various kinds of health indicators. At HHS right now there are five or six working groups measuring different kinds of performance indicators, ranging from specific to general, all related to health care reform. In some areas, measures focused on clinical outcomes are adequate for most of what a policy maker needs to know. In other areas, information about quality of life beyond clinical outcomes will be necessary because they are areas where the domains of interest extend beyond signs and symptoms or clinical benefits may not be well correlated with other outcomes of interest.

There are three critical health areas where clinical outcomes do not provide enough information: a) long-term care of people with progressive chronic illness; b) end-of-life care; and c) the care of people with severe and persistent mental disorders. In each of these critical areas, interventions can improve health status but may potentially miss things that matter most to the people and society if measures of well-being and quality of life are not included.

Layard indicated that of all groups with disabilities, those with severe and persistent mental illness are the least likely to adapt. People with mental illness indicate the following aspects to be greatest importance to them: safety, having a place to live with privacy, having work, and having relationships. Changes in the clinical symptoms of mental illness may help some of these things that matter most, but they often do not. Recent research on the most evidence-based therapies found that very few of them affect cognitive decrements of the illness and that is what is most important in terms of getting people back to work. Without treatments that focus on this aspect, it is hard to impact quality of life.

Illnesses are disruptive to people's lives. We need to be able to measure things where the policies and management structures we are trying to affect are actually going to affect quality of life. Two-way causation is problematic (e.g., people with mental health problems who have unstable housing do less well clinically but the reverse is also true). A second point is that we need measures that are concrete, clear, transparent, and can allow people to understand what is being measured. Quality-of-life interventions and evaluations that are being developed for people with severe and persistent mental illness should focus on the outcomes of work, housing, and relationships. We need to specify ways of adjusting for differences at the individual level (i.e., risk adjustment) to get at the effect of the policy and not the effects of selection.

Open Discussion

Role of Biomarkers

Results of studies that focus on a single biomarker should be interpreted cautiously. Articles are published when differences emerge and not published when there are no differences. This is a difficult area because there is no standardization of particular measurements. Steptoe has found more consistency in the relationship between momentary measures of well-being (as opposed to evaluative measures of well-being) and biological markers. That makes sense from a biological perspective because we are talking about corticolimbic connections relating the parts of the brain that are associated with emotion through the autonomic nervous system and other vegetative functions. One might expect a closer link there than to the broader constructs. However, various evaluative measures of well-being also are related to biological indicators; the mechanism of that relationship is more puzzling.

Little can be done about the relationship of biological markers to life satisfaction in terms of policy. Rebecca Blank (U.S. Department of Commerce) noted that it is possible that twin studies on well-being may help discriminate the biological from environmental factors as well as the interaction of the two (i.e., nature versus nurture). Steptoe indicated that the biological markers have a strong hereditary component and that most of the data he presented was cross-sectional. Future research should manipulate both biology and well-being.

Role of Work

The NIA has funded research with the National Institute of Child Health and Human Development that looks at interventions in the workplace to change the structure around flexibility for work-family practices. The pilot work found individuals with managers that showed higher creativity scores for managing flexibility slept 30 minutes more per night and showed fewer cardiovascular risk factors than those with less flexible managers.

Carol Graham (Brookings) provided examples from data in Latin American countries about job satisfaction and work productivity. Having job satisfaction was not found to be linked to productivity. However, doing something purposeful yesterday at work was linked to higher levels of productivity. There are different domains of well-being that may be linked to productivity.

Several participants discussed a possible ethical dilemma in that managers are interested in improving workers' well-being for the purposes of increased productivity and retention, but they are motivated to do so at the lowest possible cost. In other words, there is an incentive for managers to determine how to provide as little as possible to workers while manipulating other factors that will increase satisfaction, while not actually addressing workers' well-being.

Making people feel that their work is worthwhile is a good thing and should be a goal across industry; this is different than providing the cheapest possible benefits. Benefits packages may be a part of the reason people go to work, but it is not the reason they work when they get there. Rewards from benefits packages and wage increases wear off and the impacts dissipate. Work is not just about financial reward. A recent analysis of the British Household Panel Survey data looking at the impact of performance pay showed that earning less than £30 per hour is bad for well-being, and earning over £30 is good for well-being.

Policy Areas Informed by Well-Being Measures

Richard E. Lucas

Associate Professor, Department of Psychology, Michigan State University, East Lansing, MI

Even 10 years ago, psychologists would have dismissed the idea that well-being could be used in a policy context. There was a sense of pessimism then and even now about the extent to which life circumstances matter at all to our measures of subjective well-being. A pessimistic view is that life circumstances account for a very small percentage of the variance in happiness. In contrast, evidence from different trends of research shows that personality traits seem to matter. Personality traits are correlated moderately with well-being measures, and the well-being measures are stable over time. Twin studies show that heritability of the well-being measures is moderate; there are circumstances or events that can make people feel better or worse, but it is inevitable that we adapt to our genetic baseline, and therefore life circumstances have no lasting effects.

We have taken a second look at these pessimistic conclusions in recent years. More economists are starting to work in collaboration with psychologists. Economists bring a different perspective

on the size of effects, different analyses, and different types of data that are relied on to answer these questions. This new perspective has enabled psychologists to reevaluate and adjust their interpretations. For example, Brickman et al. (1978) was a seminal work that laid the foundation for the idea that people adapt to their life circumstances.²⁷ Brickman and his colleagues compared subjective well-being among people with spinal cord injuries, matched controls, and people who won the lottery. The authors concluded that the differences among these groups were not as large as would be expected. Since then numerous studies have looked at people with spinal cord injuries. When the size of the effect found is quantified and compared across multiple studies with standardized mean differences of effect, there is actually a large effect between the control group and the group with spinal cord injuries in Brickman et al. Most of this literature shows medium to large effects, yet these studies are rarely cited and primarily published in the rehabilitation literature. We do need to be cautious, however, because these are cross-sectional effects.

Efforts have been made to look at this question with longitudinal data. A prospective study of disability onset in the German Socio-Economic Panel started with people who did not have a disability at the beginning of the study.²⁸ The study examined if people who subsequently experienced a disability were different from the sample average before disability onset and if they return to baseline level 3 years after disability onset (defined as when the government certifies them disabled). There was no statistically significant difference among the samples before disability onset. There was a decline among those who became disabled prior to disability onset compared to those who did not develop a disability (this makes sense given the definition of disability onset). Also, there was no trend of return to baseline after onset; if anything there was a non-significant decrease over time following the onset. Differences based on severity also existed (i.e., the well-being of those with severe disabilities decreased more so than those with less severe disabilities). This type of consistent evidence shows that long-term disabilities do impact well-being in the long term.

Given that life circumstances do indeed matter, we can address the possible uses of subjective well-being measures to evaluate policy. One example is an airport noise study comparing life satisfaction in areas with high and low noise levels.²⁹ Standard approaches to evaluating the impact of noise at a particular airport are problematic because they rely on strict assumptions that can easily break down: comparing housing prices assumes people can move easily and willingness-to-pay questions assume the respondents would not adjust their ratings simply to make a point. People in different areas were surveyed about life satisfaction without them knowing that the purpose of the survey was to evaluate airport noise. Findings indicated that people living in noisy areas have lower levels of life satisfaction, but this was not reflected in housing prices. Using this information, possible policy interventions could include compensation calculations (which may be controversial) and the evaluation of the effectiveness of noise insulation. We can include well-being measures in policy analysis without explicitly saying well-

²⁷ Brickman, P., Coates, D., & Janoff-Bullman, R. (1978). Lottery winners and accident victims: Is happiness relative? *Journal of Personality and Social Psychology*, 36, 917-927.

²⁸ Lucas, R. E. (2007). Long-term disability is associated with lasting changes in subjective well-being: Evidence from two nationally representative longitudinal studies. *Journal of Personality and Social Psychology*, 92, 717-730.

²⁹ Van Praag, B. M. S. & Baarsma, B. E. (2005). Using happiness surveys to value intangibles: The case of airport noise. *The Economic Journal*, 115, 224-246.

being is an outcome goal. The reasons the use of life satisfaction works in this particular policy analysis are that it is a well-defined issue, it addresses something that is generally accepted as undesirable, and the use of well-being measures are better than alternative measures (e.g., willingness to pay).

A second example is a pollution study that compared life satisfaction of people downwind of power plants as new emissions regulations were being implemented. Evaluators wanted to assess if a reduction in SO₂ was associated with an increase in life satisfaction. In this case, the finding validates the measures (i.e., the survey was not about pollution per se and yet we can still see the effects of pollution on measures), and it affirms an already implemented policy (i.e., the new environmental regulations). However, it is unclear what the implications of this information are for policy decisions. If no effect was found, does that mean the government would not have implemented the policy? Trade-offs may become an issue as well; what action would the government take if the findings had indicated that regulations decreased pollution but increased unemployment and decreased life satisfaction?

A final example is marriage. The effect of marriage on life satisfaction is not as strong as people have suggested. The average correlations range from 0.06 to 0.09, which is weaker than the correlation with income. Cross-sectional results tend to be higher in the United States. However, longitudinal analysis looked at the happiness of people before marriage (baseline), moving into marriage (reaction), and after marriage (adaptation).³⁰ People who are going to get married started off with higher levels of life satisfaction at baseline, the levels increased at the time of marriage, but then returned to baseline after marriage.

When we look across different levels of analysis, we sometimes get different answers. A Centers for Disease Control and Prevention (CDC) surveillance survey looking at county levels of life satisfaction shows that effects differ across individual and county levels.³¹ Perhaps the goal of a policy is to increase the life satisfaction of a county and increasing the number of married people makes the county happier overall (correlation of .42 at the county level) even if all married people themselves are not necessarily happier (correlation of .21 at the individual level). The aggregated level is the biggest predictor for education (correlation of .52 at the county level and .07 at the individual level), and the effects are even stronger for education at the state level.

Well-being is associated with a variety of life circumstances, and well-being surveys provide an efficient way of evaluating the impact of these conditions on people's lives. Further, well-being measures can provide information that other approaches cannot. In some cases, the use of well-being measures is relatively straightforward, but in other cases it requires that people value well-being itself as an endpoint. Cross-sectional data may not tell us about the underlying processes, and the data we have available may provide information about the wrong level of analysis. We need to think carefully about the intended uses of these well-being measures. It may be controversial to assume that the goal of public policy is to increase everyone's well-being.

³⁰ Lucas, R. E., Clark, A., Yannis, G., & Diener, E. (2003). Reexamining adaptation and the setpoint model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology*, 84, 527-539.

³¹ Lawless & Lucas, in press.

Discussants

Rebecca Blank

Under Secretary for Economic Affairs, U.S. Department of Commerce

It may not be the business of policy or government to care about individuals' happiness. The United States is founded on the principle of "life, liberty, and the pursuit of happiness," not on attaining happiness itself (i.e., equal opportunity versus equal outcomes). The policy environment in which people have opportunities to live a satisfying life is different than an environment in which there are policies with the goal of making sure everyone actually has a satisfying life. We might measure access to education, health care, and employment and the value of those resources being available to everyone in a community, but we might not actually need to assess individuals' life satisfaction. Well-being measures are indicator variables for the things we can measure more directly and impact with policy (e.g., health care, employment, education). Happiness is not the policy objective; the policy objective should be a society in which people are able to pursue happiness.

Another issue is addressing whose happiness policy makers should be concerned with. For example, low-income interventions for those in poverty are not necessarily about increasing happiness among the poor, but about providing a set of services that satisfies the taxpayers. An evaluation of the food stamp program may find that the satisfaction of recipients is only minimally impacted (i.e., their well-being would be increased more if they were given cash instead). The vehicle of this intervention is not just about the satisfaction of the recipients, but the taxpayers as well (i.e., ensuring tax dollars are actually going toward providing pre-selected food items to families rather than not knowing where cash is spent). It is not always obvious in the political environment that the person on whom the program is focused is the person whose satisfaction is being maximized, which has implications for policy evaluation.

Dynamic patterns matter enormously. Policy analysis is often not about comparing life with and without the policy, but about a dynamic pre/post change in policy. Longitudinal data are more informative and credible for policy than cross-sectional data to examine pre/post policy effects. The change issue itself may be problematic given that people typically do not like change and that satisfaction measures around policy changes may pick up this change effect rather than the policy effect. We need to look at dynamics over time, and knowing when to measure and how to interpret the results within the changing policy context is as important as determining what to measure.

All of this is not to say that well-being measures cannot be useful, particularly where we do not have more direct measures. However, there is skepticism as to how far and in what ways these measures can be used directly for policy evaluations where the interpretations are clear. This is an extremely young literature with respect to policy evaluation, and much is unknown.

Felicia Huppert

Director of the Well-Being Institute, Department of Psychiatry, University of Cambridge, Cambridge, U.K.

We have a normal distribution of well-being. After decades of research, we know an enormous amount of information about the negative end of the range of well-being (e.g., depression, anxiety). However, experts disagree on how to define positive well-being. Flourishing is the positive end of the mental health spectrum, operationally defined by inverting the symptoms of the negative end: positive emotion, engagement, meaning, self-esteem, optimism, resilience, vitality, competence, emotional stability, and positive relationships.

The well-being module in the European Social Survey (43,000 participants across 23 countries) contains items which more or less correspond to these 10 features of flourishing. Two very clear factors emerged from a factor analysis of the well-being module data: positive attributes (emotional stability, vitality, optimism, resilience, positive emotion, self-esteem) and positive functioning (engagement, competence, meaning, positive relationship). These were quite distinct factors, and future surveys should at least be measuring these factors. When life satisfaction was added, a distinct third factor emerged: positive appraisal (life satisfaction and positive emotion). Happiness loaded completely onto the third factor. When life satisfaction and happiness are measured in general, they are highly correlated. More research is needed to determine how many components of each factor to measure and how to weight the measures.

There is a richness in the subjective well-being data that we have not been capturing thus far. GDP is a summary of many key economic variables. We need a summary of key subjective well-being variables. But as with GDP, the information provided by the individual variables is more important for policy than the summary measure alone.

Open Discussion

Well-Being Measures for Different Policy Purposes

Much of the discussion centered on feelings of skepticism or optimism about the usefulness and appropriateness of well-being measures for different policy purposes. Justin Wolfers (University of Pennsylvania) raised a question that continued to emerge in discussions over the course of the workshop: is there a standard policy question where subjective data give a different answer than objective data, and we are confident that the subjective data are correct? Several agreed that it is a good question without a clear answer at this point.

Blank posited that it is not enough to say a measure would add value. We need to be certain that the measure would add value relative to all other available measures and have an impact. An argument needs to be made that the well-being measures are specifically the ones we need, especially amid dwindling funding for data agencies. David Weir (ISR/Michigan) expressed concern that measuring well-being in the policy context will open the door to concerns that are not traditionally the business of the government (e.g., life satisfaction and marriage).

Others argued that there are many reasons from a policy maker's perspective that the government is interested in well-being, particularly where there are strong patterns of externalities and income effects. Another clear area where there is definitely value to well-being measures is the health sector. Well-being is certainly a very important outcome in addition to survival and life expectancy, especially for people receiving treatment. There is no other reasonably good economic measure (e.g., amount spent per patient) that could substitute for public views.

Well-Being Measures for a National Progress Monitoring Function

Several participants felt strongly that well-being measures, especially ones used in a national account function, ought to be understood and accepted by the public. Layard asserted that if a measure of how people feel is used, we also need to take into account how they feel about the measure. Dolan agreed that it is entirely legitimate to consult with taxpayers about how the taxpayers wish their money to be used. It would not be the only input, but it matters in determining if it is a legitimate policy goal to maximize well-being. Kahneman stressed that the presentation of the measurement to the public matters as well. Several have argued that we should have a national statistic as a descriptor of how society is doing and that it should measure change over time. Life satisfaction does not satisfy those constraints. A measure that does not vary as circumstances change is not going to be impressive to the public.

Robert Hauser (DBASSE) countered that we have existing measures that do not change, such as the percentage of Americans who move across county lines; therefore, the criterion that the measure changes is not critical. Kahneman pointed out that even though that measure is reported, it does not explain anything government does or inform policy. Economic growth in the United Kingdom is an example of an aggregate measure that does not change in an easily visible way, but is still a meaningful objective for policy. U.K. economic growth has been stable at 2.25 percent for a long time; however, it is still used in cost-benefit analyses without interventions being able to conceivably contribute to changing it.

There have been large national shifts in subjective well-being among the eastern European countries as they ceded to the European Union, providing an opportunity for a natural experiment. GDP in developed countries does not seem to drive life satisfaction, but it is possible it plays a role in developing countries.

Some argued that the primary usefulness and appropriateness of well-being measures is for targeted populations and interactions. The overall numbers are not useful. Even if the well-being numbers are stable at the national level, it would still have value if the sub-group numbers are revealing. Also, the fact that the measures have not changed to date does not necessarily mean they will never change. Satisfaction with different aspects of health services do change; it might be the case that we need to drill down into the life satisfaction measure to find targeted populations and circumstances that change and therefore are amenable to policy intervention.

We should not be looking at well-being in terms of a single measure, but rather as a dashboard of indicators. One scenario could involve the government having a lean monitoring function, while the academic sector conducts more in-depth ongoing measurement research simultaneously. We have a shorter time frame for developing and implementing specific well-being measures for the

purpose of policy evaluation. Policy problems exist now, and we need to know the social impacts of particular programs. We have a longer time frame to refine and implement well-being measures for a national progress monitoring function.

Summary and Directions

The second day of the meeting began with opening remarks by eight invited experts that addressed challenges in measurement of subjective well-being at the population level, successes and pitfalls in applications of subjective well-being measures for policy purposes, and research and data needs for the field. Their remarks were followed by open discussion about the sort of subjective well-being measures that can be practically used, how they can be practically applied to public policy making, and research needed to advance the agenda.

Paul Allin

Deputy Director, Societal Well-Being, Office for National Statistics (ONS), South Wales, United Kingdom

The U.K. government is committed to developing subjective well-being measures, embedding well-being in policy delivery, and implementing better measurements for policy appraisal led by the Cabinet office.³² There already have been references to well-being in the budget and spending review. Recent attention to measures of well-being has been spurred by politicians promoting an emphasis on the well-being of the population³³ and by international efforts, such as the Stiglitz report, that have shaped a broader agenda in a number of countries and international organizations.

The series of recommendations in the Stiglitz report provides a global framework for thinking about how and whether to implement well-being measures. The report focuses on three broad strands: 1) classical GDP; 2) objective and subjective measures of quality of life; 3) and the environment and sustainability. In terms of measuring material living standards, the Stiglitz report recommends focusing on income and consumption rather than just production; wealth jointly with income and consumption; the household perspective; the distribution of income, consumption, and wealth; and broader income measures including nonmarket activities. ONS refers to the broad sweep of the Stiglitz report as measuring national well-being.

There is interest in creating a dashboard of a limited number of indicators, not in replacing a single measure, GDP, with another single measure. The focus is on shifting the “emphasis from measuring economic production to measuring people’s well-being,” where well-being is defined

³² Evans J. and Thomas J. 2010. There’s more to life than GDP but how can we measure it? *Economic & Labour Market Review* 4(9), 29-36 (September). http://www.statistics.gov.uk/elmr/09_10/downloads/ELMR_Sep10_Thomas.pdf (accessed March 16, 2011).

Waldron S. 2010. Measuring subjective wellbeing in the UK. Office for National Statistics Working Paper. <http://www.statistics.gov.uk/articles/nojournal/workingpaper-measuringsubjectivewellbeingintheuk.pdf> (accessed March 16, 2011).

³³ See, for example, the October 25, 2010, speech by Prime Minister David Cameron at the annual meeting of the Confederation of British Industry: http://www.conservatives.com/News/Speeches/2010/10/David_Cameron_Creating_a_new_economic_dynamism.aspx (accessed March 16, 2011).

in a multidimensional way and examined within the context of sustainability.³⁴ There is an existing evidence base for broader measures that include subjective well-being. The United Kingdom now has a wealth survey that makes it possible to examine the distribution of wealth and how wealth changes. The first period of data was collected just as the recession started. The United Kingdom has already done work on estimating the economic value of unpaid work, experimental household accounts, and social capital. Now the focus should be on filling the gaps with national and local subjective well-being measures and continuing work on the value added by public services on human capital and on ecosystems accounts. The development process should be transparent and welcoming of public involvement.

The United Kingdom is taking several next steps. There will be a Cabinet office well-being conference on November 25, 2010, with the Prime Minister. Following the conference, Jil Matheson, U.K. national statistician, will launch a national debate about measuring national well-being, seeking widespread views on the dimensions of national well-being (“what matters to you?”) through online and in-person events. This conference will be followed by a well-being measurement forum in January 2011. Subjective well-being measures (covering overall life satisfaction, affect, and purpose) will be included in the ONS Integrated Household Survey administered from April 2011 to 450,000 selected respondents. The size of this dataset will allow for examination by local authority level and demographic characteristics. In addition, the U.K. omnibus survey of 12,000 people per year will headline well-being questions as well as further detail questions as a way to explore these measures. The long-term goal is to develop a core set of national well-being measures, including subjective and well-being assessments, which cover the key dimensions of national well-being and the progress of the U.K.

Allin highlighted three areas of need: 1) Fitting measures of subjective well-being into wider measures of national well-being (what the new economics foundation had proposed as “national accounts of well-being”), and addressing how to make them work and how to engage the various stakeholders. 2) Keeping measurement aligned with policy. The value of the headline measures is for framing and signaling without all the detail. 3) Addressing the implications of having a national debate around measuring well-being and subjective well-being; it is a delicate balance between reaching standards too soon and missing the opportunity for a harmonized approach, and including internationally-agreed measures. Some have made reference to premature ossification; rather, harmonization is about the benefits of managing standards and evolving them in a controlled way, not about putting them in place forever.

Somnath Chatterji

Department of Health Statistics and Informatics, World Health Organization

In 1947, the World Health Organization (WHO) defined health as a person’s physical, mental, and social well-being, not simply as the absence of infirmity. It was an aspirational definition at the time, but there is no reason to believe that the intent was to assume all of well-being is equated with health. It is imperative to separate the notion of health from well-being for the purposes of measurement.

³⁴ Stiglitz et al., 2009, p. 12.

The historical definition of “quality of life” is also problematic in that there is no consistent definition in the health literature. Some equate quality of life to functioning status and others to more subjective measures of well-being or a mixture of components. The WHO recognizes quality of life as a contextualized subjective measure and defines it as an individual’s perception of his/her position in life in the context of his/her culture and value system. However, despite this working definition, the instrument used to assess quality of life measures health outcomes, functioning, and appraisal of satisfaction with domains of life. Quality of life or well-being needs to include cognitive appraisal of life satisfaction as well as affective experience (the remembering self versus the experiencing self).

Several issues have been raised. First is the need to agree on the universe of measurement and have conceptual clarity. It is important to have a common language for well-being. Second is the issue of fitting measurement to purpose. Even with agreement on the conceptual boundaries of well-being, measurement may focus on a particular aspect or have a certain perspective depending on the purpose for which the data will be used. If the measures will be communicated for policy purposes, then the measures and definition must match the intuitive notion that people have for well-being or happiness. This is a challenge in the context of the multiple languages involved in any work conducted by the WHO, for example, how to convey the nuanced notion of interest, ensure measurement fidelity, and ensure that the construct is being captured correctly across multiple translations. This is especially difficult in large population surveys conducted in multiple countries with limited time to ask questions that can help ensure conceptual clarity.

Third is improving the fidelity of measurement. A measure must be sensitive to change and collected longitudinally in order to disentangle causality. Although health is considered a contributive factor for well-being, well-being also may be an important predictor for health outcomes. Huge datasets are needed to examine strategies for improving well-being. The size of the dataset is also dictated by the fidelity of the measurement, how sensitive that measure is to change, and how sensitive the measure is in terms of distinguishing different population subgroups of interest. Another issue concerning fidelity of measurement is framing. For example, if well-being measures are asked in the context of a health survey, it is quite possible that responses about well-being will be unusually influenced by health status; this can lead to systematic bias and there are implications for policy when multiple surveys provide contradictory results. Other fidelity of measurement issues to address include biomarkers, cross-population comparisons, contextual effects, and set points.

The WHO conducted a survey of 50,000 people in six countries using nationally representative samples that included questions of both evaluative and experienced well-being (using a version of the day reconstruction method). This sample will be followed longitudinally for 5 to 10 years. Preliminary analyses of the cross-sectional data indicate that older people are more likely to report higher levels of life satisfaction than younger people. This is preliminary data; it could be that the people who survived are happier, which is consistent with what has been reported in other studies. There is a clear relationship between life satisfaction and household wealth, social networks, and chronic illness. A similar pattern emerges when looking at the duration weighted net affect derived from the day reconstruction method variables. The difference is that the income gradient is much steeper with overall life satisfaction than with the duration weighted net affect. Additionally, the duration weighted net affect has an impact on health. The survey will

include measures of biomarkers, hopefully heart rate variability, and other markers of emotive states. These preliminary data show that the determinants of life satisfaction and experienced life are very similar, which raises the question of whether or not measures of both are needed.

Robert M. Groves

Director, Census Bureau, U.S. Department of Commerce

Groves summarized the recommendations that have been made at the workshop thus far: 1) actively model and analyze the items we currently have; 2) conduct research on measurement until we understand key properties; 3) invent a conceptual framework that fully integrates with GDP in some way; and 4) enter the domain of official statistics with large samples. The goal is to determine which of these strategies is appropriate and the usefulness of subjective well-being measures in the larger societal assessment.

Measurement is clearly an issue. The high variability of estimates from sample surveys is a function of both sample size and reliability of the items. From a psychometric perspective, creating indices of correlated indicators of the same construct can reduce error variance. A multi-item index is also desirable because of the multidimensionality of the basic constructs.

Groves identified four critical components of an evolutionary process toward the use of subjective well-being measures. The first component is deep and rich longitudinal measurement of subjective well-being which would be best accomplished in the academic realm. The survey would not be just for academics interested in understanding well-being for their own purposes, but also for contributing to a larger national goal of identifying subsets of items for other monitoring devices. The smaller set of items contributing to the larger national goal may change over time as improvements are made. The other purpose of this component would be very active modeling through item response theory (IRT) and other approaches to identify subset items as well as causal modeling of hypothesized mechanisms.

The second component of the evolutionary process is for field experiments to follow causal model assertions on important policy levers. The field experiments should focus on a nexus of attributes that are on the candidate list for policy intervention. One might envision a great number of field experiments, but those in a policy space that could be manipulated should be overrepresented in the research program.

The third component is conceptual – an articulation of a theoretical structure that motivates and integrates the measurement and the statistics derived from them.

The fourth component is the initial preparation for official statistics. After core index items are identified by the longitudinal survey, these would be included on key Federal government surveys. Even though the Federal government is committed through the Office of Management and Budget (OMB) to minimize redundancy, context effects found in the literature are such that these monitoring surveys ought to be multiple in nature. The same items on different surveys may yield different response; they will not be redundant, rather they will add to learning about the measures.

Groves proposed that all four of these components commence concurrently and that a commitment is made to the evolutionary process. If only one strategy is employed at a time, it will slow the overall development process and the end goal of generating credible national statistics. This approach demands a structure including the three components of the evolutionary process brokered by a Subjective Well-Being Research Board. This proposed research board would be responsible for coordination and cooperation across the three levels of studies being conducted (academic longitudinal survey, field experiments, and Federal monitoring surveys). The research board should: 1) have a budget and decision-making authority; 2) be able to intervene in the longitudinal studies to direct focus to national goals; 3) coordinate the selection of subsets of measures used in the Federal monitoring surveys; and 4) suggest field experiments to program agencies, as appropriate.

Steven Landefeld

Director, Bureau of Economic Analysis, U.S. Department of Commerce

Landefeld agreed with the assertion that in order to be useful, subjective well-being measures should react to policy and events over time, but it is not clear that the broad ladder-of-life or life satisfaction measures are responsive in that way. People do tend to find some effect, but it is relatively short-lived. It may be more the case that people return to a genetic or cultural set point in terms of well-being measures over time.

A measure likely to be adopted for policy purposes must be simple and react the way policy makers and the public expect it to react. For example, the Gallup-Healthways Well-Being Index provided a limited response in both depth and duration to the financial crisis. Measures of well-being dipped after the fall of Lehman Brothers but rebounded after only 5 months, which may indicate that after a temporary event, people revert to a baseline level of well-being. The current recession has been unusual in its pervasiveness; it has affected the entire population and has been difficult to overcome. After more than 2 years, productivity is back up, but unemployment remains high. The asset and housing crises continue to affect everyone planning for retirement. The Michigan Consumer Sentiment Index, an indicator of subjective well-being that tries to measure this phenomenon, is still 30 percent below its pre-recession peak.

There have been advances in the measurement of well-being and understanding the relationship between well-being and income, wealth, and other variables. It was not clear to Landefeld that official statisticians should adopt such measures for guidance of policy, especially in the context of reduced and shrinking resources. The need for broader measures can be addressed first by filling the gaps in existing national accounts. For example, as Stevenson and Wolfers (2008) have pointed out, by looking at the distribution of income statistics one can better understand why measures of U.S. well-being did not track with GDP per capita (between 1972 and 2008), in part because median income was not rising during that period as were mean income and GDP per capita.³⁵ Including that kind of information will be extraordinarily important, and it represents relatively “low-hanging fruit.” That is not to say it will be easy and there will be significant challenges in reconciling the distributions of income from household data and tax data and

³⁵ See Stevenson B. & Wolfers J. (2008). Economic growth and subjective well-being: Reassessing the Easterlin paradox. *Brookings Papers on Economic Activity* (Spring), 1-87. Retrieved March 16, 2011 from <http://bpp.wharton.upenn.edu/jwolfers/Papers/EasterlinParadox.pdf>.

expanding to national accounts. Information on the distribution of saving, consumption, and wealth needs to be included as well, as recommended in the Stiglitz report and discussed by Allin.

Another data gap to fill concerns measures of traditional economic sustainability, for example, better measures of leveraging, of the overall quantitative magnitude of the housing bubble, and the unsustainable trends between asset prices and income. More of this type of information can be derived from existing data or from post-financial crisis data emerging from regulators. Traditional economic sustainability includes the notion of sustainable GDP, or GDP less an allowance for capital used in production, or the same concept in the area of investment—gross investment versus net investment. This distinction between gross and net measures turns out to be quite important for the U.S. in assessing the sustainability of policies over economic cycles. We also want to invest in statistical infrastructure for measuring longer term sustainability. For example, it may be possible to construct much more detailed data on the energy sector that could be linked to data on emissions to better address environmental sustainability.

Other extensions of the national accounts include satellite accounts for health expenditures, household production and investments in human capital, innovation, and research and development. Including these items would be relatively inexpensive and mainly build on existing data or a better integration of the existing data. There will be some acquisition of new data in the post-financial crisis period, but it should be inexpensive relative to new data collections and surveys.

These efforts are not to suggest that national accounts are the only answer. There are plans for collaboration with social well-being researchers for specific policy issues. Targets should be items that are important to economic policy. Examples include examining links between health effectiveness and spending; energy and the environment; and household production and time use. The BEA is collaborating with the Energy Information Administration and with outside researchers on energy and the environment. BEA also could collaborate with BLS in producing household production type estimates using the BLS time use survey, and with academic and other researchers to extend these household measures to welfare measures within the type of economic framework used by Jorgenson (1997) and by Jones and Klenow (2010), which could in turn be integrated with a new BLS module on well-being.³⁶ Such collaborative work could begin to accomplish needed research on the linkages between well-being, economic activities, and time use. In summary, while incremental progress can be made, the overall informational requirements can be immense and not easily or quickly resolved.

Christine McGuire

Research and Development, Department of Health (England)

McGuire raised three points based on the perspective of working closely with national policy makers to identify short-, medium-, and long-term research needs to inform policy development and evaluation within a finite budget. First, evidence informs policy in a number of ways.

³⁶ See Jorgenson D. (1997). *Welfare: Volume 2, Measuring social welfare*. Cambridge, Massachusetts: The MIT Press and Jones C. & Klenow P. (2010). Beyond GDP? Welfare across countries and time. *NBER Working Paper No. 16352*. Accessed March 16, 2011 from <http://www.nber.org/papers/w16352>.

Research evidence can inform problem identification, policy options, the development of policy, policy evaluation, monitoring, and methodological development. In terms of policy appraisal, the demands are immediate. Well-being measures that will be ready in 2 years may not be an attractive proposition. This will have implications for communication.

Second, often policy research has to be tailored for the particular policy needs to be useful. Given keen competition for research funds, compelling arguments need to be made to justify funding for particular purposes. Questions about the value of specific well-being measures over others, and the need for different levels of sophistication in measures must be considered in the context of how best to use limited resources.

Finally, marketing and communication are key. Policy makers have to plan for future needs, but they are under immense pressure to deliver now. It will be helpful to have clear arguments laid out in simple language that demonstrate what well-being measures would do for policy makers in the short, medium, and long term. Additionally, the wider research community needs to respond to efforts to integrate well-being measures across research projects for the purpose of informing policy; education and communication efforts may be needed to help researchers respond. Providing funding for research is not enough; the research community needs to understand how to integrate well-being measures as a matter of course.

Katherine Wallman

Chief Statistician, U.S. Office of Management and Budget

The current state of the science on measures of subjective well-being can be attributed to a number of efforts, including the pioneering work by Krueger and Kahneman, the 2010 supplement to the American Time Use Survey, and several efforts ongoing by the BEA. This workshop and hopefully a future panel will bring us farther along this route.

U.S. agencies producing official statistics know they must answer a set of questions as part of the OMB approval process. These questions apply to measures of subjective well-being as well and feed into next steps.

1. What is the goal of the measurement and what are the intended uses of the information?
2. What questions need to be asked?
3. What level of detail is needed (e.g., geography, population groups) and with what frequency?
4. What core items are needed? What can be implemented quickly? The OMB does approve replication on different surveys if it is done for comparability as well as contextual reasons.
5. What measurement vehicle is needed to carry out this work? Echoing the outline Groves set out, how much of it needs to be on an ongoing Federal survey and how much can be done in partnership with the academic and research community?
6. How much of it needs to be part of official statistics?

The OMB recognizes the need to collaborate with academic colleagues to answer many of these questions.

David Weir

Director, Health and Retirement Study, Institute for Social Research, University of Michigan, Ann Arbor

The Health and Retirement Study (HRS) is a longitudinal survey conducted in the academic sector of more than 22,000 Americans over the age of 50. The HRS is not an official government survey, although it is federally funded by the NIA and the Social Security Administration. The HRS is not positioned to be “the” national study, but it can become one of a constellation of studies informing issues of national interest. There has been an explosion of new measures in the last 5 years, including subjective, psychological, and psychosocial measures that have greatly enriched the study. Last year the HRS added experience-based assessment items from the Gallup measures and life satisfaction.

From his perspective as an economist, Weir noted two intellectual challenges to the roadmap that Groves proposed: 1) National accounts capture only factors that can be measured in a market economy; they do not value the externalities of productive activity, such as environmental degradation, nor do they account for factors such as the entry of women into the paid labor force; 2) Economists at the end of the 19th century decided that measuring W_{it} , the well-being of person i at time t , was impossible and furthermore, unnecessary to the axioms of economic science. If one were to look at the variance properties of W_{it} , the i is primarily individual variation—how people perceive and report their experiences. Using the cross-sectional variation in W_i is probably a dangerous guide for policy focus, because people who feel unhappy would be the focus of all efforts when we know that that is more of an individual characteristic than something that can be impacted. Time t also has problems because it tends not to show a long-term trend or more persistent variation as expected. Although it is fine to seek better measures that have more desirable properties, Weir argued that we do not *need* to measure cardinal utility in order to address the shortcomings of national income accounts.

Economists assume that people know what makes them happy and they pursue it, and that the value of nonmarket activity can be understood using revealed preference type techniques (e.g., willingness to pay, compensating differential measures). However, this strategy unravels if people in fact do not know how much happiness they will derive from something they do not currently have or are somehow blocked from acting on those preferences. This is an area where better measurement would be helpful, and where we might be able to learn from our European counterparts. A fitting goal moving forward would be to look for a place in the American context where the problems of measurement can be contained by a well-designed study and the use of good measures to permit comparison of a subjective well-being metric with more conventional means of policy assessment.

David Halpern

Deputy Director, Institute for Government, London, United Kingdom

Well-being is driven by multiple unexplained factors that are not included in current accounts, yet these uncaptured factors might inform marginal choices in several policy areas. One of the most interesting developments of Allin’s work concerns small area variations, promoting debate and discussion about why particular areas have higher well-being. In some ways, this

information can empower local democracy. Halpern identified some practical next steps:

1. Well-being measurement needs to be taken seriously. Of note, the Prime Minister is going to launch a large-scale initiative on this topic, and it will be important in the coming weeks and months to promote informed discussion.
2. A key issue is confirming that well-being is something that can be moved, and hence amenable to influence through policy levers. High on the list of next steps would be intervention studies that demonstrate it is possible to move subjective well-being measures.
3. There is an urgent need to establish a prototype. It would be fantastic if other countries have comparable measures.
4. If the measures had to be reduced to one type of question, it should be about life satisfaction. Of course, questions about the meaning of life are also important because they hint at one's likely life trajectory. More field work and factor analyses are needed for data reduction and the best way to explain the variance.
5. Work on well-being does not negate the need to revisit the GDP-plus type models. There also are issues about cross-generational effects and on the environment and so on. Over a longer time cycle, the chief statistician will work with other parts of government to drive the effort to be more sophisticated about GDP, including incorporating subjective well-being measures.

The current administration in the United Kingdom is sympathetic to the democratic element of this approach and the sense that citizen preference should contribute to the approach used. Halpern feels more resolute after this workshop that we should be concerned not only with individual momentary choice, but also with the choices we make collectively about what constitutes a good society.

Open Discussion

Goals: How Well-Being Should or Could be Used to Inform Policy

Although the use of well-being measures for monitoring purposes is important, Conal Smith (OECD) considered the emphasis to be rightly on the potential for well-being measures to shape policy development and appraisal. The demand for subjective well-being information is driven by policy. Rowlatt stressed that in the United Kingdom there is a keen sense of urgency and huge political interest in measures of well-being; it would be unfortunate to miss this moment of opportunity. Following the Stiglitz report, there was a groundswell of support across the OECD for this as well. Suzman expressed concern that given the differences of perspectives, implementing measures of well-being right now to make policy decisions may be premature and that the topic is best suited for intensive research until we better understand the different facets of wellbeing, their determinants and how they differ from economic wellbeing and consumer confidence.

However, Smith outlined three compelling reasons for pursuing the use of well-being measures to inform policy at this time. First, based on the literature it seems difficult to attribute differences in well-being to culture. The coefficient on what drives well-being is important to

policy, even if the aggregate measure does not change significantly over time. We need the ability to look at unrelated policy domains and collect some, even if noisy, empirical evidence of where government policy can make an impact relative to dollars spent. Second, both specific well-being measures tailored to policy decisions and higher level measures are important. The higher level measures, such as those used for monitoring, can be of value even if they are noisy. Third, there is a relatively low opportunity cost to implementing well-being measures at this time.

Understanding the mechanistic pathways for how well-being drives health might lead to interventions for improving well-being by preventing poor health. Another example is unemployment. There is a movement at the U.S. Department of Labor to measure the effectiveness of unemployment programs with the outcome measure of getting people employed. However, an equally compelling goal might be to mitigate the misery of the unemployed; currently we do not have a plan to include that or a way to measure it.

Perhaps well-being is not simply a passive response to a policy, but how policy can empower people to pursue the things that matter to them because that is what will lead to sustainable well-being. Huppert referenced the work of Brian Little, a Canadian psychologist, which has shown that there is a relationship between pursuit of personal projects and well-being. Sustainable well-being is related to our daily ability to engage in activities that matter to us. Based on his findings, Little concluded that the happiness of pursuit may be more important than the pursuit of happiness.

Wolfers cautioned against setting the bar for a national indicator of well-being higher than is done for GDP. The use of well-being measures for the purpose of monitoring is of value even if we do not understand everything that drives well-being. We do not understand everything that drives GDP yet it is still useful. Some are concerned that a well-being indicator will communicate something contradictory to other indicators when in fact we do want the well-being indicators to be able to tell a different story. There is no doubt that well-being moves; what we do not yet know is what it is about a country that makes well-being higher or lower. Well-being measures did not behave in a predictable way in terms of the recession; perhaps it is not that there is a problem with the indicator, but that it is telling us something different about this particular recession.

Methodological Issues

Workshop participants discussed several methodological issues. General concerns about how to include well-being measures include the time line, the impact of the messenger of the survey, question ordering, context effects, sampling design, language translation, data reduction, and the role existing Gallup measures can play.

From a pure methodological perspective, much more work is needed to identify the best possible set of questions for various purposes before implementation. However, many felt that this would cause undue delay and that there will be value in implementation now, even if the sets of measures evolve over time with additional research. Dolan asserted that there is a difference between relevance to policy makers and the underlying academic rigor. There is a general

consensus that we value three types of measures because they capture different but important domains of well-being: 1) evaluation (i.e., ladder of life, life satisfaction), 2) experienced life (i.e., day reconstruction method), and 3) eudemonics (i.e., meaning of life, life worthwhile). We have yet to resolve the types of measures needed and appropriate for specific policy purposes. For the purposes of cost-benefit analysis, we may only need the high-level questions. However, it is unlikely that the ladder-of-life measures will be responsive to policy interventions. We need more detailed individual assessments of certain areas on which specific policy interventions focus (i.e., understanding the well-being of people experiencing a specific health problem). The sampling will need to focus more around clusters and networks. It is important to consider measures for the purpose of monitoring, but we also need to discuss the need for more specificity in terms of measures for the purpose of informing policy.

Groves agreed that question ordering may be important; however, the reality is that the well-being questions will not be put consistently at the beginning of multiple surveys. The issue of context effects provides impetus for including the same set of well-being measures on multiple vehicles so that the context effects, which are inevitable, can be quantified.

Dolan offered that we should look at how well-being changes across a range of measures. Subjective well-being measures allow us to remove ourselves from the focusing effect question by simply asking about happiness, gathering the right-hand-side variables, and then regressing the variables to find out what is important. This would be an important contribution and would free us from the focusing effects that occur when we ask about specific aspects assumed to be important to people's lives.

The process of reducing the universe of measures of well-being for a particular construct or domain to a shorter set of common questions was discussed. Halpern suggested that a data reduction exercise and/or factor loading of the measures are important in order to settle on a short set of measures in a systematic way that is efficient in capturing the largest possible variance. Groves clarified that alternative constructs are not a matter of empirical analysis solely; constructs should be agreed upon by the brightest minds around the table in a process like the NIH consensus model. Different constructs would be considered for different purposes. The empirical task is to decide on the best set of items to measure the agreed upon constructs using IRT models to determine the best single- and two-item indicator of the construct. This work would begin with the harvesting of datasets.

Gale Muller (Gallup) shared information about the history of the Gallup poll, the well-being index, and the process of creating the measures used by Gallup. Gallup has developed several items to measure various aspects of well-being, tested items that do not work, crafted sets of measures with input from Kahneman and others, and continued to work on these measures. Gallup is constrained by the need for items that work world-wide across multiple languages. The implication is that its capability to drill down and focus on changes in smaller regions, such as the work Allin spoke about, is limited. However, in one case study it looked at particular regions within a country and was able to see changes in well-being in relation to future unrest, well in advance of a failed state. There was a healthy correlation between regional changes in well-being and future unrest. This finding provides additional support to the hypothesis that smaller areas

may reveal a great deal of information and enable us to see trends in well-being over time among sub-populations.

Gallup's well-being measures include the ladder of life and measures of affect; its items are in the public domain. Building the specific items is an art and a science and takes considerable time and energy. Gallup is eager to collaborate and share what it has learned thus far with academics and government representatives in its efforts to adapt well-being measures for the purpose of informing policy.

Wolfers raised the possibility of using measures from the Gallup-Healthways Well-Being Index³⁷ as they stand in order to maximize time and international comparability. Several participants had concerns about this strategy. Nielsen pointed out that there is ongoing discussion within Gallup to revise and improve its well-being measures; if we simply took what it has now, we would miss out on those enhancements. There is a danger in borrowing what already exists given that there is a developing consensus of other dimensions that need to be tapped. Landefeld cautioned that while using a product we already have would be helpful for trend growth and comparisons, the measures do not serve our purposes because they are too aggregated and too weighted to cultural and biological markers, which make them unresponsive to policy. Given the discussion about the importance of using targeted measures for policy, there is an opportunity cost to moving forward with existing measures. Muller agreed that measure development is an ongoing process and simply taking what Gallup has probably would not best serve policy purposes for individual countries. However, Gallup can offer a starting place. It will be important to look at the key items on the dashboard that are driving the overall numbers because those will be more important in terms of making policy decisions.

General Concerns

The main obstacle to having well-being inform policy is the lack of knowledge of what determines well-being, which is in glaring contrast to what we know about what determines income and employment. There is a need for robust causal models of determination of well-being. Significant progress on improving well-being will require a critical mass of multiple lines of research addressing the causal models.

Conal Smith argued that there is a closing window of opportunity for international collaboration to agree on a set of well-being measures that will allow for comparisons. Several countries are moving forward with this on the heels of the Stiglitz report. For example, Eurostat is making changes to the European Union Statistics on Income and Living Conditions (EU-SILC) survey for 2013. Although we may not want to tie our hands given it is a rapidly changing field, we may not have another opportunity to implement official cross-country measures that are comparable. We need to identify and implement core measures that are essential on the basis of what we know today. If 6 to 12 OECD countries put well-being measures in official statistics and they are all different, then it could be very difficult to change in the future. There is a lot of inertia and internal stakeholders involved.

³⁷ Information about the Gallup-Healthways Well-Being Index can be found at <http://www.well-beingindex.com/> (accessed March 16, 2011).

Allin disagreed somewhat with Smith's assertion. The EU-SILC plans are taking place outside of regulatory procedures, and at a recent international conference organized by the EU, several countries were looking to adopt the SILC measures. In general, however, it is important for us to understand how well-being measures will work for a particular country before we can have international measures. There is great value in learning from each other and understanding where comparisons can be made, but we need to ascertain the extent of the need for comparisons in the first place. Inertia is there in part for good reasons. The system of national accounts that exists now is not the same system used after World War II. There is a process for managing change, and we can expect that subjective well-being measures would fall into that process over time. The major benefit of the OECD's work is to provide information on what is going on internationally, especially on what works, and to determine the need for international collaboration.

Stevenson and Wolfers offered an example of a counterintuitive finding that points to the possibility that subjective data may tell a richer story than objective data. The well-being of women world-wide has fallen compared to that of men over the past 30 years, and there is no reasonable explanation for why this has happened. Conservatives blame women's movements or workforce entry. However, the same decline is seen in countries where women have not advanced as much as others. This trend is seen across all women, regardless of disaggregation. Abraham raised the possibility that women may be more empowered today and have higher expectations for their lives; because feelings of well-being are relative to expectations, it may not be surprising that women report lower well-being if their higher expectations are not being met.

Suzman confirmed that implementing both types of well-being measures in experimental modules in longitudinal studies is an appropriate step at this time, as is including measures in cross-sectional studies such as the American Time Use Survey (ATUS). He observed that harmonizing questions across surveys and countries to allow for comparative analyses has obvious value. However the resolution of many remaining issues must be accompanied by a strong research program on the measurement of the different types of well-being and their consequences for health and policy. The NIA has announced their first such competition in this area. Placed in the historical context of the development of the GDP and the more recent National Health Accounts, the NIA probably will need to continue for several decades its current well-being research initiative at both basic and applied levels.

Recommended Next Steps

Several workshop participants offered recommendations for next steps during the open discussion period that ranged from conservative actions designed to contribute to an evolving research agenda to more aggressive activities to implement well-being measures now while there is a window of opportunity politically. A number of specific recommendations were generally supported:

1. Continue dialogue and maintain a focus on this topic
2. Continue close collaboration between the NIA and ESRC
3. Form a CNSTAT panel on this topic that will issue a report within 1 year

4. Support efforts by the OECD and others to compile a universe of well-being measures being used world-wide
5. Continue supporting research on measuring well-being (e.g., longitudinal panel studies, testing questions in experimental modules, field experiments)

Recommendations that were offered but not necessarily supported by all participants are as follows:

1. Commit to building consensus on agreed-upon constructs and measures
2. Implement a set of well-being measures in national statistics, based on what we know now, in collaboration with international efforts to maximize comparability

APPENDIX 1

Workshop Agenda

Meeting Goals

This meeting will explore research needs and practical challenges surrounding the integration of measures of subjective well-being into the planning and evaluation of policies by local and national governments and agencies. The workshop will bring together leading academic and policy personnel from the United States and United Kingdom to examine the potential for measures of subjective well-being to inform the design and evaluation of specific social and economic programs.

November 8, Room 201

- 8:45 a.m. **Introduction**
- **Katharine Abraham, Chair**
- NIA
- ESRC
- (Breakfast available outside meeting room 201)
- 9:15 **Topic Overview.** An introduction to measurement approaches, highlighting developments in measurement, concerns about existing measures, and issues relevant to policy application. This introduction will frame the discussion and highlight some of the challenges in measurement as we focus on policy relevance and applications of this science.
- **Daniel Kahneman** – U.S. Perspective
- **Paul Dolan** – U.K. Perspective
- Discussants: **Jacqui Smith and Amanda Rowlett**
- 10:30 **Open Discussion**
- 11:10 Break
- 11:30 **Developing Policy Relevant Frameworks for Cost-Benefit Analysis, Pitfalls, Research, and Data Needs**
Mental health and well-being over the life course
- **Richard Layard**
- Discussant: **Amanda Sacker**
- 12:00 p.m. **Open Discussion**
- 12:30 Working Lunch to Discuss Afternoon Sessions (bring lunch back to room 201)

- 1:30 ***Policy Areas Potentially Informed by Well-Being Measures***
Health and Medical Care – evidence, challenges, potential pitfalls, research, and data needs
- **Andrew Steptoe**
- Discussants: **Susan Cartwright and Richard Frank**
- 2:00 ***Open Discussion***
- 2:45 Break
- 3:00 ***Policy Areas Potentially Informed by Well-Being Measures*** (cont.)
Social, Economic, and Environmental (employment, education, poverty, housing, marriage, crime, civic engagement, social support programs, commuting, pollution control)
- **Richard Lucas**
- Discussants: **Rebecca Blank** (employment), **Felicia Huppert** (education)
- 3:30 ***Discussion and Integration***
- 5:00 p.m. **ADJOURN**
-

November 9, Room 201

- 9:00 a.m. ***Summary and Directions.*** Implications of policy presentations on measurement needs, research and data agenda, policy maker-research interactions. Discussion will focus on (1) Challenges in measurement of subjective well-being at the population level, (2) Successes and pitfalls in applications of SWB measures for policy purposes, and (3) Research and data needs for the field.

(Breakfast available outside meeting room 201)

Opening Remarks (reflections on day 1; questions raised for day 2)

- **Paul Allin**
- **Somnath Chatterji**
- **David Halpern**
- **Bob Groves**
- **Steve Landefeld**
- **Christine McGuire**
- **Katherine Wallman**
- **David Weir**

- 11:00 Break

Open Discussion:

- What sort of subjective well-being measures exist that can be practically used (and are likely to have a general acceptance in the academic, policy making, and wider public spheres)?
- How can they be practically applied to public policy making?

- To what degree do we need to apply different notions and uses of subjective well-being in different policy areas (e.g., employment, social, environment policies), and why?
- What research is needed to advance this agenda?

(Lunch—boxed sandwiches available, and lunch tickets for the third floor Atrium)

1:00 p.m. **ADJOURN**

Appendix 2

List of Participants

Katharine G. Abraham, University of Maryland, College Park
Dorinda Allard, U.S. Department of Commerce
Paul Allin, U.K. Office for National Statistics
Rebecca Blank, U.S. Department of Commerce
Susan Cartwright, Lancaster University
Somnath Chatterji, World Health Organization
Sarah Curtis, University of Durham (by videoconference)
Paul Dolan, The London School of Economics and Political Science
Richard Frank, U.S. Department of Health and Human Services
Vicki Freedman, University of Michigan, Ann Arbor
Carol Graham, The Brookings Institution
Robert M. Groves, U.S. Census Bureau
David Halpern, U.K. Institute for Government
Brian A. Harris-Kojetin, U.S. Office of Management and Budget
Felicia A. Huppert, University of Cambridge
Jeannette L. Johnson, National Institute on Aging
Ruta Kadonoff, U.S. Department of Health and Human Services
Daniel Kahneman, Princeton University
Chandra Keller-Allen, Rose Li and Associates, Inc.
Christian Kroll, Harvard University
J. Steven Landefeld, U.S. Department of Commerce
Alfonso R. Latoni, National Institute on Aging
Richard Layard, The London School of Economics and Political Science
Ruth Lee, Research Councils U.K., U.S. Office
Rose Maria Li, Rose Li and Associates, Inc.
Richard E. Lucas, Michigan State University
Stuart Maudsley, National Institute on Aging
Christine McGuire, Department of Health (England)
Gale Muller, Gallup Corporation
Lis Nielsen, National Institute on Aging
Amanda Rowlatt, U.K. Department for Work and Pensions
Philip Rubin, Yale University
Amanda Sacker, University of Essex
Conal Smith, Organisation for Economic Co-operation and Development
Jacqui Smith, University of Michigan
Andrew Steptoe, University College London
Betsy Stevenson, U.S. Department of Labor
Cass R. Sunstein, U.S. Office of Management and Budget
Richard Suzman, National Institute on Aging
Antonio Terracciano, National Institute on Aging
Joy Todd, U.K. Economic and Social Research Council

Katherine K. Wallman, U.S. Office of Management and Budget

David R. Weir, University of Michigan, Ann Arbor

Justin Wolfers, University of Pennsylvania

National Research Council of the the National Academies Staff

Constance F. Citro, Director, Committee on National Statistics

Michael L. Cohen, Senior Program Officer, Committee on National Statistics

Daniel L. Cork, Program Officer, Committee on National Statistics

Robert M. Hauser, Interim Executive Director, Division of Behavioral and Social Sciences

Christopher D. Mackie, Study Director, Committee on National Statistics

Anthony S. Mann, Program Associate, Committee on National Statistics

Krisztina Marton, Senior Program Officer, Committee on National Statistics

Keiko Ono, Senior Program Associate, Committee on Population

Jane Ross, Director, Center for Economic, Governance and International Studies

Miron Straf, Deputy Executive Director, Division of Behavioral and Social Sciences

Barbara A. Wanchisen, Senior Board Director, Board on Behavioral, Cognitive, and Sensory Sciences