UNIVERSIDADE FEDERAL DE CIÊNCIAS DA SAÚDE DE PORTO ALEGRE

Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde

Fernanda Beatriz Junges da Silva

STRATEGIES FOR PRICING SUPPLEMENTARY HEALTH PLANS IN BRAZIL: NAVIGATING COMPLEXITIES AND ENHANCING ACCESSIBILITY

Porto Alegre
2025

Fernanda Beatriz Junges da Silva

STRATEGIES FOR PRICING SUPPLEMENTARY HEALTH PLANS IN BRAZIL: NAVIGATING COMPLEXITIES AND ENHANCING ACCESSIBILITY

Dissertação apresentada ao Programa de Pós-graduação em Tecnologias da informação e Gestão em Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre como requisito para conclusão de mestrado.

Orientador: Prof. Dr. Mauro Mastella

Porto Alegre

2025

Catalogação na Publicação

Junges da Silva, Fernanda Beatriz
STRATEGIES FOR PRICING SUPPLEMENTARY HEALTH PLANS IN BRAZIL NAVIGATING COMPLEXITIES AND ENHANCING ACCESSIBILITY / Fernanda Beatriz Junges da Silva. -- 2025.
86 f. : il., graf., tab. ; 30 cm.

Dissertação (mestrado) -- Universidade Federal de Ciências da Saúde de Porto Alegre, Programa de Pós-Graduação em Tecnologias da Informação e Gestão em Saúde, 2025.

Orientador(a): Mauro Mastella.

1. Precificação de Planos de Saúde. 2. Modelagem Preditiva. 3. Risco Atuarial. 4. Saúde Suplementar. I. Título.

Sistema de Geração de Ficha Catalográfica da UFCSPA com os dados fornecidos pelo(a) autor(a).

AGRADECIMENTOS

Chegar até aqui foi uma jornada intensa, cheia de desafios e aprendizados, e nada disso seria possível sem as pessoas que caminharam ao meu lado.

Em primeiro lugar, agradeço aos meus pais, José Eduardo Marques e Deisi Junges, pelo amor incondicional, pela força silenciosa e pela fé inabalável em mim. Vocês me ensinaram a acreditar nos meus sonhos e a lutar por eles, mesmo quando o caminho parecia difícil. Esta conquista é tão minha quanto de vocês.

Ao João, meu namorado e parceiro de vida, obrigada por ser abrigo nos dias difíceis e comemoração nos dias felizes. Sua presença tornou essa caminhada mais leve, mais bonita e mais cheia de sentido. Ter o teu apoio me deu forças para seguir em frente. Te amo!

Sou imensamente grata também aos meus avós Lécio e Noeli, e ao meu tio Fábio Junges, pelo carinho, pelas palavras de incentivo e pela torcida constante.
Deixo um agradecimento especial aos meus avós Thereza e José, que, mesmo de longe, continuam me acompanhando e inspirando de uma forma que palavras não conseguem descrever.

Agradeço de coração ao meu orientador, Mauro Mastella, pela orientação cuidadosa, paciência e confiança depositada em mim ao longo desse percurso. Sua dedicação fez toda a diferença para que este trabalho chegasse até aqui.

Aos colegas e amigos que estiveram comigo nessa jornada, obrigada por cada conversa, cada risada e cada gesto de apoio que tornaram tudo mais leve.

À banca avaliadora, agradeço pela disponibilidade e pelas contribuições valiosas para o aprimoramento deste trabalho.

E a todos os professores e funcionários do PPGTIGS, obrigada por compartilharem conhecimento e por ajudarem a construir a profissional que hoje sou.

Por fim, agradeço a Deus pela força, pela saúde e por me conduzir com amor e esperança em todos os momentos.

Este é apenas o começo de uma nova etapa. Levo comigo cada aprendizado, cada apoio e uma imensa gratidão a todos que fizeram parte dessa história.

Muito obrigada!

RESUMO

Introdução: O mercado de planos de saúde suplementar no Brasil é complexo devido a mudanças populacionais, regulamentação e dificuldades fiscais. Embora o Sistema Único de Saúde (SUS) forneça cuidados fundamentais, a crescente demanda por planos de saúde privados reflete lacunas na acessibilidade e na capacidade de pagamento. A sustentabilidade financeira foi uma preocupação fundamental, uma vez que o número de Operadoras de Planos de Saúde diminuiu de 2.037 para 696 entre 2000 e 2022. O cálculo atuarial é central para a gestão do risco de subscrição e da viabilidade a longo prazo. Os avanços na modelagem preditiva apresentam benefícios potenciais em termos de melhoria da política de preços, alcançando um equilíbrio entre sustentabilidade e acessibilidade.

Objetivos: Este estudo teve como objetivo analisar a eficácia de uma metodologia de precificação de planos de saúde suplementar e desenvolver um protótipo para sua implementação. Especificamente, buscou (I) identificar metodologias de precificação prevalentes por meio de uma revisão integrativa, (II) avaliar a sensibilidade da metodologia selecionada em múltiplos cenários e (III) desenvolver um protótipo em conformidade com as regulamentações da ANS.

Métodos: Uma revisão sistemática da literatura identificou metodologias-chave e variáveis preditivas na previsão de custos de planos de saúde. Com base nesses achados, um protótipo foi desenvolvido integrando análise de dados históricos e regressão linear múltipla. O desempenho do modelo foi validado por meio de um estudo de caso, testando diferentes margens de segurança para avaliar sua eficácia na previsão de taxas de sinistralidade e custos de planos de saúde.

Resultados: O protótipo demonstrou fortes capacidades preditivas, particularmente com dados históricos recentes e uma margem de segurança de 99%, melhorando a precisão da precificação para planos corporativos. No entanto, a incorporação de dados mais antigos reduziu a precisão devido à dinâmica do mercado e à inflação médica. O Plano Corporativo Regional apresentou a maior taxa de sucesso preditivo (88% dos cenários atingindo a taxa de sinistralidade alvo), enquanto o Plano Corporativo Nacional exibiu maior variabilidade de custos, exigindo refinamento.

Conclusões: O estudo apresenta um modelo de precificação atuarial que aprimora a sustentabilidade financeira em planos de saúde suplementar. Os resultados indicam que uma abordagem atuarial estruturada melhora a previsibilidade de custos, auxiliando as operadoras a manter o equilíbrio financeiro. Pesquisas futuras devem explorar modelos híbridos combinando técnicas estatísticas com aprendizado de máquina e desenvolver ferramentas de apoio à decisão para uma gestão mais eficaz de planos de saúde.

Palavras-chave: Risco Atuarial; Precificação de Planos de Saúde; Modelagem Preditiva; Sustentabilidade Financeira; Saúde Suplementar.

ABSTRACT

Introduction: Brazil's supplementary health insurance market is complicated by population changes, regulation, and fiscal difficulties. Although the Unified Health System (SUS) provides fundamental care, growing demand for private health plans reflects gaps in accessibility and affordability. Financial sustainability was a key concern as the number of Health Plan Operators decreased from 2,037 to 696 from 2000 to 2022. Actuarial calculation is central to the management of underwriting risk and long-run viability. Predictive modeling advancements hold potential benefits in terms of improving price policy, striking a balance between sustainability and affordability.

Objectives: This study aimed to analyze the efficacy of a supplementary health plan pricing methodology and develop a prototype for its implementation. Specifically, it sought to (I) identify prevalent pricing methodologies through an integrative review, (II) assess the sensitivity of the selected methodology across multiple scenarios and (III) develop a prototype in accordance with ANS regulations.

Methods: A systematic literature review identified key methodologies and predictive variables in health plan cost forecasting. Based on these findings, a prototype was developed integrating historical data analysis and multiple linear regression. The model's performance was validated through a case study, testing different safety margins to evaluate its effectiveness in predicting loss ratios and health plan costs.

Results: The prototype demonstrated strong predictive capabilities, particularly with recent historical data and a 99% safety margin, improving pricing accuracy for corporate plans. However, incorporating older data reduced accuracy due to market dynamics and medical inflation. The Regional Corporate Plan showed the highest predictive success rate (88% of scenarios meeting the target loss ratio), while the National Corporate Plan exhibited greater cost variability, requiring refinement.

Conclusions: The study presents an actuarial pricing model that enhances financial sustainability in supplementary health plans. Results indicate that a structured actuarial approach improves cost predictability, helping operators maintain financial balance. Future research should explore hybrid models combining statistical techniques with machine learning and develop decision-support tools for more effective health plan management.

Keywords: Actuarial Risk; Health Plan Pricing; Predictive Modeling; Financial Sustainability; Supplementary Health Insurance.

SUMARY

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

1. INTRODUCTION

The context of supplementary health in Brazil is multifaceted and heterogeneous, the result of a dynamic interrelation of constitutional rules, market forces, and actuarial issues. The constitutional affirmation of a right to health as a fundamental guarantee to all citizens imposes a huge burden on the State and gave origin to the Unified Health System (SUS). Nevertheless, difficulties in the service network of SUS and the remuneration system have increased the appeal of supplementary plans of health (Paim, 2018).

The supplementary health market protected by supplementary health insurance plans has a rich history stretching back to the period around the 1960s in Brazil (CELLA, 2019). Its trajectory of growth is exceptional and is attested to in the figures: from around 30.9 million beneficiaries as of December 2000 to a record-breaking 48.9 million as of December 2021, i.e., a rate of increase of 60% (ANS, 2022). Of particular note is the fact that in this group, there are 33.6 million enrolled in collective business plans, a reflection of the huge impact exerted by group health coverage on the scenario of supplementary health.

Although it has grown a lot, the coverage of supplementary plans of health in Brazil is still narrow and reached a limit of 22.98% of the population as estimated by IBGE as of December 2021. Revealing a gap in accessibility and affordability to a considerable part of society was the result.

The reduction in the number of Health Plan Operators (HPOs) from 2,037 as of December of 2000 to 696 as of June of 2022 contributes another source of complexity to the story. The reduction is officially explained by insolvency, brought about by issues like rising claims ratios and demographic changes as well as fee adjustment and difficulties in health plan pricing (ARAÚJO; SILVA, 2018; SÁ; MACIEL JÚNIOR; REINALDO, 2017). The decrease highlights the fine line operators are challenged to maintain between providing adequate cover and ensuring financial viability.

In the context of actuarial risk management, attention is on ensuring standards of economic and financial stability, with a focus on liquidity, solvency, and plan equilibria (PREVIC, 2012). Underwriting risk, the driving force in the building of actuarial models, is central to technical provisions and price estimation (ANS, 2022). The importance it has highlights the challenges in calculating such values and underpins the difficulties operators are challenged to surmount to minimize uncertainties in the supplementary health market—especially following recent modifications to ANS's ROL.

Although conventional means of calculating the rate of premiums have been used and been adequate, with consideration of frequency of utilization and average price, there is a dawning recognition of the importance of different methodologies. These differentials are less widely used in the supplementary health market and have the promise of a more variable representation of healthcare costs and, as a result, better refinement of the price strategies operators utilize (BUENO, 2017).

Comprehending the characteristics of Brazil's supplementary health market by way of research is fundamental because it illuminates a scope of problems ranging from legal to market and actuarial issues. The detailed study acts as a compelling resource to policymakers and healthcare practitioners and informs the crafting of methodologies meant to promote the fairness and resilience of the healthcare system. Conversely, the lack of such research may perpetuate gaps in healthcare accessibility and coverage and compromise the country's ability to provide adequate healthcare to all its denizens.

Ignoring the financial and actuarial side of the supplementary system poses grave risks—harm to consumers and market instability. Lacking a deep understanding of such issues, crafting effective policy to protect operators' and the larger healthcare system's financial health becomes virtually impossible. In a way, ignoring such research not only handicaps progress in improving healthcare but also risks subjecting all stakeholders—policymakers, providers, and users—to harm.

Overall, Brazil's supplementary healthcare landscape constitutes a rich and complicated mosaic of legal mandates, market forces, and actuarial issues. The keys to navigating such a maze are balancing growth and inclusiveness—providing benefits to a larger population from supplementary plans while controlling the actuarial risks associated with an expanding market.

2. OBJECTIVES

This section outlines the general and specific objectives of the study.

2.1. General Objective

This study aimed to analyze the efficacy of a specific supplementary health plan pricing methodology and to develop a prototype to promote its implementation.

2.2. Specific Objectives

I. Identify the most prevalent pricing methodologies used for supplementary health plans through an integrative exploring the rationales for using each method.

II. Develop a prototype for putting into service the methodology in accordance with ANS regulations, considering the needs of different types of health plan operators.

III. Examine the sensitivity of the selected pricing methodology to multiple cenarios through a case study.

3. JUSTIFICATIVE

This study was informed by the importance of the supplementary health market and the intrinsic high risks entailed by the industry (CERRI, 2016). Significantly, the recent approval of Senate Bill No. 2033/2022 ushered in a prominent change, shifting ANS's roll from exhaustive to exemplary. This was to the effect that healthcare operators were required to cover non-ANS procedures, adding sophistication to the actuarial undertaking of risk estimation and health plan price setting (BRASIL, 2022). Subject to the requirement to cover procedures out of ANS's roll, actuaries were presented with additional challenges in making an accurate estimation of risks and finding a proper price for the setting of health plans.

The supplementary health market in Brazil was a key pillar of the country's healthcare system and served a beneficiary base of above 48.9 million individuals (ANS, 2022). Notably, despite its significance, the market was formulated to pose countless actuarial risks to operators.

In the dynamic market landscape, determination of the optimal pricing model was critical to ensuring the viability of operators. This research investigated alternative methodologies of price setting that had the potential to increase the precision and efficacy of price setting in the Brazilian supplementary health market.

The field of price setting of health plans was vast with studies investigating numerous methodologies (e.g., Wynia et al., 2003). Of note, however, a larger share focused on conventional methodologies that were predicated on frequency of consumption and average cost (e.g., Dowling, 2008).

Usual and customary price set pay rates on the basis of the prevailing charges of comparable providers operating in the same location. This was a simple methodology whose simplicity marginally succeeded in diverging into geographic disparity and price inflation.

Besides, fee-for-service price set reimbursement to providers on a per-service basis. This system rewarded excess procedures and drove up costs by failing to account for the result of therapy (McClellan et al., 1994).

In contrast, diagnosis-related groups pricing assigned classes to patients by diagnosis and associated each class with fixed payments. This system fostered efficiency by motivating hospitals to cure their patients. That being said, it rewarded hospitals treating iller patients because the therapy costs of such individuals were typically higher (Robinson et al., 2014).

Familiarity with such conventional methodologies was important for a number of reasons. Firstly, it established a basis of understanding the context of Brazil's health plan pricing. Secondly, it emphasized the major areas of research to target when conducting research on the subject. Lastly, it presented a jumping ground to build on to formulate ideas on how to enhance research and innovation in the area (e.g., Cutler et al., 2013).

Although past methods had provably worked, it had become increasingly accepted that they may fail to capture the nuances and peculiar risks associated with Brazil's supplementary market. The same was especially considering the recent transformations ANS has undertaken. Thus, the research added value by delving into innovative methodologies of pricing that would better keep pace with the changing landscape of Brazil's supplementary market of additional health.

The research was instrumental to the academic community as it was in a niche area of scant research. Pricing of the health plans was core and was hardly discussed in academics. The lacuna was glaring and was mostly driven by the lack of specialists in the field of actuarial sciences. According to data as of December 2023, Brazil had a total of around 5,954 members of the Brazilian Institute of Actuaries (IBA). Of all the areas identified, around 4,464 were active, comprising around 75% of the total and a relatively small percentage of whom concentrated on the supplementary health sector.

The importance of the research was emphasized by the intrinsic level of complicatedness in dealing with the sector indicated by the deficiency of dedicated experts and scholars. The majority of actuaries had not centered their studies on this particular domain, and as a result, a huge shortage of academic research focused on the area of health plan price determinations resulted. Besides filling the gap left by existing research, the research presented a solid foundation to build up knowledge in the practice of the actuarial sciences and especially supplementary health. The circumstances of legislative overhaul and the growing challenges in the industry further intensified the necessity to investigate the effectiveness of the methods of price determination.

The shortage of academic research on the vital topic served to highlight the uniqueness and necessity of the research. Through a comparison of methods like Monte Carlo simulations, Collective Risk Theory, and Credibility Theory, the research not only added to available knowledge but also gave useful insights to professionals and scholars interested in streamlining price determination in a significantly influential sector of society. Ultimately, this paper did more than fill a void—it contributed significantly to the build-up and firm establishment of knowledge at the intersection of actuarial sciences and supplementary health.

The research not only contributes to the world of academics but also makes a considerable impact on operators of health plans. A detailed scrutiny of methods of price determination benefited both the operators and the beneficiaries and created a win-win relationship that could attract a large and considerable number of clients to the operators.

The importance to the operators of health plans was in defining the efficacy of price methodologies. This gave operators a better definition of the efficiency of their methods so that they could better predict healthcare costs required by patients. In gaining a better understanding of these costs, operators were better positioned to modify their billing to maintain better alignment of their revenues with the costs of conducting the business.

This better comprehension optimized the financial management of operators but also had direct relevance to the beneficiaries. In having better and costing-alignment prices, the benefits to beneficiaries were more stable and customized plans to their health, as expressed by quality services and by extension stronger customers' confidence and loyalty.

The impact of the study was thus positive and reached beyond the academic community to directly address key players in the health plans market. The study was enriched theoretically but also fostered sustainability and efficiency in operations in real-world settings and cemented the value of the study to multiple interested parties.

4. THEORETICAL FRAMEWORK

In this chapter, the main principles and theoretical foundations essential for understanding the methods and tools used in this research are explained. It is divided into two major parts. The first section, "Healthcare Costs," consists of four subtopics that address various aspects of healthcare costs and their impact. The second section, "Related Studies," presents a collection of prior research relevant to the study's objectives. This structure provides a strong theoretical basis to guide the study's progress.

4.1. Healthcare cost

Managing healthcare costs effectively is critical for the administration of health systems and the development of sound public policies. Authors such as Drummond et al. (2020), Oliveira et al. (2020), and França et al. (2019) define healthcare costs as the monetary value of resources used to provide health services. This definition includes both direct costs (such as medicines, supplies, and healthcare workers) and indirect costs (such as productivity loss from illness or disability).

Healthcare costs can be categorized in several ways: by service type (outpatient or inpatient), resource type (personnel, materials), or funding source (public or private) (Greene, 2017). Despite their differences, all analyses seek to quantify the economic value of resources consumed in delivering healthcare (Neumann et al., 2000).

Healthcare cost analysis serves multiple purposes: evaluating service efficiency, guiding investment decisions, and shaping public policy. However, measuring these costs can be challenging due to the variety of services involved, the difficulty of linking resources to outcomes, and issues with data accuracy (O'Brien et al., 2007).

This study adopts the definition of healthcare costs by Drummond et al. (2020): "the monetary value of resources used to produce healthcare services, including direct and indirect costs." This comprehensive definition captures the multiple factors influencing healthcare expenses.

4.1.1. Healthcare System

Ensuring access to quality healthcare remains a major challenge globally. Brazil addresses this with a two-tiered system that serves its diverse population. The public sector, through the Unified Health System (UHS), guarantees universal coverage based on principles of equity and comprehensiveness. Alongside UHS, a supplementary private insurance market offers additional healthcare options for those who can afford them. Managing this dual system and maintaining financial sustainability is an ongoing challenge in Brazil.

4.1.2. Unified Health System (UHS)

The UHS was established by Brazil's 1988 Constitution to provide universal healthcare coverage. It operates based on universality (healthcare for all), equity (care based on need rather than ability to pay), and comprehensiveness (a wide range of services) (Brazil, 1990).

UHS is decentralized and managed jointly by federal, state, and municipal governments (Brazil, 1990). It relies on a mix of government-run facilities and a network of public, philanthropic, and private providers (Brazil, 1990).

Funding for UHS comes from several sources: earmarked federal revenues, sector-specific taxes on companies, and a share of revenues from supplementary health plans (Brazil, 1990, 1993, 1998). Despite these streams, financing remains a concern, requiring more public investment (IPEA, 2023). UHS supports primary care (Basic Healthcare Units and Family Health Teams), specialized care (hospitals and clinics), medication and supply distribution, and health programs (immunization, chronic disease management, and health promotion) (Brazil, 2023).

4.1.3. Supplementary health System:

While UHS ensures basic healthcare access, some Brazilians seek faster service or more options through supplementary health plans, regulated by the ANS (ANS, 2023).

These plans operate on a self-financing model. Beneficiaries pay monthly premiums, which vary based on the plan type, age, and location (ANS, 2023). Plans may also include out-of-pocket costs like copayments, deductibles, and administrative fees. Revenues fund healthcare expenses, operations, marketing, and, for for-profit companies, shareholder returns (ANS, 2023). Although flexible and customizable, these plans come at significant costs that individuals must carefully consider.

4.1.4. Challenges

Despite offering universal coverage, Brazil’s healthcare system faces significant challenges. UHS struggles with underfunding, leading to long wait times and unequal resource distribution across regions. Ensuring consistent quality of care nationwide remains difficult. Meanwhile, supplementary health plans, while offering quicker and more extensive services, can be prohibitively expensive for many. Some plans may also exclude people with pre-existing conditions, raising concerns about fairness.

4.2. Related Studies

A deep understanding of healthcare costs and modeling is vital for ensuring the efficiency and sustainability of healthcare systems. Although little research has focused specifically on cost dynamics within supplementary healthcare, three important studies provide insight.

The first study by Kshirsagar et al. (2021) highlights the advantages of machine learning models over traditional regression approaches for predicting healthcare costs, achieving a 20% improvement. This study paves the way for future research, especially into algorithms connected to the Gini index.

Cao and Castel (2020) focus on evaluating modeling techniques, particularly weighted logistic regression. Their study emphasizes the importance of factors like gender, age, and complex health conditions in predicting high-cost patients, helping refine strategies for identifying and managing resource-intensive cases.

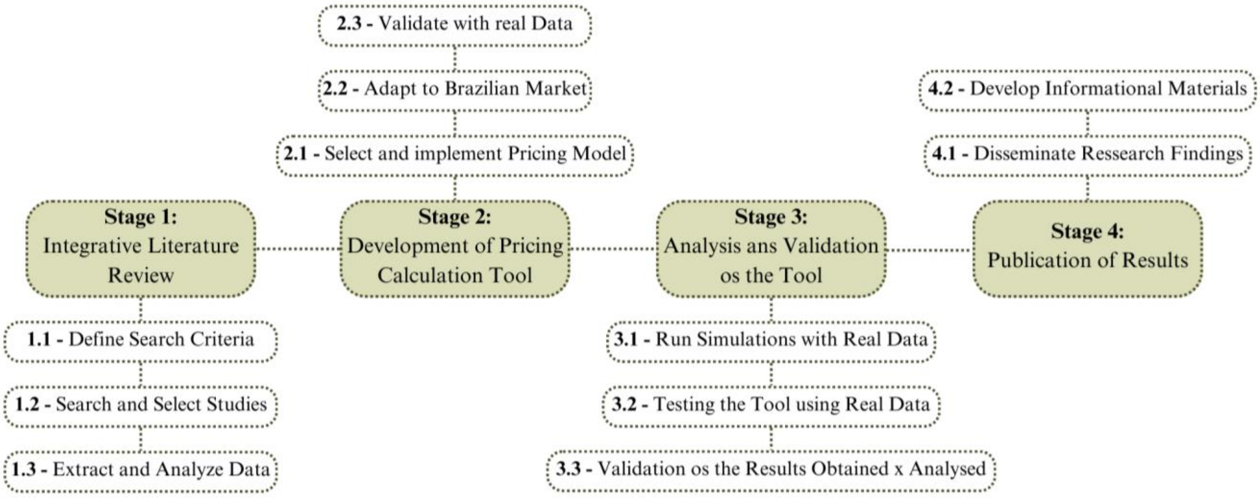
Xu et al. (2019) conduct an extensive analysis of the costs associated with out-of-network healthcare services. Using logistic and generalized linear regression models, they find a notable rise in direct costs, especially in emergency department visits from 2008 to 2013. They suggest investigating the causes of this trend and exploring policy solutions to improve efficiency in out-of-network care.

This research stands apart through a broad, integrative literature review comparing various cost prediction models across geographic boundaries. It aims to lay the foundation for developing a comprehensive health plan pricing methodology that incorporates international best practices while aligning with ANS regulations. Future efforts include implementing this methodology in a real-world system to deliver tangible benefits to healthcare providers, ultimately enhancing cost management efficiency and sustainability in the supplementary healthcare sector.

5. MATERIALS AND METHODS

In pursuit of the defined objectives, this research unfolds across four structured stages, as depicted Figure 1.

Figure 1 - Research design



Source: Author's own work

The first stage consists of an integrative literature review, aiming to synthesize the existing body of knowledge while identifying critical gaps in health plan pricing methodologies. Specific keywords and databases were strategically selected to explore studies related to healthcare pricing models, demographic aging, and the integration of technology in health services. Following the review, the key findings were analyzed to pinpoint limitations in current practices and highlight opportunities for methodological improvement.

Building upon these insights, the second stage is dedicated to the development of a specialized pricing calculation tool tailored specifically to the Brazilian healthcare market. This phase involves selecting the most suitable model from the literature, adapting it to comply with national regulations and accounting for cost structures unique to the local context.

The third stage centers on validation through real-world simulation. The tool will be rigorously tested using actual health plan data, simulating various market scenarios to evaluate its responsiveness, accuracy, and potential limitations. Comparative analyses will be conducted against real market performance data to benchmark the tool’s practical viability.

Finally, the fourth stage is devoted to disseminating the research findings. Outputs will include drafting the dissertation, submitting articles to peer-reviewed scientific journals, and presenting at academic conferences. Additionally, targeted informational materials will be developed for stakeholders such as health plan managers and consumers, ensuring that the practical applications of the research extend beyond the academic sphere.

It is important to emphasize that this study adopts a case study methodology, applied specifically to the health insurance operator under analysis. A case study, as defined by Yin (2018), is particularly suited for investigating contemporary phenomena within their real-life contexts, especially when the boundaries between phenomenon and context are not clearly defined.

By focusing on a single operator, this study narrows its scope to a contextualized examination, allowing a deep and targeted analysis of specific operational realities. This approach enables the development and testing of the pricing tool in a manner that is both realistic and tailored to the nuanced challenges of the sector (Stake, 1995).

The case study design is especially valuable here because it bridges theory and practice. It allows theoretical findings from the literature to be grounded in the practical realities of a functioning health insurance system, enhancing both the depth of the analysis and the relevance of the recommendations. Through this method, the study seeks to generate actionable insights for the operator while contributing to the broader academic understanding of health plan pricing (Creswell, 2013; Patton, 2015).

5.1. Integrative Literature Review

The Integrative Literature Review is designed to uncover the complexities of pricing methodologies in supplementary health, utilizing a qualitative approach to gain a deeper understanding of the subject under investigation (BEUREN, 2003). This preference for a qualitative methodology is intentional, chosen for its precision in exploring the intricacies of the research problem at hand. In alignment with the objectives of descriptive research, the review seeks to make a meaningful contribution to the understanding of pricing methods in supplementary health (GIL, 2008). This foundation sets the stage for a thorough examination of pricing methodologies in supplementary health services.

The methodology for the integrative review is structured and systematic, governed by a detailed protocol that outlines search strategies, data extraction methods, and analysis procedures. The review applies comprehensive search criteria across various databases, rigorously selecting studies that align with the research objectives, ensuring methodological integrity and relevance. This process facilitates the extraction of essential data, including study designs, methodologies used in pricing health plans, and key findings related to supplementary health pricing.

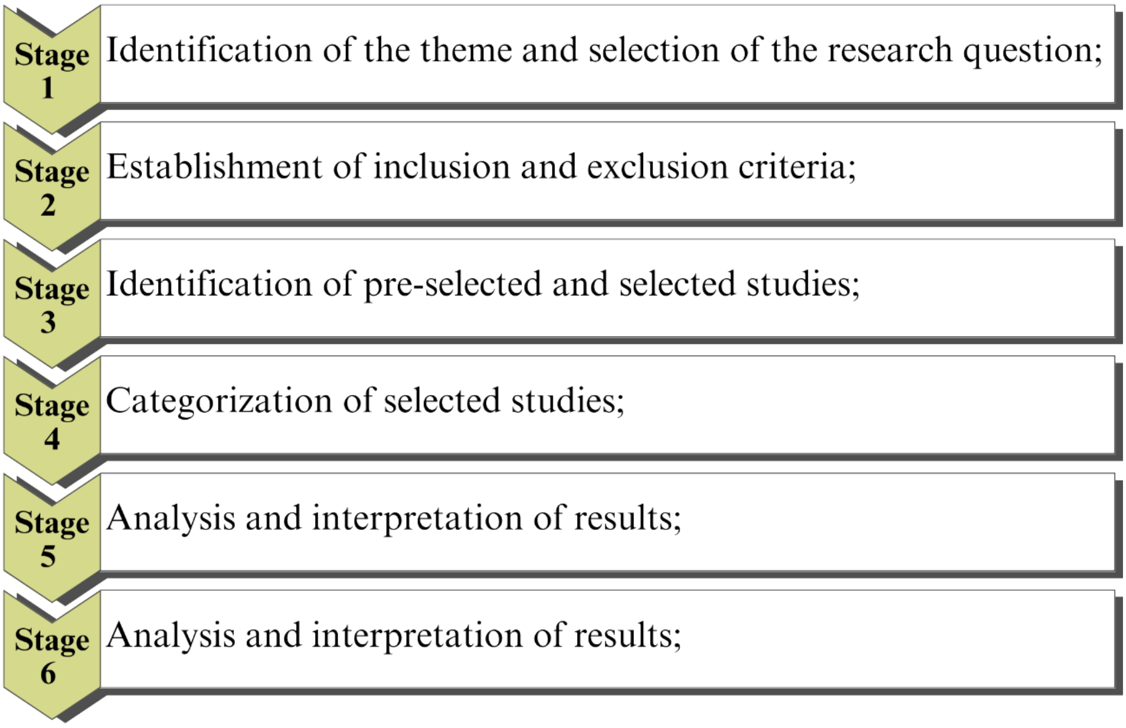
Each included study undergoes a thorough evaluation to assess its methodological rigor and potential biases, employing established tools and frameworks for comprehensive analysis.

Techniques such as thematic analysis and meta-analysis are employed to identify patterns and correlations within the findings. To minimize publication bias, proactive measures are taken, including the inclusion of unpublished studies and conducting sensitivity analyses to verify the reliability of the results.

By adhering to established reporting guidelines, the integrative review will offer a clear and transparent presentation of the outcomes, providing valuable insights into the complexities of supplementary health pricing methodologies. This methodical approach significantly contributes to the existing body of knowledge, advancing the understanding of pricing strategies in the supplementary health sector.

An integrative literature review strategy, as outlined by WEBSTER and WATSON (2002), will be used to address the gap in understanding health plan pricing methodologies. The primary goal of this approach is to synthesize existing knowledge and offer a comprehensive analysis of prior studies, paving the way for new insights and a deeper understanding of the complexities involved in pricing methodologies (BOTELHO; CUNHA; MACEDO, 2011; ERCOLE; MELO; ALCOFORADO, 2014). Figure 2 will serve as a visual representation of the methods utilized throughout the integrative review process.

Figure 2 - Integrative Review Process



Source: Adapted from Botelho, Cunha and Macedo (2011)

The research will explore the following search descriptors: 1) (Cost Forecast OR Cost Estimate OR "Pricing") AND ("Health Plans" OR "Supplementary Health" OR "Health Insurance Plans") AND ("Methodology" OR "Methods"); 2) (Cost Forecasting OR Cost Estimation OR Pricing) AND (Health Plans OR Supplementary Health OR Health Insurance Plans) AND (Methodology OR Methods) AND Actuarial.

The selected databases for this investigation include Scopus, Web of Science, ScienceDirect, and PubMed, chosen for their relevance to the research topic and their extensive coverage of academic publications. The selection process will focus on titles and abstracts to ensure the retrieved articles are directly relevant to the research question, thereby minimizing irrelevant results. This approach will aid in achieving a more comprehensive understanding of supplementary health pricing methodologies.

5.2. Pricing Calculation Tool

The methodology for determining health insurance pricing builds upon the integrative review presented in the previous chapter. This review provides a comprehensive perspective on existing pricing methods, highlighting prevalent practices and their underlying rationale. These insights serve as the basis for developing a precise and practical pricing model tailored to the context of this study.

The health insurance operator analyzed in this study is a cooperative with decades of experience, headquartered in the state of Rio Grande do Sul (RS). Dedicated to promoting the well-being of its associated physicians and beneficiaries, the organization offers a comprehensive portfolio of health plans, including individual and corporate coverage. In addition to medical plans, it provides dental care services and operates a state-of-the-art hospital equipped with advanced technology, designed to deliver high-quality care across various specialties. The operator's extensive healthcare network includes accredited hospitals, clinics, laboratories, and specialized facilities, ensuring accessibility and excellence in medical services. Its commitment to innovation and preventive care further highlights its role as a leader in the regional healthcare sector, serving a diverse and growing population of beneficiaries. The use of the operator's data for the purposes of this study was formally authorized in accordance with the guidelines established by the cooperative.

The analysis focuses on a corporate health insurance scheme encompassing a range of coverage options, including outpatient services, hospital care with obstetrics, collective accommodation, and multiple copayment structures. Data is segmented by age groups and healthcare cost categories, adhering to RDC nº 28/2000 standards set by the National Health

Agency (ANS). The dataset includes premium revenues and expenses for Office Consultations, exams, therapies, and other outpatient and inpatient services, strictly following ANS guidelines.

Data extraction, supported by QlikView software, was conducted in June 2023 and January 2024, covering the period from 2018 to 2024. This timeframe captures the substantial impact of the Covid-19 pandemic on healthcare costs, particularly in 2020 and 2021. The pricing model is developed entirely in Excel, serving as a tool for structuring data, performing calculations, and identifying key cost trends. The model employs 12-month periods, enabling detailed comparisons between premiums and healthcare expenses.

The variables used in the study remain unadjusted to preserve the integrity of the collected data. These variables include financial metrics, such as total healthcare costs and beneficiary copayments (both in Brazilian Reais), as well as operational metrics like service volume by procedure type (e.g., emergency consultations, exams, hospitalizations, and therapies). Other variables include the total number of service guides processed, the number of active plan beneficiaries, and age group segmentation based on the 10 categories defined by ANS. Additionally, temporal variables, such as year and month, allow for the analysis of seasonal and long-term trends, including the pandemic's effects. The dataset also distinguishes plans by product type, geographical coverage (regional or national), and target audience (individuals or companies), capturing structural variations.

Excel is used not only for initial structuring and analysis but also for the implementation of the pricing model prototype. This tool enables regression analyses, advanced calculations, and the visualization of cost trends, ensuring that the model remains practical and accessible. The results, formatted in Excel, ensure compatibility with supplementary healthcare data frameworks and integration with other industry tools. This analysis provides insights into cost dynamics, trends, and fluctuations over time, contributing to a deeper understanding of the corporate health insurance plan.

This robust methodology, fully implemented in Excel, delivers an effective tool for health plan pricing, enhancing resource management and supporting data-driven decision-making in the healthcare sector.

5.3. Analysis and validation of the tool

The evaluation of the Pricing Calculation Tool constitutes a critical phase to ensure its accuracy, reliability, and practical applicability. This stage is essential for verifying the tool's effectiveness in forecasting healthcare costs and supporting the determination of health insurance plan pricing with precision. Initially, the tool will be subjected to a comprehensive analysis through the use of real healthcare plan data. After the data input, pricing simulations will be conducted for the forecast period, and the simulated results will be compared against actual observed costs, enabling a direct validation of the tool's predictive capacity.

The validation process will focus on assessing the tool's ability to accurately reproduce healthcare cost patterns and pricing structures, while also evaluating its consistency and stability across different periods and datasets. Sensitivity analyses will be performed to examine the robustness of the tool's outputs under varying assumptions and external conditions, allowing an assessment of its adaptability to the dynamic and often unpredictable nature of the healthcare market.

By conducting this analysis and validation, the study aims to instill confidence in the tool's performance, ensuring that it provides a reliable foundation for decision-making in health plan pricing. Confirming its accuracy and reliability will enable the formulation of more efficient resource management strategies and support data-driven financial planning in the healthcare sector.

5.4. Scenario analysis

The rise in hospital costs is a major challenge in healthcare — and that's exactly why scenario analysis matters so much. The reality is that many complex and constantly shifting factors impact these costs. If research only looks at a single baseline scenario, it risks missing the bigger picture. Traditional models often assume ideal conditions, which rarely hold up when faced with real-world changes in hospital expenses.

That's why it's so important to look at different scenarios — optimistic, pessimistic, and baseline — to build smarter cost management strategies. In a best-case scenario, improvements in the economy and new technologies might help keep expenses in check. But let's be honest: that also means relying on factors we don't fully control. On the flip side, a pessimistic scenario — like an economic downturn or rising operational costs — calls for careful planning. It's about being ready with preventive actions and contingency plans to cushion the impact (Bradfield et al., 2005).

Scenario analysis isn't just useful for extremes. It helps us understand what's happening even in more typical conditions. Things like short- and long-term trends, demographic shifts, and emerging tech all play a role in shaping the hospital cost landscape (Van der Heijden, 2005).

That's the focus of this research: to go beyond traditional models and really explore the many possible futures for hospital costs. Because only by understanding these dynamic variables can we build more effective and resilient healthcare strategies.

5.4.1. Methodology and Scenario Definition

The construction of the projected scenarios is based on the historical variation in the utilization of different healthcare service categories. The goal is to simulate possible future behaviors based on recent past trends. For this purpose, the annual average percentage change in service volume was calculated by type of care, using two datasets:

- A complete dataset including all years in the historical series, including 2020 (an atypical year due to the COVID-19 pandemic);
- An adjusted dataset excluding 2020, in order to eliminate distortions caused by the sharp drop in utilization during the pandemic.

Based on this information, five analytical scenarios were defined:

1. **Baseline Scenario:** Assumes that variations in service volume will continue according to the historical average, including 2020 data. This represents a continuation of observed trends.
2. **Pessimistic Scenario 1:** Assumes that service volume will increase according to the historical average, excluding 2020. This scenario accounts for a potentially higher demand for services if the pandemic's suppression effect on utilization was temporary.
3. **Pessimistic Scenario 2:** Assumes that service volume will increase at the rates calculated including 2020, but external economic or epidemiological factors will exacerbate the cost increases. This scenario considers an adverse environment where rising demand coincides with increasing unit costs.
4. **Optimistic Scenario 1:** Assumes that instead of increasing at the historical average rate, service volume will decrease by the same percentage. This scenario considers potential improvements in healthcare efficiency, prevention measures, or economic conditions leading to reduced demand.
5. **Optimistic Scenario 2:** Similar to the first optimistic scenario but applies the reduced rates calculated without 2020 data. This assumes an even greater reduction in service volume due to a return to pre-pandemic healthcare consumption patterns and efficiency improvements.

5.4.2. Discussion and Implications

Each scenario offers important insights for managing hospital costs. Pessimistic scenarios draw attention to the financial risks tied to the growing demand for healthcare, underlining the necessity for proactive cost-saving measures, such as improving efficiency and exploring alternative reimbursement models. On the other hand, optimistic scenarios suggest that potential cost reductions could occur through systemic changes, like the broader adoption of telemedicine or a greater focus on preventive care, which could help reduce overall healthcare use. The baseline scenario, meanwhile, provides a reference point for evaluating shifts in both costs and service demand (Schoemaker, 1995).

For those making decisions in healthcare planning, understanding these scenarios is vital, as it allows for anticipating financial challenges and shaping policies based on evidence. Scenario analysis should be seen as an ongoing process—constantly updated with new information and adjusted as healthcare trends evolve (Godet & Roubelat, 1996). By taking a comprehensive approach, this research explores different possible outcomes under various conditions, factoring in the connections between technological progress, demographic changes, economic pressures, and other relevant influences. This method offers a balanced understanding of the potential variations in hospital costs, pointing out both risks and opportunities.

Ultimately, scenario analysis helps identify not only challenges but also areas for improvement, efficiency gains, and innovation. By thoroughly exploring different possibilities, this research seeks to offer valuable insights that can inform strategies to manage costs more effectively, strengthening healthcare organizations' ability to cope with the complexities they may face.

6. RESULT ANALISYS

6.1. Integrative Systematic Literature Review

6.1.1. Units of Analysis

The systematic review was initiated based on the descriptors mentioned above, resulting in a total of 952 returns. Initially, 166 duplicate studies were identified and excluded, leaving 786 unique studies. Among these, 42 articles were subjected to exclusion criteria, including 6 studies in other languages, 12 non-article studies, and 24 studies published before the year 2000. Therefore, 744 studies remained for content analysis. After the title analysis, 679 studies were rejected, leaving 65 studies for abstract analysis, of which 36 were rejected. After this process, 19 studies remained for full analysis, and subsequently, 4 were excluded, resulting in 15 selected articles.

6.1.2. Data Collection Techniques

Data collection was performed in the first semester of 2023 through direct searches using the keywords specified in the selected databases. The main data of each article, including title, authors, keywords, abstract, type of document, and journal of publication, were collected in the Microsoft Excel and Rayyan programs. The data were then processed to eliminate duplicates and identify valid articles for review.

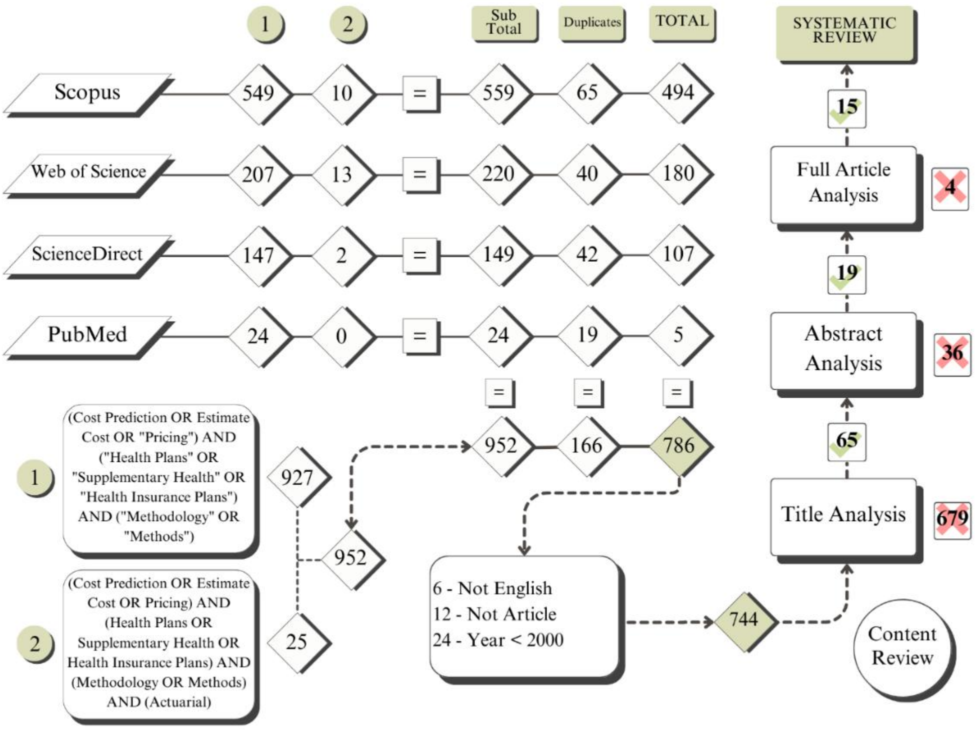
6.1.3. Results

The content analysis was initiated with a vast array of descriptors, resulting in a total of 952 returns. Among the first group of descriptors, 549 studies belonged to the Scopus database, 207 to the Web of Science, 147 to Science Direct, and 24 to PubMed, totaling 927 studies. As for the second group of descriptors, 10 studies were found in the Scopus database, 13 in the Web of Science, 2 in Science Direct, and none in PubMed, totaling 25 studies.

After identifying 166 duplicated studies, 786 unique studies remained. Among these, 6 studies in other languages, 12 that were not articles, and 24 published before the year 2000 were excluded. Therefore, 744 studies were left for content analysis. Of these, 679 were rejected after the title analysis, as they were not aligned with the established objectives for this study. Out of the remaining 65 studies, they moved on to the next stage of abstract analysis, of which 36 were rejected for not addressing cost projection or for only presenting a retrospective view of costs, rather than the intended prospective view in this research. 29 studies proceeded to full-text analysis, where 4 additional studies were excluded as they were only explanatory articles, without pointing out positive or negative aspects of the methodology used or comparisons with other methodologies.

Ultimately, 15 studies were included in the final analysis. These selected articles formed the basis for the methodological framework of this research, providing relevant insights into cost projection methodologies in the health sector. The entire selection and exclusion process is visually represented in the Flowchart in Figure 3 which systematically outlines each stage, from the identification of descriptors to the final selection of articles. This graphical representation ensures a transparent, structured, and replicable methodological approach, reinforcing the rigor and reliability of the research.

Figure 3 - Flowchart in identification and selection in studies



Source: elaborated by the authors (2023)

In this study, the results were structured through a systematic approach consisting of three distinct phases. The first phase involved filtering the 15 selected articles using the previously described methodology. In the second phase, a structured summary of the research findings was compiled, capturing essential details such as the year of publication, journal, research objectives, identified gaps, and the results obtained. These studies are organized chronologically in Table 1, providing a clear timeline of methodological advancements in health insurance cost projection.

Table 1 - First Author, year, title, journal, objective, future research/gaps, results obtained

Nº	FIRST AUTHOR	YEAR	TITLE	JOURNAL	OBJECTIVE	FUTURE RESEARCH/GAPS	RESULTS OBTAINED
1	STEENHUIS	2022	The potential risk of using historic claims to set bundled payment prices: the case of physical therapy after lower extremity joint replacement	Bmc Health Services Research	Predict the factors that explain the variation in the utilization of post-discharge PT after TKA or THA for osteoarthritis patients	The study highlights that the lack of historical data on high-value services can lead to an underestimation of the appropriate price;	If enabling factors are not considered in the risk adjustment of the bundle price, they may cause historic claims-based pricing methods to over- or underestimate appropriate post-discharge primary care PT use
2	KSHIRSAGAR	2021	Accurate and Interpretable Machine Learning for Transparent Pricing of Health Insurance Plans	35th AAAI Conference On Artificial Intelligence, AAAI 2021	Evaluate the ability of the machine learning model to predict costs	Develop algorithms involving the Gini index	The machine learning models showed 20% better results than the regression model compared
3	CAO	2020	Statistical prediction of high-cost claimants using commercial health plan data	Lecture Notes In Networks And Systems	Evaluate weighted logistic regression modeling techniques to maximize the accuracy of prospectively identifying high-cost claimants.	-	Male gender, increasing age, and complex conditions proved to be very relevant factors in predicting high costs
4	HARRISON	2020	Economic Outcomes of insurer-led care management for high-cost patients	American Journal Of Managed Care	Evaluate the impact of an insurer-led care management program on cost and utilization outcomes for high-need, high-cost Medicaid enrollees.	The study highlights the need for future research to design more effective care management programs.	The study found that there were no meaningful differences in total costs or visit frequency between Community-Based Care Management enrollees and non-CBCM enrollees in the 12-month post-enrollment period.
5	XU	2019	Cost-Sharing Payments for Out-of-Network Care in Commercially Insured Adults	American Journal Of Managed Care	The study used logistic and generalized linear regression models to estimate the cost-sharing amounts and out-of-pocket spending for out-of-network care.	Examine the reasons for the increase in out-of-pocket spending for out-of-network care, particularly in ED settings, and explore potential policy solutions to address this issue	The study found that out-of-pocket spending for out-of-network care increased substantially from 2008 to 2013, particularly for emergency department (ED) visits.
6	PARK	2018	Alternative evaluation Metrics for risk adjustment methods	Health Economics (United Kingdom)	Directly compare three board classes of estimators	Simulate health plans' behaviors in response to both group-level and individual-level residual risks conditional on risk-adjusted payments	Parametric regression estimators showed higher tail distribution prediction accuracy and individual-level prediction accuracy
7	POLITI	2018	A Comparison Between Subjective and Objective Methods of Predicting Health Care Expenses to Support Consumer's Health Insurance Plan Choice	MDM Policy And Practice	Determine which method is more effective in personalizing electronic tools that help consumers select health insurance plans based on their estimated health care utilization. (1 - quantitative healthcare utilization predictions and 2 - subjective-health status predictions.)	A larger sample size may have been useful, particularly for the follow-up data, given the ranges of variables discussed above	Both methods have limitations in predicting unexpected events such as new diagnoses of chronic illnesses or emergencies that arise.

Nº	FIRST AUTHOR	YEAR	TITLE	JOURNAL	OBJECTIVE	FUTURE RESEARCH/GAPS	RESULTS OBTAINED
8	MARTEI	2018	Methodology to Forecast Volume and Cost of Cancer Drugs in Low- and Middle-Income Countries	Journal Of Global Oncology	Provide information on forecasting the volume and cost of cancer drugs in low- and middle-income countries	The study does not account for the potential impact of drug resistance or changes in treatment guidelines on drug volumes and costs.	The results show that the methodology used can provide important information about the cost distribution of cancer drugs that were not previously appreciated
9	MORID	2017	Supervised Learning Methods for Predicting Healthcare Costs: Systematic Literature Review and Empirical Evaluation	Amia Symposium. Annual Symposium Proceedings.	Literature review on health cost forecasting	Experiments on different data sets from different institutions and regions; Analyze the amount charged as well as the disbursed amount paid out by the patients to see which approaches work best for each cost metric type	The method with the best overall performance was Boosted Trees, but for individuals with higher costs, the linear regression method performed better.
10	JOHNSON	2010	Gibbs sampling for a Bayesian hierarchical general linear model	Electronic Journal Of Statistics	Establish verifiable conditions under which the Gibbs sampler is geometrically ergodic	-	Gibbs samplers with different initial values produced similar estimates and required similar simulation effort
11	LANG	2010	Willingness to Pay for Lung Cancer Treatment	Value In Health	Evaluate the willingness of lung cancer patients to pay for a hypothetical new drug that could cure the disease.	The study did not test the reliability of the contingent valuation approach used.	The Willingness to Pay was influenced by various factors, including income and health status.
12	ROBST	2006	Estimation of a hedonic pricing model for Medigap insurance	Health Services Research	Examine premiums paid by beneficiaries for Medigap supplemental coverage. Average premiums charged by insurers are reported, as well as premiums by enrollee age and gender, and additional policy characteristics.	One limitation was the use of ordinary least squares regressions, which have strong assumptions regarding the distribution of the dependent variable and residuals.	Premiums policies were positively associated with the cost of health care services and the use of substitute goods and negatively associated with practice patterns. The study did not find a significant relationship between premiums and the risk-adjusted measure of health.
13	CECIL	2006	Relationship of the use and costs of physician office visits and prescription drugs to travel distance and increases in member cost share	Journal Of Managed Care Pharmacy	Help readers better understand the factors driving prescription drug costs and how to manage them effectively.	Further investigations that contain a broader variance in price changes for the bundle of products and services in their analysis could be clarifying for the health insurance industry	The price elasticity of demand for medical office visits was -0.20, indicating that a 10% increase in copayment would lead to a 2% decrease in office visits.
14	MARK	2003	Risk Adjustment for People with Chronic Conditions in Private Sector Health Plans	Medical Decision Making	Test the main risk adjustment systems	Study the results obtained in different health plans	One of the best performers was the regression-based method
15	PACALA	2003	Using self-reported data to predict expenditures for the health care of older people	Journal Of The American Geriatrics Society	Create and test a method for using self-reported data to predict future expenditures for the health care of older people.	-	Self-reported data may predict future costs more accurately than administrative data about beneficiaries

Source: elaborated by the authors (2023)

Table 1 illustrates the evolution of research in the field, from traditional statistical models to more sophisticated predictive techniques. Recent studies, such as those by Cao and Castel (2020), emphasize the application of machine learning in cost prediction and the economic impact evaluation of healthcare policies. Additionally, research by McClellan and Hackbarth (1994) and Cutler, McClellan, and Merrill (2013) discuss the application of regression models for risk adjustment and future expense estimation, highlighting the limitations of approaches based solely on historical claims data. The literature indicates that predictive models combining machine learning and advanced regression techniques have demonstrated greater accuracy in cost estimation, underscoring the need for more dynamic and adaptable methodologies (YIN, 2018).

Earlier studies focused on analyzing cost determinants and individuals' willingness to pay for healthcare services, often considering factors such as income, age, and medical conditions in influencing consumption behaviors (FRANÇA; GADELHA; TRAVASSOS, 2019). Several works, such as those by Neumann, Olsen, and McNamee (2000), compare pricing models and risk adjustment methodologies, including linear and non-linear regression approaches, to evaluate their effectiveness in different healthcare systems and insurance structures. Additionally, research on consumer decision-making in health insurance plan selection examines how cost-sharing mechanisms affect service utilization and financial accessibility (O'BRIEN et al., 2007).

Other significant contributions include using self-reported data as an alternative to administrative claims records for forecasting healthcare expenses. Studies such as those by Steenhuis et al. (2022) assess the reliability of patient-provided information and its potential to improve cost prediction models through adjusted regressions. Additionally, research on bundled payments and insurer-led care management programs investigates their effectiveness in controlling expenses for high-cost patients, offering insights into cost-containment strategies (WYNIA; AUSTIN; CUTLER, 2003).

The methodological advancements presented in the literature highlight a growing trend in incorporating enhanced regression models, machine learning, and more sophisticated risk adjustment techniques (YIN, 2018). These approaches enable greater accuracy in cost forecasting and the development of more effective policies for financing supplementary healthcare.

The third phase of the study aimed to address the research questions by analyzing the methodologies applied in cost prediction, the variables considered in each approach, and the justifications provided for selecting these methods. Table 2 offers a structured overview of the methodologies used across the selected studies, highlighting the diversity of statistical and machine learning techniques applied in health insurance cost forecasting.

Table 2 - Methodology studied, variables analyzed, and results obtained

Nº	FIRST AUTHOR	STUDIED METHODOLOGY	VARIABLES ANALYZED	JUSTIFICATION FOR CHOSEN METHOD
1	STEENHUIS	Multilevel linear regression	Predisposing and necessity factors, facilitating factors. Variables included in the templates are not specifically mentioned in the text	Regression is a statistical technique commonly used to analyze the relationship between a dependent variable and several independent variables. In addition, multivariate regression allows controlling the effect of several independent variables simultaneously, which is important when investigating the relationship between multiple factors and a specific result.
2	KSHIRSAGAR	Machine learning models	Costs, age, sex, CIDs, procedure codes, types of care	-
3	CAO	Weighted Logistic Regression	Age, gender, BMI, previous year costs	It is capable of handling binary data (high or low utilization of health services) and is useful for predicting rare events such as high claims costs.
4	HARRISON	Bayesian hierarchical zero-inflated gamma linear regression	Age, gender, race/ethnicity, neighborhood income category, Comorbidity Index score, and years of continuous enrollment in the plan	They are appropriate for analyzing cost and utilization data, which are typically skewed and have many zero values.
5	XU	Logistic and generalized linear regression	Plan characteristics, rural residency, year-fixed effects, and state-fixed effects. Age and sex were also considered in the algorithm constructing the HCC risk scores	They are commonly used statistical methods to estimate the effects of various factors
6	PARK	Regression estimators; Machine learning estimators; Distributional estimators	Inpatient, outpatient, and prescription drug claims, diagnosis, and procedure codes	It is fast and easy to implement, but also because it presents results on the scale of interest, which facilitates the interpretation of its coefficients.
7	POLITI	Linear models	Based on age, gender, and reported health conditions	This statistical method is well-suited for analyzing longitudinal data with repeated measures, which is the type of data used in this study.
8	MARTEI	Exponential smoothing method with error adjustment for the forecast	National cancer registry or GLOBOCAN estimates	It is a simple and scalable approach that can be applied in countries with limited cancer registry data.

Nº	FIRST AUTHOR	STUDIED METHODOLOGY	VARIABLES ANALYZED	JUSTIFICATION FOR CHOSEN METHOD
9	MORID	Linear Regression, regression decision tree, vector regression model, learning algorithm, learning algorithm	Age, Sex, Diagnosis groups, count of claims with diagnosis codes from each group, Procedure groups, Drug groups, Count of members' diagnoses, procedures, Drugs, Gender, Age	The authors identified five supervised learning methods that have been used in previous studies to predict healthcare costs. They are commonly used in health cost prediction studies and have been reported to perform well in previous studies.
10	JOHNSON	Bayesian Regression and Gibbs Sampler	Average expense and average income	It is a widely used and versatile statistical technique that can be applied to many different types of data.
11	LANG	Contingent valuation method (CVM), interval regression, and multiple linear regression	Sex, age, marital status, education, religion, job status, health status, income, family care, nursing assistant, dietary supplement, and smoking	These methods are commonly used in contingent valuation studies
12	ROBST	Ordinary least squares regressions	Premiums paid, cost of health services, patterns of practice and utilization, classification of care.	OLS regressions have strong assumptions regarding the distribution of the dependent variable and residuals
13	CECIL	Last squares dummy variable regression model	Copayment, distance from the member's residence to the physician's office	It is a widely used and accepted method for estimating demand and measuring price elasticity. Regression analysis allows researchers to control other factors that may affect demand, such as distance, age, gender, and other benefit categories.
14	MARK	Five different types of prospective risk adjustment systems, three regression models that added case-mix indicators from the DCG or ACG software packages, and one model that used HCCs weights	Age, gender, and area wage index values	They are the most common and widely used in the risk adjustment literature.
15	PACALA	Logistic and linear regression	Beneficiaries' health, lifestyle, use of health care, and demographic characteristics	It is a statistical technique commonly used to analyze the relationship between a dependent variable and one or more independent variables.

Source: elaborated by the authors, 2023

The most employed methodologies include regression models, machine learning algorithms, and Bayesian statistical approaches. Regression models, such as multilevel linear regression, logistic regression, and generalized linear regression, were widely used due to their ability to assess relationships between cost-related variables while accounting for multiple influencing factors (DOWLING, 2008; CUTLER; MCCLELLAN; MERRILL, 2013). Some studies implemented Bayesian hierarchical models and Gibbs sampling, which are particularly effective for analyzing highly skewed cost data or datasets with many zero values (NEUMANN; OLSEN; MCNAMEE, 2000; O'BRIEN et al., 2007). Additionally, machine learning techniques, such as decision trees and vector regression models, were explored for their predictive accuracy and adaptability to complex datasets (KSHIRSAGAR et al., 2020).

The analyzed variables varied significantly across studies, reflecting the diversity of approaches in cost estimation. Commonly considered factors include demographic variables (age, gender, race/ethnicity), health-related variables (chronic conditions, comorbidities, diagnosis codes), and insurance-related factors (plan characteristics, utilization patterns, and procedure codes) (CERRI, 2016; FRANÇA; GADELHA; TRAVASSOS, 2019). Some studies also incorporated economic variables (income, average expenses, copayments, and premiums paid), as well as geographic and accessibility indicators (residence proximity to healthcare facilities) (PAIM, 2018).

The rationale for the chosen methodologies highlights the advantages and limitations of different statistical and machine learning approaches. Regression models were favored for their interpretability and ability to control multiple independent variables, making them useful for evaluating cost determinants and risk adjustment strategies (GIL, 2008; YIN, 2018). On the other hand, machine learning models were recognized for their superior predictive accuracy, particularly in complex datasets where traditional regression methods may fall short (KSHIRSAGAR et al., 2020; TEEHUIS et al., 2022). Bayesian methods and hierarchical models were applied in cases where cost data exhibited high variability or required a probabilistic estimation approach (NEUMANN; OLSEN; MCNAMEE, 2000).

The analysis presented in Table 2 highlights the evolution of methodologies in health insurance cost prediction, demonstrating a transition from traditional statistical techniques to more advanced predictive models. These findings emphasize the importance of selecting appropriate methodologies based on the nature of the data, the specific objectives of each study, and the need for continuous methodological refinement to improve cost prediction accuracy in healthcare financing (YIN, 2018; ROBINSON; EPSTEIN; SHAH, 2014).

6.1.4. Discussion

In this study, we employed a powerful analytical technique known as the Word Cloud (WC) to analyze our results. Word Clouds are visually appealing representations of word frequencies in each text, where the size of each word corresponds to its frequency or significance. They have become a popular option for examining texts and communicating qualitative research findings (VASCONCELLOS-SILVA; ARAUJO-JORGE, 2019). The Word Cloud technique allowed us to conduct a fascinating lexical analysis, providing a graphical display of word occurrences and frequencies (KREUTZ et al., 2020).

In our investigation, we constructed the WC by extracting words from the abstracts of the 15 analyzed papers. The WC was created using the website www.wordclouds.com. Figure 3 illustrates the WC, showing that the generated results are in alignment with our research theme.

Figure 4 – Word cloud

Source: elaborated by the authors (2023) through Word clouds

Key terms like "health," "risk," and "expenditures" emerged as the most prominent and frequently occurring words. Additionally, we observed that words like "model," "regression," and "prediction" hold significant importance and are among the most extensively researched topics, as indicated in the subsequent stages of our study. The WC we developed facilitated a quick and intuitive analysis, effectively highlighting common themes present in our research results, a feature underscored by DePaolo and Wilkinson (2014).

6.1.5. Methodology studied

In this systematic review, a total of 15 articles were examined, each aiming to address various healthcare-related questions, with a particular focus on predicting healthcare costs.

Among the methodologies employed by the authors for cost prediction, regression analysis emerged as the most widely utilized, being chosen by 13 researchers. Regression analysis is a powerful statistical tool that enables the investigation of relationships between variables and accurate predictions based on observed data, (Gelman, Carlin, Stern, and Rubin, 2004;) making it a highly valuable and effective approach for predicting healthcare utilization and associated costs.

Additionally, other methodologies such as machine learning models and forecasting methods were also applied by different authors to explore their research objectives related to cost prediction. This impressive diversity of analytical approaches illustrates the researchers' commitment to employing sophisticated methods to forecast healthcare costs and effectively address essential questions in the healthcare field.

Out of the 13 articles that used regression, an impressive majority of 11 authors chose this method exclusively. However, the two authors adopted a more exploratory approach by combining regression with other cost prediction methods. They intended to compare the effectiveness of different approaches and gain deeper insights into the intricacies of cost prediction models. This comparative analysis provided a comprehensive understanding of the efficiency of various predictive techniques in healthcare cost forecasting.

6.1.6. Variables analyzed

Among the various studies reviewed, several variables were identified for predicting healthcare costs. The most frequently cited variables were age (appearing in 11 of the articles), gender (present in 10 of the studies), and costs (mentioned in 8 of the articles). Other commonly used variables included diagnoses (cited in 6 articles), procedures (found in 4 studies), and health plan characteristics (mentioned in 3 of the articles).

The age and gender of patients have been widely recognized as relevant factors for predicting healthcare costs, as health needs and utilization patterns can vary significantly among different age and gender groups. Costs, on the other hand, are fundamental variables for the economic and financial analysis of healthcare services, enabling an understanding of the resources required for medical care and identifying areas that demand greater attention and investment.

In addition to the mentioned variables, other studies also included information about diagnoses and procedures performed on patients. These pieces of information are crucial for understanding individuals' health conditions and the type of treatment they are receiving, thereby allowing for a better estimation of future costs based on their healthcare needs. In summary, the analysis of the studies revealed a variety of variables used for predicting healthcare costs, with age, gender, and costs being prominently cited by researchers.

6.1.7. Summary of results

Among the various methodologies analyzed, regression emerged as a dominant statistical technique, frequently employed by researchers to explore the relationship between dependent and independent variables. Its versatility lies in its ability to investigate how multiple factors simultaneously influence specific outcomes, making it particularly suitable for the complexity of predicting healthcare costs in the supplementary health sector.

Machine learning models have also gained prominence, though specific justifications for their use were often not detailed in the reviewed studies. These techniques stand out for their capacity to handle complex datasets, process extensive amounts of data, and deliver accurate predictions through iterative model training. For instance, weighted logistic regression was highlighted for its effectiveness in addressing binary data and predicting rare but critical events, such as high claims costs. This characteristic makes it an ideal choice for scenarios where infrequent outcomes significantly impact cost patterns.

Bayesian hierarchical zero-inflated gamma linear regression proved to be particularly adept at analyzing skewed and zero-inflated datasets typical of healthcare cost and utilization. Similarly, generalized linear regression and distributive estimators were recognized for their ease of implementation and the clarity of their results, offering actionable insights that are highly valued by researchers and decision-makers alike.

Other methodologies were also applied in more specific contexts. Linear models demonstrated their utility in analyzing longitudinal data with repeated measures, while exponential smoothing with error adjustment provided a scalable approach to predicting the volume and cost of specific drugs. Decision tree regression and machine learning estimators further enriched the landscape of predictive techniques, showcasing the adaptability of different methodologies to varied contexts within healthcare cost analysis.

The supplementary health market in Brazil has shown consistent growth yet concerns about the insolvency of health plan operators persist, primarily due to underwriting risks such as pricing inadequacies. Through a systematic integrative literature review conducted across four databases, 15 relevant studies were identified, offering a comprehensive perspective on methodologies for predicting healthcare costs.

Regression analysis consistently emerged as the most widely recognized and accepted approach across the reviewed studies. This finding aligns with the conclusions of authors such as Park, Morid, and Mark, who emphasized the superior reliability of regression techniques in healthcare cost prediction. Furthermore, age, gender, and historical costs were consistently identified as fundamental variables in predictive models, underscoring their critical role in accurately forecasting healthcare expenses.

In addition to regression-based approaches, the exploration of alternative methodologies reflects the growing interest in enhancing predictive accuracy. Techniques like machine learning and Bayesian methods not only complement traditional regression but also push the boundaries of predictive modeling in supplementary health.

The findings from this study are of significant relevance to health plan operators and decision-makers in the supplementary health market. By offering a solid knowledge base, the methodologies analyzed in this study have the potential to improve actuarial pricing processes and foster more informed decision-making. This, in turn, contributes to the sustainability and efficiency of health operations.

In conclusion, the results of this study can be used as a valuable guide for advancing healthcare cost prediction. By integrating insights from various methodologies, it supports the development of more robust pricing strategies and ensures the availability of accessible, effective, and sustainable healthcare services for the population reliant on supplementary health systems.

6.2. Development of the Health Plan Pricing Prototype

Following the systematic literature review, which identified the most relevant and up-to-date approaches to health plan pricing, this chapter presents the development of the pricing prototype based on the theoretical knowledge previously discussed. The data used for this model were provided by a health insurance operator headquartered in the state of Rio Grande do Sul, whose operations were outlined at the beginning of this study. The datasets, separated by product type (Regional Family Plan, Regional Business Plan, and National Corporate Plans), were organized and structured in Excel spreadsheets. From this foundation, it was possible to begin building the pricing model, as detailed in the following sections.

6.2.1. Data Extraction and Structuring

The prototype's development began with the extraction and organization of databases, individualized for each product (Regional Family Plan, Regional Business Plan, and Nacional Corporate Plans), considering their respective characteristics. This separation emphasized the importance of tailoring the analysis to the unique features of each product. To facilitate analysis, additional columns were created to calculate the total events per care category and the total copayment amount paid. These data, although not present in the original databases, were essential for modeling and were easily derived from the available information using formulas within Excel.

Data organization followed the guidelines of the National Supplementary Health Agency (ANS), which regulates the collection and submission of information by health plan operators (ANS, 2020). The ANS establishes a minimum dataset that must be reported, including information about beneficiaries, healthcare providers, and healthcare events.

6.2.2. Conversion of Categorical Data to Numerical Data

A crucial step was the conversion of the categorical variables "type of care" and "age group" into numerical representations. This transformation was necessary to enable the application of linear regression, which requires numerical inputs. The conversion followed a sequential order, assigning integers to each category, as shown in the following table:

Table 3 – Numerical Encoding of Care Categories and Age Groups

Category	Code	ANS Age Group	Code
Emergency Consultations	1	0 to 18	1
Office Consultations	2	19 to 23	2

Other Medical-Hospital Expenses	3	24 to 28	3
Complex Exams	4	29 to 33	4
Simple Exams	5	34 to 38	5
Hospitalization	6	39 to 43	6
Other Outpatient Care	7	44 to 48	7
Complex Therapies	8	49 to 53	8
Special Therapies	9	54 to 58	9
Simple Therapies	10	Over 59	10

Source: elaborated by the authors, 2025

The conversion of categorical data to numerical data is a common practice in data analysis, allowing the application of mathematical and statistical models that require numerical inputs (Field, 2018).

6.2.3. Multiple Linear Regression Modeling

The choice of multiple linear regression as the methodological approach in this study is grounded in both theoretical appropriateness and empirical consistency, as evidenced in the integrative literature review. Among the 15 studies analyzed, several adopted regression-based models to estimate healthcare costs, identify key cost drivers, and assess the impact of beneficiary characteristics on spending behavior.

Studies such as Morid (2017) and Politi (2018) emphasize that linear models are effective for predicting healthcare costs, especially when the goal is to quantify the relationship between multiple independent variables and a continuous cost outcome. Park (2018) further supports the use of regression estimators for their interpretability and performance in risk adjustment contexts, particularly when dealing with claims data that reflect real-world healthcare consumption.

Additionally, Xu (2019) and Lang (2010) used linear and generalized linear models to estimate out-of-pocket expenditures, showing that such methods are suitable for modeling cost-related outcomes even in the presence of complex covariates. Although more sophisticated approaches such as machine learning have been explored (Kshirsagar, 2021; Morid, 2017), linear regression models retain a comparative advantage due to their interpretability, transparency, and ease of operationalization, factors particularly relevant for private health operators and regulatory agencies such as ANS.

The independent variables included in the model were selected based on their consistent presence as cost predictors across the reviewed studies, as seen on Table 4.

Table 4 – Description of Selected Independent Variables for Cost Prediction Modeling

Category	***Code***	***Based on***	***Unit of Measurement***
Emergency Consultations	EC	Steenhuis (2022), Cecil (2006)	Number of consultations
Office Consultations	OC	Steenhuis (2022), Cecil (2006)	Number of consultations
Complex Exams	EX	Martei (2018), Morid (2017)	Number of procedures
Simple Exams	EX	Martei (2018), Morid (2017)	Number of procedures
Hospitalization	HOSP	Steenhuis (2022), Cecil (2006)	Number of admissions
Other Outpatient Care	AMB	Martei (2018), Morid (2017)	Number of procedures
Complex Therapies	TH	Steenhuis (2022)	Number of sessions
Special Therapies	TH	Steenhuis (2022)	Number of sessions
Simple Therapies	TH	Steenhuis (2022)	Number of sessions
Attendance Guides	AGD	Steenhuis (2022)	Number of guides issued
Co-payment	CP	Xu (2019), Cecil (2006)	R$ (monetary value)
Type of Care	TC	Robst (2006), Harrison (2020)	Category
Age Group	AG	Cao (2020), Lang (2010)	Age range according to ANS
Total Healthcare Cost	THC	Model output variable	R$ (monetary value)
Intercept	β_0	Gujarati and Porter (2009)	-
Error Term	ε	Gujarati and Porter (2009)	-

Source: elaborated by the authors, 2025

Age group (AG) is highlighted in studies by Cao (2020) and Lang (2010) as a major determinant of healthcare cost variation, due to its association with chronic conditions and service demand.

- Type of care (TC), distinguishing outpatient from inpatient services, aligns with categorizations used in studies like Robst (2006) and Harrison (2020), which examined care settings and their differential cost impacts.
- Medical consultations (OC, EC) and hospitalizations (HOSP) were shown to be significant cost drivers in Steenhuis (2022) and Cecil (2006), reflecting both frequency and intensity of healthcare utilization.
- Ambulatory procedures (AMB) and diagnostic exams (EX) follow similar logic, supported by Martei (2018) and Morid (2017), where specific categories of procedures significantly contributed to cost projections.
- Therapies (TH) and attendance guides (AGD) were included to account for the diversity and volume of non-hospital procedures and administrative burden, respectively—factors noted as underexplored yet relevant in Steenhuis (2022).

▪ Total co-payment (CP) was included following findings from Xu (2019) and Cecil (2006), which show that cost-sharing mechanisms influence both access behavior and overall spending patterns.

In summary, the adoption of multiple linear regression and the selected explanatory variables are solidly supported by the literature. The model integrates demographic, clinical, financial, and operational dimensions of healthcare utilization, offering a comprehensive and interpretable framework for cost prediction in private health plans, aligned with ANS technical requirements and academic best practices.

Equation 1: General formulation of the multiple linear regression model

$$\mathrm{THC} = \beta^0 + \beta^1 * \mathrm{AG} + \beta^2 * \mathrm{TC} + \beta^3 * \mathrm{OC} + \beta^4 * \mathrm{EC} + \beta^5 * \mathrm{AMB} + \beta^6 * \mathrm{HOSP} + \beta^7 * \mathrm{EX} + \beta^8 * \mathrm{TH} + \beta^9 * \mathrm{AGD} + \beta^{10} * \mathrm{GE} + \beta^{11} * \mathrm{CP} + \varepsilon$$

Source: elaborated by the authors, 2025

Where:

- **THC**: *Total Healthcare Cost* – Total monthly cost incurred by the operator for healthcare services (R$).
- **β_0**: *Intercept* – Expected cost when all independent variables are zero.
- **AG**: *Age Group* – Age range of the beneficiary according to ANS classification.
- **TC**: *Type of Care* – Category of healthcare services utilized.
- **OC**: *Office Consultations* – Number of scheduled outpatient consultations.
- **EC**: *Emergency Consultations* – Number of consultations performed in emergency or urgent care.
- **AMB**: *Ambulatory Services* – Number of outpatient services not classified as consultations or exams.
- **HOSP**: *Hospitalizations* – Number of inpatient admissions.
- **EX**: *Exams* – Combined number of simple and complex exams performed.
- **TH**: *Therapies* – Combined number of therapy sessions, including simple, complex, and special therapies.
- **AGD**: *Attendance Guides* – Number of healthcare service authorization guides issued.
- **GE**: *General Events* – Total number of healthcare procedures utilized by the beneficiary.
- **CP**: *Co-payment* – Total monthly amount paid by the beneficiary as part of cost-sharing (R$).

- **ε**: *Error Term* – Random variation in THC not explained by the independent variables.

The regression analysis was conducted separately for each health plan product, resulting in distinct equations tailored to the characteristics of each group. The coefficients were obtained using the data analysis tool in Excel, specifically the linear regression function. These coefficients were then used to calculate the estimated costs for each combination of age group and care type.

As highlighted by Gujarati and Porter (2009), multiple linear regression is a widely used statistical method for modeling the relationship between a continuous dependent variable and multiple independent variables, a technique commonly applied in economic and actuarial analyses. The selection of independent variables for this study was guided by existing literature on the factors influencing health plan costs, as well as the availability of relevant data.

6.2.4. Cost and Utilization Indicator Calculations

Based on the costs estimated by regression, several indicators were calculated for each combination of age group and care type:

- Copayment: Represents the portion of healthcare expenses paid directly by the beneficiary as part of the cost-sharing arrangement. It is calculated as the difference between the total cost of care and the amount covered by the operator without considering copay.
- Copayment Recovery Percentage: It indicates the proportion of the total care cost that was recovered from the beneficiary through copayment, reflecting the extent to which cost-sharing contributes to offsetting the operator's expenses. This metric is calculated as shown above:

Equation 2: Copayment

$$Copayment = \frac{Copayment}{Care\ Cost} * 100$$

Source: elaborated by the authors, 2025

- Frequency: Represents the number of events per beneficiary per year. This metric is calculated as shown above:

Equation 3: Frequency

$$Frequency = \frac{Number\ of\ Events}{Number\ of\ Beneficiaries}$$

Source: elaborated by the authors, 2025

- Average Cost per Procedure: Represents the total cost of the procedure divided by the number of events. This metric is calculated as shown above:

Equation 4: Average Cost per Procedure

$$Average\ Cost\ per\ Procedure = \frac{Total\ Cost}{Number\ of\ Events}$$

Source: elaborated by the authors, 2025

- Risk Premium: Represents the product of frequency by the average cost per procedure. The premium, in the actuarial context, represents the value that the beneficiary must pay to have the right to health plan coverage (Bowers et al., 1997). This metric is calculated as shown above:

Equation 5: Risk Premium

$$Risk\ Premium = \frac{Frequency}{Average\ Cost\ per\ Procedure}$$

Source: elaborated by the authors, 2025

These indicators are important for analyzing the use of healthcare services and for evaluating the financial performance of health plans (Novaes, 2000).

6.2.5. Financial Adjustment by IPCA

To adjust the observed costs for inflation, the general National Consumer Price Index (IPCA) was applied. The adjusted cost was calculated by multiplying the observed cost by (1 + IPCA), where IPCA is the index for the corresponding period. The risk premium was also adjusted by IPCA.

IPCA is Brazil's official inflation index, calculated by the Brazilian Institute of Geography and Statistics (IBGE). The use of IPCA to adjust observed costs allows considering the impact of inflation on health plan costs (IBGE, 2023).

6.2.6. Calculation of Safety Margin and Pure Premium

The safety margin (SM) was calculated for each age group, reflecting the variation in costs over the months analyzed. The equation used was:

Equation 6: Safety Margin

$$MSE\ = \frac{(Z_{value}\ (\frac{\sigma}{\sqrt{n}}))}{\mu}$$

Source: elaborated by the authors, 2025

Where:

Z_{value}: Z_{value} corresponding to the confidence level (80% or 99%). The z-value is a statistical value that represents the number of standard deviations a given value is from the mean. In a normal distribution, the z-value can be used to calculate the probability of a given value occurring. For example, a z-value of 1.96 corresponds to a 95% confidence level, meaning there is a 95% probability of a value being within 1.96 standard deviations of the mean (Triola, 2018).

n: Number of observations - In this case, the number of months observed.

μ: Average cost - The average of the costs per month observed throughout the year for the specific age group.

σ: Standard deviation of costs in the age group - The standard deviation of costs observed throughout the year for the specific age group.

The total risk premium was obtained by summing the risk premiums per care type, reflecting the total expected cost based on the use of healthcare services. The pure premium was calculated by applying the SM to the total risk premium, adding a safety margin to cover unexpected variations in costs. The adjusted pure premium was obtained after adjusting the variations between age groups to the parameters of ANS RN 63, ensuring that the variation in premiums between age groups is in compliance with ANS regulation.

6.2.7. Calculation of Commercial Premium and Comparison with Observed Costs

The final commercial premium was calculated by adding loading costs (administrative, commercial, profit margin, and taxes) to the adjusted pure premium. This premium represents the value that will be effectively charged to beneficiaries, encompassing all expenses and the operator's profit margin.

To ensure compliance with regulatory standards, the limits for registration in the Actuarial Technical Note (NTA) were defined following the guidelines of National Supplementary Health Agency (ANS) Normative Resolution (RN) No. 564. The NTA, now referred to as the Product Registration Technical Note (NTRP), is a technical document detailing the actuarial calculations used in health plan pricing, and RN 564 establishes the parameters for its registration with the ANS.

Normative Resolution No. 564, of December 16, 2022, addresses the Product Registration Technical Note (NTRP), justification for the initial price formation of supplementary health care products, and requirements for obtaining product registration with the ANS. This standard details the information that must be included in the NTRP, including

Annexes II-A and II-B, which present the actuarial calculations and assumptions used in pricing.

The limits for registration in the NTRP were calculated as follows, as established by RN 564:

Maximum Limit: The maximum limit allowed for registering the premium in the NTRP is 30% above the commercial premium. This limit aims to prevent the charging of excessively high premiums from beneficiaries.

Minimum Limit: The minimum limit is defined as the greater value between the adjusted pure premium and 70% of the commercial premium. This limit ensures that the premium registered in the NTRP is sufficient to cover healthcare costs and a portion of the operator's administrative expenses.

This methodology for calculating the limits ensures that the premiums registered in the NTRP are within an acceptable range, protecting both beneficiaries and the operator.

The calculated premiums were compared with the costs observed in the following period to assess pricing efficiency. This comparison allows verifying if the calculated premiums were sufficient to cover the costs incurred in the following period, ensuring the plan's financial sustainability.

6.2.8. Preparation of Annexes of ANS RN 564 and Adjustment to RN 563

The final results were organized in the Annexes of ANS RN 564. RN 564 establishes the format and minimum information that must be included in the actuarial technical notes, now called NTRP, which are documents that prove the adequacy of the premiums charged by health plan operators.

The organization of the results in these annexes ensures that the NTRP is complete and in compliance with ANS requirements, facilitating the document's analysis and registration.

Additionally, the adjusted pure premium was obtained after adjusting the variations between age groups to the parameters of ANS RN 563. Normative Resolution No. 563, of December 15, 2022, addresses the limits to be observed for the adoption of price variation by age group in private health care plans contracted from January 1, 2004. This standard establishes the maximum percentages of premium variation between the 10 (ten) age groups, ensuring that the premium variation between age groups complies with ANS regulations.

RN 563 revoked RN 63, updating the rules for price variation by age group. The main provisions of RN 563 include:

- Adoption of 10 (ten) age groups, according to the table specified in the standard.

- The value set for the last age group cannot be more than six times the value of the first age group.
- The accumulated variation between the seventh and tenth age groups cannot be greater than the accumulated variation between the first and seventh age groups.
- Variations by age group change cannot present negative percentages.

By adjusting the adjusted pure premiums to the parameters of RN 563, the operator ensures that the premium variation between age groups is within the limits established by the ANS, protecting beneficiaries from excessive premium increases due to age.

The combination of compliance with RN 564 and RN 563 ensures that health plan pricing is carried out transparently, fairly, and in compliance with ANS regulatory standards.

6.2.1. Prototype summary flowchart

The development of the Health Plan Pricing Prototype followed a systematic and structured approach, which is visually summarized in

Figure 5 – Prototype Summary Flowchart. This flowchart presents all key stages involved in the prototype's construction, from the initial data preparation to the final adjustments for regulatory compliance.

The process begins with the extraction and structuring of data, individualized by product type and aligned with ANS (2020) guidelines. Then, categorical variables are converted into numerical values to allow the application of multiple linear regression modeling, enabling cost estimation per beneficiary.

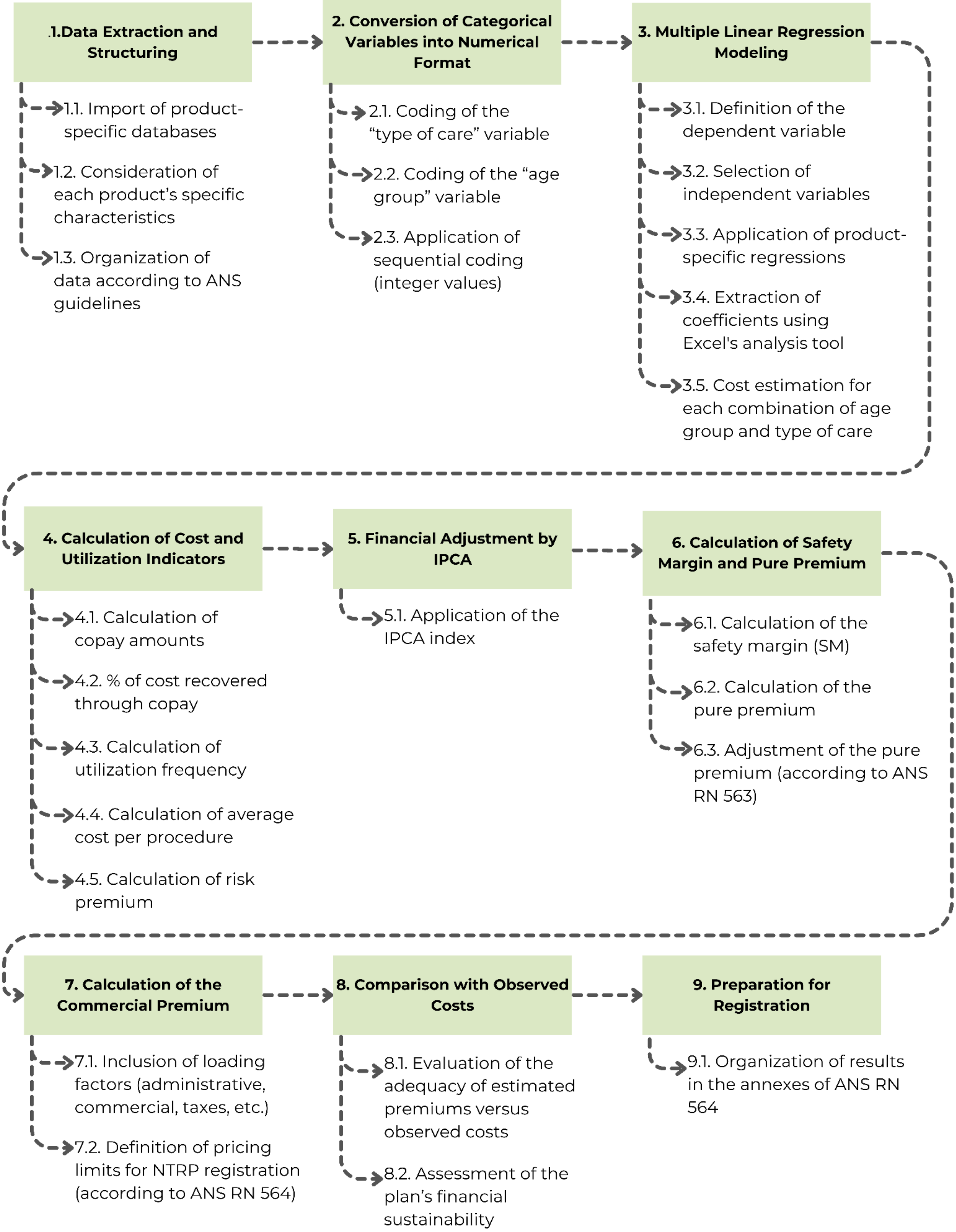
Subsequently, cost and utilization indicators are calculated to support the pricing structure. These are followed by a financial adjustment using the IPCA index, and the computation of the safety margin and pure premium, which are later adjusted according to ANS Normative Resolution No. 563.

Based on these adjusted values, the commercial premium is determined, including administrative and operational loadings. A comparison with observed costs is conducted to validate the adequacy of the pricing model. Finally, the results are structured in accordance with ANS RN 564 for proper registration.

This summarized flow, shown in

Figure 5, provides a clear and organized overview of the prototype's logic and implementation process.

Figure 5 – Prototype summary flowchart



Source: elaborated by the authors, 2025

6.3. Pricing Methodology

Once the prototype structure was defined and the dataset was organized and coded accordingly, the next step involved applying statistical methods to analyze the data and support the construction of pricing models. The following section presents the descriptive analysis of healthcare utilization and cost patterns across different plan types, providing a comprehensive understanding of the dataset's characteristics and variability. This stage is fundamental to identifying trends and differences among the plans, thereby informing the model design process.

6.3.1. Descriptive analysis

Descriptive statistics are essential for analyzing healthcare data, enabling the summarization and interpretation of large datasets. As highlighted by Montgomery and Runger (2021), these statistics are instrumental in data-driven decision-making, providing key insights into patterns, variations, and trends. This study analyzes total healthcare costs, cost-sharing, and clinical event frequencies across different health insurance plan types. For each plan, we calculated the mean, median, standard deviation, minimum and maximum values, and the coefficient of variation (CV) to assess data variability and dispersion, facilitating comparisons across plans.

The coefficient of variation (CV) quantifies the relative variability of data points in relation to the mean. Calculated as the ratio of the standard deviation to the mean (often expressed as a percentage), the CV allows for comparison across datasets with different units or scales. A lower CV indicates less variability, with values $\leq 15\%$ considered low, $15\% < CV \leq 30\%$ moderate, and $CV > 30\%$ high (Brunner & Langer, 2019). This metric enables comparison of data dispersion, even when variables are expressed in different units of measurement. All costs are reported on a monthly basis unless otherwise stated.

The CV ranges used to classify data variability (low, moderate, high) are based on statistical conventions and provide general guidelines for interpreting dispersion. While these thresholds may vary slightly depending on the specific field of study, they are widely used in the literature and offer a practical approach to assessing variability.

The following sections present the findings for each health insurance plan. It is important to note that "general events" represent the total number of healthcare procedures, including simple exams, office consultations, therapies, and hospitalizations. This definition is crucial for understanding the scope of the analysis, as it encompasses a wide range of services utilized by beneficiaries. The dataset analyzed has been previously described.

The overall average total healthcare cost is R$ 11,802,296 (SD: R$ 2,748,055), indicating considerable expenditure variation. The average cost-sharing amount is R$ 632,678 (CV: 32%), suggesting high variability in patient out-of-pocket payments. General events averaged 73,730 occurrences per month, with simple exams (46,046) and office consultations (13,667) being the most common, reflecting strong demand for routine care. Simple exams (CV: 19%) and office consultations (CV: 17%) exhibited moderate variability, while complex therapies (CV: 38%) and hospitalizations (CV: 16%) showed greater dispersion, reflecting diverse patient needs and treatment intensities. These cost and utilization metrics are summarized in Table 5.

All monetary values presented refer to monthly averages, calculated over a six-year period (January 2019 to December 2024), and are expressed in Brazilian reais (R$). Similarly, all non-monetary values refer to the monthly average quantity of procedures performed during the same period and are expressed as number of events (e.g., consultations, exams, therapies, hospitalizations). This distinction ensures clarity when interpreting both financial and operational dimensions of health plan utilization.

To ensure consistency and support further analyses, the variables included in the descriptive statistics are identified below with their respective acronyms and units of measurement:

- THC (Total Healthcare Cost) – Total monthly cost incurred by the operator for all healthcare services (R$).
- CP (Co-payment) – Total monthly amount paid by beneficiaries as part of cost-sharing mechanisms (R$).
- GE (General Events) – Total number of healthcare service events recorded per month (number of procedures).
- EC (Emergency Consultations) – Number of emergency or urgent care consultations per month (number of procedures).
- OC (Office Consultations) – Number of scheduled outpatient consultations (number of procedures).
- CE (Complex Exams) – Number of high-complexity diagnostic procedures performed monthly (number of procedures).
- SE (Simple Exams) – Number of low-complexity diagnostic tests conducted monthly (number of procedures).
- HOSP (Hospitalizations) – Number of inpatient hospital admissions per month (number of procedures).

- AMB (Ambulatory Services) – Number of outpatient services not classified as consultations or exams (number of procedures).
- CT (Complex Therapies) – Number of complex therapeutic sessions, such as chemotherapy or dialysis (number of procedures).
- ST (Special Therapies) – Number of sessions involving specialized therapies, particularly those used in the treatment of autism spectrum disorder (e.g., ABA) (number of procedures).
- SIT (Simple Therapies) – Number of basic therapy sessions, including physical, occupational, and speech therapies (number of procedures).

These variables were derived directly from the operator's internal records and processed in accordance with the standards established by the National Supplementary Health Agency (ANS). The descriptive statistics summarized here serve as the analytical foundation for the pricing modeling and forecasting presented in the subsequent sections.

Table 5 – Descriptive statistics of the general dataset

	THC	CP	GE	EC	OC
Mean	R$ 11.802.296	R$ 632.678	73.730	3.917	13.667
Median	R$ 11.553.714	R$ 604.616	73.619	4.094	13.699
Standard Deviation	R$ 2.748.055	R$ 201.107	13.774	1.193	2.376
Minimum	R$ 6.808.169	R$ 220.314	30.120	1.351	6.811
Maximum	R$ 18.699.715	R$ 1.114.218	101.278	6.431	18.444
Coefficient of Variation	23%	32%	19%	30%	17%

	CE	SE	HOSP	AMB	CT	ST	SIT
Mean	1.684	46.046	310	1.412	176	79	6.440
Median	1.688	46.314	314	1.419	160	35	6.495
Standard Deviation	482	8.539	49	327	67	95	1.413
Minimum	457	17.839	168	499	79	1	2.709
Maximum	2.377	64.037	399	2.332	362	373	9.277
Coefficient of Variation	29%	19%	16%	23%	38%	121%	22%

Source: elaborated by the authors, 2025

A notable aspect of the dataset is the substantial variation observed in special therapies, which present the highest coefficient of variation (121%). This trend can be attributed to the increasing demand for treatments related to Autism Spectrum Disorder (ASD), particularly Applied Behavior Analysis (ABA) and other intensive therapeutic interventions. Studies have shown that the costs associated with ASD-related therapies have been rising due to the growing recognition of early interventions as essential for long-term developmental outcomes (Karami

Matin et al., 2022). Additionally, research indicates that the financial burden of ASD treatment is significantly higher than that of other chronic conditions, as many children require continuous and multidisciplinary care over extended periods (Cidav et al., 2017). These factors contribute to the observed volatility in special therapies, reinforcing the need for more precise forecasting and policy adjustments to support affected families.

Family plan beneficiaries incurred an average total healthcare cost of R$ 5,465,014 (CV: 22%), showing moderate cost variability, lower than the overall dataset. The average cost-sharing amount was R$ 205,628 (SD: R$ 55,522), indicating relatively consistent out-of-pocket costs. General events averaged 32,108, primarily simple exams (20,230) and office consultations (5,852). General events (CV: 16%) and office consultations (CV: 13%) showed relatively consistent utilization, while simple exams (CV: 16%) exhibited moderate variability. Complex therapies (CV: 41%) and special therapies (CV: 110%) showed high variability. As shown in Table 6, family plan beneficiaries exhibit a distinct pattern of healthcare service utilization compared to the overall dataset. For instance, the variability in "complex therapies" is notably higher for family plan beneficiaries (CV: 41%) than in the general dataset (CV: 38%), suggesting a wider range of complex healthcare needs within this group.

Table 6 – Descriptive statistics of the family plan

	THC	CP	GE	EC	OC
Mean	R$ 5.465.014	R$ 205.628	R$ 32.108	1.492	5.852
Median	R$ 5.408.366	R$ 206.683	R$ 32.779	1.515	6.012
Standard Deviation	R$ 1.191.253	R$ 55.522	R$ 5.027	373	766
Minimum	R$ 3.041.038	R$ 70.172	R$ 13.962	639	3.245
Maximum	R$ 8.629.909	R$ 321.028	R$ 41.317	2.367	7.275
Coefficient of Variation	22%	27%	16%	25%	13%

	CE	SE	HOSP	AMB	CT	ST	SIT
Mean	765	20.230	146	587	87	47	2.901
Median	769	20.353	148	600	81	21	2.981
Standard Deviation	219	3.322	24	121	36	52	577
Minimum	178	8.276	72	208	29	1	1.186
Maximum	1.067	26.980	194	888	194	201	3.953
Coefficient of Variation	29%	16%	16%	21%	41%	110%	20%

Source: elaborated by the authors, 2025

The regional corporate plan had an average total healthcare cost of R$ 4,360,884 (CV: 14%), indicating low cost variability, lower than the overall dataset. The average cost-sharing amount was R$ 219,099 (SD: R$ 50,432), reflecting relatively consistent out-of-pocket costs. General events averaged 29,259, mainly simple exams (18,517) and office consultations (5,314). General events (CV: 15%) and office consultations (CV: 14%) showed stable utilization, while simple therapies (CV: 19%) and complex therapies (CV: 34%) showed greater variability. Special therapies (CV: 103%) also demonstrated high variability. Refer to Table 7 for a detailed breakdown of the regional corporate plan data.

Table 7 – Descriptive statistics of the regional corporate plan

	THC	CP	GE	EC	OC
Mean	R$ 4.360.884	R$ 219.099	R$ 29.259	1.493	5.308
Median	R$ 4.336.909	R$ 228.098	R$ 30.022	1.459	5.313
Standard Deviation	R$ 612.944	R$ 50.432	R$ 4.419	398	751
Minimum	R$ 2.808.764	R$ 97.180	R$ 12.823	589	2.919
Maximum	R$ 5.649.432	R$ 345.166	R$ 37.707	2.534	7.046
Coefficient of Variation	14%	23%	15%	27%	14%

	CE	SE	HOSP	AMB	CT	ST	SIT
Mean	677	18.517	117	554	68	22	2.510
Median	704	18.925	118	558	65	15	2.509
Standard Deviation	177	2.929	19	109	23	22	477
Minimum	176	7.442	75	216	30	1	1.243
Maximum	921	24.185	160	835	127	90	3.587
Coefficient of Variation	26%	16%	16%	20%	34%	103%	19%

Source: elaborated by the authors, 2025

The national corporate plan showed the highest variability in total healthcare costs, averaging R$ 1,976,398 (CV: 57%). The average cost-sharing amount was R$ 207,951 (CV: 55%), also exhibiting high variability. General events averaged 12,364, predominantly simple exams (7,300) and office consultations (2,507). General events (CV: 50%) and office consultations (CV: 59%) showed some variability, but simple therapies (CV: 64%), complex therapies (CV: 76%), and special therapies (CV: 113%) all exhibited high variability. Table 8 provides the detailed metrics for the national corporate plan. It is worth noting that the national corporate plan exhibits the highest variability across most service categories, particularly in "special therapies" (CV: 113%), highlighting the diverse healthcare needs and potentially unequal distribution of healthcare resources within this beneficiary group.

Table 8 – Descriptive statistics of the national corporate plan

	THC	CP	GE	EC	OC
Mean	R$ 1.976.398	R$ 207.951	R$ 12.364	932	2.507
Median	R$ 1.421.027	R$ 159.781	R$ 9.622	553	1.772
Standard Deviation	R$ 1.145.853	R$ 115.313	R$ 6.200	659	1.470
Minimum	R$ 717.193	R$ 52.962	R$ 3.335	123	647
Maximum	R$ 4.495.256	R$ 479.063	R$ 24.503	2.309	5.421

Coefficient of Variation	14%	23%	15%	27%	14%

	CE	SE	HOSP	AMB	CT	ST	SIT
Mean	242	7.300	47	270	21	23	1.029
Median	200	6.273	38	204	15	12	729
Standard Deviation	111	3.186	23	148	16	26	663
Minimum	56	2.121	12	75	3	1	280
Maximum	490	13.811	89	612	67	112	2.546
Coefficient of Variation	46%	44%	48%	55%	76%	113%	64%

Source: elaborated by the authors, 2025

The descriptive statistical analysis conducted in this chapter allowed for the identification of significant differences among the analyzed health insurance plans. As highlighted by Cameron and Trivedi (2005), descriptive statistics are essential for understanding healthcare service utilization patterns and can aid in formulating more effective policies for managing healthcare costs. This assertion is supported by other authors, such as Sanhueza et al. (2017), who emphasize the importance of descriptive statistics for analyzing health data and identifying relevant trends and patterns for decision-making. The variability observed in the coefficients of variation suggests that certain healthcare services exhibit greater dispersion in utilization, which may reflect differences in beneficiary profiles or plan structures. This variability is an important aspect to consider when analyzing costs and healthcare service utilization, as pointed out by Glied and Lleras-Muney (2008), who discuss the importance of taking into account the heterogeneity of beneficiaries when analyzing health data.

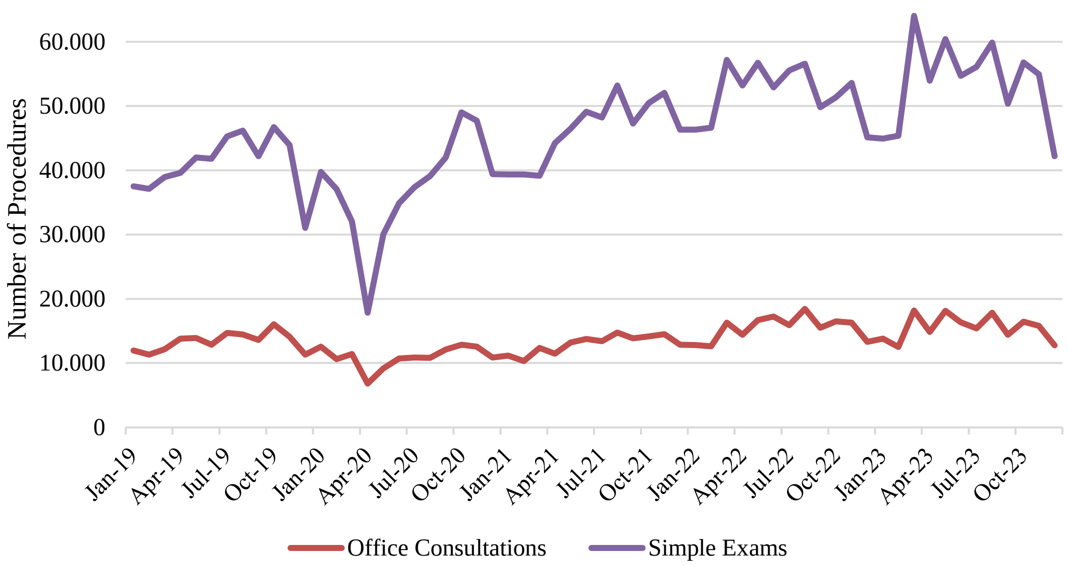
The family and regional corporate plans show significantly lower coefficients of variation, indicating greater stability in the data and lower volatility in the analyzed variables. This behavior suggests consistency in the estimates and can be interpreted as an indicator of reliability for these plan samples. This lower variability can be attributed to several factors, such as the homogeneity of beneficiary groups, the nature of services offered by the plans, or even the way data were collected and processed, as discussed by Barros and Victora (2006). In contrast, the national corporate plan shows the highest coefficients of variation, reflecting greater volatility in the data. High coefficients of variation are often associated with greater uncertainty and dispersion in the analyzed variables, which can reduce the reliability of conclusions drawn from the sample. This greater variability may be the result of a combination of factors, such as the diversity of beneficiaries, the complexity of services offered by the plan, or the presence of atypical events that affect costs and service utilization, as highlighted by Farias et al. (2010). The presence of a high coefficient of variation may indicate that the data is more dispersed around the mean, making statistical analysis more susceptible to distortions and increasing the need for caution when interpreting the results. This need for caution is emphasized by other authors, such as Altman and Bland (1995), who discuss the importance of considering the variability of data when interpreting statistical results and avoiding hasty generalizations.

6.3.2. Temporal Trends in Healthcare Utilization

To deepen the understanding of these patterns, the following section presents a temporal analysis of selected variables related to healthcare service utilization. By observing their behavior over time, it becomes possible to identify trends, seasonal effects, and potential anomalies that may impact the performance and sustainability of each plan. This longitudinal perspective allows for a more dynamic and contextualized view of the data, complementing the static analysis previously discussed. Such visualizations are particularly useful for highlighting the evolution of key indicators—such as total cost, co-payment, number of procedures, and types of services used—enabling more informed interpretations and facilitating strategic decision-making.

Figure 6 presents the monthly evolution of the number of office consultations and simple exams over the analyzed period. It can be observed that the volume of consultations remained relatively stable, with a slight drop in April 2020, directly attributed to the restrictions imposed by the COVID-19 pandemic. In contrast, simple exams exhibited a more volatile trajectory, with significant fluctuations over time. In addition to the initial decline in April 2020, this category reached volumes ranging from a low of approximately 18,000 to a peak exceeding 64,000 in certain months. This pattern indicates a higher sensitivity of simple exams to seasonal, operational, or contextual factors, which warrants special attention when modeling related costs. However, as these procedures typically have a low unit cost, their variability tends to have a limited impact on total healthcare expenditures.

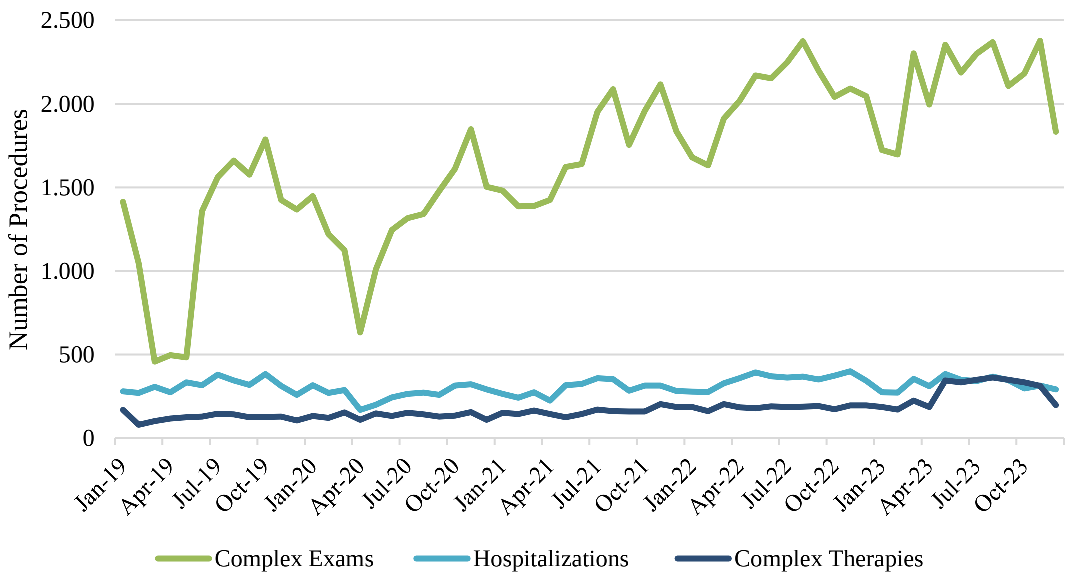
Figure 6 – Number of Office Consultations and Simples Exams



Source: elaborated by the authors, 2025

Figure 7 presents the monthly trends in the number of complex exams, hospitalizations, and complex therapies. All three categories experienced a notable drop in April 2020, reflecting the immediate impact of the COVID-19 pandemic. The number of hospitalizations and complex therapies remained relatively stable over time, with hospitalizations consistently exceeding complex therapies until April 2023. From that point onward, complex therapies surpassed hospitalizations in volume and continued to remain higher. In contrast, the number of complex exams showed considerable variation throughout the period, with a clear upward trend, indicating increasing demand for these diagnostic procedures.

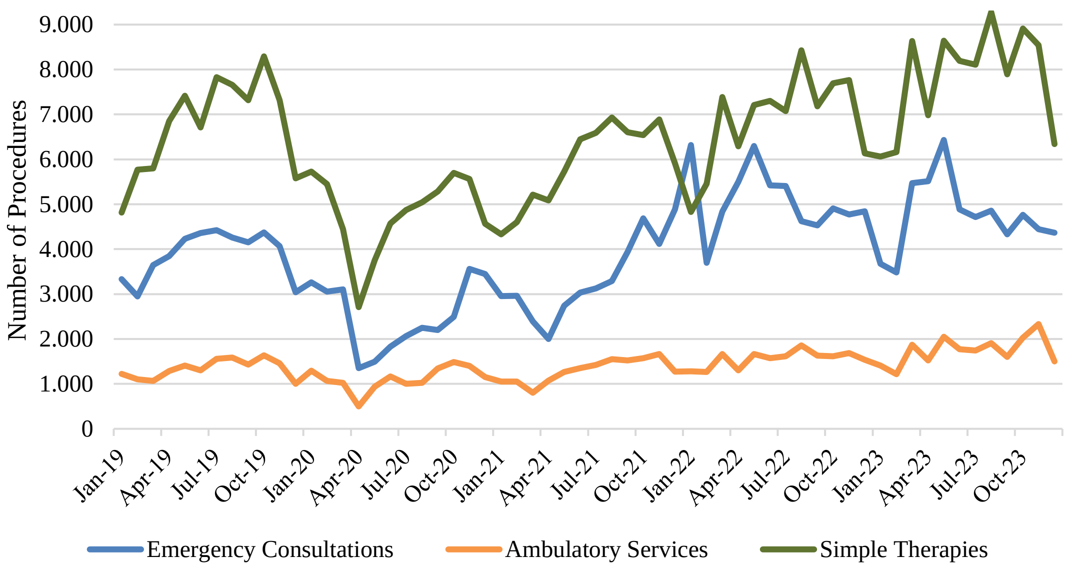
Figure 7 – Number of Complex Exams, Complex Terapies and Hospitalizations



Source: elaborated by the authors, 2025

Figure 8 shows the monthly evolution in the number of emergency consultations, ambulatory services, and simple therapies. All three categories experienced a drop in April 2020, consistent with the impact of the COVID-19 pandemic. Ambulatory services demonstrated greater stability throughout the analyzed period, with only minor fluctuations. In contrast, both emergency consultations and simple therapies exhibited more pronounced oscillations over time, yet followed similar trends, moving in parallel throughout the observed months.

Figure 8 – Number of Emergency Consultations, Ambulatory Services and Simples Terapies



Source: elaborated by the authors, 2025

6.3.3. Scenarios Definition

Based on the five scenarios previously defined in this study, simulations were conducted to estimate the projected behavior of healthcare costs and risk premiums under different conditions of service utilization. The applied variation rates for each type of care are shown in Table 9 below:

Table 9 - Average variations per services

SERVICE TYPE	AVERAGE VARIATION	AVERAGE VARIATION (EXCLUDING 2020)
Emergency Consultations	10.8%	10.5%
Office Consultations	5.1%	5.6%
Complex Exams	15.3%	21.2%
Simple Exams	7.7%	9.4%
Hospitalization	1.7%	1.8%

Other Ambulatory Services	8.0%	9.7%
Complex Therapies	23.4%	31.7%
Special Therapies	23.4%	31.7%
Simple Therapies	5.9%	5.7%

Source: Author's own work

The goal of these simulations is to assess the financial adequacy and resilience of different health plans when faced with both favorable and adverse utilization patterns. This analysis allows for identifying the most sensitive products, the implications of underpricing or overpricing, and the necessary pricing adjustments to maintain financial sustainability across different demand scenarios.

In the following sections, the results of these projections are presented in detail, showing the behavior of each product type—Family, Regional Corporate, and National Corporate—under the five distinct scenarios, with comparisons to established risk thresholds and safety margins.

6.3.4. Results

This chapter presents the results of the simulations for different plans and scenarios. We use two distinct approaches: (i) annual projections based on the immediately preceding year and (ii) simulations with historical data (2019–2024). Additionally, we calculate the results considering two safety margins (SM) – 80% and 99% – to reflect different levels of caution in the projections. Each set of simulations covers five scenarios (Baseline, Pessimistic Scenario 1, Pessimistic Scenario 2, Optimistic Scenario 1, and Optimistic Scenario 2) applied to three plans: Regional Family Plan, Regional Business Plan, and Nacional Corporate Plans .

Initially, we perform the simulations with a safety margin of 80%, using the data from the immediately preceding year for each forecast. The tables below present the results for each of the plans.

Table 10 – Family Regional Plan SM 80% (year/year)

Scenario /Year	2020	2021	2022	2023	2024
Baseline Scenario	82,03%	88,00%	89,22%	99,15%	116,54%
Pessimistic Scenario 1	88,53%	94,86%	96,45%	107,91%	119,75%
Pessimistic Scenario 2	90,18%	96,72%	98,46%	110,47%	122,26%
Optimistic Scenario 1	75,30%	81,00%	81,89%	90,09%	101,25%
Optimistic Scenario 2	73,65%	79,14%	79,87%	87,53%	98,74%

Source: elaborated by the authors, 2025

In Table 10, it is observed that for the Regional Family Plan, the projected values show a growth trend over the years, especially in the Baseline scenario, which ranges from 82.03% in 2020 to 116.54% in 2024. The pessimistic scenarios (1 and 2) show projections slightly higher than the Baseline, while the optimistic scenarios indicate lower values, demonstrating the sensitivity of this plan to variations in the scenario assumptions.

Table 11 – Regional Corporate Plans SM 80% (year/year)

Scenario /Year	2020	2021	2022	2023	2024
Baseline Scenario	89,41%	70,65%	64,87%	84,17%	69,43%
Pessimistic Scenario 1	95,67%	75,25%	69,87%	90,46%	73,88%
Pessimistic Scenario 2	97,16%	76,44%	71,19%	92,20%	75,29%
Optimistic Scenario 1	82,64%	65,51%	59,74%	77,28%	62,89%
Optimistic Scenario 2	81,16%	64,32%	58,41%	75,54%	61,47%

Source: elaborated by the authors, 2025

Table 11 presents the results for the Regional Business Plan, where the values demonstrate greater oscillation. Notably, the Baseline scenario begins with 89.41% in 2020 but drops to 64.87% in 2022, partially recovers in 2023, and falls again in 2024. This volatility highlights the higher uncertainty inherent in this plan, with the pessimistic and optimistic scenarios reflecting variations that can significantly influence decision-making.

Table 12 – Nacional Corporate Plans SM 80% (year/year)

Scenario /Year	2020	2021	2022	2023	2024
Baseline Scenario	113,75%	88,08%	74,75%	66,08%	64,96%
Pessimistic Scenario 1	120,43%	92,17%	79,56%	70,28%	68,34%
Pessimistic Scenario 2	122,49%	93,91%	80,96%	71,58%	69,45%
Optimistic Scenario 1	103,44%	78,63%	68,07%	59,80%	58,62%
Optimistic Scenario 2	101,38%	76,89%	66,67%	58,50%	57,51%

Source: elaborated by the authors, 2025

Table 12 demonstrates that the Nacional Corporate Plans begins with high projections (e.g., 113.75% in the 2020 Baseline scenario) but shows a sharp decline over time, reaching 64.96% in 2024. This downward trend occurs across all scenarios, suggesting that as the forecast horizon widens, projections tend to adjust to lower values, possibly reflecting the volatility of the national environment.

Following this, simulations with a higher safety margin (99%) performs, again using data from the immediately preceding year.

Table 13 – Family Regional Plan SM 99% (year/year)

Scenario /Year	2020	2021	2022	2023	2024
Baseline Scenario	64,85%	69,42%	72,62%	82,39%	93,72%
Pessimistic Scenario 1	69,99%	74,84%	78,50%	89,67%	96,30%
Pessimistic Scenario 2	71,30%	76,30%	80,14%	91,80%	98,32%
Optimistic Scenario 1	59,53%	63,90%	66,65%	74,86%	81,42%
Optimistic Scenario 2	58,23%	62,44%	65,01%	72,74%	79,40%

Source: elaborated by the authors, 2025

Table 13 shows that increasing the safety margin to 99% results in significantly lower projections for the Regional Family Plan. For example, the Baseline scenario records 64.85% in 2020 and 93.72% in 2024, values lower than those observed with the 80% margin. This reduction is consistent across all scenarios, demonstrating the direct influence of a more conservative approach on the projected results.

Table 14 – Regional Corporate Plans SM 99% (year/year)

Scenario /Year	2020	2021	2022	2023	2024
Baseline Scenario	73,87%	58,11%	52,17%	68,85%	58,47%
Pessimistic Scenario 1	79,04%	61,89%	56,19%	73,99%	62,21%
Pessimistic Scenario 2	80,27%	62,87%	57,26%	75,42%	63,40%
Optimistic Scenario 1	68,27%	53,89%	48,04%	63,22%	52,96%
Optimistic Scenario 2	67,05%	52,91%	46,97%	61,79%	51,77%

Source: elaborated by the authors, 2025

Table 14 demonstrates that, under a 99% safety margin, projections for the Regional Corporate Plans undergo a sharp reduction compared to the values obtained with the 80% margin. The Baseline scenario, for example, drops from 73.87% in 2020 to 58.47% in 2024, suggesting that the increased caution adopted tends to moderate performance expectations, keeping values within a narrower range.

Table 15 – Nacional Corporate Plans SM 99% (year/year)

Scenario /Year	2020	2021	2022	2023	2024
Baseline Scenario	87,47%	64,73%	52,34%	48,01%	52,29%

Pessimistic Scenario 1	92,60%	67,74%	55,71%	51,06%	55,00%
Pessimistic Scenario 2	94,19%	69,02%	56,69%	52,01%	55,89%
Optimistic Scenario 1	79,54%	57,79%	47,66%	43,45%	47,18%
Optimistic Scenario 2	77,95%	56,51%	46,68%	42,50%	46,29%

Source: elaborated by the authors, 2025

Table 15 shows that projections for the Nacional Corporate Plans also adjust to lower values when the safety margin increases to 99%. The Baseline scenario presents a drop from 87.47% in 2020 to 52.29% in 2024, maintaining the previously observed downward trend. This moderation in the numbers demonstrates the robustness of the conservative approach, which seeks to reduce uncertainty in the projections.

Subsequently, simulations based on data from more than one year were used to compose the forecasts, covering the period from 2019 to 2024. For each year, all available historical data prior to the year in question were employed. Initially, this approach applies with a safety margin of 80%.

Table 16 – Family Regional Plan SM 80% (years/year)

Scenario /Year	2021	2022	2023	2024
Baseline Scenario	100,28%	96,76%	110,15%	126,81%
Pessimistic Scenario 1	108,11%	104,60%	119,88%	130,30%
Pessimistic Scenario 2	110,22%	106,78%	122,73%	133,03%
Optimistic Scenario 1	92,30%	88,80%	100,09%	110,17%
Optimistic Scenario 2	90,19%	86,62%	97,24%	107,44%

Source: elaborated by the authors, 2025

Table 16 demonstrates that by using a more robust set of historical data and maintaining an 80% margin, the Regional Family Plan presents projections that adjust to slightly higher values than those obtained in simulations based on only one previous year. The Baseline scenario, for example, varies from 100.28% in 2021 to 126.81% in 2024, indicating that the incorporation of a broader historical perspective contributes to projections with greater amplitude.

Table 17 – Regional Corporate Plans SM 80% (years/year)

Scenario /Year	2021	2022	2023	2024
Baseline Scenario	93,96%	80,67%	92,70%	90,07%
Pessimistic Scenario 1	100,07%	86,88%	99,62%	95,83%
Pessimistic Scenario 2	101,66%	88,53%	101,54%	97,66%
Optimistic Scenario 1	87,13%	74,28%	85,11%	81,58%

Optimistic Scenario 2	85,55%	72,63%	83,19%	79,74%

Source: elaborated by the authors, 2025

As illustrated in Table 17, for the Regional Business Plan, the use of multiple years of data, combined with an 80% margin, generates projections that, while showing some oscillation, remain relatively stable. The Baseline scenario, for example, oscillates between 93.96% in 2021 and 90.07% in 2024, demonstrating that historical aggregation can reduce abrupt variations, favoring a more consistent forecast.

Table 18 – Nacional Corporate Plans SM 80% (years/year)

Scenario /Year	2021	2022	2023	2024
Baseline Scenario	123,63%	107,82%	117,80%	106,47%
Pessimistic Scenario 1	129,36%	114,76%	125,29%	111,99%
Pessimistic Scenario 2	131,81%	116,78%	127,61%	113,81%
Optimistic Scenario 1	110,36%	98,18%	106,61%	96,08%
Optimistic Scenario 2	107,91%	96,17%	104,29%	94,25%

Source: elaborated by the authors, 2025

Table 18 shows that projections for the Nacional Corporate Plans using historical data with an 80% margin reveal initially high values, with the Baseline scenario reaching 123.63% in 2021, followed by a downward trend to 106.47% in 2024. This historical approach allows for a better understanding of seasonal variations and contributes to an estimate that, while still high, exhibits less volatility compared to simulations based on a single year.

Subsequently, the same methodology of using historical data applies, but with a more rigorous safety margin of 99%.

Table 19 – Family Regional Plan SM 99% (years/year)

Scenario /Year	2021	2022	2023	2024
Baseline Scenario	83,40%	83,43%	96,64%	111,25%
Pessimistic Scenario 1	89,91%	90,19%	105,18%	114,31%
Pessimistic Scenario 2	91,66%	92,07%	107,68%	116,70%
Optimistic Scenario 1	76,77%	76,57%	87,81%	96,64%
Optimistic Scenario 2	75,01%	74,69%	85,32%	94,25%

Source: elaborated by the authors, 2025

As illustrated in Table 19, when adopting a 99% safety margin in the composition of historical data for the Regional Family Plan, the projections become more conservative. The

Baseline scenario, for example, projects 83.40% in 2021, reaching 111.25% in 2024, values that are lower than those obtained with the 80% margin. This reduction highlights the impact of a more cautious approach in obtaining the estimates.

Table 20 – Regional Corporate Plans SM 99% (years/year)

Scenario /Year	2021	2022	2023	2024
Baseline Scenario	80,49%	70,05%	81,90%	80,95%
Pessimistic Scenario 1	85,72%	75,45%	88,01%	86,12%
Pessimistic Scenario 2	87,08%	76,88%	89,71%	87,77%
Optimistic Scenario 1	74,63%	64,51%	75,19%	73,31%
Optimistic Scenario 2	73,28%	63,07%	73,50%	71,67%

Source: elaborated by the authors, 2025

Table 20 shows that the results of the Regional Corporate Plans with historical data and a 99% margin demonstrate moderate stability, but with projections lower than those obtained with an 80% margin. For example, the Baseline scenario remains around 80% over the years, reinforcing that the adoption of a high safety margin imposes stricter limits on projections.

Table 21 – Nacional Corporate Plans SM 99% (years/year)

Scenario /Year	2021	2022	2023	2024
Baseline Scenario	99,32%	85,79%	94,79%	89,09%
Pessimistic Scenario 1	103,92%	91,30%	100,81%	93,71%
Pessimistic Scenario 2	105,89%	92,91%	102,68%	95,23%
Optimistic Scenario 1	88,66%	78,12%	85,78%	80,39%
Optimistic Scenario 2	86,69%	76,51%	83,91%	78,87%

Source: elaborated by the authors, 2025

Finally, Table 21 presents the results for the Nacional Corporate Plans considering historical data and a 99% safety margin. It is observed that the Baseline scenario starts with 99.32% in 2021, falling to 89.09% in 2024, while the other scenarios (pessimistic and optimistic) follow a similar trend of moderation. This approach demonstrates how the combination of a robust historical base and a high safety margin results in more cautious projections, providing a conservative estimate of future performance.

Based on the data presented, the analysis of the simulations reveals crucial insights about the dynamics of health plans and the effectiveness of the projection approaches, especially in relation to the target loss ratio of 75%.

When evaluating the annual projections, the analysis notes that the 99% safety margin consistently produces more conservative estimates compared to the 80% margin. For the Regional Family Plan, the values projected with a 99% safety margin range from 58.23% to 98.32%, while with 80%, the values oscillate between 73.65% and 122.26%, indicating a trend of underestimation with the 80% margin. In the Nacional Corporate Plans, the 99% margin also demonstrates lower estimates, ranging from 42.50% to 94.19%, in contrast to 57.51% to 122.49% for the 80% margin. The Regional Corporate Plans follows the same trend, with values of 46.97% to 80.27% for 99% and 58.41% to 97.16% for 80%.

Considering the target loss ratio of 75%, the analysis observes that the projections with a 99% margin are more frequently adequate. Specifically, in the Regional Family Plan, the optimistic scenarios (C3 and C4) with a 99% safety margin present values close to or below the target in some years. In the Nacional Corporate Plans, all scenarios with a 99% safety margin show values below the target from 2021 onwards. In the Regional Business Plan, several scenarios with a 99% safety margin also align with the target.

Regarding the simulations that use data from multiple years, the results are less consistent. Only the simulation for 2022 of the Regional Family Plan, with a 99% safety margin and the fifth scenario, showed a loss ratio close to the target. In the Regional Business Plan, scenarios C3 and C4 with a 99% safety margin in all years and all scenarios in 2022 with an 80% safety margin proved adequate.

These results indicate that, to reach the target loss ratio of 75%, the projection approach based on the immediately preceding year with a 99% safety margin is more effective. The use of historical data from multiple years did not demonstrate significant improvements in the projections, and in many cases, resulted in less accurate estimates. The analysis highlights the importance of choosing the safety margin and the need to consider the particularities of each plan in strategic decision-making.

6.4. Pricing Deviation Analysis

This section provides a detailed analysis of pricing results, highlighting their impact on plan accessibility and sustainability. To achieve this, graphs with reference lines indicating the loss ratio target are used, allowing an assessment of pricing accuracy and identifying necessary adjustments.

The analysis considers different frequency percentiles (80% and 99%), which represent varying confidence levels in the results. The 80th percentile indicates an 80% probability that the projected results will be achieved or exceeded, while the 99th percentile reflects a higher level of security. The choice between these percentiles reflects the operator's risk appetite: a lower percentile (80%) offers higher profit potential but also greater risk of losses, while a higher percentile (99%) ensures greater financial stability, with lower profit margins.

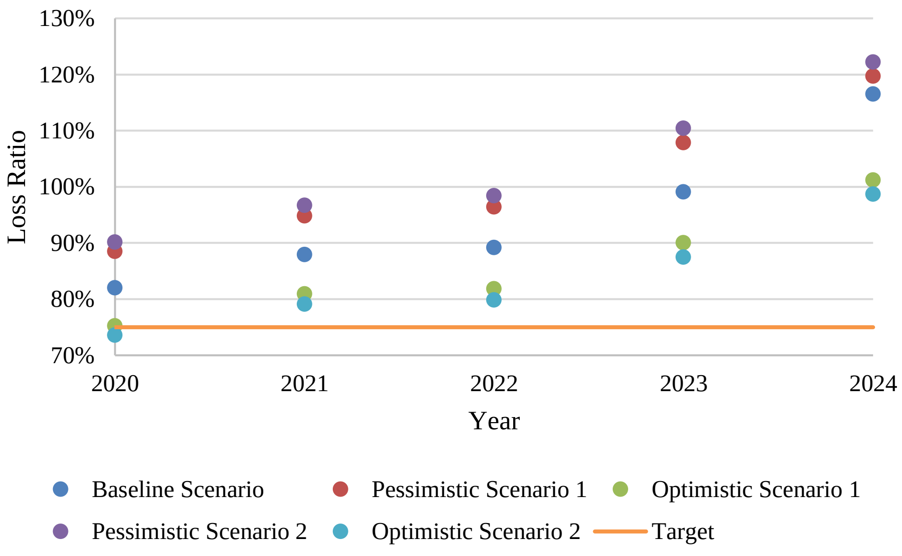
The results are presented in graphs illustrating the evolution of different scenarios and percentiles from 2020 to 2024. For comparison purposes, a 75% target was established as the desired performance threshold. Results below this threshold indicate a favorable scenario for the operator, while higher values signal the need for adjustments to avoid financial imbalances.

6.4.1. Family Regional Plans

The analysis of family regional plans assesses how different projections (optimistic and pessimistic) impact their performance, demonstrating the sensitivity of the pricing methodology to variations in assumptions.

Figure 5 presents the results for the 80% percentile. It is observed that, in all scenarios, values remain above the 75% target throughout the analyzed period (2020-2024), indicating an unfavorable performance for the operator. There is a progressive worsening trend, reinforcing the need for pricing adjustments and strategic review. Even in the most optimistic scenarios, costs exceed revenues, emphasizing the urgency of corrective measures.

Figure 9 – Family Regional Plan SM 80%

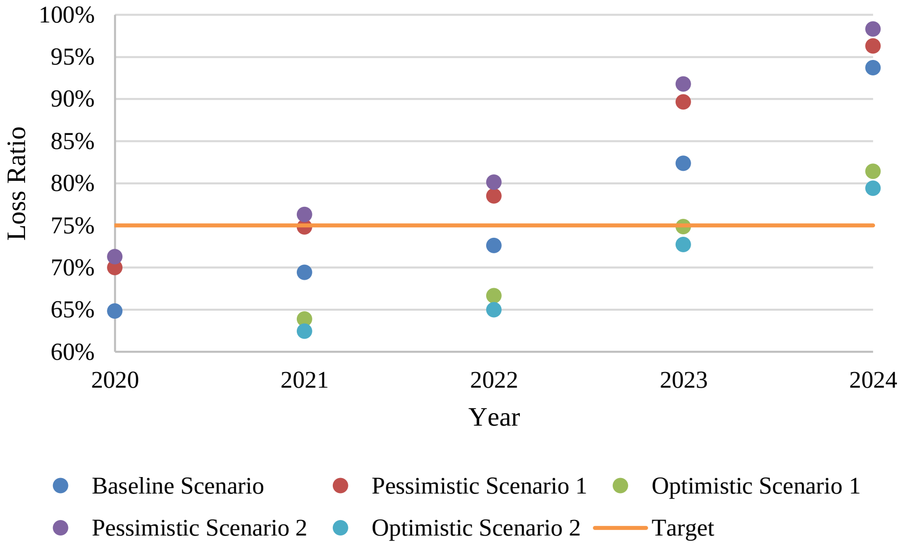


Source: elaborated by the authors, 2025

On the other hand, Figure 6, which represents the 99% percentile, reveals a different behavior. The optimistic scenarios show results below the target for most of the period, suggesting a more favorable performance for the operator. The baseline scenario also remains below the target until 2022 but surpasses it in subsequent years, signaling an increased risk. Conversely, the pessimistic scenarios demonstrate greater predictability and cost control, staying below the target for nearly the entire period.

The comparison between percentiles highlights the impact of the safety margin on result stability. The 99% percentile exhibits less volatility, whereas the 80% percentile consistently presents deficits, reinforcing the importance of adopting more conservative pricing margins to mitigate financial risks.

Figure 10 – Family Regional Plan SM 99%

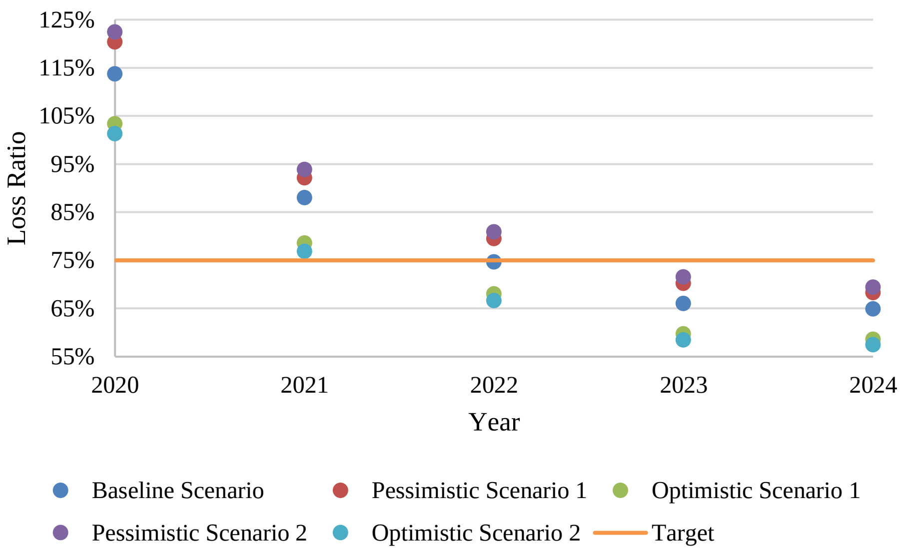


Source: elaborated by the authors, 2025

6.4.2. National Corporate Plans

Figure 7, related to the 80% percentile, reveals a downward trend in results from 2020 to 2024, contrasting with the increasing trajectory observed in family plans. From 2023 onward, all scenarios—optimistic, pessimistic, and baseline—fall below the 75% target, indicating a positive performance for the operator. This improvement is more pronounced in the pessimistic and baseline scenarios, highlighting the resilience of corporate plans even under adverse conditions.

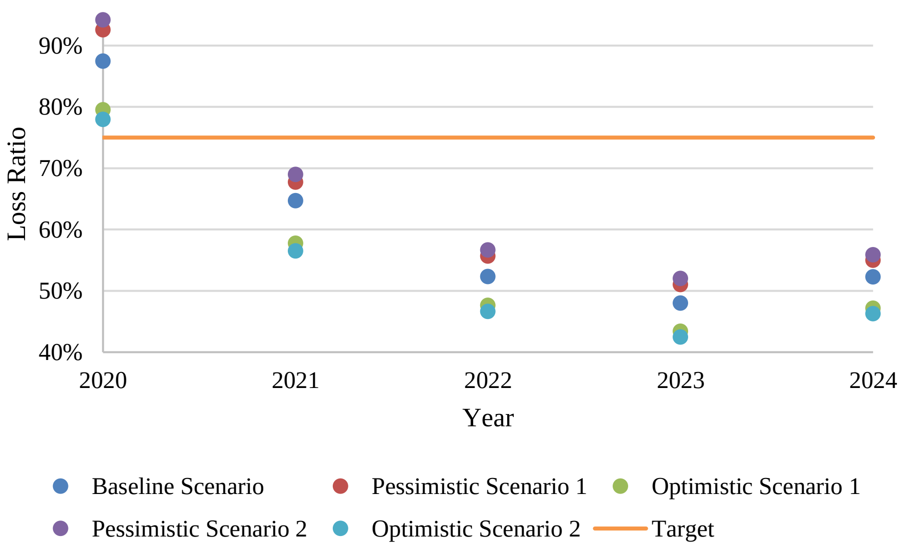
Figure 11 – National Corporate Plans SM 80%



Source: elaborated by the authors, 2025

Figure 8, representing the 99% percentile, maintains a downward trend but with less volatility. Optimistic scenarios remain close to the target, ensuring profitability throughout the period, though with a slight decreasing trend. Pessimistic and baseline scenarios remain below the target in a more stable manner, reinforcing the financial security of these plans. Since 2021, all results have stayed below the target, confirming the ability of corporate plans to withstand market fluctuations.

Figure 12 – National Corporate Plans SM 99%

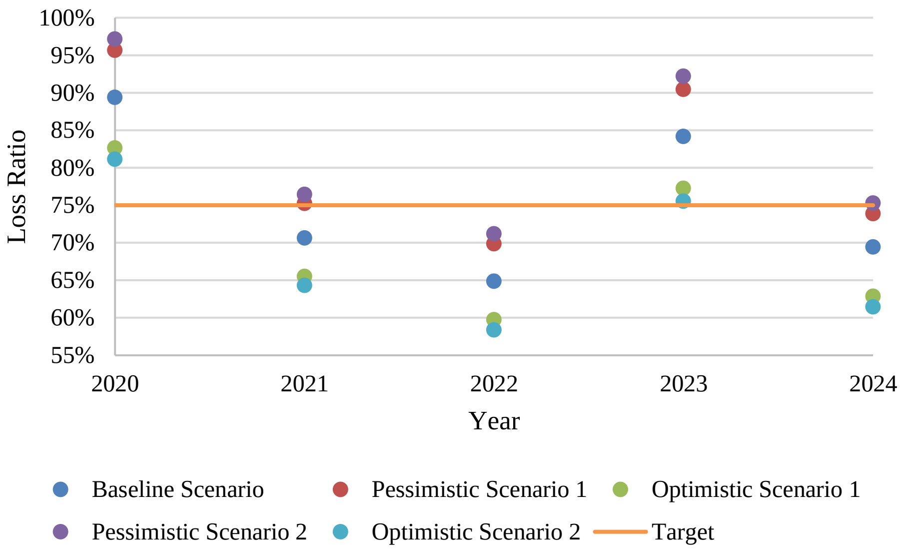


Source: elaborated by the authors, 2025

6.4.3. Regional Corporate Plans

Figure 9 presents a general downward trend in results over the years, with a more pronounced decline starting in 2020. Initially, all scenarios exceed the 75% target, indicating an unfavorable performance. However, between 2021 and 2024, most scenarios present results below the target, demonstrating a favorable evolution for the operator.

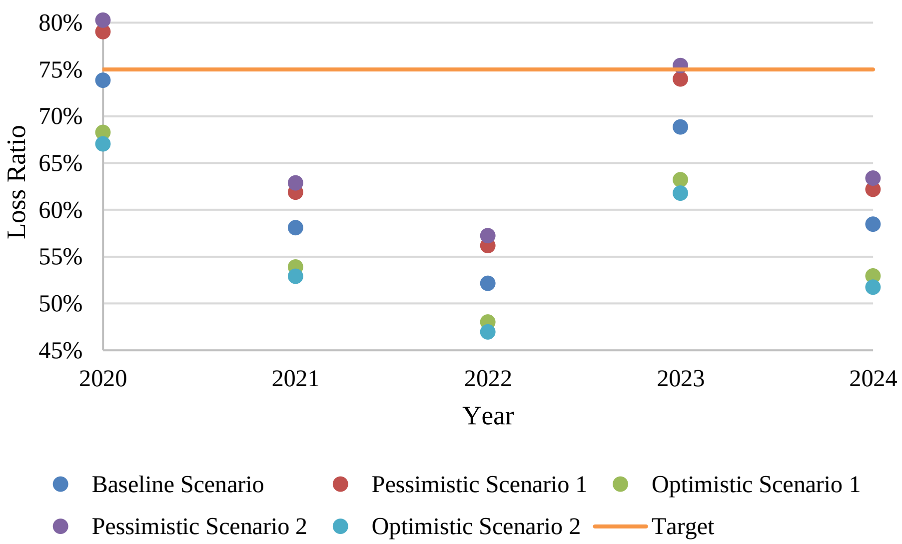
Figure 13 – Regional Corporate Plans SM 80%



Source: elaborated by the authors, 2025

Figure 10, referring to the 99% percentile, shows greater stability, with results below the target in most analyzed years. This performance reinforces that a more conservative approach to pricing protects against adverse fluctuations and ensures profitability.

Figure 14 – Regional Corporate Plans SM 99%



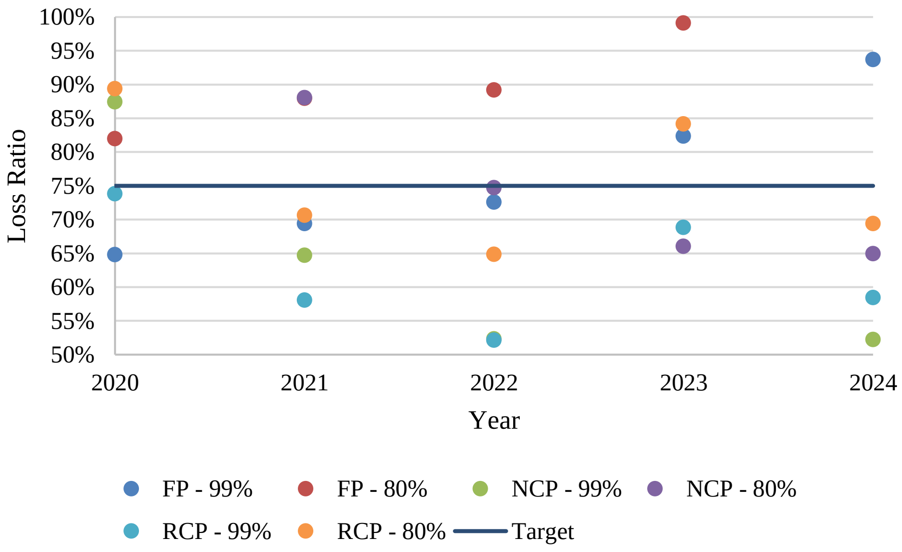
Source: elaborated by the authors, 2025

6.4.4. Baseline Scenario

The analysis of the baseline scenario, presented in Figure 11, assesses the expected performance of the health plans under standard market conditions. Results show that approximately 64% of the cases achieve a loss ratio below the 75% target, indicating a generally favorable outlook for the operator. This outcome suggests that the current pricing strategy ensures the financial sustainability of most plans, although specific segments may still require targeted adjustments. The legend for the following figures is described as follows:

- FP – 80%: Family Plans with an 80% safety margin;
- FP – 99%: Family Plans with a 99% safety margin;
- RCP – 80%: Regional Corporate Plans with an 80% safety margin;
- RCP – 99%: Regional Corporate Plans with a 99% safety margin;
- NCP – 80%: National Corporate Plans with an 80% safety margin;
- NCP – 99%: National Corporate Plans with a 99% safety margin.

Figure 15 – Base Scenario



Source: elaborated by the authors, 2025

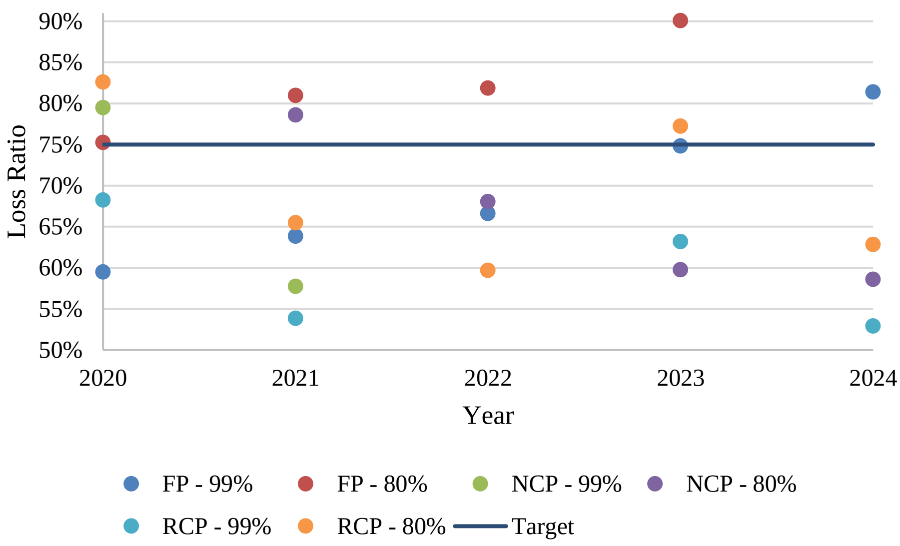
6.4.5. Optimistic and Pessimistic Scenarios

The evaluation of optimistic and pessimistic scenarios provides insights into the range of expected variations in results.

The optimistic scenarios (Figures 12 and 13) demonstrate that approximately 60% of cases present results below the target, indicating favorable conditions for the operator.

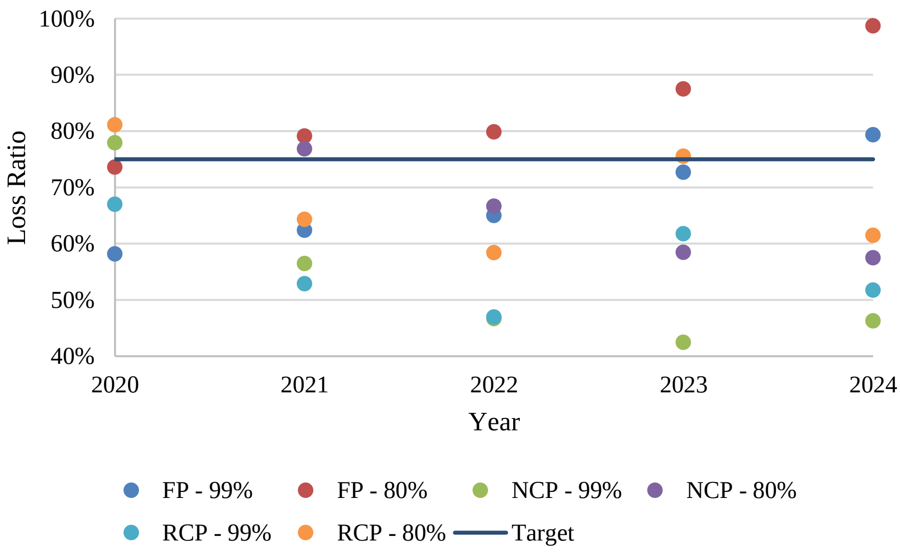
However, there is significant variation between percentiles, highlighting the need for continuous monitoring.

Figure 16 – Optimistic Scenario 1



Source: elaborated by the authors, 2025

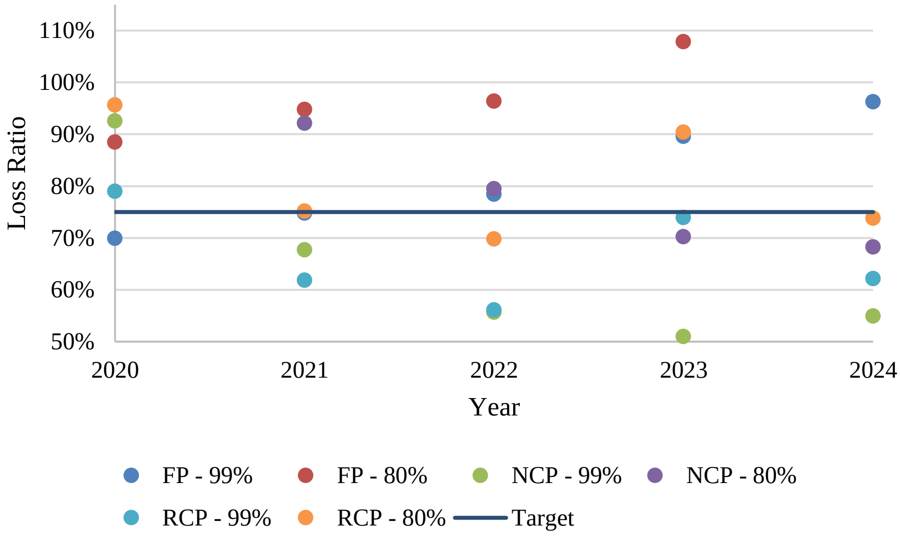
Figure 17 – Optimistic Scenario 2



Source: elaborated by the authors, 2025

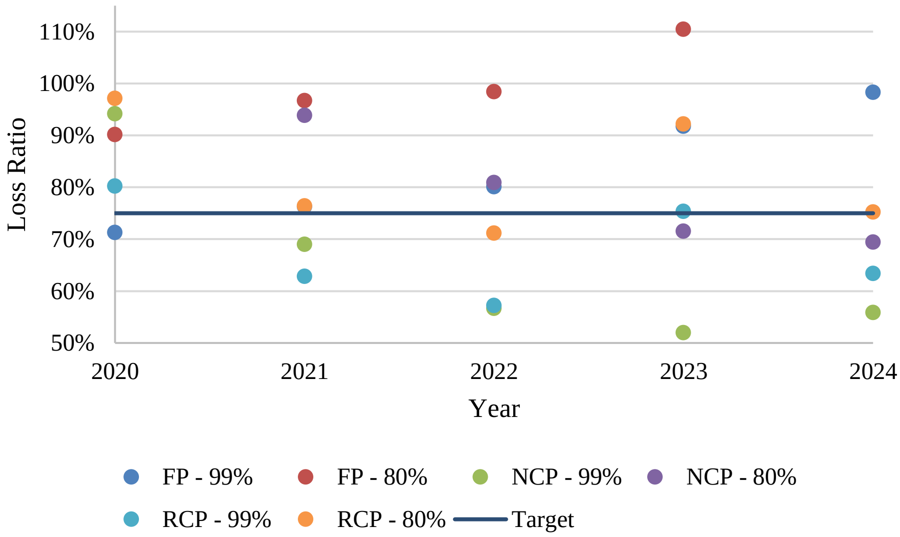
Conversely, the pessimistic scenarios (Figure 18 and Figure 19) show that, even under adverse conditions, approximately 47% of cases remain below the target, suggesting that the operator still faces challenges. The variability between percentiles indicates that financial risk is higher for family plans, while corporate plans exhibit greater stability.

Figure 18 – Pessimistic Scenario 1



Source: elaborated by the authors, 2025

Figure 19 – Pessimistic Scenario 2



Source: elaborated by the authors, 2025

The scenario analysis reveals key insights into the financial sustainability and pricing adequacy across different health plan segments.

Family Plans exhibited unfavorable results under the 80% safety margin, surpassing the 75% target and signaling the need for pricing adjustments. This vulnerability is largely due to the regulatory framework: the annual adjustment of individual and family plans is determined by the ANS and does not reflect the actual cost behavior or utilization patterns of each specific operator. Since the index is based on industry-wide averages, it often fails to restore financial balance for operators with higher-than-average claim ratios, particularly when their beneficiary profiles are older or more costly. As a result, the authorized adjustment may prove insufficient to offset real healthcare expenses, compromising the financial viability of this segment.

In contrast, National and Regional Corporate Plans showed more stable performance, generally remaining below the 75% target, with the 99% margin further enhancing resilience against market volatility.

The baseline scenario indicated that the majority of plans (64%) are projected to be profitable under normal conditions, though some still require fine-tuning in pricing. Optimistic scenarios (with 60–63% of plans below target) suggest growth opportunities but highlight the balance needed between profitability and exposure to risk. Pessimistic scenarios (with only 37–47% of plans below target) revealed greater volatility—particularly in the Family segment—emphasizing the importance of adopting robust risk mitigation strategies.

Overall, the adoption of a 99% safety margin significantly improves financial stability and provides a stronger buffer against adverse events. To ensure the long-term sustainability of health plans, a precise and adaptive pricing strategy is essential—one that balances accessibility and profitability while addressing regulatory constraints and protecting against future financial imbalances.

6.4.6. Highlight of Positive Results

In this study, a scenario is considered successful when the predicted loss ratio is equal to or less than 75%, which aligns with the cost containment target established for the health plans. Scenarios with loss ratios above this threshold are classified as unsuccessful, as they indicate a financial outcome that may compromise the operator’s sustainability.

To evaluate the model's effectiveness, the % Success Rate was calculated for each set of simulations. This metric is defined as the proportion of scenarios in which the predicted loss ratio met the target (i.e., $\leqslant$ 75%) relative to the total number of simulated scenarios. In formula terms:

Equation 7: % Success Rate

$$\%\,Success\ Rate\ = \frac{FreNumber\ of\ successful\ scenarios\ quency}{Total\ number\ of\ scenarios} * 100$$

Source: elaborated by the authors, 2025

The analysis of loss ratios under this criterion showed promising results, especially when the model used only the data from the immediately preceding year for its estimations.

This approach yielded significantly improved predictive accuracy, suggesting that more recent data better captures current cost dynamics and care utilization patterns.

The following table summarizes the absolute and relative performance across plan types, categorized by the safety margin applied (99% and 80%):

Table 22 - Loss Ratio Performance by Plan and Safety Margin (Using Previous Year Data)

PLAN	SAFETY MARGIN	SCENARIOS ≤ 75% LOSS RATIO	% SUCCESS RATE
Family Plan	99%	14	56 %
Family Plan	80%	1	4 %
National Corporate Plan	99%	20	80 %
National Corporate Plan	80%	0	0 %
Regional Corporate Plan	99%	22	88 %
Regional Corporate Plan	80%	12	48 %
Total	**-**	**69**	**46%**

Source: elaborated by the authors, 2025

As shown in Table 22, the Regional Corporate Plan exhibited the highest success rate, with 22 out of 25 scenarios achieving the target loss ratio when a 99% safety margin was applied. This suggests strong alignment between the model's predictions and actual costs for this segment. The National Corporate Plan also performed well with the 99% margin, achieving success in 80% of scenarios, although it did not reach the target in any scenario with the 80% safety margin—indicating that more conservative thresholds are essential for this group.

The use of only the immediately preceding year as a reference proved more effective than including all historical years. This may reflect the diminishing relevance of older data, which can be affected by changes in medical costs, regulation, and market behavior.

The performance decline is evident in Table 23:

Table 23 - Loss Ratio Performance by Plan and Safety Margin (Using All Previous Year Data)

PLAN	SAFETY MARGIN	SCENARIOS ≤ 75% LOSS RATIO	% SUCCESS RATE
Family Plan	99%	1	4 %
Family Plan	80%	0	0 %
National Corporate Plan	99%	0	0 %
National Corporate Plan	80%	0	0 %

Regional Corporate Plan	99%	8	32 %
Regional Corporate Plan	80%	2	8 %
Total	**-**	**11**	**7,33%**

Source: elaborated by the authors, 2025

This comparison reinforces that the inclusion of older data may reduce model performance, likely due to changes in the healthcare environment. The Regional Corporate Plan remained the most resilient across both datasets, while the National Corporate Plan showed greater volatility in its results. This is consistent with the high standard deviation observed in several cost-related variables for this group, suggesting internal heterogeneity in healthcare utilization patterns.

7. CONCLUSION

This dissertation explored the development of an innovative prototype for actuarial calculation in health plans, aiming to optimize pricing and ensure the financial sustainability of operators through a detailed case study. The construction of the prototype began with a systematic literature review, which highlighted the complexity of the supplementary health market in Brazil and the need for robust predictive models for cost management.

The systematic review identified the main methodologies and predictor variables used in health cost forecasting. Multiple linear regression emerged as a widely applied and effective statistical technique, capable of modeling the relationship between health plan costs and various independent variables, such as age group, type of care, and service utilization. The relevance of demographic data, historical costs, and plan characteristics was also highlighted, providing a solid foundation for constructing the predictive model.

Based on the evidence from the systematic review, a prototype was developed that combines the analysis of robust historical data with the application of multiple linear regression. The use of real data from a case study, covering different products and age groups, allowed the model to be validated in a practical context, demonstrating its potential to assist operators in strategic decision-making.

The results validated the effectiveness of the prototype, demonstrating its ability to model the total costs of health plans with reasonable accuracy, especially when used with recent historical data and a higher safety margin. Multiple linear regression allowed for the estimation of costs for various combinations of variables, providing a solid foundation for plan pricing.

7.1. Effects of Safety Margin on Premiums

The choice of safety margin had a significant impact on premiums, with increases ranging from 20% to 30% when moving from an 80% to a 99% safety margin. Compliance with ANS Resolutions 563 and 564 was ensured, guaranteeing that the variation in prices by age group and the registration of products met regulatory standards. Co-participation was found to be a moderating factor in the excessive use of health services, although its effect was more limited for higher age profiles.

7.2. Economic Sustainability and Challenges

The economic sustainability of health plans was identified as an ongoing challenge, influenced by the age structure and the utilization profile of beneficiaries. The decision to incorporate new beneficiary masses must be based on detailed actuarial analyses, considering different scenarios and safety margins. Despite the advances achieved, it is essential to recognize the limitations of the case study. The use of historical data may not capture atypical future events, such as pandemics or disruptive technological advances, which can significantly impact health plan costs. Moreover, the accuracy of the model depends on the quality and scope of the available data, emphasizing the importance of proper data collection and recording.

7.3. Future Research and Development

For future research, it is recommended to explore the development of hybrid models that combine regression techniques with machine learning algorithms, to enhance predictive capacity and identify complex patterns within the data. Additionally, studying the influence of medical inflation is crucial, as this factor could have a significant impact on health plan costs. Models that incorporate this variable would improve the accuracy of projections.

The creation of decision support tools—specifically software applications with intuitive interfaces designed to assist actuaries—represents a significant avenue for future development. These tools could facilitate easier operation, data visualization, and graphical analysis.

Another recommendation is the development of a continuous monitoring system for costs and health service utilization, with performance indicators and automatic alerts. This would help operators identify deviations early and take corrective measures. Integrating the prototype with other management tools, such as management information systems and data analysis platforms, could also optimize workflows and facilitate decision-making.

7.4. Contributions to the Field

In summary, this dissertation contributes to the advancement of knowledge in the field of health plan actuarial science by presenting an innovative and adaptable prototype for actuarial calculation, validated through a case study. The results obtained are expected to assist operators in improving financial management and offering fairer and more accessible health plans to the population.

Compared to previous studies, this work aligns with the general trend of incorporating data-driven and statistically grounded methodologies to predict healthcare costs and utilization patterns. Several studies in the literature—such as those by Morid (2017), Kshirsagar (2021), and Park (2018)—have explored machine learning and regression-based methods to forecast expenditures and adjust pricing mechanisms in health plans. This dissertation, however, distinguishes itself by proposing a flexible actuarial prototype that integrates multiple scenarios derived from historical service variation, rather than relying exclusively on cost predictors or machine learning outputs.

While studies like Cao (2020) and Xu (2019) emphasized individual-level predictors such as age, gender, and previous claims to estimate high-cost utilization, this dissertation takes a broader approach by simulating cost behavior under varying macro-utilization conditions. This provides a complementary perspective—more focused on aggregate product performance under different stress scenarios rather than identifying specific high-risk individuals.

Moreover, studies such as Steenhuis (2022) and Politi (2018) raised concerns about the limitations of historical data for prospective pricing, which this study also acknowledges by incorporating pandemic-related distortions into the scenario construction. The use of scenarios excluding the year 2020 addresses precisely the kind of temporal bias those authors highlighted, reinforcing the relevance of adapting risk models to contextual changes.

Another distinguishing aspect of this study is its practical focus on safety margins (e.g., 80% vs. 99%) as a tool for financial resilience, something not explicitly modeled in most of the referenced studies. While authors like Harrison (2020) and Johnson (2010) explore advanced statistical techniques (e.g., Bayesian regression, zero-inflated models), this dissertation prioritizes methodological simplicity and applicability, especially for small and medium-sized operators who may lack access to complex systems or large data sets.

Finally, this prototype offers a transparent and replicable framework that aligns with the growing demand—highlighted by studies like Kshirsagar (2021) and Park (2018)—for interpretable and operational tools in pricing and cost forecasting. By balancing robustness with accessibility, the methodology proposed here contributes to bridging the gap between actuarial theory and real-world application in regulated markets.

8. REFERENCES

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR (ANS). (2006). **Resolução Normativa nº 28, de 27 de março de 2006**. Dispõe sobre a regulamentação do registro de produtos de planos de saúde.

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR (ANS). (2020). **Padrão TISS - Troca de Informação em Saúde Suplementar**.

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR (ANS). (2022). **Resolução Normativa nº 563**, de 15 de dezembro de 2022.

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR (ANS). (2022). **Resolução Normativa nº 564**, de 15 de dezembro de 2022.

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR. **Agência Nacional de Saúde Suplementar**. Brasília, DF: ANS, 2023.

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR. **Caderno de Informação da Saúde Suplementar**. Rio de Janeiro, n. 2022, p. 67, 2022. Disponível em: <https://www.gov.br/ans/pt-br

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR. **Tabnet Informações em Saúde Suplementar**. Rio de Janeiro, 2022. Disponível em: http://www.ans.gov.br/anstabnet/index.htm.

ALTIERI, E. H. **Modelo de cálculo da necessidade de capital para cobrir os riscos de subscrição de operações não vida**. *Revista Brasileira de Risco e Seguro*, v. 9, n. 17, p. 1–46, 2013.

ALTMAN, D. G.; BLAND, J. M. **Basic statistics: why do we need confidence intervals?** *British Medical Journal*, v. 311, n. 6910, p. 1072-1073, 1995.

AMERICAN ACADEMY OF FAMILY PHYSICIANS. **Fee-for-Service vs. Value-Based Care**. 2023.

AMERICAN HOSPITAL ASSOCIATION. **Understanding Health Care Pricing**. 2023.

ARAÚJO, Â. A. S.; SILVA, J. R. S. **Análise de tendência da sinistralidade e impacto na diminuição do número de operadoras de saúde suplementar no Brasil**. *Revista Brasileira de Economia da Saúde*, v. 23, n. 8, p. 2763–2770, 2018.

BARROS, A. J. D.; VICTORA, C. G. **Introdução aos conceitos de epidemiologia**. 4. ed. Pelotas: Editora da UFPEL, 2006.

BERGMANN, M. et al. **Understanding health service utilization and health system costs: A systematic review**. *International Journal of Health Economics*, 2018.

BEUREN, I. M. **Trajetória da construção de um trabalho monográfico em contabilidade**. In: ______. *Como elaborar trabalhos monográficos em contabilidade: teoria e prática*. 2. ed. São Paulo: Atlas, 2003.

BOOTH, A.; SUTTON, A.; PAPAIOANNOU, D. **Systematic approaches to a successful literature review**. 2. ed. London: Sage, 2016.

BOTELHO, L. L. R.; CUNHA, C. C. A.; MACEDO, M. **O método da revisão integrativa nos estudos organizacionais**. *Gestão e Sociedade*, v. 11, p. 5–20, 2011.

BOWERS, N. L.; GERBER, H. U.; HICKMAN, J. C.; JONES, D. A.; NESBITT, C. J. (1997). **Actuarial mathematics**. Society of Actuaries.

BRADFIELD, R.; WRIGHT, G.; BURT, G.; CAIRNS, G.; VAN DER HEIJDEN, K. **The origins and evolution of scenario techniques in long-range business planning**. *Futures*, v. 37, n. 8, p. 795-812, 2005.

BRASIL. **Projeto de Lei nº 2033/2022**. Brasília, DF, 2022.

BROUWER, W. B. F. et al. **A framework for the assessment of health care costs and outcomes**. *Health Economics*, v. 28, n. 9, p. 1014-1025, 2019.

BRUNNER, T.; LANGER, G. **Economic evaluation of healthcare interventions: Methods, challenges, and applications**. *Health Economics Review*, v. 9, p. 19-33, 2019. DOI: 10.1186/s13561-019-0217-4.

BUENO, L. P. **Métodos estatísticos básicos para seguros em geral**. Rio de Janeiro: Fundação Escola Nacional de Seguros, 2019.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: Methods and Applications**. Cambridge: Cambridge University Press, 2005.

CAMERON, A. C.; TRIVEDI, P. K. **Regression analysis of count data**. Cambridge: Cambridge University Press, 2013.

CAO, A. Z.; CASTEL, L. D. H. **Statistical prediction of high-cost claimants using commercial health plan data**. In: *Lecture Notes in Networks and Systems*. Springer, 2020. v. 69, p. 904–912.

CAO, X.; LIU, Y.; LUO, Z.; ZHAO, D.; LI, Y.; ZHAO, Y. **Statistical prediction of high-cost claimants using commercial health plan data**. In: *Lecture Notes in Networks and Systems*, v. 106, p. 431-439, 2020.

CECIL, Dean; CHO, Sung-Hye; CHEN, Min. **Relationship of the use and costs of physician office visits and prescription drugs to travel distance and increases in member cost share**. *Journal of Managed Care Pharmacy*, v. 12, p. 256-263, 2006.

CELLA, J. L. D. J. **Efeitos da crise econômica de 2014-2016 sobre o mercado brasileiro de saúde suplementar**. 2019. 264 f. Dissertação (Mestrado em Economia) – Universidade Presbiteriana Mackenzie, São Paulo, 2019.

CENTERS FOR MEDICARE & MEDICAID SERVICES. **Medicare Diagnosis Related Groups (MS-DRGs)**. 2023.

CERRI, G. G. **The challenges of health care in Brazil**. *Revista de Medicina*, v. 95, n. spe1, p. 60, 29 ago. 2016.

CIDAV, Z.; MARSHALL, J.; SPERLING, R.; MANDELL, D. S. **The cost of intervention for autism spectrum disorder: A comparison of ABA and eclectic treatments**. *Journal of Autism and Developmental Disorders*, v. 47, n. 8, p. 2415–2424, 2017. DOI: 10.1007/s10803-017-3176-5.

CONASEMS. **Conselho Nacional de Secretarias Municipais de Saúde**. Brasília, DF: CONASEMS, 2023.

CRESWELL, J. W. **Qualitative inquiry and research design: choosing among five approaches**. 3. ed. Porto Alegre: Penso, 2013. Cap. 4, p. 77-110.

CUTLER, D. M.; MCCLELLAN, M.; MERRILL, J. **Excess spending on health care**. *Health Affairs*, v. 32, n. 1, p. 7-17, 2013.

DOWLING, M. **Health insurance pricing**. New York: John Wiley & Sons, 2008. 344 p.

ERCOLE, F. F.; MELO, L. S.; ALCOFORADO, C. L. G. C. **Revisão integrativa versus revisão sistemática**. *Revista Mineira de Enfermagem*, v. 18, p. 9–12, 2014.

FARIAS, N. P. et al. **Análise da variabilidade dos custos hospitalares no estado do Rio Grande do Sul**. *Revista de Saúde Pública*, v. 44, n. 6, p. 1091-1100, 2010.

FIELD, A. (2018). **Discovering statistics using IBM SPSS statistics**. Sage publications.

FINK, A. **Conducting research literature reviews: from the internet to paper**. 4. ed. Thousand Oaks: Sage, 2014. Cap. 3, p. 45-80.

FRANÇA, G. V.; GADELHA, C. A.; TRAVASSOS, C. **O financiamento da saúde no Brasil: desafios e perspectivas**. *Ciência & Saúde Coletiva*, v. 24, n. 10, p. 3443-3454, 2019.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GLIED, S. A.; LLERAS-MUNEY, A. **Health insurance and medical care expenditures: evidence from the 1965 Medicare expansion**. *The Quarterly Journal of Economics*, v. 123, n. 1, p. 259-302, 2008.

GODET, M.; ROUBELAT, F. **Creating the future: The use and misuse of scenarios**. *Long Range Planning*, v. 29, n. 2, p. 164-171, 1996.

GREENE, S. F. **Health care economics**. Boston: Pearson, 2017. p. 212.

GUJARATI, D. N.; PORTER, D. C. (2009). **Basic econometrics**. McGraw-Hill Education.

HARRISON, Alexander et al. **Economic outcomes of insurer-led care management for high-cost patients**. *American Journal of Managed Care*, v. 26, p. 115-122, 2020.

INSTITUTO BRASILEIRO DE ATUÁRIA. **Instituto Brasileiro de Atuária**. Rio de Janeiro, RJ: IBA, 2023.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2023). **Índice Nacional de Preços ao Consumidor Amplo (IPCA)**.

JOHNSON, Neil; LEE, Daniel; COX, David. **Gibbs sampling for a Bayesian hierarchical general linear model**. *Electronic Journal of Statistics*, v. 4, p. 304-320, 2010.

KARAMI MATIN, B.; YOUSEFI, M.; REZAPOUR, A.; KHOJASTEHMEHR, R.; EBADI FARD AZAR, F.; RASHIDIAN, A. **Economic burden of autism spectrum disorders across the lifespan: A systematic review**. *Archives of Public Health*, v. 80, n. 1, 2022. DOI: 10.1186/s13690-022-00799-8.

KSHIRSAGAR, Rohun et al. **Accurate and interpretable machine learning for transparent pricing of health insurance plans**. In: *35th AAAI Conference on Artificial Intelligence, AAAI 2021*. p. 15127-15136, 2021.

LANG, Gregor; WONG, Chi-Keung. **Willingness to pay for lung cancer treatment**. *Value in Health*, v. 13, p. 542-550, 2010.

MAGALHÃES, J. C. **Precificação de planos de saúde**. 2. ed. Rio de Janeiro: Elsevier, 2023.

MAHAR, M. T. et al. **The impact of health plan design on medical care utilization: A systematic review**. *Journal of Health Economics*, v. 47, p. 117-129, 2017.

MARK, G.; SMITH, E.; BROOKS, E. **Risk adjustment for people with chronic conditions in private sector health plans**. *Medical Decision Making*, v. 23, p. 29-42, 2003.

MARTEI, Giovanni et al. **Methodology to forecast volume and cost of cancer drugs in low- and middle-income countries**. *Journal of Global Oncology*, v. 4, p. 64-75, 2018.

MCCLELLAN, M.; HACKBARTH, G. **Cost and quality of health care in the United States**. *The New England Journal of Medicine*, v. 330, n. 21, p. 1500-1506, 1994.

MINISTÉRIO DA SAÚDE. **Ministério da Saúde**. Brasília, DF: Ministério da Saúde, 2023.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied statistics and probability for engineers**. New York: Wiley, 2021.

MORID, Ali; ZHANG, Wenjia; MIKHAIL, A.; XU, Xing. **Supervised learning methods for predicting healthcare costs: systematic literature review and empirical evaluation**. In: *AMIA Annual Symposium Proceedings*, 2017. p. 741-750.

NATIONAL ASSOCIATION OF HEALTH UNDERWRITERS. **Usual, Customary and Reasonable (UCR) Fees**. 2023.

NEUMANN, P. J.; OLSEN, J. A.; MCNAMEE, H. **Cost analysis in health care evaluation**. *Medical Decision Making*, v. 20, n. 3 Suppl, p. 6A-33A, 2000.

NOVAES, H. M. D. (2000). Avaliação econômica de tecnologia em saúde: desafios para o sistema de saúde brasileiro. **Ciência & Saúde Coletiva**, 5(2), 305-319.

O'BRIEN, B. J.; NORMAND, G. T.; THIEMANN, E.; ANDREWS, J. **Conceptual and methodological issues in the measurement of healthcare costs**. *Medical Care Research and Review*, v. 64, n. 1, p. 61-100, 2007.

PACALA, James; WILLIAMS, Dan. **Using self-reported data to predict expenditures for the health care of older people**. *Journal of the American Geriatrics Society*, v. 51, p. 1236-1241, 2003.

PAIM, J. S. **Thirty years of the unified health system (UHS)**. *Ciência & Saúde Coletiva*, v. 23, n. 6, p. 1723–1728, 2018.

PARK, Seungwoo et al. **Alternative evaluation metrics for risk adjustment methods**. *Health Economics (United Kingdom)*, v. 27, p. 185-196, 2018.

PATTON, M. Q. **Qualitative research & evaluation methods**. 4. ed. Thousand Oaks: Sage, 2015. v. 1, p. 120-135.

PELL, J. P. et al. **Statistical analysis of healthcare utilization data: Methods and issues**. *BMC Health Services Research*, v. 18, p. 645, 2018.

POLITI, Mary C.; WONG, Estelle K.; WILSON, Sheri L.; KAMAT, Preeti; SINGH, Rahul P. **A comparison between subjective and objective methods of predicting health care expenses to support consumer's health insurance plan choice**. *MDM Policy and Practice*, v. 3, p. 1-10, 2018.

PREVIC. **Instrução No 03, de 10 de outubro de 2012. Dispõe**. v. 10, n. 9, p. 32, 2012.

RANDS, A. **Gestão de planos de saúde**. 5. ed. São Paulo: Atlas, 2023.

ROBINSON, J. C.; EPSTEIN, A. M.; SHAH, M. H. **The unintended consequences of DRG-based reimbursement**. *New England Journal of Medicine*, v. 371, n. 17, p. 1610-1612, 2014.

ROBST, John. **Estimation of a hedonic pricing model for Medigap insurance**. *Health Services Research*, v. 41, p. 627-644, 2006.

SÁ, M.; MACIEL JÚNIOR, J.; REINALDO, L. **Processo de Ruína Finito: um estudo de caso na saúde suplementar no Brasil**. *Revista Evidenciação Contábil & Finanças*, v. 5, n. 2, p. 88–103, 2017.

SANHUEZA, S. et al. **Utilización de estadísticas descriptivas en la investigación en salud: una revisión metodológica**. *Revista Médica de Chile*, v. 145, n. 3, p. 381-388, 2017.

SCHOEMAKER, P. J. H. **Scenario planning: A tool for strategic thinking**. *MIT Sloan Management Review*, v. 36, n. 2, p. 25-40, 1995.

STAKE, R. E. **The art of case study research**. Thousand Oaks: Sage, 1995. Cap. 5, p. 101-135.

STEENHUIS, Sander et al. **The potential risk of using historic claims to set bundled payment prices: the case of physical therapy after lower extremity joint replacement**. *BMC Health Services Research*, v. 22, p. 1061, 2022.

TRIOLA, M. F. (2018). **Elementary statistics**. Pearson.

VAN DER HEIJDEN, K. **Scenarios: The art of strategic conversation**. New York: John Wiley & Sons, 2005.

WEBSTER, J.; WATSON, T. R. **Analyzing the past to prepare for the future: Writing a literature review**. *MIS Quarterly*, v. 26, n. 2, p. xiii-xxiii, 2002.

XU, Ming; LUO, Liyang; ZHANG, Yuhong. **Cost-sharing payments for out-of-network care in commercially insured adults**. *American Journal of Managed Care*, v. 25, p. 631-638, 2019.

YIN, R. K. **Case study research and applications: design and methods**. 6. ed. Thousand Oaks: Sage, 2018. Cap. 2, p. 33-72.

ANEXOS

Anexo IIA – RN 564

ANEXO II - A

Dados Estatísticos da Nota Técnica de Registro de Produto

CNPJ da Operadora:

Período de Análise: ____/_______ a ____/_______

Abrangência do preço do produto: ______________________

Região: _______________

Tabela de Itens de Despesa

N.º de Linha	Dados do Plano				Item de Despesa Assistencial:_______________					
			Faixas Etárias							
	Nome do Plano	N.º de Registro do Plano	Idade Inicial	Idade Final	N.º de Expostos	N.º de Eventos	Freqüência de Utilização	Total de Despesa Assistencial	Valor Médio do Item de Despesa Assistencial	Despesa Assistencial por Exposto
	A	B	C	D	E	F	G = F / E	H	I = H / F	J = G x I = H / E
1										
2										
3										
4										
5										
6										
7										

Anexo IIB – RN 564

ANEXO II-B

Dados Estatísticos da Nota Técnica de Registro de Produto - Total

CNPJ da Operadora:

Período de Análise:____/_______a____/______________

Abrangência do preço do produto: ___________________

Região: _____________

Tabela de Totais

N.º de Linha	Dados do Plano				Total dos Itens de Despesa						
			Faixas Etárias								
	Nome do Plano	Nº de Registro do Plano	Idade Inicial	Idade Final	Despesa Assistencial por Exposto	Recuperação de Co-participação	Recuperação de Seguro	Recuperação de Resseguro e Co-seguro	Despesa Assistencial Líquida por Exposto	Margem de Segurança Estatística por Exposto	Despesa Assistêncial Líquida por Exposto com Margem de Segurança Estatística por Exposto
	A	B	C	D	E	F	G	H	I = E - F - G - H	J	K = I + J
1											
2											
3											
4											
5											
6											
7											

N.º de Linha	Dados do Plano				N.º de Beneficiários	Total dos Itens de Despesa					Lucro		Mensalidade
			Faixas Etárias			Despesas não Assistenciais por Beneficiário							
						Demais Despesas da Carteira de Planos							
	Nome do Plano	Nº de Registro do Plano	Idade Inicial	Idade Final		Despesas de Comercialização por Beneficiário	Outras Despesas da Carteira de Planos por Beneficiário	Despesas Adminis-trativas por Beneficiário	Prestação de outros Serviços por Beneficiário	Despesa Total por Beneficiário	Valor da Margem de Lucro por Beneficiário	Ajuste	Valor Comercial da Mensalidade
	A	B	C	D	L	M	N	O	P	Q = K + M + N + O + P	R	S	T = Q + R + S
1													
2													
3													
4													
5													
6													
7													