MILLIMAN WHITE PAPER

# Role of data in transforming UK private medical insurance analytics

Joanne Buckle, FIA
Natasha Singhal

## Health insurance value chain and analytics

Insurers generally hold complete data sets which are sufficient for aggregated and high-level analysis, as necessary for regulatory reporting[1].Using advanced analytics for clinical insights or detailed trend analysis requires a combination of clinical, financial and demographic data relating to the members of a population.

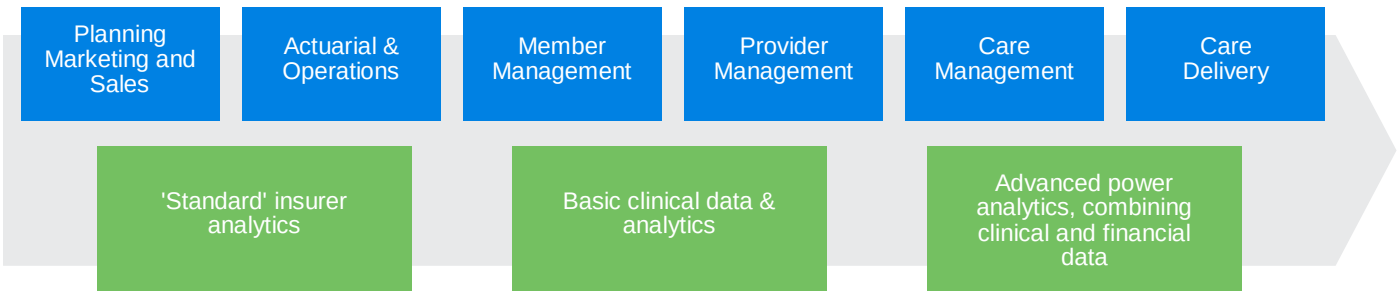
As insurers move from pure reimbursement services through to provider management, care management and aspects of care delivery, they are using new, more powerful analytic tools.

Figure 1 shows how analytics and advanced services are intertwined within the health insurance value chain.

Why is it that data quality and all the various characteristics of data are so important? Most simply it is because only with data confidence do we achieve decision confidence—the ability to make and implement key decisions based on credible and accurate data. In one form or another this is generally a well-known fact, yet achieving it is not an easy feat.

Data quality and integrity are of more importance now than at any previous time in history. As insurers and providers delve into advanced analytics, each hopes to achieve more precise decisions for applications ranging from risk and finance to making individual patient care decisions.

**FIGURE 1: HEALTH INSURANCE VALUE CHAIN**




## Elevating quality by applying data assessment tools

The sheer volume and complexity of healthcare data can create bias, hinder analysis and impair decision making. Often health service data contains input errors and incomplete records from service providers, leading to errors in interpretation by data users. Given the significance of decisions, more weight and attention needs to be placed on assuring quality data before approaching analytics. Elevating the importance of data quality is a responsibility for any analyst and the reason Milliman invests in data assessment tools to assure data confidence.

The drive and ability to push analytics past traditional actuarial and operational appraisals is dependent on the completeness and quality of the clinical and financial data held by the insurer. The data creates the foundation for the results, which in turn gives direction on how services and products are engineered.

A good quality data set, over and above the minimum required to meet regulatory requirements:

- Allows automation as analytics can be embedded in operational processes if data confidence is present.
- Enables faster delivery of key business results with confidence and additional insights/information at the push of a button.
- Increases use of health analytics to improve the overall business model, such as:
  - Risk stratification and segmentation as well as impactability modelling for population health management
  - Standardised analytical output serving various stakeholders.
  - Benchmarks from insurer-owned data to improve underwriting performance.
  - Provider profiling based on historical data.
  - Pricing and product development.
  - Designing, implementing and monitoring care management programmes.

[1] Solvency II reporting within the EU prescribes a certain level of data requirements to allow for solvency calculations as per guidelines.

## Data quality tool

How can the quality of healthcare data be assured? What are the ideal characteristics for a healthcare data set for the purposes of health analytics? The list of characteristics required for today's technology-driven environment can be collated as data that is credible, reliable, valid, structured, relevant, timely, clear, complete, rigorously evaluated for bias and statistically accurate.

We present the use of a proprietary data quality tool developed to assure data quality is properly vetted prior to actuarial analysis. The data quality tool was applied to data collected from five different private medical insurance (PMI) insurers in the UK, representing over half of the PMI market. Below we share results achieved by applying the data quality tool and discuss ways in which high-quality PMI data can be used to achieve decision confidence.

Our data quality tool comprises four overarching concentrations, each containing a number of specific checks. Figure 2 describes each of the four concentrations.

For each concentration, a specific threshold test is applied. If data presents values more extreme than our thresholds, then the level of discrepancy within the data set will be considered too high. If a test conforms to our threshold levels, a score of 1 indicates a 'pass' on the test, otherwise no points are awarded. For extremely important tests, such as availability of clinical codes within the data set, if the actual results are significantly inferior to that which we would expect in a high quality data set, a score of -1 is given to emphasise the importance of particular fields in advanced healthcare analytics.

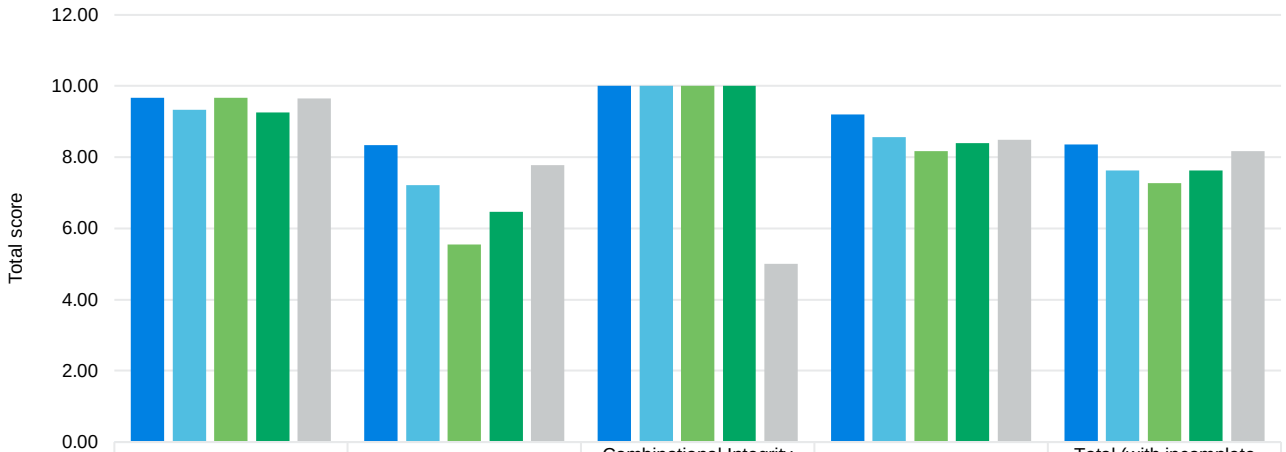
Key fields that we test from the data include billed, allowed and paid amounts within the claims data. Dates within claims data, such as claims incurred, claims paid and claims admission and discharge dates are used, alongside clinical data within the claims data, such as diagnosis and procedure coding fields. We also look at policy dates and member characteristics, such as age, gender, excess levels and duration.

Figure 3 shows our results for the five data sets used in this tool.

**FIGURE 2: DATA QUALITY TOOL: 4 CRITICAL CONCENTRATIONS**

| | |
|---|---|
| **Data validation, edits and thresholds** | Reviewing key data fields, the values within them and requirements for record level edits for data accuracy e.g. data within period of analysis, reasonable entries within fields. |
| **Data audit** | Assessing the credibility of data within each field, distributions of claims and exposure fields and erroneous entries within the data, e.g. negative exposure. |
| **Combinational integrity** | Coherency within datasets to allow the merging of claims and membership information without statistically significant loss of data. Whether data has allowed us to study a member longitudinally within a dataset. |
| **Reconciliation** | Ease with which data can be reconciled with financial and accounting information and to control totals. |

**FIGURE 3: DATA QUALITY TOOL SCORES[2]**




| | Data Validations | Data Audits | Combinational Integrity & Reconciliation | Total | Total (with incomplete tests) |
|---|---|---|---|---|---|
| Insurer A | 9.67 | 8.33 | 10.00 | 9.20 | 8.36 |
| Insurer B | 9.33 | 7.22 | 10.00 | 8.57 | 7.64 |
| Insurer C | 9.67 | 5.56 | 10.00 | 8.16 | 7.27 |
| Insurer D | 9.26 | 6.47 | 10.00 | 8.40 | 7.64 |
| Insurer E | 9.66 | 7.78 | 5.00 | 8.49 | 8.18 |

All scores are standardised to a maximum possible score of 10 for each category, including the total. Overall we did not find that any one data set scored very low, though insurer E scored low in the combinational integrity test because it was not possible to link claims and exposure data together.

It can be seen that, generally, at a total level, all data sets ultimately scored within a similar region. At a more granular level it is possible to highlight the strengths and areas of improvement for each data set.

Interpretation of results, by concentration:

### DATA VALIDATION

Overall most insurers scored well on the data validation tests, showing these data sets are in general accurate and complete.

### DATA AUDIT

There are lower-than-expected scores on data audit for almost all insurers. Data audit tests whether the data is suitable for advanced analytics and whether the analytical outcomes will be reliable, given any lack of information within the data set. This is the main area of improvement to enhance data sets and allow for further growth in the types of analytics possible. Areas of improvement that would significantly improve advanced analytical capabilities included:

1. Complete and accurate coding of clinical conditions. This can be done by using standardised code sets, such as the Clinical Coding & Schedule Development (CCSD) code set and the International Statistical Classification of Diseases and Related Health Problems (ICD-10) code sets.

   A complete set of clinical codes is imperative for advanced clinical analysis. Where the quality of clinical codes within a data set was below our threshold due to having too many admissions with missing codes or having a proportion of invalid codes above our threshold, we awarded negative marks to highlight its importance.
2. Improved maintenance of admission and discharge dates for hospital visits. Allowing for accurate calculations of the length of a hospital stay. Possible solutions could include avoiding grouping together multiple visits, which may skew the length of stay to a longer duration than actually experienced.
3. Maintaining similar field descriptions over time. For example, if the name of a benefit is updated in one year with no real change to the actual benefit, an update to previous years will ensure consistency within the data set and avoid confusion and inconsistencies in future years.

### COMBINATIONAL INTEGRITY AND RECONCILIATION

Overall we found all but one of the data sets possessed combinational integrity, allowing the matching of membership data to claims experience. Not being able to match policyholder characteristics to their claims experience meant the data set had to be excluded from any advanced analytics.

[2] Total with incomplete tests includes the tests that were not possible to complete due to data sets missing data variables. We provide two scores to avoid penalising where an insurer is unable to capture certain information, or it is not required for its specific benefit plan.

## Conclusions

Our results indicate that an improvement in the quality of the actual data held and its completeness would have the greatest impact in improving the ability of actuaries and insurers to carry out advanced operational and clinical analytics. For example, with more comprehensive clinical fields, greater precision in member segmentation can be achieved. This yields more extensive and reliable care management processes, in turn leading to further efficiencies and better measurement of outcomes.

The costs of lacklustre data, limited and inadequate for the intended purposes, are noticeable and will be a significant shortcoming for the overall market. Research predicts poor-quality data can cost businesses on average 12% of revenue a year[3], ranging from impacting marketing and sales to compliance and most certainly risk management.

Inadequate data is costly because it is time-consuming to work with and often expensive to address. For example, even if information technology (IT) systems are updated for an error in collecting data on a risk factor, it will be at least a year before enough data is collected to make any analysis on experience from this factor reliable.

Fixing errors as soon as they are identified should become standard practice. Ignoring errors is costly, requires more manual workarounds with higher error risks and complicates the move to modernise for advanced automation.

Through focusing efforts on improving data collection and by analysing areas of improvement in data quality, PMI insurers can effectively improve the advanced analytical capabilities of their data to improve business models and decisions in the longer run. Insurers can improve answers to questions on provider profiling, risk prediction, attractive benefit design options and testing return on investments. Data quality tools allow the identification of specific improvements, based on the questions asked and allow for action-based data quality improvements to produce meaningful analytical improvements.

## How Milliman can help

Our consultants have experience in advanced analytics on numerous medical claims data sets globally, including the building of advanced risk prediction and underwriting models, testing of return on investment for care management interventions, provider profiling and population segmentation. We can leverage our experience and international presence to help our clients to better understand the costs and consequences of poor data quality and build processes to improve the collection and use of clinical and claims data.

If you have any questions or comments on this paper, or on any other issues affecting data protection, please contact any of the consultants below or your usual Milliman consultant.

[3] Source: https://econsultancy.com/the-cost-of-bad-data-stats/


Milliman

Milliman is among the world's largest providers of actuarial and related products and services. The firm has consulting practices in life insurance and financial services, property & casualty insurance, healthcare, and employee benefits. Founded in 1947, Milliman is an independent firm with offices in major cities around the globe.

milliman.com

**CONTACT**

Joanne Buckle
joanne.buckle@milliman.com

Natasha Singhal
natasha.singhal@milliman.com