CHAPTER 4: METHODOLOGY

To fulfill the objectives of the dissertation and to empirically test the hypotheses proposed in the preceding section it was decided to conduct two studies. The first study used the data collected from the national survey done by the ministry of health known National Family Health Survey – 3. The second study was done with the help of primary data collected via a survey conducted in one urban and one rural area of the state of West Bengal in India. This section describes in details about the data used for both the studies in terms of their source, sampling design, operationalization of the variables as well as the model specifications and data analysis techniques used.

4.1 STUDY 1

4.1.1. Data Source

For the purpose of study 1, the data from the National Family Health Surveys-3 (NFHS-3) has been utilized. It is the latest in the series of NFHS conducted by the Ministry of Health and Family Welfare (MoHFW), Government of India (GOI) in coordination with International Institute for Population Sciences (IIPS), Mumbai, India and Macro International, Maryland, USA. This survey was conducted in the year 2005-06 and it covered all the 29 states of India. This survey used standardized questionnaires and employed a face to face interview technique with adults for collection of data. This survey was aimed to primarily collect data pertaining to the health care, reproductive and nutrition issues of Indian citizens. Additionally it collected demographic and other socially relevant information of respondents for aiding in policy decisions.
4.1.2. Questionnaires

NFHS-3 used three questionnaires that comprise questionnaires for men, women and households separately. These questionnaires are modeled upon the standard DHS (Demographic and Health Surveys) questionnaires used around the world for collecting data on population, nutrition and health by Measure DHS project\(^2\). The final content for NFHS-3 questionnaires was developed with the help of workshops and meetings attended by health experts, health ministry officials and related government agencies. Each of these questionnaires were bilingual in nature (English and main language of the state). These questionnaires were structured and questions were close ended in nature.

\(^{2}\)The MEASURE Demographic and Health Surveys (DHS) Project is responsible for collecting and disseminating accurate, nationally representative data on health and population in developing countries. The project is implemented by ICF International and is funded by the United States Agency for International Development (USAID) with contributions from other donors such as UNICEF, UNFPA, WHO, and UNAIDS (Source: http://en.wikipedia.org/wiki/Measure_dhs; http://www.measuredhs.com/).
4.1.3. Sampling Design

The NFHS-3 survey was cross sectional in nature. It followed a multi stage clustered sampling approach for collection of data across all the states. It followed two separate approaches for collecting data from rural and urban areas respectively. For rural area it followed a two stage sample design approach. Depending on the list of villages obtained from the national census conducted in 2001, using the probability proportional to population size (PPS) technique, villages were shortlisted in the first stage. Then within the selected villages, using a systematic random sampling technique, households were selected for collecting required data in the second stage. For the urban areas a three stage sample design method was followed. The first stage consisted of selection of wards based upon probability proportional to population size. From each of the selected wards, in the second stage one census enumeration block (CEB) was randomly selected. For the third and final stage of the sample design, within the selected CEB, using systematic random sampling technique households were selected for collecting the data.

4.1.4. Survey and Study Sample

NFHS-3 survey collected data in two phases from November, 2005 to August, 2006 across all the 29 states of India. Sample for the survey consisted of 124,385 women aged 15 – 49 and 74,369 men aged 15 – 54 years. Since, the focus of the present study is on identifying the determinants of healthcare services choice made by an individual, the sample for this study has been chosen from NFHS-3 data as follows. First, only those individuals were considered who had utilized any type of healthcare services for curative purpose for their own treatment in the past three months from the date of the interview. This reduced the eligible sample size to 30,603 individuals considering males and females together. A working individual is assumed to be more
likely to make the decision regarding healthcare services utilization for himself or herself as compared to somebody who is not working. Hence, the present study concentrates only on those individuals who were working (in some occupation) the time of the survey. This reduces the sample size to the final figure of 16,728 for further analysis.

Table No.3: Number of eligible sample respondents for Study 1 from overall NFHS – 3 dataset

<table>
<thead>
<tr>
<th>Eligible Respondents</th>
<th>Urban</th>
<th>Rural</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFHS-3 Overall Sample</td>
<td>79016</td>
<td>119738</td>
<td>198754</td>
</tr>
<tr>
<td>NFHS-3 Sample for Study 1</td>
<td>8260</td>
<td>8468</td>
<td>16728</td>
</tr>
</tbody>
</table>

4.1.5. Variables & Measures

This sub section describes the variables that have been used in this study for empirically testing the proposed framework and the related hypotheses mentioned in the earlier chapter. Data for these variables have been taken from the NFHS-3 survey. Data for some of the variables was re-coded as per the requirements of this study.

Dependent Variable

The dependent variable for this study is a categorical variable that reflects the type of healthcare service provider utilized by a working adult individual for himself or herself. The data for this has been directly taken from NFHS-3 questionnaire. In the first step, the respondent was asked whether or not that respondent had visited any medical facility in the last three months preceding the date of survey. Only those who responded positively are considered as eligible respondents. In the next step, they are asked what type of facility they visited most recently. There are 14 different categories of health facilities in the NFHS-3 questionnaire. To facilitate data analysis
that will help answer the research questions, the data on dependent variable has been re-coded under public or private healthcare facility. Only those respondents have been considered for final analyses who have sought medical treatment for themselves during the said time frame. For the utilization of any type of private health care facility, the response was coded as “0” whereas for any type of public healthcare usage it was coded as “1”.

*Independent Variables*

The independent variables for this study consist of age, gender, education, income, religion, caste, marital status, exposure to mass media and accessibility.

Age: The data on age was taken as-is from the NFHS-3 dataset. It was recorded in years and is a continuous variable.

Gender: NFHS-3 had different questionnaires for male and female individuals. For the purpose of this study, the separate datasets were merged. Gender was coded as male or female for each respondent accordingly. It has been considered as a dichotomous categorical variable. Hence, dummy coding was used. Males were considered the reference category and coded as “0”, whereas females were coded as “1”.

Education: NFHS-3 records the education of an individual in both categorical as well as continuous form. For categorical form, education was recorded in four distinct categories in terms of the levels of education completed by an individual. For the continuous form, the number of years of education completed by an individual was recorded. For the sake of easier interpretation of the results, the categorical form of education was considered for analysis. In view of the proposed hypotheses related to education, an individual with no education was considered the reference category and hence coded as “0”. The three other categories of
education levels completed by an individual i.e. primary, secondary and higher than secondary were coded as “1”, “2”, and “3” respectively.

Income: NFHS-3 does not collect data on the income of an individual. As a result standard of living index was used as a proxy for income. Similar proxy was used by Thind (2004) in his study. This standard of living index, also called as SLI is calculated with the help of asset scores. A series of questions are asked to an individual regarding various kinds of assets possessed by individual as well as living conditions of that individual. Using the method of principal component analysis, depending upon the answers provided, asset scores are generated for each individual. Based upon these scores, an individual is categorized as either having a low, medium or high standard of living. For the purpose of this study, SLI was used as a proxy for income of an individual. Individuals coming under the high and medium SLI were re-coded as the reference category i.e. “0”. Individuals belonging to low SLI were coded as “1”.

Caste: Caste represents an intra-religion classification that is widespread in Indian society. It is recorded in background information for individuals in NFHS-3 data. The data on caste was divided into three categories. Individuals belonging to the general caste and other backward caste were clubbed together and treated as the reference category hence coded “0”. Those belonging to scheduled caste or tribe (SC/ST) were clubbed together and coded as “1”. The reason for this coding categorization is based on the fact that the general and OBC caste people in India are socially as well as economically more powerful as compared to the vulnerable SC/ST caste people. Hence the utilization patterns of general and OBC people are quite similar as opposed to SC/ST categories.
Marital Status: Marital status is reported under seven different categories in NFHS-3. Since currently married and never married represent 98.4% of the overall data, all other subcategories of widowed, divorced, separated and deserted were clubbed under the category of unmarried individuals. Marital status was thus converted into a dichotomous variable, wherein never married and others formed the base category and were coded as “0”. Currently married individuals formed the reference category, coded as “1”.

Exposure to mass media: Exposure of an individual to television, radio and newspapers were classified under the mass media exposure variable. NFHS-3 records information on frequency of access to these mass mediums by an individual. Individuals who have daily access to radio, television and/or newspapers were classified as high mass media exposure individuals and were coded as “1”. Individuals whose access to these medium was nil or irregular were treated as base category and coded as “0”.

Accessibility: NFHS -3 does not directly collect information on accessibility to health services. Hence, Anganwadi services coverage of a ward for urban area and that of a village for rural area were treated as the proxy for accessibility variable. As a result, accessibility was converted into a dichotomous variable with covered or not covered under Anganwadi services as the two categories. “Not covered under Anganwadi services” was treated as the base category and coded as “0” whereas “covered under Anganwadi services” was coded as “1”.

---

3Anganwadi is a government sponsored child-care and mother-care center in India. The Anganwadi system is mainly managed by the Anganwadi worker. She is a health worker chosen from the community and given 4 months training in health, nutrition and child-care. She is in charge of an Anganwadi which covers a population of 1000. There are an estimated 1.053 million Anganwadi centers employing 1.8 million mostly-female workers and helpers across the country. They provide outreach services to poor families in need of immunization, healthy food, clean water, clean toilets and a learning environment for infants, toddlers and pre-schoolers. They also provide similar services for expectant and nursing mothers. According to government figures, Anganwadis reach about 58.1 million children and 10.23 million pregnant or lactating women. Anganwadis are India’s primary tool against the scourges of child malnourishment, infant mortality and curbing preventable diseases. (Source: http://www.aanganwadi.org/, 2013)
4.1.6. Model Specification

Since the target variable is dichotomous in nature the following logistic regression model was proposed and tested:

\[ \text{Logit} (p_i) = \beta_0 + \beta_1 \text{(Age)} + \beta_2 \text{(Gender)} + \beta_3 \text{(Education)} + \beta_4 \text{(Mass Media Exposure)} + \beta_5 \text{(Marital Status)} + \beta_6 \text{(Access)} + \beta_7 \text{(Income)} + \beta_8 \text{(Caste)} + r_i \]

Where:

\( p_i \) = Probability of observing the response for an individual i utilizing public healthcare services

\[ \text{Logit} (p_i) = \log \left[ \frac{p_i}{1 - p_i} \right] = \log \text{odds of utilizing the public healthcare services} \]

\( \beta_0, \beta_1, \ldots, \beta_8 \) = intercepts and slopes coefficients and \( r_i \) = residual for individual i; \( r_i \sim N (0, \sigma^2) \).
4.1.7. Data Analysis

The data analysis was done in a phased manner. In the first phase a descriptive analysis was done wherein the frequencies have been reported for each of the variables of the study done. In the second phase multiple logistic regression analysis was carried out with the above mentioned model for the entire national data. The same analysis was done after dividing the sample into urban and rural areas.

4.2. STUDY 2

An important drawback of the studies based on secondary data is that the analysis and conclusions are a function of pre-existing data on a fixed set of variables. The study of the influence of all the relevant variables may therefore not be possible when secondary data is used. For the study proposed, the secondary data does not contain information on some variables (e.g. cost of healthcare, insurance, perceived service quality) that are part of the proposed model and are considered important for the choice of healthcare service provider. To overcome this limitation another study is suggested that intends to collect primary data from one urban and one rural area of India in order to study the influence of all the relevant variables. Additionally the second study, though on a limited scale, would also serve the purpose of the validation of the results obtained from the first study.

4.2.1. Data Source

For the purpose of study 2, the data collected from the field has been utilized. This survey was conducted in two phases from March 2013 to May 2013. It was done in the state of West Bengal in India. This survey used a questionnaire designed specifically for the purpose of this study. A
face to face interview technique with working adults was employed for collection of data from one urban area (Panihati) and one rural area (Haripal) of West Bengal.

4.2.2. Questionnaire Design

For the study 2, a separate questionnaire was designed and used (Refer to Appendix). This was due to the fact that NFHS-3 survey does not collect information on a few variables that are part of the dissertation framework to be empirically tested. These variables are health insurance, income, cost of treatment, distance between closest government hospital and residence of respondent and overall service quality perception of respondent regarding the service utilized. The questionnaire was designed to include all the variables that were part of the conceptual framework. The information for health service utilized by the target respondents was self reported in the questionnaire.

The questionnaire was designed in a pattern similar to the NFHS-3 survey questionnaire used. The questionnaire was structured and the questions were close ended. The questionnaire minimized the error caused due to recall bias by restricting the respondent to recall any health care use only till the last one month from the date of survey being conducted (Kurbat, Shevell, & Rips, 1998). Keeping in mind that the literacy levels of individuals would vary, and specific information regarding health care service use was required the questionnaires were not self administered. Rather a face to face interview technique was conducted and the researcher filled in the responses on behalf of each of the respondents irrespective of urban or rural sample. Accuracy of the information filled was ensured by checking in all the responses twice by the researcher. Once, immediately after filling in the response while the respondent was answering and again after the entire questionnaire was filled, all the responses were verified with the
respondent. Although, this meant more time was needed for each questionnaire to be filled but reliability and consistency in the quality of data gathered was made sure. Since, the questionnaire was administered by the author using the interview technique; consistency was maintained so as to minimize any kind of error caused due to interviewer bias. Additionally, since this survey is self sponsored hence any kind of auspices bias is not present.

Prior to taking the questionnaire to the field, the draft of the questionnaire was pre-tested by the researcher using a convenience sample of 30 respondents each from rural and urban areas. The time needed to fill in the questionnaire was checked. Additionally, any issues arising out while conducting a face to face interview procedure for filling out the questionnaire were noted and corrected. Accordingly the final questionnaire was prepared after taking note of the suggestions and comments received.

4.2.3. Study Design and Survey Sample

This study is cross sectional in design. For collecting data, a multistage sampling process was followed for urban as well as rural sample. Target population was adult working (any occupation) individuals who had utilized healthcare services for curative purpose for themselves in the past one month from the date of interview.

*Urban Area:* Panihati Municipal Corporation was selected for the urban area sample. This area is selected as this municipality is one of the largest in terms of area as well as population density in West Bengal. As a result, it is a representative sample of the urban population of India. Panihati consists of 35 wards with 71,857 households (Census, 2011). In the first stage, a single ward i.e. Ward No. 25 was chosen randomly from these 35 wards. It has 1671 households listed under it as per census data. In the second stage, households within this ward were selected based on
simple random sampling. The questionnaire was administered to one eligible individual (>18 years and working) from each household for collecting the relevant information. Face to face interview was conducted for this purpose. The total number of respondents that were interviewed was 316. The final number of the usable responses after the data cleaning process was 290.

**Rural Area:** Hooghly district was selected out of the 19 districts from the state of West Bengal and Haripal sub-district was chosen randomly out of it. This sub district consists of 48,130 households and had a population of 2,35,494 individuals (Census, 2011). From the census website of Indian government, the list of villages under Haripal sub district of West Bengal was taken out. For the first stage of sampling, the villages of Haripal, Nalikul and Bandipur were selected randomly as the rural area for the sample selection. In the next stage, households within these villages were selected based on simple random sampling. The questionnaire was administered to one eligible individual (>18 years and working) from each household for collecting the relevant information. Face to face interview was conducted for this purpose. The total number of respondents that were interviewed was 413. The final number of the usable responses after the data cleaning was 375.

**4.2.4. Variables & Measures**

*Dependent Variable*

The dependent variable for this study is the type of healthcare service provider utilized by an adult working individual for himself or herself under the condition of illness. It has been operationalized similar to the standard NFHS-3 questionnaire.

It was specifically observed during the data collection process that a considerable number of study sample utilized the Indian medicine service providers classified as AYUSH by the
government of India. Hence, any analysis that does not capture the strength of this additional option would remain incomplete in understanding the healthcare services utilization for Indian healthcare system. As a result, multinomial logistic regression analysis was required to be carried out on the sample obtained from the second study to capture all the three major categories of health service providers that are available to an individual in the event of an illness.

For the utilization of any type of public health care facility, the response was coded as “0” whereas for any type of private healthcare usage it was coded as “1” and for Indian medicine provider usage it was coded as “2”.

*Independent Variables*

The independent variables for this study consist of age, gender, education, income, religion, caste, marital status, exposure to mass media, distance between residence and closest government hospital, health insurance, cost of treatment, and overall service quality perception of healthcare services received at the facility of usage.

A few of the independent variables have been operationalized as per NFHS-3 questionnaires strictly. The details about them have already been mentioned above under the first study. The following part of this sub-section describes the operationalization of the remaining variables.

Education: It was treated as a categorical variable with three main categories. Those with atleast primary education or less than that (i.e. class 5th) were treated as the reference category and coded as “0”. Those with education up to higher secondary level (Upto class 12th) were coded as “1” and those having graduation and beyond were coded as “2”.
Income: For this study, income was treated as a continuous variable. Respondents were asked about their approximate average monthly income in Indian rupees.

Mass media exposure: For capturing the information on mass media exposure, the respondents were asked about the frequency of their exposure to the mediums of print, television and radio. If any of the individuals had daily exposure to any or all of the stated mediums above then they were classified under high mass media exposure. The remaining respondents were classified as low or no mass media exposure depending on their frequency of exposure. Thus this variable is a dichotomous variable with low/no exposure coded as the reference category “0” and high exposure to mass media coded as “1”.

Cost of treatment: For the variable of cost of treatment, the respondents were asked about the average cost of treatment in rupees borne by the respondent per illness episode. This variable is continuous and have been operationalized similar to studies done in India by earlier authors (Majumder, 2006).

Distance between residence and government hospital: Distance has been used on a number of occasions by previous authors (Lawson, 2004; Hjortsberg, 2003; Hutchinson, 2001) to capture accessibility. Here it has been captured in the form of distance to be travelled between the respondents’ place of residence and the nearest government hospital. Government hospital was chosen since it is believed to have the minimum capabilities required to treat patients. It was reported in kilometers.

Health Insurance: For the variable of health insurance status of an individual, the operationalization has been done using the measure of Cevallos & Chi (2010). A dichotomous variable was created wherein an individual was asked whether he/she has any kind of health
insurance. If the response was positive it was coded as “1” otherwise uninsured were coded as “0”.

Overall service quality perception: The importance of the perceived quality of healthcare services providers influencing the healthcare services utilized have been empirically found by many authors in the past (Pinto & Udwadia, 2010; Dilip & Duggal, 2004; Lindelow, 2004; Mariko, 2003; Mwabu et al. 1993). To capture the perceived service quality of the healthcare services provider utilized by the respondents, a five point scale was presented to them. It ranged from very poor to excellent and they were asked to rate the perceived quality of services utilized by them.

4.2.5. Model Specification

Since the target variable is polychotomous in nature wherein the dependent variable has three categories multinomial logistic regression models was proposed and tested. Two separate logit models dealing with the three categories of the dependent variable are given in details below.

A. Public healthcare services vs. Private healthcare services

Logit (p_i) = \beta_0 + \beta_1 (Age) + \beta_2 (Gender) + \beta_3 (Education) + \beta_4 (Religion) + \beta_5 (Marital Status) + \beta_6 (Caste) + \beta_7 (Income) + \beta_8 (Mass Media Exposure) + \beta_9 (Health Insurance Coverage) + \beta_{10} (Cost of treatment) + \beta_{11} (Distance between residence and closest govt. hospital) + \beta_{12} (perceived overall service quality) + r_i

Where:

p_i = Probability of observing the response for an individual i utilizing private healthcare services

Logit (p_i) = \log [p_i/(1 - p_i)] = log odds of utilizing the public healthcare service,
\[ \beta_0, \beta_1, \ldots, \beta_{12} = \text{intercepts and slope coefficients}; r_j = \text{residual for individual } i; r_i \sim N(0, \sigma^2) \]

**B. Public healthcare services vs. Indian Medicine Provider**

\[ \text{Logit } (p_i) = \beta_0 + \beta_1 \text{ (Age)} + \beta_2 \text{ (Gender)} + \beta_3 \text{ (Education)} + \beta_4 \text{ (Religion)} + \beta_5 \text{ (Marital Status)} + \beta_6 \text{ (Caste)} + \beta_7 \text{ (Income)} + \beta_8 \text{ (Mass Media Exposure)} + \beta_9 \text{ (Health Insurance Coverage)} + \beta_{10} \text{ (Cost of treatment)} + \beta_{11} \text{ (Distance between residence and closest govt. hospital)} + \beta_{12} \text{ (perceived overall service quality)} + r_i \]

**Where:**

\[ p_i = \text{Probability of observing the response for an individual } i \text{ utilizing Indian medicine providers healthcare services} \]

\[ \text{Logit } (p_i) = \log \left[ \frac{p_i}{1 - p_i} \right] = \log \text{ odds of observing the response capable of taking any real value;} \]

\[ \beta_0, \beta_1, \ldots, \beta_{12} = \text{intercepts and slopes estimated}; r_i = \text{residual for individual } i; r_i \sim N(0, \sigma^2) \]

**4.2.6. Data Analysis**

Similar to the first study, a descriptive analysis was carried out for the aggregate as well as urban and rural area samples separately. Then the data was analyzed using multinomial logistic regression analysis for the aggregate data as well as urban and rural area samples separately as was done in the first study.
TABLE 5: Coding scheme and description of variables used for study 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>Label</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Current age – respondent (Years)</td>
<td>NA</td>
</tr>
<tr>
<td>Gender</td>
<td>Gender</td>
<td>0:Male; 1:Female</td>
</tr>
<tr>
<td>Education</td>
<td>Highest educational level</td>
<td>0: No Education/Primary (Upto V) 1: Higher Secondary (Upto XII) 2: Graduate/PG/Higher than PG</td>
</tr>
<tr>
<td>Religion</td>
<td>Religion</td>
<td>0: Muslims 1: Hindu/Others</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Current marital status</td>
<td>0: Unmarried &amp; Others; 1: married</td>
</tr>
<tr>
<td>Caste</td>
<td>Type of caste or tribe</td>
<td>0: General Category/OBC 1: SC/ST</td>
</tr>
<tr>
<td>Income</td>
<td>Average monthly Income (INR)</td>
<td>NA</td>
</tr>
<tr>
<td>Exposure to Mass Media</td>
<td>Frequency of reading newspaper or magazine</td>
<td>0: No/Less mass media exposure (Weekly/No); 1: High mass media exposure (Daily)</td>
</tr>
<tr>
<td></td>
<td>Frequency of listening to radio</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Frequency of watching television</td>
<td></td>
</tr>
<tr>
<td>Cost of Treatment</td>
<td>Cost of treatment for the concerned visit to the facility (INR): cost of consultancy + cost of medicines + charges - for diagnostic tests if any</td>
<td>NA</td>
</tr>
<tr>
<td>Health Insurance</td>
<td>Coverage of the respondent under any type of health insurance</td>
<td>0: No 1: Yes</td>
</tr>
<tr>
<td>Access</td>
<td>Distance between residence and nearest government hospital (Kms)</td>
<td>NA</td>
</tr>
<tr>
<td>Overall Perceived Service Quality</td>
<td>Perceived service quality of the services received from the provider utilized</td>
<td>1: Very Poor 2: Poor 3: Average 4: Good 5: Excellent</td>
</tr>
<tr>
<td>FacilityType (Usage)</td>
<td>Type of facility visited</td>
<td>0: Public: Government/Municipal Hospital; Government Dispensary; UHC, CHC, PHC, Anganwadi/ICDS; Mobile clinic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1: Private: Private Hospital; Private clinic/dispensary; Pharmacy/drug store; Trust/NGO Hospital; Trust/NGO Dispensary/clinic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2: Indian Medicine Provider: Vaidya; Hakim; Homeopath</td>
</tr>
</tbody>
</table>