p–ISSN: 2723 – 6609 e-ISSN: 2745-5254
Vol. 5, No. 12, December 2024 http://jist.publikasiindonesia.id/

Case Study of Claim Data and Participant Data in Indonesian Insurance Companies

Chris Solontio[1], Achmad Nizar Hidayanto[2]
Universitas Indonesia, Indonesia
Email: chris.solontio21@ui.ac.id[1*], nizar@sci.ui.ac.id[2]
*Correspondence

Keywords: data management; data quality; health insurance; data quality dimensions.

ABSTRACT

Data has become an asset for insurance companies, especially health insurance, which has many benefits so management needs to realize the importance of good data quality to avoid the impact of poor data quality. In this study, data quality measurement will be carried out by observation to see the total amount of invalid data from the data dimensions, namely, accuracy, completeness, and consistency on the connection between claim data and participant data as well as findings from all data from the case study site of this research. In addition to the data analysis, interviews were conducted with the IT, Customer Retention, Operations, and Actuarial teams which are directly related to data flow and data processing to make analyses that assist management in making decisions. From the results of the analysis and interviews that have been conducted, there are still inappropriate data and obstacles faced by users in dealing with poor data quality. From the results of the analysis, management needs to form a data govenance team that has responsibility for the entire data flow and maintains data quality. Later, the data set that has been managed will have a positive impact on other teams in terms of analyzing trends or fraud in a faster time, helping in the creation of data warehouses, the application of artificial intelligence (AI) and digital transformation as a form of company improvement to insurance policyholders.

Introduction

According to the results of the Populix survey, the majority of Indonesians already have BPJS Kesehatan insurance of 83% and private insurance 38% (Data Indonesia, n.d.). One of the most owned private insurances is health insurance (Katadata, N.D.). This makes the medical data record on the use of health insurance will increase over time. The increase in health data in insurance companies makes management want to use the data set to provide improved services to participants, carry out digital transformation or competitive business competition such as product offerings by looking at the segmentation of participants or providing the best premiums. Therefore, one of the things

that every company must have is the quality of the data that is recapped every day, has good data quality, no data silos and is well integrated between claim data, products, and participant data. From what has been explained, this study will use one of the life insurance companies to assess the quality of the data owned by the life insurance company.

This life insurance company is one of the largest life insurance companies in Indonesia that has a wide range of hospital or clinic partners and has two main health products, namely managed care and indemnity. Both health products provide health protection for employees who work in a company. In this study, the quality of the data from the managed care product will be assessed, where this product is a tiered health insurance product whose services must go through a first-level health facility, namely a clinic or health center, then if further action is needed, a referral will be given to the hospital for outpatient treatment at the advanced level or inpatient with a note that the diagnosis and services provided to the participants need to be recorded. From each of these steps, each provider needs to create an SJP number to be used as a unique number from the medical recap journey from the first level, advanced level to the need for hospitalization carried out by the participant.

In the face of today's business challenges, the company has the initiative to improve its services to be the best and help management quickly make decisions through digital transformation. Some of these initiatives have been carried out such as accelerating the claim process, creating dashboards for the company's internal and assisting management in making decisions from cash flow results, claim behavior and others. Meanwhile, initiatives that have not been implemented are the creation of a data warehouse, and the use of artificial intelligence (AI) in detecting fraud. From the initiative that has been running, there are obstacles where the process takes approximately 3 hours because it needs to validate data because there is still incomplete, siloed and inconsistent data. Meanwhile, initiatives that have not been implemented are also experiencing obstacles because they cannot be executed due to poor data conditions. Therefore, there are obstacles in data quality.

Data quality is one of the components in data management that can produce efficient operational processes, help decision-making, create data warehouses, and have a positive influence on customer satisfaction (David, 2011). In maintaining data quality, every company needs to define data dimensions to know the impact of poor data quality on costs, reputation, compliance regulations, and so on (Askham, Cook, Doyle, Fereday, & Gibson, 2013). In the insurance business line, poor data quality in general can lead to losses in operational and strategic costs for hidden costs and direct costs (Haug, Zachariassen, and van Liempd, 2011). hidden costs and direct costs from the effects of

poor-quality data on operational activities and strategies for insurance companies as shown in Figure 1.

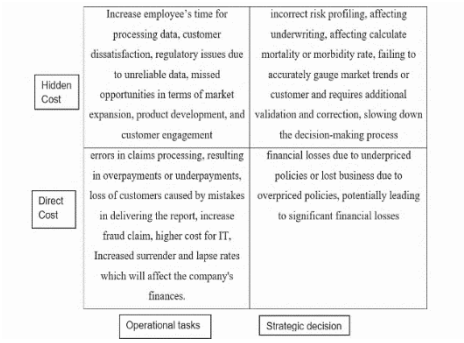

Figure 1
Consequences of Poor Data Quality in Insurance

In data quality for the information age, Thomas Rednan formulated a set of data quality dimensions that are rooted in data structures. In addition, the dimensions of data cannot be determined the same in every business area but can vary in each company depending on the characteristics of the data or the use of data by the company. The importance of data dimensions as a tool to measure data quality in an organization, many researchers conduct this research by conducting literature reviews or case studies in several business lines such as Table 1.

Table 1
Data Quality Research Area

Data Quality Research Area	Reference
Government	Government Organization (Nulhusna, Taufiq, & Ruldeviyani, 2022), BPS-Statistics Indonesia (Rahmawati and Ruldeviyani, 2019), Malaysian Public (Zarina Abdul Jabar et al., 2022), State Electricity Company (Purnomoadi et al., 2023).
Financial industry	PT BPI (Sunandar and Nizar Hidayanto, 2022).
Airport services	PT JAS (Wahyudi and Isa, 2023).
Education	Institute of Statistics (Wijayanti et al., 2018).
Factory	Paper Factory (Xu, Zhang, and Shi, 2022).
Medical or health	Electronic Medical Record (Miguel et al., 2021).

Method

In assessing the quality of this data, observations will be made with profiling that has been carried out in Mary's research by forming several roles that are carried out and looking at the number of invalid data that occurs in claim data and participant data by paying attention to the dimensions of data quality that will be used based on data characteristics, namely accuracy, completeness, and consistency. In addition, the interview process was carried out in several related units whose work activities are directly related to claim data and participants to find out the causes of poor data quality and the resulting impact.

The limitation of this study is that the claim data analyzed is claim data that occurred at a provider (hospital/clinic) in 2022 and the amount of data provided was 125 thousand data consisting of 38 thousand SJP data, 9 thousand claim participants, and three providers. SJP is a letter of guarantee for patients after health services are carried out at a hospital or clinic. In data processing, it will use excel and Python to view invalid data from participant data and claim data.

Results and Discussion

Based on the six roles that have been formed from claim data and participant data to assess the quality of the data, invalid data from each data dimension is found as shown in Table 3. The results were:

1. For the first role, no invalid data was found on the receipt date and delivery date data by looking at the length of stay.
2. Participant Status with a diagnosis was found to be 2 SJPs invalid with information on the status of child participants but the diagnosis was Z34 which is a diagnosis of normal pregnancy surveillance.
3. Based on the list of diagnoses per gender based on ICD X on the ICD10CM website, it was found that 13 SJPs were invalid in the data claiming a relationship between gender and diagnosis (Table 4).
4. The type of service with a diagnosis of 12,307 SJP is inconsistent where the service does not have details of the type of medication administered.
5. There is an inconsistency between the participant plan and the participant data where there are 36 participants who have 2 different plans for example participant number 00001 but have silver and blue insurance plans
6. It was found that 14 participants whose age range did not match their membership status (Table 2).

Table 2
Invalid Data on ICD X Diagnosis By Gender

Gender	***Diagnose Invalid***
Male	O74, N91, N89, N80, and N93
Female	C61, D17.6, and N47

Table 3
Invalid data on age range Based on Participant Status

Status	**Age Range**	**Total Participants**
spouse	0 - 5	1
	6 - 12	2
	13 - 16	6
child	36 - 45	4
	46 - 55	1

In addition to the results of data analysis, interviews were conducted with departments that work directly or manage data, namely information technology, customer retention, operations, and economics to find out the effects and consequences of poor data quality. From the results of the interview, several points were found as follows:

Table 4
Total Invalid Data Based on Data Quality Dimension

It	*Roles*	*Accuracy*	*Completeness*	*Consistency*
1.	Service date and discharge date (hospitalization)	0	0	0
2.	Participant status with ICD X Diagnosis	2	0	0
3.	Gender with ICD X Diagnosis	13	0	0
4.	Types of services with ICD X Diagnosis	0	0	12.307
5.	Participant plan and participant number	0	0	36
6.	Participant status and age range	14	0	0

Information Technology

There is no single department in the IT field that is specifically responsible for maintaining the quality of the data in the system. Although for now the IT service team and business analyst work together in carrying out their duties as MIS and data engineers, this makes the team overwhelmed because there is a double task of the main task of the two teams in maintaining the existing data, sometimes only relying on findings from users if anomalous data is found when processing data and in terms of validation it also takes a long time.

Customer Retention

Anomalies in the date of birth data of participants and their family members caused by the absence of format locking on the date of birth from a file manually uploaded to the system caused an error in the participant's date of birth.

Operational

The data processing for the analysis material takes a long time because inconsistent and non-uniform data are found, especially in the outpatient or inpatient service detail codes caused by each hospital with different inputs does not have a uniform code for the service detail code.

Actuarial

It takes a long time in the data validation section because there is an inconsistency between participant data and claim data, so it must be validated by the data owner or IT who provides the data. This has an impact on determining premium extensions for participants who are not fast and analyzing claims for additional purposes of audit reports, as well as providing recommendations and information on inflation that occurs to the pricing team in designing a baseline percentage of one of the components of premium calculation and other reports needed by the actuarial team.

From the results of the observations and interviews, this data quality problem is caused by:

1. no team is fully responsible for the data. Although at this time the service IT team is appointed to be responsible for the data on the server, it is not optimal because of the double work carried out by the service IT team and business analysts, it adds to the tasks of the related team as the person in charge of data as a task as an MIS team or data engineer.
2. There are no rules for validation in the system. One of them is the existence of gender-inappropriate age data and inconsistencies in a criterion, as stated by the operational team that there is no format locking for the date of birth of the Excel file that is manually uploaded to the system, causing errors in the participant's date of birth.
3. There is no standardization of service detail code data from the data that comes in from the hospital to the company's system that has been provided. This happens because the data has not been standardized by the operational team to IT. This needs to be done immediately because uniform service details will help easier analysis for data working units such as underwriting and pricing teams and operations teams.

From point 1, there are no regulations in the company that govern the overall data management to avoid poor data quality.

Conclusion

After analyzing the data quality, it was found that the data quality problem arose due to the lack of awareness from the management and the absence of a data management team that was specifically responsible for all the data elements or outputs that would be generated from the company's data. This needs to be a special concern for management because insurance companies are financial companies like banks where data has become important to improve services and avoid fraud. From these problems, there are three suggestions to help insurance companies in improving the quality of their data, namely:

1. In answering the problems faced by the IT team, the IT team needs a special team to pay attention to data consisting of data engineers and data stewards. According to DMBOK, this data team has responsibility for all important elements in data management in the company where its tasks include creating data architecture, data quality standards, metadata, data file documentation against data definitions or criteria, updating master data or reference data, data warehouse, and business intelligence.
2. The next stage is if a data governance team has been formed, the company can use a vendor such as Collibra, or create its own system that will maintain and directly identify incoming data to maintain good data quality.
3. Because there is already a dashboard created, it is necessary to create a data warehouse in order to speed up the process of creating a dashboard or creating a data mart for every user who wants to use data from a direct database.

Bibliography

Askham, N., Cook, D., Doyle, M., Fereday, H., & Gibson, M. (2013). *The Six Primary Dimensions For Data Quality Assessment*. DAMA International.

Data Indonesia. (n.d.). *Survei Mayoritas Warga Indonesia Miliki Asuransi Kesehatan.*

David, Loshin. (2011). *The Practitioner's Guide to Data Quality Improvement*. Knowledge Integrity, Inc.

Haug, Anders, Zachariassen, Frederik, & van Liempd, Dennis. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, *4*(2), 168–193. https://doi.org/10.3926/jiem.2011.v4n2.p168-193

katadata. (n.d.). Ini Jenis Asuransi Swasta yang Paling Banyak Dimiliki Penduduk Indonesia.

Miguel, Diaz Iturry, Solange N., Alves Souza, Marcia, Ito, & Da Silva Suzana Alves. (2021). Data Quality in health records: A literature review. *Iberian Conference on Information Systems and Technologies, CISTI*, (June), 1–6. https://doi.org/10.23919/CISTI52073.2021.9476536

Nulhusna, Rizqa, Taufiq, Nur Fajar, & Ruldeviyani, Yova. (2022). Strategy to Improve Data Quality Management: A Case Study of Master Data at Government Organization in Indonesia. *Proceeding - 2022 International Symposium on Information Technology and Digital Innovation: Technology Innovation During Pandemic, ISITDI 2022*, 150–155. https://doi.org/10.1109/ISITDI55734.2022.9944466

Purnomoadi, Andreas, Sari, Ira M., Anna Maria, J. S., Fiddiansyah, Dimas B., Saputro, Nuriyanto E., & Sofan Hadi, M. S. (2023). A Method to Quantify Data Quality in Asset Health Indices Model. *Proceedings of 2023 4th International Conference on High Voltage Engineering and Power Systems, ICHVEPS 2023*, 16–20. https://doi.org/10.1109/ICHVEPS58902.2023.10257386

Rahmawati, Sinta Denovi, & Ruldeviyani, Yova. (2019). Data Quality Management Strategy to Improve the Quality of Worker's Wage and Income Data: A Case Study in BPS-Statistics Indonesia, 2018. *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*. https://doi.org/10.1109/ICIC47613.2019.8985803

Sunandar, Nandang, & Nizar Hidayanto, Achmad. (2022). Data Quality Management Improvement: Case Studi PT BPI. *2022 IEEE World AI IoT Congress, AIIoT 2022*, 53–58. https://doi.org/10.1109/AIIoT54504.2022.9817195

Wahyudi, Tedi, & Isa, Sani Muhammad. (2023). Data Quality Assessment Using Tdqm Framework: a Case Study of Pt Aid. *Journal of Theoretical and Applied Information Technology*, *101*(9), 3576–3589.

Wijayanti, Wiluyaningtyas, Hidayanto, Achmad Nizar, Wilantika, Nori, Adawati, Infaz Rizki, & Yudhoatmojo, Satrio B. (2018). Data quality assessment on higher education: A case study of institute of statistics. *2018 International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2018*, 231–236. https://doi.org/10.1109/ISRITI.2018.8864476

Xu, Dongting, Zhang, Zhisheng, & Shi, Jinfei. (2022). A Data Quality Assessment and Control Method in Multiple Products Manufacturing Process. *2022 5th International Conference on Data Science and Information Technology, DSIT 2022 - Proceedings*, 1–5. https://doi.org/10.1109/DSIT55514.2022.9943883

Zarina Abdul Jabar, Zam, Wook, Muslihah, Zakaria, Omar, Ramli, Suzaimah, & Afiza Mat Razali, Noor. (2022). Establishment of Big Data Analytics Application Model for Malaysian Public Sector: an Expert Validation. *Edpacs*, *65*(4), 1–20. https://doi.org/10.1080/07366981.2021.1958736