Chapter III  
Data Base and Research Methodology

This chapter deals with the detailed discussion of data base and research methodology applied for the achievement of objectives. Basically, for the purpose of present study two set of data were collected and used, which include both primary as well as secondary sources. The study is mainly based upon the use of secondary data, except to examine the customers’ perception regarding health insurance. The present chapter is divided into five sections according to the objectives of the study; each section aims at discussion of data base and research methodology applied for the achievement of particular objective. Section A deals with the data base and research methodology used to evaluate the performance of health insurance business of general insurance companies in India. Section B deals with the data base and research methodology used for the comparative study of health insurance plans offered by general insurance companies in India. Section C covers the data base and research methodology used to examine the working and performance of Community Health Insurance (CHI) in India. Section D deals with the data base and research methodology used to examine the role played by Third Party Administrators (TPAs) with reference to role defined by IRDA. Section E covers the data base and research methodology used to examine customers’ perception regarding health insurance.

Section A

3.1 Data Base and Research Methodology for the Performance Evaluation of Health Insurance Business of General Insurance Companies in India.

In order to achieve the above stated objective, the study was conducted with Indian general insurance companies and covered a period of 8 years from 2002-03 to 2009-10. Presently, more than 20 general insurance companies are operating in India out of which 10 (including 4 public sector insurance companies) were selected on the basis of base year 2002 i.e. the companies which are providing health insurance since 2002 are forming a part of the present study. The basic reason behind the selection of base year as 2002 lies in the fact that, this is the exclusive year in which general insurance companies
have started the extension of health insurance products with the integration of Third Party Administrators (TPAs) in order to meet the requirement of IRDA (Third Party Administrator-Health Services) Regulation, 2001, which was introduced in order to infuse a new management system and to regulate the health care services and costs. Accordingly, the companies under the scope of study include:

- Tata AIG General Insurance Company Limited;
- IFFCO Tokio General Insurance Company Limited;
- Reliance General Insurance Company Limited;
- Royal Sundram Alliance Insurance Company Limited;
- Bajaj Allianz General Insurance Company Limited;
- ICICI Lombard General Insurance Company Limited;
- United India Insurance Company Limited;
- Oriental Insurance Company Limited;
- New India Assurance Company Limited; and
- National Insurance Company Limited.

Moreover, in order to achieve the above stated objective, the detailed information regarding health insurance was drawn mainly from the annual reports of the insurance companies under consideration. For this, the use of respective sites of the general insurance companies was made. Beside this, the use of statistical year book of IRDA and other publications related with the insurance was made for the collection of certain facts and figures necessary for the purpose of performance evaluation. The data collected was analyzed with the help of followings:

- Data Envelopment Analysis (DEA)
- Malmquist Total Factor Productivity Analysis (MTFPA)

All this was done with the help of DEAP software package.

3.1.1 **Data Envelopment Analysis (DEA):** It is a mathematical programming approach which estimates the frontier by solving a series of linear programming problems. The efficiency of the each firm is measured with respect to distance from the frontier. Generally, the efficiency scores ranges from 0 to 1 and a firm having efficiency score equal to 1 referred as a fully efficient firm or a firm operating on the frontier and a firm having efficiency score less than 1 referred as inefficient firm or a firm operating away
from the frontier. The firms or companies whose efficiency is measured with the help of DEA are called Decision Making Units (DMUs). DMUs efficiency is obtained by solving a set of linear programming problems which can be categorized in to following according to treatment of slack:

(a) **One-stage DEA** which uses the following equation and calculate slack residually:

\[
\begin{align*}
\min_{\theta, \lambda} & \quad \frac{y_i}{\theta} \\
\text{st} & \quad y_i + X\lambda \geq 0, \\
& \quad \theta_i - \lambda_i \geq 0, \\
& \quad \lambda \geq 0.
\end{align*}
\]

…………………………………… (1)

Where \( \theta \) is a scalar and \( \lambda \) is a Nx1 vector of constants. This envelopment form involves fewer constraints than the multiplier form \((K+M < N+1)\), and hence is generally preferred to solve. The value \( \theta \) obtained will be the efficiency score for the i-th DMU. It will satisfy \( \theta \leq 1 \), with a value of 1 indicating a point on the frontier and hence a technical efficient DMU, according to Farrell (1957) definition. Moreover the linear programming problem must be solved N times, once for each DMU in the sample.

(b) **Two-stage DEA** which uses the following equation:

\[
\begin{align*}
\min_{\theta, \lambda} & \quad \frac{y_i}{\theta} \\
\text{st} & \quad y_i + X\lambda - OS = 0, \\
& \quad \theta_i - \lambda_i - IS = 0, \\
& \quad \lambda \geq 0, OS \geq 0, IS \geq 0.
\end{align*}
\]

…………………………………… (2)

Where OS is an Mx1 vector of output slack, IS is a Kx1 vector of input slacks and M1 and K1 are Mx1 and Kx1 vector of ones, respectively. In this case, \( \theta \) is not variable rather its value is taken from the first stage results. Moreover it is required to be solved for each of the N DMU’s involved. However there are two problems associated with this: firstly, sum of slack maximized rather than minimized. Secondly, it is not invariant to units of measurement. The alternative to get rid of these problems is the use of multi-stage DEA.
(c) **Multi-stage DEA**, where we conduct a sequence of radial LP’s to identify the efficient projected point. This method is computationally more demanding than the other two methods above specified due to the fact that it identifies efficient projected points which have input and output mixes which are as similar as possible to those of the inefficient points, and that it is also invariant to units of measurement (*Coehli, T.J. 1996*). Hence we have used multi-stage method over the other two alternatives.

The above stated equations are used in DEA to evaluate the efficiencies of the DMUs. In our paper, efficiency is defined as technical efficiency. Further, only the measurement of technical efficiency is not enough, so we worked ahead and measured the technical efficiency both with Constant Return to Scale (CRS) as well as with Variable Return to Scale (VRS). The technical efficiency which is measured with VRS is also known as Pure Technical Efficiency. Besides this, scale efficiency has also been calculated for all the firms during the period under consideration.

- **Technical Efficiency**: Technical efficiency can be regarded as the product of pure technical efficiency and scale efficiency. It reflects the ability of the firm to obtain the maximum output from a given set of input or the efficiency with which inputs are transformed into output or just the output/input ratio.

- **Pure Technical Efficiency**: In pure technical efficiency, production line with variable return to scale is used. From the viewpoint of economics, this will release the restrictions of scale. Therefore, the inefficiency only lies in the factors such as productivity, resource allocation and management.

- **Scale Efficiency**: In contrary to the case of pure technical efficiency, only the factor of scale is effective here, while the factors of productivity, resource allocation and management are excluded.


**3.1.2 Malmquist Total Factor Productivity Analysis (MTFPA)**: Further, in order to measure the productivity as well as change in productivity of health insurance business of general insurance companies in India, the use of Malmquist Total Factor Productivity Analysis (MTFPA) was made. It is a form of DEA, which helps in measurement of
productivity as well as productivity change. In other words, DEA helps in the measurement of productivity with the use of Malmquist index summary. This summary provided with Efficiency Change (EC), Technological Change (TC), Pure Technical Efficiency Change (PTEC), Scale Efficiency Change (SEC) and Total Factor Productivity Change (TFPCH). Overall, it provided with Total Factor Productivity Change (TFPC), which comprises of EC, TC, PTEC and SECE. The growth of TFP has two major components: technological change/technical progress and efficiency change. Technological change/technical progress is represented by a shift in the production frontier while efficiency change is based upon an index of a firm’s efficiency relative to past and future frontiers (Coehli T.J. 1996).

If we let \( D_t (x_t,y_t) \) be the distance from the origin for a firm with an input vector of \( x_t \) and an output vector of \( y_t \) where \( t \) represents time and the subscript \( s \) for the firm is omitted for clarity, then,

\[
M^t = \frac{D^t(x_t,y_t)}{D^t(x_t^{t+1},y_t^{t+1})}
\]

.................................................. (3)

The ratio of the two distances, \( M_t \), is the Malmquist productivity index. This index suffers from a problem that it depends upon the starting values. For example, if we examined the Malmquist index with respect to the period \( t + 1 \) frontier, we have

\[
M^{t+1} = \frac{D^{t+1}(x_t^{t+1},y_t^{t+1})}{D^{t+1}(x_t^{t+1},y_t^{t+1})}
\]

.................................................. (4)

So, to avoid an arbitrary choice of which frontier to choose, we take the geometric mean, which yields the Malmquist index of total factor productivity,

\[
M(x_t^{t+1},y_t^{t+1},x_t^{t+1},y_t^{t+1}) = \left[ \frac{D^t(x_t^{t+1},y_t^{t+1})}{D^{t+1}(x_t^{t+1},y_t^{t+1})} \times \frac{D^{t+1}(x_t^{t+1},y_t^{t+1})}{D^{t+1}(x_t^{t+1},y_t^{t+1})} \right]^{1/2}
\]

.................................................. (5)

The change in efficiency is thus the ratio of the distance from the frontier in period \( t \) to the distance in period \( t + 1 \). If technical efficiency increases, the ratio will be greater than one and, if it decreases, the ratio will be less than one.
To calculate technical change, we examine how the firm uses its inputs to produce outputs in periods $t$ and $t + 1$ and how the input/output bundles change over time. Technical change is computed as follows:

$$TC(x^{t+1},y^{t+1},x^t,y^t) = \left[\frac{D^{t+1}(x^{t+1},y^{t+1})}{D^t(x^{t+1},y^{t+1})}\right] \times \frac{D^{t+1}(x^t,y^t)}{D^t(x^t,y^t)}^{1/2}$$

................................. (6)

If favorable technological change exists, the frontier will moved to the left, and both output bundles will be farther from the period $t + 1$ frontier than they are from the period $t$ frontier. Once again, a ratio greater than one indicates progressive technical change.

(Source: Coehli T.J. 1996)

3.1.3 Measurement of Inputs and Outputs

The result of Data Envelopment Analysis (DEA) and Malmquist Total Factor Productivity Analysis (MTFA) depends heavily upon the variable of the input and output specified. Therefore for the success of study, an accurate selection of variable of input and output is necessary. With the review of literature, it came into notice that researchers have general agreement on the selection of input’s indicators. Generally labor, capital and materials are selected as shown in Grace and Timme (1992), Gardner and Grace (1993), Cummins and Zi (1998).

3.1.3.1 The variables of inputs which are considered under the present study are as:

(a) Equity Capital (X1): According to the theory of corporate finance, financial capital can be authorized capital or invested capital. The invested capital includes long term debt and equity capital. However, reserve constitutes almost all the long term debt of insurers. Debt which constitutes the part of the invested capital does not remain same. Therefore, we cannot say that insurers use this long term debt to support their business. So in this study, we have taken equity share capital as the first indicator of input.

(b) Amount of Labor (X2): As the insurance companies do not constitute the part of the manufacturing industries rather form a part of the financial service industry, so it does not uses raw material as primary input. In financial service industry, the cost of labor is the most important input. The main difficulty regarding this input was to get salary data of
insurers, which is not feasible. So alternatively, we have taken amount of commission, agents’ fees, referral and other expenditure as the second indicator of input.

**3.1.3.2 The variables of outputs which are considered under the present study are as:**

As far as the variable of outputs is concerned wide diversity found in literature, because of the fact that defining and measuring outputs in the insurance industry is a challenging task. In various studies three commonly used outputs are: premium income, weighted sum of activities, and incurred benefits plus additions-to-reserves. In most of the previous studies net written premiums or net earned premiums have been used as proxies for outputs. Doherty (1981); Yuengert (1993); Cummins and Zi (1998) thinks premium cannot reflect the quantity of output. However, Houston and Simon (1970) thinks premium to insurers is what income to manufacturers, and therefore premium can be regarded as an appropriate indicator of output. Similar arguments appeared in: Praetz (1980); Fields and Murphy (1989); Grace and Timme (1992); Gardner and Grace (1993); Rai (1996); Diacon (2001); and Li (2005). They consider premium as an appropriate indicator of output, assuming the product is homogeneous and competitive pressures compel all insurers to charge the same price.

Finally, we have selected **two indicators of inputs as equity capital and labour (including commission, agents’ fee, referral and other expenditure)**; and **one indicator of output as net health insurance premium**, which is also in accordance with the assumption of DEA, that number of Decisions Making Units (DMUs) should be three time of number of inputs and outputs.

### Section B

#### 3.2 Data Base and Research Methodology for the Comparative Study of Health Insurance Plans Offered by General Insurance Companies in India.

For the purpose of comparative study of health insurance plans, the companies under its scope are same as that used in first objective. In other words, the health insurance plans of all the general insurance companies taken above were considered for comparative study. Moreover, the study is entirely based upon the use of secondary data, which is mainly drawn from the respective sites of the general insurance companies.
under consideration. The comparative study of health insurance plans was exclusively done with reference to inclusion/coverage and exclusion/non coverage under these. Beside this, the following variables and sub variables were forming a part of comparative study are shown in table 3.1.

**Table 3.1**

**Variables and Sub Variables Forming a Part of Comparative study**

<table>
<thead>
<tr>
<th>S.No</th>
<th>Variables</th>
<th>Sub variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What is total number of health insurance plans offered?</td>
<td></td>
</tr>
</tbody>
</table>
| 2    | What are inclusions/coverage/benefits under health insurance plans offered? | ✓ Medical expenses  
✓ Hospitalization coverage  
✓ Non hospitalization coverage/domiciliary hospitalization  
✓ Day care expenses  
✓ Ambulance charges  
✓ Compensation for death  
✓ Total/partial disablement/injury  
✓ Additional benefits/discounts etc. |
| 3    | What are exclusions/non coverage under health insurance plans offered?     | ✓ Pre-existing illness  
✓ Dental treatment  
✓ Disease within 30 days  
✓ Disease within 90 days  
✓ Disease within 120 days  
✓ HIV/AIDS  
✓ Suicide/Attempt to suicide  
✓ Natural calamities, war, perils |

As far as the analysis of data collected above is concerned, the use of content analysis was made. Beside this, the use of percentages, independent sample t-test and one-way ANOVA was made in order to draw the meaningful inference. This was done with the help of SPSS software package as well as use of MS-Excel 2007 was made.
3.2.1 Content Analysis: It is a research tool used to determine the presence of certain words or concepts within texts or sets of texts. Researchers quantify and analyze the presence, meanings and relationships of such words and concepts, then make inferences about the messages within the texts, the writer(s), the audience, and even the culture and time of which these are a part. To conduct a content analysis on any such text, the text is coded or broken down, into manageable categories on a variety of levels--word, word sense, phrase, sentence, or theme--and then examined using one of content analysis' basic methods: conceptual analysis or relational analysis conceptual analysis, a concept is chosen for examination and the number of its occurrences within the text recorded. Because terms may be implicit as well as explicit, it is important to clearly define implicit terms before the beginning of the counting process. Relational analysis, like conceptual analysis, begins with the act of identifying concepts present in a given text or set of texts. However, relational analysis seeks to go beyond presence by exploring the relationships between the concepts identified. In other words, the focus of relational analysis is to look for semantic, or meaningful, relationships. Individual concepts, in and of themselves, are viewed as having no inherent meaning. Source: Colorado State University site: (http://writing.colostate.edu/guides/research/content/pop2a.cfm assessed as on 17th April 2011).

3.2.2 Independent sample t-test: The Independent samples t-test procedure compares means for two groups of cases. For t-test, the observations should be independent, random samples from normal distributions with the same population variance.

3.2.3 One-way ANOVA: It is a technique used to compare the mean of more than two populations, such as comparing the yield of crop from several varieties of seeds, the gasoline mileage of four automobiles, the smoking habits of five groups of company’s employees and so on. In such circumstances one generally does not want to consider all possible combinations of two populations at a time for that would require great number of test before we would be able to arrive at decision. This would also consume lot of time and money, and even then certain relationships may be left unidentified. Therefore, one utilizes the ANOVA technique and through it, investigates the differences among the mean of all population simultaneously. In other word the basic principle of ANOVA is to
test the differences among the mean of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between the samples (Kothari, 2004, pp. 256-258).

Section C

3.3 Data Base and Research Methodology to Examine the Working and Performance of Community Health Insurance (CHI)

In order to achieve the above stated objective, the use of past studies undertaken by the various researchers in respect of Community Health Insurance (CHI) in India was made. Besides this, various studies conducted by International Labour Organization (ILO) have also provided a base for present study. According to one of the recent study conducted by ILO in 2009 has shown that more than 100 community health insurance schemes are operating in India. These schemes are initiated by Non-Governmental Organizations (NGOs), Community-Based Organization (CBO), Public Departments (PD), Health Providers (HP), Private Trusts (PT), Micro-Finance Institutions (MFIs), Public-Private Trusts (P-PT) and Trade Unions (TU). For the purpose of present study, we have taken 100 such schemes which are operating in India and the detailed list of all these is given in annexure II. The criteria behind the selection of schemes for the present study was availability of information regarding variable explaining and affecting their working and performance. All these variables are given in table 3.2.

Table 3.2

Variables Forming a Part of Study to Examine the Working and Performance of Community Health Insurance (CHI)

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✔ Working with or without the linkage of insurance companies;</td>
</tr>
<tr>
<td>2</td>
<td>✔ Working with owned fund or receiving assistance/subsidies;</td>
</tr>
<tr>
<td>3</td>
<td>✔ Area of operation or working;</td>
</tr>
<tr>
<td>4</td>
<td>✔ Scope/number of services provided; and</td>
</tr>
<tr>
<td>5</td>
<td>✔ Present and potential coverage by the CHI schemes</td>
</tr>
</tbody>
</table>
Hence on the basis of availability of information regarding above stated variables, the selection of CHI schemes were made for the purpose of present study. The information for the same is collected mainly from the respective sites of CHI providers as well as the use of various studies conducted by ILO and other researchers were also made.

As far as the analysis of data collected is concerned, the use of percentages, mean, standard deviation was made. Beside this, the use of cross-tabulation, chi-square, correlation and independent sample t-test has been made to draw the meaningful inference from the study. All this was done with the help of SPSS software package.

3.3.1 Cross-tabulation: A statistical technique that describes the two or more variables simultaneously and result in the table that reflect the joint distribution of two or more variables that have a limited number of categories or distinct value. In general, the margins of cross-tabulation show the same information as the frequency tables for each of the variables. Cross-tabulation tables are also called contingency tables and the cross-tabulation between two variables is also known as bivariate cross-tabulation. Because of two variables has been cross-classified, percentage can be computed either column wise, based on the column totals or row wise based on the row totals. The variable to be considered in row or column depends upon which will be considered as dependent or independent variable (Malhotra, 2007, pp. 278-279).

3.3.2 Correlation: It quantifies the degree of association between two variables or the strength of linear relationship between two variables and also indicates the direction of the relationship. The correlation coefficient, r, measure the strength of linear relationship. The value of r is between +1 and -1. The values of r close to +1 or -1 represent a strong linear relation. The value of r close to 0 means that the linear association is very weak. It could be that there is No association at all, or the relationship is non linear (Tyrrell, 2009, pp. 64).

3.3.3 Independent sample t-test: The Independent samples t-test procedure compares means for two groups of cases. For t-test, the observations should be independent, random samples from normal distributions with the same population variance.
3.3.4 Chi-square: The Chi-square statistics is used to test the statistical significance of the observed association in a cross-tabulation. It assists us in determining whether a systematic association exists between two variables. The null hypothesis $H_0$ is that there is no association between the variables. The test is conducted by computing the cell frequencies that would be accepted if no association were present between the variables, given the existing row and column totals. These expected cell frequencies, denoted $f_e$, are then compared to the actual observed frequencies $f_o$, found in the cross-tabulation to calculate the chi-square statistics. The greater the discrepancies between the expected and actual frequencies, the larger the value of the statistic. Assume the cross-tabulation has $r$ rows and $c$ columns and a random sample of $n$ observations. Then the expected frequency for each cell can be calculated by using a simple formula:

$$f_e = \frac{n_r n_c}{n}$$  

.............................................. (7)

Where, $n_r =$ total number in the row; $n_c =$ total number in the column; $n =$ total sample size.

Then the value of chi-square is calculated as follows:

$$\chi^2 = \sum_{\text{all cells}} \frac{(f_o - f_e)^2}{f_e}$$  

.............................................. (8)

To determine the systematic association exists, the probability of obtaining a value of chi-square as large as or larger than one calculated from the cross tabulation is estimated. An important characteristic of the chi-square statistics is the number of degrees of freedom (df) associated with it. In general, the number of degree of freedom is equal to the number of observations less than number of constraints needed to calculate a statistical term. In the case of chi-square statistic associated with a cross-tabulation, the number of degree of freedom is equal to the product of number of rows ($r$) less one and the number of columns ($c$) less one. That is, $df = (r - 1) \times (c - 1)$. The null hypothesis $H_0$ of number of association between the two variables will be rejected only when the calculated value of the test statistics is greater than the critical value of chi-square distribution with the appropriate degree of freedom (Malhotra, 2007, pp.504-507).
Table 3.3
Key Statistics Associated with Chi-square

| Phi coefficient | It is used to measure the strength of association in the special case of a table with two rows and two columns (a $2 \times 2$ table). It is calculated as:
| $\phi = \frac{\lambda^2}{\sqrt{\lambda^2 + n}}$ |
| It takes the value of 0 when there is no association and when the variables are perfectly associated, phi assumes the value of 1. |

| Contingency Coefficient | It is used to measure the strength of association in a table of any size. It is calculated as:
| $c = \frac{\lambda^2}{\sqrt{\lambda^2 + n}}$ |
| It varies between 0 and 1. The value of 0 occur in the case of no association (i.e. the variables are statistical independent), but the maximum value 1 is very rare come. |

| Lambda Coefficient | It is a measure of the percentage improvement in predicting the value of dependent variable, given the value of the independent variable in the contingency table analysis. Lambda also varies between 0 and 1. |


Section D

3.4 Data Base and Research Methodology to Examine the Role Played by Third Party Administrators (TPAs) with Reference to Role Defined by IRDA

In order to achieve the above stated objective, a sound theoretical framework was developed on the basis of studies already undertaken in this area. In other words, the achievement of above objective was driven mainly by review process. Beside this, the use of regulation of IRDA for TPAs was made in order to examine the conditions, code of conduct/role theoretically defined by it. Thereafter, the analysis of role in practice played by TPAs was made on the basis of information extraction from earlier studies and
from their respective sites. Further, evaluation was made, and that too theoretically, to examine the two fold aspects i.e. parity between the role defined by IRDA and role in practice played by TPAs; and deviation between role defined IRDA and role in practice played by TPAs.

Section E

3.5 Data Base and Research Methodology to Examine the Customers’ Perception Regarding Health Insurance

For the purpose of above stated objective, the specified area was selected on the assumption that specific area based studies expected to give more meaningful and significant information. Accordingly, for the examination of customers’ perception regarding health insurance, the present study was done in Punjab. Further, it was planned to give true representation to three belts of Punjab, viz., Majha, Doaba and Malwa. Hence, one district from each of three belts was selected. The districts included in sample were Amritsar from Majha, Jalandhar from Doaba and Ludhiana from Malwa. Thereafter, selection of sample of respondents was made by following random sampling and on the whole a sample size of 600 respondents was planned from the general public. The general public was mainly considered, as it is expected to include existing customers as well as potential customers. In the view of fact that in the present study general public has been considered as unit of investigation, a sample framework consisting of equal number of respondents from each of the district was taken. In other words, the questionnaire were got filled from 600 respondents (200 respondents from each of the district), out of which 563 was found to be suitable for the purpose of analysis.

The data has been collected from the general public by administering self-structured, pre-tested questionnaire from them. The preliminary draft of the questionnaire was pretested on 50 respondents. This helped in improving the questionnaire and also gave an indication as to kind of responses that would be forthcoming; accordingly with few addition and deletion, the final questionnaire (given in annexure I) was developed and used for collection of information from respondents. The questionnaire for data collection contains questions relating to awareness level, sources of awareness, subscribed or not subscribed health insurance, reasons for same, type of policy, which
factor motivate them to buy and obstruct them to buy, satisfaction level and services etc. The questions were open ended, dichotomous and offering multiple choice. Besides this, some questions required ranking or rating or alternatives.

The analysis of data collected above was carried out by using simple frequencies, multiple frequencies and percentages for multiple responses as well as Weighted Average Scores (WAS) has been calculated. Beside this, the use of factor analysis and chi-square was made to draw the meaningful inference from the study. All this was done with the help of SPSS software package.

3.5.1 **Weighted Average Scores (WAS):** In the present study, WAS calculated in order to find out the importance attached by respondents towards various reasons to opt for health insurance as well as for claim logged aspects.

3.5.2 **Chi-square:** The Chi-square statistics is used to test the statistical significance of the observed association in a cross-tabulation. The null hypothesis \( H_0 \) of number of association between the two variables will be rejected only when the calculated value of the test statistics is greater than the critical value of chi-square distribution with the appropriate degree of freedom (Malhotra, 2007, pp. 504-507).

3.5.3 **Factor Analysis:** Factor Analysis is a general name denoting a class of procedures primarily used for data reduction and summarization. In research, there may be large number of variables, most of which are correlated and which must be reduced to manageable level. Relationship among set of many interrelated variables are examined and represented with the help of factor analysis. The approach used in the factor analysis is “Principle Component Analysis”. In this component analysis, the total variance in the data is considered. The diagonal of the correlation matrix consists of unities and full variance is bought in to factor matrix. It determines the minimum number of factors that will account for maximum variance in the data for use in subsequent multivariate analysis. The factors are also called principal components. Although the initial or unrotated factor matrix indicates the relationship between the factors and individual variables, it seldom results in factors that can be interpreted, because the factors are correlated with many variables. Hence the variance explained by each factor is
redistributed by rotation. The method used for rotation in this study is “Varimax”. It is a method of factor rotation that minimizes the numbers of variables with high loading on a factor, thereby enhancing the interpretability of the factors. There are number of terms/statistics associated with the factor analysis which are described in table 3.4.

### Table 3.4
**Key Statistics Associated with Factor Analysis**

<table>
<thead>
<tr>
<th>Key statistics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communalities</td>
<td>It is the amount of variance a variable share with all other variable considered.</td>
</tr>
<tr>
<td>Eigen value</td>
<td>It represents the total variance explained by each factor.</td>
</tr>
<tr>
<td>Factor loading</td>
<td>It is the simple correlation between the variables and factors.</td>
</tr>
<tr>
<td>Factor matrix</td>
<td>It contains the factor loading of all the variables on all the factors extracted.</td>
</tr>
<tr>
<td>Factor score</td>
<td>These are composite scores estimated for each respondent on the derived factor.</td>
</tr>
<tr>
<td>Percentage of variance</td>
<td>This is the percentage of variance attributed to each factor.</td>
</tr>
<tr>
<td>Residual</td>
<td>It is the difference between the observed correlations, as given in input correlation matrix, and the reproduced correlation, as estimated from factor matrix.</td>
</tr>
<tr>
<td>Kaiser-Meyer-Olin</td>
<td>It is measure of sampling adequacy and is used to examine the appropriateness of the factor analysis.</td>
</tr>
<tr>
<td>Barlett’s Test</td>
<td>It is a test statistics used to examine the hypothesis that the variables are uncorrelated in the population.</td>
</tr>
</tbody>
</table>

(Source: Malhotra, 2007, pp. 639-649)

Before the application of factor analysis, reliability of the scale items has been checked with the help of Cronbach’s alpha, which assess the internal consistency of entire scale.
3.6 Limitations of the Study

Research being never ending process makes ground for further researchers. Obviously, all studies and researchers have their own limitations and this study is no exception as such despite its theoretical and practical relevance the study does suffer from limitations. These limitations are as:

- The study is confined to some selected general insurance companies in India, while the inclusion of other may differentiate and would provide more appropriate results.
- Though utmost care has been taken while selecting the variables of inputs and outputs for performance evaluation, still the inclusion of some other variables might influence the results.
- The approach of the study in relation Community Health Insurance (CHI) is narrow, as it is just based on the availability of information on certain variables with respect to these. Alternatively, the inclusion of other variables explaining and affecting their working and performance can be considered for better generalization of findings.
- The study is also based on primary data collected with the help of predesigned questionnaire suffer from basic limitation of lack of honest and true response. Moreover, the primary data is collected only from the major cities of Punjab. Hence, the study suffers from regional bias and results derived from primary survey may not be applicable to other part of the country.
- Further, an increased sample size and multi-city sampling can be considered for further research for better generalization of findings.