IRDES

Document de travail
Working paper

Comparability of Health Care Responsiveness in Europe Using anchoring vignettes from SHARE

Nicolas Sirven (Irdes)
Brigitte Santos-Eggimann (Iumsp)
Jacques Spagnoli (Iumsp)

DT n° 15

September 2008

Institut de recherche et documentation en économie de la santé

IRDES - Association Loi de 1901 - 10 rue Vauvenargues - 75018 Paris - Tél. : 01 53 93 43 00 - Fax : 01 53 93 43 50 - www.irdes.fr

Comparability of Health Care Responsiveness in Europe

using anchoring vignettes from SHARE*

Nicolas Sirven†, Brigitte Santos-Eggimann‡, Jacques Spagnoli§

Abstract

The aim of this paper is to measure and to correct for the potential incomparability of responses to the SHARE survey on health care responsiveness. A parametric approach based on the use of anchoring vignettes is applied to cross-sectional data (2006-07) in ten European countries. More than 6,000 respondents aged 50 years old and over were asked to assess the quality of health care responsiveness in three domains: waiting time for medical treatment, quality of the conditions in visited health facilities, and communication and involvement in decisions about the treatment. Chopit models estimates suggest that reporting heterogenity is influenced by both individual (socio-economic, health) and national characteristics. Although correction for differential item functioning does not considerably modify countries ranking after controlling for the usual covariates, about two thirds of the respondents' self-assessments have been re-scaled in each domain. Our results suggest that reporting heterogenity tends to overestimate health care responsiveness for 'time to wait for treatment', whereas it seems to underestimate people's self-assessment in the two other domains.

Keywords: Anchoring Vignettes, Cross-Country Comparison, Chopit Model

JEL Classification Codes: I11, C81, C42

*This study is part of the COMPARE project. The authors are grateful to Arthur van Soest, Hendrick Jürges, and Renske Kok for usefull comments and suggestions.

†*Corresponding author*, Research Fellow, (PhD), IRDES (Institute for research and information on health economics) 10, rue Vauvenargues. 75018 Paris (France). Tel. +33 (0)1 53 93 43 00. Fax. +33 (0)1 53 93 43 50. Email: sirven@irdes.fr

‡Associate Professor (MD, Dr PH, MPH), IUMSP (Institute of Social and Preventive Medicine, University of Lausanne) 52, route de Berne. CH-1010 Lausanne. Tél. +41 21 651 01 24. Fax. +41 21 6510139. Email: Brigitte.Santos-Eggimann@chuv.ch

§Research Fellow (MsC), IUMSP (Institute of Social and Preventive Medicine, University of Lausanne) 52, route de Berne. CH-1010 Lausanne. Tél. +41 21 651 01 24. Fax. +41 21 6510139. Email: Jacques.Spagnoli@chuv.ch

1 Introduction

Cross-national evaluations of health care systems have long relied essentially on indicators of expenditure (e.g. proportion of Gross National Product invested in health) and health (e.g. life expectancy at birth, level of avoidable mortality, subjective health), with health considered as the main, if not only, outcome. More recently, health care responsiveness, or the extent to which the process of care delivery matches patients' expectations (Murray *et al.*, 2004), was added as an important criterion for evaluating health care systems and specific indicators were integrated in the WHO World Health Report 2000 (WHO, 2000). Health survey respondents are now increasingly asked about their experience of access to care. A range of dimensions have been identified for the responsiveness concept: respect of autonomy, confidentiality, dignity, prompt attention, communication, social consideration, as well as the quality of basic amenities, the choice and continuity of care, are measured in the world health survey (WHO, 2002).

The evidence of differences in responsiveness of national health care systems (Coulter & Jenkinson, 2005; Schoen *et al.*, 2004) points to a potential for improvements in the quality of health care delivery. Geographic variations must, however, be interpreted with caution due to the subjectivity of questions on satisfaction with access to health care. Expectations and satisfaction with access may be influenced by a range of socio-demographic or health variables, which distribution is not equivalent across countries and consequently necessitate some adjustment. Responses to subjective questions are also likely to reflect a different 'response style' in regions or countries, reflecting historical, cultural or environmental circumstances (Hausdorf *et al.*, 2008).

The vignettes method, designed for correcting subjective responses and thus avoiding the effects of a response style bias, is currently applied in WHO surveys (King *et al.*, 2004). The Survey of Health, Aging and Retirement in Europe (SHARE) drawn inspiration from these experiences and used vignettes in the first wave (2004) as a tool to take into account the effect of national variations in response style regarding subjective health (cf. Jürges, 2007) and work disability. In the second wave (2006-07), new vignettes were developped in order to compare people's expectations towards the health care system.

The aim of this exploratory study is to compare health care responsiveness in ten countries participating in SHARE (wave 2), adjusting for a range of socio-demographic and health factors, to describe the effect of correction for differences in response styles using a set of vignettes, and to investigate such differences at the aggregated level.

2 Data

This study explores the potential incomparability of responses on health care responsiveness using cross-section of individual-level data from the Survey on Heath, Ageing and Retirement in Europe. SHARE has been developed on the basis of prior successful experiments which are the Health and Retirement Study (HRS) in the United States, and the English Longitudinal Survey of Ageing (ELSA). This European bi-annual longitudinal survey aimed at carrying out international comparisons and analysis of economic

and social problems related to ageing.[1] Alongside with SHARE, the COMPARE project collects survey data based on the anchoring vignettes method (King *et al.*, 2004) in order to create internationally comparable measures of several dimensions of the quality of life (of which health care responsiveness is one dimension). Notice that this study is based on preliminary SHARE wave 2 raw data released in the first trimester of 2008 (v0.1). Analyses were conducted on all subjects aged 50 year old and over in 2007 who participated in SHARE wave 2 and its vignettes supplement in ten countries, including questions on health care responsiveness and corresponding vignettes. The overall sample size amounted to 6,233 individual respondents. It ranged from 365 (France) to 1,034 (Germany) at the country level. The proportion of men was 45.1% across countries and the mean age was 65 years.

Questions and vignettes on health care responsiveness derived from three of the eight dimensions defined by the WHO for population health surveys: waiting time for medical treatment, quality of the conditions in visited health facilities, and communication and involvement in decisions about the treatment. They were part of a self-administered drop-off questionnaire filled after completion of the SHARE main interview. For each dimension, subjects responded to one general question reporting their own experience and then provided their opinion for one vignette presenting the specific situation of an hypothetical individual. Self-assessment of questions and vignettes follows a five items scale from very good to very bad ('conditions of the health facilities' and 'communication about treatment') or from very short to very long ('time to wait').

Respondents aged 65 and over evaluated a second vignette for each dimension of health care responsiveness. It appears that only 25.4% of them could rank the two vignettes for 'conditions of the health facilities' in a consistent way. This figure is slightly higher for 'time to wait for treatment' (34.2%) and for 'communication about treatment' (54.6%), but remains quite low anyway. In other words, the oldest sub-sample of the population (who were the ones asked to assess two vignettes) may have experienced difficulties in understanding the vignettes since they reported inconsistently ordered responses. King & Wand (2002:5) argue that "differences between hypothesized ordering of the researchers and the consensus ordering may fruitfully be used for diagnosing problems in the survey instruments, particularly when translating the questions for use in different languages." An alternative hypothesis is that multiple vignettes may be correctly interpreted by respondents but generate response inconsistencies when they describe different circumstances (e.g. waiting for care in case of acute versus chronic conditions), or different subdimensions. Owing to this consideration, we choose to analyse, for each of the three dimensions of health care responsiveness, the one single vignette available for the whole COMPARE sample of subjects aged 50+. Models estimations were carried out on pooled countries responses to benefit from statistical properties of large samples.

Socio-demographic and health variables used for adjustment were abstracted from the SHARE main interview. They included the usual covariates like country dummies and some usual individual characteristics such as gender, age, and the level of education. Three more specific variables have been retained in the analysis in order to take into account the influence of health conditions on health care satisfaction. They are (i) a

[1] For further details, cf. Börsch-Supan & Jürges (2005) and www.share-project.org

dummy based on the Euro-d scale (Prince *et al.*, 1999; Dewey & Prince, 2005) which provides a standard measure of the symptoms of depression, (ii) a dummy indicating if the respondent has difficulties in Katz' basic activities of daily living (ADL) or Lawton's instrumental activities of daily living (IADL), and (iii) a dummy taking the value 1 if the respondent reported he ever smoked daily and drank alcohol within the last 3 months.

Additional data on public expenditures on health *per capita* (USD at 2000 PPP rates) between 2000 and 2006 are taken from OECD Health Data 2008. An average measure of these was computed since figures are not always available for all countries at all times. We shall look at the simple correlation of this variable with an aggregated variable of health care satisfaction, respectively made out of (i) potentially incomparable, and (ii) comparable responses. This should indicate if the use of vignettes signficantly reduces response heterogeneity between countries.

3 Method

3.1 The model

This study is based on King *et al.* (2004:192) who have designed "a method of directly measuring the incomparability of responses to survey questions, and then correcting for it." The core idea being that if respondents evaluate in a different way the same hypothetical situation (vignette), this is evidence of response scale differences—i.e., differential item functioning (DIF). To correct for DIF with a parametric approach, the standard ordered probit model (or oprobit) for health care ratings is extended to a joint compound hierarchical probit model (or chopit). The main differences are that in the latter model (i) vignettes provide information about a common reference to self assessed questions, and (ii) thresholds for responses to both self-assessed and vignettes questions may vary by country, individual characteristics, or health conditions, etc. As King *et al.* (2004:197) put it "[i]n broad outline, our model can be thought of as a generalization of the commonly used ordered probit model, where we model DIF via threshold variation, with the vignettes providing the key information." In the detail, the chopit model consists of two components.[2]

First, the self-assessment response y_i for respondent i is modelled as an ordinal probit model with underlying response

$$y_i^* = x_i^{'}\beta + \epsilon_i$$

where x_i are covariates, β are fixed effects and $e_i \backsim N(0,1)$ is a residual error term. The observed responses $k = 1, ..., 5$ are generated via threshold model with person-specific thresholds τ_i^k

$$y_i = k, \text{ if } \tau_i^{k-1} \leq y_i^* < \tau_i^k$$

where $-\infty = \tau_i^0 < \tau_i^1 < ... < \tau_i^5 = +\infty$. The thresholds are modelled as

$$\tau_i^1 = \alpha^{1'} x_i$$

$$\tau_i^k = \tau_i^{k-1} exp(\alpha^{k'} v_i), \text{ with } k = 2, 3, 4$$

[2]The formal presentation of the model heavily draws on Rabe-Hesketh *et al.* (2002).

where v_i are covariates, and α^k the parameters to estimate. Here the underlying response y_i^* can be interpreted as the true perceived health care responsiveness of respondent i, on a scale that is comparable across individuals.

Second, the vignettes response z_{ij} for respondent i to vignette j (where $j = 1$ in our case) is assumed to only differ from θ_j (the 'true perception' of the hypothetical person described in the vignette) by a random term $u_{ij} \backsim N(0, \sigma^2)$ with underlying response

$$z_i^* = \theta_j + u_{ij}, \text{ with } j = 1 \text{ vignette here}$$

where e_i and $u_i j$ are assumed to be independent of each other and of the covariates. Notice that the parametric approach can be applied with only one vignette. As King & Wand (2007:14) put it "we do not require that each respondent give unique answers to each vignette in the set or that all respondents rank the vignettes in the same order. We only need assume that respondents understand the vignettes on a common scale apart from random perceptual, response, and sampling error. The ties and inconsistencies that result from these types of errors violate no assumptions of our methodology." In addition to this 'vignette equivalence' assumption[3], we assume 'vignette consistency' by applying the same thresholds in the vignette component as in the self-assessment component, i.e.

$$z_{ij} = k, \text{ if } \tau_i^{k-1} \leq z_i^* < \tau_i^k$$

where $-\infty = \tau_i^0 < \tau_i^1 < ... < \tau_i^5 = +\infty$. Here again the thresholds are modelled as

$$\tau_i^1 = \alpha^{1'} x_i$$

$$\tau_i^k = \tau_i^{k-1} exp(\alpha^{k'} v_i), k = 2, 3, 4$$

The same covariates are used to both determine self-assessment and vignette thresholds and the observed responses to health care self-assessment, so that $v_i = x_i$. Once the chopit model is estimated for each domain of health care, the analysis of cross-national differences can be based either on the comparison between ordered probit and chopit estimates, or on counterfactual simulations derived from the chopit model. The latter method is especially usefull to investigate the influence of aggregated variables on people's self-assessment.

3.2 Counterfactual simulations

For each country's sample, the estimated chopit models are applied to the overall SHARE wave 2 sample (25,862 individuals) to obtain predictions.[4] In the detail, four counterfactual distributions of self-assessed responses were simulated, based upon the DIF corrected estimates and using respectively:

1. The country's own parmeters in self-assessment and threshold equations (this is similar to the sample distribution);

[3]The assumption of vignette equivalence means that there is a 'true' (objective) actual level of health care responsiveness underlying each hypothetical situation described in the vignettes. In other words, the domain levels represented in each vignette are understood in the same way by all respondents, irrespective of their country of residence or other socio-demographic variables.

[4]Comparison of vignettes sample with the overall SHARE population (wave 2) do not indicate important differences in standard socio-economic characteristics (age, gender, education, etc.). This suggest that our method may be subject to minor selection bias.

2. The country's own parmeters in self-assessment, but using the benchmark country's parameters (e.g. Germany's) for the thresholds (i.e. giving everyone in different countries the thresholds of a similar person in Germany). This distribution is the counterfactual of "if everyone in the sample would understood the question as the Germans." It thus gives people's self-assessment without cross-country reporting heterogeneity;

3. The benchmark country's self-assessment parameters, but the own country's thresholds;

4. The benchmark country's self-assessment parameters and thresholds.

Choosing Germany as the benchmark (since it has the largest sample of people responding to the vignettes) means the four counterfactual distributions are the same for this country. For other countries, however, 1 – 2 shows the effect of DIF (with reference to Germany); whereas 2 – 4 and 1 – 3 give two different ways of looking at the genuine differences in health care responsiveness in the country under study and Germany (cf. Kapteyn, Smith & van Soest, 2007). Notice that cross-country differences in 2 – 4 are due to the change in the influence of country dummies in self-assessment, so they can be analysed with regard to national macroeconomic contexts. We found in particular that change in people's self-assessment (2 – 4) is correlated with public expenditures on health. As a consequence, we decided to investigate differences in health care responsiveness (with and without DIF correction, i.e. counterfactuals 1 and 2) with respect to macro data on health care (as a proxy for 2 – 4). All estimations are based on Rabe-Hesketh's *et al.* (2002) program using the GLLAMM procedure for Stata. All analyses were performed using Stata software (StataCorp., 2005).

4 Results

4.1 Model estimates

In order to test for DIF and to facilitate comparisons with previous research on this topic, Tables 1 and 2 respectively report ordered probit and chopit estimations for health care responsiveness. By and large, results from the former model confirm standard findings in the literature (Rice, Robone & Smith, 2008). For instance, Table 1 indicates that, without DIF correction, age is significantly associated with better ratings in any of the three domains of health care, suggesting that people have less expectations about the system as life expectancy is reducing. Health status is also an important determinant of self-assessed questions in the oprobit models since respondents with depression symptoms (Euro-d) systematically report worse health care responsiveness, whatever the domain is. Parameter estimates of country dummies in Table 1 indicate that most countries (apart from Belgium and the Czech Republic) have higher rates of respondents unsatisfied with health care responsiveness as compared to Germany.

Tables 1-2 about here

Results from chopit models (Table 2) are quite different from oprobit ones since age remains significant only for 'conditions of the health facilities', and neither gender nor education level are significant anymore. This indicates that correcting for cultural

differences plays an important role for these covariates. More precisely, the effect of age, gender, and education on responsiveness appears to be over-estimated when reporting heterogeneity is not taken into account. On the contrary, Euro-d is the only individual characteristic to be significant in both oprobit and chopit estimations for any health care domain—meaning that depression symptoms affect people's rating independently from cultural differences. After adjusting for reporting heterogeneity, we still observe a negative and significant effect of depressive symptoms, though reduced for the three domains and item combinations. This indicates that the positive relationship between depression and responsiveness is over-estimated if reporting heterogeneity in response style by Euro-d is not accounted for. It is noticeable that thresholds covariates in the chopit models generally appear to have some explanatory power and hence improve the overall fit of the model.

Changes in countries ranking in health care responsiveness are given by comparing country dummies coefficients between oprobit and chopit. Looking at estimates for 'time to wait for treatment' indicates a good stability for both the most satisfied countries (Belgium, the Czech Rep., and France) and the less satisfied (Poland). Only Denmark accounted for a scaling up in four ranks (from the 8th position to the 4th).[5] Comparable results from the two other domains were observed. Countries where health care responsiveness is believed to be lower remain the same for 'conditions of the health facilities' (The Czech Rep., Italy, Poland), and 'communication about treatment' (Italy, France, Poland) after DIF correction. Appart from Spain—that has been re-scaled up by four ranks in the Chopit model for 'conditions of the health facilities'—the structure of cross-country self-assessed responsiveness is more or less the same before and after reporting heterogeneity is taken into account. By and large, correction for DIF does not considerably modify country rankings in health care responsiveness.

Although more detailed differences between oprobit and chopit estimates can be derived straight away from coefficients value and p-value in Tables 1 and 2, for simplicity reasons, we propose to focus on cross-country comparisons based on simulated data derived from the estimated models.

4.2 Systems responsiveness and public spendings

With the aim to evaluate people's appreciation of national health care systems, cross-country comparisons are often based on models estimates or on predictions derived from these models—e.g. the share of respondents in country m reporting 'conditions of the health facilites' are good or very good. However, neither cross-country average ratings of each domain nor the share of respondents reporting good or very good health care

[5]Different standards in this particular level of responsiveness explain why the ranking of Denmark is affected when the chopit model is estimated instead of the ordered probit model. In the estimation of the first threshold parameter of 'time to wait for treatment', (τ^1) the country dummy for Denmark is found to be 0.199 and significant. This means that the Danes have a lower standard for high ratings in this domain of health care. On the other hand, the country dummies for Denmark for thresholds 2 and 3 are negative and significant, which means that Danes have a higher standard for what constitutes these levels of responsiveness than German respondents. Notice that the dummy for threshold 4 is not significant, which means that for the worst ratings in health care, Danes have the same standards as German. Applying the same standards to everyone thus consists in re-scaling Danish respondents in the five categories according to the influence the modelled thresholds.

responsiveness, are necessarily good indexes since a large part of the people may rate the system as 'moderate'. Because it comprises responses without a clear-cut position, this middle category is often difficult to interpret—especially when researchers try to figure out if respondents are overall satisfied or not.

In order to provide a better index of health care responsiveness, one could derive for each country a 'balance statistic' (Theil, 1952) from the share of people reporting (very) good health care responsiveness minus the share of those who rate the system as (very) bad.[6] In theory, values for this index are bounded between -100 (all respondents consider the system is bad or very bad) and +100 (everyone believes the system is good or very good). Cross-country correlations between public health expenditures and health care responsiveness could both help investigating (i) the determinants of health care quality or satisfaction at the macro level, and (ii) the power of DIF for explaining international differences in the level of satisfaction with health care.

Figures 1-3 about here

Figures 1 to 3 indicate a positive correlation between government spending in health *per capita* and the balance statistic for health care responsiveness. These figures suggest that 'Conditions of the health facilities' is the health care domain people consider as the most satisfied with. Figure 1 indicates that the balance statistic only displays negative values for most countries in the case of 'time to wait for treatment'; meaning that most respondents are unsatisfied with this aspect of care in all countries except Germany, Belgium, or—after DIF correction—the Czeck Rep. Although correction for DIF (counterfactual 2) does not meaningfully modify country rankings (according to the values of the balance statistics) for any domain of health care under study, it leads to better R^2 values (Figures 1 and 2) and higher coefficient slopes (Figure 2 and 3) than counterfactual 1. The raise of the slope in Figures 2 and 3 suggests that, on average, people appear more sensitive to public spendings (the more, the better) than what we could have thought if reporting heterogenity were not accounted for. On the contrary, a comparable slope but a decrease in the constant in Figure 1 suggests that DIF could overestimate people's satisfaction with self-assessment in 'time to wait for treatment'.

Table 3 about here

Gaining confidence from the previous results requires to rely on statistics based upon a larger set of observations. Since the balance statistic is made out of aggregated data, two-sample proportion tests in each domain could only be applied respectively to the share of people reporting (very) good health care responsiveness, and to the share of those who rate the system as (very) bad. Table 3 reports the overall differences between counterfactual distributions 1 – 2. For instance, correction for DIF leads to a significant decrease in the share of people satisfied with 'time to wait for treatment' from 20.9% to 17%. Since the share of unsatisfied simultaneously climbs from 35.5% up to 39.5%, it seems that reporting heterogeneity does tend to overestimate people's satisfaction with 'time to wait for treatment'. Applying the same reasoning to 'conditions of the health facilities' and 'communication with doctors' indicates, on the contrary, that DIF tended to underestimate these two domains of health care. These simple tests support the results provided in Figures 1 to 3.

[6] For a detailed discussion and a survey on empirical methods on the evaluation of people's satisfaction, see Pesaran & Weale (2006)

5 Conclusion

This study used anchoring vignettes to produce internationally comparable data in three domains of individuals' self-assessment of health care responsiveness (the extent to which the process of care delivery matches patients' expectations). About two thirds of the respondents' self-assessments have been re-scaled in each domain—just like if they all had the same understanding of the questions and the same values, cultural beliefs, etc. Chopit models estimates suggest that reporting heterogenity (i) is more prominent in some countries compared to others in Europe (Germany being the benchmark country), varies across health care domains and across individuals within each country, and (ii) can be explained to some extent by both individual (socio-economic) and national characteristics. The use of counterfactual distributions of individual's responses to health care responsiveness helped investigate the genuine difference (i.e. without DIF) between countries. The positive correlation between the amounts of public health expenditures *per capita* and health care satisfaction is generally stronger after correction for DIF. Although we have to keep in mind that these results at the macro level should be taken with great care since they are based on small samples, confidence in these results can be gained from simple proportion tests. They suggest that reporting heterogenity tends to overestimate health care responsiveness for 'time to wait for treatment', whereas DIF seems to underestimate satisfaction with 'conditions of the health facilities' and 'communication with doctors' in Europe. The general results are encouraging. More surveys using anchoring vignettes in different countries, on a regular time-base, are a necessary evolutionnary step to make meet research issues and public health policies in Europe.

References

Börsch-Supan A. & Jürges H. (2005) Eds. The Survey of Health, Aging, and Retirement in Europe-Methodology. Manheim Research Institute for the Economics of Ageing (Germany).

Coulter A. & Jenkinson C. (2005). European patients' views on the responsiveness of health systems and healthcare providers. *European Journal of Public Health*, 15:355–360.

Dewey M.E. & Prince M.J. (2005). "Mental health" in Boersch-Supan, A. *et al.* (Eds.), Health, ageing and retirement in Europe: first results from SHARE. Manheim, Research Institute for the Economics of Ageing, pp.108–117.

Hausdorf K. *et al.* (2008). Rating access to health care: are there differences according to geographical region? *Australian and New Zealand Journal of Public Health*, 32:246–249.

Jürges H. (2007). True health differences vs. response style: exploring cross-country differences in self-reported health. *Health Economics*, 16:163–178.

Kapteyn A., Smith J.P. & van Soest A. (2007). Vignettes and self-reports of work disability in the United States and The Netherlands. *American Economic Review*, 97: 461–473.

King G., Murray C., Salomon J. &. Tandon A. (2004). Enhancing the validity and crosscultural comparability of measurement in survey research. *American Political Science Review*, 98(1): 567–583.

Kristensen N. & Johanson E. (2008). New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Economics*, 15(1):96–117.

Murray C. & Frenk J. (2000). A framework for assessing the performance of health systems. *Bulletin of the World Health Organization*, 78: 717–731

Pesaran M. & Weale M. (2006). Survey expectations. in Elliott, G., Granger, C.W.J., & Timmermann, A. (Eds.). Handbook of Economic Forecasting. North-Holland. (1): 715–776.

Prince M.J. *et al.* (1999). Development of the Euro-d scale – a European Union initiative to compare symptoms of depression in 14 European centres. *British Journal of Psychiatry*, 174: 330–338.

Rabe-Hesketh S., Skrondal A. & Pickles A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2: 1–21.

Rice N., Robone S. & Smith P.C. (2008). The measurement and comparison of health system responsiveness. University of York (U.K.), HEDG Working Paper 08/05, mimeo.

Schoen C. *et al.* (2004). Primary care and health system performance: adults' experiences in five countries. *Health Affairs*; W4:487–503.

StataCorp. (2005). Stata Statistical Software: Release 9.2. College Station, TX: StataCorp LP.

Theil H. (1952). On the time shape of economic microvariables and the Munich Business test. *Revue de l'Institut International de Statistique*, 20.

World Health Organization (2000). The World health Report 2000. Health systems: improving performance. Geneva: WHO.

World Health Organization (2001). Background paper for the technical consultation on responsiveness concepts and measurement. Geneva: WHO, 13–14 September 2001.

Table 1: OPROBIT Estimates for health care responsiveness

	Time to wait		**Conditions**		**Communication**	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Woman	-0.042	0.029	-0.060**	0.030	-0.022	0.029
Age	-0.009***	0.002	-0.010***	0.002	-0.010***	0.002
Education	-0.045**	0.015	-0.021	0.015	-0.032**	0.015
Euro-D	0.106**	0.033	0.161***	0.035	0.177***	0.034
ADL or IADL	0.128***	0.030	0.047	0.032	0.043	0.031
Smoke & drink	0.088**	0.034	0.043	0.036	0.039	0.035
Germany	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sweden	0.631***	0.065	0.407***	0.067	-0.013	0.066
Netherlands	0.204**	0.065	0.163**	0.069	-0.023	0.067
Spain	0.680***	0.067	0.563***	0.069	0.270***	0.067
Italy	0.586***	0.056	1.145***	0.059	0.533***	0.057
France	0.129*	0.068	0.331***	0.071	0.271***	0.069
Denmark	0.709***	0.051	0.225***	0.053	-0.025	0.052
Belgium	-0.292***	0.056	0.088	0.060	0.154**	0.058
Czechia	0.035	0.050	0.651***	0.053	0.176**	0.051
Poland	1.050***	0.061	1.026***	0.062	0.440***	0.061
Constant 1	-2.208***	0.141	-1.721***	0.145	-2.006***	0.143
Constant 2	-1.206***	0.139	0.077***	0.144	-0.390***	0.141
Constant 3	0.098***	0.138	1.339***	0.145	0.776***	0.141
Constant 4	1.274***	0.139	2.223***	0.150	1.724***	0.146
N	6043		6013		6023	
Log likelihood	-7627.9		-6408.1		-6990.4	

Significance levels: * p<0.10 ** p<0.05 *** p<0.01

Table 2: CHOPIT estimates for health care responsiveness

	Time to wait		**Conditions**		**Communication**	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Self-assement						
Woman	-0.060	0.040	-0.137**	0.046	-0.033	0.044
Age	-0.003	0.002	-0.009***	0.003	-0.002	0.002
Education	0.026	0.020	0.006	0.023	-0.025	0.022
Euro-D	0.078*	0.046	0.116**	0.052	0.141**	0.051
ADL or IADL	0.132**	0.042	0.063	0.048	0.037	0.047
Smoke & drink	0.105**	0.047	-0.031	0.054	0.041	0.053
Germany	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sweden	0.911***	0.090	0.216**	0.107	-0.450***	0.104
Netherlands	0.582***	0.092	0.203*	0.112	-0.378***	0.107
Spain	0.793***	0.093	-0.220*	0.117	0.184*	0.101
Italy	0.820***	0.079	1.122***	0.089	0.580***	0.086
France	0.349***	0.095	0.032	0.116	0.629***	0.103
Denmark	0.441***	0.070	0.135	0.086	-0.151*	0.080
Belgium	0.049	0.079	-0.302**	0.103	0.154*	0.087
Czechia	0.239**	0.070	1.030***	0.081	0.331***	0.077
Poland	1.214***	0.085	1.307***	0.092	0.667***	0.091
teta	0.506**	0.191	1.099***	0.220	0.686**	0.211
Threshold 1						
Woman	-0.015	0.051	-0.105*	0.058	0.023	0.050
Age	0.003	0.003	0.003	0.003	0.008**	0.003
Education	0.053**	0.025	0.024	0.029	0.039	0.025
Euro-D	0.014	0.059	-0.017	0.067	0.000	0.058
ADL or IADL	0.008	0.053	0.041	0.060	-0.011	0.053

Continued on next page...

... Table 2 continued

	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Smoke & drink	0.058	0.061	-0.116*	0.070	0.011	0.060
Germany	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sweden	0.581***	0.112	0.323**	0.125	-0.064	0.117
Netherlands	0.071	0.143	-0.209	0.152	-0.619***	0.145
Spain	0.368**	0.128	-0.411**	0.148	0.278**	0.117
Italy	0.424***	0.107	0.371**	0.128	0.343**	0.102
France	0.541***	0.114	-0.146	0.146	0.673***	0.111
Denmark	0.199**	0.097	0.139	0.105	0.077	0.091
Belgium	0.497***	0.095	-0.177	0.121	0.295**	0.099
Czechia	0.323***	0.091	0.823***	0.097	0.494***	0.086
Poland	0.387**	0.121	0.653***	0.127	0.311**	0.111
Constant	-2.102***	0.237	-2.069***	0.215	-2.293***	0.208
Threshold 2						
Woman	-0.014	0.043	0.016	0.028	-0.039	0.026
Age	0.003	0.002	-0.003*	0.002	0.000	0.001
Education	-0.012	0.022	0.000	0.014	-0.032**	0.014
Euro-D	-0.035	0.051	-0.030	0.033	-0.020	0.032
ADL or IADL	0.019	0.045	-0.007	0.028	0.012	0.028
Smoke & drink	-0.052	0.054	0.036	0.033	-0.002	0.032
Germany	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sweden	-0.278**	0.103	-0.472***	0.067	-0.358***	0.065
Netherlands	0.120	0.104	0.123**	0.058	0.149**	0.061
Spain	-0.379**	0.127	-0.323***	0.071	-0.337***	0.067
Italy	-0.193**	0.091	-0.355***	0.071	-0.260***	0.056
France	-0.459***	0.109	-0.154**	0.065	-0.266***	0.062
Denmark	-0.353***	0.091	-0.202***	0.046	-0.161***	0.046
Belgium	-0.155**	0.073	-0.197***	0.052	-0.242***	0.052
Czechia	-0.132*	0.072	-0.379***	0.048	-0.317***	0.047
Poland	-0.035	0.099	-0.285***	0.067	-0.066	0.057
Constant	0.033	0.206	0.928***	0.130	0.745***	0.127
Threshold 3						
Woman	-0.003	0.023	0.022	0.027	0.044	0.028
Age	0.001	0.001	0.000	0.001	-0.001	0.002
Education	0.041**	0.012	0.016	0.014	0.018	0.014
Euro-D	0.001	0.027	0.042	0.030	-0.015	0.031
ADL or IADL	-0.005	0.025	-0.067**	0.028	-0.037	0.029
Smoke & drink	0.018	0.028	-0.024	0.032	-0.022	0.033
Germany	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sweden	-0.049	0.052	-0.049	0.065	0.018	0.065
Netherlands	0.203***	0.047	0.037	0.065	0.139**	0.063
Spain	0.081	0.053	0.040	0.069	0.122**	0.062
Italy	-0.059	0.046	0.144**	0.050	0.084	0.052
France	0.089*	0.052	0.060	0.067	0.041	0.065
Denmark	-0.194***	0.044	-0.057	0.052	-0.073	0.052
Belgium	-0.022	0.045	-0.026	0.062	-0.041	0.056
Czechia	-0.002	0.039	0.029	0.047	0.103**	0.048
Poland	-0.352***	0.058	-0.032	0.053	-0.080	0.058
Constant	0.084	0.113	0.178	0.130	0.071	0.133
Threshold 4						
Woman	0.053*	0.032	0.009	0.038	0.007	0.044
Age	-0.001	0.002	0.003	0.002	0.002	0.003
Education	-0.002	0.017	0.032*	0.019	-0.020	0.023
Euro-D	-0.038	0.036	-0.039	0.043	-0.146**	0.050
ADL or IADL	-0.090**	0.033	0.039	0.039	-0.002	0.046
Smoke & drink	-0.025	0.036	0.014	0.045	-0.049	0.052

Continued on next page...

... Table 2 continued

	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Germany	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Sweden	0.000	0.074	0.197**	0.083	0.069	0.090
Netherlands	-0.152	0.093	0.014	0.111	0.065	0.105
Spain	0.059	0.072	0.154*	0.081	-0.007	0.104
Italy	0.178**	0.063	-0.158*	0.081	-0.072	0.088
France	-0.052	0.086	0.030	0.092	-0.115	0.120
Denmark	0.062	0.055	0.069	0.074	0.008	0.079
Belgium	-0.043	0.080	0.131*	0.076	0.035	0.088
Czechia	0.039	0.063	-0.058	0.081	0.010	0.085
Poland	0.288***	0.060	-0.129	0.085	0.152	0.095
Constant	0.191	0.155	-0.079	0.184	0.214	0.220
log sigma	-0.057	0.153	-0.031	-0.019	0.084	-0.016
N	12086		12026		12046	
Log likelihood	-14955.1		-13416.6		-14781.9	

Significance levels: * p<0.10 ** p<0.05 *** p<0.01

Table 3: Two-sample test of proportion

Country	with DIF (Counterfact. 1)	Without DIF (Counterfact. 2)	Diff (mean)	Std. Err.	Z	P-value
Time to wait for treatment						
Satisfied	0,209	0,170	-0,039***	0,003	-11,280	0,000
Unsatisfied	0,355	0,395	0,040***	0,004	9,429	0,000
Conditions of the health facilities						
Satisfied	0,602	0,683	0,081***	0,004	19,118	0,000
Unsatisfied	0,081	0,068	-0,013***	0,002	-5,462	0,000
Communication with doctors						
Satisfied	0,549	0,585	0,037***	0,004	8,413	0,000
Unsatisfied	0,111	0,110	-0,001	0,003	-0,351	0,363

Significance levels: * p<0.10 ** p<0.05 *** p<0.01

Figure 1: Satisfaction with time to wait for treatment

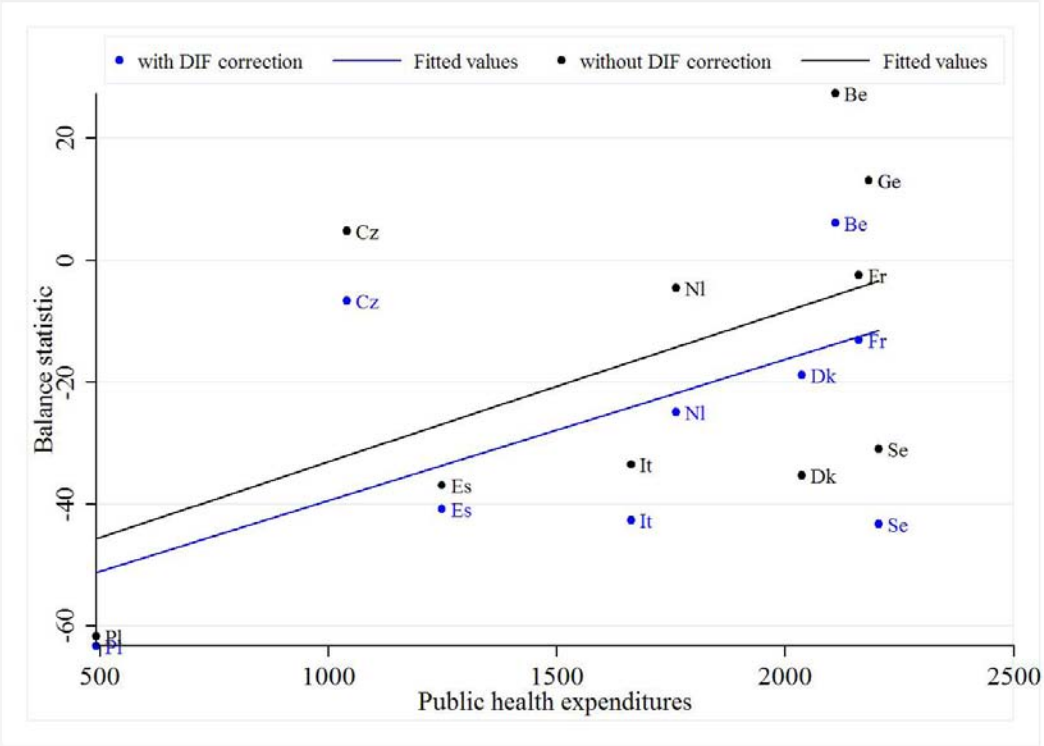


Figure 2: Satisfaction with the conditions of the health facilities

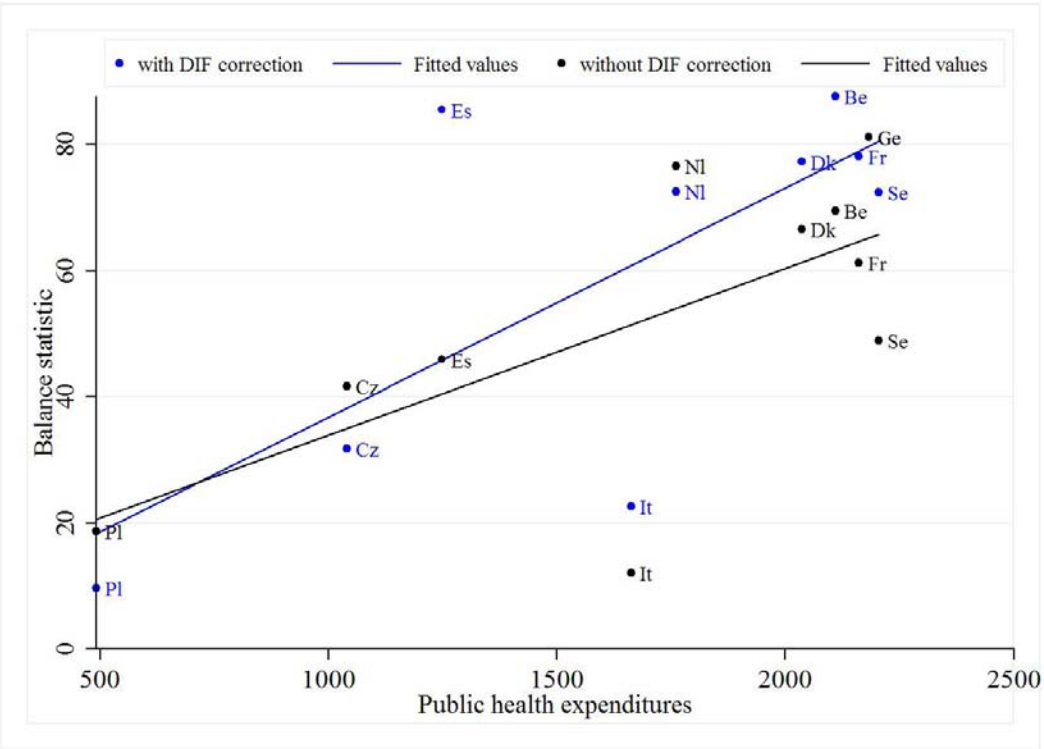
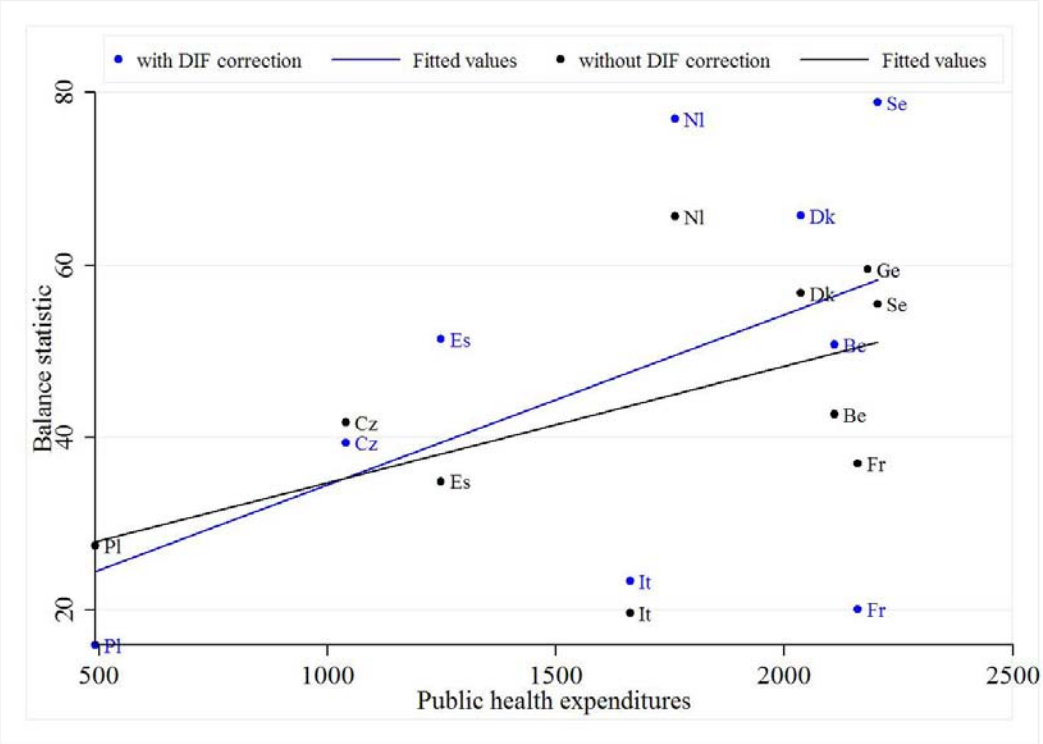


Figure 3: Satisfaction with communication with doctors



Comparability of Health Care Responsiveness in Europe Using Anchoring Vignettes from SHARE

Nicolas Sirven (Irdes), Brigitte Santos-Eggimann (Iumsp), Jacques Spagnoli (Iumsp)

The aim of this paper is to measure and to correct for the potential incomparability of responses to the SHARE survey on health care responsiveness.
A parametric approach based on the use of anchoring vignettes is applied to cross-sectional data (2006-07) in ten European countries. More than 6,000 respondents aged 50 years old and over were asked to assess the quality of health care responsiveness in three domains: waiting time for medical treatment, quality of the conditions in visited health facilities, and communication and involvement in decisions about the treatment. Chopit models estimates suggest that reporting heterogenity is in uenced by both individual (socio-economic, health) and national characteristics. Although correction for dierential item functioning does not considerably modify countries ranking after controlling for the usual covariates, about two thirds of the respondents' self-assessments have been re-scaled in each domain. Our results suggest that reporting heterogenity tends to overestimate health care responsiveness for `time to wait for treatment', whereas it seems to underestimate people's self-assessment in the two other domains

Rendre comparable la satisfaction des soins en Europe à partir des vignettes-étalons de SHARE

Nicolas Sirven (Irdes), Brigitte Santos-Eggimann (Iumsp), Jacques Spagnoli (Iumsp)

L'objectif de cet article est d'évaluer, puis de corriger un biais potentiel de déclaration concernant des questions sur la satisfaction des soins dans l'enquête SHARE. Une approche paramétrique utilisant des vignettes-étalons est appliquée à des données en coupe (2006-07) dans 10 pays Européens. Plus de 6.000 répondants âgés de 50 ans ou plus, ont été interrogés sur leur perception de la qualité des soins dans trois domaines : le temps d'attente avant la prise en charge médicale, la qualité des installations et la communication avec le personnel médical en ce qui concerne le traitement. Les estimations à partir de modèles CHOPIT suggèrent que les différences dans les façons de répondre sont influencées par des facteurs individuels (socio-économiques, santé) et des caractéristiques nationales. Bien que la correction du biais de déclaration ne modifie pas fondamentalement le classement des pays, près de deux tiers des réponses des individus ont ré-échelonnées (pour chaque domaine envisagé). Nos résultats suggèrent que la satisfaction des soins est surestimée à cause du biais de déclaration dans le cas du « temps d'attente », alors qu'il surestime les déclarations des individus dans les deux autres domaines.

IRDES
www.irdes.fr

ISSN : 2102-6386