RESEARCH METHODOLOGY

Health care has always been a problem area for India, a nation with a large population and a larger percentage of this population living below the poverty line. In such a situation insurance becomes an important issue in the country. But surprisingly, for a country with the 5th largest economy, insurance in India has not been the sector that has taken off, considering its immense potential. From the literature review it has been observed that there are many areas in the field of health insurance that are unexplored or need to be explored in different context. This study aims to fill important theoretical and methodological gaps that were found.

3.1(i) The objectives of the study are hence as follows:

1. To study the motives behind the changing face of insurance sector in India.

2. To study the current scenario in insurance, future positions bottlenecks that could be faced by health insurance due to changing face of insurance sector in India.

3. To analyze the effect of opening up of insurance sector on health insurance in India.

4. To review the health insurance scene in India and consumer’s perspectives on health insurance.

5. To analyze future prospects of health insurance in India.

6. To study the role of IRDA in marketing the important beginnings of the changes having significant implications for the health sector in India.

7. To analyze the factors responsible for health insurance purchase decision and also factors affecting the extent of purchase decision of health insurance in India.

3.1(ii) Data Collection

The research design of the current study is descriptive and qualitative in nature. Both primary and secondary data has been used in the study.
First six objectives of the study are studied on the basis of secondary data obtained through various publications of central and state government; technical and trade journals; research journals; books, magazines and newspapers; reports and publications of various bodies like IRDA, FICCI, Ministry of health, Tariff Advisory Committee, Insurance Information Institute, WHO etc.; annual reports prepared by insurance companies operating in India; reports prepared by various research scholars, universities, economists etc.; public records and statistics; and other sources of published information like internet. Reliability, suitability and adequacy of data were checked before conducting the study.

Primary data is collected through questionnaire. The data was collected on different socio economic variables like income, gender, education etc and also related to people’s perception about buying health insurance on interval scale data. These factors and variables are identified on the basis of interview and discussion with policy holders and other stakeholders of health insurance.

3.1 (iii) Sampling

Five districts in Haryana, i.e Rohtak, Hisar, Gurgaon, Kurukshetra and Sonipat, were selected for survey. All 19 general insurance companies are offering health insurance to the people living in and nearby these districts. Here people can get treatment from renowned hospital of that area. The main target segment for the study is middle income group with three sub segments as lower-middle, middle-middle and upper-middle class. For selecting unit of analysis to do the survey, we can take either individual as a unit of analysis or consumer unit as a unit of analysis or family as a unit of analysis. In this study household has been taken as unit of analysis. It has been done because health is a family decision and health of one family member may affect health of another family member. The household has been defined as a group of persons normally living together and taking food from the same kitchen excluding persons who are not related by family or legal arrangements. Total sample of 400 was taken out of which 369 households have responded. Stratified Random Sampling method is used to select the respondents so that we can have all categories of people in the sample, for example people who have bought insurance, people who have not bought insurance and also people who have renewed insurance policy later.

3.1(iv) Hypotheses formulation
While analysing the insurance purchase decision and estimating the extent of health insurance purchase decision (which is only possible if house hold has already taken decision to buy insurance) certain hypotheses have been made by the researcher. Three hypotheses have been made, they are:

Regarding Factors affecting decision of health insurance purchase (to buy or not to buy health insurance) it has been hypothesised that decision to purchase health insurance is linearly related to the income of the house hold and secondly it is also linearly related to the health insurance expenditure as a ratio of total expenditure.

1) \( H_0: Z = a + b I \); where \( I \) is the income of the household and \( Z \) is the decision to buy or not to buy. \( Z \) will be 1 if health insurance is bought and 0 if not purchased.(Null Hypotheses)

\( H_1: Z \neq a + b I \) (Alternate Hypotheses)

2) \( H_0: Z = a + b H.E \); where \( H.E \) is the Health Expenditure to total expenditure ratio of the household and \( Z \) is the decision to buy or not to buy. \( Z \) will be 1 if health insurance is bought and 0 if not purchased.(Null Hypotheses)

\( H_1: Z \neq a + b H.E \) (Alternate Hypotheses)

3) \( H_0: E = a + b I \); where \( I \) is income of the Household and \( E \) is extent of health insurance purchase. (Null Hypotheses)

\( H_1: E = a + b I \) (Alternate Hypotheses)

Econometric analysis is used to find the factors affecting health insurance purchase decision. The decision for buying insurance has been formulated in two interrelated choices. First, the choice is related to buy or not to buy health insurance. Second, if decision is to buy the insurance then to buy for how many people in the family and for what kind of coverage i.e. extent or amount of insurance. In each choice, the decision maker faces a finite and exhaustive set of mutually exclusive alternatives; the qualitative choice models are applied to describe these two choices. Heckman two stage model is used to solve the problem of selection bias. This model uses bivariate normal model for the selection equation, and ordinary least squares to estimate the behavioural equation with the selected sample.

In the analysis first households are segregated into those who have taken the insurance and those who have not. Then those who have bought insurance are analysed to see to what extent
households have bought insurance. In the framework of sample selection, one equation specifies whether or not household is buying insurance and another equation determines the extent of buying insurance.

The basic idea of a sample selection model is that the outcome variable, $y$, is only observed if some criterion, defined with respect to a variable $z$, is met. The common form of the model has two stages. In the first stage, a dichotomous variable $z$ determines whether or not $y$ is observed, $y$ being observed only if $z=1$.

In the second stage, the expected value of $y$ is modelled, conditional on its being observed. So, $z$ a dummy variable is observed, which is a realization of an unobserved continuous variable $z^*$, having a normally distributed, independent error $e$, with mean zero and a constant variance $\sigma^2$. For values of $z=1$, $y$ is observed, which is the observed realization of second latent variable (and model that with some independent variable $x$ and get a vector coefficients beta), $y^*$, which has a normally distributed, independent error, $u$, with a mean zero and a constant variance $\sigma^2$. The two errors are assumed to have a correlation $\rho$. The joint distribution of $u$ and $e$ is bivariate normal.

In first equation dependent variable is a binary variable, which takes the value of one if household has bought health insurance and zero if it has not bought health insurance. Since dependent variable can only take two values simple OLS is not used rather binary discrete choice modelling is used here. Here the observed dependent variable is discrete. An appropriate estimator in this case can be either the Probit or Logit maximum likelihood estimator. PROBIT model is used in this study to analyze first equation related to health insurance purchase decision.
In the second equation, where only the observations where households having bought health insurance have been used, the dependent variable is measuring extent of health insurance purchase and is a continuous variable. Therefore, OLS can be used in this case. In second equation Inverse Mills Ratio (IMR) is used as one of the independent variable, which is calculated in first equation. IMR is a "correction term" for the bias that arises from the selectivity bias problem. Heckman characterized the sample selection problem as a special case of omitted variable problem with IMR being the omitted variable.

In the case of independent variables in both equations other than income and health expenditure, many control variables have been used. These control variables are as follows:

- Age
- Gender
- Education
- Number of Children
- Number of people in family

Eight qualitative factors have been identified which effect the health insurance purchase decision and are used as independent variables while analysing the model. Respondents were asked to rate these variables on the scale of one to five where five is for strongest. These variables have been used as interval variables. The factors used in the study are:

- Cost
- Quality
- Accessibility
- Coverage of illness
- Coverage of Services
- Trust
- Illness expenditure
• Knowledge about insurance

The models which were used for the two equations are as follows:

**Equation 1 (Probit)**

\[
(\text{Insurance Purchase}) = \beta_0 + \beta_1 (\text{Income}) + \beta_2 (\text{Income}^2) + \beta_3 (\text{Income}^3) + \beta_4 (\text{Health Expenditure/Total Expenditure}) + \beta_5 (\text{Age1}) + \beta_6 (\text{Age2}) + \beta_7 (\text{Age3}) + \beta_8 (\text{Age4}) + \beta_9 (\text{Education}) + \beta_{10} (\text{Gender}) + \beta_{11} (\text{Hospitalisation Cost}) + \beta_{12} (\text{Other Health Costs}) + \beta_{13} (\text{Cost}) + \beta_{14} (\text{Quality of Care}) + \beta_{15} (\text{Accessibility}) + \beta_{16} (\text{Coverage of services}) + \beta_{17} (\text{Coverage of illness}) + \beta_{18} (\text{Trust}) + \beta_{19} (\text{Knowledge about insurance}) + \beta_{20} (\text{Illness expenditure}) + \epsilon
\]

**Equation 2 (Outcome Equation)**

\[
(\text{Extent of Insurance Purchase}) = \beta_0 + \beta_1 (\text{Income}) + \beta_2 (\text{Income}^2) + \beta_3 (\text{Income}^3) + \beta_4 (\text{Health Expenditure/Total Expenditure}) + \beta_5 (\text{Age1}) + \beta_6 (\text{Age2}) + \beta_7 (\text{Age3}) + \beta_8 (\text{Age4}) + \beta_9 (\text{Gender}) + \beta_{10} (\text{Child}) + \beta_{11} (\text{Hospitalization Cost}) + \beta_{12} (\text{Other Health Costs}) + \beta_{13} (\text{Cost}) + \beta_{14} (\text{Quality of Care}) + \beta_{15} (\text{Accessibility}) + \beta_{16} (\text{Coverage of illness}) + \beta_{17} (\text{Coverage of services}) + \beta_{18} (\text{Trust}) + \beta_{19} (\text{Illness expenditure}) + \beta_{20} (\text{lambda}) + \gamma_i
\]

(\text{Lambda here denotes Inverse mills Ratio})

To find the factors determining the probability of health insurance purchase and at the same time once decided to purchase health insurance which factors will affect the extent of insurance purchase, two separate equations are used. Heckman two step method is used to take care of sample selection bias. In this model, the first equation is a discrete choice model using probit which is related to probability of buying insurance and second equation is a simple OLS where the extent of health insurance purchase has been analysed.

**3.2 LIMITATIONS OF THE STUDY**

Tabulation in the report is generated from the data supplied by the TPAs only for the period of 2007-08. Any data in the recorded form is unavailable. Data on health insurance is highly unorganised. Aggregate data published in this are from the transaction level data supplied by the TPAs only. These may or may not match with individual insurer’s company level aggregate data of health coverage as insurers provide health coverage under different schemes which may or may not be serviced through TPAs.
Claims figures reflect pure claim cost. Expenses/interest are not included. The tables are indicative and contextual. Findings are not complete/conclusive and anyone making use of the same is advised to make his/her independent assessment of the accuracy.

There could be errors and omissions.

3.3 SELECT REFERENCES

Allen, T. Harell; *New Methods in Social Science research*, Economic and political weekly, 1986 pp 67-71


Gopal, M.H., *Research Reporting in Social Sciences*, Dharwar: Karnataka University, 1995 pp- 17-23


