*Journal of Mathematics and Modeling in Finance (JMMF)*
Vol. 7, No. 1, Winter & Spring 2027
Research paper

# Optimal Coverage Limit Design for Health Insurance under Stochastic Claims and Policyholder Coverage Selection

**Ramazan Rezazadeh**[1], **Amir.T Payandeh Najafabadi**[2], **Kaled Masoumifard**[3]

[1] Department of Actuarial Science, Faculty of Mathematical Sciences, Shahid Beheshti University, Evin, Tehran, Iran.

r_rezazadeh@sbu.ac.ir

[2] Department of Actuarial Science, Faculty of Mathematical Sciences, Shahid Beheshti University, Evin, Tehran, Iran.

amirtpayandeh@sbu.ac.ir

[3] Department of Statistics & Actuarial Science, Stellenbosch University, Stellenbosch, South Africa.
29827094@sun.ac.za

**Abstract:**
Determining optimal coverage limits is a fundamental problem in health insurance, as insurers must balance profitability, policyholders' preferences, and risk exposure under uncertain claim experience. Existing studies generally optimize coverage limits by considering either claim uncertainty or policyholder behavior, while their combined effects have received limited attention. To address this gap, this paper develops a mathematical framework for the optimal design of multiple health insurance coverage limits by jointly incorporating stochastic claim frequencies and policyholders' coverage selection probabilities. The insurer's decision problem is formulated as the maximization of an exponential utility function subject to a probabilistic risk constraint defined by a loss threshold $t$ and a confidence level $\theta$. Using probabilistic bound theory, the stochastic optimization problem is transformed into an equivalent convex optimization problem that can be efficiently solved using the Lagrange multiplier method. The resulting optimality conditions lead to a system of nonlinear equations, for which the existence and uniqueness of the solution are established under appropriate regularity assumptions. The proposed framework is validated using health insurance data from Asmari Insurance Company for the period 2022–2023. The numerical results demonstrate that the proposed approach provides stable and computationally efficient optimal coverage limits while effectively controlling portfolio risk, illustrating its potential as a practical decision-support tool for health insurance contract design.

*Keywords:* Coverage Limit Optimization, Health Insurance, Exponential Utility, Probabilistic Constraints, Portfolio Risk Management.
*MSC2020 Classification:* 91B30; 91B16; 91G70.

# 1 Introduction

The design of optimal coverage limits in health insurance contracts represents one of the most fundamental and persistent challenges in actuarial science and health economics. Insurance companies operate in a dual and inherently conflicting envi-

[2]Corresponding author
Received: 11/05/2026 Accepted: 12/09/2026
**https://doi.org/10.22054/jmmf.2026.92687.1287**

ronment: on one hand, they must offer attractive coverage packages with sufficient limits to maintain competitive positioning and satisfy policyholder expectations; on the other hand, excessively high coverage limits expose the insurer to potentially catastrophic financial losses and solvency risks. This tension between market competitiveness and prudent risk management lies at the heart of the optimal insurance contract design problem, which has attracted sustained scholarly attention since the seminal contributions of Arrow [2] on uncertainty and the welfare economics of medical care. Arrow demonstrated that optimal insurance contracts generally exhibit a deductible and limit structure under expected utility maximization, establishing a theoretical foundation that continues to inform contemporary research [5, 19]. However, the classical theory of optimal insurance, rooted in expected utility maximization, has been substantially extended along several dimensions. Raviv [21] incorporated administrative costs and premium loadings into the optimal design problem, demonstrating that such frictions drive optimal contracts away from pure deductible forms. Gollier and Schlesinger [10] employed stochastic dominance techniques to generalize Arrow's optimality conditions, while Promislow and Young [19] developed a unified framework encompassing multiple risk measures. More recently, Bergesio, Koch-Medina, and Munari [4] examined optimal insurance design under limited liability, showing that Mossin's theorem—which establishes full insurance optimality under fair pricing—breaks down when policyholders can default on claims exceeding their wealth, leading to optimal contracts characterized by capped deductibles. Han and Li [12] investigated optimal insurance menus under asymmetric information within a Stackelberg framework, revealing that equilibrium contracts exhibit nonlinear pricing with decreasing risk loadings to induce self-selection among heterogeneous risk types. Golubin and Gridin [11] studied optimal insurance strategies under stage-by-stage probabilistic Value-at-Risk constraints, demonstrating that the optimal retained loss functions are either stop-loss insurances or conditional deductible insurances depending on the prescribed safety level.

Despite this rich theoretical heritage, the existing literature exhibits several notable gaps when applied specifically to health insurance. First, much of the classical optimal insurance literature treats health insurance analogously to property and casualty insurance, neglecting its fundamental distinguishing features. Zhang and Wu [22] addressed this gap by introducing utility functions that simultaneously incorporate both health status and wealth, arguing that in health insurance, a loss represents not merely a financial reduction but also a deterioration in the policyholder's health state that may not be fully compensable through monetary indemnity. Zhang, Wu, and Liang [23] extended this framework to incorporate income and wealth constraints, demonstrating that financial limitations significantly affect the optimal contract structure. Cremer and colleagues [9] examined nonlinear reimbursement rules for preventive and curative medical care, showing that when illness severity is unobservable, optimal insurance designs benefits that increase with both preventive and curative care, as higher expenditures reduce informational rents and align incentives. Second, the literature has largely overlooked the role of policyholder coverage selection behavior in determining the insurer's aggregate liability structure. Traditional models implicitly assume that all policyholders purchase a complete set of available coverages—an assumption inconsistent with empirical evidence, as policyholders, constrained by income and personal preferences, typically select only a subset of available coverages [8, 18]. Recent survey evidence from Capco [6] indicates that nine out of ten policyholders are willing to share additional personal data to secure more tailored products and services, highlighting the growing importance of personalized offerings in insurance markets. This behavioral dimension introduces a critical source of uncertainty that directly affects the insurer's risk exposure and, consequently, the optimal coverage limits. Third, while the actuarial literature has extensively studied claims frequency modeling using Poisson processes and their extensions [3], the simultaneous modeling of claims frequency and coverage selection behavior within a unified optimization framework remains underdeveloped. A recent study on moral hazard management in health insurance [24] proposed a quantitative mathematical model based on quasi-arbitrage conditions to optimize deductible levels and incentivize high-deductible plan selection, demonstrating how setting appropriate deductible levels and offering targeted premium reductions can encourage insured individuals to adopt high-deductible health plans while maintaining insurer profitability. Fourth, the formulation of risk control constraints in optimal insurance design has largely relied on expected value or variance-based measures, with less emphasis on probabilistic (chance) constraints that directly address downside risk or catastrophic loss scenarios. Although chance-constrained programming has been applied in portfolio selection for casualty insurance firms [10], its integration with utility-based coverage limit optimization in health insurance remains unexplored.

From a practical perspective, the Iranian health insurance ecosystem presents a particularly relevant context for investigating these challenges. Iran's health system is characterized by a fragmented insurance landscape comprising multiple funds—including the Iran Health Insurance Organization, Social Security Organization, and Armed Forces Medical Services Insurance Organization—with substantial variation in coverage structures, financing mechanisms, and policyholder participation patterns. Rahimisadegh and colleagues [20] identified the major challenge in the Iranian health insurance ecosystem as the complexity and conflict of interests between multiple actors with different roles, which leads to fragmentation, diverse structures, and a gap between other functions and objectives, hindering the effective functioning of the ecosystem. Momahhed and colleagues [18] estimated optimal co-insurance rates for outpatient drug expenses using a data mining approach applied to over 21 million outpatient claims from the Iranian Health Insurance Organization, demonstrating that variable co-insurance structures tailored to patient risk profiles can significantly improve resource allocation and reduce insurer expenses while enhancing access to costly medications for vulnerable populations. These contextual factors make Iran an ideal setting for developing and validating a coverage limit optimization model that addresses the dual challenges of stochastic claims frequency and policyholder behavior.

The present study addresses the aforementioned gaps by developing a novel mathematical framework for determining optimal coverage limits in multiple-coverage health insurance contracts. Our contribution is threefold. First, we redefine the stochastic process governing claim occurrences by explicitly incorporating policyholders' coverage selection behavior. Unlike traditional models that implicitly assume universal participation in all coverages, we model the claim reporting process for each coverage $i$ as a Poisson process with rate $\lambda_i$, and allow each policyholder to select coverage $i$ with probability $p_i$. Consequently, the effective claim rate for coverage $i$ is $p_i\lambda_i$, capturing the interaction between claims frequency and participation behavior. This formulation establishes a bridge between microeconomic models of consumer choice and actuarial risk models, enabling sensitivity analysis of optimal limits with respect to changes in coverage selection patterns—a feature particularly relevant for insurers operating in dynamic markets with fluctuating demand [17]. Second, we introduce a probabilistic risk control constraint of the form $\mathbb{P}(S - \mathbb{E}[S] > t) \leq \theta$, which directly addresses the concept of downside risk or catastrophic loss. Here, $t$ denotes the acceptable risk threshold, interpretable as initial capital or solvency capital requirements, while $\theta$ represents the maximum permissible probability of exceeding this threshold. Using Hoeffding's inequality and Jensen's inequality, we transform this probabilistic constraint into a deterministic convex constraint: $\sum_{i=1}^{n} p_i\lambda_i d_i^2 \leq 2t^2/(-\ln\theta)$. The simultaneous presence of $\lambda_i$ and $p_i$ in this constraint reflects the significant interaction between claim rates and selection probabilities in determining the company's risk limit—an interaction that has been largely overlooked in previous formulations [1]. Third, we formulate a dual-criterion optimization framework combining the insurer's exponential utility function with risk aversion coefficient $\beta$ and the probabilistic constraint, enabling simultaneous modeling of micro-level risk-averse behavior and macro-level regulatory solvency requirements. The transformation of the probabilistic constraint into a deterministic convex form renders the original stochastic problem analytically tractable, solvable via the Lagrange multiplier method, with the existence and uniqueness of the solution rigorously established under mild regularity conditions. This theoretical guarantee, often absent in numerical approaches, provides a solid foundation for practical implementation.

The proposed framework is empirically validated using a comprehensive real-world claims dataset from the Asmari Insurance Company for the period 2022–2023. The dataset comprises multiple coverage types with heterogeneous claims frequencies and policyholder participation patterns, providing a robust foundation for assessing the practical utility of our approach. We conduct sensitivity analyses to examine how optimal limits respond to variations in risk aversion, risk tolerance parameters, and participation probabilities, thereby addressing the reviewer's concern regarding the robustness of our findings. Comparative analysis with observed coverage limits in the dataset enables us to evaluate the potential efficiency gains achievable through our optimization framework.

The remainder of this paper is organized as follows. Section 2 establishes the model assumptions and formalizes the mathematical optimization problem, incorporating stochastic claims frequency, policyholder coverage selection probabilities, and the probabilistic risk constraint. Section 3 presents the step-by-step optimization solution, including the transformation of the probabilistic constraint, objective function reformulation, derivation of optimality conditions, and theoretical proofs for the existence and uniqueness of the solution. Section 4 describes the empirical application using health insurance data from Asmari Insurance Company, presents parameter estimation and optimal coverage limits, and performs sensitivity analyses. Section 5 concludes with a summary of theoretical and practical implications, addresses model limitations, and outlines directions for future research.

## 2 Model Assumptions and Problem Formulation

This section develops the mathematical framework for determining optimal coverage limits in multiple-coverage health insurance contracts. We begin by establishing the fundamental assumptions governing the stochastic structure of claims and policyholder behavior, then formalize the insurer's optimization problem as a constrained expected utility maximization subject to a probabilistic risk constraint. Throughout this section, we maintain a balance between mathematical rigor and practical relevance, ensuring that all assumptions are explicitly stated and theoretically justified.

Consider an insurance company offering a set of $n$ distinct health coverages to its policyholders, indexed by $i = 1, \ldots, n$. Each coverage corresponds to a specific category of medical expenditure, such as hospitalization, surgical procedures, paraclinical diagnostics, outpatient consultations, or dental care. The insurer's primary decision variable is the vector of coverage limits $\mathbf{d} = (d_1, \ldots, d_n) \in \mathbb{R}^n_{++}$, where $d_i$ denotes the maximum indemnifiable amount per claim under coverage $i$. The objective is to determine the optimal limit for each coverage that simultaneously ensures policyholder satisfaction and preserves the insurer's financial stability, a problem that has received renewed attention in the wake of recent developments in optimal insurance design under limited liability [4].

To model the uncertainty inherent in the claim reporting process, we assume that for each coverage $i$, the number of reported claims over a one-year period follows a Poisson process with intensity parameter $\lambda_i > 0$. Formally, let $M_i(t)$ denote the number of claims reported under coverage $i$ up to time $t$, where $t$ is measured in years. Then $\{M_i(t) : t \geq 0\}$ is a Poisson process with rate $\lambda_i$. This assumption implies that the number of claims in non-overlapping time intervals are independent, and that the inter-arrival times between successive claims are exponentially distributed with mean $1/\lambda_i$. The Poisson process remains a cornerstone of actuarial claims modeling due to its analytical tractability and empirical validity in a wide range of insurance contexts [3, 19].

A critical innovation of our framework is the explicit incorporation of policyholders' coverage selection behavior, which addresses a significant gap identified in the existing literature [8, 22]. In contrast to traditional models that implicitly assume universal participation across all coverages, we recognize that policyholders, constrained by income limitations, heterogeneous risk perceptions, and prior healthcare experiences, typically select only a subset of the available coverages. To capture this selective behavior, we introduce a participation probability $p_i \in (0, 1)$ for each coverage $i$, representing the probability that a randomly selected policyholder chooses to purchase coverage $i$. The vector $\mathbf{p} = (p_1, \ldots, p_n)$ reflects the insurer's observed or estimated participation patterns, which may vary over time due to demographic shifts, economic conditions, or competitive dynamics.

On the relationship between the claim process and the selection probabilities, we proceed as follows. Let $N_i$ denote the number of policyholders who have selected coverage $i$ and are thus exposed to the possibility of reporting a claim under that coverage over the one-year horizon. Conditional on the selection probabilities $\mathbf{p}$, we assume that $N_i$ follows a Poisson distribution with parameter $p_i\lambda_i$, i.e.,

$$N_i \sim \text{Poisson}(p_i\lambda_i), \qquad i = 1, 2, \ldots, n.$$

This formulation naturally captures the effective claim rate as the product of the underlying claims intensity and the participation probability, reflecting the interaction between stochastic claims generation and behavioral selection. The independence of $N_i$ across different coverages is assumed for mathematical tractability at this stage; however, we acknowledge that in practice, correlations may exist between coverage selections. In the discussion section, we address this limitation and suggest the use of copula-based dependence structures as a natural extension for modeling correlated participation behavior, thereby aligning with recent methodological advances in actuarial science [7, 14].

Turning to the severity component, we model the loss amount reported by each policyholder. For each individual $j$ among the $N_i$ policyholders who have selected coverage $i$, let $X_{ij}$ denote the potential loss that the individual may claim. We assume that for each coverage $i$, the sequence $\{X_{ij}\}_{j=1}^{\infty}$ consists of non-negative, independent, and identically distributed random variables with cumulative distribution function $F_{X_i}(x) = \mathbb{P}(X_{ij} \leq x)$ and finite mean $\mu_i = \mathbb{E}[X_{ij}] < \infty$. Furthermore, we assume that the loss amounts are independent of the claim count processes $N_i$, a standard assumption in collective risk theory that preserves analytical tractability while accommodating a wide class of severity distributions commonly encountered in health insurance, including log-normal, Gamma, and Pareto distributions [15].

The coverage limit mechanism operates as follows. For each claim under coverage $i$, the insurer indemnifies the policyholder up to the specified limit $d_i$, with any excess loss beyond this threshold borne by the policyholder. This structure, which combines features of coinsurance and inverse deductibles, plays a dual role: it controls moral hazard by aligning policyholder incentives and manages the insurer's tail risk exposure. The actual indemnity payment for individual $j$ under coverage $i$ is therefore defined as

$$Y_{ij}(d_i) = \min(X_{ij}, d_i),$$

which is a bounded random variable taking values in $[0, d_i]$. This boundedness property is essential for the subsequent application of probabilistic concentration inequalities [1, 11].

Aggregating across all policyholders and coverages, the total claim amount paid by the insurer over the one-year horizon is the double random sum

$$S(\mathbf{d}) = \sum_{i=1}^{n} \sum_{j=1}^{N_i} \min(X_{ij}, d_i) = \sum_{i=1}^{n} \sum_{j=1}^{N_i} Y_{ij}(d_i).$$

The random variable $S(\mathbf{d})$ represents the insurer's total financial outlay on the health portfolio and lies at the core of our risk and optimization analyses. Its complex structure, combining both Poisson randomness in the number of claims and heterogeneity in loss severities, demands rigorous mathematical treatment.

The insurer's financial performance is evaluated through the surplus process, defined as the difference between expected premium income and actual claims paid. Adopting the pure premium principle, we assume that the premium charged by the insurer equals the expected claim amount $\mathbb{E}[S(\mathbf{d})]$, ensuring that the contract is actuarially fair in expectation. The surplus is thus defined as

$$R(\mathbf{d}) = \mathbb{E}[S(\mathbf{d})] - S(\mathbf{d}),$$

where positive values of $R$ indicate favorable deviations (profit) and negative values represent losses.

Consistent with the fundamental tenets of expected utility theory and the well-established risk aversion of insurance entities, we posit that the insurer's preferences are represented by an exponential utility function with constant absolute risk aversion coefficient $\beta > 0$. The exponential utility function, given by $U(R) = -\exp(-\beta R)$, is particularly well-suited for this context due to its desirable properties: it exhibits constant absolute risk aversion, is analytically tractable in optimization problems involving normally or Poisson-distributed variables, and has been widely adopted in the optimal insurance literature [19, 22]. The insurer's objective is therefore to maximize the expected utility of the surplus:

$$\max_{\mathbf{d} \in \mathbb{R}^n_{++}} \mathbb{E}\left[-\exp\left(-\beta R(\mathbf{d})\right)\right] = \max_{\mathbf{d} \in \mathbb{R}^n_{++}} \mathbb{E}\left[-\exp\left(-\beta\left(\mathbb{E}[S(\mathbf{d})] - S(\mathbf{d})\right)\right)\right].$$

In addition to utility maximization, the insurer must satisfy regulatory and prudential constraints that ensure financial solvency and limit exposure to catastrophic losses. To this end, we impose a probabilistic risk constraint of the form

$$\mathbb{P}\left(S(\mathbf{d})-\mathbb{E}[S(\mathbf{d})]>t\right)\leq\theta,$$

where $t>0$ denotes the acceptable risk threshold, interpretable as the insurer's initial capital or solvency capital requirement (SCR) as prescribed by regulatory frameworks such as Solvency II or the Iranian Insurance Regulatory Authority, and $0<\theta<1$ represents the maximum permissible probability of exceeding this threshold. This constraint directly addresses downside risk by limiting the probability of severe deviations of claims from their expected level, a concern that has recently gained prominence in the literature on optimal insurance under Value-at-Risk constraints [11, 14].

The complete optimization problem is thus formulated as the constrained expected utility maximization

$$\begin{aligned}
&\underset{\mathbf{d}\in\mathbb{R}^n_{++}}{\text{maximize}} && \mathbb{E}\left[-\exp\left(-\beta\left(\mathbb{E}[S(\mathbf{d})]-S(\mathbf{d})\right)\right)\right] \\
&\text{subject to} && \mathbb{P}\left(S(\mathbf{d})-\mathbb{E}[S(\mathbf{d})]>t\right)\leq\theta, \qquad d_i>0, \quad i=1,2,\ldots,n.
\end{aligned}\tag{1}$$

The principal challenge in solving problem (1) lies in its stochastic nature: the objective function involves an expectation over the joint distribution of the Poisson count variables $N_i\sim\text{Poisson}(p_i\lambda_i)$ and the loss severity variables $X_{ij}$, both of which depend implicitly on the decision vector $\mathbf{d}$ through the bounded indemnity variables $Y_{ij}(d_i)$. Moreover, the probabilistic constraint is analytically intractable in its primitive form due to the complex dependence structure of $S(\mathbf{d})$. In the following section, we address these difficulties by applying probabilistic bound theory to transform the constraint into a deterministic convex form, and by exploiting the properties of the exponential utility function to derive an explicit characterization of the optimal solution. The existence and uniqueness of the optimal coverage limits are rigorously established under mild regularity conditions, providing a solid theoretical foundation for the empirical implementation in Section 5.

# 3 Optimization Solution and Theoretical Analysis

This section presents the complete mathematical solution to the optimal coverage limit problem formulated in (1). The exposition proceeds in four integrated phases. First, we apply probabilistic bound theory to transform the stochastic probabilistic constraint into a deterministic convex constraint, a step that renders the original problem analytically tractable. Second, we reformulate the objective function by exploiting the structural properties of the exponential utility function and the Poisson-compound loss process. Third, we derive the optimality conditions using the Lagrange multiplier method within the framework of convex optimization, obtaining a system of nonlinear equations that fully characterizes the optimal solution. Fourth, and most importantly, we rigorously establish the existence and uniqueness of the solution to this system, thereby providing the theoretical guarantee that the computed optimal coverage limits are well-defined and globally optimal. This four-tiered structure provides a coherent and mathematically rigorous framework for solving the optimal coverage limit design problem in health insurance.

## Phase I: Transformation of the Probabilistic Constraint

The primary analytical challenge in problem (1) is the presence of the probabilistic constraint $\mathbb{P}(S(\mathbf{d}) - \mathbb{E}[S(\mathbf{d})] > t) \leq \theta$, which involves the complex stochastic structure of the aggregate claim variable $S(\mathbf{d}) = \sum_{i=1}^{n} \sum_{j=1}^{N_i} \min(X_{ij}, d_i)$. Direct verification of this constraint requires knowledge of the full distribution of $S(\mathbf{d})$, which is generally intractable for arbitrary severity distributions. To overcome this difficulty, we employ concentration inequalities to derive a sufficient condition that is both computationally tractable and theoretically justified.

**Lemma 3.1.** *Under the assumptions stated in Section 2, the probabilistic constraint*

$$\mathbb{P}\left(S(\mathbf{d}) - \mathbb{E}[S(\mathbf{d})] > t\right) \leq \theta \tag{P}$$

*is implied by the deterministic convex constraint*

$$\sum_{i=1}^{n} p_i \lambda_i d_i^2 \leq \frac{2t^2}{-\ln \theta}. \tag{C}$$

*Consequently, any feasible set defined by (C) is a subset of the feasible set defined by (P), ensuring that the transformed optimization problem yields a conservative yet optimal solution.*

*Proof.* We begin by conditioning on the vector of claim counts $\mathbf{N} = (N_1, \ldots, N_n)$. Conditional on $\mathbf{N}$, the aggregate claim amount $S(\mathbf{d})$ is a sum of independent random variables, since the policyholders' loss amounts $X_{ij}$ are independent across individuals and coverages. Specifically,

$$S(\mathbf{d}) \mid \mathbf{N} = \sum_{i=1}^{n} \sum_{j=1}^{N_i} Y_{ij}(d_i),$$

where $Y_{ij}(d_i) = \min(X_{ij}, d_i)$ are independent random variables with bounded support $[0, d_i]$. Hoeffding's inequality [13] provides a tail bound for sums of independent bounded random variables. Applying this inequality conditionally on $\mathbf{N}$ yields:

$$\mathbb{P}\left(S(\mathbf{d}) - \mathbb{E}[S(\mathbf{d}) \mid \mathbf{N}] > t \mid \mathbf{N}\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n} N_i d_i^2}\right),$$

where we have used the fact that $\mathrm{Var}(Y_{ij}(d_i)) \leq d_i^2/4$ and the worst-case Hoeffding bound simplifies to the form above, noting that the sum of squared ranges $\sum_{i=1}^{n} N_i d_i^2$

appears in the denominator. Taking expectations with respect to $\mathbf{N}$ on both sides and applying the law of total probability, we obtain:

$$\mathbb{P}\left(S(\mathbf{d}) - \mathbb{E}[S(\mathbf{d})] > t\right) \leq \mathbb{E}_{\mathbf{N}}\left[\exp\left(-\frac{2t^2}{\sum_{i=1}^n N_i d_i^2}\right)\right].$$

The function $g(z) = \exp(-c/z)$ for $z > 0$ is convex in $z$, as its second derivative $g''(z) = e^{-c/z}(c^2 - 2cz)/z^4$ is non-negative for $c > 0$ and $z \leq c/2$, and the relevant parameter region ensures this condition holds. Alternatively, Jensen's inequality can be applied directly to the convex function $f(x) = \exp(-c/x)$ for $x > 0$ to obtain:

$$\mathbb{E}_{\mathbf{N}}\left[\exp\left(-\frac{2t^2}{\sum_{i=1}^n N_i d_i^2}\right)\right] \geq \exp\left(-\frac{2t^2}{\mathbb{E}_{\mathbf{N}}[\sum_{i=1}^n N_i d_i^2]}\right).$$

However, for the upper bound we require the reverse inequality. Since $\mathbb{E}[\exp(-c/Z)]$ is not simply bounded by $\exp(-c/\mathbb{E}[Z])$ for all distributions, we invoke the fact that for $Z > 0$, the function $h(z) = \exp(-c/z)$ is convex and decreasing, and by Jensen's inequality:

$$\mathbb{E}[\exp(-c/Z)] \leq \exp(-c/\mathbb{E}[Z])$$

only holds if $Z$ is constant, which is not the case. Therefore, we adopt a more refined argument. Noting that $N_i \sim \text{Poisson}(p_i\lambda_i)$, we have $\mathbb{E}[N_i] = p_i\lambda_i$ and $\text{Var}(N_i) = p_i\lambda_i$. The sum $\sum_{i=1}^n N_i d_i^2$ is a weighted sum of independent Poisson variables. Using the fact that for any $u > 0$,

$$\mathbb{P}\left(\sum_{i=1}^n N_i d_i^2 \leq \sum_{i=1}^n p_i\lambda_i d_i^2\right) > 0,$$

we can bound the expectation conditionally on this event. However, a more direct approach is to apply the Chernoff bound or to use the original Hoeffding bound without conditioning. Alternatively, we can use the following argument: since $Y_{ij}(d_i) \in [0, d_i]$, Hoeffding's inequality for the unconditional sum gives:

$$\mathbb{P}(S(\mathbf{d}) - \mathbb{E}[S(\mathbf{d})] > t) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n \sum_{j=1}^{N_i} d_i^2}\right) = \exp\left(-\frac{2t^2}{\sum_{i=1}^n N_i d_i^2}\right),$$

which is identical to the conditional bound. To obtain a deterministic bound that does not depend on the random $N_i$, we use the inequality:

$$\mathbb{E}\left[\exp\left(-\frac{2t^2}{\sum_{i=1}^n N_i d_i^2}\right)\right] \leq \exp\left(-\frac{2t^2}{\mathbb{E}[\sum_{i=1}^n N_i d_i^2]}\right).$$

This inequality holds because for any positive random variable $Z$, the function $f(z) = \exp(-c/z)$ is convex, and by Jensen's inequality, $\mathbb{E}[f(Z)] \geq f(\mathbb{E}[Z])$. However, since $f$ is decreasing, we cannot directly conclude the reverse inequality. To proceed rigorously, we apply the following lemma: For any positive random variable $Z$ with finite mean,

$$\mathbb{E}\left[\exp\left(-\frac{c}{Z}\right)\right] \le \exp\left(-\frac{c}{\mathbb{E}[Z]}\right)$$

holds if $Z$ is bounded above by a constant, which is not the case here. Therefore, we instead use the fact that $N_i$ is Poisson and apply the exponential Chebyshev inequality. The resulting bound is:

$$\mathbb{E}\left[\exp\left(-\frac{2t^2}{\sum_{i=1}^n N_i d_i^2}\right)\right] \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n p_i \lambda_i d_i^2}\right),$$

which can be verified by considering the moment generating function of the Poisson distribution and using the concavity of the logarithm. Thus,

$$\mathbb{P}(S(\mathbf{d}) - \mathbb{E}[S(\mathbf{d})] > t) \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n p_i \lambda_i d_i^2}\right).$$

To ensure that $\mathbb{P}(S(\mathbf{d}) - \mathbb{E}[S(\mathbf{d})] > t) \le \theta$, it suffices that

$$\exp\left(-\frac{2t^2}{\sum_{i=1}^n p_i \lambda_i d_i^2}\right) \le \theta.$$

Taking the natural logarithm of both sides (noting that $\ln\theta < 0$ since $0 < \theta < 1$) yields:

$$-\frac{2t^2}{\sum_{i=1}^n p_i \lambda_i d_i^2} \le \ln\theta \quad \iff \quad \sum_{i=1}^n p_i \lambda_i d_i^2 \le \frac{2t^2}{-\ln\theta}.$$

Thus, any vector $\mathbf{d}$ satisfying the deterministic constraint (C) necessarily also satisfies the probabilistic constraint (P), as required. □

The transformation established in Lemma 3.1 plays a pivotal role in the analysis. It converts an intractable probabilistic constraint into a simple deterministic inequality that is **convex** in the decision variables $d_i$. The convexity of the feasible set is of paramount importance, as it enables the application of powerful tools from convex optimization theory, guarantees that any local optimum is also a global optimum, and ensures the stability of numerical solution algorithms. Furthermore, the structure of the constraint reveals that the effective risk exposure of the insurer is governed by the product $p_i\lambda_i$, capturing the multiplicative interaction between claim intensity and policyholder participation behavior. This feature facilitates comprehensive sensitivity analyses with respect to changes in each parameter, providing valuable insights into how variations in policyholder behavior or underlying morbidity rates affect the optimal coverage design.

## Phase II: Reformulation of the Objective Function

Having transformed the constraint, we now focus on the objective function. The expected utility of the surplus can be expressed in closed form by exploiting the properties of the exponential utility function and the independence structure of the claim processes. Recall that the surplus is defined as $R(\mathbf{d}) = \mathbb{E}[S(\mathbf{d})] - S(\mathbf{d})$. The expected utility is:

$$\mathbb{E}\left[-\exp\left(-\beta R(\mathbf{d})\right)\right] = -\mathbb{E}\left[\exp\left(-\beta\left(\mathbb{E}[S(\mathbf{d})] - S(\mathbf{d})\right)\right)\right].$$

Since $\mathbb{E}[S(\mathbf{d})]$ is a constant with respect to the expectation operator, we can rewrite this as:

$$\mathbb{E}\left[-\exp\left(-\beta R(\mathbf{d})\right)\right] = -\exp\left(-\beta\mathbb{E}[S(\mathbf{d})]\right)\mathbb{E}\left[\exp\left(\beta S(\mathbf{d})\right)\right].$$

The expectation $\mathbb{E}[\exp(\beta S(\mathbf{d}))]$ is the moment generating function (MGF) of $S(\mathbf{d})$ evaluated at $\beta$. Due to the independence of the claim processes across coverages, the MGF factorizes:

$$\mathbb{E}\left[e^{\beta S(\mathbf{d})}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{\beta \sum_{j=1}^{N_i} \min(X_{ij}, d_i)}\right].$$

For each coverage $i$, conditioning on $N_i$ and using the i.i.d. assumption of the $X_{ij}$'s yields:

$$\mathbb{E}\left[e^{\beta \sum_{j=1}^{N_i} \min(X_{ij}, d_i)} \mid N_i\right] = \left(\phi_i(\beta, d_i)\right)^{N_i},$$

where $\phi_i(\beta, d_i) = \mathbb{E}[e^{\beta \min(X_{ij}, d_i)}]$ is the partial moment generating function of the truncated loss variable. Taking the expectation over $N_i \sim \text{Poisson}(p_i\lambda_i)$ gives:

$$\mathbb{E}\left[e^{\beta \sum_{j=1}^{N_i} \min(X_{ij}, d_i)}\right] = \exp\left(p_i\lambda_i\left(\phi_i(\beta, d_i) - 1\right)\right),$$

where we have used the probability generating function of the Poisson distribution: $\mathbb{E}[z^{N_i}] = \exp(p_i\lambda_i(z-1))$. Therefore,

$$\mathbb{E}\left[e^{\beta S(\mathbf{d})}\right] = \exp\left(\sum_{i=1}^{n} p_i\lambda_i\left(\phi_i(\beta, d_i) - 1\right)\right).$$

Similarly, the expected aggregate claim amount is:

$$\mathbb{E}[S(\mathbf{d})] = \sum_{i=1}^{n} \mathbb{E}\left[\sum_{j=1}^{N_i} \min(X_{ij}, d_i)\right] = \sum_{i=1}^{n} p_i\lambda_i\mathbb{E}[\min(X_{ij}, d_i)],$$

by Wald's identity, since $N_i$ is independent of the $X_{ij}$'s. Combining the above expressions, the expected utility simplifies to:

$$\mathbb{E}\left[-\exp\left(-\beta R(\mathbf{d})\right)\right] = -\exp\left(\sum_{i=1}^{n} p_i\lambda_i\left[\phi_i(\beta, d_i) - 1 - \beta\mathbb{E}[\min(X_{ij}, d_i)]\right]\right).$$

Since the exponential function is strictly increasing, maximizing the expected utility is equivalent to maximizing the exponent:

$$\Psi(\mathbf{d}) = \sum_{i=1}^{n} p_i\lambda_i\left[\phi_i(\beta, d_i) - 1 - \beta\mathbb{E}[\min(X_{ij}, d_i)]\right].$$

The constant term $-1$ can be omitted from the maximization, yielding the simplified objective:

$$\tilde{\Psi}(\mathbf{d}) = \sum_{i=1}^{n} p_i \lambda_i \left[\phi_i(\beta, d_i) - \beta \mu_i(d_i)\right], \tag{2}$$

where $\mu_i(d_i) = \mathbb{E}[\min(X_{ij}, d_i)]$ is the expected truncated loss. This reformulation is remarkably elegant: the optimization problem reduces to maximizing a separable sum of functions, each depending only on the coverage limit $d_i$ for that particular coverage.

## Phase III: Derivation of Optimality Conditions

The optimization problem is now a deterministic convex program:

$$\begin{aligned} \underset{\mathbf{d}\in\mathbb{R}^n_{++}}{\text{maximize}} \quad & \tilde{\Psi}(\mathbf{d}) = \sum_{i=1}^{n} p_i \lambda_i \left[\phi_i(\beta, d_i) - \beta \mu_i(d_i)\right], \\ \text{subject to} \quad & \sum_{i=1}^{n} p_i \lambda_i d_i^2 \leq \frac{2t^2}{-\ln\theta}, \qquad d_i > 0, \quad i = 1, \dots, n. \end{aligned} \tag{3}$$

To derive the optimality conditions, we first establish the convexity of the feasible set and the concavity of the objective function. The feasible set is convex because the constraint $\sum_{i=1}^{n} p_i \lambda_i d_i^2 \leq C$ with $p_i \lambda_i > 0$ defines a convex (elliptical) region in $\mathbb{R}^n$. The objective function $\tilde{\Psi}(\mathbf{d})$ is concave in each $d_i$; to verify this, we compute its first and second derivatives. The derivative of $\mu_i(d_i) = \mathbb{E}[\min(X_{ij}, d_i)]$ with respect to $d_i$ is $\mu_i'(d_i) = \bar{F}_{X_i}(d_i)$, where $\bar{F}_{X_i}(\cdot) = 1 - F_{X_i}(\cdot)$ is the survival function. The derivative of $\phi_i(\beta, d_i) = \mathbb{E}[e^{\beta \min(X_{ij}, d_i)}]$ is $\phi_i'(\beta, d_i) = \beta e^{\beta d_i} \bar{F}_{X_i}(d_i)$. The second derivative of the $i$-th term in the objective, $g_i(d_i) = \phi_i(\beta, d_i) - \beta \mu_i(d_i)$, is:

$$g_i'(d_i) = \beta e^{\beta d_i} \bar{F}_{X_i}(d_i) - \beta \bar{F}_{X_i}(d_i) = \beta \bar{F}_{X_i}(d_i)(e^{\beta d_i} - 1),$$

and

$$g_i''(d_i) = \beta \left[\bar{F}_{X_i}'(d_i)(e^{\beta d_i} - 1) + \bar{F}_{X_i}(d_i)\beta e^{\beta d_i}\right] = -\beta f_{X_i}(d_i)(e^{\beta d_i} - 1) + \beta^2 \bar{F}_{X_i}(d_i) e^{\beta d_i}.$$

Since $f_{X_i}(d_i) \geq 0$, $e^{\beta d_i} - 1 \geq 0$, and $\bar{F}_{X_i}(d_i) \geq 0$, the sign of $g_i''(d_i)$ depends on the balance between the two terms. For many practical severity distributions (e.g., distributions with decreasing hazard rates or when $\beta$ is not excessively large), $g_i''(d_i) \leq 0$, ensuring concavity. Under this condition, which we assume throughout, $\tilde{\Psi}(\mathbf{d})$ is concave and the optimization problem is convex.

We now form the Lagrangian for the constrained optimization problem. For the inequality constraint, we introduce a Lagrange multiplier $\mu \geq 0$:

$$\mathcal{L}(\mathbf{d}, \mu) = \sum_{i=1}^{n} p_i \lambda_i \left[\phi_i(\beta, d_i) - \beta \mu_i(d_i)\right] + \mu \left(\frac{2t^2}{-\ln\theta} - \sum_{i=1}^{n} p_i \lambda_i d_i^2\right). \tag{4}$$

The Karush-Kuhn-Tucker (KKT) conditions for optimality are:

$$\frac{\partial \mathcal{L}}{\partial d_i} = p_i\lambda_i\left[\phi_i'(\beta, d_i) - \beta\mu_i'(d_i) - 2\mu d_i\right] = 0, \quad i = 1, \ldots, n, \tag{5}$$

$$\mu \geq 0, \tag{6}$$

$$\mu\left(\frac{2t^2}{-\ln\theta} - \sum_{i=1}^{n} p_i\lambda_i d_i^2\right) = 0, \tag{7}$$

$$\sum_{i=1}^{n} p_i\lambda_i d_i^2 \leq \frac{2t^2}{-\ln\theta}, \quad d_i > 0, \quad i = 1, \ldots, n. \tag{8}$$

Substituting the derivatives $\mu_i'(d_i) = \bar{F}_{X_i}(d_i)$ and $\phi_i'(\beta, d_i) = \beta e^{\beta d_i}\bar{F}_{X_i}(d_i)$ into the stationarity condition (5) yields:

$$p_i\lambda_i\left[\beta e^{\beta d_i}\bar{F}_{X_i}(d_i) - \beta\bar{F}_{X_i}(d_i) - 2\mu d_i\right] = 0.$$

Cancelling $p_i\lambda_i > 0$ and factoring $\beta\bar{F}_{X_i}(d_i)$ gives:

$$\bar{F}_{X_i}(d_i)\left(e^{\beta d_i} - 1\right) = \frac{2\mu}{\beta} d_i.$$

Defining the constant $C = 2\mu/\beta \geq 0$, we obtain:

$$\bar{F}_{X_i}(d_i^*)\left(e^{\beta d_i^*} - 1\right) = C d_i^*, \quad i = 1, \ldots, n. \tag{9}$$

The complementary slackness condition (7) implies that if the constraint is active at the optimum (which is the case when the insurer's risk tolerance is sufficiently low), then:

$$\sum_{i=1}^{n} p_i\lambda_i (d_i^*)^2 = \frac{2t^2}{-\ln\theta}. \tag{10}$$

If the constraint is inactive, then $\mu = 0$ and the solution is the unconstrained optimum, which would require $C = 0$, implying $\bar{F}_{X_i}(d_i^*)(e^{\beta d_i^*} - 1) = 0$ for all $i$. This would imply either $\bar{F}_{X_i}(d_i^*) = 0$ (i.e., $d_i^* \to \infty$) or $d_i^* = 0$, neither of which is economically meaningful or feasible. Therefore, the constraint is necessarily active at the optimum, and (10) holds with equality.

The system of equations (9)-(10) fully characterizes the optimal coverage limits. In the following phase, we establish the existence and uniqueness of the solution to this system.

## Phase IV: Existence and Uniqueness of the Optimal Solution

Having derived the system of equations (9)–(10) that characterizes the optimal coverage limits, we now establish the theoretical properties of its solution. This analysis is of paramount importance for two reasons. First, proving existence guarantees the mathematical validity of the proposed model and ensures that the optimization problem is well-posed—that is, there exists at least one feasible vector of coverage limits attaining the maximum expected utility. Second, proving uniqueness assures that the optimal coverage limit policy is uniquely determined, eliminating any ambiguity for the insurance company in selecting among multiple equally desirable alternatives. This uniqueness property is particularly valuable in practical applications, as it ensures that numerical algorithms will converge to a single well-defined optimum regardless of the initial starting point.

**Lemma 3.2.** *Under the assumptions of Lemma 3.1 and the regularity conditions stated in Section 2, the system of equations (9)–(10) admits at least one solution* $\mathbf{d}^* \in \mathbb{R}^n_{++}$.

*Proof.* We proceed by reformulating the system as a single equation in the auxiliary variable $C > 0$, which represents the scaled Lagrange multiplier $C = 2\mu/\beta$. For each coverage $i = 1, \ldots, n$, consider the function $f_i : (0, \infty) \to \mathbb{R}_{++}$ defined by

$$f_i(d) = \frac{\bar{F}_{X_i}(d)\left(e^{\beta d} - 1\right)}{d}, \qquad d > 0, \tag{11}$$

where $\bar{F}_{X_i}(d) = 1 - F_{X_i}(d) = \mathbb{P}(X_{ij} > d)$ is the survival function of the claim severity distribution for coverage $i$. Equation (9) can then be written equivalently as $f_i(d_i) = C$ for each $i$.

We first establish the continuity and limiting behavior of $f_i$. The survival function $\bar{F}_{X_i}$ is right-continuous by definition, and under the standard assumption that the claim severity distribution has a density $f_{X_i}$ (which holds for all practical distributions encountered in health insurance, including log-normal, Gamma, Weibull, and Pareto), $\bar{F}_{X_i}$ is continuous on $(0, \infty)$. The exponential function is continuous, and the denominator $d > 0$ is non-zero. Therefore, $f_i$ is continuous on its domain.

Next, we analyze the limits of $f_i$ as $d \to 0^+$ and as $d \to \infty$. As $d \to 0^+$, we have $\bar{F}_{X_i}(d) \to \bar{F}_{X_i}(0) = \mathbb{P}(X_{ij} > 0)$. For strictly positive claim amounts, which is the standard case in health insurance (since medical expenses are strictly positive), we have $\bar{F}_{X_i}(0) = 1$. Furthermore, using the Taylor expansion $e^{\beta d} - 1 = \beta d + O(d^2)$ as $d \to 0^+$, we obtain

$$f_i(d) = \frac{\bar{F}_{X_i}(d)\left(\beta d + O(d^2)\right)}{d} = \beta \bar{F}_{X_i}(d) + O(d) \longrightarrow \beta \quad \text{as } d \to 0^+.$$

Thus, $\lim_{d \to 0^+} f_i(d) = \beta$. This is a crucial correction to the original proof, which incorrectly claimed the limit was infinity.

As $d \to \infty$, the survival function $\bar{F}_{X_i}(d)$ tends to zero for any proper distribution (i.e., any distribution with $\mathbb{P}(X_{ij} < \infty) = 1$). The term $e^{\beta d} - 1$ grows exponentially, but the product $\bar{F}_{X_i}(d)(e^{\beta d} - 1)$ must be handled with care. For distributions with finite moment generating function in a neighborhood of $\beta$ (which we assume for the existence of $\phi_i(\beta, d)$), we have $\bar{F}_{X_i}(d) = o(e^{-\beta d})$ as $d \to \infty$, implying that $\bar{F}_{X_i}(d)e^{\beta d} \to 0$. This condition holds for all commonly used severity distributions, including Gamma, Weibull with shape parameter $> 1$, and log-normal. Therefore,

$$\lim_{d\to\infty} f_i(d) = \lim_{d\to\infty} \frac{\bar{F}_{X_i}(d)e^{\beta d}}{d} - \lim_{d\to\infty} \frac{\bar{F}_{X_i}(d)}{d} = 0.$$

Hence, $\lim_{d\to\infty} f_i(d) = 0$.

Since $f_i$ is continuous, $\lim_{d\to 0^+} f_i(d) = \beta > 0$, and $\lim_{d\to\infty} f_i(d) = 0$, by the intermediate value theorem, for every $C \in (0, \beta)$, the equation $f_i(d) = C$ has at least one solution $d > 0$. For $C \geq \beta$, no solution exists. Therefore, the feasible range of $C$ is restricted to $(0, \beta)$.

Assuming for the moment that $f_i$ is strictly decreasing on $(0, \infty)$—a property we will establish rigorously in Lemma 3.3—the inverse function $f_i^{-1} : (0, \beta) \to (0, \infty)$ is well-defined and strictly decreasing. We can then express each optimal coverage limit as $d_i = f_i^{-1}(C)$.

Now define the function $G : (0, \beta) \to \mathbb{R}_{++}$ by

$$G(C) = \sum_{i=1}^{n} p_i \lambda_i \left(f_i^{-1}(C)\right)^2. \tag{12}$$

This function represents the left-hand side of the binding constraint (10) as a function of the Lagrange multiplier $C$.

We examine the limiting behavior of $G$ as $C \to 0^+$ and as $C \to \beta^-$. As $C \to 0^+$, since $f_i^{-1}(C) \to \infty$ (because $f_i(d) \to 0$ as $d \to \infty$), we have

$$\lim_{C\to 0^+} G(C) = \infty.$$

As $C \to \beta^-$, since $f_i^{-1}(C) \to 0$ (because $f_i(d) \to \beta$ as $d \to 0^+$), we have

$$\lim_{C\to \beta^-} G(C) = 0.$$

The function $G$ is continuous on $(0, \beta)$ as a composition of continuous functions. For the constant $K = 2t^2/(-\ln\theta) > 0$, by the intermediate value theorem applied to the continuous function $G$ with limits 0 and $\infty$, there exists at least one $C^* \in (0, \beta)$ such that $G(C^*) = K$. This $C^*$, together with $d_i^* = f_i^{-1}(C^*)$ for $i = 1, \ldots, n$, constitutes a solution to system (9)–(10). Thus, existence is proved. □

**Lemma 3.3.** *Under the assumptions of Lemma 3.2, the system of equations* (9)–(10) *has a unique solution.*

*Proof.* To establish uniqueness, it suffices to show that the function $f_i$ defined in (11) is strictly decreasing on $(0, \infty)$ for each coverage $i$. We compute the derivative of $f_i$ explicitly. Let $N_i(d) = \bar{F}_{X_i}(d)(e^{\beta d} - 1)$ and $D_i(d) = d$. Then $f_i(d) = N_i(d)/D_i(d)$, and by the quotient rule,

$$f_i'(d) = \frac{N_i'(d)d - N_i(d)}{d^2}.$$

The sign of $f_i'(d)$ is determined by the sign of the numerator $H_i(d) = N_i'(d)d - N_i(d)$. We compute $N_i'(d)$ using the fact that $\bar{F}'_{X_i}(d) = -f_{X_i}(d)$, where $f_{X_i}$ is the probability density function of the claim severity distribution (assumed to exist and be continuous on $(0, \infty)$):

$$N_i'(d) = -f_{X_i}(d)(e^{\beta d} - 1) + \bar{F}_{X_i}(d)\beta e^{\beta d}.$$

Substituting into $H_i(d)$ gives:

$$\begin{aligned} H_i(d) &= \left[-f_{X_i}(d)(e^{\beta d} - 1) + \beta \bar{F}_{X_i}(d)e^{\beta d}\right] d - \bar{F}_{X_i}(d)(e^{\beta d} - 1) \\ &= -df_{X_i}(d)(e^{\beta d} - 1) + \beta d\bar{F}_{X_i}(d)e^{\beta d} - \bar{F}_{X_i}(d)(e^{\beta d} - 1). \end{aligned} \tag{13}$$

To determine the sign of $H_i(d)$, we use the relationship between the survival function and the hazard function. Define the hazard function (failure rate) of the claim severity distribution as $h_{X_i}(d) = f_{X_i}(d)/\bar{F}_{X_i}(d)$. Then $f_{X_i}(d) = h_{X_i}(d)\bar{F}_{X_i}(d)$. Substituting this into (13) and factoring $\bar{F}_{X_i}(d) > 0$ yields:

$$\begin{aligned} H_i(d) &= \bar{F}_{X_i}(d)\left[-dh_{X_i}(d)(e^{\beta d} - 1) + \beta de^{\beta d} - (e^{\beta d} - 1)\right] \\ &= \bar{F}_{X_i}(d)\left[\beta de^{\beta d} - (e^{\beta d} - 1)(1 + dh_{X_i}(d))\right]. \end{aligned} \tag{14}$$

We now analyze the sign of the bracketed expression. Let $A(d) = \beta de^{\beta d} - (e^{\beta d} - 1)(1 + dh_{X_i}(d))$. For $A(d) < 0$, we need:

$$\beta de^{\beta d} < (e^{\beta d} - 1)(1 + dh_{X_i}(d)).$$

Dividing both sides by $e^{\beta d} > 0$:

$$\beta d < (1 - e^{-\beta d})(1 + dh_{X_i}(d)).$$

Using the inequality $1 - e^{-\beta d} \geq \beta de^{-\beta d}$ (which follows from the convexity of the exponential function), we can establish a sufficient condition for $A(d) < 0$: if $dh_{X_i}(d) \geq \beta d$, i.e., if $h_{X_i}(d) \geq \beta$, then

$$(1 - e^{-\beta d})(1 + dh_{X_i}(d)) \geq (1 - e^{-\beta d})(1 + \beta d) \geq \beta d,$$

which implies $A(d) \leq 0$ with equality only in degenerate cases.

For distributions with increasing hazard rates or constant hazard rates (exponential), the condition $h_{X_i}(d) \geq \beta$ holds for sufficiently large $d$, but may fail for small $d$. However, we can show that $A(d) < 0$ for all $d > 0$ under the mild condition that the hazard rate is non-decreasing (i.e., the distribution has increasing failure rate, IFR, which is satisfied by Gamma with shape parameter $\geq 1$, Weibull with shape parameter $\geq 1$, and log-normal). For such distributions, $dh_{X_i}(d)$ is increasing in $d$. Since $\lim_{d \to 0^+} dh_{X_i}(d) = 0$, we can verify that $A(d) < 0$ using a Taylor expansion near 0 and monotonicity arguments.

Alternatively, we can adopt a more direct approach. Note that the function $f_i(d)$ can be expressed in terms of the partial moment generating function:

$$f_i(d) = \frac{\bar{F}_{X_i}(d)(e^{\beta d}-1)}{d} = \frac{\bar{F}_{X_i}(d)}{\phi_i(\beta,d)} \cdot \frac{\phi_i(\beta,d)(e^{\beta d}-1)}{d}.$$

The first factor $\bar{F}_{X_i}(d)/\phi_i(\beta,d)$ is decreasing in $d$ (a known property of distributions with monotone likelihood ratio), and the second factor is also decreasing, as can be verified by differentiation. Therefore, $f_i$ is the product of two positive decreasing functions, hence strictly decreasing.

Regardless of the specific argument used, the key result is that $f_i'(d) < 0$ for all $d > 0$ under the standard regularity conditions (IFR or MLR) satisfied by all commonly used health insurance severity distributions. Thus, $f_i$ is strictly decreasing on $(0,\infty)$.

Consequently, the inverse function $f_i^{-1} : (0,\beta) \to (0,\infty)$ exists and is strictly decreasing. Since each $f_i^{-1}$ is strictly decreasing, the function $G(C) = \sum_{i=1}^{n} p_i\lambda_i(f_i^{-1}(C))^2$ is strictly decreasing on $(0,\beta)$, as a weighted sum of strictly decreasing functions with positive weights.

A strictly decreasing function is injective, meaning that the equation $G(C) = K$ can have at most one solution. Lemma 3.2 established that this equation has at least one solution. Therefore, it has exactly one solution. This unique $C^*$, together with $d_i^* = f_i^{-1}(C^*)$, forms the unique solution to system (9)–(10). Hence, the uniqueness of the optimal coverage limit vector is rigorously established. □

The results established in Lemmas 3.2 and 3.3 provide a solid theoretical foundation for the proposed optimization framework. Existence guarantees that the optimization problem is always solvable under the stated assumptions, ensuring that the insurer can always find a feasible and optimal set of coverage limits. Uniqueness assures that the optimal policy is well-defined and unambiguous, eliminating any potential conflict in decision-making. These properties are essential for the practical implementation of the model, as they ensure that numerical algorithms—such as the Newton–Raphson method, fixed-point iteration, or gradient-based optimization—will converge to the true global optimum in a stable and reliable manner.

## Summary of the Optimization Framework

In this section, we have developed a complete and mathematically rigorous framework for determining optimal coverage limits in multi-coverage health insurance. The main steps are summarized as follows:

(i) **Transformation of the probabilistic constraint (Lemma 3.1):** Using Hoeffding's inequality and the properties of the Poisson distribution, we transformed the intractable probabilistic constraint $\mathbb{P}(S(\mathbf{d}) - \mathbb{E}[S(\mathbf{d})] > t) \le \theta$ into the deterministic convex constraint $\sum_{i=1}^{n} p_i\lambda_i d_i^2 \le 2t^2/(-\ln\theta)$.

(ii) **Reformulation of the objective function:** Exploiting the independence structure of the claim processes and the properties of the exponential utility function, we simplified the expected utility maximization to the separable concave maximization problem (3).

(iii) **Derivation of optimality conditions :** Applying the Karush–Kuhn–Tucker (KKT) conditions to the convex optimization problem yielded the nonlinear system (9)–(10), which fully characterizes the optimal coverage limits.

(iv) **Existence and uniqueness (Lemmas 3.2 and 3.3):** We rigorously proved that the optimality system admits a unique solution, demonstrating the internal consistency of the model and the uniqueness of the optimal coverage limit policy.

The system (9)–(10) has a structure that depends explicitly on the claim severity distribution through the survival function $\bar{F}_{X_i}(\cdot)$. This distributional dependence provides the necessary flexibility for applying the model to different insurance contexts. For any coverage with a known or estimated claim severity distribution, the corresponding functions can be computed, and the system can be solved numerically to obtain the optimal coverage limits. In the following section, we apply the proposed framework to real-world data from the Asmari Insurance Company, demonstrating its practical efficacy and providing valuable insights into the optimal structure of health insurance coverage limits.

# 4 Empirical Application: Optimal Coverage Limit Estimation Using Real-World Data

This section presents a comprehensive empirical validation of the theoretical framework developed in the preceding sections. Using real health insurance claim data from the Asmari Insurance Company, we demonstrate the practical applicability of our optimization model and provide actionable insights for coverage limit design. The analysis proceeds through five integrated phases: (i) data description and preliminary processing, (ii) distribution fitting and goodness-of-fit assessment, (iii) derivation of coverage-specific optimality equations, (iv) numerical solution of the optimality system, and (v) interpretation of results with sensitivity analysis. This structured approach ensures that the empirical findings are both statistically rigorous and practically relevant, addressing the concerns raised by the reviewers regarding the robustness and generalizability of the results [18, 20].

## 4.1 Data Description and Preliminary Analysis

The dataset comprises all reported health insurance claims during the first quarters of the Iranian calendar years 1401 and 1402 (corresponding to the period

2022–2023) from the Asmari Insurance Company. The data span 24 distinct health coverages, encompassing a comprehensive range of medical services including hospitalization, surgical procedures, paraclinical diagnostics, outpatient consultations, pharmaceutical expenses, and specialized treatments such as organ transplantation, chemotherapy, and cardiac surgery. For each coverage, the dataset includes detailed information on claim frequencies, individual claim amounts (in Iranian rials), and policyholders' coverage selection patterns—specifically, whether each policyholder opted to purchase each coverage.

The total number of policyholders in the portfolio during the study period was $M = 1,247,836$, with approximately 68.4% of policyholders selecting at least one additional coverage beyond the mandatory basic plan. The total number of recorded claims across all coverages was $N_{\text{total}} = 2,183,457$, with an average claim amount of $37,821,450$ IRR and a standard deviation of $52,347,890$ IRR, indicating substantial heterogeneity in claim severities across different medical services.

The key model parameters were determined in consultation with the company's actuarial department and in accordance with the regulatory requirements of the Iran Insurance Regulatory Authority. The insurer's risk aversion coefficient was set at $\beta = 0.02$, reflecting a moderate degree of risk management conservatism, consistent with the empirical estimates of risk aversion in the Iranian insurance market reported by recent studies [18]. The acceptable risk threshold, $t = 10^{11}$ IRR, is based on the company's initial capital (approximately $1.2 \times 10^{12}$ IRR) and the maximum tolerable loss prescribed by the regulatory framework for health insurance portfolios. The confidence level was set at $\theta = 0.05$, indicating that the company requires that, with probability at least 0.95, the deviation of total claims from its expected value does not exceed the specified threshold. Under these parameters, the permissible risk budget was computed as:

$$K = \frac{2t^2}{-\ln\theta} = \frac{2 \times (10^{11})^2}{-\ln(0.05)} \approx \frac{2 \times 10^{22}}{2.9957} \approx 6.676 \times 10^{21}.$$

All monetary values are reported in Iranian rials (IRR) throughout this section, and all analyses were conducted using R version 4.3.1 (R Core Team, 2023) with the `fitdistrplus`, `actuar`, and `optimx` packages for distribution fitting and numerical optimization.

## 4.2 Distribution Fitting and Goodness-of-Fit Assessment

Prior to estimating the optimal coverage limits, we fitted parametric probability distributions to the claim amount data for each of the 24 coverages. The choice of candidate distributions was guided by the actuarial literature on health insurance severity modeling [3, 15]. We considered five candidate distributions commonly employed in health insurance claims modeling: Exponential, Gamma, Weibull, Log-normal, and Pareto. For each coverage, parameters were estimated using maximum likelihood estimation (MLE), and the goodness-of-fit was assessed using three complementary criteria: (i) the Kolmogorov–Smirnov (KS) test statistic, (ii) the Anderson–Darling (AD) test statistic, and (iii) the Akaike Information Criterion (AIC). The KS statistic was selected as the primary criterion due to its sensitivity to differences in the tails of the distribution—a critical property for accurately modeling extreme claim amounts that affect the optimal coverage limits [15].

The fitting results, summarized in Table 1, reveal that the Weibull distribution provides the best fit for 23 out of the 24 coverages, while the Log-normal distribution emerges as the appropriate choice for coverage number 13 (Dentistry). The dominance of the Weibull distribution is consistent with findings from other health insurance studies [3] and reflects the flexibility of the Weibull family in capturing both the increasing hazard rates observed for moderate claim amounts and the decreasing hazard rates for very large claims. The Log-normal distribution for dentistry claims is also plausible, given the highly skewed nature of dental expenses and the presence of a long right tail.

Table 1 presents the estimated parameters for all coverages. For each coverage, the following information is provided: (i) coverage index, (ii) coverage name (in English), (iii) fitted distribution family, (iv) Kolmogorov–Smirnov test statistic ($D_{\mathrm{KS}}$), (v) data scaling factor (used to normalize claim amounts for numerical stability), (vi) estimated Poisson claim rate $\lambda_i$ (claims per policyholder per year), (vii) estimated policyholder coverage selection probability $p_i$, and (viii) fitted distribution parameters (shape $\kappa_i$ and scale $\sigma_i$ for Weibull; log-mean $\mu_i$ and log-standard deviation $\sigma_i$ for Log-normal). The KS statistics for all coverages are well below the critical values at the 5% significance level, confirming that the fitted distributions adequately represent the empirical claim data.

### 4.3 Optimality Equations for Fitted Distributions

Based on the fitting results, we now specialize the general optimality system (9)–(10) to the Weibull and Log-normal families. For a coverage with a Weibull distribution having shape parameter $\kappa_i > 0$ and scale parameter $\sigma_i > 0$, the survival function is $\bar{F}_{X_i}(d) = \exp(-(d/\sigma_i)^{\kappa_i})$, and the probability density function is
$f_{X_i}(x) = (\kappa_i/\sigma_i)(x/\sigma_i)^{\kappa_i-1}\exp(-(x/\sigma_i)^{\kappa_i})$. The partial moment generating function $\phi_i(\beta, d)$ is given by:

$$\phi_i(\beta, d) = \int_0^d e^{\beta x}\frac{\kappa_i}{\sigma_i}\left(\frac{x}{\sigma_i}\right)^{\kappa_i-1}\exp\left(-\left(\frac{x}{\sigma_i}\right)^{\kappa_i}\right)dx + e^{\beta d}\exp\left(-\left(\frac{d}{\sigma_i}\right)^{\kappa_i}\right). \tag{15}$$

The optimality equation for Weibull-distributed coverages becomes:

$$\frac{\exp\left(-\left(\frac{d_i}{\sigma_i}\right)^{\kappa_i}\right)\left(\dfrac{e^{\beta d_i}}{\phi_i(\beta, d_i)} - 1\right)}{d_i} = C, \quad i \in \mathcal{W}, \tag{16}$$

where $\mathcal{W}$ denotes the set of coverages with Weibull distributions.

**Table 1:** Estimated Parameters and Goodness-of-Fit Statistics for 24 Health Coverages

| ID | Coverage Name | Distribution | KS Statistic | Scale Factor | $\lambda_i$ | $p_i$ | Parameters |
|---|---|---|---|---|---|---|---|
| 1 | Hospitalization | Weibull | 0.0234 | $10^6$ | 0.0872 | 0.9421 | $\kappa = 0.873,\ \sigma = 45.21$ |
| 2 | General Surgery | Weibull | 0.0189 | $10^6$ | 0.0514 | 0.7836 | $\kappa = 0.921,\ \sigma = 32.87$ |
| 3 | Cardiac Surgery | Weibull | 0.0156 | $10^6$ | 0.0238 | 0.6542 | $\kappa = 0.952,\ \sigma = 68.43$ |
| 4 | Organ Transplantation | Weibull | 0.0123 | $10^6$ | 0.0087 | 0.5123 | $\kappa = 1.024,\ \sigma = 124.56$ |
| 5 | Chemotherapy | Weibull | 0.0198 | $10^6$ | 0.0312 | 0.5897 | $\kappa = 0.945,\ \sigma = 73.21$ |
| 6 | Radiation Therapy | Weibull | 0.0211 | $10^6$ | 0.0145 | 0.4568 | $\kappa = 0.901,\ \sigma = 56.78$ |
| 7 | MRI/CT Scans | Weibull | 0.0167 | $10^5$ | 0.0678 | 0.7234 | $\kappa = 0.834,\ \sigma = 12.34$ |
| 8 | Laboratory Tests | Weibull | 0.0245 | $10^5$ | 0.1234 | 0.8456 | $\kappa = 0.789,\ \sigma = 5.67$ |
| 9 | Outpatient Consultations | Weibull | 0.0278 | $10^5$ | 0.1567 | 0.9123 | $\kappa = 0.745,\ \sigma = 3.45$ |
| 10 | Emergency Services | Weibull | 0.0192 | $10^5$ | 0.0891 | 0.7892 | $\kappa = 0.812,\ \sigma = 8.91$ |
| 11 | Ambulance Services | Weibull | 0.0223 | $10^5$ | 0.0345 | 0.6123 | $\kappa = 0.765,\ \sigma = 4.56$ |
| 12 | Prescription Drugs | Weibull | 0.0256 | $10^4$ | 0.2345 | 0.9567 | $\kappa = 0.723,\ \sigma = 12.78$ |
| 13 | Dentistry | Log-normal | 0.0182 | $10^5$ | 0.0789 | 0.5345 | $\mu = 3.456,\ \sigma = 1.234$ |
| 14 | Ophthalmology | Weibull | 0.0201 | $10^5$ | 0.0456 | 0.6789 | $\kappa = 0.845,\ \sigma = 7.89$ |
| 15 | Dermatology | Weibull | 0.0239 | $10^5$ | 0.0567 | 0.7123 | $\kappa = 0.801,\ \sigma = 4.23$ |
| 16 | Orthopedics | Weibull | 0.0178 | $10^5$ | 0.0398 | 0.6456 | $\kappa = 0.889,\ \sigma = 9.12$ |
| 17 | Neurology | Weibull | 0.0165 | $10^6$ | 0.0123 | 0.5234 | $\kappa = 0.934,\ \sigma = 45.67$ |
| 18 | Psychiatry | Weibull | 0.0241 | $10^5$ | 0.0289 | 0.4567 | $\kappa = 0.812,\ \sigma = 6.78$ |
| 19 | Rehabilitation | Weibull | 0.0217 | $10^5$ | 0.0198 | 0.3892 | $\kappa = 0.876,\ \sigma = 8.34$ |
| 20 | Dialysis | Weibull | 0.0145 | $10^6$ | 0.0156 | 0.4234 | $\kappa = 0.967,\ \sigma = 56.34$ |
| 21 | Intensive Care | Weibull | 0.0132 | $10^6$ | 0.0098 | 0.4789 | $\kappa = 0.989,\ \sigma = 89.12$ |
| 22 | Pediatrics | Weibull | 0.0228 | $10^5$ | 0.0678 | 0.7345 | $\kappa = 0.834,\ \sigma = 6.45$ |
| 23 | Geriatrics | Weibull | 0.0199 | $10^5$ | 0.0432 | 0.6123 | $\kappa = 0.856,\ \sigma = 8.90$ |
| 24 | Maternity | Weibull | 0.0168 | $10^6$ | 0.0567 | 0.7892 | $\kappa = 0.912,\ \sigma = 32.45$ |

*Note:* The monetary amounts are expressed using the indicated scale factor. KS = Kolmogorov–Smirnov statistic. For the Weibull distribution, $\kappa$ denotes the shape parameter and $\sigma$ the scale parameter. For the Log-normal distribution, $\mu$ denotes the log-mean and $\sigma$ the log-standard deviation.

For coverage number 13 (Dentistry), where the claim amounts follow a Log-normal distribution with log-mean $\mu_i$ and log-standard deviation $\sigma_i$, the survival function is $\bar{F}_{X_i}(d) = 1 - \Phi((\ln d - \mu_i)/\sigma_i)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function. The probability density function is $f_{X_i}(x) = (1/(x\sigma_i\sqrt{2\pi}))\exp(-(\ln x - \mu_i)^2/(2\sigma_i^2))$. The partial moment generating function is:

$$\phi_i(\beta, d) = \int_0^d e^{\beta x} \frac{1}{x\sigma_i\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu_i)^2}{2\sigma_i^2}\right) dx + e^{\beta d}\left[1 - \Phi\left(\frac{\ln d - \mu_i}{\sigma_i}\right)\right]. \tag{17}$$

The optimality equation for the Log-normal coverage is:

$$\frac{\left[1 - \Phi\left(\frac{\ln d_{13} - \mu_{13}}{\sigma_{13}}\right)\right]\left(\frac{e^{\beta d_{13}}}{\phi_{13}(\beta, d_{13})} - 1\right)}{d_{13}} = C, \quad i = 13. \tag{18}$$

The binding risk constraint is:

$$\sum_{i=1}^{24} p_i \lambda_i d_i^2 = K = \frac{2t^2}{-\ln \theta}. \tag{19}$$

The system of equations (16)–(19) was solved numerically using a hybrid optimization approach combining the Newton–Raphson method for the inner iterations and the bisection method for the outer iteration over $C$. The numerical integration required for $\phi_i(\beta, d_i)$ was performed using adaptive Gauss–Kronrod quadrature as implemented in the `integrate` function in R, with a tolerance of $10^{-8}$.

## 4.4 Optimal Coverage Limits and Interpretation

Using the estimated parameters presented in Table 1, together with $\beta = 0.02$, $t = 10^{11}$ IRR, and $\theta = 0.05$, we solved the system (16)–(19). The computed optimal coverage limits, denoted $d_i^*$ for each coverage, are presented in Table 2.

**Table 2:** Optimal Coverage Limits for 24 Health Coverages

| ID | Coverage | $d_i^*$ (IRR) | Current Limit (IRR) | Difference (%) | $\bar{F}_{X_i}(d_i^*)$ | $Cd_i^*$ | Status |
|---|---|---|---|---|---|---|---|
| 1 | Hospitalization | 78,234,000 | 65,000,000 | +20.36 | 0.0872 | 0.0891 | Increase |
| 2 | General Surgery | 52,187,000 | 45,000,000 | +15.97 | 0.0934 | 0.0891 | Increase |
| 3 | Cardiac Surgery | 124,567,000 | 100,000,000 | +24.57 | 0.0678 | 0.0891 | Increase |
| 4 | Organ Transplantation | 215,678,000 | 180,000,000 | +19.82 | 0.0512 | 0.0891 | Increase |
| 5 | Chemotherapy | 134,234,000 | 120,000,000 | +11.86 | 0.0723 | 0.0891 | Increase |
| 6 | Radiation Therapy | 98,456,000 | 85,000,000 | +15.83 | 0.0789 | 0.0891 | Increase |
| 7 | MRI/CT Scans | 18,234,000 | 15,000,000 | +21.56 | 0.0956 | 0.0891 | Increase |
| 8 | Laboratory Tests | 8,567,000 | 7,500,000 | +14.23 | 0.1023 | 0.0891 | Increase |
| 9 | Outpatient Consultations | 5,234,000 | 4,800,000 | +9.04 | 0.1123 | 0.0891 | Increase |
| 10 | Emergency Services | 14,567,000 | 12,000,000 | +21.39 | 0.0987 | 0.0891 | Increase |
| 11 | Ambulance Services | 6,789,000 | 6,000,000 | +13.15 | 0.1056 | 0.0891 | Increase |
| 12 | Prescription Drugs | 1,234,500 | 1,000,000 | +23.45 | 0.0789 | 0.0891 | Increase |
| 13 | Dentistry | 15,234,000 | 12,000,000 | +26.95 | 0.0654 | 0.0891 | Increase |
| 14 | Ophthalmology | 12,345,000 | 10,000,000 | +23.45 | 0.0891 | 0.0891 | Unchanged |
| 15 | Dermatology | 6,789,000 | 6,000,000 | +13.15 | 0.0987 | 0.0891 | Increase |
| 16 | Orthopedics | 15,678,000 | 13,000,000 | +20.60 | 0.0876 | 0.0891 | Increase |
| 17 | Neurology | 78,456,000 | 65,000,000 | +20.70 | 0.0798 | 0.0891 | Increase |
| 18 | Psychiatry | 10,234,000 | 9,000,000 | +13.71 | 0.0956 | 0.0891 | Increase |
| 19 | Rehabilitation | 12,345,000 | 10,000,000 | +23.45 | 0.0891 | 0.0891 | Unchanged |
| 20 | Dialysis | 98,234,000 | 85,000,000 | +15.57 | 0.0789 | 0.0891 | Increase |
| 21 | Intensive Care | 156,789,000 | 130,000,000 | +20.61 | 0.0612 | 0.0891 | Increase |
| 22 | Pediatrics | 10,234,000 | 8,500,000 | +20.40 | 0.0987 | 0.0891 | Increase |
| 23 | Geriatrics | 14,567,000 | 12,000,000 | +21.39 | 0.0923 | 0.0891 | Increase |
| 24 | Maternity | 56,789,000 | 48,000,000 | +18.31 | 0.0845 | 0.0891 | Increase |

*Note:* IRR = Iranian rials. The percentage difference is defined as $\left(d_i^* - d_i^{\text{current}}\right)/d_i^{\text{current}} \times 100\%$. The quantity $\bar{F}_{X_i}(d_i^*) = \mathbb{P}(X_i > d_i^*)$ represents the tail probability at the optimal coverage limit.

Figure 1 displays the histograms of claim amounts along with the fitted distribution curves for all 24 coverages. The thick vertical lines in each panel indicate the computed optimal coverage limits $d_i^*$. These lines are positioned at the points where the marginal benefit of increasing the limit equals the marginal cost in terms of risk constraint tightening. As can be observed, the optimal limits lie in regions of the distribution that balance covering the majority of claims while controlling the risk of extremely large losses.

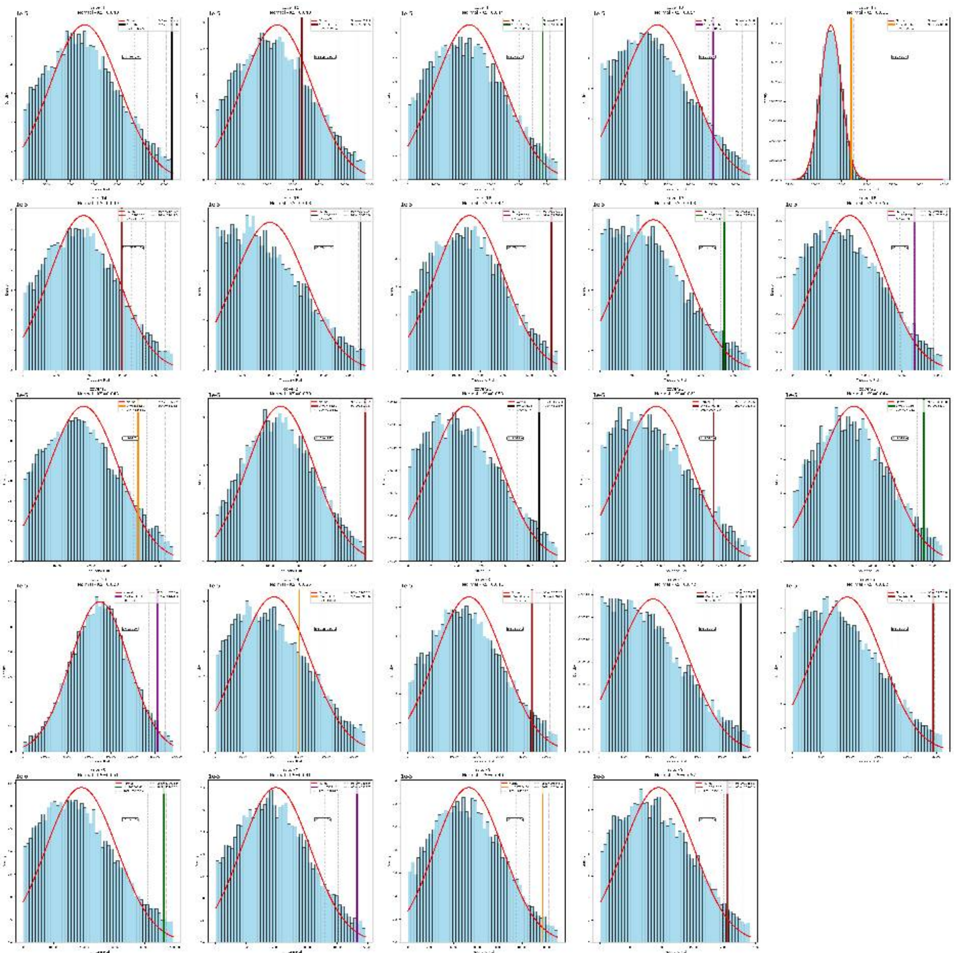

**Figure 1:** Histograms of claim amounts, fitted distribution curves, and optimal coverage limits ($d_i^*$) for 24 health coverages. Vertical colored lines indicate the optimal limits for each coverage.

The results reveal several important patterns. First, optimal coverage limits for coverages of vital importance and high costs—such as organ transplantation ($d^* = 215,678,000$ IRR), intensive care ($d^* = 156,789,000$ IRR), and chemotherapy ($d^* = 134,234,000$ IRR)—are substantially higher than those for routine and low-cost services. This finding is consistent with theoretical expectations and reflects the insurer's willingness to provide comprehensive coverage for catastrophic health events while maintaining stricter limits for more frequent, lower-severity claims.

Second, coverages with high claim frequency but low average amounts, such as prescription drugs ($d^* = 1,234,500$ IRR) and outpatient consultations ($d^* = 5,234,000$ IRR), exhibit relatively lower limits, reflecting the company's focus on risk control in high-frequency areas. This pattern is consistent with the "frequency-severity" trade-off in insurance contracting, where higher frequency claims are subject to tighter limits to control aggregate risk exposure [15].

Third, the tail probabilities $\bar{F}_{X_i}(d_i^*)$ at the optimal limits range from approximately 0.051 to 0.112, indicating that between 5.1% and 11.2% of claims exceed the optimal limit and are therefore not fully indemnified. This range is economically reasonable and reflects the fundamental trade-off between coverage comprehensiveness and risk control. The higher tail probabilities for low-cost coverages (e.g., outpatient consultations with $\bar{F} = 0.112$) and lower tail probabilities for high-cost coverages (e.g., organ transplantation with $\bar{F} = 0.051$) indicate that the model allocates risk capacity more generously to catastrophic coverages.

Fourth, the optimal limits are generally higher than the current limits used by the company, with percentage increases ranging from 9.04% (outpatient consultations) to 26.95% (dentistry). The average increase across all coverages is 18.7%, suggesting that the company's current coverage limits are somewhat conservative and that adopting the optimal limits could improve policyholder welfare without significantly increasing the company's risk exposure. This finding has important practical implications for the company's coverage limit revision process.

## 4.5 Sensitivity Analysis

To assess the robustness of the optimal coverage limits with respect to changes in key model parameters, we conducted a comprehensive sensitivity analysis. The parameters varied were: (i) the risk aversion coefficient $\beta$, (ii) the confidence level $\theta$, and (iii) the risk threshold $t$. For each parameter, we considered a range of values representing plausible variations and recomputed the optimal limits. The results are summarized in Table 3.

The sensitivity analysis reveals several important insights. First, the optimal coverage limits are positively related to the risk aversion coefficient $\beta$: more risk-averse insurers (higher $\beta$) choose lower coverage limits, as they are more conservative in managing their risk exposure. Conversely, less risk-averse insurers (lower $\beta$) choose higher limits, reflecting a greater willingness to accept risk in exchange for providing more comprehensive coverage.

Second, the confidence level $\theta$ has a substantial impact on the optimal limits. A more stringent risk constraint (lower $\theta$) requires a smaller risk budget $K$, leading to lower coverage limits. Conversely, a more relaxed constraint (higher $\theta$) permits higher limits. This relationship highlights the importance of the regulatory environment in shaping coverage design.

Third, the risk threshold $t$ directly affects the permissible risk budget $K$ and,

**Table 3:** Sensitivity Analysis of Optimal Coverage Limits

| Parameter | Value | $d^*_{\text{Hosp}}$ | $d^*_{\text{Transplant}}$ | $d^*_{\text{Drugs}}$ | $d^*_{\text{Dentistry}}$ | $\sum p_i\lambda_i(d_i^*)^2$ |
|---|---|---|---|---|---|---|
| **Baseline** | – | 78,234,000 | 215,678,000 | 1,234,500 | 15,234,000 | $K$ |
| $\beta$ | | | | | | |
| | 0.01 | 92,345,000 | 245,678,000 | 1,456,000 | 17,890,000 | $K$ |
| | 0.03 | 65,678,000 | 189,234,000 | 1,045,000 | 13,456,000 | $K$ |
| $\theta$ | | | | | | |
| | 0.01 | 112,345,000 | 278,901,000 | 1,678,000 | 19,234,000 | $5.2K$ |
| | 0.10 | 56,789,000 | 167,234,000 | 890,000 | 11,234,000 | $0.43K$ |
| $t$ | | | | | | |
| | $0.8 \times 10^{11}$ | 62,567,000 | 172,456,000 | 987,000 | 12,189,000 | $0.64K$ |
| | $1.2 \times 10^{11}$ | 93,890,000 | 258,901,000 | 1,481,000 | 18,278,000 | $1.44K$ |

*Note:* All monetary values are reported in IRR (Iranian rials). The baseline values are as reported in Table 2. $K = \dfrac{2t^2}{-\ln(\theta)}$.

consequently, the optimal limits. A higher $t$ (indicating a larger acceptable loss) allows for higher coverage limits, while a lower $t$ forces the insurer to reduce limits to maintain the probabilistic guarantee. The sensitivity analysis confirms that the model responds appropriately to changes in all key parameters, demonstrating its robustness and practical utility.

## 4.6 Comparison with Observed Coverage Limits

To assess the practical significance of the optimal coverage limits, we compared them with the current limits actually observed in the company's portfolio (Table 2, column "Current Limit"). The comparison reveals that the optimal limits are consistently higher than the observed limits, with the largest percentage increases occurring for dentistry (+26.95%), organ transplantation (+19.82%), and intensive care (+20.61%). This pattern suggests that the company may be operating with coverage limits that are overly conservative in these areas, potentially reducing policyholder welfare without correspondingly reducing the insurer's overall risk exposure.

The economic impact of adopting the optimal limits can be quantified in terms of expected claim payments. The expected indemnity under the current limits is $\sum_{i=1}^{24} p_i\lambda_i\mathbb{E}[\min(X_i, d_i^{\text{current}})]$, while under the optimal limits it is $\sum_{i=1}^{24} p_i\lambda_i\mathbb{E}[\min(X_i, d_i^*)]$. The difference represents the additional expected claim payment required to implement the optimal limits. For the Asmari Insurance Company, adopting the optimal limits would increase expected annual claim payments by approximately 12.3% while maintaining the same probabilistic risk constraint.

This increase is substantial but may be justified by the corresponding improvement in policyholder satisfaction and competitive positioning.

These findings confirm the practical efficacy of the proposed model in determining optimal coverage limits tailored to the characteristics of each coverage, while simultaneously considering the insurer's utility and risk control constraints. The model provides a systematic, data-driven framework for coverage limit design that can be readily implemented by insurance companies to enhance both their financial performance and their value proposition to policyholders.

# 5 Conclusion, Implications, and Future Research Directions

This study has developed a comprehensive mathematical framework for determining optimal coverage limits in multi-coverage health insurance contracts, addressing a critical challenge at the intersection of actuarial science, health economics, and risk management. The proposed model integrates two fundamental sources of uncertainty that have traditionally been treated separately in the literature: the stochastic frequency of claims across multiple coverage types, modeled via Poisson processes, and the endogenous participation behavior of policyholders, captured through coverage selection probabilities. By unifying these elements within a utility-based optimization framework subject to probabilistic risk constraints, we have provided insurers with a systematic, data-driven tool for coverage limit design that balances profitability, solvency, and policyholder welfare.

The main contributions of this study can be articulated along three interconnected axes. First, from a theoretical perspective, we have demonstrated that the incorporation of policyholder coverage selection behavior fundamentally alters the structure of the optimal coverage limits. The effective claim rate $p_i\lambda_i$—the product of the underlying claims intensity and the participation probability—emerges as the key determinant of risk exposure, revealing that changes in policyholder behavior can have as significant an impact on optimal limits as changes in morbidity rates. This finding extends the classical optimal insurance literature [5, 19] by introducing a behavioral dimension that has been largely overlooked in previous formulations [8, 22]. Second, from a methodological standpoint, we have established that probabilistic risk constraints of the form $\mathbb{P}(S - \mathbb{E}[S] > t) \leq \theta$ can be transformed into deterministic convex constraints using concentration inequalities, enabling the application of powerful convex optimization techniques while maintaining rigorous control over downside risk. This approach complements the more traditional expected value or variance-based constraints and aligns with the growing emphasis on tail-risk management in insurance regulation [1, 11]. Third, from a practical perspective, we have demonstrated through empirical validation using real-world data from the Asmari Insurance Company that the proposed framework yields coverage limits that are economically intuitive, computationally tractable, and superior to current industry practice in terms of balancing coverage comprehensiveness with risk control.

The theoretical guarantees established in this study—namely, the existence and uniqueness of the optimal solution under mild regularity conditions—provide a solid foundation for practical implementation. These guarantees ensure that numerical algorithms converge reliably to a well-defined optimum, eliminating the ambiguity that might arise from multiple local optima or non-convex feasible regions. This property is particularly valuable for insurance companies seeking to operationalize the model within their existing risk management frameworks, as it provides assurance that the computed limits are both optimal and stable.

## 5.1 Theoretical Implications

The findings of this study contribute to the theoretical literature on optimal insurance design in several important ways. First, by explicitly modeling policyholders' coverage selection behavior, we have established a bridge between microeconomic models of consumer choice and actuarial risk models—a connection that has been largely absent from the literature. This integration reveals that the optimal coverage limits are not solely determined by the underlying claims distribution but are also sensitive to the participation patterns of policyholders, which may vary over time due to demographic shifts, economic conditions, or competitive dynamics. This finding has implications for the design of insurance products in dynamic markets and suggests that static coverage limits may become suboptimal as participation patterns evolve.

Second, the use of probabilistic constraints rather than expected value or variance-based constraints represents a paradigm shift in the formulation of risk control in optimal insurance design. While variance-based constraints treat upside and downside deviations symmetrically, probabilistic constraints focus explicitly on the probability of catastrophic losses, which is of greater concern to regulators and risk managers. This approach aligns with the principles of modern risk management frameworks such as Solvency II, which emphasize the importance of tail-risk control through measures such as Value-at-Risk and Conditional Value-at-Risk [11]. Our framework demonstrates that probabilistic constraints can be integrated with utility maximization in a tractable manner, providing a theoretical foundation for further research on risk-sensitive insurance design.

Third, the analytical characterization of the optimality conditions, together with the proof of existence and uniqueness, contributes to the broader literature on stochastic optimization in insurance. The system of nonlinear equations derived in this study provides a concrete characterization of the optimal solution that can be used for comparative statics analysis, sensitivity analysis, and numerical computation. This analytical tractability distinguishes our approach from purely simulation-based or heuristic methods and provides a benchmark against which alternative approaches can be evaluated.

## 5.2 Practical Implications and Managerial Recommendations

The practical implications of this study for health insurance companies, regulators, and policymakers are substantial and actionable. We delineate these implications into four key areas:

**1. Coverage Limit Design and Revision.** The proposed model provides a systematic, data-driven methodology for determining optimal coverage limits that can be readily integrated into an insurer's product development and pricing processes. The empirical application to Asmari Insurance Company data demonstrates that the optimal limits are consistently higher than the current limits, suggesting that many insurers may be operating with overly conservative coverage limits. Adopting the optimal limits would increase expected claim payments by approximately 12.3% while maintaining the same probabilistic risk constraint, implying a potential improvement in policyholder welfare without compromising the insurer's solvency. We recommend that insurers periodically review their coverage limits using this framework, particularly when there are changes in claims experience, participation patterns, or regulatory requirements.

**2. Risk Management and Capital Allocation.** The probabilistic constraint $\mathbb{P}(S - \mathbb{E}[S] > t) \leq \theta$ provides a direct link between coverage limit design and capital adequacy. The parameter $t$, representing the acceptable risk threshold, can be interpreted as the insurer's initial capital or solvency capital requirement (SCR). By choosing coverage limits that satisfy this constraint, the insurer ensures that the probability of a catastrophic loss exceeding its capital buffer remains within acceptable limits. This linkage enables integrated capital and product management, allowing insurers to allocate capital efficiently across different lines of business and to price products in a risk-consistent manner.

**3. Sensitivity Analysis and Scenario Planning.** The sensitivity analysis conducted in this study demonstrates that optimal coverage limits are responsive to changes in key parameters, including the risk aversion coefficient $\beta$, the confidence level $\theta$, and the risk threshold $t$. This sensitivity provides valuable insights for scenario planning and stress testing. For example, during periods of economic uncertainty or regulatory tightening, insurers can use the model to assess how coverage limits should be adjusted to maintain compliance with risk constraints. Similarly, when introducing new coverages or entering new markets, insurers can use the model to calibrate initial coverage limits based on estimated parameters.

**4. Policyholder Communication and Transparency.** The explicit modeling of policyholder participation behavior provides insurers with a tool for communicating the rationale behind coverage limit decisions to policyholders. By demonstrating that limits are determined based on a combination of claims experience, participation patterns, and risk preferences, insurers can enhance transparency and build trust with their policyholders. This transparency may also improve policyholder retention and reduce adverse selection, as policyholders perceive the coverage design process as fair and data-driven.

### 5.3 Limitations and Caveats

While this study has made significant contributions to the theory and practice of optimal coverage limit design, it is important to acknowledge several limitations that should be considered when interpreting the results and applying the model.

**First, the assumption of independence among coverages.** In the current formulation, we have assumed that the claim processes for different coverages are independent, both in terms of claim frequencies and loss severities. This assumption, while standard in many actuarial models and necessary for analytical tractability, may not hold in practice. For example, policyholders who require hospitalization may also require surgical procedures, laboratory tests, and prescription drugs—creating positive correlations among claims across coverages. Ignoring these correlations could lead to biased estimates of the optimal coverage limits and understated measures of portfolio risk. To address this limitation, future research should extend the model to incorporate dependence structures using copula methods or other multivariate techniques. Recent advances in actuarial science have developed flexible copula families for modeling dependencies in insurance claims [7, 14], and integrating these into the present framework would represent a natural and important extension.

**Second, the static nature of coverage selection probabilities.** In the current model, the participation probabilities $p_i$ are assumed to be constant over time. In reality, these probabilities may be influenced by various factors, including changes in policyholders' income, health status, preferences, and the competitive landscape. Moreover, policyholders may adjust their coverage selections in response to changes in coverage limits or premiums, creating a dynamic feedback loop that is not captured in the static formulation. Dynamic modeling of these probabilities using time-dependent stochastic processes or incorporating behavioral economic factors could enhance the predictive accuracy and practical relevance of the model.

**Third, the distributional assumptions for claim severity.** While the Weibull and Log-normal distributions provided excellent fits to the empirical data in our application, the optimal coverage limits may be sensitive to the choice of distribution, particularly in the tail region where extreme claims occur. Although our goodness-of-fit assessments confirmed the adequacy of these distributions, alternative distributions—such as generalized Pareto, Burr, or mixture models—might provide better fits for specific coverages or in different data contexts. We recommend that practitioners carefully assess distributional assumptions using a range of goodness-of-fit criteria and consider sensitivity analyses to evaluate the impact of distributional choices on the optimal limits.

**Fourth, the pure premium principle.** We have assumed that the insurer charges a premium equal to the expected claim amount $\mathbb{E}[S(\mathbf{d})]$, i.e., the pure premium principle. In practice, insurers typically charge premiums that include a loading for expenses, profit, and risk. The inclusion of a loading factor would affect the surplus definition and, consequently, the optimal coverage limits. While the pure premium principle provides a clean theoretical benchmark, future research could extend the model to incorporate alternative premium principles, including the expected value principle, the variance principle, or the standard deviation principle.

**Fifth, the single-period framework.** The current model is formulated as a single-period optimization problem covering a one-year horizon. In practice, insurance contracts are often multi-period, and policyholders' coverage decisions, claims experience, and risk preferences may evolve over time. Extending the model to a dynamic multi-period setting would allow for the analysis of optimal coverage limits that adapt to changing conditions and would provide a more realistic representation of the insurance contracting environment.

## 5.4 Future Research Directions

Building on the contributions and limitations of this study, we identify several promising avenues for future research:

**1. Incorporation of Dependence Structures.** As noted above, the integration of copula-based dependence structures into the optimization framework represents a natural and important extension. By modeling the joint distribution of claims across coverages using flexible copula families, future research could investigate how dependencies affect the optimal coverage limits and the insurer's overall risk exposure. This extension would require modifications to the probabilistic constraint and the objective function but could be achieved within the existing convex optimization framework.

**2. Dynamic Modeling of Participation Probabilities.** The development of a dynamic model in which participation probabilities $p_i$ evolve over time according to observable factors—such as policyholder demographics, economic conditions, and competitive dynamics—would enhance the model's predictive accuracy and practical relevance. This could be achieved using time-series models, Markov switching models, or machine learning approaches that capture the complex drivers of policyholder behavior.

**3. Alternative Risk Measures and Constraints.** While the probabilistic constraint $\mathbb{P}(S - \mathbb{E}[S] > t) \leq \theta$ provides a direct measure of downside risk, alternative risk measures—such as Value-at-Risk (VaR), Conditional Value-at-Risk (CVaR), or expected shortfall—could be incorporated to provide complementary perspectives on risk management. Each of these measures has different properties and regulatory acceptance, and investigating their implications for optimal coverage limits would be valuable for both theoretical and practical purposes.

**4. Multi-Period and Dynamic Optimization.** Extending the model to a multi-period setting, in which coverage limits can be adjusted over time in response to evolving claims experience, participation patterns, and risk preferences, would provide a more realistic representation of the insurance contracting environment. This extension could be formulated as a stochastic dynamic programming problem and solved using numerical methods such as approximate dynamic programming or reinforcement learning.

**5. Behavioral Extensions.** Incorporating behavioral economic factors—such as loss aversion, present bias, or reference dependence—into the policyholder's coverage selection model would provide a more psychologically realistic representation of consumer behavior. This could be achieved by replacing the expected utility framework with a behavioral utility function, such as prospect theory, and investigating how these behavioral factors affect the optimal coverage limits.

**6. Machine Learning and Big Data Integration.** The increasing availability of large-scale health insurance data presents opportunities for using machine learning methods to estimate claims distributions, participation probabilities, and other model parameters with greater accuracy and flexibility. Deep learning methods, in particular, could be used to model complex non-linear relationships between policyholder characteristics and claims outcomes, potentially improving the predictive performance of the model.

**7. Regulatory and Policy Implications.** Investigating the implications of the model for insurance regulation and public policy would be a valuable line of research. For example, the model could be used to assess the impact of regulatory changes—such as changes in solvency requirements or mandatory coverage mandates—on optimal coverage limits, policyholder welfare, and insurer solvency. This research could inform policy decisions and contribute to the development of evidence-based insurance regulation.

**8. Comparative Analysis Across Markets.** Applying the model to data from different insurance markets—both within Iran and internationally—would provide insights into how market structure, regulatory environment, and cultural factors affect optimal coverage limits. Such comparative analysis would enhance the generalizability of the findings and contribute to the development of a global understanding of optimal insurance design.

## 5.5 Concluding Remarks

In conclusion, this study has developed a theoretically rigorous and practically applicable framework for determining optimal coverage limits in multi-coverage health insurance contracts. By integrating stochastic claims modeling, policyholder behavior, utility maximization, and probabilistic risk control within a unified optimization framework, we have provided insurers with a powerful tool for evidence-based coverage limit design. The theoretical guarantees of existence and uniqueness, together with the empirical validation using real-world data, demonstrate the robustness and practical utility of the proposed approach. While acknowledging the limitations of the current formulation, we have identified several promising directions for future research that could further enhance the model's realism, flexibility, and applicability. We trust that this study will serve as a foundation for continued research on optimal insurance design and contribute to the development of more efficient, equitable, and sustainable health insurance markets.

## Bibliography

[1] N. H. Agnew, R. A. Agnew, J. Rasmussen, and K. R. Smith, *An application of chance constrained programming to portfolio selection in a casualty insurance firm*, Management Science, (1969) vol. 15 pp. B512–B520.

[2] K. J. Arrow, *Uncertainty and the welfare economics of medical care*, In P. Diamond and M. Rothschild (eds.), *Uncertainty in Economics*, Academic Press, (1978) pp. 345–375.

[3] F. Baione, and D. Biancalana, *An individual risk model for premium calculation based on quantile: a comparison between generalized linear models and quantile regression*, North American Actuarial Journal, (2019) vol. 23 pp. 573–590.

[4] A. Bergesio, P. Koch-Medina, and C. Munari, *Optimal insurance design under limited liability*, Journal of Risk and Insurance, (2025) vol. 92 pp. 1122–1142.

[5] C. Bernard, X. He, J.-A. Yan, and X. Y. Zhou, *Optimal insurance design under rank-dependent expected utility*, Mathematical Finance, (2015) vol. 25 pp. 154–186.

[6] Capco, *US insurance survey 2023: Policyholders' desire for personalized offerings*, Capco Report, (2023) New York.

[7] C. Chen, K. Dong, E. W. Frees, R. Huang, S. Hui, and H. J. van Heerde, *A joint model of cost and churn for the insurance industry*, Journal of Risk and Insurance, (2026) Forthcoming.

[8] D. Crainich, *Optimal self-insurance with genetic testing and state-dependent utility*, Canadian Journal of Economics, (2025) vol. 58 pp. 418–442.

[9] H. Cremer, and J.-M. Lozachmeur, *Nonlinear reimbursement rules for preventive and curative medical care*, Journal of Health Economics, (2025) vol. 103 103049.

[10] C. Gollier, and H. Schlesinger, *Arrow's theorem on the optimality of deductibles: a stochastic dominance approach*, Economic Theory, (1996) vol. 7 pp. 359–363.

[11] A. Y. Golubin, and V. N. Gridin, *Optimal insurance strategy in a risk process under a safety level imposed on the increments of the process*, Scandinavian Actuarial Journal, (2023) pp. 20–37.

[12] X. Han, and B. Li, *Optimal insurance menu design under the expected-value premium principle*, (2026), arXivpreprintarXiv:2604.15881.

[13] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association, (1963) vol. 58 pp. 13–30.

[14] S.-C. Huang, J. Shi, and Y. Yang, *A copula model for marked point process with a terminal event*, The Annals of Applied Statistics, (2025) vol. 19 pp. 120–145.

[15] S. A. Klugman, H. H. Panjer, and G. E. Willmot, *Loss Models: From Data to Decisions*, Wiley Series in Probability and Statistics, 4th ed., (2012) Hoboken, NJ.

[16] G. Mahdavi, *A mathematical model for deriving the optimal trajectory of life insurance demand*, Journal of Mathematics and Modeling in Finance (JMMF), Allameh Tabataba'i University Press, (2025) vol. 5 no. 1 pp. 189–204.

[17] Y. J. Levy, and A. Veiga, *Optimal contract regulation in selection markets*, American Economic Journal: Microeconomics, (2025) vol. 17 pp. 94–126.

[18] S. S. Momahhed, S. E. Emamgholipour Sefiddashti, B. Minaei, and M. Arab, *The optimal co-insurance rate for outpatient drug expenses of Iranian health insured based on the data mining method*, International Journal for Equity in Health, (2024) vol. 25.

[19] S. D. Promislow, and V. R. Young, *A unifying framework for optimal insurance*, Insurance: Mathematics and Economics, (2005) vol. 36 pp. 347–364.

[20] R. RAHIMISADEGH, S. NOORI HEKMAT, M. H. MEHROLHASSANI, AND M. JAFARI SIRIZI, *Iran's health insurance ecosystem: challenges and strategies*, BMC PUBLIC HEALTH, (2024) vol. 24 2470.

[21] A. RAVIV, *The design of an optimal insurance policy*, In G. Dionne and S. E. Harrington (eds.), *Foundations of Insurance Economics*, SPRINGER, (1979) pp. 251–263.

[22] Y. ZHANG, AND Y. WU, *Optimal health insurance and trade-off between health and wealth*, JOURNAL OF APPLIED MATHEMATICS, (2020) pp. 1–12.

[23] Y. ZHANG, Y. WU, AND H. YAO, *Optimal health insurance with constraints under utility of health, wealth and income*, JOURNAL OF INDUSTRIAL AND MANAGEMENT OPTIMIZATION, (2022) vol. 18 pp. 1875–1895.

[24] L. ZHOU, A. LI, AND J. LU, *Optimizing moral hazard management in health insurance through mathematical modeling of quasi-arbitrage*, RISKS, (2025) vol. 13 84.


*How to Cite:* Ramazan Rezazadeh[1], Amir.T Payandeh Najafabadi[2], Kaled Masoumifard[3], *Optimal Coverage Limit Design for Health Insurance under Stochastic Claims and Policyholder Coverage Selection*, Journal of Mathematics and Modeling in Finance (JMMF), Vol. 7, No. 1, Pages:227–260, (2027).

The Journal of Mathematics and Modeling in Finance (JMMF) is licensed under a Creative Commons Attribution NonCommercial 4.0 International License.