

# Health systems performance: how can we secure robust international comparison?

Nigel Rice, Silvana Robone, Peter C. Smith

## Summary

A fundamental concern in making international comparisons in public services is the extent to which population expectations, norms and cultural characteristics influence respondent ratings of performance. Where these vary across otherwise comparable groups of individuals they are likely to obfuscate attempts at objective comparison. In the presence of subjective ratings of performance, this study considers methods that can be used to secure robust international comparison of health system responsiveness. In summary, we find:

- variation in the way respondents interpret and use the response categories available in survey data and that this variation can be explained in part by income and education and to a lesser degree age and gender;
- substantial variation in the reporting of responsiveness across countries which is associated with large differences in reporting behaviour, reflecting cultural norms and expectations;
- that accounting for differences in the use of response categories across countries markedly influences the ranking of country performance.

This document describes briefly the concept of responsiveness, the data analysed and methods used, and sets out the policy implications.

## Background

International comparison of performance has become one of the most influential levers for change in the provision of public services. However, international comparison is intrinsically difficult. Perhaps one of the most challenging areas in which to secure meaningful comparison is the notion of health system 'responsiveness'. This can be defined as the extent to which health services are aligned with user preferences in domains such as patient autonomy, choice and quality of amenities and can be seen as an elaboration of the more widely discussed concept of 'user satisfaction'.

Data on responsiveness is categorical and self-reported giving rise to the possibility, both within - and more notably - across countries, that individuals differ in their subjective rating of a fixed level of performance. We make use of data taken from the World Health Survey to investigate methods aimed at enhancing cross-country comparisons of health system performance. In addressing this aim we:

- interrogate individual-level survey data on health service responsiveness to inform cross-national comparison of health system performance;
- examine how individuals vary in their use of response categories available in surveys when answering questions on performance and how this differs across populations and population sub-groups;

This research was funded by the Economic and Social Research Council (ESRC) under the Public Services Programme, grant number RES-166-25-0038 and grant number RES-060-25-0045. We would like to thank the World Health Organization for providing access to the World Health Survey.

The views expressed are those of the authors and not necessarily of the funders.



**Table 1: Examples of domain questions and vignettes used in the World Health Survey**

---

<b>Communication – item: clarity of communication</b>	<b>Respondent own rating:</b> How would you rate: 1 – how clearly health care providers explained things to you? 2 – the time you get to ask questions about your health problems or treatment? Example vignette: [Rose] cannot write or read. She went to the doctor because she was feeling dizzy. The doctor didn't have time to answer her questions or to explain anything. He sent her away with a piece of paper without telling her what it said. Q1: How would you rate her experience of how clearly health care providers explained things to her? Q2: How would you rate her experience of getting enough time to ask questions about her health problem or treatment?
---	---

---

- explore the use of information contained in responses to a set of questions on vignettes as a method to adjust survey reports of health system performance;
- assess the impact on cross-country comparison of adjusting subjective survey reports of performance.

### **What is health system responsiveness?**

The concept of responsiveness as a measure of health systems performance was developed and promoted by the World Health Organisation (WHO). It relates to a health system's ability to respond to the legitimate expectations of potential users about non-health enhancing aspects of care. Together with health improvement and fairness of financial contribution, it has been suggested as an intrinsic goal of health system performance (Murray and Frenk, 2000). The concept of responsiveness covers a set of non-clinical and non-financial dimensions of quality of care that reflect respect for human dignity and interpersonal aspects of the care process (Valentine et al., 2009). These are measured across eight domains chosen to reflect the goals for health care processes and systems valued highly by individuals in their contact with health systems. The domains are: autonomy, choice, clarity of communication, confidentiality of personal information, dignity, prompt attention, quality of basic amenities and access to family and community support.

### **What data do we use?**

Perhaps the most ambitious attempt to date to measure and compare health systems responsiveness is the World Health Survey. The World Health Survey was an initiative launched by the WHO in 2001 aimed at strengthening national capacity to monitor critical health outputs and outcomes through the fielding of a comparable household survey instrument. Seventy countries participated in the World Health Survey with samples drawn from nationally representative

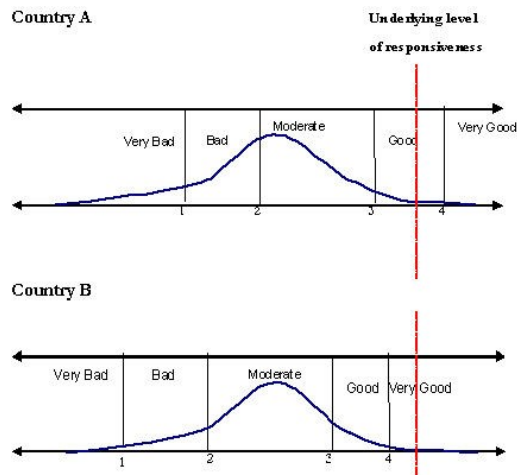
frames resulting in sample sizes of between 600 and 10,000 respondents per country. Our analysis exploited the survey module on responsiveness. The survey questionnaire asks respondents to rate their most recent (in the previous year) experience of contact with the health system within each of the eight domains.

These questions are supplemented by a set of anchoring vignettes for each domain. Anchoring vignettes offer objective descriptions of the experiences of hypothetical individuals when accessing health services. Respondents are asked to rate both their own experiences and the experiences described by the vignettes on the categorical scale: 'very good', 'good', 'moderate', 'bad' and 'very bad'. Table 1 provides an example of a survey question in the domain 'Clarity of communication' together with an example vignette.

### **What are our methods?**

It is natural to think of good or poor performance as having different meanings for different people. Accordingly, individuals may differ in their subjective ratings of any given objective level of performance. This is a source of differential reporting behaviour and is a phenomenon likely to be particularly pronounced when comparing performance across countries.

Differences in reporting behaviour can be shown diagrammatically using the example in Figure 1. Assume individuals in country A and country B are asked to rate the responsiveness of their health systems according to the above five-point scale and that the positioning of the threshold which individuals apply to separate the categories are indexed 1 to 4. Underlying (true) responsiveness in the two countries is identical, as depicted by the dashed vertical line; however, variation in reporting behaviour results in respondents in country A applying a different set of thresholds to the underlying response scale compared to respondents in country B. This reflects differences in expectations and norms.



**Figure 1: Example of differential reporting behaviour**

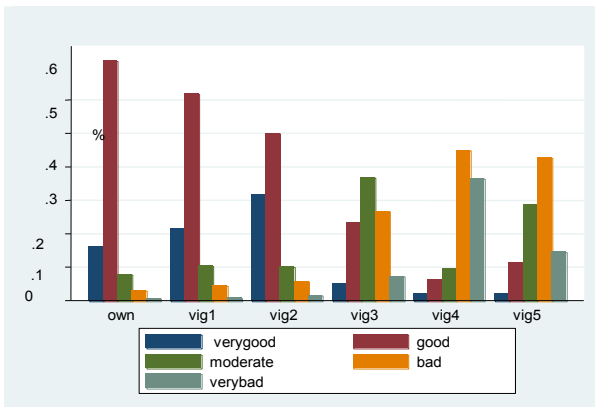
A casual inspection of the ratings in the two countries would suggest that individuals in country A face poorer health system responsiveness compared to individuals in country B.

Our research makes use of what has been termed the HOPIT model, which exploits information provided by anchoring vignettes to account for systematic variation in reporting behaviour. Given that the vignettes are fixed and pre-determined, observed variation across individuals in the rating of the vignettes can be attributed to differences in reporting behaviour. Figure 2 illustrates the issue by presenting individuals' reports of their own experiences of contact with the health system together with ratings of five vignettes in the domain 'clarity of communication'. It is notable that there is no unanimity amongst individuals in how they rate the vignettes, even though each vignette describes a fixed level of responsiveness. This provides prima-facie evidence of the existence of variation in reporting behaviour.

Information provided by the set of vignettes allows us to investigate the determinants of reporting behaviour as a function of individual characteristics such as age, gender, income, education and the country of the respondent. Knowledge of such characteristics can then be used to adjust a respondents' self-reports of experiences with health services to produce ratings purged of reporting behaviour. Anchoring reporting behaviour to a common scale (some chosen benchmark country) allows us to produce more comparable cross-country rankings of health system responsiveness.

### What are our findings?

Our findings illustrate that systematic variation in reporting behaviour exists in the World Health Survey both across individuals within countries and more prominently across countries. Results indicate that reporting behaviour is related to individual characteristics of income and education and to a lesser extent age and gender.



**Figure 2: Variation in reporting behaviour (Clarity of communication)**

**Table 2: Ranking of system responsiveness pre and post adjustment for variation in reporting behaviour (domain of Dignity)**

Rank	Raw Frequencies	Anchored Frequencies
1	Austria (61.9%)	Finland (55.1%)
2	Denmark (61.0%)	Denmark (54.6%)
3	Sweden (55.8%)	Sweden (54.5%)
4	Czech Rep. (52.9%)	Belgium (42.9%)
5	UK (51.4%)	France (40.3%)
6	Greece (51.0%)	UK (39.9%)
7	Finland (49.3%)	Netherlands (38.8%)
8	Hungary (47.8%)	Uruguay (35.6%)
9	France (47.6%)	Czech Rep. (32.2%)
10	Ireland (45.7%)	Estonia (28.5%)

Substantial variation in the reporting of responsiveness is observed across countries, in part associated with large differences in reporting behaviour. Correcting for reporting behaviour using the anchoring methodology has a marked impact on the ranking of country performance. Table 2 provides an example and compares the frequencies of reporting 'very good' responsiveness observed in the raw data with those obtained once adjustment for variation in reporting behaviour has been undertaken using the HOPIT model. The table shows the ranking of countries both before and after adjustment for a group categorised as 'high' using the Human Development Index (defined by the United Nations Development Program). The HOPIT model anchors responses to a chosen benchmark, in this case Mexico, so that reporting behaviour is related to a single common scale.

The results show considerable movement in the ranking of certain countries once adjustment for reporting behaviour has been undertaken. It is notable, for example, that Austria, ranked first when comparing raw frequencies of reporting 'very good' responsiveness, moves out of the top ten once reporting behaviour is benchmarked against that observed in Mexico. Indeed only six of the original top ten countries remain in the top ten once adjustment has taken place.

## Discussion

Data on public sector performance is often categorical and self-reported, giving rise to the possibility of differential reporting behaviour, both within and, more notably, across countries. The information provided by responses to vignettes offers the possibility of adjusting such data for variations in reporting behaviour and anchoring to a common scale, offering a more comparable basis on which to undertake cross-country analyses of performance. A comparison of the relative performance of countries obtained through the use of raw data compared with data adjusted for variation in reporting behaviour shows considerable differences, confirming the importance of such adjustments.

We have therefore concluded that the use of anchoring vignettes is a potentially valuable tool for promoting comparability in cross-country analyses. International comparison continues to be a key instrument for initiating policy change. However, for comparisons to be credible the possibility that individuals from diverse cultural backgrounds may vary in their reporting of objective levels of performance needs to be accommodated. Only after such differences have been accounted for can secure and persuasive comparative analyses take place. Our study demonstrates that the anchoring vignette approach offers a feasible and reliable approach towards placing subjective rankings on a consistent basis.

## What does this mean for policymakers?

Our results also provide national and international organisations and policymakers with a deeper understanding of the magnitude of potential variations in performance of public services across countries. The substantial changes in rankings that occur when adjusting for differences in reporting behaviour suggest that simple cross-national surveys of subjective ratings should be treated with considerable caution. They indicate that serious consideration should be given to incorporating vignettes into international surveys of subjective assessments of public services, not necessarily limited to health care.

Furthermore, our research has brought to light systematic reporting variations within countries, based on characteristics such as educational status. This suggests that the vignette approach also has potential applicability *within* countries when investigating geographical or social class variations in user experiences. Consideration should therefore be given to including vignettes in future UK attitudinal surveys, particularly those interested in identifying variations in user experiences across different social groups.

## References

- Murray, C.J.L. and Frenk, J. (2000) A framework for assessing the performance of health systems. *B World Health Org.*, 78, 717-731.
- Valentine, N.B., Prasad, A., Rice, N., Robone, S. and Chatterji S. (2009) Health systems responsiveness - a measure of the acceptability of health care processes and systems In: *Performance measurement for health system improvement: experiences, challenges and prospects* (eds P. Smith, E. Mossialos, S. Leatherman). London: WHO European Regional Office.

Further details of this research can be found in: Rice, N., Robone, S., Smith, P.C. (2009) International comparison of public sector performance: The use of anchoring vignettes to adjust self-reported data. *Evaluation*; 16(1): 81-101