Comparative Analysis Report
Rate Review Data

November 2012

New Hampshire Insurance Department

This Project Report and Analysis for the New Hampshire Insurance Department is prepared for David Sky, New Hampshire Insurance Department, 21 South Fruit Street, Suite 14, Concord, NH 03301.

Table of Contents

Executive Summary

ReportingMD was contracted by the New Hampshire Insurance Department to provide a comparative analysis for the rate filing process. Specifically, we were asked to review NHCHIS, Internet – State Interface Technology Enhancement (I-Site), Rate Filing Template, Line of Business (LOB), and System for Electronic Rate Form Filing (SERFF), New Hampshire Hospital Association (NHHA), Supplemental Data, Insurance Commissioner's State Based System (SBS), and New Hampshire Comprehensive Health Information System (NHCHIS). These systems were catalogued and can be found in attached Exhibits. In addition, ReportingMD reviewed external and public systems for data that might help with identifying rate review process trends.

Currently insurance carriers submit redundant information to satisfy multiple state and national programs. The staff of NHID rigorously collects and compiles the information and builds localized ad hoc reports to use in their analyses. ReportingMD compared data sets and reviewed and catalogued information to isolate and understand data used in the rate review filing process. The issues associated with data collection and reporting are directly related to consistency, accuracy, relevance, and quality. We scored each data set based on this information.

Our analysis and cataloguing of the NHID has given ReportingMD insight into the issues and barriers with the data used in the rate filing process. Our intent was to create consistency in the data sets, reduce duplication of efforts and establish a rate review reporting system that can create standardized reports with trending analyses. To accomplish this we recommend expanding some NHID data sets, while eliminating others and building a consolidated data warehouse for more comprehensive NHID reporting.

Narrative

Methodology

The methodologies employed during the discovery phase of the project included: data source reviewing, scoring, cataloging, and interviewing data source owners and NHID employees, as well as internal meetings and discussion groups. Primary data sources were first identified and catalogued. In most cases, the primary data sources were readily available and their review was completed with very few obstacles. However, reviewing a data source based on screen forms, user documentation, and other limited scope documents provides limited insight into the more substantial data issues facing the NHID rate review process. For example, ReportingMD did not work with the data sets to verify quality and therefore our understanding of the data integrity issues was limited to NHID staff that used the data. ReportingMD did score each data source based on relevancy, accessibility, quality, and consistency.

Key Definitions:

Data source: any entity that collects and manages the data used by NHID.

Data warehouse: A collection of data that has referential integrity and relational structures between data points. A data warehouse has normalized data that is imported using consistent reliable methodologies.

KPI: Key Performance Indicators

Repository: Collection of related and unrelated data from disparate systems which may not have relational data sets.

ODBC: Open Database Connectivity

OCR: Optical Content Reader

Catalogue

ReportingMD's rate review analysis focused on detailed cataloguing and evaluating of the primary data sources. We have provided a detailed description of each data source as well as an analysis of the data source with regard to importance, relevance, accessibility, consistency, and quality. Combined with the data catalog, this analysis defines the data building blocks for toolsets that can dramatically improve the understanding of the multitude of significant data points and other, not so obvious, drivers of the premium rate review process.

The data catalogue contains: Field Name, Description, and Data Type for the data sources catalogued. Due to the limits imposed by data access security issues to some data sources, yellow fields in the data columns represent fields in which information was not available. Many data sources like LOB, supplemental, and SERFF did not have a database schema or data dictionaries to identify the data schema relationships.

I-Site in particular contains thousands of tables and tens/hundreds of thousands of columns of data. To catalogue such an extensive data set would be exceptionally time consuming and of little value once completed. Instead, ReportingMD defined the types of reports and the nature of the data in the three major areas of that application: financial, compliance, and market complaints. We have also catalogued some of the tables as an example of the data (see Data Catalogue).

Other data sources were assessed but we had limited access or technical support, which made them difficult to define, i.e. SERFF. We employed a "screen scraping" methodology to capture data sources where schemas were not available. The limitations of this type of "data capture" reduced our technical understanding of the information. Without a view into the database structure, column (field) names and data types could not be defined so we defined a best guess description of the screen field elements as we reviewed them.

For secondary data sources we provided a detailed description of the nature and contents of the data. We also evaluated each secondary data source based on the same criterion as the primary data source.

Scoring

Our analysis of each data source consisted of evaluating the data using defined criteria. Ratings were assigned to each data source by multiple members of our staff and then evaluated by the team to derive a consensus. Ratings are our own personal assessments based on the skill sets of our team as defined within our original proposal. This evaluation process provides a high level view of each data set's most salient points.

When calculating the overall data set scores, we used the following weighting system, which resulted in a maximum score of 70:

       1.) Quality x 2

       2.) Availability x 1

       3.) Relevance x 3

       4.) Consistency x 1

A data source evaluation can cross more than one criterion when evaluating any given set of data. A definition of the evaluation criteria, in order of importance, is as follows:

Relevance

As we have defined above, most of the rate review process currently relies on financial data. Therefore, each data source that contained financial data was given high importance but not so much as to limit the evaluation of other types of data. We have included multiple, varied, data sources (especially in the Research Section) as part of this evaluation to broaden the scope of the potential datasets that could be derived when the data are combined, cross-referenced, evaluated, aggregated, and reported.

We have identified relevant data to warrant inclusion into the rate review process. What is important is the quality of the data and whether it is reasonably accessible? The remaining criteria: accessibility, quality, and consistency are complex to both qualify and quantify.

Accessibility

Can we access the data easily? During our discovery period, getting at the data/data schemas was made difficult by the following: privacy issues, accessibility (infrastructure) issues, competitive issues, financials issues, legal issues, and more. Once we began to evaluate the relevance of each data source, we faced significant accessibility constraints. We have provided an evaluation of each data source based on our level of accessibility and any limitations to the value of the data as a result.

Data sources that cannot be accessed directly have limited value. Data sources with access limited to application/web screens and "user" tools expose some data, but that is marginalized due to the difficulty in capturing or integrating it with other data. The hurdles usually associated with the capture of this type of data are laborious, costly, lacking in content, and prone to error. For example, consider the SERFF data source that contains unique data that are not available elsewhere, however, access to the data is limited to a website. The data cannot be copied or captured using any other method than screen copies, which violates the site's usability clause and makes it very difficult to aggregate and report. Hence, no matter how unique this data may be, they are still unavailable. Let's consider for a moment that the screen capture method does not violate the usability clause. Imagine you need several hundred (or quite possibly several thousand) screen captures to get a complete data set. While difficult, imagine the data changes on a weekly, monthly or quarterly basis. Accessibility, or the lack thereof, can be a hindrance to a data analysis.

When data is captured in structured data fields, end users will have the capability of report writing and exporting. Labor reporting costs will vary based on the data collection method and level of reporting. If a system has static .pdf based reports versus ad hoc reporting capability, the cost is very different. Direct

data exchange, either permanently through some type of trust relationship and/or shared data sources or a semi-permanent connection like web services or file transfer, is the best-case accessibility option when evaluating a data source. The data still must satisfy all of the criteria in that it is relevant, consistent, and of high quality; although accessibility is the key element. Accessibility is critical when trying to determine and maintain the quality of the data and it promotes strong data consistency for integration purposes.

Quality

For the purposes of our evaluation we have decided to split data quality into two parts: quality and consistency. Separating quality and consistency provides for two distinct views into the value of the data. Data quality topics include: accuracy, completeness and reliability and it is affected by the way data is entered, stored and managed.

Not only is there a need to capture quality data, but it needs to be maintained. We have evaluated the primary data sources with this in mind. How good is the data? Can we rely on the data? Integration of data from multiple sources requires that data sources are, and will remain, accurate and reliable.

When an organization sets out to capture data, it has to define the data sets and sources very carefully. The integrity of the data quality is crucial to the operational and transactional processes. Business reporting and business intelligence relies on this foundation upon which to build the reports, analytics, dashboards and other decision tools.

Consistency

Consistency of data is paramount when combining multiple data sources in a data warehouse. Data consistency encapsulates the validity, usability, and integrity of the data. When data is combined from multiple sources, it is considered to be point-in-time consistent. Point-in-time consistent means that all interrelated data sources must be coalesced to a similar set of consistencies to be able to recreate any

moment with high accuracy, repeatability, and sameness. For example, the LOB data is collected once a year and is reported typically by the insurance carrier's financial department, while rate review requests may be completed by an underwriter department of the same insurance carrier. Both groups used inconsistent methods to complete their business analysis and submit answers to the LOB survey and the rate review. Therefore, the result is inconsistent data collection methods, making it impossible to know which information is more accurate.

Data collection methods used to build the data catalog varied by the type and nature of the data source. Some data source owners used Information Technology professionals or organizations that were able to provide data description and documentation tools.   This type of data source collection model scored high in our criteria. While data collection models that did not use consistent shared data models scored lower.

Data Collection

NHID collects data from several resources for their reporting and analyses of insurance industry data. An overwhelming majority of the information comes from insurance company claims, financial, and membership data. Much of the data resides across various state and national programs in data repositories. Repositories are a collection of data that may or may not interrelate because the data is dumped in from disparate systems. Often these repositories on their own may or may not have quality, accessiblity, or relevant reportable data for the rate review process. In comparing the data sources, repositories, and data warehouses for NHID rate review process, various data collection methods are used.

Survey

Line of Business (LOB), Supplemental Data, SBS, Rate Filing Template, and SERFF are all examples of data collected by means of survey (see Data Catalogue). Typically, data from surveys are limited to the interpretation of the person completing the survey. Therefore, this type of data collection method can be inconsistent from carrier to carrier and from employee to employee of a singular carrier. NHID needs reliable consistent points of data when regulating premiums. Additionally, the effort and time it takes to collect and aggregate the individual carrier data surveys impacts timeliness and dissemination of the information.

The rate review process requires NHID approval for a change in carrier premium. This requires information to be collected in a timely and consistent manner. Given that the rate review process occurs as frequently as quarterly, the data collection points need to be available on a similar time line, thus making survey tools a natural choice for NHID data collection methods. However, the data needs to be collected using well established industry standards. Another example of survey tools is the rate filing template which has been developed in Microsoft Excel and is a survey based tool with defined standards for collecting data.

Data Extraction

When data is extracted from financial, claims, and membership systems the data is quite granular, therefore, the use of extract data sets are typically more diverse. Examples of extracted data sets are I-Site and Omnipoint (All Payer). Typically, reports can be generated to meet various needs of NHID. However, the extraction, load, and translation of this type of data collection needs to be consistent. Therefore, the industry has set standards like HL7 to help with comparable data sets. For example, NAIC (National Association of Insurance Controllers) has I-Site, which contains the most complete data source available when considering financial and financial performance data on insurance companies.

Case Studies

Case studies or reports identifying trends in the market place are often sighted as justification for a change in premium. Examples of this are Kaiser and Healthcare.gov, which have case studies, healthcare market trends, and market analysis (see Data Sources Section). Case studies are typically narrowly focused and conducted to meet specific objectives, which may not be related to the rate review process. Case studies are time consuming to understand the methodology used in order to determine the applicability of the case study.

Interviews

Interviews are a natural informal process of the rate review. For example, data collected through correspondence in SERFF informally presents opportunities to interview the carrier. Many of the questions and requests for more information often lead to better understanding of the carrier premium request and their ability to respond in a timely manner. This type of question and response for data leads to an informal interview with the carrier. While interviewed data are a more qualitative data collection approach, it is unclear what effect interviews have on the rate review process

Data Sources

There were two types of data sources that were reviewed: main data sources and researched data sources; each of which were identified by NHID at the start of the project. ReportingMD was asked to identify other relevant sources. We have reviewed the data sources and applied a relative weighted score to each of the data sources (see Methodology Section).

Main Data Source List:

Data Source	Description	Owner	Score
NHCHIS	NH Comprehensive Healthcare Information System	NHCHIS	65
I-Site	Internet - State Interface Technology Enhancement	NAIC	65
SERFF	System For Electronic Rate and Form Filing	NAIC	59
Rate Filing Template	Electronic collection of rate filing data	State of NH	58
LOB	Line of Business Survey	State of NH	50
NHHA	New Hampshire Hospital Association	NH Hospital Association	43
Supplemental Data	Additional requests for data made to insurance companies	State of NH	58
SBS	Insurance Commissioner's State Based Systems	State of NH	48

Researched Data Source List:

Data Source	Description	Owner	Score
CMS	Centers for Medicaid and Medicare Services	Federal Government	58
US Census		Federal Government	40
MEPS	Medical Expenditure Panel Survey		38
Kaiser	The Henry J. Kaiser Family Foundation	Private Corporation	32
HealthCare.Gov	Healthcare.gov	Healthcare.gov	10
DRED	Department of Resources and Economic Development	State of NH	NA
NHDHHS	NH Department of Health and Human Services	State of NH	NA

Accessibility/Consistency

Valuable data is consistent in content and accessibility. Consistent data has standard formats for frequent loading of new content. Data sources become more accessible when ownership, privacy, and business objectives are less conflicting. Consistency will have a significant impact on the quality of the data source. NHID has been using a limited section of data sources like I-site and SERFF because the consistency and accessibility is largely unclear. Both data sources have the capability to run more detailed or custom reporting, however, understanding the consistency (frequency of new data made available) of these data sets requires more available resources than NHID currently has.

In recent years, data has gained an enormous role in even the smallest of enterprises. The collection, cleansing, aggregation, and dissemination of data are big business. At the same time, financial and patient privacy laws (Sarbanes Oxley/HIPAA, etc.) have had a "suppressive" effect on those types of data being readily available or disseminated even though business demand has increased dramatically.

In order for data to be useful, it must be available. In order for data to be effective, it must be accurate and consistent. Data in non-structured fields is difficult and time consuming to analyze when aggregated, and/or integrated with other data sources. For example, the rate filing report is a .pdf document that has non-structured data fields and therefore can only be analyzed one request at a time. What if you wanted to know how many rate reviews had a request of a 10% increase or greater in 2012? This would currently be a laborious task to complete from a PDF document.

While there are many standardized reports (especially in the financial areas) most data is in formats that are designed for and limited in scope and quality. In many cases, especially with larger and more established vendors, those limitations do not wholly remove the value of the data; instead, they just limit the manipulation and integration of that data through pre-defined data models. For example, SERFF has a section where you can enter reporting parameters based on a predefined data set. This is helpful if all of

your data resides in the predefined data set. If your data was not accessible for reporting, it would have less value.

Other methods of data capture may still apply even though they are more difficult, expensive, and error prone methods. Such methods include: screen/scrape; report data export; and optical content readers (OCRs).

Structured data that is captured in defined data fields is only one part of making data consistent and accessible. Once the data is in a structured reportable format, one needs to decide what to do with it? Can we consistently import the data by day month, quarter, and year? How accurate is the data and can we augment it with data from another source to improve its value? Is the data set relational to other data sets? Can we aggregate and define it?

In most cases, the real value of data is in its scope. Data with a narrow scope produces limited or focused results. One of the main qualities of good data is its ability to integrate with other types of data into a master data collection or "data warehouse". A data warehouse is the collection and aggregation of data from a variety of sources that can be queried and reported on. The larger and more complete the warehouse, the larger and more complete the reporting (and value) and analysis of the results. Data warehouses tend to be built with a broad scope in mind.

Regardless, one underlying principle remains true: data quality is paramount. Not only is the capture and validation of the data important but, how the data is consolidated, analyzed, and reported creates quality and consistency concerns. When combining data sources, generally the data needs to be transformed to maintain consistency (see Recommendations).

Main Data Sources

Below is a description and evaluation for each of the Main Data Sources:

NH Comprehensive Healthcare Information System (NHCHIS)

Overall Score: (65)

Ratings: – Relevance: (10); Accessibility: (10); Quality: (10); Consistency: (5)

Description:

The New Hampshire Comprehensive Health Care Information System (CHIS) was created by NH state statute to make health care data "available as a resource for insurers, employers, providers, purchasers of health care, and state agencies to continuously review health care utilization, expenditures, and performance in New Hampshire and to enhance the ability of New Hampshire consumers and employers to make informed and cost-effective health care choices." The statute also required that the New Hampshire Insurance Department (NHID) and the NH Department of Health and Human Services (NH DHHS) partner on the project. The same legislation that created the CHIS also enacted statutes that mandated that health insurance carriers submit their encrypted health care claims data and Health Employer Data and Information Set (HEDIS) data to the state.

Positives:

This project has detailed insurance claims data sources. CHIS reflects good access to theoretically consistent data. The data can be consolidated into a centralized data warehouse. NHCHIS is the beginning of that project and based on our review, it is the ideal candidate to be the focal point of any data collection efforts supported and participated in by NHID. CHIS can bring tremendous strength and focus to the data collection, data warehousing, and reporting processes. By having a centralized and coordinated effort, the State of NH will benefit from many overlaps in need, support and other functionalities.

Negatives:

The CHIS lacks quality consistent data sets. Data is refreshed every quarter. Data sets grow very large, making accessibility for longitudinal and trend reporting difficult. Data sets are sent to one person at NHID and are loaded independently on local machines and thereby localized data bases and reports are created specific to a singular user.

I-Site

Overall Score: (65)

Ratings: – Relevance: (10); Accessibility: (6); Quality: (10); Consistency: (9)

Description:

NAIC I-SITE is an online interface designed for state insurance departments to obtain comprehensive financial, market conduct, producer licensing, and securities information. I-SITE offers regulators access to NAIC database information including Summary Reports, Batch Reports, and Detailed Lookup Reports. The integration of SBS and I-SITE ensures the seamless access to data between I-SITE and SBS.

I-SITE raised some unique issues when we tried to catalogue it. There was a data dictionary of sorts, which only proved to be enormous and unwieldy. There are tens of thousands of columns of data. To catalogue all of those columns of data would have taken significant time and resulted in no perceptible value (who's going to review them all?). What we did is catalog the company tables as they were much smaller and easy to capture. This is to provide a snapshot into the data types and structures available. We then followed that with a section that contains a list with descriptions of all of the main reports we felt significant to the rate review process (see catalogue). I-SITE's website is built for the consumption of reported data and, when used only through the website, is all that is available.

Positives:

This is a focused data source specifically created to support state insurance commissioners. It has comprehensive, detailed data within its subject areas. The data has historical depth, and great consistency. While we did not run data quality tests, we do believe there is great accuracy. Some data sources are provided by the actual insurance companies.

I-SITE data can be augmented with the All Payer database to provide more comprehensive reports. For example, the state could report profit or loss by insurance carrier for immunizations, emergency room visits, etc. across multiple years. The main KPIs for the rate review process are based on financial, aggregated claims incurred and membership data

Negatives:

We have three issues with I-SITE data:

1.) I-SITE is overwhelming in its depth and complexity. While there are many tools, reports, and descriptions, it is nearly impossible for all but highly trained users to assess and aggregate answers to specific questions without poring over reports or writing complex queries to do so. It is unreasonable to expect NHID rate review analysts to have that skill set. There is ODBC (Open Database Connectivity – an industry standard) access to the data and that is why we gave them a 6 and not a lower score. ODBC access is a good option for small data capture jobs and reporting. Given the volume of data in I-SITE, ODBC would not function well for foundational data sources for I-SITE.

2.) NAIC protects their data and access to it quite strenuously. While there are some data-definition tools available, they are essentially useless due to the data complexity. However, NAIC makes their living off of this data and NHID could be a significant customer to NAIC given the right circumstances. It might be possible for NHID to negotiate a data extraction/feed agreement with NAIC.

3.) It will be costly to do. Again, the depth and breadth of the data comes into play. It will take significant time and technical resources to define data sets needed for the rate review reporting system. Even if they agree to a total dump of the data, it will still require those resources to distill that data into NHID systems for reporting. Transformation and update processes will have to be built and maintained to load the data set at regular intervals.

SERFF

Overall Score: (59)

Ratings: – Relevance: (8); Accessibility: (6); Quality: (10); Consistency: (9)

Description:

The SERFF system is designed to enable companies to send specified information for states to receive, comment on and approve or reject insurance industry rate filings. SERFF offers a decentralized point-to-point, web-based electronic filing system. SERFF facilitates communication, management, analysis and electronic storage of documents and supporting information. The system is designed to improve the efficiency of the rate and form filing and approval process and to reduce the time and cost involved in making regulatory filings. It also provides up-to-date filing requirements when they are needed.

Positives:

SERFF would also make for a good foundational data source. While a lot of the data is communication type in nature, the metrics surrounding those communications and form filings could prove valuable as a data source. By capturing document metrics, we can minimize the time spent in SERFF and potentially build links to specific document type and their histories to further minimize time spent in SERFF doing document management tasks.

Negatives:

SERFF is a document management and retrieval system. While that type of system is certainly useful and fairly straightforward to build, the design and integration with a data warehouse is outside the scope and nature of this document.

SERFF is also owned and operated by NAIC. We expect that any accessibility issues would be the same or similar to what we discuss below with regard to I-SITE data and access.

Rate Filing Template

Overall Score: (58)

Ratings: – Relevance: (9); Accessibility: (6); Quality: (8); Consistency: (9)

Description:

This template is used for rate filing, which includes: rate change and enrollment, history of rate change; rate changes distribution components of rate change, HMO, POS, PPO, Indemnity and other plan designs and benefit level rate adjustment, historical experience, experience rate development, final trend assumptions, historical administrative costs, administrative charges, retention charges, illustrative rates, rating factors, base rate PMPM development for Standard Plan type, company financial information, loss ratio exhibit individual market, loss ratio exhibit small group market, and standard plan types.

Positives:

The rate review template attempts to standardize reporting for the rate review process. Its base platform is Microsoft Excel spread sheets. The tool set calculates key performance indicators (KPI) at the singular submission of the rate review.

Negatives:

Trending analysis would need to happen by compiling multiple rate reviews. The data cannot be aggregated and compared. The template is a survey based tool that leaves some interpretation open to the rate filer. Using Excel as a data collection and reporting tool is 20 year old technology that produces singular filed results.

Line-of-Business Data (LOB)

Overall Score: (50)

Ratings: – Relevance: (9); Accessibility: (5); Quality: (6); Consistency: (6)

Description:

Line of Business Database is based in excel. Template Excel spreadsheets are sent to insurance companies annually to collect information about number of subscribers by product, carrier key contact, premiums, if they marketed in NH in the prior year, and a brief questionnaire. Once the Information is received by NHID, the carrier is published on the NHID website.

Positives:

The data source provides general overview of the carrier's products and size of their business in NH.   It is a good way to quickly review the carrier's size and impact in NH.

Negatives:

The length of time it takes to create the database may be problematic for the state to have current relevant information. The data is accessible by NHID employees. Only one employee at a time can review the data unless multiple copies are distributed. It would be very difficult to query or analyze the data for comparative analysis.

The data quality is only as good as the person who pulled the data at the insurance company. It would be difficult to cross check or compare the data with other sources. The loading and compiling of the data is manual (copy and paste) which may lead to a level of inconsistency.

The following reasons present in inconsistencies:

1) The length of time it takes to collect data, and then be compiled by NHID

2) Data is extracted by the carrier in inconsistent ways

3) Timing of when the data is extracted.

New Hampshire Hospital Association (NHHA)

Overall Score: (43)

Ratings: – Relevance: (8); Accessibility: (5); Quality: (5); Consistency: (4)

Description:

NHHA collects data from member hospitals in NH and provides reports on their website. Data is collected in several reporting areas: inpatient reporting; ambulatory surgery; ambulatory care; and selected specialty reporting.

Reporting topics include: admissions; admissions types; payers; market share patient origin; numbers and types of procedures performed and when; and select patient demographics data.

They also compile state wide utilization and financial data for all 26 acute care hospitals in NH.

Positives:

NHHA data would provide a good source of utilization and clinical data to support claims financial data. The product and patient mix may have a direct effect (or not) on certain reported financial results both from insurance companies directly and to further detail NAIC data. The focus on only NH institutions would help clarify the data as it specifically applies to the claims based services provided at the hospital level by insurance companies.

Negatives:

It is unclear that all hospitals in NH participate in a consistent manner. The data collection methods are unknown and the data may have some inconsistency and potentially be outdated by the time it is published. There are no data feeds available.

Supplemental Data

Overall Score: (58)

Ratings: – Relevance: (8); Accessibility: (7); Quality: (9); Consistency: (9)

Description:

The data is submitted in the Supplemental Report is critical to understanding and evaluating the New

Hampshire's health insurance market. Specifically, this data uses financial reports to determine solvency

of accident and health carriers. This data set is only applicable for more than 2400 or more covered life

months.

Positives:

The data set provides a good general overview of the number of claims, lives, benefits and services

provided by the carrier.

Negatives:

The data are not due to the state until July 15th. At this time it is already 6 month old data that NHID

needs to base its rate review process. The data is submitted in a MS Excel workbook, thereby limiting the

applicability of comparing 100 of entities and aggregating the data for reporting. The data is sent via

email to NHID, which creates data management problems.

SBS- State Based System

Overall Score: (48)

Ratings: – Relevance: (5); Accessibility: (6); Quality: (9); Consistency: (9)

Description: SBS Company licensing data source is provided as part of I-Site by the National association of Insurance Commissioners. The data source provides company demographics, name, address, business lines, deposits, contacts, branch and mergers. The data source is very high level and identifies the current licensing and financial status of the insurance carrier.

Positives:

SBS data is available in structured fields, which allows users to access the data they need with relative ease. The system has consistency but it is less relevant for the rate review process.

Negatives:

The data may be restricted at some levels within NHID. The information contained is very high level and would not support KPIs needed for the rate review process.

Researched Data Sources

A description and evaluation for each of the Researched Data Sources:

Centers for Medicare & Medicaid Services (CMS)

Overall Score: (58)

Ratings: Relevance: (8); Accessibility: (7); Quality: (9); Consistency: (9)

Description:

The National Medicare/Medicaid database. CMS provides an extremely rich Research, Statistics, Data & Systems site that contains extensive data that has been collected through patient billing of services, laboratory testing, and procedures by providers and institutions. It also contains the results of numerous surveys, the data collected through the Physician Quality Reporting System (PQRS), and significant data collected both historically and continuously though many other programs and methods. Both data view/reporting tools and direct data extraction/purchase opportunities are provided.

The scope of the data is extensive and covers such topics as: Research; Actuarial Studies; Consumer Assessments/Research/Initiatives; Outcomes Surveys; Enrollment population data and statistics; Geographic Variations; and Clinical & Lab data and more...

Positives:

The scope and historical depth and breadth of data available are comprehensive and detailed. The data are available in a variety of formats. Most of the data are available as a matter of public record although fees are charged and vary based on type of data, amount of data, use of the data, and organization type.

Other positives:

1.) Both clinical and financial data is available

2.) The Data Quality and Consistency values are high

3.) Tremendous historical depth

4.) Would provide rich analytical content for the Medicare/Medicaid portion of both financial and population demographics data

5.) Continued PQRS/Meaningful Use initiatives will provide a powerful and rich data source for years to come

Negatives:

There is little negative that can be said about such a comprehensive data source, however, there are a few barriers to effective use of this data source:

1.) Volume and scope of data. The very size and depth of the data contained is intimidating and would require careful analysis to determine what to extract and how often. The sheer volume of data could have significant impact on data warehouse architectural requirements.

2.) It was not readily apparent that there are direct data linking options. Data extraction may only be by request and may be limited in size, timeframe, and format per request.

3.) It contains only Medicare/Medicaid data.

US Census Bureau

Overall Score: (40)

Ratings: Relevance: (3); Accessibility: (7); Quality: (8); Consistency: (8)

Description:

The US Census Bureau serves as the nation's leading source of quality data about the US people and

economy. Population and housing data is collected every 10 years and economic data every 5 years.

Positives:

The US Census Bureau can provide extensive population demographics.

Negatives:

Volume of data contained; interval between collection periods

Medical Expenditure Panel Survey (MEPS)

Overall Score: (38)

Ratings: Relevance: (6); Accessibility: (5); Quality: (5); Consistency: (5)

Description:

The Medical Expenditure Panel Survey (MEPS) is a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. MEPS is the most complete source of data on the cost and use of health care and health insurance coverage.

MEPS currently has two major components for which data are released: the Household Component and the Insurance Component. The Household Component data are based on questionnaires fielded to individual household members and their medical providers. The Insurance Component estimates come from a survey of employers conducted to collect health insurance plan information.

Positives:

The Extensive Insurance Component section is based on US Census Bureau data. Data are not publically available but may be accessible to NHID. This MEPS data has a summary of data tables on employer-based health insurance with querying tools. Statistics about survey methods; data collection methods; sample sizes, etc.   According to their website, they have: "The most complete source of data on the cost and use of health care and health insurance coverage."

Negatives:

Limited scope survey based data. In some cases, the data may be a replication of Census Bureau data. Due to confidentiality and privacy issues, data may be difficult to access/re-use.

The Henry J. Kaiser Family Foundation

Overall Score: (32)

Ratings: Relevance: (3); Accessibility: (3); Quality: (6); Consistency: (8)

Description:

A leader in health policy analysis, health journalism and communication, the Kaiser Family Foundation is dedicated to filling the need for trusted, independent information on the major health issues facing our nation and its people. Kaiser is a non-profit, private operating foundation focusing on the major health care issues facing the U.S., as well as the U.S. role in global health policy. Unlike grant-making foundations, Kaiser develops and runs its own research and communications programs, sometimes in partnership with other non-profit research organizations or major media companies.

Kaiser serves as a non-partisan source of facts, information, and analysis for policymakers, the media, the health care community, and the public. "Our product is information, always provided free of charge "— from the most sophisticated policy research, to basic facts and numbers, to information young people can use to improve their health or elderly people can use to understand their Medicare benefits.

Positives:

Kaiser offers a wealth of data. Subjects include: Health Reform; Medicare/Medicaid; Insurance Costs; Uninsured Costs to Healthcare; Sate Health Policy; HIC/AIDS; Global Health Issues; Minority & Women's health; and Healthcare coverage in the media. Coverage of those issues appears to be extensive and detailed.

Negatives:

Almost all data is study based, which means narrow in focus. Additionally all data appears to be exclusively in .pdf format, which is difficult to capture. Because of the institutional research, the

community forum/journalism focuses; and limited access to complete datasets we consider Kaiser as

more of an ancillary source. Were the project to take on the role of archival document storage and

retrieval, there may be much more content here of value.

Healthcare.gov

Overall Score: (10)

Ratings: Relevance: (1); Accessibility: (1); Quality: (2); Consistency: (2)

Description:

A consumer oriented site sponsored by U.S. Department of Health & Human Services. Contains

consumer oriented subject matter: find insurance; understand insurance purchasing; healthcare laws and

consumer protection; and prevention and wellness.

Positives:

May offer some limited view into the consumer side of insurance plans and purchasing.

Negatives:

Limited value. No data sets available. No data other than consumer information.

Review & Analysis

KPI

Key Performance Indicators (KPI's) are used to evaluate and review the premium rate. A combination of

membership, claims, and financial data values are used for the rate review process. How well a company

performs financially and its financial stability are essential, baseline values. Some of the KPIs data

sources identified during our discovery process were:

KPI Data Sources

KPI Data Sources	Data Source
Financial	
Loss ratios	I-SITE
Profit vs. administrative costs	I-SITE
Enrollment statistics	CHIS
Product mix/lines of business	LOB, CHIS
Product markets and penetration	CHIS
Market share analysis	I_SITE
Premium rate	Rate Filing Template, CHIS
Coverage plan details	SERFF, CHIS
Carrier claim experience	I-SITE, NHHA, CHIS
Clinical (claims based data)	
Admits/discharges	NHHA, CHIS
Procedure mix	NHHA, CHIS
Medicare/Medicaid population mix	NHHA, CMS, CHIS
Patient demographics (age, gender, location, income)	NHHA, CMS, CHIS, US
Membership (PMPM)	
Administrative	Rate Filing Template
Premium	LOB, Rate Filing Template
Deductible	Rate Filling Template
Inpatient, discharge, professional	Rate Filling Template

There was enough data discovered to support the above KPIs. The difficulty is getting access and

integrating that data. The same data fields can exist in multiple systems, but the data result can be very

different between systems. This presents a problem when developing a consistent set of reports to approve a rate filing.

To properly create a system that supports the above KPI data sources with clarity and detail will require data from multiple sources to be collected, distilled, transformed and coordinated. Based on our review the data and technology are available to build a comprehensive reporting system centralized in one application.

Conclusions

Data is dispersed in a variety of formats and data sources for use by NHID rate review filing. Today, comparing and using these data sources can have conflicting results. Financial, payer claims, and membership data are the main data sets and there are numerous sources available for each. The same data fields are scatted across multiple data sources. Consolidation of the current disjointed collection of data would require an extensive, laborious, and error prone process for use in the current rate review filings. It is unreasonable to expect to consolidate the data from these sources because the collection methods are non-standardized for rate review filing. Any efforts to incorporate even a few select current data sources would not be capable of supporting the rate review process. The complex raw data values of the KPI and reporting timelines make consolidation of existing data sources a useless endeavor that would produce few discernible positive results. However, data may be collected from non-aggregated data sets from a few existing sources to create a data warehouse. The existing sources will need to have their data sets expanded to support this.

The NHID rate review process has developed a rate filing template that has a clear set of data points which define the needed KPI.   The KPI in the Rate Filing Template is collected and entered in a survey based Microsoft Excel tool. As stated in the data collection section above, survey based tools can be interpreted by the filer, therefore creating inconsistent data entry points. The sheer volume of data points collected in the rate filing template (over 270 data points for one plan) creates a data reporting and collection burden on the carrier that is repetitive by the carriers with other reporting mechanism to NAIC and the state. Even if the data can be collected in an accurate consistent standard, it will only be reportable for that singular rate filing for the carrier. In other words, the data cannot be aggregated and compared easily amongst carriers and across time for longitudinal analyses without considerable effort in consolidating the Excel spreadsheets into a data warehouse.

NHCHIS data that are expanded can become one of the primary data sources in the rate review process. Finally, the NHCHIS data are claims-based allowing for the data to be aggregated and reported, benchmarked, compared, trended, and reportable.

I –Site data sets carry considerable information that can be used for analyses. We have supplied a sampling of the I-site data in the catalogue section. The I-Site data can have regular extraction routines established to load into another reporting system.   I-Site KPI data sources should be identified for potential extraction.

Currently, NHID rate review process has some ad hoc data analysis processes, but they are singular in nature. Given that non-aggregated, financial, claims and membership data exist in NHCHIS and I-Site the opportunity to build aggregated rate review standard set of reports by carrier and plan trended over time exists. SERFF is essentially a rate filing collection tool that allows for filing tracking and collection source for the carrier and NHID. The process and tool seem adequate to continue to collect the rate filing request. The main source of concern is identifying a way to verify the filing submitted by the carrier with consistency, accessibility, and accuracy. ReporitngMD's recommendations address this issue.

Recommendations

ReportingMD found that NHID staff was not restricted from use of the data sets we reviewed, but there are limitations to the use of NHID data sets by staff. First, some staff spend months compiling data sets like LOB that could be handle in minutes in by loading the data into a data base. These staff could be better utilized by analyzing data rather than compiling.   The NHID has limited number of technical report writers. Some staff has a dual purpose as a programmatic analyst and a data programmer/analyst. To work with larger data sets like CHIS and I-Site requires data analysts. There are significant barriers to using the current NHID data sets. One must have solid understanding of insurance data sets, the format of the data set (sql, pdf, Excel), and the technical (understanding relational databases) skills to be able to create reports that can support anlaysis for the rate review.. For example, some of the data sets reviewed are written documents in a pdf format. Any staff given access can read these, while others are in a SQL database that only staff with understating of relational database reporting can understand or review. Thereby, there are limitations to data access given staff abilities. Additionally, access barriers related to ease of getting data. For example, NHCHIS data sets seem to be loaded locally on machines and used independently.   While many of the staff seek the use of similar data they would need to replicate and store the same data on their machine and begin to build reports locally and specifically for their use. We recommend that the NHCHIS data are installed on a server and shared through the use of ODBC drivers, thereby connecting natively.   Finally, we recommend that NHID staff the best understand a data set share their knowledge with other staff. A few hours a month to provide cross training and knowledge sharing could help staff to increase the use of all data sets.

We recommend expanding the NHCHIS to have expanded financial and member ship data. The membership data in NHCHIS, especially, needs to include member benefit detail like case management, Chiropractic DME, ER, etc. as well as employer unique identifiers. Expanding the NHCHIS data set to include or derive (if it exist) financial data, specifically carrier's medical expense, hospital expense, surgical expense data, quality improvement expense, and net revenue, would make a more comprehensive

data warehouse and replace other data sets in I-site and LOB. Additionally, we recommend the data

extraction of a well-defined set of I-Site data points. NHCHIS with the proper expanded mandate could

become a comprehensive data source needed for NHID Rate Review and LOB survey. Much of the

NHCHIS data, once aggregated, can produce the same information that is in LOB and Supplemental data

so it will be possible to eliminate these and save considerable staff time and insurance industry effort. In

the future, NHCHIS could become a data source for data sharing for other NHID and even more DHHS

programs. Again, we do not recommend creating a repository (collection of various data sources) from

all the data sources we reviewed.

ReportingMD recommends incorporating I-site data and NHCHIS data into the rate review filing process.

To accomplish this NHID will need to take transactional data sets in NHCHIS and I-site so that claims,

membership, and financial data sets can be aggregated into standardized reports, which provide the same

detail found in the rate filing template, LOB, and Supplemental data. Ultimately, this would reduce the

number of data sets for NHID and allow for better concentration of staff learning and absorption efforts of

NHCHIS and I-site data.   In addition, this should help with the reduction of data requests from the

carriers and create less conflicting points of data for NHID staff to use. Specifically, because the rate

filing process requires a controlled collection of data, NHID staff can thereby create aggregated data

reports to the rate filing criteria. This reporting system is recommended to be built in a data warehouse

and can be built from a combined data extraction of NHCHIS and I-site. A data warehouse by its

definition has accessible data that is easily disseminated through a standard set of reports for trending and

comparable analysis over time and by carrier. Once the data are loaded into the data warehouse, a rate

review automated report package can be created.

A data warehouse can be connected to by several different types of ad hoc reporting tools including,

Microsoft Access, SAS, Cognos, and Crystal Reports. Currently members of NHID are skilled in using

SAS and Microsoft Access, therefore additional report building training will be less expensive for NHID.

Based on our data review we recommend the following next steps:

1.) Review of the information contained in this document with Rate Review Analysts; technical information management staff; and senior management.

2.) Work with I-Site regarding I-site data extraction data sources for rate review.

3.) Define the data sets that need to be expanded with NHCHIS.

4.) Create an ETL (extract, transform, and load) process to capture, cleanse, consolidate, aggregate and populate the data warehouse with additional membership and financial data for the rate review process.

5.) Create a set of standard reports from warehoused data to support the rate review reporting process

6.) Use training and user feedback loop to verify and approve reports and train end users to perform ad hoc reporting.

7.) Review accessibility of data sets available and provide training on using the data sets to create greater opportunity for data analysis for all NHID Staff.

Exhibits

The Main Data Source Catalogue

A catalogue of the Main Data Sources. Each data source contains (when available): Column (field) name,

column description, and column data type.

The Researched Data Sources

A catalogue of the Researched Data Sources. Each data source contains a description of the nature and

types of data available.