SOA Research Institute

Actuarial Innovation and Technology

Alternative Data Usage in Life and Health Insurance | Evidence from Australia

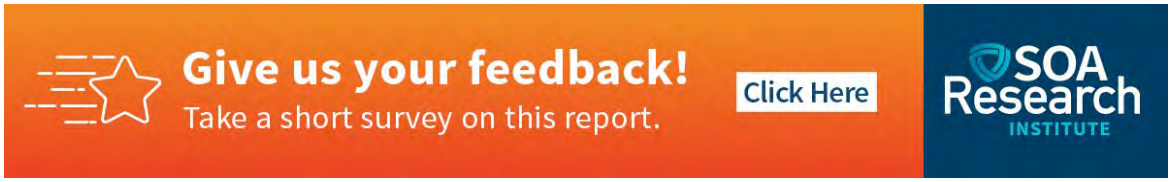
October | 2023

Alternative Data Usage in Life and Health Insurance | Evidence from Australia

AUTHORS Zhan Wang, FIAA, CERA

Han Li, PhD., AIAA

SPONSOR Actuarial Innovation and Technology Strategic Research Program Steering Committee



CONTENTS

Alternative Data Use in Life and Health Insurance | Evidence from Australia

Use of this Report

The comments and views presented in this report are summarized from the interviews and do not necessarily represent the views of the organizations for whom the participants work. It is important to note that these comments have not been verified by the researchers, and not all comments from the interviews are included in this report.

"In the report, quotes from the participants are shown in this format."

Executive Summary

Alternative data sources beyond policy in-force data and claims data are increasingly seen as a valuable resource for life and health insurance. These non-traditional data sources can have advantages due to, among other factors, the speed of data collection, increased data accuracy and granularity, and the provision of information on lifestyle and behavioral characteristics. However, such data sources must be carefully evaluated to ascertain if the benefits outweigh the costs and by how much. This research utilized an interview-based approach that delves into the utilization of alternative data in the Australian life and health insurance markets.

To provide context for the interviews that took place, we explained the key elements of the Australian life and health insurance markets, along with product characteristics. Key observations included tight regulations, limited range of products, concentrated market share, and industry sustainability concerns in recent times.

Drawing from one-on-one interviews conducted with 20 experienced practitioners in the field, we present a comprehensive overview of commonly used alternative datasets categorized into five main groups with detailed explanations on some highlighted case studies. Furthermore, we engaged in discussions regarding the viewpoints collected, covering potential opportunities, existing challenges, and the future prospects of alternative data usage within the Australian insurance industry.

We draw the following conclusions based on the interview responses:

- Public non-personal data are the most used datasets by practitioners in the Australian life and health insurance sectors, although there is a tendency and willingness to also utilize non-public personal data.
- The objectives of using alternative data include customer engagement and interaction, product development, experience monitoring, segmentation analysis, and competitor analysis.
- Besides data availability, other key challenges in using alternative data encompass management resource constraints and a lack of a comprehensive framework for utilizing alternative data.

While acknowledging that the development of alternative data usage might transpire gradually over time, we are enthusiastic about witnessing the collaboration of the Australian insurance sector as a whole to harness the power of data for the greater good of society. Some of the key takeaways from this research are summarized below:

- The changing landscape in data and data sharing is inevitable, and this has potential to improve the insurance industry for both the insurers and policyholders.
- The development of alternative data usage may occur gradually and incrementally through trial and error.
- There should be a secure system in place to handle commercially sensitive information and ensure the privacy and security of the data.
- It is crucial for the insurance industry to collaborate and harness the power of data for the greater good of society, ensuring that data are used ethically and responsibly, respecting the privacy and rights of individuals, and protecting vulnerable segments of the population. .

Section 1: Introduction and Interview Methodology

In Australia, the life and health insurance industries have been predominantly relying on policy in-force and claims data for many decades. These datasets provide useful insights to the characteristics and experience of the insured population, which are important input into underwriting, pricing, and profitability management. However, there are certain limitations to these datasets, including:

- Infrequent data collection points: most data are collected at policy commencement and/or at time of claims. This is especially the case with Yearly Renewable life insurance policies. Changes in policyholders' circumstances are not updated frequently in the datasets.
- Low data quality: life and health insurance data are often collected in an ad hoc and unstructured manner, for example, claims data may include multiple interviews (both in person and over the phone), lengthy case notes and unstructured texts, which are difficult to analyze and draw insights from.
- Limited information about other characteristics: traditional datasets contain information relevant to life and health insurance policies, while little is known, for example, about policyholders' lifestyle and behavioral characteristics.

Access to accurate and timely data is vital for the insurance industry. With the help of technological advancement, alternative data has been used increasingly in the insurance industry, especially in General Insurance. In Australia, for example, some auto insurance companies are actively promoting pricing factors based on a wider range of individual characteristics and the way people using their car (see youi[1], "Insurance for Individuals"). Life and health insurance companies have also started to explore this field in recent years. Life insurers are looking to use electronic health records, biometrics, and genomics to replace or supplement the data collected in the traditional underwriting process. Recently, the use of these alternative datasets has been accelerated due to government lockdown measures imposed during the COVID-19 pandemic. Insurers also introduced complimentary wellness programs to support policyholders' wellbeing, through which fitness and other lifestyle information could be tracked and collected (see AIA Vitality[2] and MLC on Track[3] as examples).

While it is evident that certain alternative datasets can improve insurance pricing and underwriting process and improve customer experience, the use of alternative data is still an emerging field for life and health insurers. Take the Australian life insurance industry as an example. Most insurers only recently started investing in data warehouse and analytical tools. Moreover, utilizing alternative data generally comes at a price: besides potential data quality issues, there are also privacy and ethical concerns involved with alternative data. Moreover, the risk and potential costs of a data breach are significant and could impact future franchise value. The collection, processing, and analysis of data may also require additional resources from the insurance companies. Enhancements to the cost benefit analysis of such data could be beneficial to identify and quantify both the strengths and weaknesses of different alternative datasets, and guide insurers on whether and how alternative data should be applied. For this project, we employed an interview-based approach to explore and evaluate different alternative datasets used in the life and health insurance industries.

Forty-one professionals working in life insurance or health insurance related companies were first approached through an initial email invite, which briefly described the context, purpose, and approach of this study. Further communication on more details of the study was provided, and 20 participants were successfully interviewed.

[1] www.youi.com.au
[2] https://www.aia.com.au/en/individual/aia-vitality.html
[3] https://www.mlcinsurance.com.au/about-us/media/mlc-life-insurance-launches-next-phase-of-award-winning-health-and-wellness-program

All interviewees were furnished with an interview questionnaire prior to the scheduled interview for better preparation. Interviews were conducted either in person or via virtual meetings, each time with one participant and two interviewers. Interviews started with a short introduction of the study to recap the purpose and deliverable of the study. Interviewees' responses were recorded by interviewers and notes were later consolidated and sent to interviewees for confirmation. Confirmed responses were summarized and analyzed to draw relevant insights.

Section 2: Market Overview

It is important to understand the context of the interviews that took place. This section explains the key elements of the Australian life and health markets landscape and product characteristics.

The life insurance market in Australia is dominated by a few large insurers with a limited range of products and distribution models. The industry is highly regulated. In recent years, some segments of the market have experienced substantial losses, leading to significant merger-and-acquisition activities.

While there are more insurers in the Australian Private Health insurance market, the market share is also dominated by a few large ones. The industry is tightly regulated by the "community rating" rule[4]. While product coverages are largely similar, insurers intend to distinguish themselves through different distribution channels and targeted marketing campaigns.

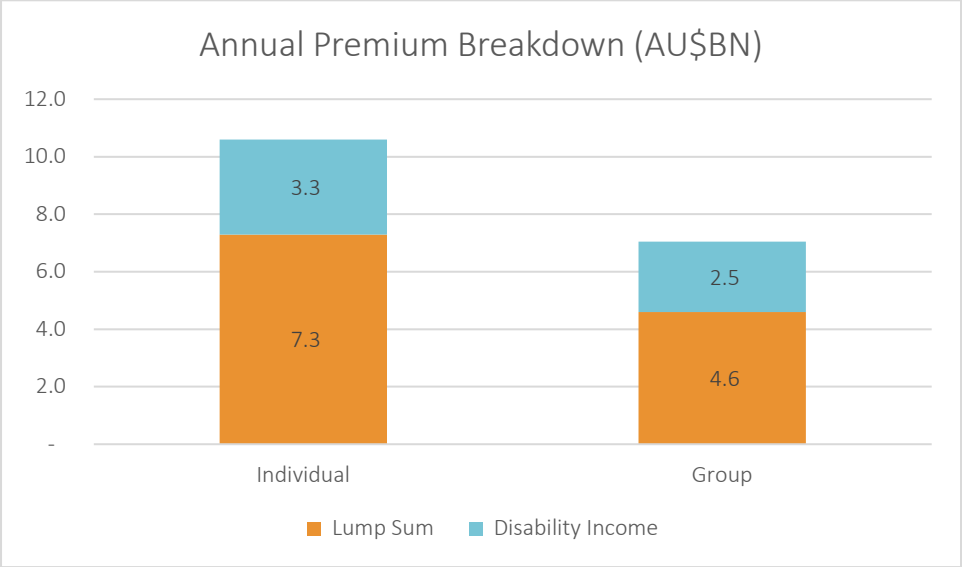
2.1 THE AUSTRALIAN LIFE INSURANCE MARKET

2.1.1 TYPES OF PRODUCTS AND DISTRIBUTION CHANNELS

The main products offered are Yearly Renewal Term policies, which cover risks of death, disability, and critical illness. These products can be categorized into "Lump Sum" and "Disability Income" benefits. "Lump Sum" (LS) benefits pay claims upon Death, Total and Permanent Disability (TPD), and diagnosis of Trauma conditions (Trauma), while "Disability Income" (DI) benefit provides a regular payment in the event of either permanent or temporary disability.

Figure 1

TOTAL ANNUALIZED PREMIUMS INFORCE BREAKDOWN BY DISTRIBUTION CHANNEL AND PRODUCT TYPE AS OF DECEMBER 2022



[4] Definition of community rating: A rule that prevents health insurers from varying premiums within a geographic area based on age, gender, health status or other factors.

The main distribution channels in Australia can be categorized into "Individual" and "Group." "Individual," or "Retail," refers to policies sold through financial advisers or direct marketing, while the "Group" segment consists mainly of insurance offered to members of superannuation (pension) funds.

As of December 2022, the Annualized Premium In force (API) was more than AU$18 billion, with AU$11 billion for the "Individual" segment and AU$7 billion for the "Group" segment[5]. The split of API into LS and DI is shown in figure 1.

Death, TPD, Trauma, and DI benefits are usually referred to as "Risk Products" collectively. While there are products with participation features, they are mostly closed books and no longer open to new business. Annuity products are also available, however, the overall coverage is low. The rest of this section will focus on "Risk Products."

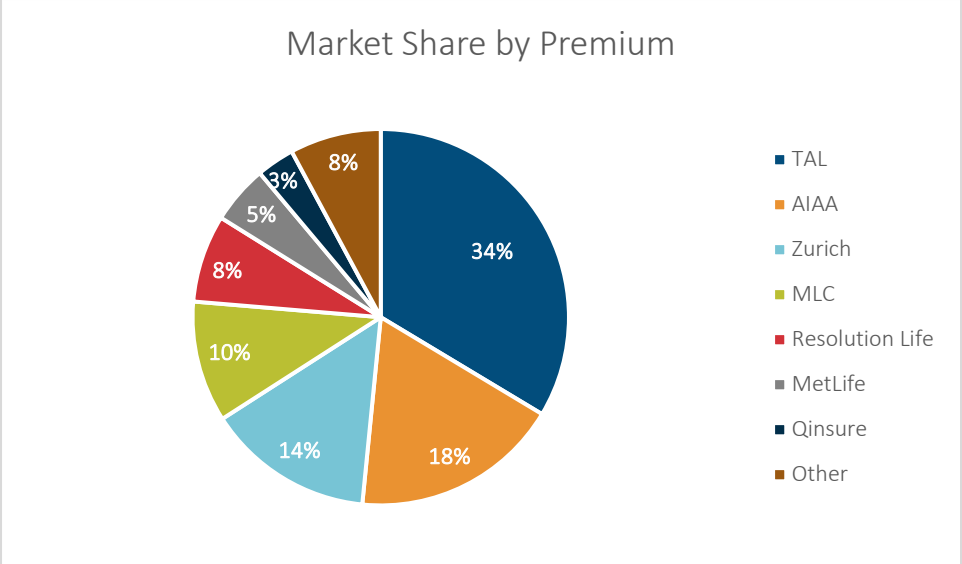
2.1.2 MARKET SHARE

There are 17 registered Life Insurance companies, seven Reinsurance companies, and ten Friendly Societies[6] in the Australian market[7]. The Life Insurance Act 1995 is the overarching legislation, and the industry is regulated by the Australian Prudential Regulation Authority (APRA).

The Risk Product market is dominated by a few large insurers, especially after significant merger-and-acquisition activities in the last ten years. The top five insurers, TAL, AIA Australia, Zurich Australia, MLC, and Resolution Life have more than 80% of the market share, measured by API. Other smaller insurers make up less than 10% of the market share. See figure 2 for more information on market share.

Figure 2

MARKET SHARE MEASURED BY API AS OF DECEMBER 2022



[5] https://www.apra.gov.au/life-insurance-claims-and-disputes-statistics

[6] Friendly Societies were established by community groups in the 1830s and evolved into member-focused providers of financial services, healthcare, retirement living, aged and home care services, transport, pharmacies and other fraternal services. They provide savings, investment and insurance products.

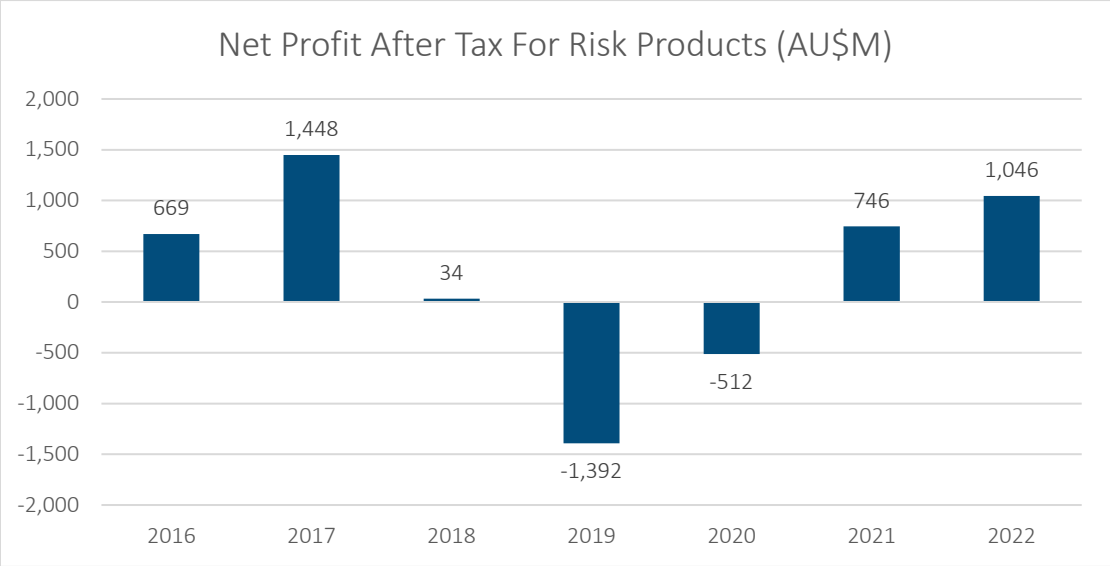
[7] https://www.apra.gov.au/registers-of-life-insurance-companies-and-friendly-societies

2.1.3 RECENT INDUSTRY PROFITABILITY

Several participants noted that senior management's commitments to invest in alternative data are limited due to resource constraints. They pointed out that the Australian life insurance industry has faced profitability and sustainability challenges in recent years, which has made it challenging to systematically invest in alternative data usage and analytics. As shown in figure 3, the Australian Life Insurance Industry experienced volatile profitability in recent years for these Risk Products and reported significant losses in 2019 (-AU$1.4 billion) and 2020 (-AU$0.5 billion) as shown by APRA statistics. The Net Profit After Tax was AU$2.0 billion for the seven-year period[8].

Figure 3

NET PROFIT AFTER TAX FOR RISK PRODUCTS - YEAR ENDING DECEMBER 2016 TO DECEMBER 2022



2.2 THE AUSTRALIAN PRIVATE HEALTH INSURANCE MARKET

There are 34 private health insurance providers in the Australian market, regulated by the Australian Prudential Regulation Authority (APRA). The Australian private health insurance industry is concentrated and dominated by a few large insurers, with over 80% of the market share held by the top five insurers.

The Australian private health insurance industry is tightly regulated, which has, to a significant degree, limited the room for pricing differentiation of private health insurance products. Examples of these regulations include guaranteed acceptance, where the private health insurer cannot reject any policyholders regardless of their health status, and community rating, where the private health insurer cannot price differently based on an individual policyholder's age, health condition, or claim history. To improve sustainability of the industry, regulators have also imposed risk equalization, which transfers funds from private health insurers with lower-than-average claim costs to insurers with higher-than-average claim costs, further supporting the implementation of community rating.

Based on quarterly private health insurance industry performance statistics released by APRA, the industry reported a net profit after tax of AU$1.2 billion with a premium revenue of AU$26.9 billion by the end of December 2022. The

[8] https://www.apra.gov.au/quarterly-life-insurance-performance-statistics

reported profitability decreased by 34% (or AU$0.6 billion) from 2021 to 2022[9]. See table 1 for additional information.

Table 1
KEY PERFORMANCE METRICS FOR THE YEAR ENDING DECEMBER 2022

	Year Ending December 2021	Year Ending December 2022	Yearly Change
Premium Revenue	AU$26.4 bn	AU $26.9 bn	1.8%
Fund Benefits (Claims)	AU $21.9 bn	AU $22.2 bn	1.6%
Gross Margin	17.2%	17.3%	+0.1 pp
Management Expenses	AU $2.5 bn	AU $2.7 bn	9.2%
Net Margin	7.7%	7.1%	-0.6 pp
Net Profit After Tax	AU $1.8 bn	AU $1.2 bn	-33.6%

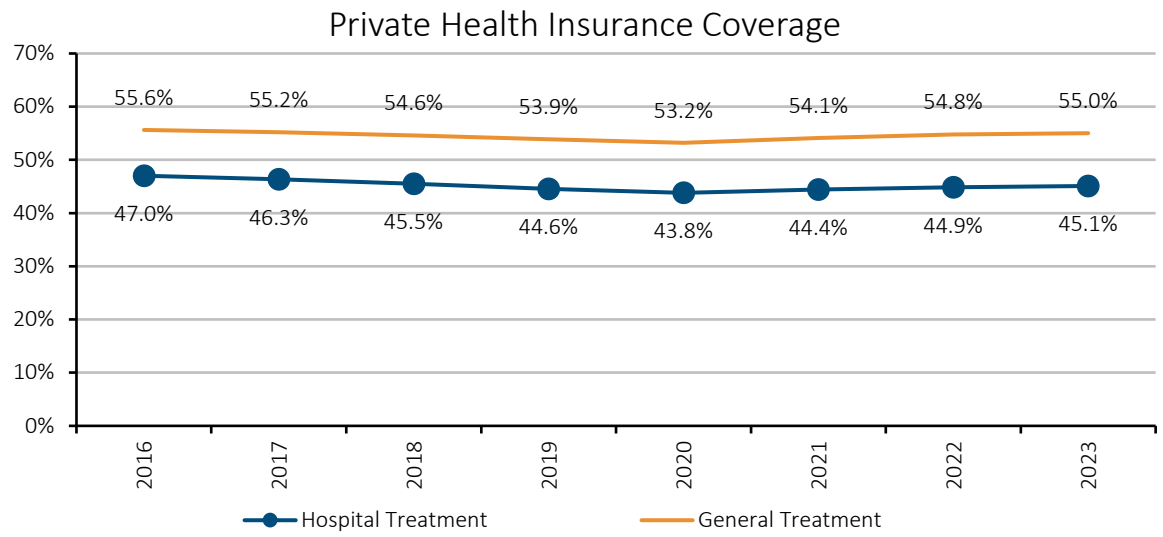
The main private health covers offered are hospital covers (or policies) and extra covers (or polices). As suggested by the cover names, hospital cover pays claims for hospital treatments such as the use of an operating theater, in-patient accommodations, rehabilitation, hospital psychiatric services, and palliative care, while extra cover pays claims for general treatments received such as general dental, chiropractic, and emergency ambulance services.

Figure 4 shows the private health insurance coverage for the year ending March 2016 to March 2023[10]. As of December 2023, over 45% of the Australian population has obtained hospital treatment coverage, and 55% of the population has obtained general treatment coverage.

[9] https://www.apra.gov.au/news-and-publications/apra-releases-quarterly-private-health-insurance-statistics-for-december-2022
[10] https://www.apra.gov.au/quarterly-private-health-insurance-statistics

Figure 4

PRIVATE HEALTH INSURANCE COVERAGE FOR YEAR ENDING MARCH 2016 TO MARCH 2023



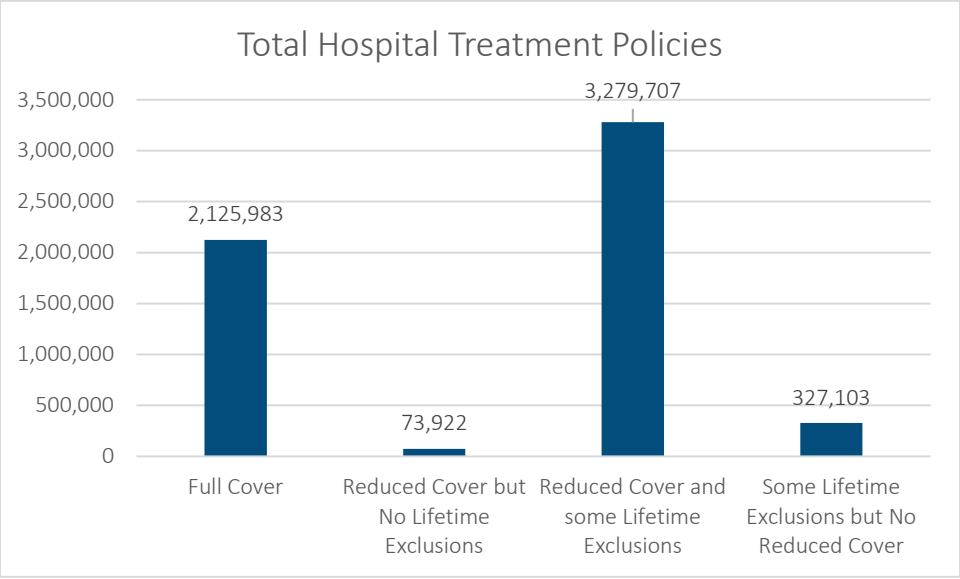
2.2.1 TYPES OF PRODUCTS

While product coverages are largely similar, insurers try to distinguish themselves through distribution channels and targeted marketing campaigns. For example, they may offer different cover packages, such as single cover, couples cover, family cover, and single parents cover, allowing flexibility over the selection of coverage levels to target specific policyholder cohorts. Insurers can increase competitiveness by partnering with other industries to offer signup benefits, such as cash rebate, waive of waiting periods, rewards from partnered companies and more.

As suggested by APRA quarterly private health insurance membership and benefit statistics (see figure 5), the most popular total hospital treatment cover is reduced cover with some lifetime exclusions, followed by full cover.

Figure 5

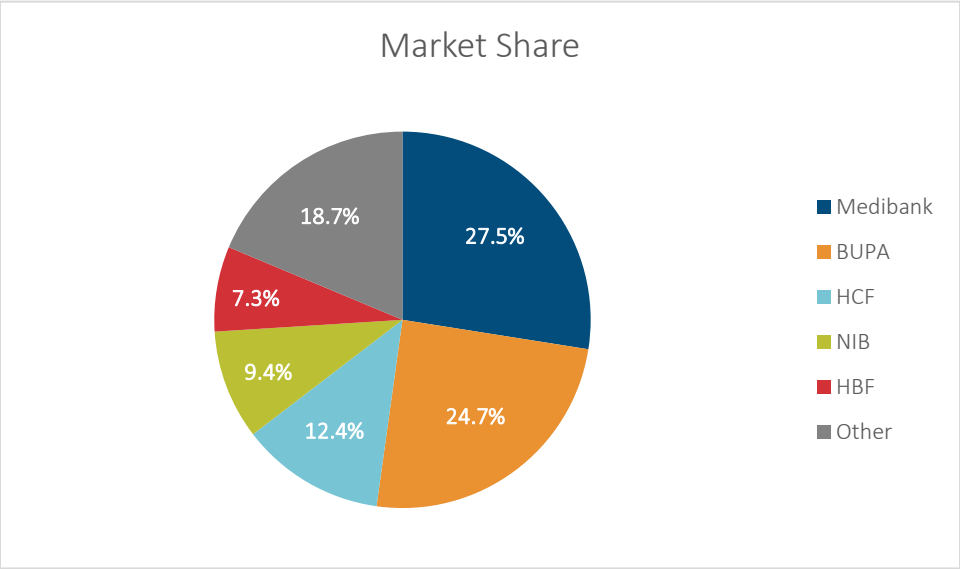
TOTAL NUMBER OF HOSPITAL TREATMENT POLICIES BY TYPE OF COVER AS OF DECEMBER 2022



2.2.2 MARKET SHARE

Figure 6

MARKET SHARE AS OF 30 JUNE 2022



Based on the Private Health Insurance Ombudsman State of The Health Funds Report 2022, there are currently 23 open membership insurers and 11 restricted membership insurers. The Australian Private Health Insurance market is considered relatively concentrated, with the top five private health insurers accounting for over 80% of the market share by premiums. See figure 6 for more information.

Section 3: Participants' Professional Backgrounds

For this research, we interviewed 20 participants across a wide range of companies in Australia. Most of them are actuaries with extensive experience in either Life Insurance or Private Health Insurance industry. A few participants have experience working in adjacent industries, such as Public Health, General Insurance, and Superannuation.

In our research, 16 of the 20 participants have more than ten years of experience working in the industry, and four have 0-9 years of experience.

Figures 7 and 8 show the market segment and type of organization the participants are working in, respectively. Eleven of the 20 participants have experience in Life insurance, six have experience in Private Health Insurance, while three have experience in adjacent industries. More than half of them work for a large insurer in Australia, three work for a large reinsurer, and two work for a consulting firm.

The participants are currently in or have held roles at various seniority levels, including Chief Actuary, Appointed Actuary, Chief Risk Officer, Head of Pricing, Senior Actuary, Actuary, and Analyst. They have experience in various practice areas, such as pricing, valuation, and data analytics. Many participants have experience in more than one practice area. These are summarized in figure 9.

Figure 7

PARTICIPANTS' EXPERIENCE BY MARKET SEGMENTS

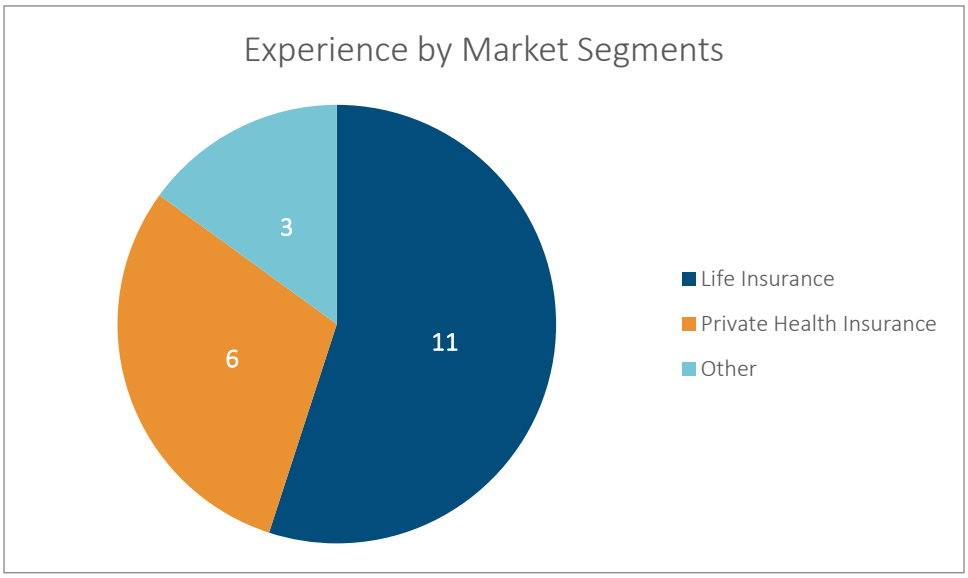


Figure 8

TYPES OF ORGANIZATION PARTICIPANTS CURRENTLY WORK IN

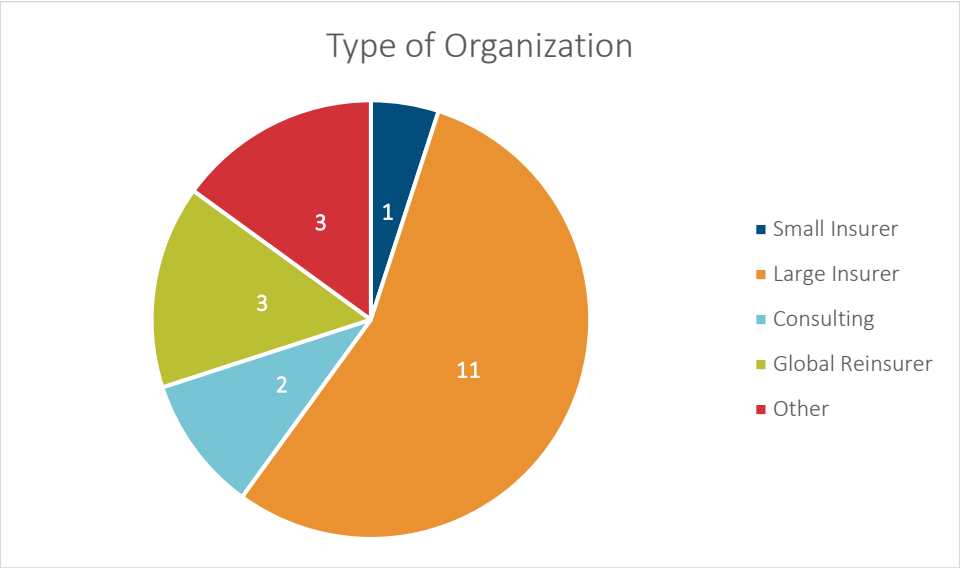


Figure 9

PARTICIPANTS' AREAS OF EXPERIENCE (MULTIPLE EXPERIENCE AREAS COULD BE SELECTED FOR EACH PARTICIPANT)

Section 4: Interview Responses

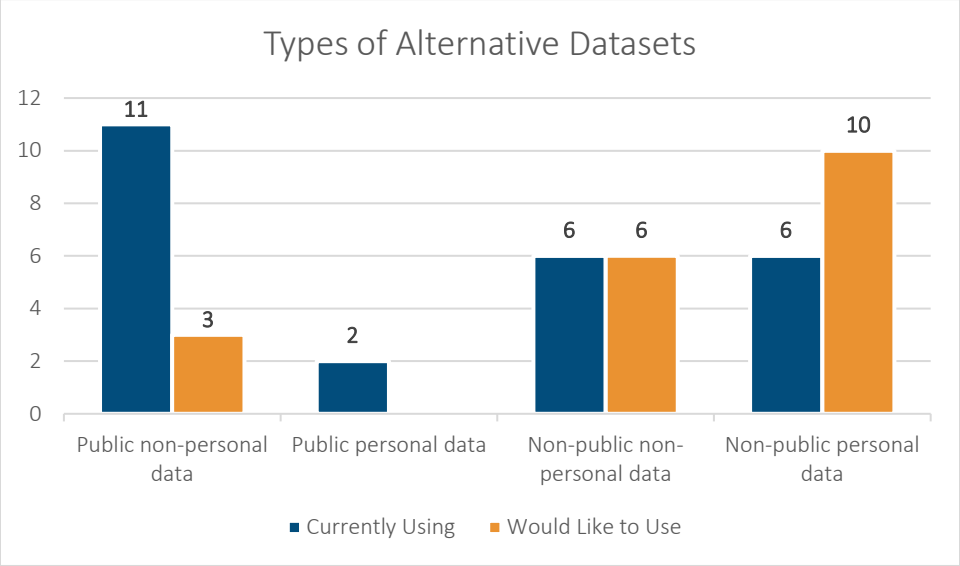
The participants were asked questions regarding their views and experience with alternative data usage. The interview included both multiple-choice and open-ended questions and asked for applicable examples. This section of the report summarizes the interview responses to these questions, while section 5 includes a few detailed case studies.

4.1 DEFINITION AND TYPE OF ALTERNATIVE DATA

The interview started with asking participants to define Alternative Data in their own way. The questionnaire then provided the following definition: "*information sources other than policy in-force and underwriting data*," for the purpose of this research. Most of the participants provided an aligning definition. They also provided a description of alternative data along the lines of "*usually unstructured, not client specific, not used as direct input in the modelling process, not from traditional sources*." Seventeen out of 20 participants stated that they have used alternative data in their practice. These participants then answered the question regarding the type of Alternative Data used. The most used type was "Public non-personal data"[11] mentioned by 11 participants. See figure 10 for more information.

Figure 10

TYPES OF ALTERNATIVE DATA USED BY PARTICIPANTS AND TYPES THEY WOULD LIKE TO HAVE (MULTIPLE OPTIONS COULD BE SELECTED)



When asked about the type of Alternative Data that they would like to use, but that was not yet available, 10 participants expressed interest in "Non-public personal" data, and six in "Non-public non-personal" data.

4.2 APPLICATION OF ALTERNATIVE DATA

When asked about the application of Alternative Data usages, a wide range of areas were identified by the participants. The most mentioned areas included customer engagement and interaction, product development,

[11] "Public" and "Non-Public" refers to whether the dataset is publicly available or not, and "Personal" and "Non-Personal" refers to whether data attributes are associated with individuals or aggregated.

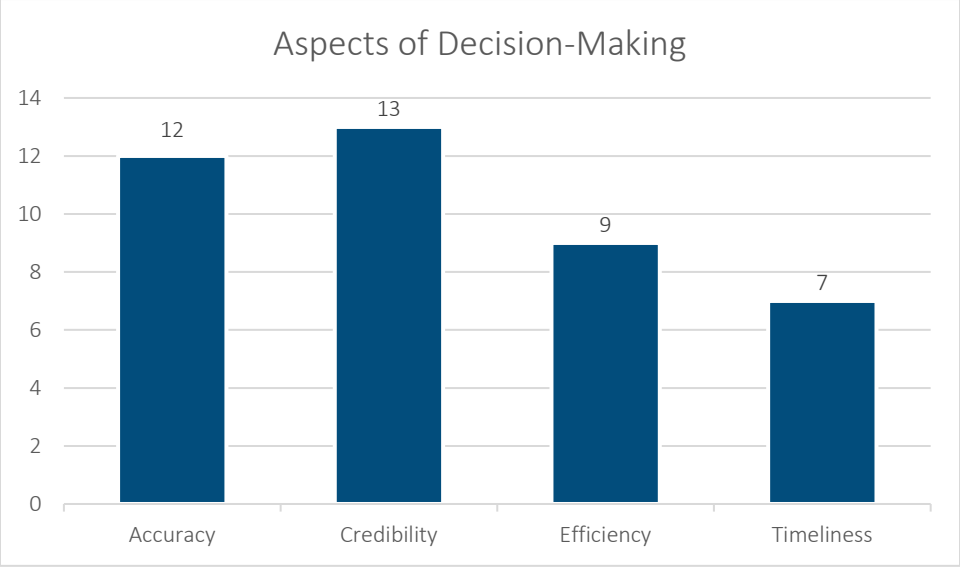
experience monitoring, segmentation analysis, and competitor analysis. Figure 11 provides a collection of participants' idea on desinated areas for alternative usage. Larger fonts within the word cloud represent terms and phrases that were more frequently cited by the participants, highlighting their significance and prevalence in the responses gathered.

Figure 11

WORD CLOUD ON APPLICATION AREAS OF ALTERNATIVE DATA (MULTIPLE ENTRIES ALLOWED)

Figure 12

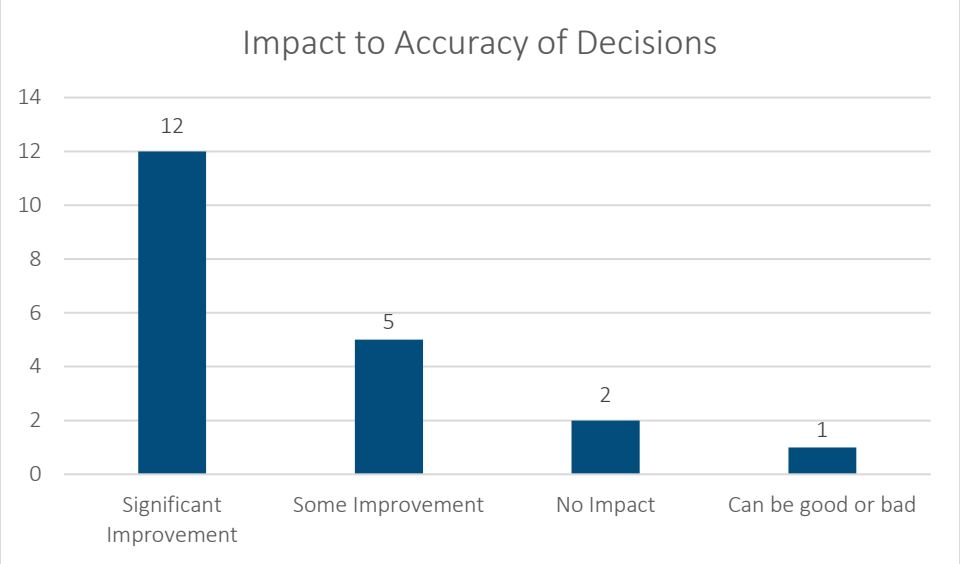
ASPECTS OF DECISION-MAKING THAT COULD POTENTIALLY BE INFLUENCED BY ALTERNATIVE DATA (MULTIPLE OPTIONS COULD BE SELECTED)



As suggested by figure 12, more than half of the participants think that Alternative Data usage can influence the accuracy and credibility of decision-making, while a few think it also influences efficiency and timeliness.

Figure 13

IMPACT OF ALTERNATIVE DATA ON ACCURACY OF DECISIONS MADE



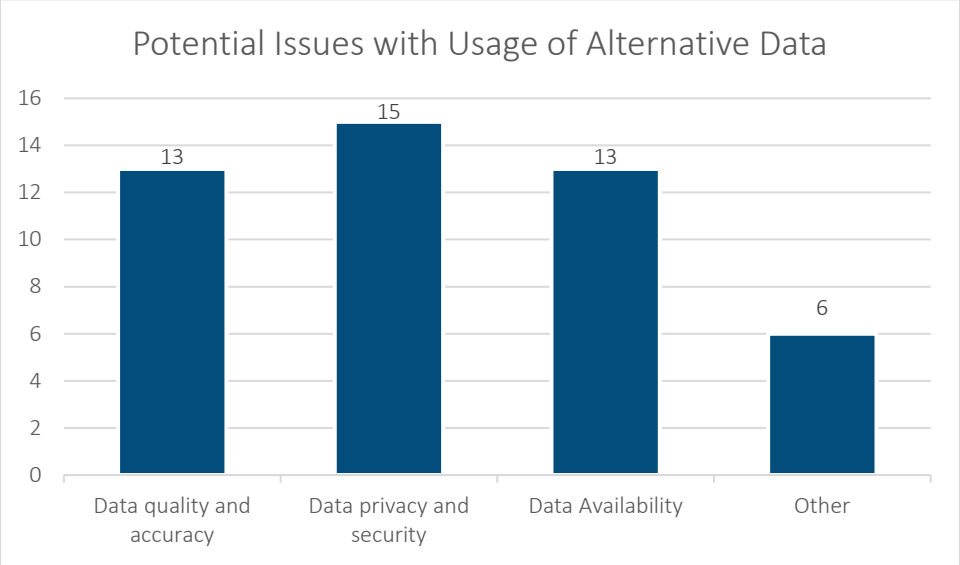
As shown in figure 13, a vast majority of participants thought that using alternative data would significantly improve the accuracy of decisions made in their practice areas, while others suggested that alternative data used would provide some or no improvement to decision-making in regard to accuracy. One participant indicated that using alternative data could be both good and bad for decision-making. It should be noted that these are personal opinions regarding the impact of alternative data usage, which may or may not be based on actual practice. A health insurance practitioner explained how hospital eligibility check data, as an alternative dataset, could be used to enhance the timeliness and predictive accuracy of claim projections (see detailed discussions in section 5.1). Additionally, several participants from life insurance companies expressed that, if postcode-level morbidity and mortality information become more available, the accuracy of their estimations and predictions would likely improve (for more discussions on postcode level data, see section 5.5).

4.3 CONCERNS AND CHALLENGES

The participants identified challenges and concerns in using Alternative Data, most notably with regard to data availability, data privacy and security, and data quality and accuracy. See figure 14 for the breakdown. It was mentioned that there could be difficulties in accessing and getting consent from policyholders to use personal data beyond what was collected during the insurance application and/or claim processes. In addition, regulatory requirements may restrict the use of certain data. Alternative Data volume can be much larger compared to traditional datasets. Concerns were raised regarding whether there are sufficient resources, capacities, budgets, and skilled employees in the organization to store and handle the larger volume of the data.

Figure 14

POTENTIAL ISSUES WITH ALTERNATIVE DATA USAGE (MULTIPLE OPTIONS COULD BE SELECTED)



Some participants also raised concerns about data reliability and consistency. It was highlighted that a controlled process may not have been applied to collect and process alternative datasets, which could be attributed to resourcing limitations and a lack of comprehensive framework for utilizing alternative data. For example, the definition of variables in the alternative data could have changed over time without adequate documentation, which adds complexity to the approach for processing and interpreting the data.

A few participants elaborated on the challenge of transforming alternative data usage into strategies and operations to create commercial benefits. To begin with, there could be many alternative data sources available, but there exists a notable level of uncertainty regarding the practical value added by these alternative datasets. Users need to have a clear purpose in choosing such datasets, and carefully examine their relevance and credibility, to determine the weight placed on the alternative data.

The interview asked participants to define bias in their own way. Various types of definition were provided, which have been summarized into:

- Cognitive bias, such as confirming bias, status-quo bias, and anchoring bias, usually originating from people's own experience and preferences.
- Unfair treatments against a certain cohort of members.
- Putting heavier weighting towards certain aspects of the business operation, potentially due to strategic initiatives or relative lobbying powers among business functions.
- Divergence between the intended outcome with the ones produced by data and model.

When asked about bias towards "Race, Political opinion, and Sexual orientation." Participants identified datasets that could be used as a proxy for socioeconomic status or race, such as postcode-level data and credit card data. However, the majority of participants think pricing bias is not a significant issue in the Australian insurance industry, due to restrictions on life insurance pricing factors, and community rating where risk rating is not allowed for health insurance.

Section 5: Case Studies

Participants provided various interesting examples of alternative data use. This section includes details of selected examples as case studies.

5.1 NLP AND UNSTRUCTURED DATA

A common methodology mentioned during the interviews is Natural Language Processing (NLP) and its various applications in unstructured data. NLP is an emerging field in artificial intelligence (AI), which provides the ability to automatically read, understand, and derive meaning from human language. NLP encompasses a wide range of tasks and applications, including text classification and sentiment analysis, as well as speech recognition and synthesis. In this section, we discuss some examples of the unstructured data usage provided by the participants.

5.1.1 CLAIMS DATA

Claims data are a mixture of structured and unstructured data. The structured component, including claim amount, date of event and date of notification, and policyholder's details, can be obtained via the internal querying system. Unlike structured claims data, unstructured claims data generally does not have a specific format and can be presented as text-based claim descriptions/notes, medical/doctor reports, telephone voice recordings, and images and photographs.

NLP methods are often utilized to extract important information from these unstructured datasets. Several participants from both life and health insurance markets expressed the importance of unstructured claims data and acknowledged the value added by analyzing such data. We summarize their comments on the benefit of using unstructured claims data in the following points:

- Improving customer experience during claims assessment which, in turn, aids customer retention and modelling.
- Helping claim assessors better understand and monitor the complexity and severity of the claims, which helps triage and prioritize claims management, as well as resource allocation.
- Enriching predictive modelling by providing more accurate estimation/prediction of claims duration in an actuarial control cycle and allowing for changes in circumstances.

A significant proportion of participants mentioned that identifying features in unstructured data can provide new insights, help explain the complexity of claims, and improve resource allocation in claim management. High machinery power is expected since volume is high for unstructured data. Machine learning techniques are often involved in processing and modelling unstructured data. Some participants expressed the view that NLP techniques could be used to predict propensity and solve classification problems, and there are off-the-shelf statistical software/packages available. However, a few participants commented that the analysis of unstructured data has not been used for actuarial assumption setting purposes at this stage, but rather to supplement the traditional actuarial analysis and validate/confirm the initial view of claim features.

Several participants from the health insurance market highlighted the fact that different types of claims data usually come from various sources. We will use Australian private health insurance as an example. A large number of policies consist of “Extra Cover” and “Hospital Cover.” “Extra Cover” claims can be categorized into three types: paper-based claims, claims through mobile apps, and claims submitted by medical practices such as dental clinics. “Hospital Cover” claims generally come from medical providers and can be either paper-based claims or submitted via electronic platforms. One participant stated that hospital data could be used as good indicators of future claim volumes and assist monthly claim valuation. A good example is the eligibility check data from hospitals, which usually comes 1-2 months before the procedure/claim occurs. This data has been proven to be strongly correlated with incoming claims. However, it should be noted that eligibility check data are largely unstructured, incomplete, or

duplicated. It was also mentioned that due to the COVID-19 pandemic, health insurers also had to estimate deferred claim liabilities as some elective procedures were delayed or cancelled during that time.

Finally, it should be noted that several participants argued that the NLP frameworks developed for claim management could also be adopted for life insurance underwriting. In terms of incorporating alternative information into the risk assessment process, actuaries can either wait for claims experience to emerge to make projections based on historical data, or assess whether alternative data can be used as a suitable proxy for an underwriting outcome. The latter approach might enable the incorporation of alternative data sources more quickly. One participant mentioned that techniques such as NLP have recently been used in the underwriting process to transform text information into structured data or remove unnecessary information from unstructured data (e.g., extract only relevant medical disclosures from medical notes rather than the whole medical disclosure).

5.1.2 PERSONAL-LEVEL FINANCIAL DATA

During the interview process, a few participants listed personal-level financial data (such as credit card data and bank transaction data) as data they would like to have, but did not yet have access to. They claimed that such financial data could infer members' interests and lifestyle behaviors and, thus, assist the insurer in market segmentation and targeting specific market segments[12]. Personal-level financial data are likely to contain a mixture of text-based and number-based information and can be highly unstructured, which makes NLP a powerful tool in modeling.

Ethics concerns were also raised by some participants, as personal-level financial data can contain sensitive information. One participant mentioned the "Open data movement" in Australia, which grants access to personal data held by participating organizations to other organizations. For instance, banks may provide insurers with financial transaction records. In exchange for granting access, individuals may receive a discount on their premiums, while those who decline to participate may face higher premiums. However, this movement may create fairness issues and social problems, as some people may not have access to certain data (for example, some people do not own credit cards) or may not be digitally savvy.

5.1.3 CUSTOMER ENGAGEMENT DATA

A number of participants mentioned customer engagement data in their interviews and commented that such information could be useful for customer modeling, analysis, and retention purposes. For insurers, there is internal data on customer interactions and activities from multiple sources in the format of phone system data, call center notes, advisor activities, online interactions, and mails/emails. These data are usually unstructured in nature. Audio data, such as call center conversations, can be converted into text data to gain insights. Again, NLP is a useful tool for analyzing such customer engagement data.

A participant from the life insurance market claimed that consumer interaction records, specifically how frequent consumers contact the company, are used internally for lapse analysis but not yet for pricing purposes. In certain circumstances, a greater degree of interaction from existing customers indicates a lower probability of lapse.

A participant from the health insurance market stated that there might be incentives for health insurers to keep members for a long period of time to improve policy lifetime value and profitability. Thus, it is important to conduct

[12] In the US, it is known that global credit reporting agencies such as Equifax have significant biases in their credit scoring systems. These biases include penalizing individuals residing in food deserts and those who receive traffic violations, which could be attributed to factors like racism and police targeting. Consequently, some regulators have prohibited the use of such credit scores.

customer sentiment analysis to understand customer communications and interactions with the company, aiming to enhance customer experience and improve retention rates.

Some participants also mentioned that analysis on member engagement activities could enable targeted and personalized member communications and, in turn, assist in customer retention. Only a few participants raised concerns about privacy issues while collecting and analyzing customer engagement data.

5.2 HEALTH AND WELLBEING DATA

A majority of participants highlighted the importance of personal-level health and wellbeing data for both life and health insurance. For instance, one participant from the superannuation (pension) industry mentioned that the data they would like to get, but did not yet have access to, included information from self-declaration questionnaires and screening questions, as well as data on general health status of the members. The participant felt that such information is important, especially when the members are considering annuity products, usually closer to retirement age. However, not all members are financially literate to understand that it may be in their best interest to disclose health conditions to obtain better annuity pricing.

During the interview process, two types of health and wellbeing data were particularly mentioned by participants across the life and health industries, namely personal illness data and pre-existing conditions, and data collected from wearables devices via wellbeing programs.

5.2.1 PERSONAL ILLNESS DATA AND PRE-EXISTING HEALTH CONDITIONS

The first type of data was mentioned multiple times as data that participants wanted, but which had not yet been made available to them. The data could be either self-reported or medical information provided by partners or third parties. A handful of participants from a health insurance background emphasized the need to keep track of customers' wellbeing and health status throughout their policy and create a "clinic pathway" for their customers. As an example, two participants argued that, if historical records on pre-existing health conditions and GP visits were provided to them before a policyholder was admitted to the hospital, the health fund could take preventive initiatives and direct their customers to suitable health programs. Based on the responses from our participants, it could be concluded that customer wellbeing and health data can be used to assist in the early prevention of claims, membership engagement, claims cost estimation and management and, ultimately, help members achieve a better level of health and wellbeing.

One participant also mentioned the emergence of new home devices for professional healthcare scanning and commented that they could be used to collect more health/medical information that is otherwise difficult to obtain[13]. Over time, genetic testing might also be useful in supporting customers to better manage their health. The implication to the life insurance industry is that biometric data can potentially be used for pricing purposes, and we may be able to use biometrics to tailor the design of products or services for customers. Nevertheless, while improved access to genetic testing and biometric testing can offer valuable insights, it may also come with a cost and create anti-selection issues. For example, people may choose not to disclose relevant information to insurance companies, and it can be difficult to discover instances of non-disclosure when customers obtain health results outside the traditional medical system.

[13] Note that it is a standard practice in the US to get the list of prescription drugs and electronic medical records during life underwriting. Blood draw, medical questionnaires and doctor reports have been standard practice for decades. DMV (department of motor vehicle) records could also be accessed to see the number of accidents/tickets for behavior analysis.

5.2.2 DATA COLLECTED FROM WEARABLE DEVICES VIA WELLBEING PROGRAMS

Regarding data from wearable devices, a few participants expressed their interest in obtaining such data once it becomes available. On the other hand, it should be noted that several life and health insurers in Australia have already deployed wellbeing programs through mobile apps and wearable devices such as the iWatch and Fitbit. In some cases, free devices were provided to a selected cohort and a "benefit," in the form of a premium reduction, was offered for their participation. Personal information, including height, weight, physical activity (e.g., steps, speed, heart rate), sleep pattern, dietary habits, cholesterol, blood glucose, and blood pressure are typically collected via point-earning opportunities.

Some participants mentioned that such programs act as very effective marketing campaigns for insurers, promoting a positive brand image. Additionally, they assist insurers in market segmentation by identifying and targeting healthy individuals. Furthermore, through increased customer engagement, wellbeing programs improve customer retention, upselling, cross-selling, and customer modeling. Ultimately, there is potential for early and targeted health intervention based on data collected from the program which, in turn, can reduce future claims costs.

In fact, several participants claimed that, at the current stage for most existing wellbeing programs, the data collected is not used directly for pricing or claim assessment/underwriting purposes. Instead, it is used to support the design of the program, aiming to influence healthier lifestyle choices and potentially reduce future claims costs. In the long run, this type of program is likely to move away from solely focusing on increasing engagement and shift towards utilizing data to identify customers with specific health risks, aiding in the management of chronic diseases and providing early interventions. We noted the following phrases used frequently by participants during the interview:

- Risk segmentation
- Change customer behavior
- Attract/retain healthier customers
- Early warning system (for people at higher risk)
- Targeted interventions
- Health coaching
- Tailored pathway (to health management)

Participants also mentioned that, although the collected personal-level data have not been directly linked with mortality experience yet, they believed there is a strong correlation with the lapse rate, which is also a key consideration for life and health insurers.

However, it is worth noting that a few participants warned that consent is a big concern for the implementation of such programs.

5.3 SOCIAL MEDIA/INTERNET DATA

Quite a number of participants expressed interest in getting more information from social media/internet, including such information as lifestyle behavior, internet click history, and Google/Bing search engine history. During our interviews, comments were made on the potential of using machine learning techniques to process and analyze social media data and understand underlying behaviors, habits, and lifestyle factors. Comments were also made regarding the potential for AI development to enrich the dataset of search history. For instance, ChatGPT can engage in conversations with customers, offering deeper insights about individuals compared to the non-connected keyword searches used previously.

"Deep Learning techniques can be used to analyze digital snapshots of individuals and predict their propensity for certain health outcomes."

One participant who has a strong focus in AI/ML modeling stated that social media data could be used to detect potential misrepresentation in underwriting. For example, if a person has not been transparent about certain questions, or hasn't truly disclosed their situation in underwriting, we may be able to pick it up from social media data. The participant also mentioned that Deep Learning techniques could be used to analyze digital snapshots of individuals and predict their propensity for certain health outcomes. As a slightly different purpose, a participant with a consulting background mentioned that scraping competitive information from the internet for competitor analysis, including products and pricing, is a standardized process.

Unfortunately, based on the interview discussions, there seems to be an overall lack of access/availability for social media/internet data to be used in life and health insurance practices at the current stage. One participant mentioned a potential business partnership between search engines and insurance companies. It was also stated that direct access to the dataset may not be necessary for insurance companies. Instead, they could define the problem they would like to solve, and search engines could then offer insights or campaigns based on the search history information. In addition to collaborating with search engines, third parties such as independent research firms were also mentioned by participants as a potential source of information. One example is that keystroke data can be used to monitor customer behavior during the online insurance application process. Another example is to work with marketing research firms to monitor trends in the number of searches and clicks on personal injury law firms. Increased interaction could indicate higher disability claims driven by legal action. In fact, a few participants mentioned that their organizations had already established a dedicated internal team who works with external data providers.

5.4 DATA FROM GOVERNMENT AGENCIES (PUBLIC NON-PERSONAL DATA)

As discussed in section 4.1, "public non-personal data" was the most frequently mentioned type of alternative data used by interviewees. Among those datasets mentioned, some have been widely used in industry practices, while others are new and more progressive. It should be noted that, even for traditionally used datasets, there might be "alternative" ways of incorporating the information into insurance practices. In this section, we further categorize these public non-personal datasets into "Traditional" and "Non-traditional" subgroups.

5.4.1 TRADITIONAL

Macroeconomic Data:
During the interview process, almost all participants mentioned the usage of macroeconomic data in their practice, including underemployment rate, GDP, interest rate, economic forecast, and inflation rate (or CPI). Such data generally comes from federal government agencies, for example the Australian Bureau of Statistics (ABS) and the Reserve Bank of Australia (RBA).

Several participants mentioned that there is a positive correlation between inflation rate and lapse rate: at times of high inflation, lapse rate may increase due to affordability concerns. Participants also mentioned that macroeconomic data are generally used for monitoring purposes and provide early warnings to the company. They can be used to monitor and analyze market environment/sentiment, and such analysis can support and inform pricing decisions and product development.

Industry Data Collected and Published by APRA:
Besides the macroeconomic data published by ABS and RBA, participants also cited the Australian Prudential Regulation Authority (APRA) as an important source of industry data. APRA is an independent statutory authority

that supervises institutions across banking, insurance, and superannuation, and promotes financial system stability in Australia. It conducts industry surveys that pool information across different companies, which can be utilized for industry experience and competitor analysis. Many participants stated that such information can assist in business performance analysis, strategic planning, and product design and development.

Participants from the health insurance market mentioned that APRA collects data for the private health insurance industry, which includes information such as participation rates, member movement, claims data, cost of acquisition, risk equalization, and financial reporting data. APRA publishes granular data categorized by factors such as product type, age band, and state, as well as by company. They also provide P&L statements and balance sheets for the entire industry, as well as for different fund sizes. This information is used for benchmarking and monitoring market share, as well as for comparing a company's own experience with that of other players in the market. One participant specifically mentioned the deferred claim liability survey conducted by APRA in response to COVID-19, which collected information from the top five health insurers. The participant highlighted the usefulness of this information for smaller insurers.

Population-level Data and Health Information:
A large number of participants confirmed the usage of population-level mortality and health information in their practice, where such data are publicly available. We will now summarize the datasets used or desired from different sources.

Australian Bureau of Statistics:

- Australian population life tables
- Census data
- Cause-of-death data
- Cancer screening data (particularly interesting for Income Protection products)
- Disability data from survey of aging and disability care
- Household, Income and Labour Dynamics in Australia
- Survey of Disability, Aging and Careers
- General Social Survey

Australian Institute of Health and Welfare (AIHW):

- Sickness/prevalence data
- Disability Services National Minimum Dataset
- Australian Spinal Cord Injury Register
- Accident rate and suicide rate

Other data source for population health:

- State-level cancer registry
- Australasian Rehabilitation Outcome Centre

It should be noted that, while participants cited many of these government datasets, they also acknowledged that such data are primarily used to support or confirm the observed trends from internal experience. There are still concerns about data quality and credibility, as the data might be collected through multiple channels and lack consistent definitions. Additionally, in some cases, data was collected through surveys.

5.4.2 NON-TRADITIONAL

Migration/International Arrival Statistics and Visa Grant Numbers:
In Australia, it is mandatory for certain visa subclass holders to maintain private health insurance, including international students and visitors. Unlike the local private health segment, this international health segment is less regulated with limited industry statistics. Several participants from the health insurance industry mentioned that indicators such as migration statistics, international arrivals, and visa grant numbers are often used to project and monitor health insurance sales volumes. Visa grant numbers under certain categories can be obtained from the Department of Home Affairs, and international arrival statistics can be obtained from the ABS.

However, participants noted that these data are not directly used as inputs in actuarial modeling. Instead, they are utilized to support decision-making processes and validate findings and conclusions.

Synthetic Data based on Simulations:
One participant from health insurance mentioned overlaying census data with other datasets, such as car choices and shopping habits, to enrich the characteristics of the data and generate a "synthetic" dataset, which can be used in simulations to support marketing efforts. By subjecting this dataset to simulated scenarios and interactions, insurance companies can gain a better grasp of customer behaviors, preferences, and responses in different situations.

5.5 OTHER DATASETS MENTIONED IN THE INTERVIEW

5.5.1 POSTCODE-LEVEL MORTALITY AND MORBIDITY DATA

It is interesting to see that a significant proportion of participants mentioned that they would like to have access to granular-level population mortality and morbidity data such as mortality rate by postcode. While it was quoted that such data are available in the UK, to the best of our knowledge, insurance companies in Australia do not have access to postcode-level mortality data due to privacy concerns.

One participant specifically mentioned that the pricing of mortality based on postcode and socioeconomic status is currently implemented in the UK, but is not as advanced in Australia. This is partly due to the lack of such information, especially for the Group Life sector of the market, where postcode information may not be provided to the insurers on a membership level.

One participant with a health insurance background mentioned that postcode-level socioeconomic indicators from service providers (i.e., third party) have been used in Irish health funds. The data can be used to infer what life stage a member may be in (e.g., young family vs. retired couple). The company can then create tailored health insurance products, which can be targeted towards specific demographics and more desirable members. It should be noted that private health insurance is also community rated in Ireland with guaranteed acceptance and renewals, and there are around 400 different health insurance products in the country. The participant concluded that alternative data usage may provide insights for marketing and segmentation initiatives to gain comparative advantages.

5.5.2 RESEARCH REPORT BY ACTUARIAL PROFESSIONAL BODIES, LARGE GLOBAL (RE)INSURERS AND CONSULTING COMPANIES

Quite a few participants mentioned that they would consider research reports by actuarial professional bodies or large global (re)insurers as reliable sources of information. In Australia, the local actuarial professional body is the Actuaries Institute of Australia (AI), and there are various working groups on topical issues in insurance practice. Several participants from the life insurance market mentioned that they frequently refer to reports produced by the COVID-19 Mortality Working Group for the latest excess mortality information, particularly during the pandemic.

Research conducted by the Society of Actuaries, the Canadian Institute of Actuaries, and the Institute and Faculty of Actuaries was also mentioned by several participants during the interview, particularly for the industry life tables they developed for the insured populations. Some participants mentioned that when Australian data are not credible enough or too sparse to set assumptions (e.g., too few records, the population base not large enough, or not enough data points over a sufficiently long period of time), they will look at overseas experience as a reference. One example provided was to look at longevity experience in the UK market, as the local market has very limited information on longevity experience due to low coverage rate.

It was also mentioned that the head offices of some large reinsurers provide in-house research articles on their website. For example, the Swiss Re Institute "sigma" provides a trusted source of market information for practitioners in the re(insurance) industry on risk topics including catastrophe, mortality and longevity, and economic and financial. It should be noted that such research activities are typically carried out by the company's experts and professionals across different departments.

One participant, however, pointed out that research articles being looked at by actuaries are mainly from "visible sources" such as the Actuaries Institute, rather than academic research. It would be beneficial for the industry to see more academic research on the correlation/causation of macroeconomic variables with insurance experience.

Based on the case studies and examples above, we concluded that overall alternative data are used to a varying extent in the management of the insurance portfolio. We also discussed the industry outlook with the participants as part of the interview. Several common themes and some concerns were raised, and we have summarized them in the next section.

Section 6: Scope of Alternative Data Usage in Life and Health Insurance

During the interviews, the participants were asked to provide their thoughts and views on the outlook of alternative data usage. Out of the 20 participants, 13 were optimistic or cautiously optimistic about the prospects of alternative data usage in the industry, while the others held a more neutral view. Most acknowledged that progress in this area has been slow and will take time to develop. Some identified limitations associated with the underlying characteristics of the industry, and expressed concerns related to data ethics and governance. Key observations are summarized into the following four categories:

- The changing landscape in data availability and handling techniques;
- Risk characteristics of the products and regulatory environment have implications on what data are collected and how data are used;
- Current market sentiment and management focus; and
- Additional fairness, privacy, ethics, and governance considerations.

6.1 CHANGING LANDSCAPE IN DATA AND DATA SHARING

A few participants pointed out that there is an overwhelming number of alternative datasets available to be used. The increasing volume of data may not even originate from the insurance industry. As discussed in section 5, we have already seen wellbeing data collected via wearable devices but, currently, only a small proportion of the population has a high level of health literacy[14] in general. Over the next 5 to 10 years, technology such as scan-at-home devices will significantly reduce the cost and barrier of obtaining health and medical information. This concept of democratization of healthcare services and resources means that information will be more accessible, affordable, and equitable for all individuals.

> *"Whether you like it or not, there will be more data available to everyone, and this may not even originate from the insurance industry."*

The Australian government-led multi-agency data linkage projects, known as the "open data movement," will facilitate the development of data linkage and sharing in the insurance industry. Some participants also observed start-up organizations specializing in digital health, mental health, and other areas, that partner with insurance companies to conduct health-related research. One participant suggested the idea of a "collaboration and feedback loop," providing medical practitioners with insurers' views on the average cost of typical procedures, which could give service providers benchmark insights and highlight any "outlier" practitioners.

Some participants from the life insurance industry are of the view that the insurers need to adapt to this new environment and find ways to make use of the new information. Otherwise, the new information may exacerbate selection risks. Customers may decide not to disclosure certain health information to the insurers, or to use genetic testing results to select against certain products. The information asymmetry and anti-selection risks could potentially increase. For the health insurance industry, the new information may provide valuable insights to more targeted product development and distribution strategies.

[14] We define personal health literacy as the degree to which individuals have the ability to find, understand, and use information and services to inform health-related decisions and actions for themselves and others.

One participant thought that the industry is on the brink of disruption and was excited to be involved in utilizing alternative data. Insurance companies must adapt to remain competitive in this new environment. As we enter the era of artificial intelligence and machine learning, actuarial techniques will undergo significant advancements. Afterall, actuarial work is an applied science.

On the other hand, more data are not always a blessing, according to some participants. Determining which data to use, and how to use the data, requires knowing the right questions to ask and how to interpret the results. Some discussed the potential skill and knowledge gaps among actuaries: how to approach alternative datasets? Do we have adequate tools or modeling techniques to use these datasets? One participant was of the view that the actuarial profession has the relevant skillset to utilize alternative data. However, it may take time to have substantial growth in this area. As a profession, we need to allow for a phase of innovative thinking and exploration of different ways to use it. This will eventually lead to a better-defined skill and knowledge baseline, and then we can focus on upskilling the profession in general.

"We should not blindly trust the results solely based on alternative datasets."

A few participants pointed out that it is important to ensure that the data being used is reasonable and objective, and to be aware that the data available externally may not be exactly what you are looking for. For example, the provider of the data may have collected it for different purposes compared to the user of the data. In such cases, external data may be based on different criteria than internal data, making it hard to interpret, transform, analyze, manage, and communicate with senior management. There might be little control over data integrity, and the reliability of the data may not be entirely relevant. Data "scanning" and "triangulating" are important to confirm and check outcomes by comparing outcomes based on different data sources.

One participant highlighted that it is important to note that all assumptions come with some level of uncertainty and, after all, the main consideration in actuarial pricing is whether all material risks are captured. One should make sure that alternative data usage would add value to what we do. Alternative data may provide additional credibility and timeliness, but the impact on accuracy and efficiency could be limited.

A few participants raised the concern that there are still gaps in traditional datasets and we should try to address these before turning to alternative datasets. The Group Life segment was identified as having issues such as lack of accurate Occupation categories for many superannuation funds, and low data quality in other membership fields. An industry-wide study has not been done for this sector for many years. For the health insurance market, a few participants would like to see APRA produce certain statistics in a more consistent and standardized manner, especially for the "international health" segment.

When asked about current pricing practices in life insurance, a recurring theme was that more weighting is placed on internal and traditional datasets. This is especially true when there is a large volume of claims data, which is more relevant to the portfolio, and this leads to hesitations in bringing in external data. While the use of alternative data is evident, it is still in the early stages and is currently limited to the most basic forms. A low weight is placed on the insights, and usage is usually limited to supporting product development, customer profiling, benchmarking competitors, and/or monitoring market sentiment.

6.2 RISK CHARACTERISTICS AND REGULATORY IMPLICATIONS

Some participants felt that the Life and Private Health industry might be 5 to 10 years behind the General Insurance industry in alternative data use, especially in pricing. One key factor is the number of pricing factors used: Life Insurance pricing in Australia is limited to basic demographic information such as age, gender, smoking status, and pre-existing conditions. This limited data are collected from customers through the application process.

> *"We are limited to the contractual agreement in terms of what data we could collect from the policyholders."*

In addition, these participants explained that the underlying risks in Life Insurance are low frequency events, when compared to General Insurance. This means that claims data volume is also low. The implication is that insurers typically have only one or two interaction points with the policyholders: one at the time of application and one at claims time if a claim occurs. Alternative data could serve as the bridge to establish more interaction points with the policyholders or to provide additional insights into their lifestyles throughout the lifetime of the policy.

One participant mentioned a limitation due to the nature of the risks: for Life and Health insurance, the underlying risks are associated with individuals and their health. This may raise more sensitivity and privacy concerns when data are shared. For example, when working with insurers who provide life insurance covers to superannuation funds, the trustees of these funds need to act in the best interest of the members. They can share data with the insurers provided it is in the members' best interest. The participant felt that there are opportunities for superannuation funds and life insurers to work more closely together. One potential area is in the development of Retirement Income products. Despite recently introduced regulatory incentives for members to elect such products, the take-up rate has been low. In this case, the superannuation funds may have rich membership data, such as financial literacy, risk and investment preference, lifestyle, etc. If shared with life insurers, such information could inform product and pricing designs to better align to members' interests and retirement needs.

Participants working in Private Health emphasized the community rating environment with no underwriting, guaranteed acceptance, and guaranteed renewals. They felt that alternative data could ultimately be used to develop an "edge" in the market through insights for marketing and segmentation initiatives to gain a comparative advantage. Most of them acknowledged the dependence on legislation, regulation, and government policies: if more diversity and innovation are encouraged, there may be more usage of alternative data.

> *"(we should be) working together to use data for good."*

One participant raised the concept that insurers have a role to play in maintaining or reducing the cost of health care for the community. Alternative data usage should not only be aimed at industry profitability, but also at achieving positive outcomes for the community at large. It is crucial for the insurance industry to collaborate and harness the power of data for good.

6.3 MARKET SENTIMENT AND MANAGEMENT FOCUS

Many participants shared their observation that senior management has limited buy-in and commitment to invest in this area due to resourcing constraints. Some mentioned the fact that the Australian life insurance industry has been struggling with profitability and sustainability in recent years. There are lots of issues distracting the insurers and increasing resource constraints. This market environment has made it difficult for systematic investment in alternative data usage and analytics.

"Some management teams have been focusing on growth but not necessarily sustainable growth."

One participant emphasized the importance of recognizing the strategic potential in this area and believes that, eventually, some insurers will begin adopting and effectively using alternative data, thereby gaining competitive advantages over others. Other participants observed that there are some investments, not even significant yet, in this area. They felt that senior management collectively recognizes that there are more insights to be drawn from the data. However, the development of alternative data usage may occur gradually and incrementally through trial and error over a considerable period, and it is important to be able to trial effectively and cheap. The value does not come from the data itself, but how to deal with and connect insights to it in a commercial context. This is the more important aspect to consider.

"We should be able to and have been managing the portfolio with imperfect data. Management decisions cannot wait (for the perfect data)."

A few participants mentioned the importance of a joint effort from the whole industry. They felt that advanced technology has improved the capability to collect and handle a larger variety and volume of datasets. Data sharing across different segments and companies will be beneficial to the whole industry.

6.4 ADDITIONAL GOVERNANCE AND ETHICS CONSIDERATIONS

Some participants highlighted that data handling and storage play an important role in the usage of alternative data. Whatever modeling framework is developed for alternative data, we need to ensure that the process can be repeated over time and that the data can be accessed for multiple purposes in a systematic and consistent way. An organizational data warehouse would be beneficial, as it becomes challenging when multiple functions attempt to analyze data in different ways, leading to significant effort being spent on reconciling outcomes and internal debates. Additionally, there should be a secure system in place to handle commercially sensitive information and ensure the privacy and security of the data.

> *"A key challenge with AI-specific regulation is that it focuses on the mechanisms rather than the outcomes, potentially leading to comparative issues between automated and equivalent human decisions."*

Additional privacy and fairness concerns were raised by a few participants, especially in the context of artificial intelligence and machine learning. They identified Responsible AI as an area of interest. It was acknowledged that the Australian AI regulations are principle-based, and there are other industry bodies and organizations which provide research and guidance on ethics.

Section 7: Conclusion

In this report, we discussed the usage of alternative data in life and health insurance practices in Australia, based on our findings from interview-based research commissioned by the SOA. We first provided an overview of the life and health insurance industries in Australia, explaining the current market state, including the types of products and distribution channels, market shares across companies, and recent industry profitability profiles. In this project, we conducted interviews with 20 individuals occupying diverse roles, primarily within large life and health insurance companies in Australia. The responses from the interviewees were described and discussed in the report. In the following, we summarize the key results and address the potential limitations of the research.

7.1 SUMMARIES OF KEY RESULTS

Throughout the interviews, participants provided various interesting examples of alternative data usage and discussed potential alternative datasets that could enhance business decision-making. We summarized the key alternative datasets outlined by our participants into five broad categories:

- Unstructured data such as claims data, personal-level financial data, and customer engagement data;
- Health and wellbeing data such as personal illness data and data collected from wearable devices;
- Social media/internet data;
- Data from various government agencies; and
- Other alternative datasets such as postcode-level mortality and morbidity data.

Drawing from the perspectives of the participants regarding the outlook for alternative data usage, we provide a conclusion on the scope of alternative data utilization in the context of life and health insurance in Australia, with key highlights summarized below:

- The evolving landscape of data and data sharing within the insurance industry is undergoing inevitable transformations.
- The development of alternative data usage may occur gradually and incrementally through trial and error.
- There should be a secure system in place to handle commercially sensitive information and ensure the privacy and security of the data.
- It is crucial for the insurance industry to collaborate and harness the power of data for the greater good of society.

7.2 LIMITATIONS

For readers of this report, particularly those who would like to apply the results, it is important to understand the potential limitations of the research. Our study has been conducted within the Australian life and health insurance landscape, which implies that its direct applicability to other countries' contexts may be limited. It's important to acknowledge the contextual variations and regulatory differences that exist internationally, potentially impacting the transferability of our findings. Moreover, we would like to emphasize that the comments and perspectives presented in this report are summarized based on conducted interviews, and they may not necessarily reflect the official views of the involved companies.

Section 8: Authors and Acknowledgments

8.1 AUTHORS

Zhan Wang, Fellow of the Institute of Actuaries Australia (FIAA)

Zhan Wang is a Partner at Azuria Partners Pty Ltd, a boutique actuarial consulting business specializing in the Australia Life Insurance industry. Zhan has been working in the industry for over 13 years. She has experience across all actuarial functions including pricing, valuation, experience management, and data warehouse.

Han Li, Ph.D., Associate of the Institute of Actuaries Australia (AIAA)

Han is an Associate Professor of Actuarial Science at the University of Melbourne. She has a broad range of research interests around longevity and mortality risks, aging and retirement, and the impact of climate change on the life and health insurance industries. Specifically, much of her research expertise centers on actuarial modelling and forecasting using advanced econometric and statistical techniques.

8.2 ACKNOWLEDGEMENT

The researchers' deepest gratitude goes to those without whose efforts this project could not have come to fruition:

- the Project Oversight Group and others for their diligent work overseeing questionnaire development, analyzing and discussing respondent answers, and reviewing and editing this report for accuracy and relevance.
- the participants who generously shared their experience, insights, advice, guidance, and wisdom. Any opinions expressed may not reflect their opinions nor those of their employers. Any errors belong to the authors alone.

Project Oversight Group members:

Gabrielle Guzman, FSA, MAAA

Blake Hill, FSA, FCIA

Traci Hughes, FSA, MAAA

Maggie Lee, FIAA

Hezhong (Mark) Ma, FSA, MAAA

Michael Niemerg, FSA, MAAA

Mujtaba Ali Shaikh

Xuling (Shirly) Wang, FSA

At the Society of Actuaries Research Institute:

Korrel Crawford, Senior Research Administrator

David Schraub, FSA, MAAA, CERA, AQ, Senior Practice Research Actuary

Appendix A: Interview Questions

This section includes the list of interview questions used:

- A.1 Primary Questions: This list of questions was used in 17 of the 20 interviews.
- A.2 Additional Questions: These additional questions were only used in interviews with 3 out of the 20 participants to focus on a particular interest area.

A.1 PRIMARY QUESTIONS

How would you define "alternative data" and "standard data" in the context of life and health insurance practices?

In our study, we have defined alternative data as information sources other than policy in-force and underwriting data. Have you ever incorporated alternative data into your daily operations or decision-making processes?

- Yes
- No

If yes, what type of alternative data have you used in your practice, can you give some examples?

- Public non-personal data
- Public personal data
- Non-public non-personal data
- Non-public personal data

In what area(s) do you use these alternative data?

1. Marketing
2. Underwriting
3. c. Customer retention, upselling, cross selling, customer modelling
4. d. Other (please specify)

If no, what obstacles do you face in using alternative data in your work?

If no, what type of alternative data would you like to have in your practice, can you give some examples?

a) Public non-personal data
b) Public personal data
c) Non-public non-personal data
d) Non-public personal data

If such data become available to you, in what area(s) do you plan to use these alternative data?

a) Marketing
b) Underwriting
c) Customer retention, upselling, cross selling, customer modelling
d) Other (please specify)

In your practice, which aspects of decision-making do you believe could potentially be influenced by alternative data?

a) Accuracy
b) Credibility

c) Efficiency
d) Timeliness - how long the data remains relevant

In your opinion, how would alternative data impact the accuracy of decisions in your practice?

a) Significant improvement
b) Some improvement
c) No impact
d) Negative impact

In your opinion, what are the potential issues associated with alternative data?

a) Data quality and accuracy
b) Data privacy and security
c) Data availability
d) Other (please specify)

How do you define "bias" in your practice, and in your opinion, how significant of an issue is it?

If applicable, what data types do you believe are sensitive and could potentially introduce bias into pricing, and what measures would you take to validate your pricing model to prevent unintentional bias?

a) Race
b) Political opinion
c) Sexual orientation
d) Other (please specify)

In your opinion, what are the future prospects of alternative data usage in life and health insurance practice?

Do you see any potential impact of IFRS 17 on the usage of alternative data?

A.2 ADDITIONAL QUESTIONS (SELECTED PARTICIPANTS ONLY)

Wellness programs: can you please give an overview of the type of wellness and customer engagement programs offered by your organization, and the key objectives of these programs?

What type of data do these programs collect? Would you be able to share some examples?

Are there examples of other datasets that you are using, or in your view, potentially can be used to improve understanding and management of insurance portfolios?

In what area(s) do you use the data, or in your view, potentially can benefit from using the data? Would you be able to share some examples?

a) Marketing
b) Underwriting
c) Customer retention, upselling, cross selling, customer modelling
d) Other (please specify)

Unstructured data: Could you please provide an overview of the types of data available, their characteristics, and how they differ from other internal datasets?

How would you approach implementing NLP techniques to analyze text data?

Can you describe some common challenges or limitations that arise when applying NLP to insurance-related text data? How would you address these challenges?

a) Data quality and accuracy
b) Data privacy and security
c) Data availability
d) Other (please specify)

Machine learning applications in insurance: In your opinion, what are some additional considerations related to privacy, fairness, or other factors that are introduced by machine learning? Can you share some examples?

How would you describe the development of "alternative data" usage in the general insurance industry over the recent 5-10 years? What are the key trends and observations in the industry?

AI application in insurance: in your view, what are some of the additional privacy and fairness considerations introduced by AI? How is AI ethics regulated? Would you be able to share some examples?

Overall, alternative data usage in life and health insurance lags behind that of general insurance. In your experience, what are some of the lessons and experience that the life and health actuaries could learn from general insurance actuaries?

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org