

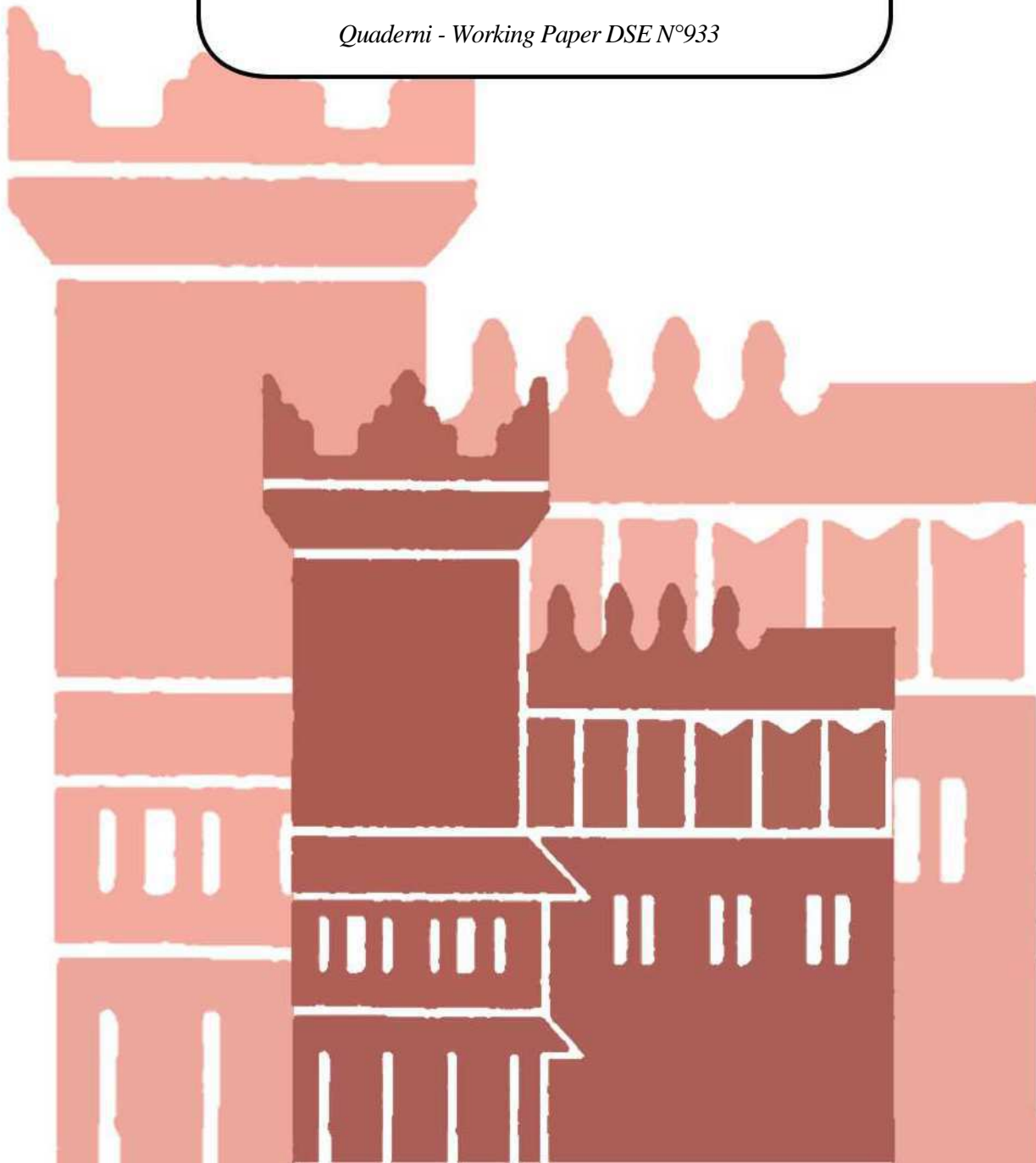
ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna  
DEPARTMENT OF ECONOMICS

**Are bad health and pain making us  
grumpy? An empirical evaluation of  
reporting heterogeneity in rating  
health system responsiveness**

Gianluca Fiorentini  
Giovanni Ragazzi  
Silvana Robone

*Quaderni - Working Paper DSE N°933*



# Are bad health and pain making us grumpy? An empirical evaluation of reporting heterogeneity in rating health system responsiveness

*Gianluca Fiorentini<sup>1</sup>, Giovanni Ragazzi<sup>2</sup> and Silvana Robone<sup>1</sup>*

<sup>1</sup> Department of Economics, University of Bologna, Italy

<sup>2</sup> Agency for Health Care and Social Services of Emilia-Romagna, Italy

## Abstract

This paper considers the influence of patients' characteristics on their evaluation of a health system's responsiveness, that is, a system's ability to respond to the legitimate expectations of potential users regarding non-health enhancing aspects of care (Valentine et al. 2003a). Since responsiveness is evaluated by patients on a categorical scale, their self-evaluation can be affected by the phenomenon of reporting heterogeneity (Rice et al. 2012). A few studies have investigated how standard socio-demographic characteristics influence the reporting style of health care users with regard to the question of the health system's responsiveness (Sirven et al. 2012, Rice et al. 2012). However, we are not aware of any studies that focus explicitly on the influence that both the patients' state of health and their experiencing of pain have on the way in which they report on system responsiveness. This paper tries to bridge this gap by using data regarding a sample of patients hospitalized in four Local Health Authorities (LHA) in Italy's Emilia-Romagna region between 2010 and 2012. These patients have evaluated 27 different aspects of the quality of care, concerning five domains of responsiveness (communication, social support, privacy, dignity and quality of facilities). Data have been stratified into five sub-samples, according to these domains. We estimate a generalized ordered probit model (Terza, 1985), an extension of the standard ordered probit model which permits the reporting behaviour of respondents to be modelled as a function of certain respondents' characteristics, which in our analysis are represented by the variables "state of health" and "pain". Our results suggest that unhealthier patients are more likely to report a lower level of responsiveness, all other things being equal, while patients experiencing pain are more likely to make use of the extreme categories of responsiveness, that is, to choose the category "completely dissatisfied" or the category "completely satisfied". These results hold across all five domains of responsiveness.

**Keywords:** Health care responsiveness, Health system performance, Reporting heterogeneity, Generalised ordered probit, Self-reported health, Pain.

**Contact Author:** Dr. Silvana Robone, Department of Economics, University of Bologna, P.zza Scaravilli 2, 40126 Bologna, Italy. Email: silvana.robone@unibo.it

**Acknowledgements:** This research was funded by the Agency for Health Care and Social Services of Emilia-Romagna. We would like to thank Matteo Lippi Bruni, Cristina Ugolini, Irene Mammi, Teresa Bago d'Uva and the participants at the Health, Econometrics and Data Group Seminar Series (University of York), for their useful comments.

**JEL Classification:** I1, I110, I190, C50, C25

## 1. Introduction

In recent years, the concept of responsiveness has been put forward as one desirable measure of the performance of health systems. Responsiveness concerns a system's ability to respond to patients' legitimate expectations regarding the non-health enhancing and non-financial aspects of health care. "Responsiveness is defined as the way in which individuals are treated and the environment in which they are treated, encompassing the notion of an individual's experience of contact with the health system" (Valentine et al. 2003a). The concept covers eight dimensions of quality of care, perceived in terms of respect for human dignity and of the interpersonal side of healthcare (Valentine et al. 2009). Human rights make reference to concepts such as respecting patient autonomy and dignity, while the interpersonal nature of care (or "client orientation") focuses on patient accommodation and the quality of basic amenities (Rice et al 2012). The eight domains chosen to represent responsiveness are as follows: *autonomy, choice, clarity of communication, confidentiality of personal information, dignity, prompt attention, quality of basic amenities* and *access to family and community support*. Table 1 provides definitions of these domains.

The evaluation of health systems' responsiveness has become an important, evidence-based means of identifying the strengths and weaknesses of health systems, of appraising their evolution over time, and of informing policy reform (Jones et al. 2011). The importance of this instrument has been confirmed at the international level by the European Ministerial Conference on Health Systems, culminating in the Tallin Charter (2008) which points to the importance that policy makers should place on the evaluation of health systems' performance (WHO 2008). At the national level, the same has been recently confirmed by recent guidelines published by the National Institute for Health and Care Excellence (NICE), a non-departmental public body within the UK Department of Health, designed to guide British policy makers in several areas of healthcare. These NICE guidelines specifically indicate the

users' perspective as an instrument with which to evaluate the UK health system (NICE 2012).

Health system responsiveness has been investigated both by adopting an international comparison perspective (see, for example, Valentine et al. 2008, Blendon et al. 2003, Robone et al 2011, Rice et al. 2012), and by more fully evaluating this performance indicator at a national level (Puentes Rosas et al. 2006, Pelzer 2009, Njeru et al. 2009, Kowal et al. 2011, Rashidian et al. 2011, Adesanya 2012). Our paper falls within the latter category, since it considers the influence of patients' characteristics on the evaluation of health system responsiveness using Italian data only.

Health system responsiveness is usually measured through the self-evaluations of respondents, which rate their experiences of health systems according to a categorical scale (usually a 5-point scale ranging from "very good" to "very bad"). One common problem is that when individuals are faced with an instrument comprising ordinal response categories, their interpretation of the response categories may systematically differ across populations or populations sub-groups, also depending on their preferences and norms (Rice et al 2010). In such a case, a given level of performance is unlikely to be rated equally by all respondents. This phenomenon has been termed "reporting heterogeneity".

A few studies have investigated how standard socio-demographic characteristics (such as gender or education) may influence the heterogeneity in the reporting of health care users about responsiveness (Puentes Rosas et al. 2006, Sirven 2012, Rice et al. 2012). The findings of such studies show that reporting heterogeneity is an issue in the case of self-reporting on the question of responsiveness. However, we are not aware of any study that specifically focus on the influence that patients' state of health and experiencing pain have on their reporting behaviour with regard to the matter of responsiveness. Valentine et al. (2003b) represents the only paper we are aware of that considers the influence of patients' self-

reported health on their reporting of responsiveness. However, they only use this relationship as a control in their regression model, and do not specifically investigate the way in which self-reported health affects the reporting behaviour of patients. Sirven et al. (2012) investigate the influence on responsiveness of much more specific health measures than self-reported health, by using a dummy based on the Euro-d scale, which is considered as a standard measure of depression (Dewey and Prince 2005), and a dummy indicating whether the respondent has difficulties with basic activities of daily living (ADL) or instrumental activities of daily living (IADL). Moreover, Sirven et al. (2012) only investigate a few of the responsiveness domains we consider in our analysis (for example, they do not consider *dignity, confidentiality* or *social support*). There is evidence in the literature regarding the fact that the experience of pain has a negative influence on patients' satisfaction with clinical outcomes (Baker et al. 2007), but not on non-clinical outcomes such as responsiveness. Our paper helps to bridge these gaps in the literature by exploring a relationship which no other study has explicitly considered before.

Our study uses a representative sample of patients (about 2,500 individuals) hospitalized in the Italian Emilia-Romagna Region. The data were collected by the Agency for Health Care and Social Services of Emilia-Romagna (ASSR) between January 2010 and December 2012. Respondents were asked to rate 5 domains of health system responsiveness (*communication, social support, privacy, dignity* and *quality of facilities*). We estimate a generalized ordered probit model (Terza, 1985), an extension of the standard ordered probit model which permits the reporting behaviour of respondents to be modelled as a function of certain characteristics of such respondent, which in our analysis are represented by the variables "state of health" and "pain". We also control for the standard socio-demographic characteristics of patients, for hospital dummy variables and for certain characteristics of treatment (medicine vs. surgery).

Our results suggest that patients in worse health, all other things being equal, are more likely to report a lower level of responsiveness than are those patients in better health, while patients experiencing pain are more likely to use extreme categories when evaluating responsiveness, that is, the categories “completely dissatisfied” or “completely satisfied”.

## **2. Institutional background**

The Italian National Health Service (SSN) is based on the principle of the universal coverage and comprehensive insurance of most health risks. It is mainly financed through general taxation, and it provides standard levels of care for the entire population. Central government funds the different Regional Health Services by means of a formula based substantially on a per capita rule, albeit adjusted to take account of certain epidemiological factors. The Regional Health Services allocate funds to Local Health Authorities (LHAs) again on a per capita basis, adjusted once again for the aforesaid epidemiological variables. The LHAs use these resources to fund all health care provided to the population under their responsibility, both through providers under their direct control, and through independent public and private healthcare service providers.

In the Emilia-Romagna Region, most providers of hospital care are publicly owned, and less than 10% of all cases are treated by private hospitals. Approximately two-thirds of Italy’s public hospitals are directly managed by the LHAs, while the remaining third are run by independent public bodies (IPH). Patients, often advised by their GPs and/or by specialists, can choose the hospital they wish to be treated at, and public hospital treatment is completely free of charge.

In Emilia-Romagna, since 1995 LHAs have funded IPHs within the Region under a contractual mechanism based on a pay-per-case rule, using the Diagnosis Related Groups (DRGs) weighting system. The effects of the pay-per-case rule can be significantly modified

due to the implementation of pre-determined constraints, both on the overall financial transfers and on the volume of activities, with reductions in the event of inappropriate hospitalisation. Those hospitals that are directly managed by the LHAs are given targets in terms of the volume of activity, and are mainly financed on a cost-reimbursement basis. Accordingly, both the IPHs and the hospitals run by the LHAs face relatively weak incentives to attract patients when they get near to the upper limits foreseen in their contracts with the Region's LHAs, while they are more strongly motivated to take care of patients from other regions or other countries, for whom no such upper financial limits are foreseen.

**Table 1.** Domains of responsiveness

<i>Autonomy</i> : respect for patients' views of what is appropriate, and allowing patients to make informed choices.
<i>Choice</i> : an individual's right or opportunity to choose a healthcare institution and health provider, and to request a second opinion and access specialist services when required.
<i>Clarity of communication</i> : the offering of a clear explanation to patients and family regarding the nature of the illness, together with details of treatment and of any available options.
<i>Confidentiality of personal information</i> : privacy in the environment in which consultations are conducted, and the concept of the privileged communication and confidentiality of medical records.
<i>Dignity</i> : the opportunity for patients to receive care in a respectful, caring, non-discriminatory setting.
<i>Prompt attention</i> : the opportunity to receive care rapidly in emergencies, or readily with short waiting times in the case of non-emergencies.
<i>Quality of basic amenities</i> : the physical environment and services often referred to as 'hotel facilities', including clean surroundings, regular maintenance, adequate furniture, sufficient ventilation and adequate space in waiting rooms.
<i>Access to family and community support</i> : the extent to which patients have access to their family and friends when receiving care, and the maintenance of regular activities (e.g. the opportunity to carry out religious and cultural practices).

**Note:** Source: Rice et al. (2012). The eight domains of responsiveness are defined by the World Health Organization (see Valentine *et al.* (2003a) for a full exposition of these domains). The response categories available to respondents are: 'very good', 'good', 'moderate', 'bad' and 'very bad'.

### 3. Data

In order to investigate our research hypothesis, we use a dataset collected by the Agency for Health Care and Social Services of Emilia-Romagna (ASSR) regarding patient satisfaction with the hospital services offered by the Italian National Health Service. The data pertain to a sample of about 2500 individuals who were admitted to nine hospitals located within the Parma and Modena LHAs, the Parma's IPHs, and to the Orthopaedic Hospital

*Istituto Ortopedico Rizzoli* (Bologna) in 2009 and 2010. The data were collected using a probabilistic approach, so as to guarantee that they could be regarded as a statistically representative sample of the proportion of inpatients treated in those wards where data were collected, and of the population in general in terms of gender. Individuals were interviewed about their satisfaction with several aspects of their interaction with the health system (29 items), which can be considered part of the “traditional” domains of responsiveness. Table 2 shows the correspondence between the “traditional” domains of responsiveness, as defined by the World Health Organization (WHO) and described in Valentine et al. (2003a), and the items comprised in the questionnaire administered by the ASSR. The domains of responsiveness included in the analysis are *communication, social support, privacy, dignity, waiting times and the quality of facilities*. The response categories were: “completely unsatisfied”, “very unsatisfied”, “unsatisfied”, “satisfied”, “very satisfied” and “completely satisfied”. Since the percentage of respondents claiming they were “completely unsatisfied” or “very unsatisfied” is extremely low (only about 2% of all respondents) we aggregate these two categories.

In order to evaluate the appropriateness of the correspondence between the domains of responsiveness as defined by the WHO, and the items of patient satisfaction expressed in the ASSR’s questionnaire, we investigate the psychometric property of “validity” in the case of the latter questionnaire. We focus on “validity” since this property allows to explore the level of homogeneity among the items which form a responsiveness domain, and the unidimensionality of the concept itself as represented by the responsiveness domain (Cortina 1993). Table 2 shows that in the ASSR questionnaire, multiple items are included for each responsiveness domain, with the exception of the domain “social support”, for which only one item is available. Cronbach alphas are the standard tools for measuring validity. A minimum value within the 0.6 to 0.7 range has been suggested for the Cronbach alpha in order to support the validity of the instrument (e.g. Labarere 2001; Steine et al. 2001). Table 3 shows Cronbach alphas computed for the responsiveness domains present in Table 2 (with the exception of *waiting times*). The values of the Cronbach alphas suggest that all of the responsiveness domains (for which is possible to apply this technique) are greater than 0.85 and lie within the desired range. The “*waiting times*” domain consists of two items only. Therefore, a Pearson’s correlation appears better suited for the purpose of evaluating validity for this domain. The correlation between the two items appears to be positive and significant, but not very strong (about 40%), suggesting that the internal validity of this domain requires



improving (Taylor 1990). Only one item corresponds to the domain of “*social support*”, therefore there is no need to check the internal validity of this domain.

The dataset also contains information on patients’ socio-economic characteristics. We control for gender, by including the dummy variable *woman* in the regression model (*man* being the reference category), and for marital status by including the dummy variable *single* (*being married* or *living with a partner* being the reference categories). We also control for education, by including the dummy variable *high education* (leaving school with a qualification lower than that of a high-school diploma, or with no qualifications at all, being the reference category), and for occupational status, by including the dummy variable *work* (not being employed being the reference category). Unfortunately, for reasons of privacy the dataset does not contain information about income or age. However, these variables are likely to be highly correlated to the occupational status and educational qualifications of respondents, and therefore their exclusion from our regression analysis should not bias our results. Women constitute about 48% of the sample. Most of the respondents are married or are living with a partner (about 55%), are retired (about 55%), and possess an educational qualification lower than that of a high-school diploma (about 61%). Full descriptive statistics on the socio-demographic characteristics of the respondents are available on request.

We also control for patients’ place of residence, including in the regression model the dummy variable *outside the province* (which is equal to 1 if patients are treated in hospitals outside the Province where they live, 0 otherwise), and the dummy variable *outside the region or country* (if patients are treated in hospitals outside their region or country of residence). Being treated in the Province of residence is the reference category. The majority of respondents reside in the Province where they are treated (about 77%), but about 9% live in Provinces other than the one where they are being treated, and about 14% live in Regions or countries other than the one where they are being treated.

The dataset contains information about the hospital where the patient is treated. Therefore, we include dummy variables at hospital level which should account for the variability in the rating of responsiveness due to hospital characteristics (the base category is the IPH of Parma). Some information is also available about the kind of treatment received by patients. Thus the regression model includes a dummy variable *surgery* (which takes a value of 1 if treatment is of a surgical nature, or of 0 otherwise).

Moreover, our regression model may include a few dummies representing the variables whose influence on responsiveness, as far as we know, has not been specifically examined to date. Thus we include the dummy variable *emergency* (which takes a value of 1 if patients

have been hospitalized through Accident & Emergency, or 0 if their hospitalization was planned), the dummy *previous hospitalization* (which takes a value of 1 if the patient has already been admitted to the same ward in the past, or 0 if the patient has never been admitted to that ward before), the dummy *poor/moderate health* (being in good or excellent health being the reference category), and the dummy *pain* (not being in a state of pain being the reference category).

**Table 2: Correspondence between the “traditional” domains of responsiveness, as defined by Valentine et al. (2003a), and the items present in the questionnaire administered by the Agency for Health Care and Social Services of Emilia-Romagna**

<i>Responsiveness domains</i>	<i>items in the questionnaire of the ASSR</i>
<b>Communication</b>	<p>how would you evaluate the information received before being hospitalized?</p> <p>how would you evaluate the information received when hospitalized?</p> <p>how would you evaluate the information received from your doctor about your state of health?</p> <p>how would you evaluate the willingness of nurses to clarify things for you?</p> <p>how would you evaluate the information received from your doctor about possible risks of your treatment?</p> <p>how would you evaluate the information received about the treatment required?</p> <p>how would you evaluate the information received about the checks ups you require following your discharge from hospital?</p> <p>how would you evaluate the information received about the symptoms to be monitored?</p> <p>how would you evaluate the information received about the staff to contact in case of need?</p> <p>how would you evaluate the information received about how to cope with physical pain?</p>
<b>Confidentiality</b>	<p>how would you evaluate the respect of your privacy when dealing with sensitive information?</p> <p>how would you evaluate the respect of your privacy when receiving confidential treatment?</p>
<b>Dignity</b>	<p>How would you evaluate the kindness and respectfulness of nurses?</p> <p>How would you evaluate the kindness and respectfulness of doctors?</p> <p>How would you evaluate the ability of nurses to make you feel comfortable?</p>
<b>Prompt attention</b>	<p>How would you evaluate the waiting time between the booking of a hospital admission and the admission itself?</p> <p>How would you evaluate the waiting time between your arrival at the ward and the admission itself?</p>
<b>Quality of facilities</b>	<p>How would you evaluate the maintenance of the facilities and rooms in the ward?</p> <p>How would you evaluate the cleanliness of the rooms and corridors?</p> <p>How would you evaluate the cleanliness of bathrooms ?</p> <p>How would you evaluate the warmness of the rooms ?</p> <p>How would you evaluate the level of quietness in the ward ?</p> <p>How would you evaluate the availability of space in the rooms?</p> <p>How would you evaluate the quality of the food ?</p> <p>How would you evaluate the choice of food on the menu?</p> <p>How would you evaluate the time of meals ?</p> <p>How would you evaluate the cleanliness of bedding?</p> <p>How would you evaluate the comfort of your bed ?</p>
<b>Social Support</b>	<p>How would you evaluate the visiting hours for relatives and friends?</p>

**Table 3: Analysis of the psychometric property of validity for the questionnaire of ASSR**

responsiveness domains	no. of items in the ASSR questionnaire	Cronbach Alpha/
		Pearson's Correlation
Communication	10	0.95
Confidentiality	3	0.85
Dignity	3	0.88
Prompt attention	2	$r(1826) = .39, p < .001$
Quality of facilities	11	0.93
Social support	1	-

#### 4. Econometric Models and Empirical Strategy

Self-reported responsiveness (SRR) is reported via an ordered categorical variable, which is assumed to represent some underlying latent scale. When individuals map the latent scale to the response categories in a consistent way, irrespective of their characteristics or circumstances, they are said to adopt homogeneous reporting behaviour. Under these circumstances, the standard ordered probit estimator, which assumes that the mapping of the latent scale to the response categories is made through a set of constant cut-points, would constitute an appropriate method of modelling the data (Rice et al 2012).

The ordered probit model (OPROBIT) can be used to model a categorical dependent variable which shows ordered multinomial outcomes for each respondent  $i$ , for example  $y_i = 1, 2, \dots, m$ . This model can be applied to SRR, which takes as its categorical outcomes ‘*completely or very unsatisfied*’, ‘*unsatisfied*’, ‘*satisfied*’, ‘*very satisfied*’ and ‘*completely satisfied*’. The model can be expressed as:

$$y_i = j \quad \text{if} \quad \mu_{j-1} < y^*_i < \mu_j, \quad j = 1, \dots, m \quad (1)$$

where the latent variable  $y^*$  is assumed to be a linear function of a vector of covariates  $x$ , plus a random error term  $\varepsilon$ :

$$y^*_i = x_i\beta + \varepsilon_i \quad \varepsilon_i \sim N(0,1) \quad (2)$$

and  $\mu_0 = -\infty, \mu_j \leq \mu_{j+1}, \mu_m = \infty$  are the cut-points which separate the categorical outcomes

If we assume that the error term is normally distributed, the probability of observing a particular value of  $y$  is:

$$P_{ij} = P(y_i = j) = \Phi(\mu_j - x_i\beta) - \Phi(\mu_{j-1} - x_i\beta), \quad j = 1, \dots, 5 \quad (3)$$

where  $\Phi(\cdot)$  is the standard normal distribution function.

The assumption of homogeneous reporting behaviour, which is made when using an ordered probit model, arises from the assumption that the cut-points  $\mu$  present in the model are constant. This assumption does not hold when respondents differ in the positioning of cut-points, when mapping the latent variable  $y^*$  to the available response categories  $y_i$ . We term this phenomenon “reporting heterogeneity”. In this case, forcing the cut-points to be constant will lead to biased estimates of the coefficients  $\beta$  in the main responsiveness equation, since these  $\beta$  will reflect both the “true” responsiveness effects and the reporting heterogeneity effects.

To acknowledge the presence of reporting heterogeneity, we estimate a generalized ordered probit model (Terza, 1985). The generalized ordered probit (GOP) model can be considered an extension of the standard ordered probit model, where the cut-points of equation (1) are modelled as functions of covariates  $z$ :

$$\mu_i^j = z_i\gamma^j \quad (4)$$

If a variable influences both the cut-point equation (4) and the main responsiveness equation (2), then these two influences cannot, in general, be separated. However, identification can be achieved by assuming that each covariate may be excluded from either the cut-point equations or the main responsiveness equation. Therefore,  $x_i$  and  $z_i$  have to be distinct vectors (Pudney and Shields 2000). The standard OPROBIT model can be considered a special case of the GOP model, with the cut-points specified as constant parameters (Pudney and Shields 2000). For more details of the GOP, see Jones et al. (2007).

We observe a so-called “parallel cut-points shift” when the covariates affect all cut-points to the same extent. However, reporting heterogeneity can be more relevant at some levels of responsiveness than at others. In these cases, covariates have a differential impact across cut-points. This is referred to as a “non-parallel shift” (Jones et al. 2007).

In the original dataset provided by the ASSR, each patient rates his/her satisfaction with 29 different aspects of responsiveness. That is, there are 29 items of responsiveness which can be stratified into 6 responsiveness domains. We could have estimated 29 different regression models, one for each item of responsiveness included in the dataset, in order to evaluate the impact of the variables in question on patients' satisfaction with responsiveness. However, the interpretation of these results would have been too complex, and probably not as meaningful. Therefore, we choose to stratify our sample into six different groups, one for each domain of responsiveness reported in Table 2. Since, as Section 4 shows, the property of validity for the domain of “*waiting times*” appears weak, we choose not to consider this domain in our analysis. Within the remaining five groups, we reshape the dataset from a “wide form” into a “long form”, by taking advantage of the fact that the level of satisfaction is recorded in the same way for each responsiveness item (using the categorical outcomes ‘*completely or very unsatisfied*’, ‘*unsatisfied*’, ‘*satisfied*’, ‘*very satisfied*’ and ‘*completely satisfied*’). Within each of the 5 responsiveness domains, we establish a variable called “satisfaction with responsiveness”, which expresses a patient’s satisfaction with the domain of responsiveness we are dealing with, regardless of which specific item of responsiveness is being considered. The specific item of responsiveness the patient is evaluating is represented by a dummy variable. Table 4 shows the percentage of individuals choosing each response category, for each responsiveness domain.

**Table 4: Frequency and percentage of individuals choosing each response category, for each responsiveness domain**

	Communication		Social Support		Privacy		Dignity		Quality of Facilities	
	freq.	%	freq.	%	freq.	%	freq.	%	freq.	%
compl. or very unsatisfied	193	1.39	16	0.98	41	1.25	51	1.02	442	2.44
unsatisfied	476	3.42	52	3.17	114	3.48	112	2.23	1,170	6.45
satisfied	5,654	40.63	707	43.08	1,360	41.53	1,384	27.6	7,845	43.23
very satisfied	3,685	26.48	403	24.56	846	25.83	1,641	32.73	4,836	26.65
completely satisfied	3,907	28.08	463	28.21	914	27.91	1,826	36.42	3,856	21.25
<b>Total</b>	<b>13,915</b>	<b>100</b>	<b>1,641</b>	<b>100</b>	<b>3,275</b>	<b>100</b>	<b>5,014</b>	<b>100</b>	<b>18,149</b>	<b>100</b>

We first estimate the standard OPROBIT as a baseline model. This model is helpful in assessing the extent to which the assumption of reporting homogeneity may bias the estimated effects on responsiveness. We then estimate the GOP model. In theory we could

model the cut-point equations as linear functions of *all* explanatory variables. However, this is not advisable since it would lead to a very heavily parameterized model (Pudney and Shields 2000). The identification of the GOP model requires excluding the covariates in the cut-point equations from the main responsiveness equation. We choose to retain, in the responsiveness equation, 1) those variables commonly adopted in the literature as drivers of responsiveness (such as socio-economic variables) (Valentine 2003b, Sirven et al. 2012, Rice et al. 2012), 2) the covariates which can genuinely affect responsiveness (such as the hospital dummies, *outside the province* and *outside the region or country* and *surgery*) 3) the other covariates which do not appear to affect the reporting behaviour of individuals across the five domains of responsiveness. To assess which variables can be included in group 3), a formal Wald test is performed in order to identify the presence of reporting heterogeneity, by following a sequential procedure similar to the one proposed by Pudney and Shields (2000). For each responsiveness domain, we first ran a GOP regression model, including in the cut-point equations all independent variables not included in groups 1) and 2), that is, *poor/moderate health*, *pain*, *emergency* and *previous hospitalization*. Results for this specification are available on request. In the case of the last two variables, the Wald tests for homogeneity in reporting behaviour do not permit the rejection of the null hypothesis for all responsiveness domains (with the exception of *emergency* for *Communication* and *previous hospitalization* for *Quality of Facilities*), while for the first two variables the tests do reject the null hypothesis. Therefore, we have retained *emergency* and *previous hospitalization* in the main responsiveness equation, and included *poor/moderate health* and *pain* in the cut-point equations. We would point out that in both OPROBIT and GOP models, given that we are dealing with a nonlinear ordered categorical dependent variable, the estimated coefficients possess qualitative content only. To provide information about the magnitude of the effects, we present marginal effects (Wooldridge 2002). In particular, we report the change in the probability of being “completely satisfied” with responsiveness due to a discrete change for the dummy variables used in our models. Note that the direction of the effect of the covariates on the probabilities of reporting the extreme outcome “completely satisfied” is determined by the sign of the coefficients (Wooldridge 2002). Inferences regarding the significance of the estimated coefficients are made with reference to the Wald tests.

## 5. Results

Table 5 reports the estimated coefficients and related standard errors for standard OPROBIT models, where satisfaction with responsiveness is regressed on the full sets of dummy variables described in the Data Section. The reference individual is male, possesses qualifications lower than a secondary school diploma, or no academic qualifications at all, is married or living with a partner, is not registered for work, lives in the Province where the hospital is based, is treated in the IPH of Parma, and receives some medical treatment. He has never been previously admitted to the ward where he receives treatments, and his current admission was planned in advance. He declares he is in good or excellent health, and is not experiencing pain. The first column of the table presents results for the *Communication* domain, the second column for *Social Support*, the third for *Confidentiality*, the fourth for *Dignity*, and the fifth for *Quality of Facilities*. Each column also shows the number of observations used to estimate each regression model, and the values assumed by the Akaike and Bayesian information criteria (AIC and BIC), which are standard measures of fit used for model selection (Green 2008).

Table 6 shows the estimated coefficients and standard errors for the GOP model. Since we have established that the categorical variables for responsiveness may take on 5 possible values, then 4 cut-points “divide” the response categories. Therefore, there are 4 cut-point equations in the GOP model. A positive (negative) sign of a coefficient means that the cut-points tend to shift rightwards (leftwards), and therefore that people have higher (lower) expectations with regard to that level of responsiveness. We choose to model the cut-points as functions of the variable *poor/moderate health* and *pain*. This choice, together with the formal statistical test described in the previous Section, also depends on the observation that in Table 5 (OPROBIT model) these variables always display a negative coefficient, which is also statistically highly significant for most of the responsiveness domains. It is quite unlikely that the Emilia-Romagna Regional Health Service is systematically less responsive to patients who are in poor or moderate health, or in pain. Therefore, the negative influence of *poor/moderate health* and *pain* on the reported levels of responsiveness is more likely to be due to a bias in the reporting of those patients who find themselves in such states, rather than to a “true” negative responsiveness effect.

The first three columns in Table 7 present the results of tests for homogeneity in reporting behaviour for the five responsiveness domains considered. For both *poor/moderate health* and *pain*,  $\chi^2$  statistics and p-values from a Wald test of the joint significance of the

estimated coefficients across the four cut-points of each model are reported. Rejection of the null hypothesis indicates that cut-points are functions of the variables in question. In addition to separate tests for each variable, the first column also reports a joint test across the two variables in question. For all domains, the null hypothesis of homogenous reporting can be rejected. The heterogeneous reporting behaviour of individuals appear to depend on both *poor/moderate health* and *pain*, with the exception of the domain of *Social Support*, where *poor/moderate health* does not appear to have any influence on the reporting style of the respondents (the p-value of the Wald test is 0.07). Columns four and five of Table 7 present the results of tests for parallel cut-point shift for the responsiveness domains in question. We can reject the null hypothesis of a parallel cut-point shift for all combinations of domains and variables, except for the combination *poor/moderate health* and *Social Support* (which appears to be characterised by a homogeneous reporting style) and *Confidentiality*.

The Akaike and Bayesian information criteria (shown in Tables 5 and 6) also suggest that the GOP model performs better than the standard OPROBIT model, since their values are lower for each responsiveness domain in the former model. Therefore, we choose to focus on the results obtained using the GOP model. When we consider the main responsiveness equation, the effect on reported responsiveness of being a woman and of having a high level of education is always negative (with the exception of *woman* for *Communication*), but seldom statistically significant. Note that in Table 5, the variable *high education* has a positive effect on responsiveness in 4 out of 5 responsiveness domains, and this effect is almost always statistically significant. Therefore, relying on the standard OPROBIT model rather than the GOP model, could produce misleading results. With regard to the gender variable, in the literature on responsiveness the findings concerning the effects of being a woman on reporting behaviour are not univocal (Sirven et al. 2012). However, if we look at the broader literature examining the question of health status, our results for gender appear to be in line with those of Bago d’Uva et al. (2008). The latter study finds that women usually have higher expectations than men with regard to their health. The effects of occupational status are almost never significant, while those of being single are negative in 4 out of 5 domains and statistically significant in 3 of them. The variables *outside the province* or *outside the region or country* always have a positive influence on the reported levels of responsiveness, and for the latter variable this influence is almost always statistically significant. As we indicated in Section 2, this effect could be accounted for by the greater financial incentives for hospitals to treat patients from other regions or countries, as the corresponding revenues are not subject to the usual financial constraints imposed by block



contracts in the case of patients from the region itself. Accordingly, hospitals may instruct personnel to be more responsive towards such external patients, in order to attract further patients from outside the region in question by means of reputational mechanisms.

Note that the signs of several coefficients for the variables *single* and *outside the province* differ from Table 5 to Table 6. Therefore, the standard OPROBIT model could lead to biased results even in the case of these variables. We find no consistent pattern, across domains, in the sign of the coefficients for the hospital dummies, and many of such coefficients are statistically highly significant. Analogously, in the case of the variables *surgery* and *emergency*, no consistent pattern in the sign of the coefficients could be found, and in this case most of the coefficients are not statistically significant. The variable *previous hospitalization* appears to have a negative influence on responsiveness.

With regard to the cut-point equations, being in a state of health worse than good, tends to have a negative, statistically significant effect on the reporting of responsiveness compared to being in good or excellent health. This is suggested by the fact that most of the coefficients which are statistically significant possess a positive sign across all cut-points. This holds for all domains, with the exception of *Confidentiality*, which displays a negative and statistically significant coefficient for the second cut-point. However, as Table 8 shows, we cannot reject the hypothesis of parallel cut-point shift in the case of this domain. When we estimate the GOP model for *Confidentiality* by making this assumption (results are available on request), the coefficient for *poor and moderate health* in the cut-point equation is 0.108, and it is statistically significant at the 99% level. Therefore, also in this domain, patients in poorer health appear to have higher expectations with regard to responsiveness. This result is in line with the findings of Sirven et al. (2012), who highlight that the reporting style of respondents is in fact affected by their state of health, since in their study respondents with depressive symptoms systematically report worse health system responsiveness. Across all domains, the variable *pain* has a positive effect on the first cut-point (the one dividing “completely or very unsatisfied” from “unsatisfied ” with responsiveness), but a negative effect on the last cut-point (dividing “very satisfied ” from “completely satisfied” with responsiveness). This implies that patients in pain tend to opt more often for the extreme categories, that is, they rate responsiveness as either extremely bad or very good more often than do patients who are not in pain.

Table 8 reports the marginal effects concerning the GOP model (the increase in the probability of choosing the “completely satisfied” category) and the related standard deviations. The marginal effects have been computed assuming that all the respondents have

the reporting style of an individual characterized by average values of health and pain (that is, we use the average values of the cut-points). All the marginal effects are statistical significant, since they have a p-value of less than 0.001. The magnitude of the effects of socio-economic variables on the evaluation of responsiveness is not particularly high, since the marginal effects for these variables vary from about 0.5% to about 5%. Similar results hold for the variables *surgery*, *emergency* and *previous hospitalization*, since their marginal effects vary from about 1% to about 3%. Marginal effects appear to be larger for *outside the province* and *outside the region or country* (which vary from about 1% to about 13%, with several higher than 5%), and for the hospital dummies (which vary from about 1% to 31%, with several higher than 5-10%).

**Table 5: Ordered probit model for self-reported responsiveness**

satisfaction with responsiveness	Communication	Social Support	Confidentiality	Dignity	Quality of Facilities
	<i>b/se</i>	<i>b/se</i>	<i>b/se</i>	<i>b/se</i>	<i>b/se</i>
woman	-0.025 0.019	0.029 0.056	0.01 0.04	-0.013 0.032	-0.047** 0.016
high education	0.153*** 0.022	0.014 0.064	0.109* 0.045	0.124*** 0.037	-0.084*** 0.019
work	-0.046* 0.023	0.074 0.069	-0.023 0.049	0.062 0.04	0.014 0.02
single	-0.043* 0.02	0.028 0.057	-0.001 0.04	-0.062 0.033	-0.027 0.017
outside the province	-0.015 0.035	-0.006 0.100	0.054 0.072	0.007 0.059	-0.006 0.03
outside the region or country	0.361*** 0.033	0.487*** 0.099	0.391*** 0.068	0.393*** 0.057	0.252*** 0.028
hospital dummy 2	0.214*** 0.047	0.17 0.114	0.247** 0.081	0.403*** 0.068	0.181*** 0.034
hospital dummy 3	-0.220*** 0.043	0.188 0.128	-0.264** 0.090	-0.215** 0.072	-0.468*** 0.037
hospital dummy 4	0.172 0.103	0.007 0.283	0.262 0.201	0.501** 0.167	0.180* 0.086
hospital dummy 5	0.479*** 0.034	0.284** 0.098	0.386*** 0.069	0.334*** 0.057	0.174*** 0.029
hospital dummy 6	-0.563*** 0.041	-0.472*** 0.119	-0.569*** 0.084	-0.588*** 0.066	-0.542*** 0.035
hospital dummy 7	-0.179** 0.068	-0.398 0.207	-0.169 0.146	0.063 0.116	0.019 0.060
hospital dummy 8	0.530*** 0.054	0.857*** 0.160	0.700*** 0.114	0.358*** 0.092	0.503*** 0.047
hospital dummy 9	-0.106** 0.032	-0.142 0.095	-0.180** 0.067	-0.129* 0.054	0.056* 0.028
surgery	-0.061* 0.027	0.058 0.080	-0.112* 0.057	-0.096* 0.046	0.023 0.023
poor/moderate health	-0.467*** 0.020	-0.266*** 0.058	-0.421*** 0.041	-0.446*** 0.033	-0.300*** 0.017
emergency	-0.142*** 0.022	-0.107 0.065	-0.05 0.046	-0.100** 0.037	-0.021 0.019
pain	-0.113*** 0.02	-0.06 0.059	-0.037 0.041	-0.102** 0.034	-0.113*** 0.017
previous hospitalization	-0.084*** 0.021	-0.150* 0.062	-0.129** 0.044	-0.032 0.035	-0.155*** 0.018
cut-point 1 constant	-2.614*** 0.046	-2.504*** 0.139	-2.538*** 0.096	-2.679*** 0.081	-2.334*** 0.037
cut-point 2 constant	-2.072*** 0.041	-1.916*** 0.117	-1.976*** 0.083	-2.207*** 0.07	-1.707*** 0.033
cut-point 3 constant	-0.391*** 0.036	-0.147 0.106	-0.288*** 0.075	-0.768*** 0.062	-0.235*** 0.031
cut-point 4 constant	0.389*** 0.037	0.563*** 0.106	0.467*** 0.076	0.166** 0.061	0.562*** 0.031
aic	32507.499	3834.510	7709.923	11558.322	45942.591
bic	32680.936	3958.781	7850.087	11708.282	46122.137
N	13915	1641	3275	5014	18149

Note: \*\*\* p<0.001, \*\* p<0.01, \* p<0.05

**Table 6: Generalized ordered probit model for self-reported responsiveness**

satisfaction with responsiveness	Communication	Social Support	Confidentiality	Dignity	Quality of Facilities
	<i>b/se</i>	<i>b/se</i>	<i>b/se</i>	<i>b/se</i>	<i>b/se</i>
woman	-0.068** -0.024	-0.035 -0.072	-0.09 -0.05	-0.03 -0.043	-0.114*** -0.02
high education	0.031 -0.027	-0.111 -0.081	-0.009 -0.057	-0.044 -0.049	-0.148*** -0.022
work	-0.082** -0.03	0.039 -0.09	-0.023 -0.062	0.026 -0.056	0.011 -0.024
single	-0.065** -0.024	0.025 -0.072	-0.035 -0.051	-0.093* -0.044	-0.061** -0.02
outside the province	0.013 -0.044	0.15 -0.127	0.248** -0.092	0.018 -0.081	0.036 -0.036
outside the region or the country	0.408*** -0.047	0.122 -0.14	0.280** -0.095	0.276** -0.089	0.142*** -0.036
hospital dummy 2	0.079 -0.061	0.08 -0.147	0.268* -0.106	0.195 -0.102	0.149*** -0.041
hospital dummy 3	-0.277*** -0.052	0.258 -0.162	-0.196 -0.108	-0.193* -0.093	-0.462*** -0.042
hospital dummy 4	0.441*** -0.127	0.113 -0.341	0.187 -0.262	0.839** -0.273	0.192 -0.104
hospital dummy 5	-0.174*** -0.05	-0.02 -0.135	0.008 -0.096	-0.225* -0.089	-0.142*** -0.036
hospital dummy 6	-0.404*** -0.047	-0.395** -0.14	-0.378*** -0.097	-0.377*** -0.081	-0.410*** -0.039
hospital dummy 7	-0.141 -0.077	-0.235 -0.232	-0.104 -0.17	0.243 -0.151	0.290*** -0.068
hospital dummy 8	0.099 -0.074	0.734** -0.237	0.459** -0.165	0.043 -0.135	0.168** -0.061
hospital dummy 9	-0.073 -0.04	-0.2 -0.118	-0.091 -0.082	-0.168* -0.072	0.127*** -0.033
surgery	-0.028 -0.035	0.068 -0.103	-0.109 -0.074	-0.072 -0.065	0.048 -0.028
emergency	-0.106*** -0.028	-0.036 -0.081	0.05 -0.058	-0.007 -0.05	0.036 -0.023
previous hospitalization	-0.02 -0.026	-0.07 -0.077	-0.068 -0.054	0.099* -0.047	-0.101*** -0.021
cut-point 1 poor/moderate health	0.171* -0.083	0.443 -0.278	0.672** -0.242	0.287 -0.191	-0.083 -0.055
pain	0.359*** -0.092	0.404 -0.238	0.653** -0.201	0.406 -0.23	0.216*** -0.058
constant	-2.858*** -0.109	-3.106*** -0.311	-3.565*** -0.285	-3.047*** -0.283	-2.491*** -0.062
cut-point 2 poor/moderate health	0.048 -0.261	-1.444 -0.807	-0.812* -0.389	-0.063 -0.546	-0.176 -0.134
pain	0.166 -0.33	-0.081 -0.355	-0.109 -0.245	0.092 -0.708	0.429** -0.163
constant	-1.539*** -0.382	-0.672 -0.472	-0.253 -0.347	-1.198 -0.862	-1.498*** -0.157
cut-point 3 poor/moderate health	0.313** -0.096	-0.083 -0.273	-0.072 -0.184	0.343 -0.206	0.290*** -0.055
pain	-0.11 -0.096	0.038 -0.305	-0.418* -0.2	-0.019 -0.203	-0.134* -0.057
constant	-0.653*** -0.113	-0.438 -0.286	-0.149 -0.196	-0.843*** -0.249	-0.437*** -0.055
cut-point 4 poor/moderate health	0.008 -0.023	0.109 -0.065	0.106* -0.046	-0.008 -0.048	0.134*** -0.019
pain	-0.171*** -0.023	-0.227*** -0.065	-0.198*** -0.047	-0.235*** -0.049	-0.155*** -0.019
constant	0.749*** -0.021	0.793*** -0.057	0.723*** -0.042	0.658*** -0.047	0.552*** -0.017
aic	17856.523	2058.335	4202.625	5580.336	28608.692
bic	18075.204	2215.024	4379.353	5769.416	28835.077
N	13915	1641	3275	5014	18149

Note: \*\*\* p<0.001, \*\* p<0.01, \* p<0.05

**Table 7: Tests of homogenous reporting and parallel cut-point shift**

	test for homogeneous reporting						test for parallel cut point shift			
	All		poor/moderate health		pain		poor/moderate health		pain	
	chi2(8)	p	chi2(4)	p	chi2(4)	p	chi2(3)	p	chi2(3)	p
Communication	297.6	<b>0.00</b>	207.6	<b>0.00</b>	68.6	<b>0.00</b>	22.0	<b>0.00</b>	68.5	<b>0.00</b>
Social Support	20.8	<b>0.01</b>	8.7	0.07	12.4	<b>0.01</b>	4.2	0.25	12.0	<b>0.01</b>
Confidentiality	45.9	<b>0.00</b>	21.5	<b>0.00</b>	25.4	<b>0.00</b>	6.9	0.07	22.2	<b>0.00</b>
Dignity	113.8	<b>0.00</b>	81.4	<b>0.00</b>	26.1	<b>0.00</b>	8.7	<b>0.03</b>	26.0	<b>0.00</b>
Quality of Facilities	300.6	<b>0.00</b>	191.1	<b>0.00</b>	90.5	<b>0.00</b>	32.8	<b>0.00</b>	88.8	<b>0.00</b>

**Table 8: Marginal effects related to the generalized oprobit model (increase in the probability of choosing the category “completely satisfied” compared to the base category) and the related standard deviations**

	Communication		Social Support		Confidentiality		Dignity		Quality of Facilities	
	coeff	st dev	coeff	st dev	coeff	st dev	coeff	st dev	coeff	st dev
woman	-0.019	0.003	-0.010	0.002	-0.027	0.004	-0.010	0.001	-0.036	0.005
higher education	0.009	0.001	-0.033	0.005	-0.003	0.000	-0.014	0.002	-0.047	0.007
work	-0.023	0.004	0.012	0.002	-0.007	0.001	0.008	0.001	0.004	0.001
single	-0.018	0.003	0.008	0.001	-0.010	0.001	-0.030	0.003	-0.019	0.003
outside the province	0.004	0.001	0.046	0.007	0.080	0.009	0.006	0.001	0.011	0.002
outside the region or country	0.129	0.014	0.037	0.005	0.090	0.010	0.094	0.008	0.046	0.007
hospital dummy 2	0.023	0.004	0.024	0.004	0.088	0.010	0.066	0.006	0.049	0.007
hospital dummy 3	-0.071	0.013	0.082	0.011	-0.056	0.009	-0.059	0.008	-0.124	0.019
hospital dummy 4	0.145	0.018	0.034	0.005	0.060	0.007	0.313	0.012	0.063	0.009
hospital dummy 5	-0.047	0.009	-0.006	0.001	0.003	0.000	-0.069	0.009	-0.043	0.007
hospital dummy 6	-0.100	0.017	-0.102	0.017	-0.101	0.015	-0.109	0.014	-0.113	0.017
hospital dummy 7	-0.038	0.007	-0.064	0.011	-0.030	0.005	0.084	0.007	0.098	0.012
hospital dummy 8	0.029	0.005	0.259	0.022	0.157	0.014	0.014	0.002	0.055	0.008
hospital dummy 9	-0.020	0.004	-0.056	0.009	-0.027	0.004	-0.052	0.006	0.041	0.006
surgery	-0.008	0.001	0.020	0.003	-0.033	0.005	-0.023	0.003	0.015	0.002
emergency	-0.030	0.005	-0.011	0.002	0.015	0.002	-0.002	0.000	0.011	0.002
previous hospitalization	-0.006	0.001	-0.020	0.003	-0.020	0.003	0.032	0.003	-0.031	0.005

Note: all the marginal effects have a p-value <0.001

## **6. Discussion**

This paper considers the influence of patients' characteristics on the evaluation of health system responsiveness, by using a sample of about 2500 in-patients collected by the ASSR between January 2010 and December 2012. The dataset we use in our analysis is not large, but it has never been explored before, and the richness of the information it contains allows us to investigate original research questions. In particular, we focus on the influence that both the state of health of the in-patients and their experiencing pain have on their reporting behaviour with regard to the question of the health system's responsiveness. Our results suggest that patients in poorer health have a tendency to report a lower level of responsiveness than do patients in better health, all other things being equal, while patients experiencing pain are more likely to make use of the extreme categories of responsiveness, that is, the "completely dissatisfied" or "completely satisfied" categories.

When we look at the magnitude of the influence on responsiveness of the regressors included in our econometric model, the aforesaid influence appears to be particularly strong and statistically significant for the hospital dummies (the probability of choosing the response category "completely satisfied" is between 1% and 31% higher, compared to the base hospital category). This empirical result suggests that an investigation of which hospital characteristics might have a "true" influence on responsiveness could be a fruitful direction for future research. Studies making international comparisons, for instance, have pointed to a positive correlation, at national level, between high health expenditure per capita, high educational qualifications and high levels of responsiveness (World Health Report 2000, Anderson G., Hussey P. 2001, Blendon et al. 2001, Valentine et al. 2003b, Robone et al. 2011). Therefore, at a more micro level, it might be interesting to investigate the relative strength of various supply-side factors affecting the responsiveness of the health system, such

as the level of hospital staff training (in particular, their training in managing interpersonal relationships), and the level of spending on non-clinical facilities or the latest technology.

We are able to adjust the dependent variables for the presence of reporting heterogeneity by estimating a generalised ordered probit model (GOP). Several recent studies addressing the issue of reporting heterogeneity when investigating responsiveness have employed the hierarchical ordered probit model (HOPIT) and have taken advantage of anchoring vignettes (see, for example, Valentine et al. 2003b, Sirven et al. 2012, Rice et al. 2012). “Vignettes represent hypothetical descriptions of fixed levels of a latent construct such as health system performance. Since the vignettes are fixed and pre-determined, any systematic variation across individuals in the rating of the vignettes can be attributed to differences in reporting behaviour” (Rice et al. 2010) and can be utilized to adjust self-reported experiences of health system performance to enhance comparability across individuals (Bago d’Uva et al. 2008).

We cannot make use of the HOPIT model in our paper, since vignettes are not available in the dataset collected by the ASSR. From a more general viewpoint, the inclusion of vignettes within the survey necessarily entails a significant increase in the cost of data collection, which is not always affordable for the institutions implementing it. Therefore, the use of the GOP model – which does not require the employment of vignettes - can be considered as a second-best strategy. Despite the identification restriction of excluding each covariate from either the cut-point equation or the main responsiveness equation, the GOP model permits adjustment for reporting heterogeneity in those cases where vignettes are not available. In our analysis, correcting for reporting heterogeneity has proved extremely important in order to obtain unbiased results in the main responsiveness equation (for certain regressors, the results obtained using the OPROBIT model vary considerably from the results obtained using the GOP model). Therefore, the use of techniques capable of addressing the

issue of reporting heterogeneity has to be suggested to policy makers for sound policy recommendations to be formulated.



## References

- Adesanya T., Gbolahan O., Ghannam, O., Miraldo M., Patel B., Verma R., Wong H., (2012).,“Exploring the responsiveness of public and private hospitals in Lagos, Nigeria”, *Journal of Public Health Research*, volume 1 (e2).
- Anderson G., Hussey P., (2001), “Comparing Health System Performance in OECD Counties”, *Health Affairs*, 20 (3).
- Bago d'Uva, T., van Doorsaer, E., Lindeboom, M., O'Donnell, O., (2008), “Does reporting heterogeneity bias the measurement of health disparities?” *Health Economics*, 17 (3), 351-375.
- Baker PN, van der Meulen, Lewsey J., Gregg P. J., “The role of pain and function in determining patient satisfaction after total knee replacement”, *The Journal of Bone and Joint Surgery*, vol. 89-B no. 7 893-900.
- Blendon R. J., M. Kim, J. M. Benson, (2001), “The Public Versus The World Health Organization On Health System Performance”, *Health Affairs*, 20 (3).
- Blendon R. J., C. Schoen, C. DesRoches, R. Osborn, K. Zapert, (2003), “Common Concerns Amid Diverse Systems: Health Care Experiences In Five Countries”, *Health Affairs*, 106-121.
- Cortina, J.M., (1993), “What is coefficient alpha? An examination of theory and applications”, *Journal of Applied Psychology*, 78, 98–104.
- Dewey M.E and Prince M.J., (2005), “Mental health” in Boersch-Supan, A. et al. (Eds.), in *Health, Aging and Retirement in Europe: first results from SHARE*, Mannheim, Research Institute for the Economics of Aging, pp 108-117.
- Green W., (2008), *Econometric analysis*, Prentice Hall, 7<sup>th</sup> Edition.
- Jones A.M, Rice N., Bago d’Uva T., Balia S., (2007), *Applied Health Econometrics*, Routledge eds.
- Jones A.M., Rice N., Robone S., Rosa-Diaz P., (2011), “Inequality and Polarization in Health Systems Responsiveness: a Cross-Country Analysis”, *Journal of Health Economics*, 30 (4), 616-625.
- Labarere, J. Francois, P. Auquier, P. Robert, C. Fourny, M, (2001), “Development of a French inpatient satisfaction questionnaire”, *International Journal for Quality in Health Care*, 13, 99–108.
- Kowal P., Naidoo N., Williams S.R., Chatterji S., (2011), “ Performance of the health system in China and Asia as measured by responsiveness”, *Health*, 3(10), 638-646.
- National Institute for Health and Clinical Excellence (NICE), (2012), “Patient experience in adult NHS services: improving the experience of care for people using adult NHS services”, London, National Clinical Guideline Centre.
- Njeru M.K., Blystad A., Nyamongo I.K., Fylkesnes K., (2009), “A critical assessment of the WHO responsiveness tool: lessons from voluntary HIV testing and counselling services in Kenya”, *BMC Health Services Research*, 9,243, 1-11.
- Pelzer K., (2009), “Patient experiences and health system responsiveness in South Africa”, *BMC Health Services Research*, 9,117, 1-12.
- Puentes Rosas E., Gómez Dantés O., Garrido Latorre F., (2006), “The treatment received by public health services users in Mexico”, *Rivista Panamericana de Salud Publica*, 19 (6), 394-402.
- Pudney, S. and Shields, M., (2000), “Gender, Race, Pay and Promotion in the British Nursing Profession: Estimation of a Generalized Ordered Probit Model”, *Journal of Applied Econometrics*, 15 (4), 367-399.

- Radishan A., Kavosi Z., Majdzadeh R., Pourreza A., Pourmalek F., Arab M., Mohammad K., (2011), "Assessing Health System Responsiveness: A Household Survey in 17th District of Tehran", *Iranian Red Crescent Medical Journal*, 13(5), 302-308.
- Rice N., Robone S., Smith P.C., (2010), "International Comparison of Public Sector Performance: The Use of Anchoring Vignettes to adjust Self-Reported Data", *Evaluation*, 16 (1), 81-101.
- Rice N, Robone S., Smith P.C, (2012), "Vignettes and health systems responsiveness in cross-country comparative analyses", *The Journal of The Royal Statistical Society, Series A*, 175 (2).
- Robone S., Rice N., Smith P., (2011), "Health systems` responsiveness and its characteristics: a cross-country comparative analysis", *Health Services Research*, 46(6), 2079-2100.
- Sirven, N., Santos-Eggimann, B., Spagnoli, J., (2012), "Comparability of Health Care Responsiveness in Europe. Using anchoring vignettes from SHARE", *Social Indicators Research*, 105(2), 255-271.
- Smith, C. (1992), "Validation of a patient satisfaction system in the United Kingdom", *Quality Assurance in Health Care*, 4, 171–177.
- Steine, S. Finset, A. Laerum, E., (2001), 'A new, brief questionnaire (PEQ) developed in primary health care for measuring patients` experience of interaction, emotion and consultation outcome.' *Family Practice*, 18(4), 410–419.
- Valentine, N. B., De Silva, A., Kawabata, K., Darby, C., Murray, C. J. L. and Evans, D., (2003a), "Health system responsiveness: concepts, domains and operationalization" in Murray CJL., Evans, DB. Eds. *Health systems performance assessment: debates, methods and empiricism*. Geneva, World Health Organisation, 573–596
- Valentine N. B., Ortiz J.P., Tandon A., Kawabata K., Evans DB., Murray CJL., (2003b), "Patient Experiences with Health Services: Population Surveys from 16 OECD Counties", in Murray CJL., Evans, DB. Eds. *Health systems performance assessment: debates, methods and empiricism*. Geneva, World Health Organisation, 643 – 652.
- Valentine N., Darby C., Bonsel G.J., (2008), "Which aspect of non-clinical quality of care are most important? Results from WHO`s general population surveys of "health system responsiveness" in 41 countries", *Social Science and Medicine*, 66, 1939-1950.
- Valentine N., Prasad A., Rice N., Robone S., Chatterji S., (2009), "Health Systems Responsiveness - a measure of the acceptability of health care processes and systems", in Smith P., Mossialos E. Leatherman S. (Eds.) *Performance measurement for health system improvement: experiences, challenges and prospects*, WHO European Regional Office, London.
- Taylor R., (1990), "Interpretation of the Correlation Coefficient: A Basic Review", *JDMS*, 1, 35-39
- Terza, J. V., (1985), "Ordinal probit: a generalization",. *Communication in Statistics*, 14(1), 1-11.
- Wooldridge, J., (2002), *Econometric analysis of cross section and panel data*, MIT Press, Cambridge, MA
- World Health Organization, (2000), *The World Health Report 2000. Health systems: improving performance*, Geneva: World Health Organization.
- World Health Organization, (2008), *The Tallin Charter: Health Systems for Health and Wealth. WHO European Ministerial Conference on Health Systems*. Tallinn, Estonia, 25-27 June, 2008.



Alma Mater Studiorum - Università di Bologna  
DEPARTMENT OF ECONOMICS

Strada Maggiore 45  
40125 Bologna - Italy  
Tel. +39 051 2092604  
Fax +39 051 2092664  
<http://www.dse.unibo.it>