Chapter 3

Research design and method

3.1. Introduction

This chapter will provide information on selection of the study area, pre-testing, sources of data, sampling design, sampling procedure, sampling frame, techniques of data collection, reference to questionnaire, reference period, data processing, methods of analysing morbidity data, study of patient's choice of a system of medicine, factors affecting utilisation of services, cross tabulation, specification of the models, and conceptual framework.

3.2. Selection of the study area

Utilisation of health services varies sharply across cultures and societies in India. Important studies based on large-scale sample survey in India have been conducted either at National level or across broad regions. Studies based on small sample are also available, but those have been conducted either in northern or southern India. By and large this region of North Bengal remains unexplored. It has been decided to cover at least two districts namely, Cooch Behar and Jalpaiguri districts of this region to get the pattern of morbidity as well as utilisation of health care. As the district hospitals and other specialised sources of care are located in the district headquarters, it has been decided to cover Cooch Behar and Jalpaiguri towns. In order to sketch health-seeking patterns of potential users of those facilities who come directly from the adjacent rural areas or from other rural health care institutions as referral cases, it has been decided to cover Block II and Block I of the Sadar Sub-divisions of the two districts respectively.

3.3. Source of data

As the study is to analyse patients' health seeking pattern as well as epidemiologic profile from the point of view of self-perceived morbidity, it has been decided to go for primary data collected through interview technique. Supplementary information on status of health, profile of the study area, etc. have been taken from RHS-RCH Phase I & II, and Census 1991 and 2001 respectively.

3.4. Pre-testing

Before finalising the questionnaire, pre-testing of a draft questionnaire has been conducted in one urban area and rural area of each of the districts. This has been very helpful to finalise the questionnaire and also to get rough ideas about incidence and prevalence of morbidity, used later to determine the sample size.

3.5. Sampling design

3.5.1. Sample size

Determination of sample size depends on a number of technical and non-technical factors. Non-technical factors are: time and resources available for the study, geographical considerations, etc. Technical factors include objective of the study, type of model to be fitted, proportion of cases having the characteristics (under study) in the population, level of margin at which the study is designed. The level of margin for the present study is 0.05 (i.e. α=0.05). The pilot study revealed that the (annual) average period prevalence rate (proportion of persons who are exposed to the event of utilisation of care) of disease for the two districts are 0.348 and 0.489 (without multiplying by 1000) in rural and urban areas respectively. The average sizes of a household in rural and urban area of the two districts are 5.455 and 5.150 respectively (according to 1991 Census).

If 'n' is the size of a sample, P = proportion of cases having the characteristics, and Q = (1-P), then size of sample is:

$$n \frac{Q}{P * \alpha^2}.$$

This will give n = 0.652 / (0.348*0.0025) ≈ 749 persons or 749 / 5.455 ≈ 140 households in rural areas and n = 0.511 / (0.489*0.0025) ≈ 418 persons or 418 / 5.150 ≈ 80 households in urban areas of each of the districts. The total sample size is 2334 persons or 440 households in the two districts. Among the large-scale household surveys, in NFHS-II, 15-60 households have been selected from each village (EIT and IIPS 1999); in RHS-RCH (Mode 1998), 20 households have been selected from each village / ward. In the present study it has been decided to select 20 households from each mouza / village / ward. This leads us to select 140/20 = 7 mouzas / villages in rural area and 80/20 = 4 wards in urban area in each district. In order to consider non-response, etc., over-sampling of 10 per cent is done. In other words, 20 + 2 = 22 households have been selected from each mouza / village / ward leading to a total of 484 households.

It is to be mentioned that after completing the survey we get 2342 persons from 440 households – 1506 from rural and 836 urban areas. However, there are 325, 158, and 483 cases or illness episodes, which have been included in the analyses in the rural, urban, and the combined categories respectively.

3.5.2. Sampling procedure

The study adopts a multistage sampling technique. Different stages are as follows:

- Stage I: Cooch Behar and Jalpaiguri districts have been selected.
- Stage II: Cooch Behar and Jalpaiguri Sadar Sub-divisions have been selected.
- Stage III: CD blocks 'Cooch Behar – II' and 'Jalpaiguri – I' have been selected.
- Stage IV: Six Mouzas / villages from each CD block have been selected randomly and 1 mouza / village from each CD block has been selected purposively (as those were covered in the pilot study). Four wards from Cooch Behar Municipality have been selected randomly. Three wards from Jalpaiguri Municipality have been selected randomly and 1 (one) ward has been selected purposively (as it was covered in the pilot study).
- Stage V: Twenty-two households from each mouza / village / ward have been selected following simple systematic sampling technique.

3.5.3. Sampling frame

The main input for the sampling frames was voter lists. Voter lists are available for different polling booths, which strictly do not follow administrative domains and trailed in alphabetical order. Compiling voter list (s) of one or more booths, sampling frames have been prepared for each village / ward taking rigorous help from either present or ex-members of the Gram Panchayats or the Municipalities or local club members or persons who are actively involved in social work at grass root level.

Table 15 shows total population of the districts of Cooch Behar and Jalpaiguri. Fourth and fifth rows show total and average size of household in the two districts, which have been used as inputs to determine the sample size. Table 16 shows selected mouzas / villages / wards in the districts with number of households.

Table 15. Population of Cooch Behar and Jalpaiguri

	Key Indicators	Cooch Behar	Jalpaiguri
1	Total Population: 2001(in thousands)	2478	3403
2	Total Population: 1991(in thousands)	2171	2801
3	Growth rate: 1991-2001 (annual exponential, %)	1.32	1.95
4	Total Households: 1991	407203	460398
5	Average size of a household: 1991	5.15	5.17

Table 16. Sampled mouzas / villages / wards & Number of households*

	Cooch Behar			Jalpaiguri	
No.	Mouza / Vill / Ward	HH	No.	Mouza / Vill / Ward	HH
0003	Sajherpar Ghoramara	735	0004	Bahadur	2537
0008	Sakuni Bala	818	0005	Patkata	4371
0010	Bararangrash	1300	0007	Kharia	11757
0011	Salmara	283	0008	Mandal Ghat	2424
0014	Uttar Sibpur	373	0009	Goralbari	3979
0045	Konamalli	477	0023	Berubari	4165
0050	Bag Bhandar	119	0029	Boalmari	1131
0003	Amartala (Kalabagan)	450	0005	Samajpara	990
0005	Goalapatti	480	0007	Telipara	748
0009	Rail Gumti	400	0008	Babupara	566
0019	Sunity Road (Debibari)	600	0024	Vivekanandapara	1100

*HH: Households, * 1991 Census, Vill: village*

3.6. Tools and techniques of data collection & Questionnaire

3.6.1. Techniques of data collection

Data has been collected through interview technique with mostly a structured and close-ended questionnaire. In one section of the questionnaire, qualitative information has been collected adopting free listing technique.

3.6.2. Different parts of questionnaire

The questionnaire has 13 sections: Identification, Household Characteristics, Background Characteristics of the Household Population, Economic Profile, Morbidity, Vital Events, Utilisation of Care, Reasons behind Choosing a Particular Type of Care / System of Medicine, Availability of Health Facilities, Accessibility to Health Care & Activity Set, Quality of Care, Expenditure, Food Habit.

3.6.3. Reference period

Data has been collected roughly for a 5-month reference period in the second half of 2003.

3.7. Methods of analysing data

The study has three facets: morbidity analysis (examination of the phenomenon of epidemiological transition, and morbidity rates), study of household's preference for a care (sketching patients' or households' cognitive structure), and estimation of contribution of different need, predisposing, and enabling factors towards utilisation of a care (multivariate analyses using binary Logistic Regression Analysis, LRA and Multiple Classification Analysis, MCA).

3.7.1. Methods of analysing morbidity statistics

In order to carry out studies on epidemiological transition data on morbidity will be classified according to the Global Burden of Disease (GBD) study 1990 (Murray and Lopez 1996).

The observed distribution will be compared with the hypothesised ones (using Chi-square statistic) to test whether epidemiological transition has taken place in rural and urban areas of Cooch Behar and Jalpaiguri districts of North Bengal.

Table 17. Classification of diseases as in Global Burden of Diseases study 1990

Cause Group	Major Categories
Group I: Communicable, maternal, perinatal, and nutritional diseases	Infectious and parasitic diseases, Respiratory infections, Maternal conditions, Conditions arising during the perinatal period, Nutritional deficiencies
Group II: Non-communicable diseases	Malignant neoplasms, Diabetes mellitus, Endocrine disorders, Neuro-psychiatric conditions, Sense organ diseases, Cardiovascular diseases, Chronic respiratory diseases, Digestive diseases, Genito-urinary diseases, Skin diseases, Musculoskeletal diseases, Congenital anomalies, Oral conditions
Group III: Injuries	Unintentional injuries, Intentional injuries

In order to compute rates of morbidity, the illnesses that exist in a population during a given time interval may first be classified as follows (see Hill 1966):

a) Illness beginning during the interval and ending during the interval.
b) Illness beginning during the interval and still existing at the end of the interval.
c) Illness existing before the beginning of the interval and ending during the interval.
d) Illness existing before the beginning of the interval and still existing at the end of the interval.

For each of the above categories we are interested to measure rates based on number of spells. We need number of illness in the first two categories to measure incidence rates.

$$\text{Incidence Rate (annual)} = \frac{I}{P} * \frac{365}{150} * 1000$$

where I is the number of new cases of illness in the 5-month reference period per 1000 average number persons living in the community during the reference period.

$$\text{Period Prevalencerate(annual)} = \frac{C}{P} * \frac{365}{150} * 1000$$

where C is total number of spells (in all the four categories) in the 5-month reference period per 1000 average number persons living in the community during the reference period. These annual rates can also be converted into monthly rates by replacing the numerator of the formulae, 365 by 30.

Morbidity rates will be computed for rural, urban, and combined categories.

3.7.2. Method of studying of patient's choice of a care

All individual responses will be tabulated according to their rank in free-lists. If there are n-numbers of opinions, those in the first, second, ..., n-th ranks will get weights as follows:

$\{(n-0)/(1+2+...+n)\}, \{(n-1)/(1+2+...+n)\}, ..., [\{n-(n-1)\}/(1+2+...+n)]$.

The underlying assumption behind such weighting system is that importance of each opinion in individual list decline linearly. Frequency of each opinion may vary sharply as all respondents may not mention all items. Total weight of each opinion will then be computed by simple aggregation. These will sharply vary according to their ranks and frequencies. All weights will then be aggregated to get the grand sum and weight of each opinion will be expressed as a proportion or share of the grand sum. These quantities are nothing but the salience of each opinion in respondents' minds.

For example (for one individual free-list only), if there are 3 opinions in favour of allopathy as: Permanent cure (in the first rank), Quick relief (in the second rank), and Reliable (in the third rank), then salience of

Permanent cure = $\{(3-0)/(1+2+3)\} = 0.500$,

Quick relief = $\{(3-1)/(1+2+3)\} = 0.330$,

Reliable = $\{(3-2)/(1+2+3)\} = 0.170$,

such that Σ Salience = 1.000.

The above results are to be comprehended as – on an average, half (50 per cent) of the space of respondent's mind is occupied by the feeling that allopathy leads to permanent cure, 33 per cent of the space is full by the feeling that it provides quick relief, and the rest (17 per cent) is occupied by the sensation that it is reliable. Though all the above factors influence respondents to choose allopathic system of medicine, the above method provides us with precise estimates of 'salience' or importance of each opinion in people's mind.

3.7.3. Methods of analysing factors affecting utilisation of services

Utilisation of services may be considered as an event (Béland 1988). In that case it will be binary in nature. We may assign it 1 if the event has occurred, 0 otherwise. Utilisation of care may have many dimensions. After going through data, we have found suitable to form two broad groups: utilisation of a care from modern source in consultation with doctors and medical specialists in one group, and utilisation from traditional source (including treatment from paramedical or supporting stuff and from any system of medicine except allopathy and homeopathy) or self-treatment or family-treatment, etc. in the other. From the above review of literature we found the following predictor variables relevant which may affect health services utilisation in North Bengal: age, gender, and caste of the morbid person, family size (size of a household), education of the head of the household, normal out-of-door trips by the head of the household, household cash income, type of illness, severity of illness, type of health facility, system of medicine, quality of care, and total direct costs or price of a care. However, as household cash income may always be not related to ability to pay health care, we plan to include some proxy measures of households' agricultural possessions and standard of living. In addition to this, as this particular region is far away from the important Indian cities, and as people of this region are compelled to travel a lot, we can examine whether this traveling habit has any bearing on utilisation of services. Finally, studies based on small sample survey could not explore the relationship between availability of health facilities and utilisation of care mainly because of common sources of care for many people. But one can consider place of residence as a proxy measure of availability (Elo 1992) with the assumption that in the rural areas health facilities are not easily available but available in urban areas. Definitions of the response and predictor variables are shown in table 16.

If P be the estimated probability of utilising a care from modern source, in probability form, the model is:

$$P = \frac{1}{1 + e^{\beta_0 + \Sigma \beta_{1i} X_{1i} + \Sigma \beta_{2i} X_{2i} + \Sigma \beta_{3i} X_{3i}}}.$$

In log odds form, the model is:

$$\text{Log}\,\Omega = \beta_0 + \sum \beta_{1i} X_{1i} + \sum \beta_{2i} X_{2i} + \sum \beta_{3i} X_{3i}.$$

In odds form the model is:

$$\Omega = \exp\left(\beta_0 + \sum \beta_{1i} X_{1i} + \sum \beta_{2i} X_{2i} + \sum \beta_{3i} X_{3i}\right),$$

where Odds $\equiv \frac{P}{1-P} \equiv \Omega$ (Retherford and Choe 1993). The equations include need (X_{1i}), predisposing (X_{2i}), and enabling factors (X_{3i}).

The results of the above logistic regression models will be transformed into simple cross tabulation of the probability of utilising any type of health care using multiple classical analysis. This will involve calculation of adjusted and unadjusted values of the response variables for each category of predictor variables.

As the indicator-coding scheme has been adopted through out the analyses, for each predictor variable results of logistic regression analysis (LRA) are available for n minus one (n – 1) categories in contrast to the n^{th} category, the reference category (the first or last as specified). The advantage of multiple classification analysis (MCA) over the LRA is that in the former both adjusted and unadjusted probability values are readily available for all the predictor variables and categories. Moreover, in the regression analysis it is customary to display statistically significant results only. However, in the MCA one can include all the β-coefficients irrespective of their statistical significance (see Retherford and Choe 1993) as each probability value in the adjusted columns are based on a set of statistically significant or insignificant β-coefficients. Unadjusted probability, in the present context, means the effect of one particular variable towards pattern of utilisation of a care when all other predictor variables are absent in the model. Adjusted probability means the effect of one particular variable towards pattern of utilisation of a care when all other predictor variables are controlled at their mean values. As a result the set of controlled variables change as we move down the table.

3.7.4. Cross tabulation

In addition to the above, cross tabulation will be done and clustered bar charts will be prepared to present and represent the characteristics of the subject, disorder, and service.

3.7.5. Conceptual framework

The conceptual framework for the study is presented below. It has been modified after Kroeger (1983). The first node stands for perceived morbidity, which interacts with predisposing, need, and enabling factors. Nodes in the third panel display a set of possible explanatory variables, which are supposed to play significant roles in determining choice of a care. Though choice of a care has many dimensions, we have kept utilisation of a care from modern source in one category and utilisation from traditional source or self treatment or family treatment, etc. in another as shown in the nodes in the fifth panel.

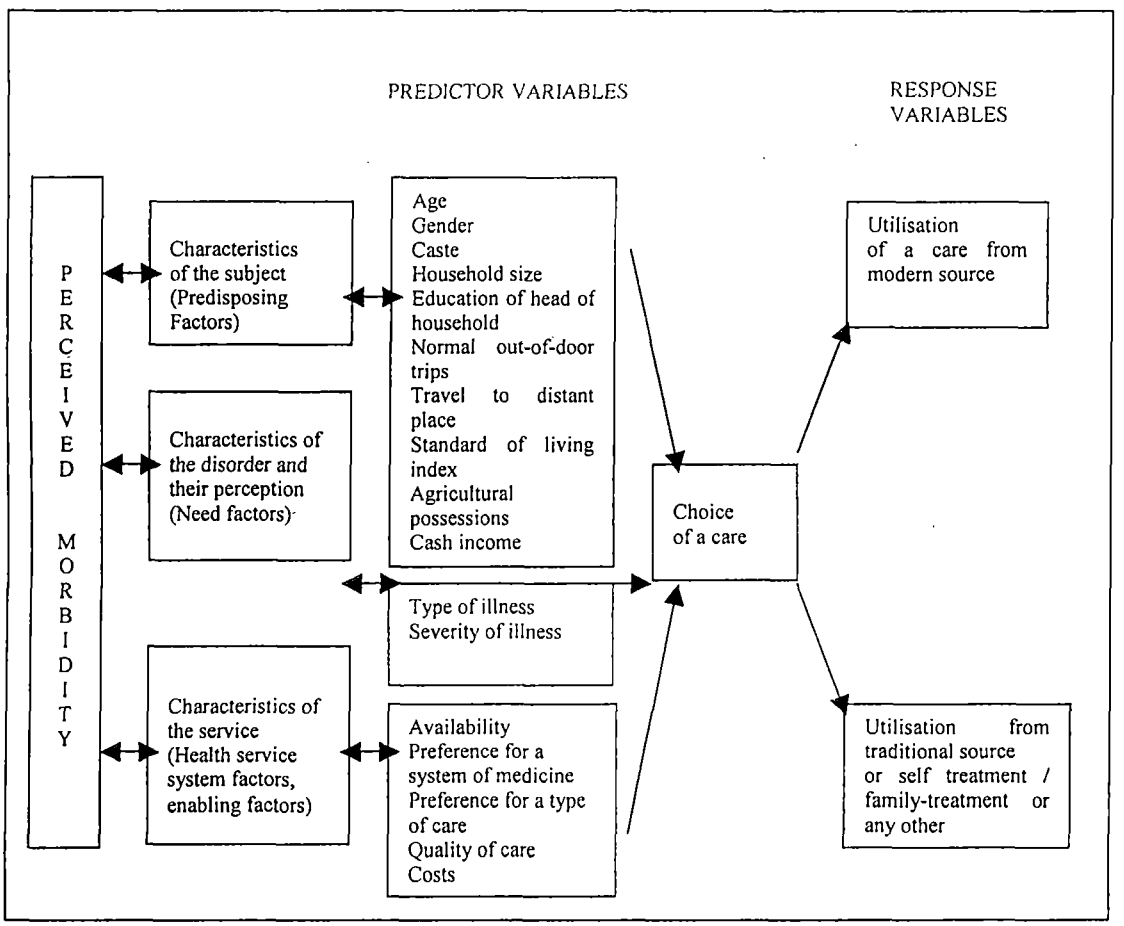

Figure 1. Conceptual framework to examine pattern of utilisation of care

3.7.6. Definition of the variables

Definitions of the response and predictor variables are shown below all of which are categorical.

Table 18. Variables in the model and definitions

Variable	Definition	Value
Utilisation	Whether the household utilised care from any modern source	1 if the event has occurred, 0 Otherwise.
Age	Age of the morbid person	1 if age 5-14 0 otherwise; 1 if age 15+ 0 otherwise.
Gender	Gender of the morbid person	1 if female, 0 otherwise.
Caste	Caste of the morbid person (General / Scheduled Caste / Tribe)	1 if general, 0 otherwise.
Family size	Number of persons in the household	1 if size ≤ 5, 0 otherwise.
Education	Education of the head of the household	1 for illiterate and up to primary, 0 for middle and above.
Normal out-of-door trips	Number of travels by the head of the household within 10 kilometres range in a month	0 if number ≤ 4, 1 otherwise.
Travel to distant places	If the head of the household travelled beyond 500 kilometres range in past three years	1 if the event has occurred, 0 Otherwise.
Standard of living	A composite index based on proportion of living rooms to persons (1 if proportion ≥ 0.5, 0 otherwise), type of house (1 if pucca or semi-pucca, 0 otherwise), type of toilet facility (1 if sanitary, 0 otherwise), audio system (1 if yes, 0 otherwise), TV (1 if yes, 0 otherwise)	1 if score >3, 0 otherwise.
Agricultural possessions	If the household possesses cultivable land, milch animals, draft animals, birds, and fruit trees. For each item the score is 1 if the household possesses it, 0 otherwise	1 if score >3, 0 otherwise.
Cash income	Household monthly cash income from all sources	1 if 2000 ≤ income ≤ 4999, 0 otherwise; 1 if income ≥5000, 0 otherwise.

Continued on page 57

Continued from page 56

Type of illness	Morbidity	1 for Group II, 0 otherwise; 1 for Group III, 0 otherwise.
Severity	How sever the attack is	1 for medium, 0 otherwise; 1 for high, 0 otherwise.
Type of facility	Public / private / other	1 for public, 0 otherwise.
System of medicine	Allopathy / Homeopathy / Traditional (Traditional: Ayurvedic, Kabiraji, etc.)	1 for allopathy, 0 otherwise; 1 for Homeopathy, 0 Otherwise.
Quality of care	Composite index on households opinion on cleanliness (yes/no), whether privacy is maintained (yes/no), service provider listen to the patient/other (yes/no), service provider talk to the patient/other (yes/no), and the household is satisfied (yes/no). For each item the score is 1 if the answer is 'Yes', 0 otherwise	1 if score >3, 0 otherwise.
Costs	Total direct cost per episode	1 if 100 ≤ cost ≤ 499, 0 otherwise; 1 if cost ≥500, 0 otherwise.