Health systems' responsiveness and reporting behaviour: Multilevel analysis of the influence of individual-level factors in 64 countries

Nicole Valentine [a, b, *], Emese Verdes-Tennant [b], Gouke Bonsel [a]

[a] Department of Public Health, Erasmus University Rotterdam, Postbus 2040, Rotterdam, The Netherlands
[b] World Health Organization, Ave. Appia 20, CH1211 Geneva, Switzerland

ARTICLE INFO

ABSTRACT

Health systems' responsiveness encompasses attributes of health system encounters valued by people and measured from the user's perspective in eight domains: dignity, autonomy, confidentiality, communication, prompt attention, social support, quality of basic amenities and choice. The literature advocates for adjusting responsiveness measures for reporting behaviour heterogeneity, which refers to differential use of the response scale by survey respondents. Reporting behaviour heterogeneity between individual respondents compromises comparability between countries and population subgroups. It can be studied through analysing responses to pre-defined vignettes — hypothetical scenarios recounting a third person's experience in a health care setting. This paper describes the first comprehensive approach to studying reporting behaviour heterogeneity using vignettes. Individual-level variables affecting reporting behaviour are grouped into three categories: (1) sociodemographic, (2) health-related and (3) health value system. We use cross-sectional data from 150 000 respondents in 64 countries from the World Health Organization's World Health Survey (2002–03). Our approach classifies effect patterns for the scale as a whole, in terms of strength and in relation to the domains. For the final eight variables selected (sex; age; education; marital status; use of inpatient services; perceived health (own); caring for close family or friends with a chronic illness; the importance of responsiveness), the strongest effects were present for education, health, caring for friends or relatives with chronic health conditions, and the importance of responsiveness. Patterns of scale elongation or contraction were more common than uniform scale shifts and were usually constant for a particular factor across domains. The dependency of individual-level reporting behaviour heterogeneity on country is greatest for prompt attention, quality of basic amenities and confidentiality domains.

© 2015 Published by Elsevier Ltd.

1. Introduction

Progress in the health sector can be evaluated by changes in population health. But different monitoring frameworks also acknowledge the importance of other indicators. For example, in 2000 the World Health Organization (WHO) estimated additional measures of people-centredness of a health system, termed responsiveness, and fairness in financial contributions. The results were widely debated at the time. However, the underlying framework was generally accepted (Anand, 2003). Recent global policy debates have again shifted the focus onto actionable indicators of health systems performance in addition to health, such as coverage, out-of-pocket expenditures and people-centredness of the health system (Evans et al., 2013).

Responsiveness encompasses the acceptability of service provision with reference to the way users are treated and the environment within which they treated. Encompassing a range of issues that are important to people, it has also been referred to as responding to a population's "legitimate expectations" regarding the characteristics of an acceptable service (Murray and Frenk, 2000; page 719). Responsiveness contributes to satisfaction, well-being, and human dignity (Gostin et al., 2003). It also has instrumental value for achieving other objectives such as improving treatment success rates (Anhang Price et al., 2014; Shiferaw et al., 2013; Zimlichman et al., 2013).

The measurement of constructs like responsiveness commonly

* Corresponding author. World Health Organization, Department of Public Health, Environmental and Social Determinants of Health, Ave. Appia 20, CH1211 Geneva, Switzerland.
E-mail addresses: valentine_nicole@hotmail.com, valentinen@who.int (N. Valentine).

requires surveys with questions for individuals with recent health service experiences (e.g. Measure Demographic and Household Survey (DHS) Antenatal Client Exit Interviews: http://www.measuredhs.com/; Consumer Assessment of Health Providers and Systems (CAHPS)). Aggregation of individuals' responses to these questions provides a user-based assessment of quality. The World Health Organization has a large publicly available dataset on responsiveness from the World Health Survey (WHS) (2002–2003). The WHS responsiveness questions covered eight 'domains', named as follows (the short name appears in parenthesis where applicable): involvement in decision-making ('autonomy'), respectful treatment ('dignity'), clear communication ('communication'), confidentiality of personal information ('confidentiality'), choice of provider ('choice'), prompt attention, quality of basic amenities, access to family and community support ('social support'). The questions on domain performance use ordinal response scales with the verbal response categories, "very good" (1), "good" (2), "moderate" (3), "bad" (4), "very bad" (5).

A central challenge to comparing responsiveness survey results is known as 'differential scale use' or 'reporting behaviour heterogeneity' (Murray et al., 2001). It refers to the differential use of the response scale by respondents, unrelated to the object measured. Reporting behaviour heterogeneity consists of a random and systematic component. If the non-random component is related to the comparator of interest (e.g. culture, socioeconomic status), it compromises comparability across countries and within countries (Rice et al., 2012). Such measurement error is not unique to responsiveness. It is also found in other subjective measures of latent constructs, like happiness, well-being or self-reported health. If significant, adjustment or rescaling of people's evaluations may be required for comparisons within and across countries.

Differential scale use arises for different reasons. Social norms may cause some populations to avoid extreme expressions of approval or disapproval. Better education may increase comprehension and judgement abilities and the corresponding differentiation of the respondent between the verbal response categories ("very good" to "very bad"). Several studies have found that education is an important determinant of reporting behaviour heterogeneity (Rice et al., 2012; Sirven et al., 2012). Other factors that affect judgement may relate to familiarity with health problems and services (Kok et al., 2012) or age (Hargraves et al., 2001). Differential use of the response scale may also arise from differing 'expectations' associated with social status (de Silva, 2000) as demarcated by income or ethnicity (e.g. Farley et al., 2011). Individuals accustomed to poorer quality experiences could have lower expectations, causing their judgements to be less harsh, even when receiving worse quality services. Another mechanism affecting scale use is related to the importance an individual assigns to an attribute of the care process. Importance fixes attitudes more firmly, creating a stronger sense of expected norms, which has been shown to change the range of scale used (Groves, 1999; Jasso, 2006).

There are different ways of addressing reporting behaviour heterogeneity. Stratification of data by particular social groups without standardization tends to ignore the problem related to scale use but can focus on views of sicker patients as "bellwethers for how well health care systems are working" (Blendon et al., 2003; page 106). Standardizing stratified results by personal characteristics found affecting patient assessments of their own experiences in regression analyses is an implicit adjustment procedure. This has also been referred to as patient-mix adjustment. It has been used in reporting by AHRQ on CAHPS surveys (Hargraves et al., 2001) and as part of the Picker Survey methodology (Holzer and Minder, 2011). But again, even if patient-mix adjustment is used, it can purge valid disparities or inequities in health care responsiveness.

An increasingly common approach to adjustment, is to characterize reporting behaviour based on respondents' answers to a separate set of questions from those concerning the respondent's experiences. These are called vignette questions. Whereas performance or assessment questions ask users how they evaluate the health systems' responsiveness during their own experiences, vignettes describe hypothetical, reference health care situations that a third person is experiencing and vignette questions request users to evaluate these situations. Vignette questions use the same response scale as performance questions and provide researchers with specific information, distinguishing scale use from actual patient experiences. This information can be used to adjust the survey respondent's rating of their own experience through standardization or other techniques (parametric or non-parametric). Vignette-based adjustment procedures have early 20th century roots: in 1948 the physicist S. L. Anderson introduced a technique for scoring slubs in a wool yarn on a five-point scale, using four reference specimens to adjust the scores of judges for systematic scoring heterogeneity (Cox, 2011). In a similar vein the WHO surveys of the Multi-Country Survey Study and WHS introduced the equivalent of reference specimens known as 'anchoring vignettes'.

The WHS vignettes have been successfully tested for eliciting a common understanding across respondents (vignette equivalence) (Rice et al., 2011; Rice et al., 2012). The data have been used to publish papers quite widely (King et al., 2013; Szwarcwald et al., 2010), which include studies on their psychometric properties (Valentine et al., 2010).

Vignettes have subsequently also been used in non-WHO surveys (e.g. SHARE 2006–07, Wisconsin Longitudinal Study) and in several specialized procedures that adjust for systematic reporting behaviour heterogeneity where their use was found to improve comparability in both WHO and non-WHO surveys (King et al., 2004; Grol-Prokopczyk et al., 2011).

This paper has the explicit goal of analysing vignettes from the World Health Survey in order to characterize how a broad range of individual-level factors affect respondents' reporting behaviour. In general, previous studies reporting adjustment procedures with anchoring vignettes have taken account of a limited range of individual-level factors from the potential array of factors to be considered. Their usual focus has produced results with fewer applications for practitioners wanting to compare responsiveness within countries to further quality improvements. The data analysed in this paper cover a *wider* range of individual-level factors than used previously in a single model, for *all* responsiveness domains, while retaining a model structure that allows for country-level effects as done elsewhere. It aims to contribute to a general adjustment framework, and associated reporting standards within countries by characterizing different aspects of reporting behaviour heterogeneity, including how to describe the observed systematic influences on the use of reporting scale. If our study reveals that individual-level variables are important, then adjustments of responsiveness results within countries is needed to ensure local comparability. This knowledge and explicit characterization of reporting behaviour patterns, will also improve the validity of between country comparisons of responsiveness.

2. Methods

2.1. Survey organization and questionnaire

The World Health Survey was a household-based survey administered in 71 countries in 2002/03 with datasets finalized in 2004/5. The surveys used stratified, multiple cluster, designs. Ethical approval was obtained from an independent ethics review