ISSN 2355-0082

# Data Quality Issues : Case Study of Claim and Insured in Indonesia Insurance Company

Chris Solontio[1], Achmad Nizar Hidayanto[2]
[1,2] Department of Information Technology, Universitas Indonesia, Jakarta Pusat, Indonesia
[1] chris.solontio21@ui.ac.id, [2] nizar@sci.ui.ac.id

Accepted 24 August 2024
Approved 7 November 2024

***Abstract*— Data has become an asset for insurance companies that have many benefits and management needs to realize the importance of 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 data dimensions, namely, accuracy, completeness and consistency of the relationship between claim data and insured, and findings from each data fields in this case study. In addition, researchers conducted interviews to find out the obstacles faced by IT, Customer Retention, Operational, and Actuary teams where they are directly related to data flow and data processing. From the results of the analysis, there is invalid data that will affect the analysis and cause obstacles faced by users according to the interview results. In the conclusion, management needs to form a data governance team to avoid poor data quality that has responsibility for data flow and maintains data quality in order to provide a positive impact such as providing the right data accuracy in data analysis and user time to be more effective in data processing, assisting in making data warehouses, applying AI and digital transformation as a form of improvement in the services provided.**

***Index Terms*— Data Management; Data Quality; Insurance; Data Quality Dimension**

## I. Introduction

According to the results of the populix survey, the majority of Indonesia's population already has BPJS Kesehatan insurance of 83% and private insurance of 38% [1]. One of the most owned private insurances is health insurance [2]. Insurance companies are now carrying out many development initiatives in the IT world such as creating a data warehouse, carrying out digital transformation such as using chatbots or processing claims quickly with the use of artificial intelligence (artificial intelligence), or the use of data analytics and data science to be able to compete between insurance companies and provide the best service for customers. This initiative can be carried out well, inseparable from the importance of a good company data condition where it must start paying attention to data management so that there is no more unreliable data.

The life insurance company that is the place of this research case study is one of the largest life insurance companies in Indonesia by providing health protection for employees of other companies and has a wide network of providers. In addition, this insurance company has two main health insurance products, namely indemnity and managed care. Of the two products, this managed care product will be the source of data used in assessing the quality of the company's data. Managed care is health insurance products that have a different process from indemnity, namely the customer must start the first treatment at a first-level health facility and then get a referral to the hospital if needed for outpatient or inpatient treatment in accordance with the plan and provisions in the policy which is an agreement between the insurance company and other companies.

In facing today's business challenges, the company has the initiative to improve its services to be the best and help management quickly in making 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 carried out are the creation of a data warehouse, and the use of 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 and inconsistent data. Meanwhile, initiatives that have not been implemented cannot be carried out due to the condition of the data. From this, there are obstacles in data management that need to be considered, especially in the quality of the data itself.

From the obstacles faced, the purpose of this study is to evaluate the impact of poor data quality on the company's operations, and customer satisfaction. Besides that, the research question that arises for this research, that is:

1. how to deal with the issue of existing data quality?
2. What quality dimensions are suitable for this insurance industry?.

One of the components of the data management framework that will be discussed is data quality [3]. Data quality can result in efficient operational processes, decision-making, data warehouse creation, and have a positive influence on customer satisfaction [4]. In maintaining data quality, every company needs to define data dimensions with the aim of knowing the impact of poor data quality on costs, reputation, compliance regulations and so on [5]. 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 [6]. Hidden fees (hidden cost) and direct costs (direct cost) of the effects of poor quality data on insurers in figures 1.

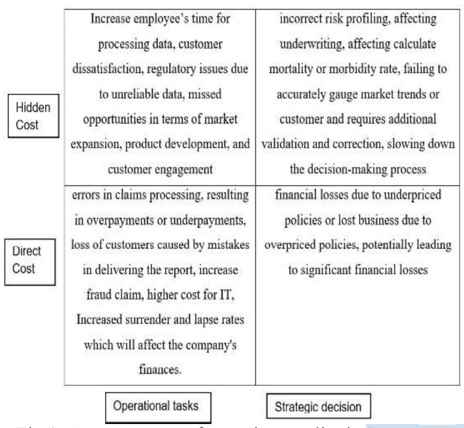


Fig 1. Consequences of poor data quality in Insurance

In Data Quality for the Information Age, Thomas Rednan formulates 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 quality and dimensions as a tool to measure data quality in an organization, many researchers conduct research on this research by conducting literature reviews or with case studies in several business lines such as Table I.

TABLE I. DATA QUALITY RESEARCH AREA

| **Research Areas of Data Quality** | **References** |
|---|---|
| Government | Government **O**rganization **[7]**, BPS-Statistics Indonesia **[8]**, Malaysian Public **[9]**, State Electricity Company **[10]**. |
| Financial Industry | PT BPI **[11]**. |
| Airport Services | PT JAS **[12]**. |
| Education | Institute of Statistic **[13]**. |
| Factory | Paper Factory **[14]**. |
| Medical or Health | Electronic Medical Record **[15]**. |

From table I, there is still limited research on quality data in insurance companies. Based on the table of impacts caused by poor quality, poor data quality will affect insurance companies, namely increased employee time to process data, customer dissatisfaction, regulatory problems due to unreliable data, slowing down the decision-making process, resulting in overpayment or underpayment, customer loss caused by errors in submitting reports, and financial losses due to policies that are not too cheap or lose business due to overly expensive policies. Based on research conducted by Chen, assessing the quality of data in the consistency dimension between MIS social insurance data and MIS workforce has the goal of improvising in terms of integrating the two databases [16]. The paper presents an evaluation using conflict classification methods, especially intra-concept conflicts and inter-concept conflicts. Intra-concept conflicts are related to data quality based on variations in data values, such as the appearance of the name "Liu Ming" in social insurance data and MIS Employment data. On the other hand, inter-concept conflict refers to the assessment of data quality exemplified by the inconsistency of the employment status of individuals between the insurance database of non-workers in the company and the database of MIS Workers who are currently employed in the company.

The next research was carried out by Haryadi, an assessment of the data quality of top banks and insurance in Europe with the aim of creating a big data process [17]. In the study, each assessment of the data quality data dimension differed between banks and insurance, where for insurance companies, Allianz data quality was carried out to see the accuracy and timeliness. The next research conducted by Mary discussed the quality of claims database data against Electronic Health Record (EHR) data which was carried out by dividing 9 roles for the relationship between data attributes to see the number of invalid data based on the compatibility dimension, completeness, and the value of the date range for which the results are still found invalid [18]. Thus, the answer to the second research question is that the most suitable data quality dimensions for the insurance industry are accuracy, completeness, and consistency.

The data quality dimension to be used based on the number of quality dimensions that will be used [15] and the results of discussions with the head of the IT division, namely completeness, consistency, and accuracy. In assessing the quality of this data, the study will apply the method used in Mary's paper by forming several roles to assess the quality of the data by seeing if any data is found invalid with several roles that have been determined. Insurance is one of the things that can be categorized in the financial services, it is the same as a bank where data quality is very important so that this research can be an aid to determine the dimensions and assess the quality of data from the characteristics of the data in insurance.

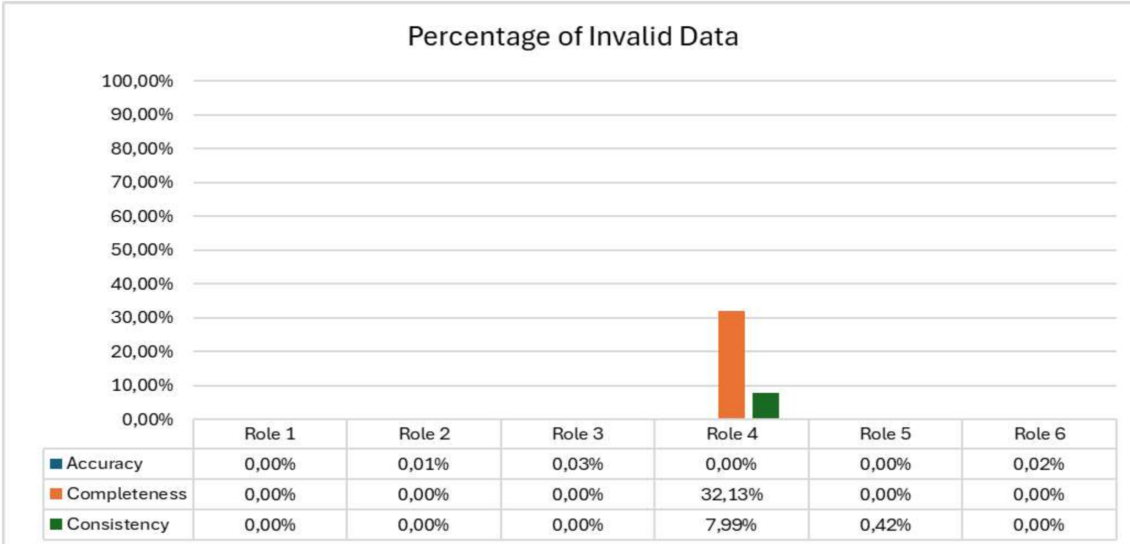


Fig 1. Consequences of poor data quality in Insurance

## II. RESEARCH METHODOLOGY

In this study, qualitative and quantitative research design will be carried out. The quantitative method will be carried out by observation with profiling data in order to find out the results of data assessments that are not valid from the data quality dimension that has been determined by forming 6 roles between two

claim data and participants [18]. Meanwhile, qualitative interviews will be conducted to units that are specifically adjacent to the use of data to add supporting data related to the problems found from the results of the assessment and the consequences generated in the company's activities.

In data processing, it will use excel and python to view invalid data from participant details and claim data. Before conducting a data assessment of the quality data dimension, 6 roles must be formed that are relevant to the interconnectedness of the two data, namely:

TABLE II. SIX ROLES TO ASSES DATA QUALITY

| Roles | Descriptions |
|---|---|
| Roles 1 | Service Date and Discharge Date for Inpatient in Claim Data. |
| Roles 2 | Participants status with ICD X in claim data. |
| Roles 3 | Gender and ICD X. |
| Roles 4 | Types of Services with ICD X. |
| Roles 5 | Participant Plan and Participant Number. |
| Roles 6 | Participant Status and Age Range. |

## III. RESULT AND DISCUSSION

### A. Result

The limitation of this study is that the claim data analyzed is claim data provided from the case study location, namely the provider history (hospital/clinic) in 2022 and the amount of data provided is 125 thousand data consisting of 38 thousand SJP (letter of guarantee for patients after health services are carried out at a hospital or clinic) data, 9 thousand participant claims and three providers.

Based on the six roles that have been formed from claim and participant data to assess the quality of the data, invalid data is found from each data dimension from roles 2 to 6 that have been formed. From figure 2, it can be concluded in detail as follows:

- Roles 1 does not find invalid data on the receipt date and delivery date data by looking at the length of stay.
- Roles 2 were found to be 2 SJPs invalid with information on the status of child participants but the diagnosis was Z34 which is a normal pregnancy surveillance diagnosis.
- Roles 3 based on a list of diagnoses per gender based on ICD X on ICD10CM website, found that 13 SJPs were invalid in data claiming a link between gender and diagnosis (Table IV).

TABLE III. INVALID DATA ON ICD X BY GENDER

| Gender | Invalid Diagnosis |
|---|---|
| Male | O74, N91, N89, N80, dan N93 |
| Female | C61, D17.6, dan N47 |

- Roles 4 types of services with 32% (12,307 SJP) incompleteness diagnoses where the service does not have details of the type of medication administered. In addition, there are 8% inconsistent data due to the writing of the type of service that does not match and there are differences in spaces, for example, anasthesi periodal operation actions (caesar / sectio) and anasthesia periodal operation actions (caesar / sectio), dengue ns 1 ag and dengue ns1 antigen.

- Roles 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.
- Roles 6 were found to be 14 participants whose age ranges did not match their membership status (Table V).

TABLE IV. INVALID DATA IN AGE RANGE BASED ON PARTICIPANT STATUS

| Status | Age Range | Total |
|---|---|---|
| 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 actuary with the aim of finding out the effects and consequences of poor data quality. From the results of the interview, several points were found as follows:

- Department of 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, but this makes the team overwhelmed because there are dual tasks from the main tasks of the two teams so that 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. In addition, another problem is the absence of regulations regarding data management and not knowing how to measure data quality.

- Customer Retention Department

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.

- Operations Department

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 and do not have a uniform code for the service detail code.

- Actuary Department

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 in determining premium extensions for customers who are not fast and analyzing Claims for additional 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 observations and interviews that have been conducted, it is found that there are data quality problems that must be faced due to two things, namely first, the absence of a team responsible for the data entering the database and the absence of regulations and an overview of the upstream to downstream process in the flow of data into the database.

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

*B. Discussion*

This paper will focus on data quality where companies must start to raise awareness of the importance of avoiding poor data quality where the consequences are very significant impact on both internal and external as explained in figure 1 in the introduction. Despite limitations in the data provided by the case study site (2022 data), connecting data analysis with interview findings reveals that the lack of a dedicated team for data management can result in declining data quality. Furthermore, a data warehouse is essential to resolve inconsistencies in health service classifications, which is a primary responsibility of a specialized data team.

Therefore, this answers the first research question that management should establish a dedicated data team to address current and future data quality issues. To effectively define this team's duties and align them with industry standards, insurance companies can refer to the DAMA-DMBOK (Data Management Body of Knowledge) framework. This framework provides a structured approach for implementing key data governance tasks, ultimately ensuring data quality while leaving data security responsibilities with the IT security team.

Based on this analysis, management can establish a data governance team composed of data governance leads, data stewards, and data engineers. This team will have four primary objectives: developing a data quality framework focused on the dimensions of accuracy, completeness, and consistency; implementing data lifecycle management to oversee data collection, processing, storage, usage, and disposal; defining data ownership by assigning clear accountability for data quality within specific departments and Work with the compliance team to create internal and external data use policies that must comply with personal data protection regulations (PDPs) Personally Identifiable Information) which has been regulated in Law No. 27 of 2022 [19].

## IV. Conclusion

In conclusion, this study identified significant data quality challenges within the insurance company, primarily due to a lack of management awareness and the absence of a dedicated data governance team. These issues impact both operational efficiency and customer satisfaction, which are critical to the company's competitive advantage in the insurance industry. Given the critical role data quality plays in supporting accurate analysis, timely decision-making, and regulatory compliance, it is essential for the company to adopt structured data governance practices.

To address these issues, the following recommendations are proposed:

- Establishing a Dedicated Data Governance Team:

This team should consist of data engineers, data stewards, and data governance leaders, with each role having clear responsibilities as in the discussion section or can be based on a data governance framework such as DAMA-DMBOK. This structure will ensure that all data management functions, from quality monitoring to data ownership, are systematically overseen.

- Leveraging Data Governance Tools:

Once the governance team is in place, the company can implement tools (e.g., Collibra) to streamline data quality checks, automate data profiling, and ensure consistency across datasets. Such tools will aid in early detection of quality issues, supporting proactive improvements in data accuracy and completeness.

- Building a Data Warehouse for Enhanced Accessibility:

Creating a centralized data warehouse will simplify data access and reporting for all users, reducing time spent on validation and enabling faster, data-driven insights. This infrastructure will also support future advancements, such as dashboards and AI-driven analytics, by providing a reliable data foundation.

By implementing these recommendations, the insurance company can significantly enhance data quality, improve customer satisfaction, and foster more reliable decision-making processes. This approach will position the company for sustainable growth and ensure readiness for future digital transformation initiatives.

## References


[1] Data Indonesia, "Survei Mayoritas Warga Indonesia Miliki Asuransi Kesehatan.", [Online]. Available: https://dataindonesia.id/varia/detail/survei-mayoritas-warga-indonesia-miliki-asuransi-kesehatan

[2] katadata, "Ini Jenis Asuransi Swasta yang Paling Banyak Dimiliki Penduduk Indonesia," 2023. [Online]. Available: https://databoks.katadata.co.id/datapublish/2023/02/01/ini-jenis-asuransi-swasta-yang-paling-banyak-dimiliki-penduduk-indonesia

[3] I. DAMA, *2017 DAMA-DMBOK Data Management Body of Knowledge, 2nd Edition-1*. Technics Publications, 2017.

[4] L. David, *The Practitioner's Guide to Data Quality Improvement*. Knowledge Integrity, Inc., 2011.

[5] N. Askham, D. Cook, M. Doyle, H. Fereday, and M. Gibson, *The Six Primary Dimensions For Data Quality Assessment*. DAMA International, 2013.

[6] A. Haug, F. Zachariassen, and D. van Liempd, "The costs of poor data quality," *J. Ind. Eng. Manag.*, vol. 4, no. 2, pp. 168–193, 2011, doi: 10.3926/jiem.2011.v4n2.p168-193.

[7] R. Nulhusna, N. F. Taufiq, and Y. Ruldeviyani, "Strategy to Improve Data Quality Management: A Case Study of Master Data at Government Organization in Indonesia," *Proceeding - 2022 Int. Symp. Inf. Technol. Digit. Innov. Technol. Innov. Dur. Pandemic, ISITDI 2022*, pp. 150–155, 2022, doi: 10.1109/ISITDI55734.2022.9944466.

[8] S. D. Rahmawati and Y. Ruldeviyani, "Data Quality Management Strategy to Improve the Quality of Worker's Wage and Income Data: A Case Study in BPS-Statistics Indonesia, 2018," *Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019*, 2019, doi: 10.1109/ICIC47613.2019.8985803.

[9] Z. Zarina Abdul Jabar, M. Wook, O. Zakaria, S. Ramli, and N. Afiza Mat Razali, "Establishment of Big Data Analytics Application Model for Malaysian Public Sector: an Expert Validation," *Edpacs*, vol. 65, no. 4, pp. 1–20, 2022, doi: 10.1080/07366981.2021.1958736.

[10] A. Purnomoadi, I. M. Sari, J. S. Anna Maria, D. B. Fiddiansyah, N. E. Saputro, and M. S. Sofan Hadi, "A Method to Quantify Data Quality in Asset Health Indices Model," *Proc. 2023 4th Int. Conf. High Volt. Eng. Power Syst. ICHVEPS 2023*, pp. 16–20, 2023, doi: 10.1109/ICHVEPS58902.2023.10257386.

[11] N. Sunandar and A. Nizar Hidayanto, "Data Quality Management Improvement: Case Studi PT BPI," *2022 IEEE World AI IoT Congr. AIIoT 2022*, pp. 53–58, 2022, doi: 10.1109/AIIoT54504.2022.9817195.

[12] T. Wahyudi and S. M. Isa, "Data Quality Assessment Using Tdqm Framework: a Case Study of Pt Aid," *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 9, pp. 3576–3589, 2023.

[13] W. Wijayanti, A. N. Hidayanto, N. Wilantika, I. R. Adawati, and S. B. Yudhoatmojo, "Data quality assessment on higher education: A case study of institute of statistics," *2018 Int. Semin. Res. Inf. Technol. Intell. Syst. ISRITI 2018*, pp. 231–236, 2018, doi: 10.1109/ISRITI.2018.8864476.

[14] D. Xu, Z. Zhang, and J. Shi, "A Data Quality Assessment and Control Method in Multiple Products Manufacturing Process," *2022 5th Int. Conf. Data Sci. Inf. Technol. DSIT 2022 - Proc.*, pp. 1–5, 2022, doi: 10.1109/DSIT55514.2022.9943883.

[15] D. I. Miguel, A. S. Solange N., I. Marcia, and Da Silva Suzana Alves, “Data Quality in health records: A literature review,” *Iber. Conf. Inf. Syst. Technol. Cist.*, no. June, pp. 1–6, 2021, doi: 10.23919/CISTI52073.2021.9476536.

[16] W. L. Chen, S. D. Zhang, and X. Gao, “Anchoring the consistency dimension of data quality using ontology in data integration,” *2009 6th Web Inf. Syst. Appl. Conf. WISA 2009*, pp. 201–205, 2009, doi: 10.1109/WISA.2009.32.

[17] A. F. Haryadi, J. Hulstijn, A. Wahyudi, H. Van Der Voort, and M. Janssen, “Antecedents of big data quality: An empirical examination in financial service organizations,” *Proc. - 2016 IEEE Int. Conf. Big Data, Big Data 2016*, pp. 116–121, 2016, doi: 10.1109/BigData.2016.7840595.

[18] M. A. Gadde, Z. Wang, M. Zozus, J. B. Talburt, and M. L. Greer, “Rules based data quality assessment on claims database,” *Stud. Health Technol. Inform.*, vol. 272, pp. 350–353, 2020, doi: 10.3233/SHTI200567.

[19] R. Indonesia, “Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi,” *Ditama Binbangkum - BPK RI*, no. 016999, pp. 1–50, 2022, [Online]. Available: https://peraturan.bpk.go.id/Home/Details/229798/uu-no-27-tahun-2022