

Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes

Joshua A Salomon, Ajay Tandon, Christopher J L Murray, for the World Health Survey Pilot Study Collaborating Group

Department of Population and International Health, Center for Population and Development Studies, Harvard School of Public Health, 9 Bow Street, Cambridge, MA 02138, USA

Joshua A Salomon
assistant professor of international health

Harvard University Global Health Initiative, 104 Mt. Auburn Street, Cambridge, MA 02138, USA

Ajay Tandon
senior research associate

Christopher J L Murray
director

Correspondence to: J A Salomon
jsalomon@hsph.harvard.edu

BMJ 2004;328:258–61

Abstract

Objective To examine differences in expectations for health using “anchoring vignettes,” which describe fixed levels of health on dimensions such as mobility.

Design Cross sectional survey of adults living in the community.

Setting China, Myanmar, Sri Lanka, Pakistan, Turkey, and United Arab Emirates.

Participants 3012 men and women aged 18 years and older (self ratings); subsample of 406 (vignette ratings).

Main outcome measures Self rated mobility levels and ratings of hypothetical vignettes using the same questions and response categories.

Results Consistent rankings of vignettes are evidence that vignettes are understood in similar ways in different settings, and internal consistency of orderings on two mobility questions indicates good comprehension. Variation in vignette ratings across age groups suggests that expectations for mobility decline with age. Comparison of responses to two different mobility questions supports the assumption that individual ratings of hypothetical vignettes relate to expectations for health in similar ways as self assessments.

Conclusions Anchoring vignettes could provide a powerful tool for understanding and adjusting for the influence of different health expectations on self ratings of health. Incorporating anchoring vignettes in surveys can improve the comparability of self reported measures.

Introduction

Valid, reliable, and comparable measures of health are critical components of the evidence base for clinical practice and health policy. Clinical trials and national surveys rely heavily on self reported measures of health,^{1–5} but interpretation of these measures is complicated by lack of comparability when different people understand and respond to a given question in different ways. Paradoxical findings have been reported in many analyses of population health surveys, suggesting that self reported measures may be misleading without adjustment for these differences.^{6–9}

Distinguishing between differences in self ratings due to actual health differences and differences due to varying norms or expectations for health is a key challenge in interpreting self reported measures of health.^{10–11} Differing expectations for health can lead to differences in the levels at which people change from using one response category to the next—that is, differences in response category cut points. For example, a 90 year old man who struggles to climb the stairs might characterise himself as having “mild difficulties” in moving around, but a 40 year old man with the same mobility might describe himself as having “moderate difficulties.” These responses are not comparable because the individuals have different response category cut points for questions about mobility.

“Anchoring vignettes” are a new component of survey instruments that can be used in conjunction with extended statistical models to position self reported responses on a common interpersonally comparable scale. We describe an application of this strategy from a series of pilot studies for the world health survey.¹²

Methods

Components of the world health survey were pilot tested in 12 countries during May and June 2002, including six countries that tested the module on health measurement (China, Myanmar, Pakistan, Sri Lanka, Turkey, and the United Arab Emirates). Researchers selected a cross section of the adult population (≥ 18 years) in each country, with an emphasis on enlisting similar numbers of men and women and getting

Mobility questions in the world health survey pilot study

Q1 Overall in the past 30 days, how much difficulty did [you/name] have with moving around?

(a) none; (b) mild; (c) moderate; (d) severe; (e) extreme

Q2 In the past 30 days, how much difficulty did [you/name] have in vigorous activities, such as running 3 km or cycling?

(a) none; (b) mild; (c) moderate; (d) severe; (e) extreme

Mobility vignettes

Paul is an active athlete who runs long distance races of 20 km twice a week and plays soccer with no problems

Mary has no problems with walking; running; or using her hands, arms, and legs. She jogs 4 km twice a week

Rob is able to walk distances of up to 200 m without any problems, but feels tired after walking one km or climbing more than one flight of stairs. He has no problems with day to day physical activities, such as carrying food from the market

Anton does not exercise. He cannot climb stairs or do other physical activities because he is obese. He is able to carry the groceries and do some light household work

Louis is able to move his arms and legs, but requires assistance in standing up from a chair or walking around the house. Any bending is painful, and lifting is impossible

Vincent has a lot of swelling in his legs due to his health condition. He has to make an effort to walk around his home as his legs feel heavy

David is paralysed from the neck down. He is confined to bed and must be fed and bathed by somebody else

Names are included as examples only. Each site developed separate sets of locally appropriate male and female names, and interviewers presented the set of names matched to each respondent's gender. See bmj.com for more vignettes.



This is the abridged version of an article that was posted on bmj.com on 23 January 2004: <http://bmj.com/cgi/doi/10.1136/bmj.37963.691632.44>

Consistency of vignette orderings and average rank correlation coefficients by country. Results are shown for the five vignettes common to all six countries

Country	Proportion consistent* rankings (%)			Average rank correlation*		
	Q1 (v global)	Q2 (v global)	Q1 v Q2	Q1 (v global)	Q2 (v global)	Q1 v Q2
China	83	76	76	0.96	0.97	0.97
Myanmar	76	72	72	0.93	0.92	0.91
Pakistan	71	81	76	0.95	0.94	0.93
Sri Lanka	71	70	79	0.84	0.85	0.88
Turkey	70	62	67	0.89	0.87	0.87
United Arab Emirates	87	85	83	0.96	0.95	0.95
All countries	76	72	74	0.90	0.89	0.92

*Because multiple vignettes may be rated in the same category, complete orderings may not be observed. Consistency of rankings and Spearman's rank order correlation coefficients were therefore computed with ties resolved in favour of the consistent ordering.

enough representation at all ages and at different levels of income and education. Researchers completed face to face surveys with one respondent per household using a standardised questionnaire translated into the local language through defined protocols.¹²

The health module included a self assessment component consisting of one to three questions pertaining to each of 12 domains, along with 15 different anchoring vignettes per domain. In this paper, we focus on the domain of mobility as an example. An anchoring vignette is a description of a concrete level on a given domain that respondents evaluate with the same questions and response scales used for self assessments on that domain (box). Vignettes are fixed (by design) across respondents so that variation in categorical responses is attributable to differences in response category cut points. The key objective in this approach is to elicit ratings for hypothetical levels on a given domain that reflect individual norms and expectations for health in approximately the same way that the self ratings do for the individuals' own levels.

We examined distributions of self assessments and vignette ratings for the two mobility items in the survey, consistency of vignette orderings, and variation in vignette ratings across age groups, countries, and the two different mobility questions.

Results

A total of 3012 respondents completed the health survey. The mean age was 41 (standard deviation 15), with a range across countries from 33 (10) in the United Arab Emirates to 49 (15) in China. A total of 1837 (61%) respondents were younger than 45, and 478 (26%) had had less than 6 years of education (see also bmj.com). Self assessed mobility ratings varied considerably between countries, with 45% (249/555 in Sri Lanka) to 85% (431/510 in the United Arab Emirates) of respondents reporting no difficulties moving around. Of the 3012 respondents, 406 (13.5%) completed the version of the questionnaire that included mobility vignettes.

Evidence on consistency of vignette orderings across respondents and internal consistency within each individual's vignette ratings on the two mobility questions suggests that comprehension of the vignette rating task is good across all sites, and that a similar understanding of the levels described in the vignettes prevails (table and bmj.com). For the two global comparisons and the internal comparison, about three quarters of responses were completely consistent with an additional 18% to 22% having only one or two rank inconsistencies in each case.

The primary purpose of including anchoring vignettes linked to self assessments is to detect and then adjust for differences in response category cut points to make categorical self reports more comparable. As an example of how vignette ratings can reveal differences in cut points that may relate to varying norms and expectations for health, figure 1 shows the distribution of ratings for one mobility vignette in different age groups for the three countries that included this vignette (Myanmar, Pakistan, and Turkey). The youngest and oldest age groups differed significantly ($P=0.001$). This example suggests that older individuals use a more lenient interpretation of the same set of response categories in describing mobility levels, which is consistent with the notion of shifting norms for health over the life course.

When survey respondents rate a series of vignettes on a domain, we can summarise the responses in different groups using stacked bar diagrams. Comparisons of vignette ratings can reveal cut point differences within and between countries, show how cut points for the same person change over time, or place cut points for multiple questions relating to the same domain on a common scale. For example, figure 2 shows the ratings for an array of 10 vignettes using the two different mobility questions. This figure shows that the second question is "more difficult" in the sense of tapping a higher level of mobility than the first; that individuals rate themselves favourably on mobility but recognise on average that the top two vignettes

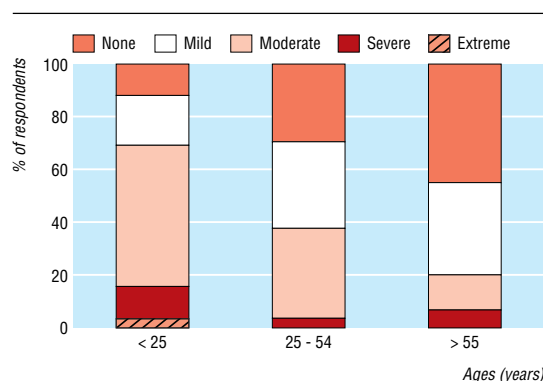


Fig 1 Variation in vignette ratings across age groups in three countries (Myanmar, Pakistan, and Turkey) (n=211). Responses are shown for the question, "[Rob] is able to walk distances of up to 200 m without any problems but feels tired after walking 1 km or climbing up more than one flight of stairs. He has no problems with day to day physical activities, such as carrying food from the market. Overall, how much difficulty does [Rob] have with moving around?"

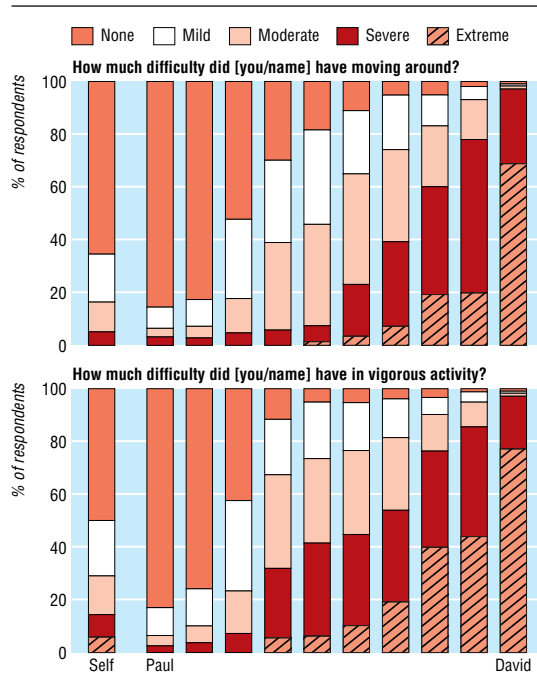


Fig 2 Self assessments and vignette ratings for two mobility questions (Q1 How much difficulty did [you/name] have with moving around? Q2 How much difficulty did [you/name] have in vigorous activities?). Pooled results are shown from six countries (China, Myanmar, Pakistan, Sri Lanka, Turkey, and United Arab Emirates) (n=3012 for self ratings, n=406 for vignettes)

describe higher levels than their own; and that respondents use the available categories similarly in providing self ratings and vignette ratings, suggested by the correspondence between the two questions on both the self assessments and vignette ratings—in both cases, individuals respond to the second question in a way that accords with tapping a higher level of difficulty. (See bmj.com for further examples.)

Discussion

Inclusion of anchoring vignettes in health surveys is part of an integrated strategy of instrument design and analysis to make self reported measures more comparable between individuals, communities, and populations.¹³ Anchoring vignettes may be applied to many different problems in which ordered categorical self report data are collected. This approach enables examination of systematic differences in categorical cut points between populations, within populations across different sociodemographic groups, or within individuals or groups over time. The anchoring vignette method also allows comparisons between different questions relating to a common domain, enabling the interpretation of responses to these related questions on a single underlying scale, providing a bridge between data collected using different instruments.

Our study shows that variation in vignette ratings for mobility can reveal differences in expectations for health—for instance, between different age groups. Formal statistical models have been introduced to allow anchoring vignette data to be used in adjusting self rated measures of health,^{14 15} but fundamental insights can be gained into differences in the use of

particular questions and their associated response categories by analysing distributions of vignette ratings, even before any models are applied. Anchoring vignettes have been developed for the world health survey for a range of different health domains, as well as for other areas that share similar methodological challenges, such as health system responsiveness and social capital. Although more work is needed to refine individual vignettes and identify those that work best, this study shows that the anchoring vignette strategy is feasible in a variety of settings and offers promise for more widespread application of the approach.

A number of limitations should be noted. Firstly, the sample size in this pilot study is small and cannot be assumed to represent general populations. Although we aim to show the types of empirical findings that are available through the use of anchoring vignettes, the data collected in the probability samples of the world health survey will allow further investigation on some of the questions that we raise. Cross validating the anchoring vignette approach will be useful—for example, using measured performance tests on selected health domains. Current understanding of the causes of differences in cut points is limited. Research on psychology and decision making has highlighted a range of biases and heuristics that shape responses to survey questions¹⁶; similar quantitative understanding of how different health expectations influence self perceptions of health and key correlates of these differences would aid interpretation of self reported measures of health.

Interest has been rising recently in the challenges of interpreting self assessments of health, relating to issues of perception versus observation and experi-

What is already known on this topic

Variation in perceptions of health and self assessments of health status may be related in part to different expectations for health

Standard methods for measuring health status do not distinguish changes in health from changes in expectations. Interpretation of self reported measures of health may be improved by using new methods that account for varying expectations

What this study adds

Application of a data collection strategy based on anchoring vignettes enables the investigation of different individual expectations for health and the adjustment of self reported measures of health to account for these differences

Empirical evidence from a multi-country survey study using the anchoring vignette strategy points to differences in health expectations across age groups and countries

By mapping responses to various questions on the same health domain to a common comparable scale, anchoring vignettes can provide a bridge between data collected using different instruments for measuring health status

ences versus expectations.⁸⁻¹⁰ As self assessments continue to play a central role in the measurement of health outcomes, including vignettes in national surveys and clinical research can improve the use of self reports by confronting important problems of interpersonal comparability.

We thank David Cutler and Gary King for useful discussions and Dan Hogan for help with research.

Contributors: See bmj.com

Funding: Analysis supported by National Institute on Aging (P01 AG17625).

Competing interests: None declared.

Ethical approval: Not needed.

- 1 Testa MA, Simonson DC. Assessment of quality-of-life outcomes. *N Engl J Med* 1996;334:835-40.
- 2 Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ* 1998;316:736-41.
- 3 Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H. Capturing the patient's view of change as a clinical outcome measure. *JAMA* 1999;282:1157-62.
- 4 Shibuya K, Hashimoto H, Yano E. Individual income, income distribution, and self rated health in Japan: cross sectional analysis of nationally representative sample. *BMJ* 2002;324:16-9.
- 5 Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ* 2002;324:1417.

- 6 Murray CJL, Chen LC. Understanding morbidity change. *Popul Dev Rev* 1992;18:481-503.
- 7 Mathers CD, Douglas RM. Measuring progress in population health and well-being. In: Eckersley R, ed. *Measuring progress: is life getting better?* Collingwood: CSIRO, 1998: 125-55.
- 8 Sen A. Health: perception versus observation. *BMJ* 2002;324:860-1.
- 9 Sadana R, Mathers CD, Lopez AD, Murray CJL, Iburg KM. Comparative analysis of more than 50 household surveys of health status. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, eds. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization, 2002.
- 10 Carr AJ, Gibson B, Robinson PG. Measuring quality of life: is quality of life determined by expectations or experience? *BMJ* 2001;322:1240-3.
- 11 Freedman VA, Martin LG. Understanding trends in functional limitations among older Americans. *Am J Public Health* 1998;88:1457-62.
- 12 World Health Organization. World Health Survey. www.who.int/whs (accessed 6 Jan 2004).
- 13 Murray CJL, Tandon A, Salomon JA, Mathers CD, Sadana R. Cross-population comparability of evidence for health policy. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, eds. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization, 2002.
- 14 Tandon A, Murray CJL, Salomon JA, King G. Statistical models for enhancing cross-population comparability. In: Murray CJL, Evans DB, eds. *Health systems performance assessment: debates, methods and empiricism*. Geneva: World Health Organization, 2003: 727-46.
- 15 King G, Murray CJL, Salomon JA, Tandon A. Enhancing the validity and cross-population comparability of measurement in survey research. *Am Polit Sci Rev* 2004;98:567-83.
- 16 Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science* 1981;211:453-8.

(Accepted 13 November 2003)

doi 10.1136/bmj.37963.691632.44

Perinatal outcome of singletons and twins after assisted conception: a systematic review of controlled studies

Frans M Helmerhorst, Denise A M Perquin, Diane Donker, Marc J N C Keirse

Abstract

Objective To compare the perinatal outcome of singleton and twin pregnancies between natural and assisted conceptions.

Design Systematic review of controlled studies published 1985-2002.

Studies reviewed 25 studies were included of which 17 had matched and 8 had non-matched controls.

Main outcome measures Very preterm birth, preterm birth, very low birth weight, low birth weight, small for gestational age, caesarean section, admission to neonatal intensive care unit, and perinatal mortality.

Results For singletons, studies with matched controls indicated a relative risk of 3.27 (95% confidence interval 2.03 to 5.28) for very preterm (<32 weeks) and 2.04 (1.80 to 2.32) for preterm (<37 weeks) birth in pregnancies after assisted conception. Relative risks were 3.00 (2.07 to 4.36) for very low birth weight (<1500 g), 1.70 (1.50 to 1.92) for low birth weight (<2500 g), 1.40 (1.15 to 1.71) for small for gestational age, 1.54 (1.44 to 1.66) for caesarean section, 1.27 (1.16 to 1.40) for admission to a neonatal intensive care unit, and 1.68 (1.11 to 2.55) for perinatal mortality. Results of the non-matched studies were similar. In matched studies of twin gestations, relative risks were 0.95 (0.78 to 1.15) for very preterm birth, 1.07 (1.02 to 1.13) for preterm birth, 0.89 (0.74 to 1.07) for very low birth weight, 1.03 (0.99 to 1.08) for low birth weight, 1.27 (0.97 to 1.65) for small for gestational age, 1.21 (1.11 to 1.32) for caesarean

section, 1.05 (1.01 to 1.09) for admission to a neonatal intensive care unit, and 0.58 (0.44 to 0.77) for perinatal mortality. The non-matched studies mostly showed similar trends.

Conclusions Singleton pregnancies from assisted reproduction have a significantly worse perinatal outcome than non-assisted singleton pregnancies, but this is less so for twin pregnancies. In twin pregnancies, perinatal mortality is about 40% lower after assisted compared with natural conception.

Introduction

There is a widespread belief that pregnancy outcome is substantially worse after assisted than after natural conception.¹⁻³ The difference, however, relates predominantly to the higher frequency of multiple pregnancies.³ The first indication that assisted singleton pregnancies may also have poorer outcomes appeared in 1985,² but it was not clear how much related to assisted reproduction or to confounders, such as maternal age and parity. Several matched cohort studies have since confirmed these findings.¹⁻⁴⁻⁸

Department of Obstetrics, Gynaecology and Reproductive Medicine, Leiden University Medical Center, NL 2300 Leiden, Netherlands
Frans M Helmerhorst
associated professor
Denise A M Perquin
registrar
Diane Donker
medical officer

Department of Obstetrics, Gynaecology and Reproductive Medicine, Flinders University and Flinders Medical Centre, Adelaide, South Australia
Marc J N C Keirse
professor

Correspondence to: F M Helmerhorst
f.m.helmerhorst@lumc.nl

BMJ 2004;328:261-4

P+ Nine tables of detailed results and a list of excluded studies can be found on bmj.com

ELPS This is the abridged version of an article that was posted on bmj.com on 23 January 2004: <http://bmj.com/cgi/doi/10.1136/bmj.37957.560278.EE>