CHAPTER -6

REPORTING BIAS IN SUBJECTIVE HEALTH MEASURES

Subjective or self reported measures of health are the most widely used measures of health. These subjective measures can be (i) binary measures that record presence or absence of illness (like the information collected by NSS on self reported ailment in the two weeks prior to the survey, or (ii) multiple category global measures of health (like information on the question: how do you rate your current overall health – poor, fair, good or very good).

It is much simpler and more economical to collect information on health by way of self reports from the respondents themselves, in comparison to obtaining such information through clinical investigations. Self reported measures are also considered to be more robust measures of health. Many studies have found the global self rated health indicator to be a good predictor of subsequent mortality or medical care utilization. Such measures also give insights into matters of human concern such as pain or depression that can never be inferred solely from physical measurements or laboratory tests.

But on many occasions it has been found that these self reported categorical data on health throw up information that is counter intuitive.

Some such examples as cited in Murray et al (2000) are given below:

In Australian national health surveys comparing the self-reported health status of Aboriginals with that of the general population, only around 12% of the Aboriginal population characterized their own health status as fair or poor, while more than 20% of the general population rated their health in these low categories. By any other major indicator of mortality and morbidity, the Aboriginal population fares much worse than the general population, which suggests that there may be important differences in the interpretation of categorical responses in the different sub-populations due to shifts in response category cut-points (Mathers and Douglas 1998).
• Residents of the state of Kerala in India—which has the lowest rates of infant and child mortality and the highest rates of literacy in India—consistently report the highest incidences of morbidity in the country (Murray 1996).

• A series of studies from the Living Standards Measurement Surveys has examined the gradient of reported illness as a function of income and found that individuals in higher income quantiles consistently report more illness than those with lower income levels (Murray and Chen 1992).

• A recent study presenting self-reported data from 12 countries in the European Union showed implausibly wide cross-country disparities in the proportion of the population reporting bad and very bad general health, ranging from a high of 19 percent in Portugal to a low of 5 percent in Ireland (Eurostat 1997). Such wide variation in levels of health within the European Union is unlikely, given other major health indicators, and cannot be explained solely by linguistic differences or measurement error.

When are people feeling sick enough to label themselves as sick persons—or healthy enough to say that they have excellent health? The above mentioned examples serve to highlight the fact that there are differences in the responses of the people to the questions that seek self reported categorical information on health. These differences may depend on cultural influences, education, personal experience of illness, age, sex, race and socioeconomic status etc.

This means that there exists reporting heterogeneity across individuals leading to cut point shifts in the response categories and these shifts occur not only in multiple category questions but even in binary questions about clearly defined physical phenomena. For example, a study from Ghana (Belcher et al. 1976) showed that missing body parts very rarely were self-reported.

Cut point shifts essentially mean that sub-groups of a population use systematically different threshold levels for the categorical health question, despite having the same level of ‘true’ health leading to a problem of comparability of self reported health data across groups or sub groups of population.
Figure 1 illustrates the primary challenge of using self-reported levels on a health status domain. For each domain, there is some true or latent scale for that domain that is, by definition, unobserved (Murray et al. 2001). So, for instance, imagine that there is a latent mobility scale, depicted in the first column of Figure 1.

Now imagine a self-reported survey question that asks respondents how much difficulty they have walking up stairs and offers five response categories: “no difficulty”, “mild difficulty”, “moderate difficulty”, “severe difficulty”, and “extreme difficulty/cannot do”. The second column in the figure shows the response category cut-points for population A. These are levels of mobility at which an individual will shift from one response category to another. The lowest cut-point in the figure shows the transition from answering, “extreme/cannot do” to “severe difficulty”. In population B, the response category cut-points are shifted relative to those in population A so that a higher level of mobility is associated with each of the response categories. Population C shows a third example with even more shift in the cut-points. The implication is dramatic. A response of “mild difficulty” walking up stairs, for instance, maps to a different level of mobility in populations A, B and C. A response of “mild difficulty” walking up stairs, for instance, maps to a different level of mobility in populations A, B and C.
This reporting heterogeneity or cut point shift may invalidate group comparisons of health status and may have important implications for the measurement of health inequalities (Murray et al 2000 and Lindeboom & Doorslaer 2003), intertemporal changes in health status and for constructing weights for QALY indices (Groot 2001, Cutler & Richardson 1994).


**Reasons for reporting bias**

State dependent reporting bias or scale of reference bias is said to occur when the respondent has been suffering from a chronic ailment for a long time and has adapted himself to discomfort and illness so that he reports himself to be in a better state of health than is expected. It could also arise if questions about health are answered relative to a certain reference group. These answers do not have a natural reference point but rather, the reference point is determined differently by different individuals depending upon his specific situation and characteristics (for eg with reference to people with more health problems, people with low income, people with no jobs etc.)

This heterogeneity in reference point of individuals has also been discussed in the context of what has been called the cultural inflation of morbidity. It has been found that while mortality rates have been decreasing, the accompanying trends in morbidity show an upward movement. Murray & Chen (1992) note that when information was recorded on ailments during the two weeks before the survey for the 28th round of NSS, reported morbidity rate for Kerala state was the highest. It was nearly three times the All India average inspite of the fact that Kerala has one of the lowest levels of mortality among Indian states. They further go on to show that although the rates are not directly comparable, but the data shows much higher rates of self reported morbidity in USA compared to Kerala. The reason proffered for this is that self perceived morbidity is a
function of both the burden of pathology and social and cultural context of the individual. Consequently, the pattern of self perceived morbidity analysed by socio economic factors may be divergent from those expected for observed morbidity.

According to Johansson (1992), the feeling of being ill is with reference to some idealized state of health. An individual’s health ideal is shaped by individual and community factors, person’s education, his contact with health services and his health ideal. All these interact with knowledge to create a perception of health. The collective community experiences may also influence health ideals. If everyone in the community has diarrhea most of the time, the health ideal for that community may not recognize diarrhea abnormal. So greater self reported morbidity is due to greater illness perception, greater awareness among people, greater contact with health services and diagnosis at early stage of the disease.

The shift in response categories has also been called “reporting error” in the literature on account of endogeneity between health and work (Keerkof & Lindeboom 2002). There may also be conscious misreporting of disability or morbidity on account of existence of some economic motives to achieve other goals such as improved sickness compensation, disability insurance etc.

**Strategies to correct for reporting bias or enhancing cross population comparability of subjective health status**

In the literature several statistical models have been formulated to pick out the extent of bias or cut point shift in these subjective categorical health variables.

Murray et al (2001) have suggested two basic categories of methods to establish the unbiased cut-points.

1. The first category of methods is to establish a scale that is strictly comparable across individuals and populations. Measurements on the comparable scale can then be used to establish the response category cut-points for each survey item.

2. The second approach is to elicit categorical responses from different groups for a fixed level on the latent scale. If the level is fixed, variation in the responses
provides information on the differences in cut-points across individuals and populations.

3. Using a comparable scale to establish response category cut-points. There are at least two main strategies for establishing a comparable scale of measurement:
   (a) the use of multiple items (i.e. questions) for measuring a particular domain; and
   (b) the incorporation of exogenous information such as a measured performance test.

a. Use of multiple items (i.e. questions) for measuring a particular domain

This body of work is often associated with the term item response theory (IRT). IRT has been used widely, for example, to establish the difficulty of different standardized scholastic test items. There are many different statistical models used in the application of IRT, and the field is rapidly expanding. The basic building block for many of these models is the one-parameter Rasch model, which is a variant of the conditional logit with a fixed effect. With more than two response categories, the Rasch model generalizes to the partial credit model (PCM) (Masters 1982). Estimation of the partial credit model is based on specification of the probability of responding in a particular category rather than in the previous category, which is modeled as an increasing function of a person’s “ability” (level on a particular domain) and a decreasing function of the item response “difficulty”. The “difficulty” parameters specify the level on the latent variable at which an individual is more likely to respond in a particular category than in the previous category.

However, it is obvious that no statistical procedure can deal with this problem in its extreme form without the addition of exogenous information. Imagine a domain such as mobility. In population A the distribution of mobility on the latent scale is normal with mean 5 and standard deviation 2 (in units with interval scale properties on some arbitrary scale). In population B, the distribution of mobility is normal with mean 8 and standard deviation of 2. In population B, all the response category cut-points on all items about mobility are exactly 3 units higher than in population A. The net result is that the distribution of responses on all items in the survey in the two populations will be
identical. In other words, population B has much higher mobility than A but the survey gives identical responses. No statistical method can identify this difference because the data are completely identical. Because we have strong prior beliefs that cut-points on items are likely to shift systematically for a domain, we suspect that the potential to establish cross-population comparability using only the underlying factor in the response data without additional information is limited. For this reason, exogenous information is needed to aid in establishing cross-population comparability.

b. The incorporation of exogenous information such as a measured performance test

Measured tests

One type of exogenous information that could be used to establish a comparable scale is a measured calibration test for a particular domain. A calibration test can establish a comparable scale if: (a) it adequately captures a domain, and (b) it can be implemented in different settings without systematic bias in the results. Such calibration tests are not feasible for a number of domains of health such as pain or affect. For some domains such as vision, hearing, cognition, mobility and others, calibration tests are feasible.

If a reliable and valid measured test exists for a domain, variation in response category cut-points for the self-reported items on these domains may be estimated using the hierarchical ordered probit (HOPIT) model (Tandon et al. 2001). The HOPIT model is a variant of the standard ordered probit model, which assumes that there is an unobserved latent variable that is normally distributed. Observed categorical responses depend on the categorical cut-points on this latent scale. The key difference between HOPIT and the standard ordered probit model is that these cutpoints are allowed to vary as a function of covariates in the HOPIT model. In essence, this means that the mapping from the underlying latent variable to the categorical responses depends on individual characteristics. As illustrated in Tandon et al. (2001), the HOPIT model is able to recover the cut-points in simulated data sets quite well using exogenous information (such as measured test results) to fix the scale of the latent variable, and consistently outperforms models that do not incorporate this exogenous information.
Hierarchical ordered probit model

Let us assume that true health is measured by a latent variable which is a function of covariates such as age, sex, and education:

\[ y^* = f(\text{age, sex, education}) \]

Since \( y^* \) is unobserved, there is an observation mechanism such that:

\[ y_i = \begin{cases} 1 & \text{if } -\infty < y^* \leq \tau_i^1 \\ 2 & \text{if } \tau_i^1 < y^* \leq \tau_i^2 \\ 3 & \text{if } \tau_i^2 < y^* \leq \tau_i^3 \\ 4 & \text{if } \tau_i^3 < y^* \leq \tau_i^4 \\ 5 & \text{if } \tau_i^4 < y^* \leq +\infty \end{cases} \]

The distinctive feature of HOPIT is that the \( \tau \)'s are subscripted so that they are different for every individual \( i \). Basically, the model now allows these response category cut-points themselves to be functions of covariates:

\[ \tau_i^k = f(\text{covariates}) \]

for \( k = 1,2,3,4 \). That is, all four cut-points can be functions of covariates in the HOPIT model. This allows the model to estimate whether or not there are response category cut-point shifts in the population (or among populations) and what is the extent of the shifts.

Identification problem

There is a problem with the HOPIT model discussed above. The problem is that the model in of itself cannot distinguish between the effects of covariates on: (a) the level of the latent variable, and (b) the response category cut-points. This is the "identification problem". This can also be illustrated using a simple example. Suppose we have two countries A and B. Country B has higher mobility than country A (i.e., its \( y^* \) is higher). Also, suppose country B has higher expectations for their health than country A (i.e., its \( \tau \)'s are higher). The model will pick up a net effect of these two tendencies. If the effect of response category cut-points in country B is dwarfed by the effect of higher mobility in
country B: net effect will be to show that the estimated latent variable of mobility in country B is lower, which would be an erroneous conclusion.

In order to deal with this identification problem, exogenous information needs to be introduced in order allow us to tease out the effect of the covariates on the latent variable from the effect on the cut-points.

**Using fixed ability comparisons to assess variation in cut-points**

For some domains of health such as pain, reliable and valid measured tests may not exist, may not be affordable, or may be unethical, even for a subsample. An alternative strategy for establishing cross-population comparability is to fix the level of health on a domain and assess variation in the response categories across individuals, groups and populations. In other words, if the level of mobility is fixed but one group says that this level maps to a response category of "no difficulty" and another says it maps to the category "some difficulty", this information can be used to assess the response category cut-points. Two strategies are available for fixed ability comparisons: (a) vignettes and (b) comparable homogeneous groups.

**Establishing cross-population comparability using vignettes**

The primary strategy for using fixed ability comparisons to establish comparable scales is the use of vignettes, as described in detail in Salomon et al. (2001). A vignette is a description of a concrete level of ability on a given domain that individuals are asked to evaluate using the same question and response scale as the self-report question on that domain. For example, one self-report question on health in the WHO survey instrument is:

*Overall in the last 30 days how much difficulty did you have with self care?* (1 = None, 2 = Mild, 3 = Moderate, 4 = Severe, 5 = Extreme/cannot do).

To assess the response category cut-points, each respondent is also asked to assess levels of self-care for hypothetical cases described with vignettes, for example: [John] cannot wash, groom or dress himself without personal help. He has no problems with feeding.
How would you rate his difficulty with self-care?

None 1  
Mild 2  
Moderate 3  
Severe 4  
Extreme/cannot do 5  

The vignette fixes a given level of self-care so that variation in the response categories is attributable to variation in the response category cut-points. When individuals are asked to evaluate a series of vignettes of varying severity, the cut-points can be evaluated using the HOPIT model (Tandon et al. 2001). The vignette version of the HOPIT model is constructed such that the dependent variable is the categorical response for a given vignette, and the independent variables are simply indicator variables for each vignette.

Using comparable homogeneous groups to establish cross-population Comparability

Another way to evaluate a fixed ability and thus variation in cut-points is to identify comparable homogeneous groups in different populations and compare their responses to an item. Recent acute changes in health status from injuries such as fractures might be used to identify reasonably comparable groups. Alternatively some lifestyle or occupational characteristic might be used such as elite athletes.

Another model for Health Reporting In The Presence Of Purposeful Bias In Health Reporting

Kerkhof & Lindeboom (2002) have estimated a model for health reporting in the situation that people may have an economic incentive for reporting themselves to be in a different state of health than the one in which they actually are, giving rise to problems in health status measurement. They have estimated their model in the Dutch context where people stand to gain from disability insurance benefits by citing poor health as a reason for their retirement from work. The model is also based upon the premise that there might be endogenous relationship between health and work both because of direct effects of health on work & vice versa and because there are unobservables that may relate observed health and work outcomes.
The key idea of their approach to analyse reporting errors is to compare the subjective health measures to an objective measure of health. Let the reported subjective measure of health be denoted by $H^s$, the true latent measure of health be denoted by $H^*$ and let an objective measure of health be denoted by $H^0$.

The basic argument in the literature considering the peculiar relationship between subjective health measures and retirement is that commonly used responses to health questions are subject to roughly two forms of possible biases. First, true health may be related to labour market status $S$ (S=Employed, Unemployed, Disabled or Early Retired). This can be a direct causal relationship, or health and labor market status could be indirectly related through unobservables. One way in which this type of endogeneity emerges if an individual's health and career are considered to result from simultaneous investment decisions regarding education, work and health. This kind of dependence of health on labour market status is referred to as type I endogeneity. Secondly, state dependent reporting behaviour could relate the observed subjective measures to the labour market status $S$. This kind of endogeneity will be denoted as type II endogeneity. The following assumptions enables one to deal with type II endogeneity, without needing to consider type I endogeneity directly. It will, however, turn out that classical, type I, endogeneity problems returns in the empirical implementation of the health reporting model.

The model for reporting behaviour of general health is as follows.

Kerkhof & Lindeboom assume that $S$ affects both $H^*$ and $H^0$ in the same manner and hence the distribution of of $H^*$ is identical for all respondents irrespective of their value of $S$. In such a scenario any difference in the subjective health measure $H^s$ that exists across labor market states ($S$), after controlling for objective health measure $H^0$ and other control variables (represented by vector $X$) must come from reporting errors. So this helps us to assess the importance of reporting errors in health responses using the following equation:

$$H^s = f(H^0, S, X, \varepsilon; \omega),$$

where, $\varepsilon$ is an unobservable and $\omega$ is a parameter vector.
S captures the reporting behaviour and X will also capture any reporting behavior that may exist across socio economic groups. State dependent reporting error can be identified by normalizing with respect to a chosen reference group.

Further, Kehrkof & Lindeboom also use the above equation to generate a cleansed health measure. The estimation of the above model requires a simultaneous estimation of the decision to work and the decision to report a particular state of health. Our interest lies in the estimation of the health reporting equation which has been estimated using the OPROBIT regression and the state dependent cut points are modeled by means of separate dummies for each level of cut point and labor market status.

**Cut Point Shift versus Index Shift**

Lindeboom and Doornaert (2003) have also developed a framework for individual reporting behaviour that enables us to formally test whether variations in responses to health questions reflect true health differences or reporting behaviour. Their health-reporting model has the distinctive feature that it allows to distinguish between different types of reporting heterogeneity: cut-point shift and index shift. Index shift refers to the situation where the individual response behaviour does not affect the shape of the SAH distribution, but rather its location. Cut-point shift refers to the situation where the shape of the distribution of the SAH changes. Stated differently, in the case of an index shift the location of the reporting thresholds changes, but their relative position remains unaltered. In the case of a cut-point shift the relative positions of the reporting thresholds change.

**A model of reporting behaviour**

The reported subjective health measure is denoted by \( H^S \) and is the respondent's answer to a question like e.g. "How good is your health in general?" with replies ranging from excellent, very good, good, fair, to poor. It is assumed that these responses are generated by a corresponding latent true health variable \( H^* \). \( H^* \) could refer to a set of latent health measures or one single measure. The reporting model relates the unobserved latent true health to the individual's responses to health questions. Of relevance for the issue of cross-population comparability is that the relationship between \( H^* \) and \( H^S \) may not be constant across populations and may even vary within populations. For instance, an individual
with a university degree and a true health level $H^*$ may state that his health is 'fair', whereas an otherwise identical but lower educated respondent may report to be in excellent health. Or stated more generally, the response labels may have a different meaning in different subgroups of the population.

The introduction of a true health concept $H^*$ is essential for the issue of cross-population comparability. Any unconditional comparison of $H^*$ (i.e. not controlling for $H^*$) across different groups of the population may in this case reflect both differences in health conditions as well as differences in reporting behaviour.

In practical situations one does not observe $H^*$, but instead one has to rely on a range of more objective measures derived from tests on the presence of various health problems, diseases and health-related impediments in performing a range of daily activities which may be called $H^0$. These could be used to proxy the (different domains of) true unobserved health ($H^*$).

The above results in the following relations of the health reporting model:

$$H^0 = f_1(H^*, X_1, \epsilon_1; \beta_1) \quad (1a)$$
$$H^* = f_2(H^0, X_2, \epsilon_2; \beta_2) \quad (1b)$$

The variables $\epsilon_1$ and $\epsilon_2$ are random variables, $f_1(.)$ describes the reporting behaviour and $f_2(.)$ the relationship between the true health and its determinants. Consequently, the effect of $X_1$ (as embodied in the $\beta_1$'s) could be viewed as pure reporting behaviour, whereas $X_2$ corrects for the dissimilarity between $H^0$ and $H^*$. In the (ideal) case where $H^*$ is adequately captured by $H^0$ $f_2$ will be an identity. $H^*$ is by definition unobserved. The difficulty lies in the assessment of the relative importance of $X_1$ (reporting behaviour) versus $X_2$ (true health effects or heterogeneity in health) in the determination of $H^0$. It will depend crucially on how the empirical model is implemented. The essence of differential reporting, cut point shift or state-dependent reporting behaviour is that $X_1$ will not only affect the mean of the index function, but will, in addition, have a direct influence on the cut-points corresponding to the different response categories. $X_2$ is included to measure the dissimilarities between $H^0$ and $H^*$ and therefore its role is expected to be different. More specifically:
The health response is now defined as:

\[ H^* = f(H^P; \alpha) + X_2' \beta_2 + \varepsilon_2 \]  

(2)

The health response is now defined as:

\[ H^i = i \iff c_{i-1} < H^* \leq c_i \quad \text{for } i = 1, \ldots, n \]  

(3a)

Where \( n \) is the number of response categories. The cut points \( c_i \) are allowed to vary with different values of \( X_1 \):

\[ c_i = g_i(X_i; \beta_i) \quad \text{for } i = 1, \ldots, n-1, \quad c_0 = -\infty, \quad c_n = \infty \]  

(3b)

In this model there may be differential response (\( H^* \)) for people with identical levels of true health, \( g_i(.) \) is left unspecified, but in its most flexible form each cut point is affected by exogenous variables in a different way. Lindeboom & Doorslaer combine 2 & 3 to get

\[ H^i = i \iff g_i(X_i; \beta_{1i}) \cdot X_2' \beta_2 < f(H^P) + \varepsilon_2 \quad \text{for } 2 \leq g_i(X_i; \beta_{1i}) \cdot X_2' \beta_2 \]  

(4a)

Now the role of \( X_1 \) and \( X_2 \) and the identification of their effects becomes clearer. The variables \( X_2 \) are part of the index function (cf. eq. 2) and determine the true level of health (\( H^* \)), whereas \( X_1 \) variables affect the thresholds and are allowed to have differential effects across thresholds. A variable that belongs in \( X_2 \), but that was mistakenly included in \( X_1 \) will shift all threshold variables in the same way. This can easily be tested by imposing the restriction that the \( \beta_i \) are equal for all \( i \). If this restriction is rejected then the \( X_1 \) variable is (at least to some extent) responsible for cut-point shift (or state dependent reporting errors). This simple test will be important because our primary aim is to determine whether variations in responses reflect true health differences or reporting behaviour.

The empirical strategy is to estimate separate ordered response models for groups in the data sample stratified according to age, sex, education, income and language. More specifically:

\[ H^i_k = l \iff \delta^i_{k,l} < f(H^P; \alpha, k) + \varepsilon_2 \leq \delta^i_{k,l} \]  

(4b)

This specification is more flexible than the usual hierarchical ordered response models like e.g. the HOPIT model. In these models one usually assumes a common vector \( \alpha \) and imposes a functional form on the function \( g_i \). For instance, Kerkhofs & Lindeboom (1995), Groot (2001) and Tandon et al (2002) take \( g_i(X_i; \beta_{1i}) = X_1' \beta_{1i} \). In this
specification of (4'), each subgroup $k$ is allowed to differ both with respect to cut-points ($\delta_i$, 's) and with respect to index effects ($\alpha_i$).

A test for differential response behaviour across different subgroups, i.e. different $\delta_i$'s and/or different $\alpha_i$'s, can be based on straightforward likelihood ratio tests. The tests compare the sum of the likelihood values for two subgroups $k$ and $k'$ with the likelihood values obtained using the total group (i.e. with equal $\delta_i$'s and $\alpha_i$'s imposed). In cases where equality of response behaviour for two groups is rejected, we can additionally test whether this is due to index shifts (different $\alpha_i$'s) or to cut-point shifts (different $\delta_i$'s). The way to implement this additional test is to estimate (4') subject to either the restriction that

$$\delta^k_i = \delta^{k'}_i, \quad \text{or} \quad \alpha^k_i = \alpha^{k'}_i,$$

for all $i$, and to compare the likelihood value of the restricted model ($\mathcal{L}_R$) with the sum of the likelihood values of the unrestricted models ($\mathcal{L}_U$). A test is based on $-2(\mathcal{L}_R - \mathcal{L}_U)$ which is $\chi^2$ distributed with degrees of freedom equal to the number of restricted parameters.

The essence of this model is captured in equation 1 and any estimation procedure HOPIT or interval regression or any other method that may be used for estimating this model.

For carrying out the above estimations it is essential to have data on objective measure of health and for the estimation models suggested by Kerkhof & Lindeboom and Lindeboom & Doorslaer we need to draw information from panel data. It is obvious that estimation of reporting bias or cut point shift is not possible with the NSS data that we have.

**Pointers to suspected measurement error or income related reporting bias in morbidity in the NSS data**

It has been observed that the incidence of morbidity is found to be positively correlated with the consumption expenditure fractiles. Higher levels of consumption expenditure deciles show higher percentage of people who are reporting ill. This is puzzling because it appears as if incidence of illness increases with the increase in the income (proxied by consumption expenditure) of an individual. Statements have been made about the
likelihood of a income related reporting bias in the morbidity data (Sarvekshan 1997, Gumber 1997). But no attempt has been made to test for the bias or quantification of this bias.

As is obvious from the discussion on strategies to check for shifts in cut points, we either need some exogenous information such as measured tests or vignettes to run the HOPIT model or we need panel data to estimate the model suggested by Kerkhof & Lindeboom (2002). We obviously do not have any such data so estimation of the bias is not possible.

**HOPIT estimation**

We exploited the fact that level 9 of the NSS data contains information on several objective measures of health ($H^0$), and two other self reported measures of health for the aged people. With one we constructed the SRII and the other measure is the conventional measure on global self reported health (SAH), where the respondents answered the question that how do they rate their health on a four point scale.

Now while even for the aged, SRII shows a positive correlation with income, while SAH shows a negative correlation between poor health and income. So apparently one could say that while there might be some element of income related reporting bias in SRII, any such bias does not seem to exist in the case of SAH.

We carried out an exercise, wherein we ran two separate HOPIT models, one for SRII and SAH where the cut points of the SRII and SAH were made a function of the per capita consumption expenditure for the sample on aged people in the data.

The objective measures of health used in both the regressions were the same. They description of the independent variables is as follows:
This is a dummy variable to show if the respondent is rich and the reference dummy is an
individual categorized as poor

male This captures the gender of the respondent, male is coded as 1 and the reference dummy is female

urban This captures the place of residence of the respondent, urban is coded as 1 and the reference dummy is
rural

married This captures the marital status of the respondent, married is coded as 1 and the reference dummy is not
married married which includes widowed, separated or not married

The following are taken as objective health indicators:

diabetes if a person suffers from diabetes, reference dummy is person not suffering from diabetes

urine if a person suffers from urinary disease, reference dummy is person not suffering from urinary disease
cough if a person suffers from cough, reference dummy is person not suffering from cough

spe_dis if a person suffers from speech disability, reference dummy is person not suffering from speech disability

her_dis if a person suffers from hearing disability, reference dummy is person not suffering from hearing disability

vis_dis if a person suffers from visual disability, reference dummy is person not suffering from visual disability
t1, t2, t3 represent the cut points for the respective categorical health variable

The results (Table 6.1) showed consistently lower thresholds (values of t1, t2, t3) of SRII categories compared to the categories for SAH. It could be said that this does indicate towards the existence of some element of reporting bias in the data. Of course there are two caveats to the above indications:

1. We have not tested for the significance of the difference between the respective cut points in the categories for SRII and SAH.

2. The HOPIT model as estimated above suffers from an identification problem on account of any exogenous information on health. And that identification problem prevents us from knowing whether the lower cut points are due difference of perception of difference in actual morbidity.
### TABLE 6.1

**HOPIT REGRESSION**

| hopit on SRII | Coef. | Std. Err. | z     | P>|z|  |
|---------------|-------|-----------|-------|-------|
| **eq1**       |       |           |       |       |
| male          | -0.002919 | 0.022002  | -0.13 | 0.894 |
| urban         | -0.125431 | 0.021294  | -5.89 | 0     |
| married       | -0.018268 | 0.02252   | -0.81 | 0.417 |
| diabetes      | 0.537433  | 0.039172  | 13.72 | 0     |
| urine         | 0.338976  | 0.048948  | 6.93  | 0     |
| cough         | 0.420226  | 0.024167  | 17.39 | 0     |
| spe_dis       | 0.286718  | 0.053212  | 5.39  | 0     |
| her_dis       | 0.050532  | 0.030242  | 1.67  | 0.095 |
| vis_dis       | 0.143636  | 0.023297  | 6.17  | 0     |
| **t1**        |       |           |       |       |
| rich          | -0.169125 | 0.022125  | -7.64 | 0     |
| cons          | 1.165137  | 0.024066  | 48.41 | 0     |
| **t2**        |       |           |       |       |
| rich          | -0.095645 | 0.02583   | -3.7  | 0     |
| cons          | 1.528476  | 0.026223  | 58.29 | 0     |
| **t3**        |       |           |       |       |
| rich          | -0.066982 | 0.029745  | -2.25 | 0.024 |
| cons          | 1.78168   | 0.028583  | 62.33 | 0     |

| hopit on SAH | Coef. | Std. Err. | z     | P>|z|  |
|--------------|-------|-----------|-------|-------|
| **eq1**      |       |           |       |       |
| male         | -0.119739 | 0.018028  | -6.64 | 0     |
| urban        | -0.184292 | 0.017613  | -10.46| 0     |
| married      | -0.151385 | 0.018584  | -8.15 | 0     |
| diabetes     | 0.407515  | 0.038921  | 10.47 | 0     |
| urine        | 0.419394  | 0.047306  | 8.87  | 0     |
| cough        | 0.421131  | 0.022072  | 19.08 | 0     |
| spe_dis      | 0.35525   | 0.04976   | 7.14  | 0     |
| her_dis      | 0.330296  | 0.026366  | 12.53 | 0     |
| vis_dis      | 0.386597  | 0.020244  | 19.1  | 0     |
| **t1**       |       |           |       |       |
| rich         | -0.076873 | 0.046478  | -1.65 | 0.098 |
| cons         | -2.229395 | 0.03774   | -59.07| 0     |
| **t2**       |       |           |       |       |
| rich         | 0.059705  | 0.024317  | 2.46  | 0.014 |
| cons         | -1.294274 | 0.023304  | -55.54| 0     |
| **t3**       |       |           |       |       |
| rich         | 0.12049   | 0.022028  | 5.47  | 0     |
| cons         | 0.977002  | 0.021498  | 45.45 | 0     |