# Health Insurance in India- Data Availability and Applicability

By Rakesh Khandelwal & Bhavna Verma

## Abstract

The objective of **t**his paper is to share the experience of a study on a sample of Indian health data set. This paper has been divided into two sections. Section I of the paper reviews the health data available with the Insurance Regulatory and Development Authority (IRDA). It draws attention to the immediate need for requiring strict compliance with standardized reporting across insurers and third party administrators (TPAs). This section also provides practical data enhancement suggestions to improve the quality and scope of analysis. The applicability of the data has been discussed in Section II of the paper. The purpose of Section II is to demonstrate the kind of analysis that can be undertaken and its use in the overall management of the health portfolio of an insurance company. The figures in all the exhibits are replaced by dummy figures to ensure confidentiality. This paper concludes by emphasizing the importance of data analysis with regard to current regulatory changes and the need for creating industry-wide benchmarks.

## Introduction

The rapid development of the healthcare sector in India is presenting several challenges to the Indian insurers providing health insurance coverage. Time and again, the need for a systematic and standardized database for the health sector has been emphasized and certain steps have been taken in this direction.

The Ministry of Communication & Information Technology (MCIT) constituted a working group to prepare the ground for the Information Technology Infrastructure for Healthcare (ITIH) in India in coordination with Apollo Health Street Ltd in October, 2002. In January, 2004, the working group recommended standards to be followed for capturing and exchanging health information. The standards covered detailed formats for Healthcare identifiers, Data elements, Messaging standards, Clinical Terminology, Minimum Data Sets and Billing Formats. A committee formed by the Insurance Regulatory Development Authority (IRDA) made recommendations for data formats and data collections from TPAs. In the follow up, the data for previous years has been collected by the Tariff Advisory Committee (TAC) from TPAs.

Further, the advent of detariffication in the general insurance industry has increased the importance of data collection and data analysis especially in the health insurance industry. There is an additional need for more effective analysis in the new competitive regime of pricing without any cross subsidy across classes. Proper data analysis is also necessitated by new regulations of the IRDA such as IBNR (Incurred but not Reported) estimation and Product Filing Requirements. It is also important to create industry wide benchmarks to enable an insurer to compare its own performance and rates with industry standards.

In this paper, we discuss the current status of health insurance data, based on the evaluation of a sample dataset. We are sharing some of our general findings in two sections.

Section I discusses the format of the current health data. This section further assesses the quality of data and explains the variability of data across different TPAs, with suggestions for data improvement.

Section II discusses the applicability of data by demonstrating the kind of analysis that can be undertaken in the context of overall management of an insurance company by replacing the actual figures with dummy figures to ensure confidentiality.

## Section I – Availability of Health Data

This section discusses prescribed health data collection formats. It further compares the completeness, accuracy and use of standardized codes in the data across different TPAs. It ends by recommending steps to improve the data quality.

### I.1. Current formats of health insurance data collection

In follow up to the recommendations of working group constituted by MCIT, the Insurance Regulatory Development Authority (IRDA) had constituted a Sub-Committee with the objective of drawing up a road map for establishing a data repository under the custodianship of the Tariff Advisory Committee (TAC). The TAC had collected three years of data from the insurers and TPAs for the purpose of evaluating the adequacy of data elements. Briefly, the following recommendations were made by the committee:

- Formats for collecting existing data as recommended by the sub-committee to be adopted for the years 2003/04, 2004/05 and 2005/06.
- The existing data would be collected from the TPAs first within a specified time schedule.
- Training of personnel from the TPAs, Insurers and TAC for using ICD classification with the help of the Government of India.
- The committee recommended that the Tariff Advisory Committee publish aggregate information regarding claims.

Data for the years 2003/04 and 2004/05 has been received from the TPAs and is under consideration. Data provided to the TAC can be described as follows:

- **Policy Data (Table A):**
  The policy level information is contained in this table. It has details like the total number of people covered under a policy, policy premium, start date and end date of policy.

- **Member data (Table B):**
  This table contains information about the individual members covered under the policy. The details include the age or date of birth of member, sum insured, gender and relationship with the insured.

- **Claims Data (Table C):**
  Information on claims made by the covered members is contained in this table. The details include the date of admission and discharge, diagnosis description and code, name of the hospital, amount paid as well as claimed and the date of payment.

- **Outstanding Claims Data (Table D):**
  This table shows the outstanding claim amount at the beginning and close of each financial year as well as the total amount paid during the year in aggregate.

We have limited our analysis to data contained in Tables A, B and C only The chart below shows the list of important fields contained in these tables.

| Table A | Table B | Table C |
|---|---|---|
| TPA Code | TPA Code | TPA Code |
| Insurer Code | Insurer Code | Insurer Code |
| U/W Office Code | Policy Number | Member Reference Key |
| Policy Number | Member Reference Key | Claim Number |
| Start Date | Date of Birth | Diagnonis Description |
| End Date | Age of Insured | Diagnosis Code |
| Product Type | Sex | Procedure Description |
| Type of Cover | Occupation | Name of the Hospital |
| Group Size | Relationship of Insured | Date of Admission |
| Policy Premium | Sum Insured | Date of Discharge |
| | Pre-existing Disease Cover | Total Amount Claimed |
| | Baby cover as part of Materni | Room & Nursing Charges |
| | Maternity Cover | Surgery Charges |
| | Floater applicable | Consultation Charges |
| | Floater amount | Investigation Charges |
| | | Medicine Charges |
| | | Miscellaneous Charges |
| | | Total Claim Paid |

## I.2. Comparative TPA analysis

There are no directives from the IRDA to establish standardized data requirements or IT systems. Therefore, the TPAs manage the policyholders details through the readily available tailor-made softwares for such purposes.. The data provided by different TPAs showed significant variation in terms of quality and consistency. In this section, we provide a descriptive account of the nature of the irregularities that we found in the health insurance dataset.

The following is an analysis of the compatibility of data sets from some TPAs and conformity with prescribed formats.

### 2.1 Data Completeness:

Completeness of data is pre-requisite for effective analysis. The variables provided in the data set should be sufficiently populated for accurate analysis. Any error or incompleteness in the data would lead to inaccurate results. There were few blank records in several data fields provided. However, major inadequacies in the data are discussed below:

1) **ICD Codes**:
ICD (International Classification of Diseases) is a standard disease based code prescribed by WHO (World Health Organization) and used across the world. A unique three digit ICD-10 code is assigned to each of the diagnosis groupings of the claimant. The claims data is grouped by these codes for analysis purposes. This coding was largely unavailable or inaccurate in many cases. The lack of prescribed ICD-10 Codes was a major shortcoming in the data. Some TPAs provided broad level code ranges rather than a unique three digit ICD-10 code.

2) **Diagnosis and Procedure Descriptions**:
Diagnosis descriptions are useful to verify the ICD codes assigned to the claims and to populate the ICD codes, in case it is unavailable. The procedure description is used to analyze whether it is medically appropriate for the diagnosis. There was a high degree of variability in the diagnosis and procedure descriptions provided by different TPAs. Some of the inconsistencies observed were:

- Diagnosis descriptions are restricted to a text limit of 20 characters as per IRDA formats. Data in the field was difficult to comprehend due to syntax errors or broad descriptions such as “Conservative Surgery”.
- One TPA provided an additional column titled “OPINION” containing combined descriptive and procedural information for some records.
- Most TPAs did not provide procedure descriptions at all, while one recorded drug names. Occasionally, diagnosis and procedure descriptions were provided interchangeably by some TPAs.

3) **Age/Date of Birth**:
The likelihood of a claim, and therefore the required premium, is affected by the age of insured. The lack of accurate date of birth or age of insured undermines the ability to do any meaningful analysis based on age. Some of the inconsistencies observed in this regard were:

- The date of birth of insured was not well-populated by most TPAs. Sometimes, the dates did not seem to be reliable as they went as back as 1900, 1901 etc.
- Some of the TPAs displayed a higher level of inaccuracy in recording ages. A substantial number of records provided age 0, which might include cases where the age was unavailable. Some records contained incorrect ages, which were either negative or over 150.

4) **Occupation**:
The occupation information helps in better understanding of the risk profile of the insured. The field for “Occupation” was blank in most cases. It was provided by very few TPAs, and for few records. It was observed that:

- Some TPAs provided broad occupational categories such as service, business and others without any further description. The field had several spelling errors.
- One TPA provided number codes in the few records that were populated.

**2.2 Data Accuracy:**

It is of utmost importance that the data be largely accurate. A high volume of inaccurate data undermines the reliability of any analysis. There were several shortcomings in the data with respect to data accuracy. Steps were taken to validate the data and to identify and rectify the inaccurate data.

1) **Reasonability Checks:**
Reasonability checks are the most important step to understand and demonstrate data quality. They are very useful from data analysis perspective as they give an early indication of which analyses can be viably done. Frequently they also indicate which other elements of data pre-processing be included in initial data enhancement efforts. This includes, but is not limited to, data cleaning, re-categorization, sorting, grouping, and calculating new variables by utilizing existing variables and reformatting. The following reasonability checks were undertaken on the data:

- Negative policy period; i.e. end date of a policy before its start date
- Length of stay less than zero; i.e. date of admission after the date of discharge
- Date of admission before the start date of its policy
- Date of admission after the end date of its policy
- Claim paid more than the corresponding amount claimed
- Inconsistency in age of Insured
- Policy premium less than zero

2) **Duplicate records:**
Duplicate records were found in both Policy table A and Member table B. This duplication leads to multiple records when an attempt is made to create a master exposure dataset containing policy as well as member details. This limitation made it difficult to evaluate the true risk profile of the lives covered under a policy. Further, there is a lack of a standard format to assign a unique policy number to each policy record. There were similar format issues with member records as well. Some TPAs used same policy numbers for group as well as individual policies.

3) **Missing policy data:**
Policy data was missing for certain claims paid. The percentage of missing records was between 30-50% for some TPAs. Such a gap undermines the entire purpose of the loss ratio analysis by different parameters.

4) **Segregation of expenses into different benefits:**
The claimed amount needs to be segregated into different categories including Room & Nursing, Surgery, Investigation, Medicine and Miscellaneous charges as per the prescribed format. However, the figures for Miscellaneous charges were unusually high for some TPAs. It might be due to the fact that some of the TPAs do not classify the total claimed amount into the various sub-categories efficiently. Such classification, if provided, can facilitate analysis of the payment patterns for different benefits.

### 2.3 Coding and Standardization:

There were inconsistency and variation among TPAs for coding certain fields, though IRDA has prescribed standard codes for some such fields. This inconsistency lead to problem while creating a single dataset for all TPAs. Some such fields are discussed below:

1) **Gender:**
A number coding is specified for the gender, but there was absence of a standard coding mechanism across TPAs. Sometimes, TPAs used incomprehensible codes such as s, w etc.

2) **Relationship of Insured**:
Standard codes were not utilized by several TPAs. There were spelling errors in simple terms such as proposer, husband, daughter, etc.

3) **Hospital and City Names:**
No standard coding is specified for these names but the expectation was that the name of any city or hospital should be entered correctly and consistently. There were serious inadequacies in this information. Spelling errors and data entry errors were rampant, resulting in the same item being spelled in multiple ways. This hampered grouping and analyzing data by provider or location.

4) **Hospital Identifier:**
Certain identifiers such as hospital registration number, PAN number and pin code of hospital are provided for in the data formats. Proper use of these prescribed fields can help deal with the issue discussed above, but most TPAs did not populate these fields.

5) **Type of cover – Group/individual:**
This field is not uniquely defined by all TPAs as prescribed. Different TPAs use individual coding mechanisms. Sometimes, the group size of policy had to be used to define the type of cover. There were cases where the group Size of 1000 had been marked as an Individual Policy. This type of categorization problem results in distortions for any analyses by this category.

### 2.4 Level of Aggregation and Compliance with IRDA standards:

Our study was limited to variables that were reasonable and sufficient for data analysis. Some of the prescribed fields were hardly populated though provided for in the TPAs' data system. Some of these fields are important from the point of view of data analysis and require more attention. These are listed below:

1) **Date of intimation to TPA**:
   This field can be used to analyze the processing period of a claim from the date of intimation to TPA to the date of actual payment of claim. Further this field can also be used to estimate the number of IBNR claims for the purpose of calculating the outstanding claim amount.

2) **New / renewal status**:
   This field can be used to calculate the rate of renewal of policies as well as the lapse rate of policies for various insurers as well as for various policy sizes. It can be further extended to renewals & lapse by region, by sum insured, policy type etc.

3) **Pre-hospitalization and post-hospitalization expenses:**
   The breakup of claim amount by pre-hospital and post-hospital expenses would help in better categorization of claim amount.

4) **Date of issue of card:** This field can be used to assess one of the efficiency parameters of the service provided to the insured by the TPA, by comparing it with the start date of the policy.

5) **Maternity Cover:**
   This can be used to calculate the expected number of maternity claims and the average cost of expected maternity claims.

6) **Pre-Existing Disease:**
   This is an important field for risk assessment and utilization analysis.

The above discussion highlights the fact that any actuarial analyses undertaken prior to a significant improvement in the data quality will be seriously compromised. Undoubtedly, the lack of adequate and accurate data must be the foremost concern for health insurers in the near future.

## I.3. Data Recommendations

In the following section, we discuss some simple solutions to the data inadequacies discussed above. Most of these can be easily rectified by the use of simple validation checks adopted by the TPAs and minor technological enhancements to their data reporting systems.

- **Data Completeness**:
  Inadequate data is useless data in most cases. To ensure that the data is complete and accurate, certain validation features need to be introduced into the TPAs' front end business application. The data must be submitted in a standardized format with mandatory fields completely filled in. Insurance companies can design a training program for all TPAs and assist the TPAs in conducting this training. This program should highlight the benefits of complete data as well.

- **Data Accuracy:**
  As far as possible, the TPAs must use drop-down menus containing only the prescribed options for entering data in different columns. Validation features must be inbuilt in the system to avoid errors discussed under "Reasonability Checks" previously, such as:
  - End Date of Policy must be after the Start Date.
  - Date of Admission must lie between the Policy Start and End Dates.
  - Date of Discharge must be after the Date of Admission.

- **Data Standardization**:
  To ensure that the data is consistent across a spectrum of data sources**,** a data field standardization document such as one recommended by working group constituted by MCIT, should be adopted by all insurers and TPAs at the earliest. In addition, companies must guide the various TPAs in achieving this standardization. This may entail conducting an audit of the TPAs current IT system, developing a gap analysis document and assisting the TPA in developing software upgradation strategy.

- **Data Efficiency:**
  It was observed that duplicate records existed for all TPAs in Policy Table A and Member Table B. It is our opinion that if a standard protocol for linking edited records with the original record were followed, the number of duplicate records would become negligible. Further it must be noted that the lack of indicative change in the policy number on renewals led to duplication of records. A unique numeric reference key for each policy (new or renewed) in Table A will ensure that data is assessed accurately. This unique key can be used subsequently in Tables B and C so as to correctly map the data.

- **Master Tables:**
  A consolidated member table with premium information after combining Table A (Policy data) and Table B (Member details) would result in better data analysis by reducing data duplication. This would enable some additional analysis such as profitability by age.

Overall, it was observed that there is wide variation in the quality of data maintained by different TPAs. There is a need to take steps to standardize the process of data coding across the industry. Further, steps should be taken to improve the degree of compliance with the fields marked as not fully populated as of now. The availability of complete and accurate data is also necessitated by the fact that due to insufficient mortality and morbidity data in India, the past experience of insurers gains utmost importance for rating purposes.

## Section II – Applicability of Health Data and Health Data Analysis

An Insurance company needs to monitor its results on a regular basis to revise its strategy for risk management and in reassessing the risk that it faces. It needs to perform the following important but not exhaustive tasks to fulfill this objective:

- Performance Monitoring and Risk Assessment
- Financial Reporting, Operations and Management
- Provider Network Analysis
- Premium Rating

A complete data set in a consistent format including all the important fields will allow doing all the above mentioned tasks.

We have undertaken a data analysis exercise on the health data supplied by an Indian Insurer. The actual figures in all the exhibits are replaced by dummy figures to ensure data confidentiality. Further, the analysis has been done on calendar year basis. The purpose of this section is to explain the kind of analysis that can be undertaken and the corresponding exhibits prepared to support the tasks listed above.

### II.1. Performance monitoring and Risk Assessment

The actual results have been analyzed by various variables like type of policy (Group or Individual), number of people covered, age of the persons covered, average claim size, length of stay in hospital, type of diagnosis etc. Analyzing the data in this way highlighted some interesting features like:

- the average use and cost of the services rendered,
- the probability distribution of aggregate claims by claim paid and
- the relative profitability of various sources of business.

This analysis also highlighted the sources from where the better quality, longer-lasting business comes from. This helps the insurer in better planning to devise schemes to retain or attract the profitable business and restructure or avoid the less profitable sources of business.

### 1.1 Premium Analysis:

The analysis of premium data is the first step in data analysis. This highlights the source and the total amount of policy premium written and the corresponding number of people covered. The Earned Premium (EP) is, then, calculated using $365^{th}$ Method. The EP calculation is done using the exact start date and end date of policy. The earned premium for a policy written during the reporting period can be calculated as follows:

$$\text{Earned Premium} = \frac{(\text{End Date of reporting period} - \text{Start date of Policy} + 1)}{365} \times \text{Written Premium}$$

The total earned premium during the year can be computed as the sum of unearned premium at the beginning of the reporting period and the total earned premium on the policies written during the period. The reporting period can be financial year starting at 1st April each year or calendar year.
The exhibit below shows the distribution of written premium into corresponding earned premium and unearned premium and the number of policies written and the corresponding people covered for each Policy year.

**Exhibit 1.1 – Written Premium Analysis**

| Policy Type | Policy Year | Number of policies | Number of people | Policy Premium | Earned Premium | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 2003 | 2004 | 2005 | 2006 |
| **Group** | **2003** | 1,000 | 50,000 | 100,000,000 | 60,000,000 | 40,000,000 | | |
| | **2004** | 1,500 | 100,000 | 150,000,000 | | 90,000,000 | 60,000,000 | |
| | **2005** | 2,000 | 150,000 | 270,000,000 | | | 162,000,000 | 108,000,000 |
| **Individual** | **2003** | 100,000 | 150,000 | 225,000,000 | 135,000,000 | 90,000,000 | | |
| | **2004** | 150,000 | 200,000 | 320,000,000 | | 192,000,000 | 128,000,000 | |
| | **2005** | 200,000 | 250,000 | 425,000,000 | | | 255,000,000 | 170,000,000 |
| **Grand Total** | | **454,500** | **900,000** | **1,490,000,000** | **195,000,000** | **412,000,000** | **605,000,000** | **278,000,000** |

The earned premium thus calculated corresponds exactly with the Incurred year analysis and helps in doing the accurate loss ratio analysis. Another advantage of this analysis is the calculation of Unearned Premium reserve (UPR) at the end of a reporting period, which needs to be shown in the annual financial statements. The UPR on a policy in force at the end of a reporting period can be calculated using the formula:

$$\text{Unearned Premium} = \frac{(\text{End Date of Policy} - \text{End Date of reporting period})}{365} \times \text{Written Premium}$$

## 1.2 Age distribution of people covered:

The total number of people covered is categorized into different age bands for both group and individual policies. The exhibit below also reflects the percentage change in the number of people covered in 2005 as compared to the corresponding number of people in 2004.

**Exhibit 1.2 – Age distribution of people covered**

| Policy Type | AgeBand | 2005 | | 2004 | | |
|---|---|---|---|---|---|---|
| | | No. of People | % of Total People | No. of People | % of Total People | % change in 2005 over 2004 |
| **Group** | **Zero** | 4,500 | 1.1% | 2,000 | 0.7% | 125.00% |
| | **1-20** | 30,000 | 7.5% | 18,000 | 6.0% | 66.67% |
| | **21-35** | 67,500 | 16.9% | 41,000 | 13.7% | 64.63% |
| | **36-45** | 15,000 | 3.8% | 10,000 | 3.3% | 50.00% |
| | **46-55** | 10,500 | 2.6% | 13,000 | 4.3% | -19.23% |
| | **56-60** | 7,500 | 1.9% | 5,000 | 1.7% | 50.00% |
| | **61-70** | 9,000 | 2.3% | 4,800 | 1.6% | 87.50% |
| | **71-75** | 3,000 | 0.8% | 1,700 | 0.6% | 76.47% |
| | **76-90** | 2,850 | 0.7% | 4,300 | 1.4% | -33.72% |
| | **> 90** | 150 | 0.0% | 200 | 0.1% | -25.00% |
| **Group Total** | | **150,000** | **37.5%** | **100,000** | **33.3%** | **50.00%** |
| **Individual** | **Zero** | 2,500 | 0.6% | 1,500 | 0.5% | 66.67% |
| | **1-20** | 70,000 | 17.5% | 52,000 | 17.3% | 34.62% |
| | **21-35** | 77,500 | 19.4% | 58,000 | 19.3% | 33.62% |
| | **36-45** | 50,000 | 12.5% | 38,000 | 12.7% | 31.58% |
| | **46-55** | 27,500 | 6.9% | 28,000 | 9.3% | -1.79% |
| | **56-60** | 10,000 | 2.5% | 8,000 | 2.7% | 25.00% |
| | **61-70** | 6,250 | 1.6% | 6,000 | 2.0% | 4.17% |
| | **71-75** | 2,000 | 0.5% | 1,800 | 0.6% | 11.11% |
| | **76-90** | 3,750 | 0.9% | 6,100 | 2.0% | -38.52% |
| | **> 90** | 500 | 0.1% | 600 | 0.2% | -16.67% |
| **Individual Total** | | **250,000** | **62.5%** | **200,000** | **66.7%** | **25.00%** |
| **Grand Total** | | **400,000** | **100.0%** | **300,000** | **100.0%** | **33.33%** |

This analysis is important to understand the age profile of all the people covered by the insurer. It helps in identifying the age bands where the number of people are growing and contracting. This analysis combined with corresponding claim data helps in calculating the average claim cost for each age band. Our analysis indicated that the age band 21-35 constitutes the most populated category under both the policy types.

### 1.3 Group size distribution of Policies written:

The policy level data is categorized by the number of people covered by the policy. The exhibit below shows the total written premium for different Group size bands and the average premium per person for each band.

## Exhibit 1.3 – Group size distribution of people covered

| Policy Type | Group Size | 2005 | | | | 2004 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Policy Premium | No. of People | Average premium per person | % of Total People | Policy Premium | No. of People | Average premium per person | % of Total People |
| Group | 0 - 19 | 5,400,000 | 750 | 7,200 | 0.2% | 1,950,000 | 1,500 | 1,300 | 0.5% |
| | 20-49 | 4,860,000 | 3,000 | 1,620 | 0.8% | 3,000,000 | 1,800 | 1,667 | 0.6% |
| | 50-99 | 6,750,000 | 4,500 | 1,500 | 1.1% | 5,250,000 | 2,800 | 1,875 | 0.9% |
| | 100-499 | 40,500,000 | 15,000 | 2,700 | 3.8% | 30,000,000 | 12,000 | 2,500 | 4.0% |
| | 500-999 | 27,000,000 | 10,500 | 2,571 | 2.6% | 13,500,000 | 8,000 | 1,688 | 2.7% |
| | 1000-4999 | 81,000,000 | 37,500 | 2,160 | 9.4% | 54,000,000 | 30,000 | 1,800 | 10.0% |
| | 5000-9999 | 36,990,000 | 26,250 | 1,409 | 6.6% | 16,800,000 | 18,900 | 889 | 6.3% |
| | > 10000 | 67,500,000 | 52,500 | 1,286 | 13.1% | 25,500,000 | 25,000 | 1,020 | 8.3% |
| **Group Total** | | **270,000,000** | **150,000** | **1,800** | **37.5%** | **150,000,000** | **100,000** | **1,500** | **33.3%** |
| **Individual** | | 425,000,000 | 250,000 | 1,700 | 62.5% | **320,000,000** | **200,000** | **1,600** | **66.7%** |
| **Grand Total** | | **695,000,000** | **400,000** | **1,738** | **100.0%** | **470,000,000** | **300,000** | **1,567** | **100.0%** |

This analysis helps in identifying the percentage share of Large groups, Small groups and Individual policies in the entire portfolio to enable the insurer to better understand the profile of the people covered. This analysis combined with the corresponding claims data can be used to analyze the profitability of different group sizes and in calculating renewal premium for the group policies. Our analysis indicated that the average premium per person decrease with the increase in group size and there is wide variation in the average premium per person for different sizes. Further, the average premium per person for group policy is lower as compared to the individual policy

### 1.4 Claims Paid by Incurred year:

The claims paid data for each Incurred Year is analyzed by the number of claims, the total amount paid and the amount paid for different benefits, including room nursing, surgery, miscellaneous etc. This exhibit is a basic claim exhibit to understand about the total amount of claims paid and the corresponding number of claims for different years.

## Exhibit 1.4 – Claims Paid Analysis by Incurred Year

| Claim Year | No. of claims | Total Claims Paid | Total Amount claimed | Room Nursing charges | Surgery charges | Consultation charges | Investigation charges | Medicine charges | Misc. charges | Average Claim Paid |
|---|---|---|---|---|---|---|---|---|---|---|
| **2003** | 60,000 | 400,000,000 | 500,000,000 | 80,000,000 | 40,000,000 | 40,000,000 | 80,000,000 | 80,000,000 | 80,000,000 | 6,667 |
| **2004** | 75,000 | 540,000,000 | 675,000,000 | 108,000,000 | 54,000,000 | 54,000,000 | 108,000,000 | 108,000,000 | 108,000,000 | 7,200 |
| **2005** | 100,000 | 665,000,000 | 831,250,000 | 133,000,000 | 66,500,000 | 66,500,000 | 133,000,000 | 133,000,000 | 133,000,000 | 6,650 |
| **Grand Total** | **235,000** | **1,605,000,000** | **2,006,250,000** | **321,000,000** | **160,500,000** | **160,500,000** | **321,000,000** | **321,000,000** | **321,000,000** | **6,830** |

This analysis is the first step in the claims data analysis. All the subsequent claims analysis use the result of this analysis to check their results. Our analysis indicated that the claims paid amount is not categorized appropriately into different benefit categories and the amount paid is categorized into miscellaneous charges at many places.

## 1.5 Claims paid Distribution:

The claims paid data for each incurred year is analyzed further to study the number of claims paid for different claims paid ranges and the corresponding proportion to the total claim amount paid. This helps to calculate the expected frequency and severity of claim amount in different claim ranges.

**Exhibit 1.5 – Claims paid Distribution**

| | 2005 | | | 2004 | | |
|---|---|---|---|---|---|---|
| Claim Paid Range | No. of Claims | Total Claims Paid | % of Total Claim Paid | No. of Claims | Total Claim Paid | % of Total Claim Paid |
| **0-1000** | 5,000 | 3,325,000 | 0.5% | 3,750 | 2,700,000 | 0.5% |
| **1000-4999** | 23,000 | 26,600,000 | 4.0% | 17,250 | 21,600,000 | 4.0% |
| **5000-9999** | 19,000 | 46,550,000 | 7.0% | 14,250 | 37,800,000 | 7.0% |
| **10000-24999** | 20,000 | 156,275,000 | 23.5% | 15,000 | 126,900,000 | 23.5% |
| **25000-49999** | 14,000 | 179,550,000 | 27.0% | 10,500 | 145,800,000 | 27.0% |
| **50000-99999** | 6,000 | 79,800,000 | 12.0% | 4,500 | 64,800,000 | 12.0% |
| **100000-199999** | 4,000 | 33,250,000 | 5.0% | 3,000 | 27,000,000 | 5.0% |
| **200000-299999** | 2,500 | 33,250,000 | 5.0% | 1,875 | 27,000,000 | 5.0% |
| **300000-399999** | 3,000 | 39,900,000 | 6.0% | 2,250 | 32,400,000 | 6.0% |
| **400000-499999** | 2,000 | 26,600,000 | 4.0% | 1,500 | 21,600,000 | 4.0% |
| **500000-999999** | 1,000 | 33,250,000 | 5.0% | 750 | 27,000,000 | 5.0% |
| **> 1000000** | 500 | 6,650,000 | 1.0% | 375 | 5,400,000 | 1.0% |
| **Grand Total** | **100,000** | **665,000,000** | **100.0%** | **75,000** | **540,000,000** | **100.0%** |

A comparative analysis shows the trend in the claim cost during the periods under study. The trend in claim amount depicted by this analysis can be used to calculate the composite trend factor which can be used in Premium rating. Our analysis indicated that the claim paid range “10000 – 50000” accounted for more than 50 percent of the total claim paid amount.

## 1.6 Claims analysis by Age bands:

The claims data is categorized to identify the age profile of the claimants. The claims data is tabulated to identify the number of claims and the amount paid for different age bands. This also helps in calculating the expected frequency and severity of claims from different age bands. As the premium rate changes with the change in the age of insured, this analysis helps in giving an indication of likely claim cost for different age bands. This analysis after normalizing the data for difference in sum insured can be used to assess the profitability of insured in different age bands.

## Exhibit 1.6 – Claims Paid Analysis by Age Bands

| Age Band | 2005 | | | | 2004 | | | |
|---|---|---|---|---|---|---|---|---|
| | No. of Claims | Amount Paid | Average claim paid | % of Total Claims | No. of Claims | Amount Paid | Average Claim Paid | % of Total Claims |
| **Zero** | 2,500 | 9,975,000 | 3,990 | 1.5% | 1,425 | 7,020,000 | 4,926 | 1.3% |
| **1-20** | 20,000 | 79,800,000 | 3,990 | 12.0% | 12,000 | 54,000,000 | 4,500 | 10.0% |
| **21-35** | 26,000 | 139,650,000 | 5,371 | 21.0% | 18,750 | 86,400,000 | 4,608 | 16.0% |
| **36-45** | 13,000 | 99,750,000 | 7,673 | 15.0% | 12,000 | 75,600,000 | 6,300 | 14.0% |
| **46-55** | 11,000 | 93,100,000 | 8,464 | 14.0% | 9,750 | 91,800,000 | 9,415 | 17.0% |
| **56-60** | 7,000 | 53,200,000 | 7,600 | 8.0% | 4,875 | 37,800,000 | 7,754 | 7.0% |
| **61-70** | 10,000 | 99,750,000 | 9,975 | 15.0% | 7,500 | 86,400,000 | 11,520 | 16.0% |
| **71-75** | 4,500 | 33,250,000 | 7,389 | 5.0% | 3,000 | 32,400,000 | 10,800 | 6.0% |
| **76-90** | 3,900 | 36,575,000 | 9,378 | 5.5% | 4,125 | 46,980,000 | 11,389 | 8.7% |
| **> 90** | 2,100 | 19,950,000 | 9,500 | 3.0% | 1,575 | 21,600,000 | 13,714 | 4.0% |
| **Grand Total** | **100,000** | **665,000,000** | **6,650** | **100.0%** | **75,000** | **540,000,000** | **7,200** | **100.0%** |

One of the difficulties faced during the analysis by age band is the absence of information to calculate the premium per member. Since the policy premium varies by age and the way data is entered into the system, the premium information is available only at the policy level (Table A). Where a group policy is issued, it is difficult to calculate the correct premium for each member covered in the policy. In the absence of premium information per member, it is not possible to calculate the correct loss ratio for different age bands.

One of the possible solutions to this problem might be to calculate the total number of people covered and the corresponding number of claims made in each age band. This can then be used to calculate the average claim cost per person covered for each age band. Since the premium rates are determined by the age bands, the premium rates for each age band can be used in calculating the estimated loss ratio analysis. But this loss ratio analysis is affected by the variation in sum insured by the member in the same age bands.

Another solution could be to add a premium index in the member table (Table B), based on the combination of the age band and the sum insured. This index can be added across all the members in the same age band to get the cumulative index. The estimated total premium can, then, be calculated using the premium rate manual. The premium, thus, calculated can be used in the loss ratio analysis by age bands.

### 1.7 Claims Analysis by Benefit Category:

The claims paid have been categorized into different benefit utilization category. This categorization has been made using the diagnosis group of each claim and putting into the appropriate category.

## Exhibit 1.7 – Claims Paid Analysis by Benefit Category

### LOS Summary by Benefit Category

**Product = ALL** **Post Date = 03/27/2006**
**Discharges from 200501to 200512**

| Inpatient HCG Category | No. of Claims | No. of Days | Amount Claimed | Claims Paid | Paid/ Claim | Paid as % of Claimed | Average LOS |
|---|---|---|---|---|---|---|---|
| Medical | 32,500 | 78,000 | 210,000,000 | 148,500,000 | 4,569 | 70.7% | 2.40 |
| Surgical | 26,000 | 97,500 | 420,000,000 | 330,000,000 | 12,692 | 78.6% | 3.75 |
| Maternity | 6,500 | 19,500 | 70,000,000 | 55,000,000 | 8,462 | 78.6% | 3.00 |
| Grand Total: | 65,000 | 195,000 | 700,000,000 | 533,500,000 | 8,208 | 76.2% | 3.00 |

This analysis helps in identifying the differences in the cost for various types of hospital cases. This analysis when compared with the past year claim data will show the trend in the mix of different benefits which is a factor in the overall composite trend factor.

## II.2. Financial Reporting, Operations and Management

An Insurance company needs to analyze financial experience, pull together all the necessary performance monitoring reports, and prepare forecasts of future operations including estimation of outstanding claim amount (IBNR). The dummy figures used under this part have no relevance with the figures used in earlier section 5.1.

### 2.1 Claims payment duration analysis:

A run-off triangle has been prepared exhibiting the cumulative number of claims paid by the end of each month since the incurred month. This helps in identifying the settlement period of claims. It can be undertaken by the type of claims, hospitals and cities to analyze further the reasons for settlement delay. The number of cumulative claims paid data has been tabulated by the duration of payment (in months) since claim occurrence. The triangle below shows the month of claim in the rows and the payment duration in months in the columns.

## Exhibit 2.1 – Number of Cumulative Claims Paid by settlement period

| No. of cumulative claims paid by the end of each month since the month of claim occurrence | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Month | Duration in months since the month of claim occurrence | | | | | | | | | | | | | | |
| Claim | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 200501 | 0 | 0 | 0 | 2 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 200502 | 0 | 0 | 3 | 8 | 10 | 10 | 10 | 11 | 11 | 13 | 13 | 14 | 15 | 15 | |
| 200503 | 0 | 14 | 27 | 30 | 34 | 34 | 34 | 34 | 35 | 35 | 38 | 38 | 38 | | |
| 200504 | 7 | 38 | 61 | 68 | 72 | 74 | 77 | 77 | 80 | 81 | 83 | 83 | | | |
| 200505 | 8 | 58 | 84 | 92 | 99 | 101 | 102 | 104 | 105 | 106 | 106 | | | | |
| 200506 | 8 | 60 | 91 | 116 | 128 | 133 | 135 | 136 | 140 | 140 | | | | | |
| 200507 | 0 | 50 | 120 | 136 | 151 | 159 | 160 | 166 | 166 | | | | | | |
| 200508 | 0 | 105 | 124 | 163 | 180 | 188 | 196 | 196 | | | | | | | |
| 200509 | 29 | 65 | 122 | 154 | 179 | 192 | 192 | | | | | | | | |
| 200510 | 0 | 38 | 133 | 168 | 191 | 191 | | | | | | | | | |
| 200511 | 1 | 29 | 113 | 145 | 145 | | | | | | | | | | |
| 200512 | 0 | 40 | 98 | 98 | | | | | | | | | | | |

There was variation among different TPAs about the duration of payment. Further, there were cases where the claim payment date was before the date of admission in hospital. But those cases were few and the total data was enough to do a meaningful analysis after clubbing all those cases as payment within the month of occurrence itself. The average payment duration during calendar year 2004 varies from 3.1 – 4.6 months.

## 2.2 IBNR Analysis:

One task that is core to all of the experience analysis is the calculation of the claim reserve often referred to as Incurred but not reported (IBNR) claims. IBNR computation is important as it needs to be shown in the annual financial reports. Further, for a group which has claims experience of, say less than 1 year and is open to renewal, an estimate of outstanding claim amount needs to be made before calculating the renewal rates. Further, the paid experience to date needs to be added with the outstanding claim amount before calculating the risk premium for premium rating..

IBNR analysis is done using a methodology based on completion factors and PMPM (per member per months) projection.

Completion Factors - The completion factor is the ratio of incurred and paid claims to estimated incurred claims. Incurred claims estimates are developed by dividing the incurred and paid claims to date for a given month by the estimated completion factor for that month. Completion factors are estimated by analyzing the completion factors for previous incurred months at the same duration (i.e., same number of months of runout from the incurred date) as the month for which the factor is being developed. Various averaging methods are used to develop these estimates.

Estimates are made on a monthly basis (i.e., either the completion factor or the claims cost PMPM is estimated for each incurred month in the lag table). Typically, completion factors are used for incurred months with enough runout so that the completion factors are credible. PMPM projection is used for the most recent incurred months where there is insufficient runout to develop meaningful completion factors.

The starting point for all of this analysis is the lag table or triangle. All the exhibits covered in this section shows the data and the result for the 12 months ended December, 2005, however, the IBNR analysis has been done using the 36 months data starting from January, 2005. The exhibit below shows the incremental lag triangle for the 12 months ending December, 2005. The claims paid data for each incurred months is plotted by the month of payment. The month of payment is shown in rows with the month of claim in the columns.

## Exhibit 2.2.A – Incremental Payments by Incurred Month

**Outstanding Claims (IBNR) Analysis**

Amt in '000

| Paid Month | Incremental Payments for the months of claim occurrence (Incurred month) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Jan-05 | Feb-05 | Mar-05 | Apr-05 | May-05 | Jun-05 | Jul-05 | Aug-05 | Sep-05 | Oct-05 | Nov-05 | Dec-05 |
| Jan-05 | 2,113 | | | | | | | | | | | |
| Feb-05 | 21,451 | 3,466 | | | | | | | | | | |
| Mar-05 | 21,186 | 23,024 | 5,096 | | | | | | | | | |
| Apr-05 | 11,261 | 16,062 | 24,019 | 4,860 | **2,000** | | | | | | | |
| May-05 | 6,666 | 14,985 | 23,827 | 26,239 | 3,477 | **1,000** | **1,200** | | | | | |
| Jun-05 | 3,880 | 7,526 | 13,601 | 25,376 | 27,439 | 3,960 | **800** | | | | | |
| Jul-05 | 2,766 | 4,575 | 7,695 | 12,904 | 25,351 | 24,652 | 1,169 | | | | | |
| Aug-05 | 1,692 | 2,321 | 3,962 | 8,631 | 13,718 | 27,485 | 29,472 | 3,910 | **1,500** | | | |
| Sep-05 | 937 | 1,417 | 2,648 | 5,035 | 9,196 | 14,799 | 28,716 | 32,984 | 3,628 | | | |
| Oct-05 | 532 | 808 | 1,296 | 2,533 | 4,460 | 7,336 | 13,443 | 26,668 | 25,139 | 4,058 | | |
| Nov-05 | 193 | 685 | 797 | 1,460 | 2,500 | 4,403 | 8,425 | 15,007 | 25,497 | 25,290 | 2,234 | |
| Dec-05 | 587 | 522 | 1,219 | 1,544 | 2,331 | 3,478 | 5,346 | 11,357 | 16,225 | 28,096 | 28,112 | 5,896 |

There were cases, highlighted as such in the above exhibit, where the payment of claim was made before the date of admission. This could be due to error in recording payments or some fraudulent claims. All of these payments were included with the first run-off month. The exhibit below shows the cumulative triangle prepared using the above incremental triangle.

## Exhibit 2.2.B – Cumulative Payments by Incurred Month

Amt in '000

**Monthly Cumulative Claims (Normalized)**

| No. of months | Cumulative Payments for the months of claim occurrence (Incurred month) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Jan-05 | Feb-05 | Mar-05 | Apr-05 | May-05 | Jun-05 | Jul-05 | Aug-05 | Sep-05 | Oct-05 | Nov-05 | Dec-05 |
| 0 | 2,113 | 3,466 | 5,096 | 4,860 | 5,477 | 4,960 | 3,169 | 3,910 | 5,128 | 4,058 | 2,234 | 5,896 |
| 1 | 23,564 | 26,490 | 29,115 | 31,099 | 32,916 | 29,612 | 32,641 | 36,893 | 30,267 | 29,348 | 30,346 | |
| 2 | 44,750 | 42,552 | 52,943 | 56,475 | 58,267 | 57,097 | 61,357 | 63,562 | 55,764 | 57,444 | | |
| 3 | 56,011 | 57,537 | 66,544 | 69,379 | 71,985 | 71,896 | 74,800 | 78,569 | 71,989 | | | |
| 4 | 62,677 | 65,062 | 74,239 | 78,010 | 81,181 | 79,232 | 83,225 | 89,926 | | | | |
| 5 | 66,557 | 69,638 | 78,200 | 83,045 | 85,641 | 83,635 | 88,571 | | | | | |
| 6 | 69,324 | 71,958 | 80,848 | 85,579 | 88,141 | 87,113 | | | | | | |
| 7 | 71,016 | 73,375 | 82,144 | 87,039 | 90,472 | | | | | | | |
| 8 | 71,953 | 74,183 | 82,941 | 88,582 | | | | | | | | |
| 9 | 72,485 | 74,868 | 84,160 | | | | | | | | | |
| 10 | 72,678 | 75,391 | | | | | | | | | | |
| 11 | 73,265 | | | | | | | | | | | |

The analysis of the lag triangle, thus, developed using 36 months data is the first step in IBNR analysis. It was assumed that the claims incurred until December 2003 has been paid in full. The development pattern for these 12 months was used to calculate the Completion Factor further for incurred months starting from January 2004.

The Completion Factor has been selected as the average of Completion Factors of 4 months out of the immediate past 6 months factors for the same duration after excluding the maximum and minimum factors out of them. Once a Completion Factor is calculated for a particular incurred month, the factor for that month is used for the calculation of factor for the next month. The factor calculation is done from the first incurred month from the left to the last incurred month in the right.

The exhibit below shows the percentage completion of payments, based on selected completion factors, by the end of corresponding number of months for each month of claim occurrence.

**Exhibit 2.2.C – Monthly Development Patterns**

| No. of | Monthly Development Patterns (Selected) for the months of claim occurrence (Incurred month) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| months | Jan-05 | Feb-05 | Mar-05 | Apr-05 | May-05 | Jun-05 | Jul-05 | Aug-05 | Sep-05 | Oct-05 | Nov-05 | Dec-05 |
| 0 | .012 | .018 | .034 | .043 | .045 | .039 | .019 | .019 | .032 | .014 | .016 | .033 |
| 1 | .216 | .286 | .312 | .309 | .290 | .236 | .249 | .307 | .270 | .251 | .334 | |
| 2 | .519 | .508 | .523 | .529 | .509 | .466 | .543 | .545 | .505 | .558 | | |
| 3 | .685 | .663 | .672 | .667 | .646 | .687 | .678 | .674 | .676 | | | |
| 4 | .767 | .766 | .806 | .770 | .787 | .790 | .781 | .782 | | | | |
| 5 | .838 | .860 | .875 | .855 | .860 | .847 | .855 | | | | | |
| 6 | .900 | .900 | .920 | .899 | .896 | .901 | | | | | | |
| 7 | .931 | .932 | .944 | .924 | .930 | | | | | | | |
| 8 | .955 | .949 | .953 | .952 | | | | | | | | |
| 9 | .969 | .967 | .968 | | | | | | | | | |
| 10 | .973 | .975 | | | | | | | | | | |
| 11 | .982 | | | | | | | | | | | |

For example, the Completion Factor of 0.855 for the month of July 2005 as above has been calculated as the average of immediate past six factors of January 2005 to June 2005 after excluding the maximum factor of 0.875 (March-05) and minimum factor of 0.838 (January-05) from the average calculation.

The completion factors so selected are used to calculate the IBNR amount. The exhibit below shows the ultimate claims amount and the total estimated loss ratio.

**Exhibit 2.2.D – Ultimate Loss Ratio Analysis**

Amt in '000

| Month | 2005 | | | | | | |
|---|---|---|---|---|---|---|---|
| | Premium | Claims Paid | IBNR | Total Estimated Claims | Completion % | Per Member per Month cost | Total Estimated Loss Ratio |
| January | 51,625 | 73,265 | 1,363 | 74,628 | 98.2% | 146 | 144.6% |
| February | 53,825 | 75,391 | 1,903 | 77,294 | 97.5% | 151 | 143.6% |
| March | 56,025 | 84,160 | 2,802 | 86,962 | 96.8% | 169 | 155.2% |
| April | 59,150 | 88,582 | 4,430 | 93,013 | 95.2% | 175 | 157.2% |
| May | 55,250 | 90,472 | 6,784 | 97,256 | 93.0% | 185 | 176.0% |
| June | 58,800 | 87,113 | 9,574 | 96,687 | 90.1% | 181 | 164.4% |
| July | 58,300 | 88,571 | 14,968 | 103,539 | 85.5% | 194 | 177.6% |
| August | 60,500 | 89,926 | 25,069 | 114,994 | 78.2% | 216 | 190.1% |
| September | 60,788 | 71,989 | 34,504 | 106,493 | 67.3% | 203 | 175.2% |
| October | 61,753 | 57,444 | 45,502 | 102,946 | 55.8% | 197 | 166.7% |
| November | 63,948 | 30,346 | 60,390 | 90,736 | 33.4% | 176 | 141.9% |
| December | 67,038 | 5,896 | 171,990 | 177,886 | 3.3% | 381 | 265.4% |
| **Grand Total** | **707,000** | **843,156** | **379,278** | **1,222,434** | | **198** | **172.9%** |

The IBNR analysis has been done separately for each TPA and for combined TPAs' data as well. The complete run-off period of claims varies across different TPAs from 15 months to 26 months. While the total percentage of claims paid by the end of 4 months

since the month of claim occurrence for all TPAs' data is 76 percent, it varies from 45 to 85 percent across different TPAs.

Depending upon the availability of data, the data can be categorized into major category by the type of policy or the type of claims for IBNR analysis. The data needs to analyzed first before undertaking the analysis. The differences in payment pattern across different TPAs needs to be looked at before making a decision about analyzing the data separately or combining All TPAs' data. It is imperative to have a large dataset having payment period of at least 36 months for better results from IBNR analysis. Further, depending upon the data availability, a quarterly IBNR analysis can also be undertaken.

### II.3. Analysis by Diagnosis Grouping and Provider Network Analysis

The claim data was used to analyze the quality and efficiency of service providers (Hospitals). The claim data was analyzed by the type of diagnosis, period of stay in the hospital and cost of service provided by them. This helps in identifying the major disease group, their claim frequency and the cost of treatment. The exhibit below categorizes the claims data by diagnosis codes and shows the total numbers of claims and the corresponding amount of total claim paid.

**Exhibit 3.1 – Claims Analysis by Diagnostic procedure**

DiagnosisCode [diagnosisanalysis.mdl-Dive A]

| DiagnosisCode | NoOfClaims | TotalAmountPaid | Avg Amount Paid |
|---|---|---|---|
| DENGUE HAEMORRHAGIC FEVER | 10 | 200,000 | 20,000 |
| DIARRHOEA AND GASTROENTERITIS OF PRESUMED INFECTIOUS ORIGIN | 120 | 750,000 | 6,250 |
| DYSPHAGIA | 20 | 300,000 | 15,000 |
| Dyspepsia | 20 | 100,000 | 5,000 |
| ESSENTIAL (PRIMARY) HYPERTENSION | 20 | 600,000 | 30,000 |
| FRACTURE OF LOWER LIMB, LEVEL UNSPECIFIED | 10 | 200,000 | 20,000 |
| FRACTURE OF UPPER LIMB, LEVEL UNSPECIFIED | 2 | 60,000 | 30,000 |
| HAEMORRHAGE, NOT ELSEWHERE CLASSIFIED | 10 | 400,000 | 40,000 |
| HEADACHE | 100 | 675,000 | 6,750 |
| Hyperplasia Of Prostate | 10 | 225,000 | 22,500 |
| Impacted Teeth | 2 | 30,000 | 15,000 |
| MALIGNANT NEOPLASM OF OVARY | 10 | 300,000 | 30,000 |

The claims data for the diagnosis category "Diarrhoea and Gastroenteritis of presumed infectious origin" has been analyzed further as below by the hospital type.

**Exhibit 3.2 – Comparative analysis of Hospitals**

HospitalName [diagnosisanalysis.mdl-Dive A]

| HospitalName | NoOfClaims | TotalAmountPaid | Avg Amount Paid |
|---|---|---|---|
| Totals | 120 | 750,000.00 | 6,250.00 |
| Hospital A | 2 | 17,196.00 | 8,598.00 |
| Hospital B | 2 | 8,186.40 | 4,093.20 |
| Hospital C | 1 | 10,800.00 | 10,800.00 |
| Hospital D | 1 | 903.20 | 903.20 |
| Hospital E | 5 | 23,580.80 | 4,716.16 |
| Hospital F | 1 | 13,680.00 | 13,680.00 |
| Hospital G | 1 | 3,180.80 | 3,180.80 |
| Hospital H | 1 | 5,669.60 | 5,669.60 |
| Hospital I | 1 | 3,455.20 | 3,455.20 |
| Hospital J | 1 | 19,016.00 | 19,016.00 |
| Hospital K | 1 | 5,400.00 | 5,400.00 |

An analysis by geographic location and ICD codes when combined can indicate which geographic areas have a higher likelihood for a particular disease; some loading can be appropriately done while doing Premium Rating when this analysis is backed by sufficient data. Analyzing the data in this way also helps in identifying the efficiency of providers and in negotiated service rates with the Hospitals. In short, it helps to:

- Develop cost per service values by hospital
- Develop average length of stay by hospital
- Investigate cost and use differences by area
- Investigate cost and use differences by TPA
- Factor into trend analysis

## II.4. Premium Rating

The basic situations in which the Health Actuary has to establish and maintain rating approaches are initial (new business) and renewal (continuing business) rating. The basic actuarial formula for calculating a renewal premium rate increase factor for a group or block of business with reliable, credible experience data available can be stated in general terms as follows:

$$\text{Group Rate Adjustment Factor (GRAF)} = \frac{\text{Group Claims Experience}}{\text{Current Rate Premium Income}} \times \text{Trend Factor}$$

**Desired Loss Ratio**

Group Rate Adjustment factor is the factor to be applied to the manual premium rate to arrive at the final premium to be charged to the group.

The Group claims experience (GCE) is the total estimated incurred amount by the group during the period under consideration. It can be calculated as the sum of actual claim amount incurred and estimated IBNR for the group. Actual claims incurred amount can

be calculated using the analysis done under the Performance measurement section. Estimated IBNR amount can be calculated using the analysis done under the financial reporting section.

The current rate premium income is the premium rate to be charged based on the current manual rates in practice on the group composition during experience period.

The Desired loss ratio (DLR) is defined to be the targeted loss ratio that an insurer needs to produce adequate revenue to cover administrative expenses and produce desired level of profit. It is estimated after taking into account the risk profile of each line of business. It is calculated as one minus the total loading (as a percentage of premiums) for commission (TPA and Agent), expenses (fixed and variable), profit loading, investment allowance, and other contingencies.

Trend factor is the estimated trend in the claim amount from one period to another. It is a combination of change in the cost of service (claim cost), utilization of service (claim frequency) and the mix of claims.

In case of reliable and credible data, the analysis can be used to calculate the premium rates for new business. The risk per premium per exposure measure can be calculated using the analysis done under the performance measurement section. The appropriate exposure could be defined as one of the following:

- Number of people covered
- Sum Insured
- Number of Policies

The risk premium so calculated needs to be normalized for parameters such as geographic location, sum insured, group characteristics, diagnosis code etc. Care must be taken while analyzing claim cost by age and gender, because variation between two different age groups or gender groups could be substantially different. It is important to adjust the historic data to reflect the change in geographic area. Different benefits can also produce significantly different claim costs. This distinction often happens because of variation in sum insured, exclusions in the plans, limitation on the specified services. The rates should also be adjusted to reflect the characteristics of various group sizes. As the size of a group decreases, the influence of individuals with serious conditions is magnified, and the cost per person can increase significantly. But as the size of group increases, the impact of individuals with serious conditions is dampened. All of these factors are critical in rating. This normalization can be done using a sophisticated data analysis software.

## Conclusion

The importance of analysis of health data is accentuated with the de-tariffing of non-life insurance starting January, 2007. Earlier, the insurers tended to look at the total corporate profits as their measure of success, versus product-by-product performance. Going ahead, each product line will need to stand more on its own performance as separate market prices dictate success and failure of organizations' product lines. It is also good public policy to require that products be priced appropriately, as this ensures more open competition as well as limits the risk of insolvency. Meanwhile, new market entrants will come onto the scene, believing they can offer better rates in one or more product areas. This will make for a highly competitive environment in 2007 and beyond.

IRDA, vide its circular dated 8th June, 2005 has issued guidelines which should be followed while calculating IBNR. It has also required filing a report of the appointed actuary on the estimation of reserve for IBNR claims at the end of each financial year along with the data, the compiled cumulative figures, the calculation sheets and the final results. Earlier, some insurance companies used to calculate IBNR by applying a factor on premium written during the year. But after the introduction of IBNR guidelines, they need to undertake proper data analysis for estimating the IBNR.

Further, IRDA vide its circular dated 28th September, 2006 has issued guidelines on "File and Use" requirements for General Insurance Products. It has required all the General Insurance companies including Health Insurance Companies to prepare and file Underwriting Policy before getting approval for any new products. It has also required that the pricing of products should be based on appropriate data and with technical justification. The impact of these regulatory changes has increased the importance of data analysis further.

Further, there is a need to take steps for the creation of an industry wide data bank. A complete and consistent industry wide dataset would result in undertaking the following analysis to be used across the industry:

- Developing benchmark cost for different benefits structure
- Analyzing the cost differentials across different geographical regions
- Creating benchmarks for payments to providers for different services
- Creating benchmarks for settlement delay
- Creating industry wise development factors to calculate estimated outstanding liability and then to assess the financial health of insurance company
- Helping in development of solvency guidelines
- Exploring the option of using other rating factors like MSA (Medical Saving Account), deductibles, co-payment, coinsurance etc.

The Tariff Advisory Committee (TAC) currently collects data from TPAs in prescribed forms. Aggressive data enhancement measures need to be taken at the earliest to facilitate effective actuarial analyses. This will also lead to appropriate product pricing and rate setting in the long-term interest of all stakeholders in the industry.

**Bibliography:**

*Indian Health Actuaries in a changing environment, Richard Kipp and Ronald G. Harris, The Actuary India, December, 2005*

*EMERING TRENDS IN THE INDIAN HEALTH INSURANCE MARKETPLACE, By Richard Kipp, M.A.A.A. and Thomas D. Snook, F.S.A., M.A.A.A.,*

*Study Material - - Institute of Actuaries, U.K.*

*IRDA Circular No. - Circular No.11/IRDA/ACTL/IBNR/2005-06 (Guidelines on estimation of IBNR Claims provision under General Insurance Business)*

*IRDA Circular No - 021/IRDA/FandU/Sep-06 (Guidelines on "File and Use" Requirements for General Insurance Products)*

*Eleventh Report, Committee on Public Undertakings 2005-06; Health Insurance- A Horizontal Study, Ministry of Finance*

*Framework for Information Technology Infrastructure in India Final Recommendation; Ministry of Information Technology*

*Group Insurance, Fourth Edition, William F. Bluhm,*

*Gupta, Indrani and Trivedi Mayur. Voluntary Insurance Sector in India: Partnering to Acheive Greater Coverage for Health. No. 95/2005*

*About the authors:*

*Rakesh Khandelwal:*

Rakesh Khandelwal is a student member of the Actuarial Society of India and the Institute of Actuaries, UK.. He has graduated in Mathematics. He is a qualified Company Secretary and a law graduate. He joined Milliman India in May, 2006. He works on a variety of health insurance projects for Indian as well as US clients. He has previously worked with X L India in the Reserving (Property & Casualty) department for three years.

*Bhavna Verma*

Bhavna Verma is a student member of the Actuarial Society of India and the Institute of Actuaries, UK. She has graduated in Mathematics from St. Stephen's College, Delhi University. Currently, she works with Milliman India, a Health Practice of Milliman Consultants and Actuaries.