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), Andreew 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 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), Ref. No. 03-006437
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 than men for a population with the associations shown in Fig. 1. It would clearly be fallacious to deduce that women

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