

35. Eriksson I, Unden AL, Elofsson S. Self-rated health. Comparisons between three different measures. Results from a population study. *International Journal of Epidemiology* 2001;30:326-33.
36. Appels A, Bsoma H, Grabauskas V, Gostautas A, Sturmans F. Self-rated health and mortality in a Lithuanian and a Dutch population. *Social Science and Medicine* 1996;42:681-9.
37. Finch BK, Hummer RA, Reindl M, Vega WA. Validity of self-rated health among Latino(a)s. *American Journal of Epidemiology* 2002;155:755-9.
38. Ferraro KF, Kelley-Moore JA. Self-rated health and mortality among black and white adults: examining the dynamic evaluation thesis. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* 2001;56:S195-205.
39. Jylha M, Guralnik JM, Ferrucci L, Jokela J, Heikkinen E. Is self-rated health comparable across cultures and genders? *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences* 1998;53:S144-52.
40. Macintyre S, Hunt K, Sweeting H. Gender differences in health: are things really as simple as they seem? *Social Science and Medicine* 1996;42:617-24.
41. Gilmore AB, McKee M, Rose R. Determinants of and inequalities in self-perceived health in Ukraine. *Social Science and Medicine* 2002;55:2177-88.
42. McDonough P, Walters V. Gender and health: reassessing patterns and explanations. *Social Science and Medicine* 2001;52:547-59.
43. Matthews S, Manor O, Power C. Social inequalities in health: are there gender differences? *Social Science and Medicine* 1999;48:49-60.
44. Robine JM. Measuring the burden of disease. *Lancet* 1998;352:757-8.
45. Morrow RH, Hyder AA, Murray CJ, Lopez AD. Measuring the burden of disease. *Lancet* 1998;352:1859-61.
46. Anand S, Hanson K. DALYs: efficiency versus equity. *World Development* 1998;26:307-10.
47. McKee M, Shkolnikov V. Understanding the toll of premature death among men in eastern Europe. *BMJ* 2001;323:1051-5.
48. Weitz E. The heroic man and the ever-changing woman: gender and politics in European communism, 1917-1950. In: Frader L, Rose J, editors. *Gender and class in modern Europe*. Ithaca (NY): Cornell University Press; 1996. p. 311-52.
49. Service R. *A history of twentieth-century Russia*. London: Allen Lane; 1997.
50. Sperling V. The "New" sexism: images of Russian women during the transition. In: Field MG, Twigg JL. *Russia's torn safety nets: Health and social welfare during the transition*. New York: St. Martin's Press; 2000:173-89.
51. Velkoff VA, Kinsella K. Russia's aging population. In: Field MG, Twigg JL. *Russia's torn safety nets: Health and social welfare during the transition*. New York: St. Martin's Press; 2000:231-50.
52. Aneshensel CS, Rutter CM, Lachenbruch PA. Social structure, stress and mental health: competing conceptual and analytical models. *American Sociological Review* 1991;56:166-78.
53. Chervyakov VV, Shkolnikov VM, Pridemore WA, McKee M. The changing nature of murder in Russia. *Social Science and Medicine* 2002;55:21-32.

Commentary

Towards valid and comparable measurement of population health

Colin D. Mathers¹

In the past two decades, considerable international effort has been put into the development of summary measures of population health that integrate information of mortality and non-fatal health outcomes (1). During the past 20 years, disability-free life expectancy (DFLE) and related measures have been calculated for many countries using self-report survey data on disability and health status (2). In this issue of the *Bulletin* (pp. 778–787), Andreev et al. compare DFLE for Russian men and women and also compare DFLE estimates for the Russian Federation with other Eastern European countries. Their results are dominated by a phenomenon seen in many developed countries: women generally report worse health than men. Thus, the large male–female gap in life expectancy in the Russian Federation is offset by worse reported health status in women.

The reporting by women of worse health, generally, than men has been seen in health surveys across many developed countries (2). Can we conclude that the health status of Russian Federation women is worse than that of Russian Federation men? Several paradoxical findings have been reported in analyses of population health surveys, suggesting that self-reported health measures may give misleading results if differences in the way people use question responses are not taken into account (3–6). This evidence has been ignored by many who use self-report-

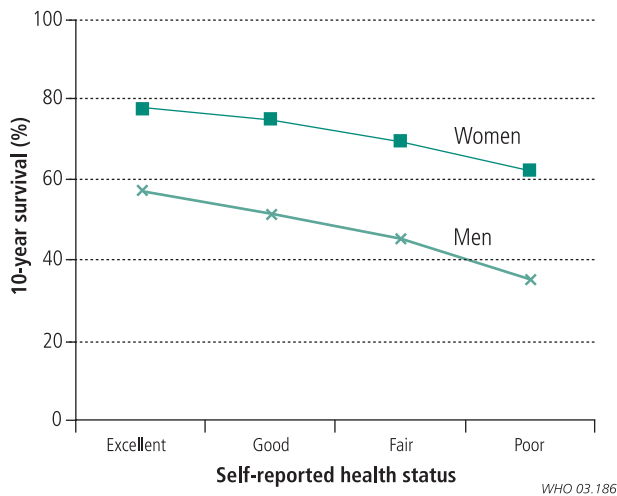
survey measures of health status to report on population health, health inequalities, or intervention outcomes. Indeed, there is substantial literature arguing that within-group correlations of self-reported health measures with other observed or measured health indicators, or with mortality risk, show the validity and comparability of such measures across groups (7–12).

Although there are, undoubtedly, correlations between self-reported health status measures and other health indicators, and there is no doubt that health status influences self-report, this does not ensure comparability of self-report measures across groups. Several studies have reported significant correlations between perceived health (with response categories such as excellent, good, fair, poor) and mortality risk within groups such as men and women, or groups defined by socioeconomic or ethnic characteristics (8, 9, 12), and argued that these correlations provide evidence that self-perceived health is a valid measure of health status. Similar arguments are made for within-group correlations with observed or measured functional indicators, with morbidity and health service utilization (7, 10, 11).

However, it is possible to have consistent associations of perceived health with survival within groups without such associations holding across groups (6). This is illustrated in Fig. 1, where survival is lower for worse perceived

¹ Coordinator, Epidemiology and Burden of Disease, Global Programme on Evidence for Health Policy, World Health Organization, Geneva, Switzerland (email: mathersc@who.int).

Fig. 1. Plots showing that within-group correlation of self-reported health status with other health indicators does not provide evidence for cross-group correlations



health in both men and women, while at the same time the survival of women with worst perceived health is better than that of men with excellent perceived health. Suppose that a population survey found most women reporting worse health than men for a population with the associations shown in Fig. 1. It would clearly be fallacious to deduce that women have worse survival (or health) than men: the indicator is not comparable across groups because women are using the response categories differently to men.

Survey developers have emphasized the importance of establishing the validity of instruments and their reliability, but

until recently, little attention had been paid to the issue of cross-population comparability (6). The latter relates fundamentally to unmeasured differences in expectations and norms for health, so that the meaning that different populations attach to the labels used for response categories in self-reported questions, such as mild, moderate, or severe, can vary greatly. Recent developments in survey methodology using measured tests and anchoring vignettes to calibrate self-report health questions hold considerable promise in addressing this problem (13).

Anchoring vignettes are short descriptions that mark fixed levels of ability (e.g. for people with different levels of mobility such as a paraplegic person or an athlete who runs 4 km each day). Survey respondents are asked to rate the vignettes for a health domain using the same question and response categories as for their self-report on their own level of health, allowing the calibration and comparison of the self-report responses. Results from the WHO Multi-country Survey Study (MCSS) carried out in 2000 and 2001 in 61 countries provides clear evidence that different populations, and groups within populations, use response categories differently to describe the same health states (14). Although the MCSS found that overall, women have somewhat worse health than men, this difference was much smaller than that reported here for the Russian Federation by Andreev et al. based on unadjusted self-report data.

Valid, reliable, and comparable measures of the health states of individuals are essential components of the evidence base for health policy. They are crucial for the measurement of health outcomes in clinical trials and the development of summary measures of population health. A strategy of including vignettes in national health surveys and clinical research may contribute to improving the interpersonal and cross-population comparability of these measures. ■

References

- Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization; 2002.
- Robine JM, Jagger C, Mathers CD, Crimmins EM, Suzman RM, editors. *Determining health expectancies*. Chichester: John Wiley & Sons; 2003.
- Johansson SR. Measuring the cultural inflation of morbidity during the decline of mortality. *Health Transition Review* 1992;2:78-89.
- Murray CJL, Chen LC. Understanding morbidity change. *Population and Development Review* 1992;18:481-503.
- Sen A. Health: perception versus observation. *BMJ* 2002;324:860-1.
- Sadana R, Mathers CD, Lopez AD, Murray CJL, Moesgaard-Iburg K. Comparative analysis of more than 50 household surveys of health status. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization; 2002.
- Hulka BS, Wheat JR. Patterns of utilization. The patient perspective. *Medical Care* 1985;23:438-60.
- Idler EL, Kasl SV, Lemke JH. Self-evaluated health and mortality among the elderly in New Haven, Connecticut, and Iowa and Washington counties, Iowa, 1982-1986. *American Journal of Epidemiology* 1990;131:91-103.
- McCallum J, Shadbolt B, Wang D. Self-rated health and survival: a seven-year follow-up study of Australian elderly. *American Journal of Public Health* 1994;84:1100-5.
- Chandola T, Jenkinson C. Validating self-rated health in different ethnic groups. *Ethnicity & Health* 2000;5:151-9.
- Manor O, Matthews S, Power C. Self-rated health and limiting longstanding illness: inter-relationships with morbidity in early adulthood. *International Journal of Epidemiology* 2001;30:600-7.
- Heistaro S, Jousilahti P, Lahelma E, Vartiainen E, Puska P. Self-rated health and mortality: a long term prospective study in eastern Finland. *Journal of Epidemiology and Community Health* 2001;55:227-32.
- Murray CJL, Tandon A, Salomon JA, Mathers CD. New approaches to enhance cross-population comparability of survey results. In: Murray CJL, Salomon JA, Mathers CD, Lopez AD, editors. *Summary measures of population health: concepts, ethics, measurement and applications*. Geneva: World Health Organization; 2002.
- Mathers CD, Murray CJL, Salomon JA, Sadana R, Tandon A, Lopez AD, et al. Healthy life expectancy: comparison of OECD countries in 2001. *Australian and New Zealand Journal of Public Health* 2003;27:5-11.