Regional Disparities in Self-Reported Health: Evidence from Chinese Older Adults

Ren Mu*

Texas A&M University

August 2010

Abstract Self-reported health (SRH) is widely used in studying health disparities despite the subjectivity inherent in individuals' interpretation of good health. With data from the pilot survey of the new China Health and Retirement Longitudinal Study, the author takes the vignette approach to control for differences in individual response scales and compares regional differences in health status among the elderly in China. The results show that regional disparities are substantially underestimated if differentials in response scales are unaccounted for. Based on common response scales, the elderly in a poor province are 15-26% more likely to report bad health than those in a wealthy province. The disparities in SRH cannot be explained by the observed differences in individual characteristics and morbidities and are completely driven by unobservables.

Keywords: self-reported health status, vignettes, regional disparities, elderly, China

JEL: I00, J14, O12, O1

* For helpful comments, the author thanks Joanna Lahey, Elaine Liu, John Strauss, Lori Taylor, Dominique van de Walle, Xiaobo Zhang, and Yiduo Zhang. Address for correspondence: The Bush School of Government and Public Policy, Texas A&M University, 4220 TAMU, College Station, TX, 77843. Email: rmu@tamu.edu; Tel: 1-979-458-8024. Fax: 1-979-845-4155.

1. Introduction

Socioeconomic inequality in health spending, use of medical service, and health outcomes has posed a critical challenge for developed and developing countries alike (e.g., van Doorslaer et al., 1997; van Doorslaer et al., 2006; Humphries and van Doorslaer, 2000; Wagstaff and van Doorslaer 2000; Wagstaff et al., 2003). A case in point involves regional health disparities in China. Under the decentralized public health financing system, Chinese local governments execute about 90% of the spending on health (World Health Organization [WHO], 2005a); but without adequate resources poor regions suffer from fewer and lower-quality services and the patient bears a higher proportion of the costs. Recent evidence shows little sign of abatement in the inequality in regional health spending (Chou and Wang, 2009), and regressive provision of health care has worsened the inequality in health outcomes. Interprovincial inequality in infant mortality and life expectancy at birth increased between 1980 and 2000 (Yip and Mahal, 2008), and the gap in infant mortality between urban and rural areas has widened (Zhang and Kanbur, 2005). Although the childhood mortality rate in developed coastal areas mirrors that of industrialized countries, the rate in most western provinces is 3-5 times higher (WHO, 2005b).

Compounding the problem of regional disparities in health, the population in China is aging rapidly because of declining fertility and increasing life expectancy. Projections have indicated that by 2030, more than 65% of disease burden is to be borne by the elderly (Chatterji et al., 2008). Because of the sagging social safety net, families will continue to fill the increasing need for old-age care. The health condition of the elderly thus affects various aspects of the lives of their children, for example, their decision to migrate (Giles and Mu, 2007). Through the family-care system, regional health disparities among the aged may perpetuate themselves in the form of inequality in opportunities for the younger generation.

Amidst the demographic trend of population aging, the Chinese government has in very recent years committed to injecting new public funds into health care and has initiated a large set of reforms in the health sector.[1] Knowledge of the distribution of the health status of the elderly is therefore essential in reforming the health care system so that it can respond effectively to the needs of an aging population while achieving the goal of equity. Notwithstanding its significance for health and welfare policies, few studies have focused on the distribution of health conditions among the Chinese elderly. The purpose of this paper is to shed light on this issue by comparing the health status of the elderly in a wealthy province in China to that of the elderly in a poor one.

An immediate challenge in the endeavor to study health disparities lies in the difficulties in measuring individual health status. Well known as essentially multidimensional, health is determined by interacting physical, mental, and emotional factors (Ware et al., 1980). The multidimensionality of health calls for using multiple health indicators in empirical studies (Strauss and Thomas, 1998), and it also partially explains why self-reported health (SRH)[2]—designed to capture an individual's subjective interpretation of his or her overall medical and functional status—is arguably the most commonly used measure of health in social science studies.[3]

[1] For studies analyzing these reforms, see articles in the 2009 Special Issue of *Health Economics*.

[2] A typical SRH question is phrased as follows: "In general, how would you rate your health?" Respondents are asked to choose a point along a 5-point scale, for example: (1) excellent, (2) very good, (3) good, (4) fair, and (5) poor.

[3] For studies using SRH to analyze how health is related to a broad range of socioeconomic outcomes, see, for example, Adams et al. (2003), Benzeval et al. (2000), Bockerman and Ilmakunnas (2009), Contoyannis et al. (2004), Deaton and Paxson (1998), Ettner (1996), and Smith (1999).

Imperfect as a measure of health, SRH is nevertheless a convenient way to solicit information on individual health and readily available in many socioeconomic surveys, where a comprehensive and objective measure of a person's general health condition is difficult if not impossible to achieve. Also a good predictor of subsequent mortality (Idler and Benyamini, 1997; Idler and Kasl, 1995)[4], SRH has independent value because it captures knowledge and interpretations not reflected in more objective indicators (Wallace and Herzog, 1995). Moreover, in medical research, health perception itself is considered one of the core concepts in measuring quality of life and clinical outcomes (Patrick and Erickson, 1993).

Despite the foregoing advantages of SRH as a measure of general health, measurement errors may potentially prevent comparisons of SRH across individuals because individual interpretations of "good health" may differ; that is, individuals with the same level of "true" health can report different health status. Such difference may relate to an individual's age, sex, education, past use of health services, personal experience with illness, individual optimism, and other cognitive biases.[5] For example, Lokshin and Ravallion (2008) found that the poor in Russia adapted to ill health by overstating their health status. Difference can also arise when SRH is used by a respondent as a justification for her or his employment decision or receipt of disability allowance (Kerkhofs and Lindeboom 1995). Furthermore, differentials in cultural and social norms can also act as confounding factors (Murray et al., 2001).

The differences in SRH across individuals, therefore, may contain disparities in real health conditions that are inseparable from the dissimilarities in the way individuals evaluate

[4] Another piece of evidence for the validity of respondents' subjective evaluation of their health was provided by Hurd and McGarry (1995), who concluded that respondents' subjective survival probability in the Health and Retirement Survey fit well in population probabilities.

[5] This is a more severe problem in cross-section studies as opposed to studies using panel data.

health or the diversity in their response scales.[6] So a meaningful comparison in SRH across groups should be preceded by an examination of whether the response scale systematically differs across and within those groups. Confirmation of scale differences then requires a rescaling of the responses so that the SRH comparison provides a valid measure of health disparities.[7]

One way to identify variations in individuals' response scales relies on their assessments of health conditions in a hypothetical vignette.[8] The vignette approach has been applied to the comparisons of self-reported work disability and specific health domains within and across countries (D'uva et al., 2008; Kapteyn et al., 2007; Salomon et al., 2004; Tandon et al., 2003). Extending the application of the vignette method to analyze SRH among the elderly in China, I take advantage of the vignette component in the pilot survey of the new China Health and Retirement Longitudinal Study (CHARLS) to achieve a valid comparison based on a common response scale adjusted for differences in individual perceptions of good health. Estimations using different sets of vignettes confirmed that the elderly in a poor province are 15-26% more likely to report bad health than those in a wealthy province. If differences in response scales were unaccounted for in the estimations, regional health disparities would be underestimated by 30-90%. It seems that in the case of China, a naïve comparison of SRH is unlikely to uncover true health differentials across regions.

[6] Recent studies also showed that respondents' answers to self-assessed health questions were affected by both the nature of the survey (particularly whether responses were written or verbal) and the sequence of questions (Clarke and Ryan, 2006; Crossley and Kennedy, 2002; Lee and Grant, 2009). These problems are relevant if multiple surveys are used to study health distributions with SRH. If respondents are surveyed with the same survey mode and are given the same sequences of questions, then these problems themselves will not invalidate the comparisons of SRH across respondents.

[7] Another argument made against using SRH in measuring health disparities is that it is measured on a discrete scale and only imperfectly captures fine gradations in health distributions. For methods and applications of imposing cardinality on SRH responses, see Lauridsen et al. (2004) and van Doorslaer and Jones (2003).

[8] An alternative approach to analyze the response heterogeneities in SRH is mapping SRH to some objective measure of health. See Etilé and Milcent (2006) and Lindeboom and van Doorslaer (2004).

An immediate question after the finding of a regional inequality in SRH is what accounts for the disparity. The second goal of this paper is to address this question by decomposing the total differences in SRH into differences due to observables, such as demographic characteristics and major disease prevalence, and differences unrelated to the observables. The decomposition results show that the observed differences cannot explain much of the health disparities; instead the regional inequality is entirely driven by unobservables. The difference in SRH unexplained by the data implies that SRH may contain more information about health condition than the objective measures such as disease prevalence does. Consistent with the notion that health is truly multidimensional, this result points to the difficulties in addressing regional health inequality.

The next section describes the data and presents descriptive statistics. Section 3 explains the econometric model used for estimating individual response scales based on vignettes. It also outlines the procedure for decomposing the total health inequality. Section 4 reports and discusses the empirical results. Section 5 concludes the paper.

2. Data and Descriptive Statistics

A. Data

The data used in the empirical analysis are from the pilot survey of the China Health and Retirement Longitudinal Study (CHARLS), which was part of a set of longitudinal aging surveys taken in the United States, England, and 19 countries in continental Europe, Korea, Japan, and India.[9] The pilot survey, conducted in Zhejiang and Gansu provinces in 2008, yielded data on roughly 1,600 households with 2,685 individuals aged 34-93. The sample was representative of

[9] CHARLS is conducted by the National School of Development (China Center for Economic Research) at Beijing University. See http://charls.ccer.edu.cn/charls/ for further details.

the elder population in each province. The data on 1,867 respondents 50 years of age or older are used in this paper.[10]

Large variations across provinces within China are exemplified by the differences between Zhejiang and Gansu. Located in the humid subtropical and economically advanced eastern coastal region, Zhejiang is traditionally known as "the land of fish and rice." At the time of the survey, it was one of the richest provinces, famous for its small businesses in industrial clusters and its bulk production of consumer goods for both domestic and international markets. It followed Shanghai and Beijing with the highest rural and urban income per capita in 2007. Gansu, located in western inland with a semiarid to arid climate, has a mining and mineral extraction-based economy. By income per capita in 2007, it ranked as the poorest province.

The CHARLS survey data confirmed a large income gap between the two provinces; they showed that household income and assets per capita were, respectively, 3.3 and 6.5 times higher in Zhejiang than in Gansu. Disparities in health care use were also evident in the data (Table 1). About 46% of the elderly in Gansu who were sick during the month prior to the survey forewent medical treatment, compared to 23% in Zhejiang. Regarding the primary reason for not seeking care, "being poor" was cited by about 25% of those who were not treated in Gansu, whereas the number was 13% for Zhejiang. Moreover "inconvenient traffic" was a more important reason in Gansu than in Zhejiang for foregoing treatment.

Given the vast and various differences between these two provinces regarding geographic features, traditions, economic activities, and access to health care, when comparing SRH in these two provinces, one should be concerned about whether people there have different views about

[10] Migration is not a concern in this sample. Half the respondents were born in the village or neighborhood in which they resided at the time of the survey. Around 89% of them lived in their birth county and 95.4% in their birth province. For the 4.6% born outside the province, the average length of residence in the current province was 32.6 years.

what constitutes good health. In other words, one should ask whether people in Gansu use systematically different response scales when evaluating their health than people in Zhejiang. If so, such difference must be taken into account before conclusions about regional differences in health status can be drawn. Before examining their views about health, turning first to the evaluations of their own health is necessary.

B. Differences in Self-Reported Health, Morbidity, and Individual Characteristics

All the respondents in the sample were asked to assess their health twice: first at the beginning of the health status section, where they were asked about their diagnosed chronic diseases and treatments, health functioning limitations, lifestyle, and health behavior; and once again at the end of the section. A different 5-point scale was used each time to measure self-reported health status. One scale included the following responses: (1) excellent, (2) very good, (3) good, (4) fair, and (5) poor; and the other, (1) very good, (2) good, (3) fair, (4) poor, and (5) very poor.[11] The order with which these two scales were used for each respondent was randomly assigned.

The distributions of SRH in the total sample and in the samples from Gansu and Zhejiang are shown separately in the upper panel of Table 2. In the total sample, about 41% of the elderly reported "fair" health, and 25% reported "poor" health. Comparing the SRH of the two provinces, one can clearly see that the elderly in Zhejiang reported better health than those in Gansu. About 78% of the elderly in Gansu reported their health as "poor" or "fair" in contrast to the 59% who did so in Zhejiang. In particular, the percentage of the elderly reporting "poor" health was 36% in Gansu but only 19% in Zhejiang. By contrast, the percentage reporting "excellent" and "very good" health was 17% in Zhejiang but 7% in Gansu. Despite the clear

[11] The author used the first 5-point scale—(1) excellent, (2) very good, (3) good, (4) fair, and (5) poor— in the analysis below. Results based on the alternative 5-point scale were consistent.

differences in the SRH of the elderly in these two provinces, little distinction emerged in terms of demographic characteristics. In both samples, the average age was about 63 with 49% being women. Average years of schooling was less than 3, and about 78% of the elderly lived in the rural area.

Regarding disease prevalence, differences between the two provinces were again very apparent. Figure 1a and Figure 1b show the locally weighted nonparametric regression lines of the prevalence rates of a selection of 10 chronic conditions for each sample.[12][13] Several of the health conditions, including lung disease, heart disease, stomach problem, stroke, and arthritis were more prevalent in Gansu than in Zhejiang for almost all ages. Others such as hypertension, high cholesterol, diabetes, and liver disease appeared to be more prevalent in Zhejiang except for the oldest ages. Albeit these differences, major diseases for these two provinces were the same. In the order of prevalence rates, they were stomach or other digestive disease, hypertension, lung disease, and heart problems. Three out of these four major diseases had higher prevalence rates in Gansu than in Zhejiang. Thus, Figure 1 suggests that the Gansu elder population may be less healthy than that of Zhejiang, an observation consistent with the average SRH reported in Table 1.

Again, the direct comparisons of SRH may not be valid if respondents in these two provinces used different response scales when evaluating their health. In the next section their vignette answers have been examined for any systematic differences in their responses.

C. Differences in Vignette Evaluations

[12] The CHARLS questionnaire listed13 health conditions, three of which with very low prevalence rates are unreported. These three conditions were cancer, psychiatric problems, and memory problems with the prevalence rates of 0.009, 0.009, and 0.015, respectively.

[13] The self-reported diseases were either diagnosed by medical professionals or self-diagnosed by the respondents. Among the listed conditions, stomach problem, lung disease, and arthritis had the highest rates of self-diagnosis at 18%, 23%, and 33%, respectively. No significant difference between the self-diagnosed rates in the two provinces was apparent.

Health vignettes describe the health conditions of a hypothetical person. The CHARLS contained vignettes for six health domains: body pain, sleep disorder, difficulty in mobility, cognition problems, shortness of breath, and emotional problems (or affect). In each domain, three vignettes questions were presented, and the ordering of the vignette descriptions proceeded from least severe to most severe. All vignettes were presented with a female or male name, and randomized across households but remained the same for all respondents within a household. The vignette evaluations were given on the 5-point scale: (1) none, (2) mild, (3) moderate, (4) severe, and (5) extreme. As examples of the vignettes, the followings were the three questions in the pain domain:

1. Zhang Jun/Wang Hong has a headache once a month that is relieved after taking a pill. During the headache he/she can carry on with his/her day-to-day affairs. Overall, in the last month, how severe were Zhang Jun/Wang Hong's body aches or pain?

2. Zhou Wei/Li Li has pain that radiates down his/her right arm and wrist during his/her day at work. This is slightly relieved in the evenings when he/she is no longer working on his/her computer. Overall, in the last month, how severe were Zhou Wei/Li Li's body aches or pain?

3. Zhao Liang/Zhou Yan has pain in his/her knees, elbows, wrists, and fingers, and the pain is present almost all the time. Although medication helps, he/she feels uncomfortable when moving around and lifting things. Overall, in the last month, how severe were Zhao Liang/Zhou Yan's body aches or pain?

Respondents in a random subsample of households[14] were asked to evaluate the health conditions of the hypothetical persons. They were given two randomly selected health domains with three vignette descriptions for each. Since the health condition of the person described in a vignette is constant across indiviudals by design, the differences in the vignette assessments can be caused only by the difference in individual response scales. The vignettes evaluations can then be used to elicit individual ratings that reflect individual perceptions, norms, and expectications about health.

Table 3 compares the vignette evaluations completed in Gansu and Zhejiang. Although the health conditions described in the vignettes were the same, distinct patterns in the evaluations were apparent for the two provinces. In each of the six domains, Gansu elders were less likely to report "none" or "mild" for the first vignette describing people with the least severe health condition than those in Zhejiang. This pattern indicates that the elderly in Gansu were softer on the people depicted in the vignettes as having relatively mild conditions. In other words, compared to the Zhejiang elderly, Gansu elders tended to overstate the severity of a mild condition. For the third vignettes, which described the most severe condition, the Gansu elderly used "none" or "mild" more often in four of the six domains (pain, sleep, cognition, and affect). For these severe conditions, they were also less likely than the Zhejiang elders to report "severe" or "extreme" in four domains (pain, mobility, cognition, and affect). Therefore, Gansu elders appeared to be harder on people depicted in the vignettes as having severe conditions, or they tended to understate the problem if it were in fact very serious. The differences in the answers for the second question, which described vignette persons with moderately severe conditions, exhibited a less clear pattern than those found for the first and the third vignettes. However,

[14] Half the total sample was supposedly randomly selected to answer the vignette component. In actuality about 40% of the sample was given the vignette component. .

even here some evidence was shown that the Gansu elderly more often used the “none” or “mild” categories than the Zhejiang elderly.

The pattern emerging from the vignette evaluations that Gansu elders understated the severity of more serious conditions implies that a direct comparison of SRH may be misleading. The commonly used two-scale health measure often categorizes the answer “fair” or “poor” as “bad health.” If the response scales of Zhejiang elderly were used for both provinces, one would expect an increase in self-reported bad health among the Gansu elderly. Therefore the difference in self-reported bad health based on a common scale would be larger than the one based on different response scales.

3. Empirical Methods

A. Estimating SRH Using Vignette Evaluations

The vignette assessments can be incorporated into the traditional ordered probit model to account for difference in individual response scale as suggested by Tandon et al. (2003) and King et al. (2004). The new model is generally known as the hierarchical ordered probit model (HOPIT). Assume the unobserved health condition varies across individuals and is denoted as Y_i^* for respondent i, and it is a linear function of the individual observed charaterstics X_i with an independent normal error term ε_i given by:

$$Y_i^* = X_i\beta + \varepsilon_i \qquad \varepsilon_i \sim N(0,1) \tag{1}$$

Included in X_i are age, gender, years of the schooling, and place of residence (rural or urban) of the respondent. I also control for various health problems, such as whether the respondent had hypertension, high cholesterol, diabetes, lung disease, heart disease, stomach problem, stroke, kidney disease, and arthritis. Recall that the 5-point scale [(1) excellent, (2) very good, (3) good, (4) fair, and (5) poor] used on the SRH question was randomly assigned to respondents either before or after they answered the health functioning questions. Studies have found that general health conditions are more positively reported when the question is placed after an objective health condtion measure (Bowling and Windsor, 2008; Lee and Grant, 2009). To account for the possible order effect, I also include a variable indicating whether or not the SRH question was placed before the health functioning questions.

The reported category of SRH for respondent *i* is given by y_i ,which has *J* ordinal responses *(J=5* in this case). Assume SRH is determined by the following reporting mechanism:

$$y_i = j \qquad if\ \tau_i^{j-1} < Y_i^* < \tau_i^j \quad j = 1, 2, \dots 5 \tag{2}$$

and the individual specific thresholds τ_i^j are molded in the following way:

$$\tau_i^j = \begin{cases} -\infty & if\ j = 0, \\ \gamma^1 X_i & if\ j = 1, \\ \tau_i^{j-1} + e^{r^j X_i} & if\ \ j = 2, 3, 4 \\ +\infty & if\ j = 5 \end{cases} \tag{3}$$

Note the thresholds are functions of individual characteristics (X_i); therefore, they vary over individuals. And the threshold parameters also vary over response categories because γ^j $(j = 1, 2, \dots 5)$ are different. The log likelihood function for SRH can thus be written as

$$l^s(\gamma, \beta) = \sum_{j=1}^{5} I(y_i = j) \log\left[\Phi\left(\tau_i^j - X_i\beta\right) - \Phi\left(\tau_i^{j-1} - X_i\beta\right)\right], \tag{4}$$

where $I(y_i = j)$ is an indicator function such that $I(y_i = j) = 1$ if $y_i = j$ and $I(y_i = j) = 0$ otherwise.

The individual specific thresholds capture the different response scales across respondents. Without vignettes evaluations, the parameters γ^j (for $j > 1$) and β can only be separately identified from the nonlinearities in the threshold model as specified in (3) or by a priori restrictions on which variables affect health and which affect reporting. Vignette evaluations allow a strong identification of γ^j and β separately through a vignette component in the model. Denote the actual level for the hypothetical person described in vignette l as θ_l, where $l = 1, \dots, L$ and L=18 because the total number of vignettes is 18 (three for each of the six domains). The perception of respondent k about the health condition of the person described in vignette l is given by

$$Z_{lk}^{*} = \theta_l + \varepsilon_{lk} \qquad \varepsilon_{lk} \sim N(0, \sigma_l^2), \tag{5}$$

where ε_{lk} is a random error and its variance is different over vignettes to allow the case that each of them may not be understood equally well. Note respondents who were asked to evlauate the vignettes were indexed by k to account for the fact that i and k may index different individuals because vignettes were asked of a subset of the total sample. The reporting of vignette evaluation was modeled using similar ordered response equations:

$$z_{lk} = j \qquad if\ \tau_k^{j-1} < Z_{lk}^{*} < \tau_k^{j} \quad j = 1, 2, \dots 5 \tag{6}$$

with five ordinal categories in the same ascending order, indicating the severity of the health problem as in the SRH question. Thresholds were determined by the same γ coefficients as in (3) with the same explanatory variables:

$$\tau_k^j = \begin{cases} -\infty & if\ j = 0, \\ \gamma^1 X_k & if\ j = 1, \\ \tau_k^{j-1} + e^{r^j X_k} & if\ j = 2, 3, 4 \\ +\infty & if\ j > 5 \end{cases} \tag{7}$$

Similarly, the log likelihood funciton for the vignett component takes the following form:

$$l^v(\gamma, \sigma_l^2) = \sum_{l=1}^{L} \sum_{j=1}^{5} 1[z_{lk} = j] \log [\Phi\left(\frac{\tau_i^j - \theta_l}{\sigma_l}\right) - \Phi\left(\frac{\tau_i^{j-1} - \theta_l}{\sigma_l}\right)] \qquad (8)$$

Since the likelihood functions in (4) and (8) contain the same parameters (γ), the efficient estimation involves the following log likelihood function:

$$l(\gamma, \sigma, \beta) = l^s(\gamma, \beta) + l^v(\gamma, \sigma_l^2) \qquad (9)$$

The parameters (γ and β) are separately estimated for the two provinces. The differences in γ capture distinctions in their response scales, and the differences in β show how individual charateristics and morbidity affect their reported health differently.

There are two assumptions underlying this approach of using vignettes to identify the threshold parameters γ^j. The first assumption, termed "response consistency," stipulates that individuals use the same response scale for SRH and for the assessment of the vignettes. It is formalized by γ^j being the same in both SRH as specified in (3) and the vignettes component in (7).[15] The second assumption is "vignette equivalence," which means that on average all respondents understand the vignettes in the same way (even though they may position the vignette differently on the scale) and any difference in their understanding is random. This assumption assures that vignette evaluations from a subsample are adequate to identify γ^j in the total sample. In the CHARLS data, the vignettes in each domain were ordered by the level of severity, with the least severe placed first and the most severe, third. Among the six health domains, the proportion of respondents whose assessments showed consistent rankings ranged from 67% for the sleep domain to 84% for the affect domain. This suggests considerable

[15] The supporting evidence for this assumption was provided by King et al. (2004) and van Soest et al. (2007).

variations of orderings among the respondents for different vignettes. The results are presented with different vignettes included in the estimations to check the robustness of the results.

B. Decomposing the Difference in SRH

It follows from the HOPIT model outlined above that the probabilities associated with the five categories in SRH are given by

$$\Pr(y_i = 1|\gamma, \beta, X) = \Phi[\tau_i^1 - X_i\beta]$$

$$\Pr(y_i = j|\gamma, \beta, X) = \Phi\big[\tau_i^j - X_i\beta\big] - \Phi\big[\tau_i^{j-1} - X_i\beta\big] \quad j = 2, 3, 4 \qquad (10)$$

$$\Pr(y_i = 5|\gamma, \beta, X) = 1 - \Phi[\tau_i^4 - X_i\beta]$$

Based on these equations, the reported health in the two provinces can be simulated and compared, using different response scales. For example, using the estimated Gansu scales ($\hat{\gamma}_g$), the differences between the two provinces in terms of percentages of reporting in each category is as follows:

$$\Delta^{g} = \widehat{\Pr}\big(y_g = j|\hat{\gamma}_{g,}\hat{\beta}_g, X_{ig}\big) - \widehat{\Pr}\big(y_z = j|\hat{\gamma}_{g,}\hat{\beta}_z, X_{iz}\big). \qquad (11)$$

Similarly, with the Zhejiang sclaes ($\hat{\gamma}_z$), the difference is

$$\Delta^{z} = \widehat{\Pr}\big(y_z = j|\hat{\gamma}_{z,}\hat{\beta}_g, X_{ig}\big) - \widehat{\Pr}\big(y_z = j|\hat{\gamma}_{z,}\hat{\beta}_z, X_{iz}\big) \qquad (12)$$

Furthermore, based on the same response scales, the above differences in (11) or (12) can be partitioned into explained and unexplained portions, following a nonlinear model version of the Blinder-Oaxaca decomposition procedure (Bauer and Sinning, 2008). In the case of using the Gansu scale, the differences are the summation of the following two components:

$$\Delta^{g} = [\widehat{\Pr}(y_g = j|\hat{\gamma}_{g,}\hat{\beta}_g, X_{ig}) - \widehat{\Pr}(y_z = j|\hat{\gamma}_{g,}\hat{\beta}_g, X_{iz})] + [\widehat{\Pr}(y_g = j|\hat{\gamma}_{g,}\hat{\beta}_g, X_{iz}) - \widehat{\Pr}(y_z = j|\hat{\gamma}_{g,}\hat{\beta}_z, X_{iz})] \tag{13}$$

The first component is the portions of differences due to differences in observable individual characteristics between the two provinces, assuming that the coefficients in Zhejiang on the determinants of self-reported health are given by the Gansu coefficients ($\hat{\beta}_g$). More specifically differences in individual demographics, such as age, gender and education, and differences in health problems are all capable of creating disparities in the reported health. The second term on the right-hand side of equation (13) shows the differential caused by the differences in unobservables. It is captured by the differences in the estimated coefficients, assuming the Gansu sample shared the same characteristics as the Zhejiang sample (X_{iz}). In order words, the second component tells what the difference between the two provinces would have been if the elders' observable characteristics had been the same in these two provinces.

Changing the reference group, one can alternatively express the decomposition by using Zhejiang coefficients in the first component and Gansu characteristics in the second:

$$\Delta^{g} = [\widehat{\Pr}(y_g = j|\hat{\gamma}_{g,}\hat{\beta}_z, X_{ig}) - \widehat{\Pr}(y_z = j|\hat{\gamma}_{g,}\hat{\beta}_z, X_{iz})] + [\widehat{\Pr}(y_g = j|\hat{\gamma}_{g,}\hat{\beta}_g, X_{ig}) - \widehat{\Pr}(y_z = j|\hat{\gamma}_{g,}\hat{\beta}_z, X_{ig})] \tag{14}$$

Equations (13) and (14) also apply to decompositions using Zhejiang scales ($\hat{\gamma}_z$) as the common scales.

4. Results

A. Estimations of SRH

The results of the response scale corrected model, using the full set of 18 vignettes are presented in Table 4 alongside those of the conventional ordered probit model, which assumes constant threshold cut-offs across individuals. The first two columns report the results for Gansu, and the rest of the table concerns Zhejiang. The full set of estimations for the cut-offs are reported in Appendix Tables 1a and 1b.[16]

Consistently across the provinces and across the two models, SRH are shown to be worse for women than for men. This result resembles findings in a number of other studies indicating that women on average report worse health than men.[17] The estimates of the gender difference for Gansu are similar in the two models, implying that worse SRH for women is not driven by systematic differences between women and men in terms of how health is reported. This result is consistent with the conclusion in Case and Paxson (2005). For Zhejiang, the estimated gender difference is larger in the HOPIT model than that in the ordered probit model, showing that gender difference will be underestimated if it is not corrected for differences in response scales. Comparing gender difference in SRH across the two provinces reveals that the gender gap is larger in Zhejiang. Age and education effects are not evident in Gansu. The age effects are significant in Zheijang, where older respondents reported worse health. It seems education is positively associated with SRH in Zhejiang, but the education effects are not significant in the HOPIT model once the differences in response scales are accounted for. For Gansu, the nonadjusted estimates show that the rural elderly have worse health, but the adjusted results imply that the rural-urban differential is negligible and statistically insignificant. The estimates for Zhejiang are weaker but also show a similar pattern. These results imply that the rural and urban differences in SRH are largely driven by heterogeneities in response scales. The appendix tables show that in both provinces, rural respondents differed from their urban counterparts in

[16] Note that the first threshold is a linear combination of the explanatory variables (equation 3), so the interpretations are straightforward. The coefficients for the other cut-offs show how the explanatory variables affect the log difference between two adjacent cut-offs.

[17] See for example Idler (2003), Molarius and Jason (2002), and Verbrugge (1989).

their use of thresholds. If such differences in response scales remain uncorrected, the rural and urban gap in SRH seems to be overestimated. Where the SRH question is placed on the questionnaire does not affect how respondents reported their health, and this is true for both provinces.

Elderly with chronicle heath problems tend to report worse health in both provinces. In Gansu the diseases that caused worse SRH include hypertension, lung problems, heart disease, stroke, stomach problems, and arthritis. The comparison of the results between the two models for Gansu shows that the impacts on SRH from hypertension, stroke, and arthritis are overestimated when the response scale effects are not taken into account. Morbidity also affects how individuals reported their health in Zhejiang (Table 1b). Evidence also shows that the morbidity effects in Zhejiang are overall underestimated with the exception of arthritis, whose effects are significant in the ordered probit model but not in the HOPIT model. Compared with Gansu, more diseases affected SRH in Zhejiang. The only health condition that does not lead to worse reported health in Zhejiang is high cholesterol. Moreover, health problems such as hypertension, diabetes, liver problems, stroke, and kidney diseases have greater impact on SRH in Zhejiang than in Gansu.

Based on the estimated coefficients for SRH and the response scales reported in Table 4 and Table A.1, the percentages of respondents in the five categories of self-reported health have been simulated as formulated in equations (9) and (10). Simulation results are reported in Table 5. The first two columns present the estimated SRH for Gansu and Zhejiang based on their own response scales. The middle two columns use Zhejiang scales, and the last columns are based on Gansu scales.

About 56% of the elderly in Gansu report bad health ("fair" or "poor"), and the number is 41% for Zhejiang. With separate response scales (but common within a province), the results show that Gansu elders are 15% more likely to report bad health than those in Zhejiang. If measured with Gansu scales, only 30% of the Zhejiang elderly would report bad health. If measured on Zhejiang scales, the proportion of Gansu elders reporting bad health would reach 68%. Once the differences in response scales are controlled for, the regional differentials in the percentage of reporting bad health changes from 15% to 26%, an increase of more than 70%.

A closer look at the five categories of SRH reveals that Gansu response scale leads to a higher concentration in the middle category ("good"), with less reporting in the two adjacent categories ("fair" and "very good"). This is somewhat consistent with the vignette assessment pattern reported in Table 3 that compared with those in Zhejiang, Gansu elders tend both to underestimate the severity of serious health problems and overestimate the less severe scenario.

B. Decomposing the Differences in SRH

Using common response scales, the differences between the two provinces can be further decomposed into differences resulting from observables and those from unobservables, as specified in equations (13) and (14). The decomposition results are reported in Table 6. The first four columns are based on Gansu scales, and the last four on the Zhejiang scales. With each common scale, Gansu parameters are first applied, then Zhejiang parameters.

Columns (1) and (2), similarly column (3) and (4), are based on the same response scales and the same parameters but differ only in individual observables. Take the results based on Gansu scales and Gansu parameter as an example: They show that 56 % of the Gansu elders and 59% of the Zhejiang elders reported bad health. The distributional differences between the two provinces in the observed individual demographics and morbidities actually dictate that a slightly higher percentage of the elderly would report bad health in Zhejiang than in Gansu. When using Zhejiang parameters (columns 3 and 4), the results are consistent, showing that differences in individual observables leads to a very small regional disparity. Obviously, the difference in the observables cannot explain the estimated 26% more elders with bad reported health in Gansu.

Corresponding to the above computations of differences resulting from observables, the differences caused by unobservables are calculated in two ways. The first one used Zhejiang as the reference group and compared the estimated distributions of health using Gansu parameters with the ones using Zhejiang parameters (column (2) and (4)). The second calculation used Gansu as the reference group (column (1) and (3)). The results are consistent across these two calculations: differences in unobservables explained all of the total differences between the percentages of bad health in the two provinces.

The conclusion that disparities in SRH are entirely driven by differences in unobservables begs more thought on what unobservables can possibly measure. The unobserved portion of the difference in SRH may have three potential sources. First, to the extent that unobserved regional differences, such as diseases that are not included in the survey, affect reported health, any resulting regional gap in health is deemed "unexplained" by a regression model. In this sense, the unobservable differences are a synonymous for data limitation. Second, over- or underreporting of observable medical conditions can also result in differences in SRH. Such misreporting, if it exists, is likely to be driven by respondents' lack of awareness of certain health problems, resulting from underutilization or poor quality of medical service. For these two scenarios, one can conclude that SRH captures more information about health conditions than the available morbidity data. The third source of unobservable differences in SRH is, of course, random errors. The decomposition cannot separately identify these three potential sources of unobserved components. It is reasonable to note, however, that addressing health disparity is a more challenging task when the decomposition results show that health disparity is mainly driven by unobservables than otherwise.

C. Robustness Checks with Different Vignettes

Because health is a multidimensional concept, all six available sets of vignettes designed to tap six different domains of health for a total of 18 vignettes are included in the above analysis. In order to check the robustness of the above results to the inclusion of different sets of vignettes, I also select a subset of the available vignettes to be included in the threshold estimation in the HOPIT model. The criteria for choosing vignettes are drawn from an underlying assumption of the vignette approach that individuals on average understand the actual levels described in the vignettes in the same way.

Table 7 contains an examination of the consistency of individual vignette ranking in each domain with the expected ordering of the vignettes for that domain. It reports the proportion of cases in which one vignette ordering is less than the other orderings through a matrix of pair wise comparisons. The expected ordering in each domain is that the rank of vignette 1 would be less than that of vignette 2, which in turn is than vignette 3. Inconsistency occurs with reversals in responses. For example, the proportion of cases, where vignette 2 is less than vignette 1 in the pain domain, is 21%. The total proportion of inconsistency in the pain domain also contains the reversed ordering of vignettes 3 and 1 and vignettes 3 and 2; and this total is 4%. Among the six domains, the sleep, cognition, and mobility domains have the highest percentage of inconsistency (33, 29, and 29% respectively), whereas the pain, breath, and affect domains have the lowest inconsistency rates (25, 17, and 16% respectively).

This information implies the existence of considerable consistency of vignette orderings and similar understanding of the severity levels described in the vignettes among the respondents; however, the degree of common understanding varied across health domains. As robustness checks to the full model with 18 vignettes, the six vignettes in two domains (breath and affect) with the lowest inconsistency are first included in the HOPIT model. I then also include three more vignettes in the pain domain. The simulation and decomposition results based on these two sets of vignettes are presented in Table 8.

With the six vignettes, the simulation results show that based on each province's own response scales, Gansu elders are 19% more likely to report bad health than their counterparts in Zhejiang. Corrected for the scale difference, the disparity reaches 25%, an increase of more than 30%. The decomposition analysis shows again that the differences in observables do not lead to differences in SRH, and all the disparities are caused by differences in unobservables. Based on the nine vignettes, the estimated health disparity is smaller—only 8% more of the elderly in Gansu reported bad health. But this is a considerable underestimation of the regional disparity: corrections of differences in response scales reveal that Gansu elders are 15% more likely to report bad health. In this case, the underestimation due to scales difference amounts to more than 90%. Regarding decomposition results, differences in observables imply that Gansu elders should have had slightly better SRH – they would have been 2-5% less likely to report bad health. The estimated disparity can be explained only by differences in unobservables, a result consistent with previous conclusions based on the 18 or six vignettes.

These robustness checks show that the estimated regional disparity can be different if vignettes of different health domains are included in the model. This finding implies that respondents may have used different scales for different health domains.[18] However, the robustness checks encourage the view that without the correction of differences in response scales, the regional disparity in SRH would be underestimated. They also show consistent results that the health differences are completely determined by unobserved factors.

5. Conclusions

Using a vignette approach, I analyze disparities in SRH among the elderly in a wealthy and a poor province in China. The results show that without correction for differences in response scales, the health differentials are underestimated, and the magnitude of the underestimation ranges from 30 to 90%, depending on various sets of vignettes included in the estimation. Specifically, the elders in the poor province are about 15 to 26% more likely to report bad heath than those in the rich province. Furthermore, this paper finds that the differentials in observed individual demographics and major morbidities cannot explain the health disparity, all of which are instead caused by unobservable differences.

Three implications can be drawn from these results. First, heterogeneities exist in the response scales of SRH across regions, and their impact on conclusions drawn about regional health disparity is substantial. For a more homogeneous region or country, response scale differences may not be so important in measuring and comparing individual SRH. But given the considerable regional and provincial variations within China, researchers and others gauging health disparity based on SRH must consider differences in response scales.

Second, the conclusion that difference in SRH is mainly caused by unobservables implies that SRH may contain more information about a respondent's health condition than the available objective indices of health, such as morbidity measures. Of course, one cannot rule out unexplained differences resulting from random errors in SRH. But given that decomposition is

[18] A similar conclusion about response scales across domains was documented in Kaptyn et al. (2007).

conducted with common response scales, it is more likely that the inability to explain disparities in SRH is caused by data limitations or by over- or underreporting of the medical conditions that are surveyed in the data. Therefore it is likely that differences in SRH may capture differences in real health.

Third, the conclusion that the elderly in the poor province are more likely to have bad health than those in the wealthy province has important policy implications. The recent policy of health insurance subsidies to rural residents in the less developed central and western provinces by the central and local governments is a welcome move to addressing regional inequalities in health, but it may not be adequate for narrowing the health gap among the elder population. For this purpose, resources allocated to economically less advantaged provinces should be more directly targeted to the elderly. Moreover, the health disparities among the elderly imply differentials in the demand for care on their household members. The need includes household members' financial resources as well as their time for physical care of the elderly. As a result, any disparities in health outcomes among the elderly are likely to affect the next generation through the family care system. Further research is needed to analyze the implications of regional health disparities among Chinese elderly with regard to the intergenerational transmission of inequality.

Table 1 Health Care Use in Gansu and Zhejiang

	Gansu	Zhejiang
Sick last month	0.381	0.251
Sick but did not seek treatment	0.461	0.234
___Already under treatment	0.152	0.078
___Illness is not serious; don't need treatment.	0.477	0.766
___ Poor	0.248	0.128
___No time	0.022	0.000
___ Inconvenient traffic	0.045	0.013
___No available treatment	0.028	0.000

Note: Data were weighted by individual weights adjusted for nonresponse.

Table 2 Self-Reported General Health Status and Individual Characteristics

	Total	Zhejiang	Gansu
Self-Reported Health Status			
___Excellent	0.026	0.028	0.022
	(0.149)	(0.149)	(0.149)
___Very Good	0.114	0.145	0.047
	(0.301)	(0.350)	(0.215)
___Good	0.207	0.234	0.151
	(0.398)	(0.424)	(0.356)
___Fair	0.407	0.400	0.421
	(0.490)	(0.490)	(0.491)
___Poor	0.246	0.193	0.359
	(0.448)	(0.399)	(0.485)
Individual Characteristics			
Gender (woman=1, man=0)	0.486	0.487	0.486
	(0.500)	(0.500)	(0.500)
Age	62.697	62.660	62.697
	(9.115)	(9.606)	(9.115)
Years of schooling	2.852	2.811	2.852
	(3.877)	(3.805)	(3.877)
Rural	0.774	0.778	0.774
	(0.399)	(0.396)	(0.399)
Number of Obs.	1867	1044	823

Note: Data were weighted by individual weights adjusted for nonresponse. Standard deviations appear in parentheses.

Figure 1a. Disease prevalence by province (I)

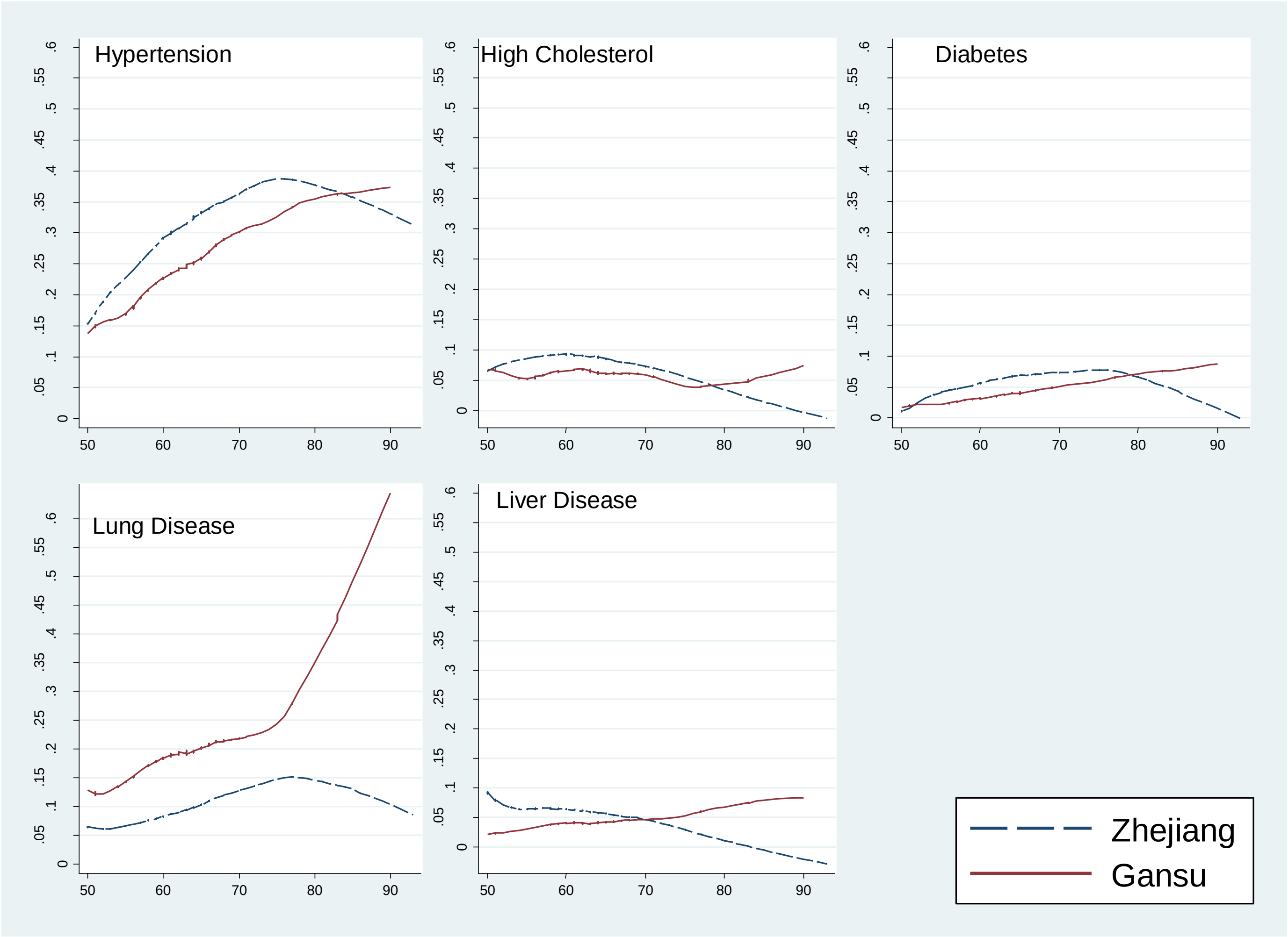

Figure 1b. Disease Prevalence by Province (II)

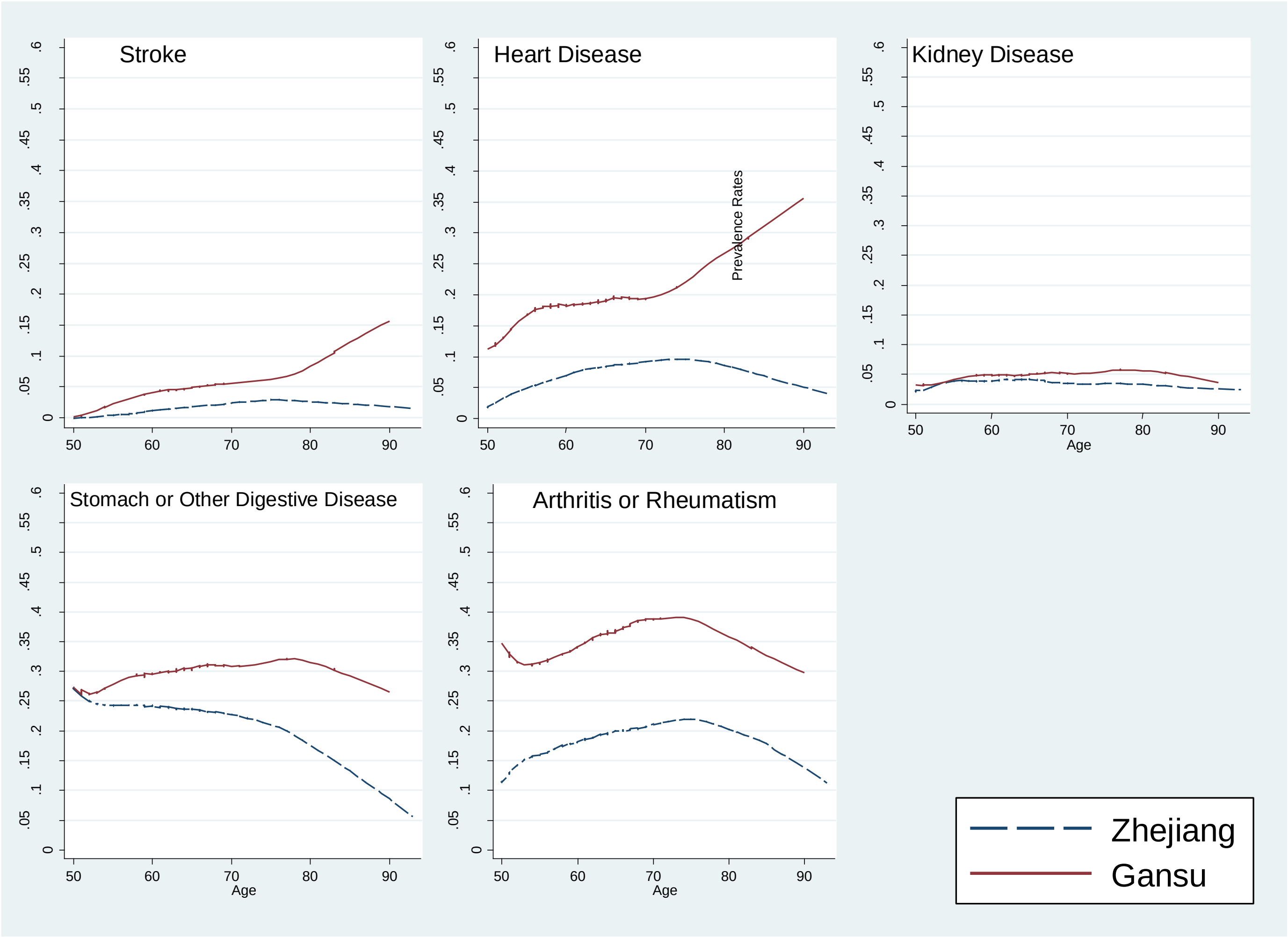

Table 3 Means of Vignette Answers

	Vignette 1		Vignette 2		Vignette 3	
	Zhejiang	Gansu	Zhejiang	Gansu	Zhejiang	Gansu
Pain (How much pain?)						
None	0.226	0.213	0.186	0.285	0.022	0.01
Mild	0.628	0.578	0.493	0.57	0.118	0.198
Moderate	0.098	0.159	0.29	0.117	0.433	0.582
Severe	0.038	0.05	0.031	0.027	0.393	0.181
Extreme	0.009	0.000	0.000	0.000	0.034	0.029
Sleep (How much difficulty?)						
None	0.218	0.144	0.101	0.015	0.044	0.081
Mild	0.54	0.46	0.311	0.263	0.164	0.231
Moderate	0.147	0.278	0.329	0.475	0.373	0.198
Severe	0.083	0.1	0.203	0.219	0.302	0.43
Extreme	0.013	0.018	0.056	0.028	0.118	0.06
Mobility (How much of a problem?)						
None	0.353	0.268	0.145	0.109	0.039	0.034
Mild	0.456	0.469	0.307	0.384	0.114	0.11
Moderate	0.136	0.153	0.379	0.298	0.242	0.34
Severe	0.054	0.101	0.162	0.192	0.457	0.441
Extreme	0.000	0.009	0.008	0.016	0.149	0.075
Cognition (How much of a problem?)						
None	0.35	0.451	0.126	0.25	0.066	0.237
Mild	0.547	0.394	0.408	0.37	0.206	0.332
Moderate	0.069	0.107	0.368	0.253	0.325	0.247
Severe	0.033	0.031	0.079	0.114	0.389	0.161
Extreme	0.000	0.017	0.02	0.013	0.014	0.022
Breathing (How much of a problem?)						
None	0.13	0.074	0.043	0.036	0.031	0.02
Mild	0.425	0.379	0.14	0.04	0.105	0.021
Moderate	0.268	0.374	0.317	0.349	0.127	0.153
Severe	0.169	0.153	0.382	0.454	0.406	0.492
Extreme	0.007	0.02	0.118	0.12	0.331	0.313
Affect (How much of a problem?)						
None	0.311	0.29	0.059	0.098	0	0.018
Mild	0.453	0.423	0.296	0.398	0.021	0.226
Moderate	0.169	0.202	0.464	0.369	0.169	0.157
Severe	0.064	0.076	0.172	0.095	0.46	0.365
Extreme	0.003	0.009	0.009	0.04	0.351	0.235

Notes: Data were weighted by individual weights adjusted for nonresponse.

Table 4 Estimations of Self-Reported Health Status by Province

	Gansu		Zhejiang	
	Ordered Probit	HOPIT	Ordered Probit	HOPIT
Female	0.209***	0.200*	0.212***	0.396***
	(0.071)	(0.109)	(0.067)	(0.085)
Age	0.000	0.003	0.010***	0.015***
	(0.005)	(0.007)	(0.004)	(0.005)
Years of schooling	-0.015	0.009	-0.020*	-0.017
	(0.010)	(0.014)	(0.011)	(0.012)
Rural	0.244**	0.073	0.159	0.006
	(0.101)	(0.136)	(0.102)	(0.111)
Reporting SRH before answering health functioning questions	0.127	0.016	0.060	-0.004
	(0.091)	(0.100)	(0.060)	(0.081)
Hypertension	0.392***	0.209	0.292***	0.251***
	(0.089)	(0.132)	(0.070)	(0.095)
High cholesterol	-0.013	0.028	0.057	0.168
	(0.190)	(0.233)	(0.144)	(0.165)
Diabetes	-0.224	-0.037	0.483***	0.587***
	(0.234)	(0.303)	(0.163)	(0.191)
Lung disease	0.495***	0.483***	0.283**	0.424***
	(0.097)	(0.140)	(0.112)	(0.153)
Liver disease	-0.027	0.169	0.632***	0.740***
	(0.217)	(0.285)	(0.118)	(0.177)
Heart disease	0.385***	0.340**	0.389**	0.477***
	(0.104)	(0.151)	(0.160)	(0.166)
Stroke	0.571***	-0.172	0.998***	1.162***
	(0.195)	(0.285)	(0.347)	(0.471)
Kidney disease	0.078	-0.239	0.545***	0.852***
	(0.216)	(0.239)	(0.148)	(0.254)
Stomach/other digestive disease	0.486***	0.426***	0.429***	0.375***
	(0.080)	(0.117)	(0.067)	(0.098)
Arthritis/rheumatism	0.171**	0.071	0.222**	0.059
	(0.073)	(0.107)	(0.104)	(0.112)
Log-likelihood	-964.071	-903.97	-1,370.7	-1,330.7
Number of obs.	823		1044	

Notes:*significant at 10%; ** significant at 5%; *** significant at 1%. A total of 712 observations were used in the estimation of the cut-offs in the hierarchy-ordered probit. Self-reported health status was coded as (1) excellent, (2) very good, (3) good, (4) fair, and (5) poor. Eighteen vignettes were included in the estimation of the hierarchy-ordered probit model.

Table 5 Simulated Self-Reported Health Status by Province

	Own Scale		Gansu Scale		Zhejiang Scale	
	Gansu	Zhejiang	Gansu	Zhejiang	Gansu	Zhejiang
Excellent	0.012	0.058	0.012	0.058	0.009	0.058
Very good	0.03	0.205	0.03	0.084	0.076	0.205
Good	0.402	0.327	0.402	0.562	0.236	0.327
Fair	0.227	0.283	0.227	0.157	0.354	0.283
Poor	0.329	0.127	0.329	0.139	0.324	0.127
Fair or Poor	0.556	0.41	0.556	0.296	0.678	0.410
Difference	0.146		0.260		0.268	

Notes: Eighteen vignettes in the six health domains (pain, sleep, mobility, cognition, breath, and affect) were included in the analysis. Individual weights adjusted for nonresponse were applied.

Table 6 Simulated Self-Reported Health Status: Differences Due to Observables and Unobservables

	Gansu Scale				Zhejiang Scale			
	Gansu Parameters		Zhejiang Parameters		Gansu Parameters		Zhejiang Parameters	
	Gansu (1)	Zhejiang (2)	Gansu (3)	Zhejiang (4)	Gansu (1)	Zhejiang (2)	Gansu (3)	Zhejiang (4)
Excellent	0.012	0.007	0.079	0.058	0.009	0.007	0.068	0.058
Very good	0.03	0.02	0.093	0.084	0.076	0.06	0.219	0.205
Good	0.402	0.379	0.545	0.562	0.236	0.213	0.326	0.327
Fair	0.227	0.225	0.154	0.157	0.354	0.368	0.267	0.283
Poor	0.329	0.369	0.127	0.139	0.324	0.352	0.12	0.127
Fair or poor	0.556	0.594	0.281	0.296	0.678	0.720	0.387	0.410
Total Difference [(1)-(4)]	0.260				0.268			
Difference due to observables [(1)-(2) or (3)-(4)]	-0.038		-0.015		-0.042		-0.023	
Difference due to unobservables [(2)-(4) or (1)-(3)]	0.298		0.275		0.310		0.291	

Notes: Eighteen vignettes in the six health domains (pain, sleep, mobility, cognition, breath, and affect) are included in the analysis. Individual weights adjusted for nonresponse are applied.

Table 7 Proportion of Cases for Which a Vignette Ranking (Row) Is Less Than Another (Column)

	Pain				Sleep		
	<1	<2	<3		<1	<2	<3
1	---	0.34	0.76	1	---	0.58	0.65
2	0.21	---	0.74	2	0.09	---	0.47
3	0.02	0.02	---	3	0.08	0.16	---
	Mobility				Cognition		
	<1	<2	<3		<1	<2	<3
1	---	0.50	0.78	1	---	0.52	0.65
2	0.13	---	0.64	2	0.10	---	0.49
3	0.08	0.08	---	3	0.05	0.14	---
	Breath				Affect		
	<1	<2	<3		<1	<2	<3
1	---	0.64	0.76	1	---	0.56	0.81
2	0.06	---	0.47	2	0.09	---	0.75
3	0.04	0.07	---	3	0.04	0.03	---

Note: Individual weights adjusted for nonresponse were applied.

Table 8 Robustness Check with Different Sets of Vignettes

	Gansu Scale				Zhejiang Scale			
	Gansu Parameters		Zhejiang Parameters		Gansu Parameters		Zhejiang Parameters	
	Gansu (1)	Zhejiang (2)	Gansu (3)	Zhejiang (4)	Gansu (1)	Zhejiang (2)	Gansu (3)	Zhejiang (4)
Panel A: Simulations with six vignettes								
Fair or Poor	0.642	0.638	0.385	0.388	0.704	0.709	0.444	0.453
Total difference based on different scales					0.189			
Total difference based on a common scale [(1)-(4)]	0.254				0.251			
Difference due to observables [(1)-(2) or (3)-(4]	0.004		-0.003		-0.005		-0.009	
Difference due to unobservables [(2)-(4) or (1)-(3)]	0.250		0.257		0.256		0.260	
Panel B: Simulations with nine vignettes								
Fair or Poor	0.511	0.55	0.352	0.366	0.583	0.629	0.411	0.433
Total difference based on different scales					0.078			
Total difference based on the same scale [(1)-(4)]	0.145				0.150			
Difference due to observables [(1)-(2) or (3)-(4)]	-0.039		-0.014		-0.046		-0.022	
Difference due to unobservables [(2)-(4) or (1)-(3)]	0.184		0.159		0.196		0.172	

Notes: Six vignettes included the vignettes in the breath and affect domains. Nine vignettes included the vignettes in the breath, affect and pain domains. Individual weights adjusted for nonresponse were applied

Appendix

Table 1a Estimations of the Thresholds in the Hierarchy Ordered Probit Model for Gansu

	γ^1		γ^2		γ^3		γ^4	
	Coefficient	s.e.	Coefficient	s.e.	Coefficient	s.e.	Coefficient	s.e.
Female	0.119*	0.070	-0.044	0.071	-0.021	0.061	0.016	0.085
Age	0.004	0.005	-0.008	0.005	0.009**	0.004	-0.002	0.005
Years of schooling	0.011	0.009	0.004	0.009	-0.002	0.007	0.032***	0.012
Rural	-0.135	0.083	0.021	0.084	0.178**	0.070	-0.432***	0.119
Reporting SRH before answering health functioning question	0.093	0.065	-0.043	0.066	-0.086	0.055	-0.155*	0.080
Hypertension	-0.028	0.086	-0.068	0.085	-0.095	0.068	-0.009	0.100
High cholesterol	0.136	0.140	-0.172	0.140	0.121	0.125	-0.159	0.178
Diabetes	-0.191	0.209	0.109	0.207	0.193	0.174	0.319	0.270
Lung disease	-0.053	0.091	0.145	0.094	-0.123*	0.071	0.023	0.104
Liver disease	0.098	0.175	-0.180	0.163	0.178	0.169	0.079	0.216
Stroke	-0.3606*	0.1897	0.0786	0.1911	-0.2556*	0.1255	-0.2659	0.1892
Heart disease	-0.153	0.104	0.059	0.105	-0.034	0.083	0.090	0.113
Kidney disease	-0.136	0.159	-0.028	0.157	-0.023	0.116	-0.196	0.159
Stomach or other digestive disease	0.057	0.073	0.001	0.075	-0.050	0.062	-0.098	0.086
Arthritis or rheumatism	-0.136*	0.071	0.069	0.072	0.008	0.058	-0.071	0.081

Notes:*significant at 10%; ** significant at 5%; *** significant at 1%. A total of 712 observations were used in the estimation of the cut-offs in the hierarchy of the ordered probit. Self-reported health status was coded as (1) excellent, (2) very good, (3) good, (4) fair, and (5) poor. Eighteen vignettes were included in the estimation of the hierarchy of the ordered probit model.

Appendix

Table 1b Estimations of the Thresholds in the Hierarchy Ordered Probit Model for Zhejiang

	γ^1		γ^2		γ^3		γ^4	
	Coefficient	s.e.	Coefficient	s.e.	Coefficient	s.e.	Coefficient	s.e.
Female	0.014	0.062	0.139**	0.064	0.032	0.050	-0.002	0.070
Age	0.007*	0.004	-0.002	0.004	-0.001	0.003	0.000	0.004
Years of schooling	-0.016*	0.009	0.005	0.009	0.011	0.007	0.030***	0.011
Rural	-0.315***	0.078	0.185**	0.078	0.000	0.065	-0.074	0.098
Reporting SRH before answering health functioning question	0.094	0.060	-0.127**	0.062	-0.032	0.048	-0.044	0.068
Hypertension	-0.068	0.069	0.127*	0.073	-0.116**	0.055	0.020	0.076
High cholesterol	-0.034	0.118	0.013	0.122	0.201*	0.103	-0.049	0.136
Diabetes	0.085	0.116	-0.190	0.117	0.166	0.108	-0.018	0.146
Lung disease	-0.061	0.109	0.079	0.114	0.022	0.089	0.137	0.123
Liver disease	-0.036	0.115	0.115	0.120	0.016	0.094	0.015	0.129
Stroke	0.269	0.2424	-0.0094	0.2683	-0.4563**	0.1769	0.4689	0.3856
Heart disease	-0.026	0.106	-0.030	0.111	0.198**	0.098	-0.094	0.126
Kidney disease	-0.064	0.153	-0.087	0.153	0.181	0.135	0.486**	0.227
Stomach or other digestive disease	0.117*	0.068	-0.146**	0.070	-0.020	0.057	-0.056	0.076
Arthritis or rheumatism	-0.005	0.085	-0.108	0.086	-0.101	0.066	0.123	0.092

Notes:*significant at 10%; ** significant at 5%; *** significant at 1%. A total of 712 observations were used in the estimation of the cut-offs in the hierarchy of the ordered probit. Self-reported health status was coded as (1) excellent, (2) very good, (3) good, (4) fair, and (5) poor. Eighteen vignettes were included in the estimation of the hierarchy of the ordered probit model.

References

Adams P, Hurd MD, McFadden D, Merrill A, Ribeiro T. Healthy, wealthy, and wise? Tests for direct causal paths between health and socio-economic status. Journal of Econometrics 2003; 112; 3-56.

Bauer T, Sinning M. An extension of the Blinder-Oaxaca decomposition to nonlinear models. Advances in Statistical Analysis 2008; 92; 197-206.

Benzeval M, Taylor J, Judge K. Evidence on the relationship between low income and poor health: Is the government doing enough? Fiscal Studies 2000; 21; 375-399.

Bockerman P, Ilmakunnas P. Unemployment and self-assessed health: Evidence from panel data. Health Economics 2009; 18; 161-179.

Bowling A, Windsor J. The effects of questions order and response-choice on self-rated health status in the English Longitudinal Study of Aging (ELSA). Journal of Epidemiology and Community Health 2008; 62; 81-85.

Case A, Paxson C. Sex differences in morbidity and mortality. Demography 2005; 42; 189-214.

Chatterji S, Kowal P, Mathers C, Naidoo N, Verdes E, Smith J, Suzman R. The health of aging populations in China and India. Health Affairs 2008, 27 (4), 1052-1063.

Chou W, Wang Z. Regional inequality in China’s health care expenditures. Health Economics 2009; 18; S137-S146.

Clarke PM, Ryan C. Self-reported health: Reliability and consequences for health inequality measurement. Health Economics 2006; 15; 645-652.

Contoyannis P, Jones AM. Socioeconomic status, health and lifestyle. Journal of Health Economics 2004; 23; 965-995.

Crossley TF, Kennedy S. The reliability of self-assessed health status. Journal of Health Economics 2002; 21; 643–658.

Deaton A, Paxson C. Aging and inequality in income and health. American Economic Review 1998; 88; 248-53.

D’uva TB, van Doorslaer E, Linderboom M, O’Donnell O. Does reporting heterogeneity bias the measurement of health disparities? Health Economics 2008; 17; 351-375.

Etilé F, Milcent C. Income-related reporting heterogeneity in self-assessed health: Evidence from France. Health Economics 2006; 15; 965–981.

Ettner SL. New evidence on the relationship between income and health. Journal of Health Economics 1996; 15; 67-86.

Giles J, Mu R. Elder parent health and the migration decision of adult children: Evidence from rural China. Demography 2007; 44-2; 265-288.

Humphries K, van Doorslaer E. Income-related health inequality in Canada. Social Science & Medicine 2000; 50; 663-671.

Hurd MD, McGarry K. Evaluation of the subjective probabilities of survival in the health and retirement study. The Journal of Human Resources 1995; 30; S268–S292.

Idler EL. Discussion: Gender differences in self-rated health, in mortality, and in the relationship between the two. The Gerontologist 2003; 43; 372-75.

Idler EL, Benyamini Y. Self-rated health and mortality: A review of twenty-seven community studies. Journal of Health and Social Behavior 1997; 38; 21-37.

Idler EL, Kasl SV. Self-ratings of health: Do they also predict change in functional ability? Journal of Gerontology 2005; 50B; S344–S353.

Kapteyn A, Smith J, van Soest A. Vignettes and self-reports of work disability in the United States and the Netherlands. American Economic Review 2007; 97-1; 461-473.

Kerkhofs M, Lindeboom M. Subjective health measures and state dependent reporting errors. Health Economics 1995; 4; 221–235.

King G, Murray C, Salomon J, Tandon A. Enhancing the validity and cross cultural comparability of measurement in survey research. American Political Science Review 2004; 98; 567-583

Lauridsen J, Christiansen T, Häkkinen U. Measuring inequality in self-reported health discussion of a recently suggested approach using Finnish data. Health Economics 2004; 13; 725–732.

Lee S, Grant D. The effect of question order on self-rated general health status in a multilingual survey context. American Journal of Epidemiology 2009; 169; 1525-1530.

Lindeboom M, van Doorslaer E. Cut-point shift and index shift in self-reported health. Journal of Health Economics 2004; 23; 1083-1099.

Lokshin M, Ravallion M. Testing for an economic gradient in health status using subjective data. Health Economics 2008; 17; 1237-1259.

Molarius A, Janson S. Self-rated health, chronic diseases, and symptoms among middle-aged and elderly men and women. Journal of Clinical Epidemiology 2002; 55; 364-70.

Murray CJL, Tandon A, Salomon J, Mathers CD. Enhancing cross-population comparability of survey results. 2001, GPE Discussion Paper No. 35, WHO/EIP, Geneva.

Patrick DL, Erickson P. Health status and health policy. Oxford University Press: New York; 1993.

Salomon J, Tandon A, Murray CJL. Comparability of self-rated health: Cross sectional multi-country survey using anchoring vignettes. British Medical Journal 2004; 32; 258–263.

Smith, J. Healthy bodies and thick wallets: The dual relation between health and economic status. Journal of Economic Perspectives 1999; 13; 192-196.

Strauss J, Thomas D. Health, nutrition and economic development. Journal of Economic Literature 1998; 36; 766-817.

Tandon A, Murray L, Salomon JA, King G. Statistical models for enhancing cross-population comparability. In: Murray L, Evans DB (Eds), Health systems performance assessment: Debates, methods, and empiricism. World Health Organization: Geneva; 2003.

van Doorslaer E, Wagstaff A, Bleichrodt H, Calonge S, Gerdtham U, Gerfin M, Geurts J, Gross L, Häkkinen U, Leu R, O'Donnell O, Propper C, Puffer F, Rodrìguez M, Sundberg G, Winkelhake O. Income-related inequalities in health: Some international comparisons. Journal of Health Economics 1997; 16, 93-112.

van Doorslaer E, Jones A. Inequalities in self-reported health: Validation of a new approach to measurement. Journal of Health Economics 2003; 22; 61–78.

van Doorslaer E, Masseria C, Koolman X. Inequalities in access to medical care by income in developed countries CAMJ 2006;174(2): 177-183.

van Soest A, Delaney L, Harmon C, Kapteyn A, Smith J. Validating the use of vignettes for subjective threshold scales. Working Paper No. 501 from Rand 2007.

Verbrugge LM. The twain meet: Empirical explanations of sex differences in health and mortality. Journal of Health and Social Behavior 1989; 30; 282-304.

Wagstaff A, van Doorslaer E. Measuring and testing for inequity in the delivery of health care. Journal of Human Resources 2000, 35(4), 716-733.

Wagstaff A, van Doorslaer E, Watanabe E. On decomposing the causes of health sector inequalities, with an application to malnutrition inequalities in Vietnam. Journal of Econometrics 2003; 112(1); 219–27.

Wallace R, Herzog AR. Special issue on the health and retirement study: Data quality and early results. The Journal of Human Resources 1995; 30; S84-S107.

Ware J, Davies-Avery A, Brook, R. Conceptualization and measurement of health status for adults in the health insurance study, vol.VI. Analysis of relationships among health status measures. Rand: Santa Monica; 1980.

World Health Organization. A health situation assessment of the People's Republic of China. 2005a.

World Health Organization. China: Health, poverty, and economic development. 2005b.

Yip W, Mahal W. The health care system of China and India: Performance and future challenges. Health Affairs 2008; 27; 921-932.

Zhang X, Kanbur R. Spatial inequality in education and health care in China. China Economic Review 2005; 16; 189–204.