

Challenges for Health Insurers & TPAs

Data enhancement and information sharing in the health insurance sector



*Alam Singh
Assistant Managing Director
Milliman*

The issue of data has presented several challenges to all the health insurance industry stakeholders. Till a couple of years ago, the primary challenge was that the data requirements and uses of data was not always fully understood by some critical stakeholders.

As the industry matured, the awareness about utility of data in building rating structures, product design and flexible pricing systems increased, however, the industry did not transfer this knowledge to action. This was partly due to the fact that the data was hard to use, it was frequently of poor quality and incomplete. In addition, each TPA captured different data in different formats, thus making data aggregation and analysis at the insurer's end a very daunting task. As early as 2005 some TPA's tried to differentiate themselves by providing value added services built around data analytics for specific insurers. Since health insurance was an insignificant part of many general insurer's portfolio, this value added service was embraced by only a few insurers. Till then none of the insurers had given data analysis much importance, very few of them even had a dedicated person responsible for health insurance.

Health insurance premium growth in excess of 30% for FY '05-'06 significantly changed the complacent attitude many insurers had adopted towards their health portfolio. As health became close to 20% of the total business handled by a leading private insurer, the industry started to pay closer attention to health. 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.

As the industry developed awareness about data usage, other critical events were occurring which would impact the current awareness level about data pertaining to health insurance. IRDA appointed DVS Sastry, to look at general insurance data issues. In addition, the process of strengthening and preparing the Tariff Advisory Committee (TAC) for functioning in a de-tariffed industry started, based on the vision that in the de-tariffed industry TAC would function as a data warehouse, supporting the industry. Among the first steps that TAC did towards this goal was to collect data from TPA's and start some initial analysis. The data collection process required that all data be submitted in a standardised format. However since standards which had been proposed earlier were not adopted, this was a capacity that TPA's & insurers will need to develop as per data reporting guidelines proposed by TAC.

Knowledge and awareness about data standards has existed in India for quite a while. The Ministry of Communication & Information Technology (MoCIT) constituted a working group to prepare the ground for the Information Technology Infrastructure for Healthcare (ITIHC) 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. Another committee formed by the Insurance Regulatory Development Authority (IRDA) and coordinated by Bearing Point made recommendations for data formats and data collections from TPAs. However, industry adoption of these standards was weak.

Although the TPA systems do have many common fields, and almost all capture those which can be defined as critical, variance is still significant. In addition, as anyone who has had the opportunity to study a large volume of data from multiple TPA's would testify, the main shortcoming in the data has been adherence to quality parameters during the data capturing stage. From my own experience of large volume analysis of an insurer's data from multiple TPA's, I have observed that although data quality is getting progressively better, many elementary errors are still common. The good news is that most of these errors can be eliminated easily by introducing validation checks and drop downs. In addition, training of the data entry staff can easily enhance data quality.

To standardize data collection, the TAC has published a Health Insurance Data Reporting Manual which contains detailed instructions on acceptable data structures. Adoption of this manual will be the first step in generating quality data. This initiative can then be expanded upon by adding data fields to enhance data depth and resultant analytics. Data fields as presented in the current structure of policy, members and claims datasets, which is recommended by TAC, has been found to be adequate for most types of analysis. Awareness about quality however is critical, therefore, the first section of the document discusses how the quality can be improved, the second section discusses benefits and mechanisms for data sharing.

Data Quality:

The TPA's have been regarded as the traditional custodians of enrollment and claims data. They 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.

Since each TPA captures data differently, TAC has now prescribed current formats of health insurance data collection. Data is currently segregated into the following structure:

- **Policy Data (Table A):** The policy level information is contained in this table. It has details such as 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:** 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.

The chart below shows the list of important fields contained in Tables A, B and C.

The challenges in data quality can be subdivided into 3 distinct issues: Data accuracy, data completeness and data standardization:

Data Accuracy:

It is of utmost importance that the data be largely accurate.

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
State Date	Date of Birth	Diagnosis 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 Maternity	Room & Nursing Charges
	Maternity Cover	Surgery Charges
	Floater application	Consultation Charges
	Floater amount	Investigation Charges
		Medicine Charges
		Miscellaneous Charges
		Total Claim Paid



Milliman

A high volume of inaccurate data undermines the reliability of any analysis. The several common shortcomings in the current data with respect to data accuracy are:

1) Reasonability Checks: Reasonability checks is 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 have to be undertaken on the data due to its current poor quality:

- 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 are usually found in both Policy table A and Member table B of all TPAs. 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 makes 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 the same policy numbers for group as well as individual policies.

3) Missing policy data: Policy data was missing for certain claims paid by most TPAs. The percentage of missing records can be as high as 30% 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.

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. Some examples of common fields which have completeness issues and the impact thereof:

1) 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.

2) Occupation: The occupation information helps in better understanding of the risk profile of the insured. The field for "Occupation" is usually blank in most TPA records.

3) 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 is largely unavailable or inaccurate in many cases. The lack of prescribed ICD-10 Codes can be a major shortcoming in the analysis of the data. Some TPAs provided broad level code ranges rather than a unique three digit ICD-10 code.

4) 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 generated by different TPAs. Some of the inconsistencies observed were:

Diagnosis descriptions are restricted to a text limit by many TPA softwares. This leads to incomplete diagnosis descriptions. Data in the field was difficult to comprehend due to syntax errors or broad descriptions such as “Conservative Surgery”. Some TPAs provide an additional column titled “OPINION”, frequently this contains a combined descriptive and procedural information for some records.

Most TPAs did not provide procedure descriptions at all, while others only recorded drug names. Occasionally, diagnosis and procedure descriptions are provided interchangeably by some TPAs.

Data Standardization:

Inconsistency and variation among industry stakeholders for populating certain fields highlights the need for a standard terminology to facilitate data collection. Lack of a structured mechanism for population common fields causes problems while creating a single dataset for all insurers or TPAs. Some fields which can easily be represented in a uniform manner by all stakeholders to enhance data usability are: gender, relationship of patient with primary policy holder, hospital and city names, unique identifier for hospitals, type of cover – group / individual. By adopting a uniform structure to represent each of these fields, the industry will lay the foundation for comprehensive analysis. This will also ensure that future initiatives based on pooling or sharing of data succeed.

Level of Aggregation and Compliance with IRDA standards: Some of the fields prescribed in the Data Reporting Manual are not always populated by TPAs. These fields are important from the point of view of data analysis and require more attention. They include- date of intimation to TPA, new / renewal status, pre-hospitalization and post-hospitalization expenses and date of issue of card.

The above discussion highlights the fact that any 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.

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:

- 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 developed by TAC, should be adopted by all insurers and TPAs within a specific time frame. 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.