When Should Risk Classification be Adopted?

Ruo Jia* Jieyu Lin†

Abstract

Modern technologies have substantially expanded the scope of risk classification, not only by distinguishing among baseline risk types but also by enabling individuals to modify their risk status through endogenous loss-prevention efforts. We develop a theoretical model to assess whether a two-dimensional risk classification technology—that reveals new public information about both baseline risk types and the efficiency of loss-prevention efforts—should be adopted. We show that risk classification is optimal if and only if the new information reveals sufficiently large heterogeneity in loss-prevention efficiencies, thereby improving prevention choices. Moreover, greater dispersion in baseline risk types raises the threshold level of heterogeneity in prevention efficiency required to justify risk classification. We evaluate the adoption of risk classification technologies under two welfare criteria: ex-post Pareto improvement and ex-ante unanimity. When individuals exhibit finite risk aversion, ex-ante unanimity is strictly weaker than Pareto improvement and therefore permits a broader set of risk classification technologies to be adopted. In addition to technology adoption, proper restriction on technologies used can always ensure Pareto improvement, even though adopting technologies without restriction cannot.

Keywords: Risk classification; Symmetric information; Loss prevention; New technology.

JEL Classifications: D81, G22, G52.

*School of Economics, Peking University, email address: ruo.jia@pku.edu.cn

†Faculty of Business, Lingnan University, email address: jieyulin@ln.edu.hk

1 Introduction

Risk classification is vital for the sustainability of insurance (Crocker & Snow, 1986; Bond & Crocker, 1991; Finkelstein et al., 2009). Insurers use risk-type-related information to segment individuals and subsequently to offer them differentiated insurance terms, ensuring the sustainability of their risk portfolios. Accordingly, the welfare implications of risk classification (or banning risk classification) have attracted sustained attention from researchers and policymakers (Handel et al., 2015; Farinha Luz et al., 2023; Nelson, 2025).

Theories of risk classification have been focusing on the asymmetric information environment, where insurers acquire individuals' private information to mitigate adverse selection (e.g., Crocker & Snow, 1986; Hoy & Witt, 2007; Rothschild, 2011). In this setting, new risk classification technologies achieve Pareto improvements by better aligning insurance terms with risk types (Hoy, 1982; Crocker & Snow, 1986).[1]

By contrast, in a symmetric information environment, where new risk-type-related information is publicly revealed, Pareto improvements cannot be achieved since risks revealed as high type suffer losses, though low types gain (Hoy, 1989; Doherty & Posey, 1998). Moreover, this reclassification risk may lead individuals to prefer remaining uninformed about risk types because the expected utility gain from being revealed as low-risk types cannot compensate for the expected utility loss of being revealed as high-risk types (Hoel et al., 2006; Crocker et al., 2025).

Recent technological advances challenge the above conventional wisdom in the symmetric information environment. Modern technologies, such as telematics and genetic testing, generate unprecedented transparency by revealing individual risk types to both insurers and policyholders simultaneously. For example, telematics monitors driving behavior in real time, producing observable risk signals shared by both parties. Genetic testing discloses predispositions to health conditions, improving the accuracy of health risk classification.

Crucially, technological advances have also enhanced individuals' ability to reduce or mitigate their risks through more effective and targeted loss prevention efforts. Wearable health devices

[1] Specifically, Hoy (1982) shows a Pareto improvement under the anticipatory equilibrium of Wilson (1977); and Crocker and Snow (1986) show the condition for Pareto improvement when subsidies across risk types are allowed, which is endorsed by Hoel and Iversen (2002) and Picard (2008). A Pareto improvement can also be achieved when loss uncertainty is thoroughly eliminated by the new information and no insurance is provided (Li, 2021), in which case, new information is revealed to the sole participant – individuals.

and smart home sensors track daily activities, detect behavioral anomalies, and offer personalized recommendations to mitigate health risks (Rashidi et al., 2011; Manogaran et al., 2018). Similarly, smart city infrastructure and autonomous driving systems help avoid accidents by providing real-time traffic updates, road condition alerts, and predictive analytics for safer routing (Li et al., 2019).

Individuals respond to new risk-type-related information with both preventive and risk-taking behaviors. Empirical evidence shows that HIV diagnoses can elicit both risky sex among other more risky behaviors (Ciancio et al., 2025; Kerwin, 2025) and greater investment in prevention, e.g., condom use (Thornton, 2008; Delavande & Kohler, 2012). These evidence, though inconsistent, suggests that new risk information may lead to modify risk status through individual behavioral changes.

Motivated by recent technological advances in risk classification and loss prevention,[2] this paper revisits the symmetric-information scenario and examines the question: ***Which types of technologies that reveal new risk-related information should be adopted for risk classification?***

We evaluate this question using two welfare benchmarks. The first is Pareto improvement: risk classification is desirable if all individuals are ex post better off relative to remaining uninformed (Hoy, 1982; Crocker & Snow, 1986). The second is ex-ante unanimity: risk classification is desirable if all individuals—differing in risk preferences—attain higher ex-ante welfare under a risk classification than under a ban of risk classification (Crocker et al., 2025).

The existing literature concludes that no risk-classification technology can achieve ex-post Pareto improvement, assuming that individuals' loss-prevention efforts have limited effects on loss probabilities in the symmetric-information scenario (Hoy, 1989; Doherty & Posey, 1998; Hoel & Iversen, 2002). When risk types are largely immutable, new information unravels pooling arrangements and imposes utility losses on individuals revealed to be high risk. We depart from this framework by analyzing *two-dimensional risk classification*, where public information reveals both baseline risk types and individuals' efficiencies in loss prevention. When such information enables more effective prevention choices, even individuals classified as high risk can benefit from improved loss-prevention choices. In this case, Pareto improvement becomes feasible. A sufficient and necessary condition is that the revealed heterogeneity in loss-prevention efficiencies be

[2]"Loss prevention" is widely used in the economics literature and is also referred to as self-protection (Ehrlich & Becker, 1972) or primary prevention in health-risk contexts (Bardey & De Donder, 2013).

sufficiently large.

Another strand of the literature studies whether individuals would ex ante support risk classification in a symmetric-information environment, using ex-ante unanimity or ex-ante welfare improvement as the benchmark (Doherty & Posey, 1998; Bardey & De Donder, 2013; Peter et al., 2017).[3] Under this criterion, risk classification is desirable if the ex-ante expected utility gain from being identified as low risk exceeds the expected utility loss from being identified as high risk, so that individuals prefer risk classification to remaining uninformed or banning it altogether. These studies identify two opposing forces: a negative effect from reclassification risk and a positive effect from improved loss-prevention decisions following risk classification. Consequently, the overall ex-ante welfare effect of risk classification is ambiguous. We advance this literature by showing that, under two-dimensional risk classification, the welfare gains from more effective loss prevention dominate the losses from reclassification risk whenever individuals' risk aversion lies within a reasonable range. As a result, ex-ante unanimity in favor of adopting new risk-classification technologies can easily be achieved among most population cohorts.

Comparing the two welfare benchmarks, Pareto improvement is sufficient for—and therefore stricter than—ex-ante unanimity. As defined here and in Hoy (1989), Pareto improvement requires that every individual be (weakly) better off in every realized state after risk classification than in the uninformed state. This criterion can be interpreted as unanimous ex-ante support for risk classification across all risk preference profiles, including individuals with infinite risk aversion. By contrast, ex-ante unanimity requires only that all individuals attain higher expected utility before risk classification. It does not preclude some individuals from being worse off in certain states—such as when revealed to be high risk ex-post—so long as their overall ex-ante expected utility increases. Hence, ex-ante unanimity is easier to satisfy than Pareto improvement: it permits adverse ex-post outcomes for some individuals, provided that expected welfare ex-ante improves for all.

To answer our research question, we develop a theoretical model: A continuum of individuals ex-ante share sharing the same expected loss probability and loss prevention efficiency, defined as the expectation of their underlying "actual" loss probabilities and corresponding "actual" loss

[3] A separate line of research examines individuals' incentives to acquire new private information (see, for example, Hoy & Polborn, 2000; Strohmenger & Wambach, 2000; Hoel et al., 2006).

prevention efficiencies. Competitive insurers offer contracts contingent on newly revealed public information about: (i) individuals' baseline risk types, i.e., loss probabilities conditional on optimal loss prevention effort prior to classification; and (ii) their loss prevention efficiencies, i.e., the rate of return on loss prevention efforts (Courbage et al., 2017). Individuals are then partitioned into N risk classes, with some revealed as higher and others as lower risks relative to the ex-ante uninformed state, and insurers offer actuarially fair, risk-contingent contracts.

Our central result is that Pareto improvement arises if and only if revealed high-risk individuals differ sufficiently in loss prevention efficiency from the pooled benchmark. In other words, welfare gains require substantial new information about prevention productivity. When high-risk individuals are revealed to have high prevention efficiency, they benefit by increasing prevention effort, sharply reducing loss probabilities and insurance premiums. When high-risk individuals are revealed to have low prevention efficiency, they benefit by reducing inefficient effort, despite facing slightly higher insurance premiums. Hence, the interaction between baseline risk types and loss prevention efficiency is crucial for high-risk individuals to gain from classification. Advances in medical treatment provide a natural analogy: when a disease becomes effectively treatable, individuals have stronger incentives to undergo early screening rather than remain uninformed.

Simulations show that a nontrivial set of risk classification technologies satisfies these welfare criteria. Among one million randomly generated two-group risk classification technologies, 22% yield a Pareto improvement (Figure 3) and 67% achieves ex-ante unanimity by raising expected utilities for all individuals relative to the uninformed benchmark, provided their absolute risk aversion does not exceed 0.1—a conservative upper bound consistent with empirical estimates (Cohen & Einav, 2007; Sydnor, 2010; Barseghyan et al., 2011). For risk classification technologies that cannot yield a Pareto improvement, we show that partial risk classification—revealed partial information provided by the risk classification technologies—help to achieve a Pareto improvement.

These results yield clear policy implications. We distinguish between two dimensions of risk-classification technologies: (i) risk-assessment and underwriting tools that reveal baseline risk types, and (ii) loss-prevention technologies that enable individuals to modify risk. When prevention technologies are underdeveloped and risk is largely immutable, expanded classification primarily redistributes surplus and should therefore be restricted. In contrast, when prevention

technologies are sufficiently advanced, risk classification can enhance prevention decisions and improve welfare, and should be encouraged.

We propose two benchmarks for regulatory evaluation. The stricter Pareto improvement criterion requires that no individual be worse off after classification; thus only a limited set of technologies satisfies this standard. The less stringent ex-ante unanimity criterion allows some individuals—such as those revealed to be high risk—to be worse off ex post, provided that they voluntarily accept reclassification risk ex ante because it raises their expected utility. Consequently, a broader set of technologies meets this benchmark. Essentially, the Pareto criterion places greater weight on ex-post fairness, ensuring that no individual bears a realized loss from classification. By contrast, the ex-ante unanimity criterion emphasizes efficiency: it permits ex-post losses for some individuals if doing so expands the set of admissible technologies and increases aggregate expected welfare. Together, these benchmarks provide policymakers with transparent and operational standards for evaluating the adoption of new risk-classification technologies in insurance markets. In addition to adopting available technologies, precise regulation on the amount of information revealed by technologies to be released to market help achieve welfare improvement.

Contribution. This paper makes three contributions.

First, we extend the theory of risk classification by endogenizing loss-prevention effort. Although earlier studies incorporated prevention decisions (Hoy, 1989; Doherty & Posey, 1998), they assumed that loss probabilities were largely immutable—an assumption consistent with the technological environment at the time—and therefore concluded that Pareto improvements are unattainable. We show that when modern technologies enable substantial and dynamic changes in risk status, risk classification under symmetric information can in fact generate Pareto improvements.

Second, we contribute to the literature on reclassification risk and the permissibility of risk classification. Prior work demonstrates that improved prevention incentives may raise ex-ante welfare (Doherty & Posey, 1998; Bardey & De Donder, 2013; Peter et al., 2017), but typically assumes homogeneous individuals with identical risk preferences, leaving the aggregate welfare effect of risk classification ambiguous. We show that, for a broad and empirically plausible range of risk aversion, the welfare gains from enhanced loss prevention behavior dominate the losses

from reclassification risk, thereby establishing conditions under which ex-ante unanimity for risk classification holds.

Third, we advance the literature on technology adoption in insurance markets. Existing studies focus primarily on how technology mitigates asymmetric information (Bardey et al., 2019; Eling et al., 2022; Holzapfel et al., 2024; Gemmo et al., 2025). We demonstrate that even in symmetric-information environments, the joint deployment of risk-assessment and loss-prevention technologies can expand the set of welfare-improving risk-classification regimes.

The remainder of the paper proceeds as follows. Section 2 presents the model. Section 3 characterizes the competitive equilibrium, derives the conditions under which Pareto improvement and ex-ante welfare improvement obtain, and reports the simulation results. Section 4 discusses policy implications. Section 5 examines several extensions. Section 6 concludes.

2 Model Setup

We consider a perfectly competitive insurance market in which a continuum of risk-averse individuals faces a potential loss of a fixed amount D and holds initial wealth W_0. Individuals differ in two dimensions: First is their loss probabilities, which depend on (i) their loss prevention effort s, (ii) the efficiency of their loss prevention technology, and (iii) a set of unobserved risk determinants (Rothschild & Stiglitz, 1976; de Garidel-Thoron, 2005). Second is their risk preference, which is their private information.

Loss prevention effort s is chosen by individuals and assumed to be observable by insurers. In contrast, the efficiency of loss prevention technology and the underlying risk determinants are initially unknown to both individuals and insurers. In this uninformed state, all individuals and insurers perceive a single loss probability function $p_0(s)$, which is strictly decreasing and convex in loss prevention effort (Ehrlich & Becker, 1972):

$$p_0'(s) < 0, \quad p_0''(s) > 0, \quad \text{and} \quad 0 < p_0(s) < 1 \quad \forall s. \tag{1}$$

Given observability of loss prevention effort and fair pricing in a competitive market, insurers offer full insurance coverage $I_0 = D$ at fair premium $m_0(s) = p_0(s)D$ (Ehrlich & Becker, 1972). Thus, each individual receives the same final wealth in both the loss and no-loss states and maxi-

mizes expected utility by making a decision on risk prevention effort ex-ante:

$$\max_s U_0(s) = u(W_0 - p_0(s)D - s), \tag{2}$$

where $u(\cdot)$ is increasing and strictly concave (i.e., $u' > 0$ and $u'' < 0$), and captures individuals' risk preference. Let s_0^* denote the optimal loss prevention effort in the uninformed benchmark.

We now introduce publicly revealed, two-dimensional information regarding individuals' loss probabilities: (i) revealed risk and (ii) efficiency of loss prevention technology. Let θ_i denote the revealed risk, defined as the baseline, static loss probability of type-i individuals when exerting risk prevention effort s_0^*, and γ_i denote their loss prevention efficiency, defined as the rate of return to individuals' loss prevention effort. In practice, θ_i is revealed through discovering some new risk determinants or more precisely assessing the value of some known risk determinants, and γ_i is determined by the loss prevention technologies deployed by individuals. The two-dimensional information sorts individuals into N groups ($N \geq 2$). A fraction λ_i of individuals are revealed to be of type i (Group i), with:

$$p_i(s) = \theta_i \exp\left(-\gamma_i(s - s_0^*)\right), \quad i = 1, 2, ..., N, \tag{3}$$

where $\theta_i = p_i(s_0^*)$ and $\gamma_i = -\frac{p_i'(s)}{p_i(s)}$ (Courbage et al., 2017).[4] The parameters (θ_i, γ_i) thus capture both the static and dynamic components of individual loss probability. Under rational expectation assumption, $p_i(s)$ $(i = 0, 1, ..., N)$ should satisfy that

$$p_0(s) = \sum_{i=1}^{N} \lambda_i p_i(s). \tag{4}$$

We denote θ_0 as the uninformed loss probability and γ_0 as the uninformed loss prevention efficiency given the optimal loss prevention effort s_0^*, that is

$$\theta_0 = p_0(s_0^*) = \sum_{i=1}^{N} \lambda_i \theta_i \text{ and } \gamma_0 = -\frac{p_0'(s_0^*)}{p_0(s_0^*)} = \sum_{i=1}^{N} \frac{\lambda_i \theta_i}{\sum_{i=1}^{N} \lambda_i \theta_i} \gamma_i, \tag{5}$$

[4] We assume a constant rate of return to loss prevention effort, following Peter (2021), which yields the exponential form of $p_i(s)$. Our results (Propositions 1–4) hold under general, non-constant efficiency functions, although the technical conditions become more complex.

where the second equalities for both θ_0 and γ_0 follows Equation (4).

Given full observability of loss prevention efforts and symmetric public information, insurers continue to offer fair-premium, full-coverage contracts to each group. Individuals choose loss prevention efforts to maximize expected utility:

$$\max_s \; U_i(s) = u(W_0 - p_i(s)D - s), \quad i = 1, 2, ..., N. \tag{6}$$

This framework allows us to evaluate whether the introduction of two-dimensional risk classification—combining revealed risk and loss prevention efficiency—can yield a Pareto improvement and ex-ante welfare improvement, even under a symmetric information environment. Specifically, a Pareto improvement requires that after risk classification, some individuals achieve a higher utility, compared to the uniformed state, and no individual has a lower utility. That is, a two-dimensional risk classification leads to a Pareto improvement if and only if

$$U_i(s_i^*) \geq U_0(s_0^*), \; \forall i \in \{1, 2, ..., N\}, \text{ with at least one strict inequality.} \tag{7}$$

An ex-ante welfare improvement requires that individuals expect to be better off when moving from the uninformed state without information and corresponding risk classification to a regime with the two-dimensional information accompanied by risk classification. That is, a two-dimensional risk classification leads to an ex-ante welfare improvement if and only if

$$\sum_{i=1}^{N} \lambda_i U_i(s_i^*) > U_0(s_0^*). \tag{8}$$

An ex-ante welfare improvement implies that individuals with different risk preference (i.e., different functional form of $u(\cdot)$) have unanimous agreement on approving risk classification when facing it.

3 Solution and Propositions

3.1 Pareto Improvement

The necessary and sufficient condition for a Pareto improvement is that those groups with revealed risk higher than the pooled risk in the uninformed state should have sufficiently different

(either low or high) loss prevention efficiency compared to the loss prevention efficiency when uninformed. This condition ensures that those revealed higher-risk group—despite facing higher insurance premiums before adjusting their loss prevention efforts—can offset this potential utility loss through improved decision-making in loss prevention efforts. In contrast, the revealed lower-risk group unambiguously benefits from both lower premiums and better loss prevention.

We distinguish between two notions of risk. The *revealed risk* θ_i reflects individuals' loss probabilities at the optimal effort level of the uninformed state s_0^*. The *risk at optimum* is based on updated loss probabilities evaluated at the new optimal loss prevention effort level s_i^*. That is, Group i is higher-risk at optimum than Group j if $p_i(s_i^*) > p_j(s_j^*)$, for $i \neq j$.

Proposition 1. *A Pareto improvement occurs following a risk classification if and only if the loss prevention efficiency of revealed higher-risk groups with $\theta_i > \theta_0$ is sufficiently different from that in the uninformed state (i.e., γ_0). That is, for any group $i \in \{1,2,...,N\}$ with $\theta_i > \theta_0$, there exist thresholds $\underline{\gamma}(\theta_i) < \gamma_0 < \overline{\gamma}(\theta_i)$ such that*

$$\gamma_i \leq \underline{\gamma}(\theta_i) \quad or \quad \gamma_i \geq \overline{\gamma}(\theta_i). \tag{9}$$

Proof: *See Appendix A.*

Intuition. Two effects shape individuals' post-classification wealth. First, risk-related information revelation leads higher-risk individuals to face higher premiums and thus lower final wealth when their loss prevention efforts have not yet changed from the uninformed optimal level s_0^* (see Handel et al., 2015). Second, information on the efficiency of loss prevention technology (i.e., γ_i) enables all individuals to optimize effort more effectively, increasing their final wealth. The net effect of the aforementioned two effects is positive for revealed lower-risk groups (i.e., $\theta_i \leq \theta_0$), who benefit on both margins. For those groups with higher revealed risk (i.e., $\theta_i > \theta_0$), however, final wealth improves if and only if the gain from optimized prevention outweighs the cost of premium increase. This occurs when their loss prevention efficiency is either sufficiently high (allowing large risk-premium reductions at a small increase of loss prevention cost) or sufficiently low (allowing large effort reductions with a small risk-premium increase) compared to the uninformed efficiency.

Proposition 2. *Groups with higher revealed risk require a correspondingly greater difference in efficiency of loss prevention technology from that in the uninformed state to achieve a Pareto improvement. Specifically, for any* $\theta_i > \theta_0$,

$$\begin{cases} \lim_{\theta_i \to \theta_0} \underline{\gamma}(\theta_i) = \gamma_0 \quad \textit{and} \quad \frac{d\underline{\gamma}(\theta_i)}{d\theta_i} < 0, \\ \lim_{\theta_i \to \theta_0} \overline{\gamma}(\theta_i) = \gamma_0 \quad \textit{and} \quad \frac{d\overline{\gamma}(\theta_i)}{d\theta_i} > 0. \end{cases} \tag{10}$$

Proof: *See Appendix B.*

Intuition. For any Group $i \in \{1, 2, ..., N\}$ with $\theta_i > \theta_0$, when the difference between θ_i and θ_0 increases, a higher premium and hence lower utility are faced by Group i, when they have not changed their loss prevention efforts from the uninformed optimal level s_0^*. To compensate, a larger benefit from prevention effort adjustment is required, necessitating a greater divergence in loss prevention efficiency across the groups.

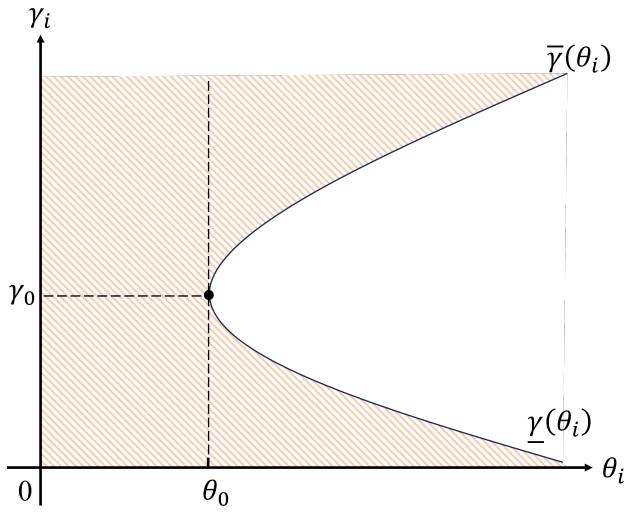


Figure 1: Pareto Improvement Domain for (θ_i, γ_i)

Figure 1 illustrates Propositions 1 and 2. A Pareto improvement requires that, for any Group $i \in \{1, 2, ..., N\}$, their risk, captured by the pair (θ_i, γ_i) lies within the shaded region in Figure 1, where (θ_0, γ_0) represented their weighted mean.[5] For groups with $\theta_i \leq \theta_0$, they benefit from risk classification independent of their loss prevention efficiency γ_i. In contrast, groups with $\theta_i > \theta_0$ benefit from risk classification if and only if their loss prevention efficiency is either higher than $\overline{\gamma}(\theta_i)$, which increases in θ_i, or lower than $\underline{\gamma}(\theta_i)$, which decreases in θ_i.

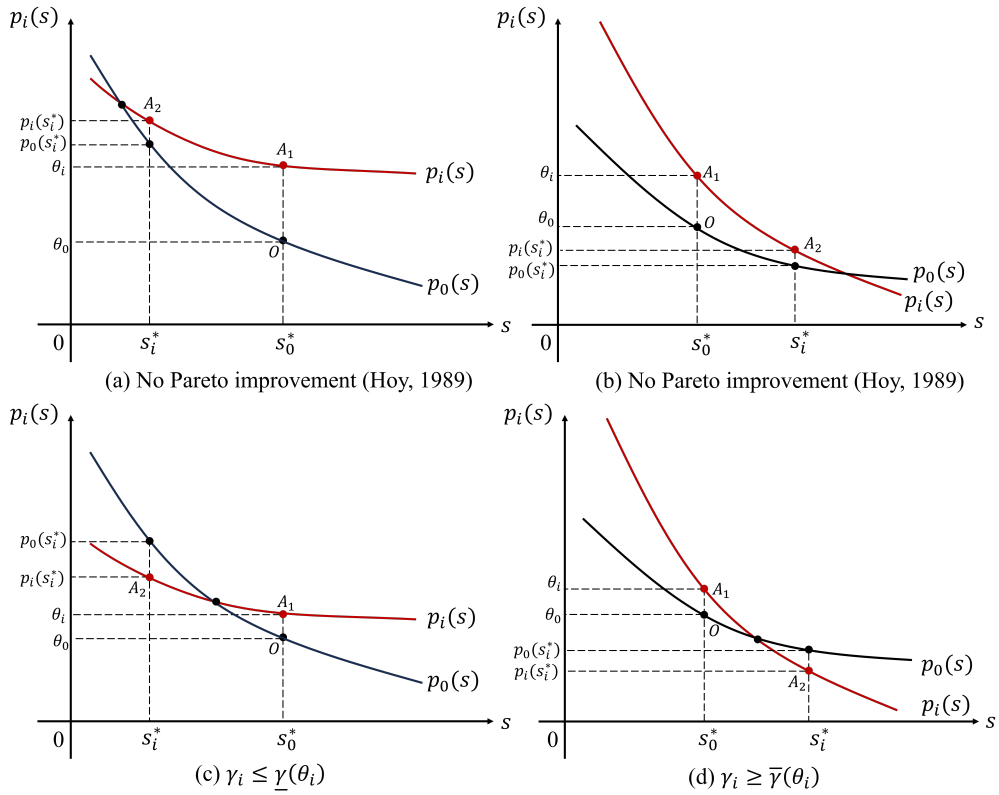
[5] In Section 3.3, we show a sizable set of risk classification can achieve Pareto improvement by simulation.

Proposition 3. *Suppose a Pareto improvement is achieved. Then, for any Group $i \in \{1,2,...,N\}$ with revealed higher risk $\theta_i > \theta_0$, their post-classification loss probability becomes lower than the uninformed loss probability; that is*

$$p_i(s_0^*) > p_0(s_0^*) \quad \text{and} \quad p_i(s_i^*) < p_0(s_i^*). \tag{11}$$

Proof: *See Appendix C.*

Figure 2 illustrates Proposition 3. In Figures 2(a) and (b), if Group i faces a higher loss probability than in the uninformed state for any level of loss prevention effort between s_0^* and s_i^*—that is, if the order between $p_i(s)$ and $p_0(s)$ remains unchanged for all s between s_0^* and s_i^*—then Group i is worse off at any loss prevention effort within this interval. Consequently, they can never achieve a higher utility than in the optimal uninformed state (Hoy, 1989).[6]



(a) No Pareto improvement (Hoy, 1989)

(b) No Pareto improvement (Hoy, 1989)

(c) $\gamma_i \leq \underline{\gamma}(\theta_i)$

(d) $\gamma_i \geq \overline{\gamma}(\theta_i)$

Figure 2: How Pareto Improvement is Achieved

Figures 2(b) and 2(c) illustrate the two possible scenarios in which Group i with $\theta_i > \theta_0$ attains

[6]This result is derived from $U_i(s_i^*) < U_0(s_i^*) \leq U_0(s_0^*)$, where the first inequality follows the order between $p_i(s)$ and $p_0(s)$ and the second follows that s_0^* is optimal uninformed loss prevention efforts. Thus, this result holds regardless of whether $s_0^* \geq s_i^*$ (as illustrated in Figure 2(a)) or $s_0^* < s_i^*$ (as illustrated in Figure 2(b)). Intuitively, it also holds under the stronger assumption that $p_i(s) > p_0(s)$ for all s, as assumed in Hoy (1989).

higher final wealth following risk classification, i.e., how the Pareto improvement is achieved. In Figure 2(b), Group i exhibits lower loss prevention efficiency, as indicated by a flatter $p_i(s)$ curve. Consequently, these individuals optimally reduce their loss prevention effort from s_0^* to s_i^*, resulting in a modest increase in loss probability (from A_1 to A_2), and achieve a lower loss probability than would be possible in the uninformed state (i.e., $p_i(s_i^*) < p_0(s_i^*)$). Figure 2(c) presents the opposite case: Group i's higher loss prevention efficiency leads to a substantial reduction in risk through increased effort (from A_1 to A_2), thereby achieving lower loss probabilities than that ex-ante, i.e., $p_i(s_i^*) < p_0(s_i^*)$.

Proposition 3 shows a reversal in the order of loss probabilities–by adjusting loss prevention efforts–between revealed high-risk groups and the uninformed state is critical to achieve the Pareto improvement. Although it might be counter-intuitive, our analysis in Section 4 demonstrates that, when there are three or more underlying risk types of individuals (even if they always follow the same order of loss probabilities given any loss prevention efforts), it is always possible to classify individuals into groups based on partial risk-type information to achieve both the reversed order of loss probabilities and thus a Pareto improvement, even though full information revelation fails to yield a Pareto improvement.[7]

3.2 Ex-ante Unanimity

In this section, individuals' preferences are described by constant absolute risk aversion (CARA) utility functions:

$$u(x) = \frac{1 - \exp(-\tau_A x)}{\tau_A}, \tag{12}$$

where the risk aversion level τ_A differs among individuals.[8] Specifically, we have the following two propositions.

Proposition 4. *There exists a risk aversion threshold $\tau_{Amax} > 0$ such that an ex-ante welfare improvement occurs for all individuals with risk aversion level τ_A lower than τ_{Amax}.*

[7] Specifically, when there are three underlying risk types of individuals that follow the same order of loss probabilities given any loss prevention efforts, revealing their exact risk types (i.e., full information revelation) cannot achieve a Pareto improvement (Hoy, 1989). Section 4 shows that it is possible to achieve a Pareto improvement by dividing individuals into two groups — revealing group membership rather than their exact risk types to individuals (i.e., partial information revelation).

[8] In Appendix G, we show that Propositions 4 and 5 hold with constant relative risk aversion (CRRA) utility functions.

Proof: *See Appendix D.*

Intuition. Risk classification has two opposite effects on individuals' ex-ante welfare. First is the reclassification risk; that is, when facing risk classification, individuals expect that they may be revealed as high risk (i.e., $\theta_i > \theta_0$) or low risk (i.e., $\theta_i < \theta_0$). The reclassification risk decreases individuals' ex-ante welfare because in the uninformed state, the risk can be transferred to insurers through the pooling premium, but after risk classification, the risk has to be born by individuals themselves through the type-dependent premium (Hoy, 1989; Doherty & Posey, 1998). The negative effect of the reclassification risk is greater if individuals are more risk-averse. Second is the improved decisions on loss prevention efforts. When facing risk classification, individuals expect to improve their decisions on loss prevention efforts after receiving information on loss prevention efficiency, which improves their ex-ante welfare. Thus, if individuals are not too risk-averse (i.e., $\tau_A < \tau_{Amax}$), the positive effect from the improved decisions on loss prevention efforts dominates the reclassification risk effect, leading to a higher ex-ante welfare for individuals when facing risk classification. That said, given any risk classification technology, the necessary and sufficient condition for an ex-ante unanimity on risk classification among individuals with different risk preferences is that no individual is excessively risk-averse.

Proposition 5. *A Pareto improvement is equivalent to an ex-ante welfare improvement for an individual, no matter how risk-averse the individual is (i.e.,* $\forall \tau_A \in (0, \infty)$*).*

Proof: *See Appendix E.*

Intuition. For an extremely risk-averse individual, a slight reduction in final, realized wealth in any revealed state compared to the uninformed state would result in large utility loss, leading to lower ex-ante welfare under reclassification risk compared to the uninformed state. Thus, if an ex-ante welfare improvement occurs for an individual with extremely high risk aversion $\tau_A \to \infty$, it implies that such outcomes (i.e., any slight reduction in final wealth after risk classification) are prevented. In other words, among all possible revealed states following risk classification, the individual obtains higher final wealth—and thus higher utility—in at least one state, and never experiences lower final wealth—and thus lower utility—in any state, i.e., a Pareto improvement.

Propositions 4 and 5 show the relationship between the two benchmarks, Pareto improvement

and ex-ante welfare improvement. Pareto improvement requires that individuals are never worse off after risk classification, no matter which risk groups individuals are revealed to be in. This is a stricter criteria than ex-ante welfare improvement since the latter allows individuals to be worse off in some groups after risk classification. When individuals are not too risk-averse, the possibility of lower final wealth (and thus lower utility) in some groups after risk classification is acceptable from an ex-ante perspective.

In the following section, by simulation, we show that (i) there is a sizable probability of Pareto improvement following random risk classification and (ii) that the maximum risk aversion is a reasonably loose requirement for an ex-ante unanimity on risk classification among individuals with risk preference heterogeneity.

3.3 Risk Classification Technologies that Achieve Welfare Improvement: A Simulation

We conduct a large-scale simulation to quantify the prevalence of welfare-improving risk-classification technologies. We fix initial wealth at $W_0 = 100$, potential loss at $D = 50$, the pooled loss probability in the uninformed state at $\theta_0 = 0.2$, optimal loss prevention effort in the uninformed state at $s_0^* = 10$, and pooled prevention efficiency in the uninformed state at $\gamma_0 = 0.1$.[9]

For each of the one million feasible risk-classification technologies $\{(\theta_i, \gamma_i, \lambda_i)\}_{i=1}^N$, revealed risks θ_i are drawn from a uniform distribution on $(0, 0.5)$, and prevention efficiencies γ_i from a lognormal distribution with $\ln \gamma_i \sim N(-3.5, 1)$, implying a mean prevention efficiency $\gamma_0 = 0.1$ and coefficient of variation equal to one. A risk classification technology is feasible if there exist population weights $\{\lambda_i\}_{i=1}^N$, with $\lambda_i \in (0,1)$ and $\sum_i \lambda_i = 1$, such that average baseline risk type and average prevention efficiency equal θ_0 and γ_0, respectively (see Equation (5)).

Figure 3 reports the fraction of risk classification technologies that satisfy ex-ante unanimity under a spectrum of risk aversion upper bounds $\tau_{A\max}$. When risks are classified into two groups (i.e., $N = 2$), 67% of technologies achieve ex-ante unanimity if absolute risk aversion does not

[9] From the first-order condition for optimal prevention in the uninformed benchmark (Equation (2)), the parameters satisfy $\theta_0 \gamma_0 D = 1$.

exceed 0.1 (Figure 3(a)), which exceeds most empirical estimates.[10] If the maximum absolute risk aversion is 0.05 (0.2), the fraction increases to 78% (decreases to 42%). As the upper bound of absolute risk aversion approaches infinity, the fraction of achieving ex-ante unanimity converges to 22%, which, by Proposition 5, coincides with the share of risk classification technologies that achieve Pareto improvement (i.e., the red dotted line in Figure 3(a)).

Figures 3(b) and 3(c) show that increasing the number of risk classes (i.e., $N = 3$ and $N = 5$) raises the likelihood of ex-ante unanimity for empirically plausible levels of risk aversion, although Pareto improvement becomes less frequent. Panels (d)–(f) report analogous results with constant relative risk aversion (CRRA) utility function, $u(x) = \frac{x^{1-\tau_R}}{1-\tau_R}$, $\tau_R \in (0,1) \cup (1,\infty)$.[11]

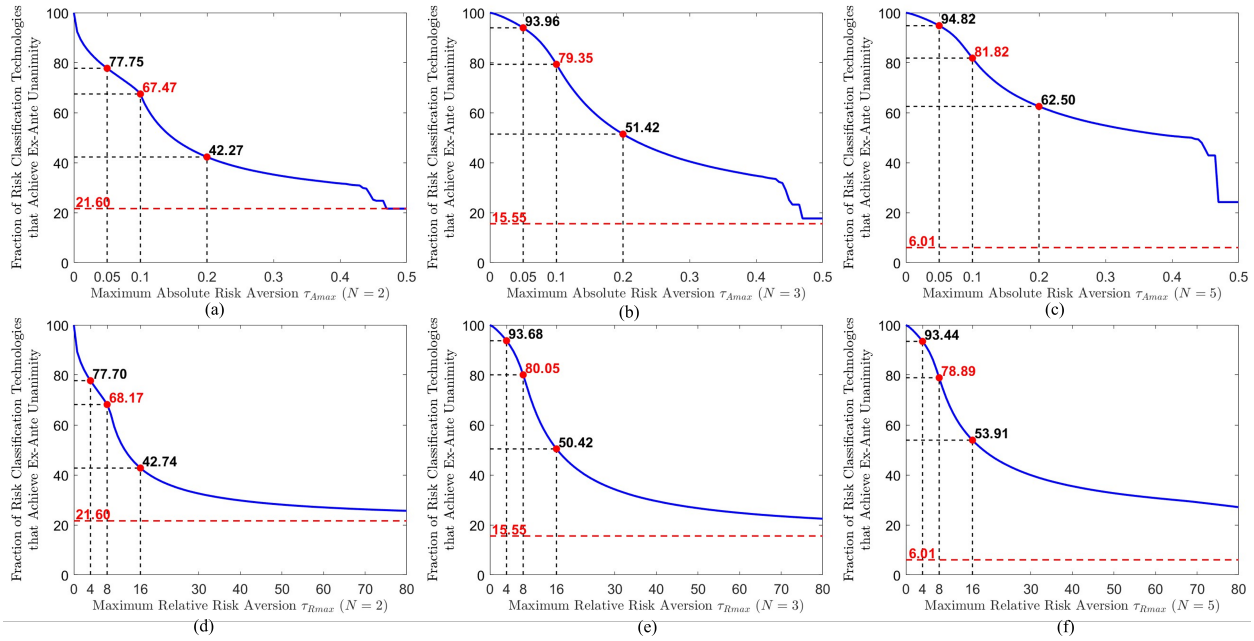


Figure 3: Fraction of 1,000,000 Random Risk Classification Technologies that Achieve Ex-Ante Unanimity

[10] An upper bound of absolute risk aversion of 0.1 exceeds most empirical estimates. For example, Cohen and Einav (2007) report a mean absolute risk aversion of 0.0067 and a 95% quantile of 0.0099; Sydnor (2010) estimate absolute risk aversion bounds between 0.002 and 0.0053; and Barseghyan et al. (2011) report estimates between 0.002 and 0.0083. See Table G1 for a summary.

[11] Because relative risk aversion equals absolute risk aversion multiplied by final wealth (Cohen & Einav, 2007), and final wealth in the uninformed benchmark is $W_0 - \theta_0 D - s_0^* = 80$, the magnitudes of our results are consistent: $\tau_{R,\max} \approx 80\tau_{Amax}$.

4 Partial Risk Classification

In this section, we further demonstrate that even if a risk classification technology cannot achieve Pareto improvement, there always exist proper partial risk classifications—individuals are classified based on partial information from the risk classification technology about their revealed risks and loss prevention efficiency—that achieve Pareto improvement.[12]

Specifically, suppose a risk classification technology if applied would divide individuals into three groups, each characterized by a baseline risk and loss prevention efficiency: (θ_H, γ_H), (θ_M, γ_M), and (θ_L, γ_L), with population shares λ_H, λ_M, and λ_L, respectively, where $\lambda_k > 0$ for $k = H, M, L$ and $\lambda_H + \lambda_M + \lambda_L = 1$.[13,14] In the uninformed state, individuals lack information about their specific revealed risk and loss prevention efficiency. Consistent with Equation (4), their loss probability function is given by

$$p_0(s) = \lambda_H p_H(s) + \lambda_M p_M(s) + \lambda_L p_L(s). \tag{13}$$

Without loss of generality, we assume

$$\gamma_H \neq \gamma_M \neq \gamma_L \quad \text{and} \quad p_H(s) > p_M(s) > p_L(s) \text{ for } \forall s \in [s_{min}, s_{max}], \tag{14}$$

where the interval $[s_{min}, s_{max}]$ ensures that all reasonable loss prevention efforts of individuals lie within this range, regardless of their knowledge of their risks.[15] Following Proposition 3, as a result of ordered loss probabilities, this risk classification technology cannot achieve Pareto improvement (Hoy, 1989).

Suppose partial information classifies a fraction $q \in (0,1)$ of individuals into Group 1 and the remaining $1-q$ into Group 2.[16] Individuals in Group i ($i = 1, 2$) learn that their probability of being

[12] Following Propositions 4 and 5, an ex-ante welfare improvement is always achievable by any partial risk classification, depending on an upper bound on individuals' risk aversion level.

[13] The result of partial risk classifications (as shown in Proposition 6) holds if there are more than three groups that would be divided by a risk classification technology if applied, but does not if there are only two groups.

[14] Some risk classification studies also consider three types of individuals in the market but with two underlying risk types—high-risk, low-risk, and uniformed type who do not know their risk types (see e.g., Doherty & Thistle, 1996; Dionne & Rothschild, 2014). Different from this paper, these studies analyze the conditions that the uninformed individuals are willing to learn their risk types.

[15] The technical expression of $[s_{min}, s_{max}]$ is provided in Appendix F.

[16] Our results hold when individuals are classified into more than two groups. Specifically, following the similar conditions shown in Proposition 6, further risk classification that subdivides either Group 1 or 2 into two additional groups also achieves a Pareto improvement.

type k is $\lambda_{ki} > 0$ ($k = H, M, L$). Consequently, Group i's revealed risk is $\theta_i = \lambda_{Hi}\theta_H + \lambda_{Mi}\theta_M + \lambda_{Li}\theta_L$, where $\lambda_{Hi} + \lambda_{Mi} + \lambda_{Li} = 1$ and $q\lambda_{k1} + (1-q)\lambda_{k2} = \lambda_k$ for $k = H, M, L$. Without loss of generality, we consider Group 1 as the revealed high-risk group, i.e., $\theta_1 > \theta_0 > \theta_2$.[17]

As analyzed following Proposition 1, the revealed low-risk group (i.e., Group 2) always benefits from the risk classification. Thus, a Pareto improvement hinges on the risk composition of Group 1, defined by $(\lambda_{H1}, \lambda_{M1}, \lambda_{L1})$, where $\lambda_{L1} = 1 - \lambda_{H1} - \lambda_{M1}$. As shown in the Figure 4, a risk classification is feasible (not necessarily achieves a Pareto improvement) if and only if $(\lambda_{H1}, \lambda_{M1})$ falls within the lightly shaded area.[18]

Proposition 6. *Among all feasible partial risk classifications, there always exists some that achieve a Pareto improvement; that is*

$$\exists(\lambda_{H1}, \lambda_{M1}) \in A, \quad U_1^*(\lambda_{H1}, \lambda_{M1}) \geq U_0^*, \tag{15}$$

except when $(\lambda_{H1}, \lambda_{M1}) = (\lambda_H, \lambda_M)$*, where A denote the set of all feasible* $(\lambda_{H1}, \lambda_{M1})$*, and* $U_1^*(\lambda_{H1}, \lambda_{M1})$ *denotes the optimal expected utility of Group* 1.

Moreover, a feasible partial risk classification achieves a Pareto improvement if $(\lambda_{H1}, \lambda_{M1})$ *satisfies*

$$\lambda_{H1} \leq f(\lambda_{M1}), \tag{16}$$

where the function $f(\lambda_{M1})$ *is the iso-utility line (i.e.,* $\forall\lambda_{M1}$*,* $U_1^*(f(\lambda_{M1}), \lambda_{M1}) = U_0^*$*) and satisfies* $\forall\lambda_{M1} \neq \lambda_M$*,* $\theta_1 = f(\lambda_{M1})\theta_H + \lambda_{M1}\theta_M + (1 - f(\lambda_{M1}) - \lambda_{M1})\theta_L > \theta_0$.

Proof: *See Appendix F.*

[17] It is possible that individuals are classified into two groups with identical revealed risks, i.e., $\theta_1 = \theta_0 = \theta_2$. In this case, a Pareto improvement is always achieved as long as the two groups have different composition, i.e., $\exists k = H, M$ or L, such that $\lambda_{k1} \neq \lambda_{k2}$.

[18] Specifically, a risk classification is feasible if $(\lambda_{H1}, \lambda_{M1})$ satisfies $\lambda_{H1}, \lambda_{M1} \geq 0$, $\lambda_{H1} + \lambda_{M1} \leq 1$, and $\lambda_{H1}\theta_H + \lambda_{M1}\theta_M + (1 - \lambda_{H1} - \lambda_{M1})\theta_L \geq \theta_0$.

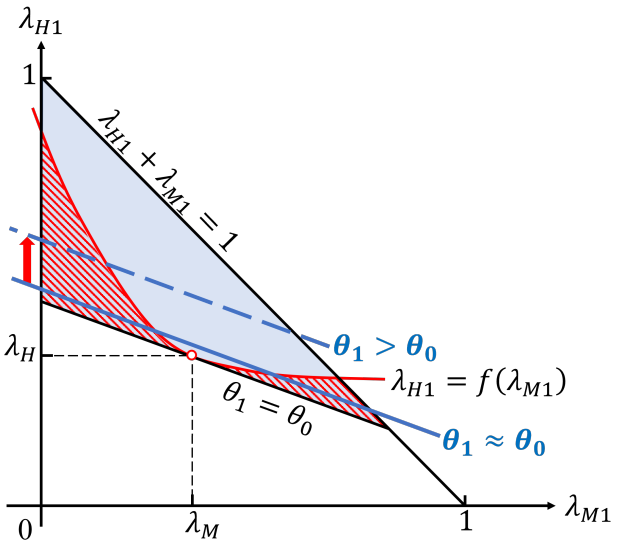


Figure 4: When Pareto Improvement is Achieved by Partial Risk Classification

Proposition 6 shows that (i) a Pareto improvement is always achievable with proper ways of partial information revelation, and (ii) it provides the conditions for a Pareto improvement. These results are illustrated in Figure 4. Among all feasible partial risk classifications, represented by all pairs of $(\lambda_{H1}, \lambda_{M1})$ within the lightly shaded area of Figure 4, those lying in the hatched area achieve a Pareto improvement.

Group 1 benefits from risk classification (i.e., a Pareto improvement is achieved) if their loss prevention efficiency is sufficiently different from the uninformed state, which requires their risk composition to be sufficiently different from the uninformed state. Specifically, when θ_1 is close to θ_0 (represented by the blue solid line $\theta_1 \approx \theta_0$ in the figure), a Pareto improvement is achieved when $(\lambda_{H1}, \lambda_{M1})$ lies at the two ends of the blue solid line and away from the risk composition of the uninformed state (i.e, (λ_H, λ_M)). As the revealed risk of Group 1 increases (i.e., illustrated by the move from the solid blue line $\theta_1 \approx \theta_0$ to the dashed blue line $\theta_1 > \theta_1$), consistent with Proposition 2, stricter conditions are necessary for a Pareto improvement, requiring a greater difference in the risk compositions between Group 1 and the uninformed state.

Our analysis of risk classification is closely related to the literature on information design. Economists are interested in analyzing the optimal information structure in markets with uncertainty (Eső & Szentes, 2007;Kamenica & Gentzkow, 2011; Bergemann & Morris, 2013; Bergemann et al., 2015) and in particular the insurance market (Garcia & Tsur, 2021; Häusle & Roth-

schild, 2024). Garcia and Tsur (2021) study the information about individuals' loss size (i.e., medical expense), while Häusle and Rothschild (2024) study the information about individuals' loss probability, where individuals' loss prevention efforts do not play a role. Our analysis achieves additional milestone by showing that some information structure about individuals' loss probability achieves a Pareto improvement if considering that loss prevention effort can significantly change risks, i.e., higher utility for all individuals regardless of their ex-post revealed state, or equivalently, regardless of individuals' risk preference from an ex-ante perspective.

5 Policy Implications

Our results imply that the welfare consequences of risk classification depend critically on the state of loss-prevention technology. When prevention tools can meaningfully alter risk status after baseline risk types are revealed, classification strengthens incentives for effective action and can thus increase utility even for individuals identified as high risk. Early detection of treatable diseases is a natural example: screening enables timely intervention and may generate Pareto improvements when treatment sufficiently mitigates risk. By contrast, when prevention or treatment is largely ineffective (e.g., for certain terminal illnesses), diagnosis primarily redistributes surplus and may reduce utility for those revealed as high risk. In such environments, restrictions on risk classification are more likely to be justified.

Even when Pareto improvement is unattainable, a risk classification technology may still win ex-ante unanimity. Individuals may experience lower utility ex post when identified as high risk, yet support risk classification ex ante because early intervention mitigates worse outcomes in later states. That is, treatment following early detection is beneficial, though not sufficient to raise ex-post utility above the uninformed benchmark. Individuals are willing to bear reclassification risk provided that their risk aversion is not excessively high. This perspective is consistent with evidence from health economics—for example, the expansion of effective HIV treatment increased individuals' willingness to undergo HIV testing (Kellerman et al., 2002).

Our framework provides policymakers with two operational criteria for evaluating risk-classification technologies. The Pareto improvement criterion requires that no individual be worse off ex post and at least some be strictly better off. This standard places strong weight on realized fairness and therefore approves only a limited set of technologies. The ex-ante unanimity criterion requires that

all individuals attain higher expected utility before risks are partitioned, even if some are worse off ex post. This standard emphasizes efficiency and permits a broader set of technologies, provided that risk aversion is finite. Consistent with this distinction, our simulations show that fewer than one quarter of randomly generated risk classification technologies achieve Pareto improvement, whereas more than two thirds reach ex-ante unanimity over a wide and empirically plausible range of risk aversion.

Under either benchmark, adopting risk classification increases social welfare. Pareto improvement ensures that aggregate ex-post utilities increase. Ex-ante unanimity implies that the sum of expected utilities—equivalently, the weighted sum of ex-post utilities across revealed states—rises, given a certain risk aversion level. Because ex-ante unanimity approves a larger set of technologies than Pareto improvement, it can generate higher aggregate welfare when many candidate technologies are available and regulators must decide on each of them. In this sense, ex-ante unanimity is the more efficiency-oriented criterion. However, it permits some individuals—typically those revealed as high risks—to be worse off ex post. Pareto improvement, by contrast, prioritizes realized fairness by protecting every individual in every state, at the cost of approving fewer welfare-enhancing risk classification technologies.

In addition to the decision on whether to approve a risk classification or not, precise regulation on risk classification technologies is also important. When a risk classification technology is available and policymakers can design the amount of information released to both individuals and insurers, Pareto improvement is always achievable, through either approving the technology or proper partial information revelation. If risk classification technologies that reveal full information but would not achieve Pareto improvement if applied, regulation that restricts the use of the technologies—especially the use of information revealed by the technologies—matters. Through proper regulation, policymakers can achieve partial risk classification and hence Pareto improvement.

Policy debates over the use of genetic and other sensitive risk information often focus on fairness concerns. Our analysis clarifies that the appropriate regulatory stance depends on whether complementary loss-prevention technologies are sufficiently developed. When prevention is limited, expanded classification should be constrained. When prevention is effective and enables

meaningful risk reduction, classification can enhance social welfare without compromising fairness. Accordingly, policymakers should evaluate not only advances in risk assessment, but also the maturity of loss-prevention technologies. As prevention technologies improve, restrictions on risk classification may warrant reassessment.

6 Extensions

6.1 A scenario of no insurance

Advancements in risk classification and loss prevention technologies are usually not invented or first applied by the insurance industry. Therefore, in this section, we show that risk classification achieves Pareto improvement not only when insurance is present but also in the absence of insurance. Conventionally, risk classification is mainly used by insurers for underwriting and pricing purposes. Modern risk identification/assessment technologies allow individuals to self-assess and then update beliefs about their own risk types, for example, genetic tests, regular health examinations, among others. Based on the updated loss probabilities, individuals update their decisions on loss prevention efforts in the absence of insurance.

We show that information revelation can still lead to a Pareto improvement in the no insurance scenario if the new information enables individuals to make sufficiently effective loss prevention decisions, which requires a large difference in loss prevention efficiency between revealed high-risk groups and the uninformed state.

Moreover, we find that Pareto improvement is easier to achieve without insurance than with insurance. The reason is that insurance covers individuals' risks (i.e., the potential loss D) and hence makes their utility less sensitive to changes in their loss prevention efforts than without insurance. Thus, a greater difference in the efficiency of loss prevention technology is required for a Pareto improvement following risk classification when having insurance, compared to the case without insurance.

Specifically, if no insurance is provided, individuals' expected utility is

$$U_{kni}(s) = p_k(s)u(W_0 - D - s) + (1 - p_k(s))u(W_0 - s),\ k = 0,1,2,...N, \tag{17}$$

where $k = 0$ denotes the expected utility of all individuals when uninformed and $k = 1,2,...N$

denotes that of individuals in Group k after information revelation. We find that Propositions 1 to 3 remain unchanged when no insurance is provided. That is, a Pareto improvement occurs if and only if the efficiency of loss prevention technology of those revealed high-risk groups with $\theta_i \geq \theta_0$ is either sufficiently lower than a threshold $\underline{\gamma}_{ni}(\theta_i)$ or sufficiently higher than a threshold $\overline{\gamma}_{ni}(\theta_i)$.

Next, we compare the thresholds $(\underline{\gamma}_{ni}(\theta_i), \overline{\gamma}_{ni}(\theta_i))$ with that in the main model $(\underline{\gamma}(\theta_i), \overline{\gamma}(\theta_i))$ when having insurance by the following simulation. Each individual faces a potential loss $D = 5$. Individuals are characterized by a constant absolute risk aversion (CARA) utility function $u(x) = 1 - \exp(-\tau x)$ with risk-aversion level $\tau = 1$ (Handel et al., 2015).[19] In the uninformed state, all individuals optimally pay loss prevention effort $s_0^* = 0.5$ and end up with loss probability $\theta_0 = 0.2$ in both the no insurance scenario and the insurance scenario in the main model, so that the expected loss is normalized as $\theta_0 D = 1$.[20]

Figure 5(a) illustrates the difference in loss prevention efficiency between revealed high-risk groups and the uninformed state required for a Pareto improvement. The red dotted and solid curves represent scenarios with insurance. The blue dotted and solid curves represent scenarios without insurance. The highest red dotted curve and the lowest red solid curve indicate that a greater difference in loss prevention efficiency is necessary for a Pareto improvement with insurance. In other words, without insurance, it becomes easier to achieve a Pareto improvement.

[19] With CARA utility function, wealth effect is excluded and thus individuals' initial wealth W_0 does not play a role.

[20] $\theta_0 = 0.2$ is calculated from the value of other parameters to ensure $s_0^* = 0.5$ is optimal in both the no insurance scenario and the main model so that we have a fair comparison between the no insurance scenario and the main model. Specifically, groups' revealed risk θ_i is defined as individuals' loss probability after risk classification when they maintain the loss prevention efforts s_0^*, i.e., $\theta_i = p_i(s_0^*)$. Hence, consistent definition on θ_i on the no insurance scenario and the main model requires the same optimal risk prevention effort level.

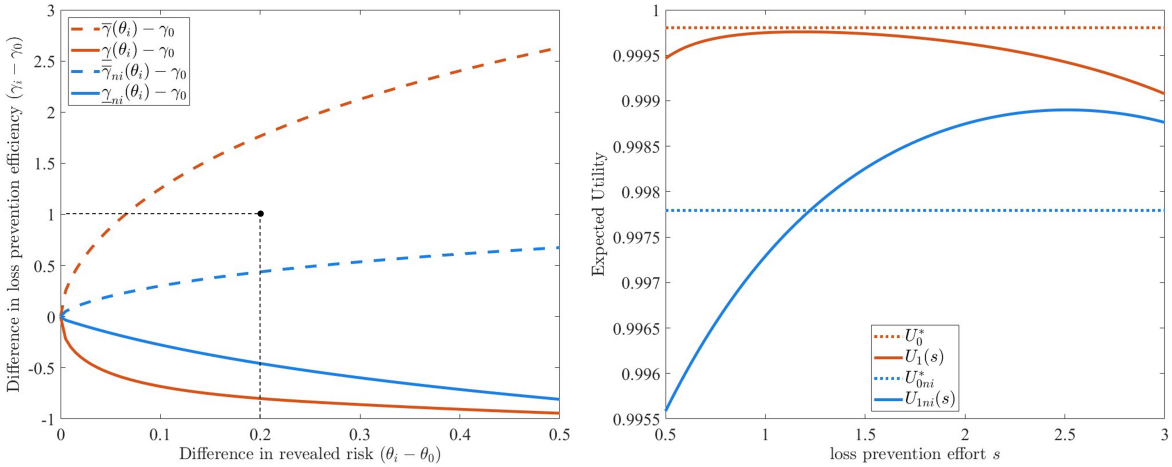


Figure 5: Conditions of Pareto improvement with and without insurance

Figure 5(b) illustrates how individuals' expected utility changes with loss prevention efforts, both with insurance (red curves) and without insurance (blue curves), given $\theta_i - \theta_0 = 0.2$ and $\gamma_i - \gamma_0 = 1$. The dark point in Figure 5(a) indicates this scenario. The dark point lies below the red dotted curve and above the black dotted curve, suggesting a group with (θ_i, γ_i) benefits from risk classification when no insurance is provided, but not with insurance. Figure 5(b) explains this: with $\gamma_i - \gamma_0 = 3$, Group-i individuals, having higher loss prevention efficiency than the uninformed state, increase their loss prevention effort from the uninformed level of 0.5. As Group-i's efforts rise from 0.5, the uninsured expected utility (blue solid curve) grows faster than the insured expected utility (red solid curve). When efforts range from 0.67 to 1, uninsured Group-i individuals achieve higher expected utility than the uninformed level (blue dotted curve). In contrast, insured Group-i individuals never surpass the insured uninformed utility (red dotted curve). Hence, a loss prevention efficiency difference greater than 3 is needed for insured high-risk individuals to counteract the utility loss from being identified as high risk.

6.2 Transferability to Credit Lending and Labor Markets

The preceding analyses are written in the terminology of insurance markets, where insurers use new, public information for risk classification. In this section, we show that similar analysis can apply naturally in other markets.

The core optimization problems in the main model in Equations (2) and (6) can be generalized

as

$$\max_s V_i(s) = W + G_i(s) - s, i = 0, 1, 2, ...N, \tag{18}$$

where $G_i(s)$ is benefit that individuals earn from paying efforts s and possesses similar characteristic of the function $-p_i(s)$. Our preceding analysis can be applied to this generalized framework and shows the value of new, public information.[21]

Take the credit lending market as an example. Some machine-learning-enhanced and AI-driven tools are applied by banks to assess non-traditional data points when evaluating credit risks for applicants, for example, applicants' social media and digital footprint data (Berg, Burg, Gombović, & Puri, 2020). Our results imply that risk classification employed by banks achieves a Pareto improvement if the new, public information involved in credit risk evaluation enables applicants to effectively change their efforts in changing their default probability. For example, heavy social network users will find it efficient to earn a higher credit score by paying credit bills on time, behaving well when booking travel and hotels, etc.

The labor market is another example, where emerging technologies and novel sources of information are reshaping the ways in which individual skills can be assessed. Online learning systems (e.g., Coursera, edX, and LinkedIn Learning) provide verifiable records of skills. Open-source project platforms (e.g., GitHub, Kaggle) offer transparent evidence of problem-solving ability. Within firms, digital collaboration tools create detailed data on task completion, communication, and teamwork, enabling a more fine-grained evaluation of soft skills. Job matching and salary calculation based on this new information achieves Pareto improvement if applicants can more effectively present their skills and expertise.

7 Conclusion

Modern risk-assessment technologies improve the identification of individual risk types, while advances in loss-prevention technologies enable individuals to meaningfully alter their risk status.

[21] $G_i(s)$ satisfies

$$G_i'(s) > 0, \quad G_i''(s) < 0, \quad \gamma_i = \frac{G_i'(s)}{G_i(s)}, \text{ for } i = 1, 2, ..., N, \quad \text{and} \quad G_0(s) = \sum_{i=1}^{N} \lambda_i G_i(s).$$

That is, individuals earn greater benefit with decreasing margin from paying greater efforts. The two-dimensional public information updates their benefit function from the uninformed $G_0(s)$ to $G_i(s)$ $(i = 1, 2, ..., N)$ and enables them to make better decisions on efforts.

These innovations generate publicly observable information shared by insurers and policyholders. Yet existing research has largely emphasized their role in reducing asymmetric information, paying less attention to how risk classification technologies shape welfare in symmetric-information settings.

We study the welfare effects of risk classification based on two-dimensional public information: individuals' baseline risk types and their loss-prevention efficiencies. When new information substantially improves prevention decisions, risk classification should be encouraged. In particular, individuals revealed as high risk may experience utility gains—either ex ante or ex post—if classification enables more effective prevention choices. Moreover, even when a risk classification technology fails to generate welfare gains, appropriately designed partial risk classification can ensure welfare improvement.

We also clarify the tradeoff between two regulatory benchmarks. Ex-ante unanimity approves a broader set of risk classification technologies and thus expands the scope for efficiency gains, whereas Pareto improvement guarantees that no individual is harmed ex post and thus emphasizes the fairness. In simulations of one million randomly generated risk classification technologies, fewer than one quarter achieve Pareto improvement, while more than two thirds satisfy ex-ante unanimity over a wide and empirically plausible range of risk aversion.

These findings yield clear policy guidance. When effective and actionable loss-prevention technologies exist, insurers and individuals should be encouraged to generate, share, and use risk-related information. When prevention is limited, expanded reclassification risk harms high-risk individuals and should be approached cautiously. In such cases, priority should be given to developing prevention technologies before broadening risk classification. Finally, when a risk classification technology is welfare-reducing, carefully designed partial risk classification can provide a viable alternative.

References

[1] Bardey, D., & De Donder, P. (2013). Genetic testing with primary prevention and moral hazard. *Journal of Health Economics*, *32*(5), 768–779.

[2] Bardey, D., De Donder, P., & Mantilla, C. (2019). How is the trade-off between adverse selection and discrimination risk affected by genetic testing? Theory and experiment. *Journal of Health Economics*, *68*, 102223.

[3] Barseghyan, L., Prince, J., & Teitelbaum, J. C. (2011). Are risk preferences stable across contexts? Evidence from insurance data. *American Economic Review*, *101*(2), 591–631.

[4] Barsky, R. B., Juster, F. T., Kimball, M. S., & Shapiro, M. D. (1997). Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study. *Quarterly Journal of Economics*, *112*(2), 537–579.

[5] Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *Review of Financial Studies*, *33*(7), 2845–2897.

[6] Bergemann, D., Brooks, B., & Morris, S. (2015). The limits of price discrimination. *American Economic Review*, *105*(3), 921–957.

[7] Bergemann, D., & Morris, S. (2013). Robust predictions in games with incomplete information. *Econometrica*, *81*(4), 1251–1308.

[8] Bond, E. W., & Crocker, K. J. (1991). Smoking, skydiving, and knitting: The endogenous categorization of risks in insurance markets with asymmetric information. *Journal of Political Economy*, *99*(1), 177–200.

[9] Chetty, R. (2006). A new method of estimating risk aversion. *American Economic Review*, *96*(5), 1821–1834.

[10] Ciancio, A., Kämpfen, F., Kohler, H.-P., & Thornton, R. (2025). Surviving bad news: health information without treatment options. *American Economic Review: Insights*, *7*(1), 1–18.

[11] Cohen, A., & Einav, L. (2007). Estimating risk preferences from deductible choice. *American Economic Review*, *97*(3), 745-788.

[12] Courbage, C., Loubergé, H., & Peter, R. (2017). Optimal prevention for multiple risks. *Journal of Risk and Insurance*, *84*(3), 899–922.

[13] Crocker, K. J., Rothschild, C., & Snow, A. (2025). The theory of risk classification. In

Handbook of insurance: Volume II (pp. 223–257).

[14] Crocker, K. J., & Snow, A. (1986). The efficiency effects of categorical discrimination in the insurance industry. *Journal of Political Economy*, *94*(2), 321–344.

[15] de Garidel-Thoron, T. (2005). Welfare-improving asymmetric information in dynamic insurance markets. *Journal of Political Economy*, *113*(1), 121–150.

[16] Delavande, A., & Kohler, H.-P. (2012). The impact of HIV testing on subjective expectations and risky behavior in Malawi. *Demography*, *49*, 1011–1036.

[17] Dionne, G., & Rothschild, C. (2014). Economic effects of risk classification bans. *Geneva Risk and Insurance Review*, *39*(2), 184–221.

[18] Doherty, N. A., & Posey, L. L. (1998). On the value of a checkup: Adverse selection, moral hazard and the value of information. *Journal of Risk and Insurance*, *65*(2), 189–211.

[19] Doherty, N. A., & Thistle, P. D. (1996). Adverse selection with endogenous information in insurance markets. *Journal of Public Economics*, *63*(1), 83-102.

[20] Ehrlich, I., & Becker, G. S. (1972). Market insurance, self-insurance, and self-protection. *Journal of Political Economy*, *80*(4), 623–648.

[21] Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., & Cullen, M. R. (2013). Selection on moral hazard in health insurance. *American Economic Review*, *103*(1), 178–219.

[22] Eling, M., Jia, R., Lin, J., & Rothschild, C. (2022). Technology heterogeneity and market structure. *Journal of Risk and Insurance*, *89*(2), 427–448.

[23] Eső, P., & Szentes, B. (2007). Optimal information disclosure in auctions and the handicap auction. *Review of Economic Studies*, *74*(3), 705–731.

[24] Farinha Luz, V., Gottardi, P., & Moreira, H. (2023). Risk classification in insurance markets with risk and preference heterogeneity. *Review of Economic Studies*, *90*(6), 3022–3082.

[25] Finkelstein, A., Poterba, J., & Rothschild, C. (2009). Redistribution by insurance market regulation: Analyzing a ban on gender-based retirement annuities. *Journal of Financial Economics*, *91*(1), 38–58.

[26] Garcia, D., & Tsur, M. (2021). Information design in competitive insurance markets. *Journal of Economic Theory*, *191*, 105160.

[27] Gemmo, I., Browne, M. J., & Gründl, H. (2025). Privacy concerns in insurance markets: Implications for market equilibria and customer utility. *Canadian Journal of Economics/Revue*

canadienne d'économique, *58*(2), 484–514.

[28] Gertner, R. (1993). Game shows and economic behavior: Risk-taking on "card sharks". *Quarterly Journal of Economics*, *108*(2), 507–521.

[29] Handel, B., Hendel, I., & Whinston, M. D. (2015). Equilibria in health exchanges: Adverse selection versus reclassification risk. *Econometrica*, *83*(4), 1261–1313.

[30] Häusle, N., & Rothschild, C. (2024). Information design in insurance markets. *Available at SSRN 4925062*.

[31] Hoel, M., & Iversen, T. (2002). Genetic testing when there is a mix of compulsory and voluntary health insurance. *Journal of Health Economics*, *21*(2), 253–270.

[32] Hoel, M., Iversen, T., Nilssen, T., & Vislie, J. (2006). Genetic testing in competitive insurance markets with repulsion from chance: A welfare analysis. *Journal of Health Economics*, *25*(5), 847–860.

[33] Holzapfel, J., Peter, R., & Richter, A. (2024). Mitigating moral hazard with usage-based insurance. *Journal of Risk and Insurance*, *91*(4), 813–839.

[34] Hoy, M. (1982). Categorizing risks in the insurance industry. *Quarterly Journal of Economics*, *97*(2), 321–336.

[35] Hoy, M. (1989). The value of screening mechanisms under alternative insurance possibilities. *Journal of Public Economics*, *39*(2), 177–206.

[36] Hoy, M., & Polborn, M. (2000). The value of genetic information in the life insurance market. *Journal of Public Economics*, *78*(3), 235–252.

[37] Hoy, M., & Witt, J. (2007). Welfare effects of banning genetic information in the life insurance market: The case of BRCA1/2 genes. *Journal of Risk and Insurance*, *74*(3), 523–546.

[38] Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, *101*(6), 2590–2615.

[39] Kellerman, S. E., Lehman, S. J., Lansky, A., Stevens, M. R., Hecht, F. M., Bindman, A. B., & Wortley, P. M. (2002). HIV testing within at-risk populations in the United States and the reasons for seeking or avoiding HIV testing. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *31*(2), 202–210.

[40] Kerwin, J. T. (2025). Scared straight or scared to death? Fatalism in response to disease

risks. *Economic Journal*, *135*(670), 1923-1941.

[41] Lau, M. I., & Yoo, H. I. (2025). Structural estimation of higher order risk preferences. *Econometrica*, *93*(5), 1855–1883.

[42] Li, J., Xiang, Y., Fang, J., Wang, W., & Pi, Y. (2019). Research on multiple sensors vehicle detection with emd-based denoising. *IEEE Internet of Things Journal*, *6*(4), 6262-6270.

[43] Li, L. (2021). Opening up the black box: Technological transparency and prevention. *Journal of Risk and Insurance*, *88*(3), 665–693.

[44] Manogaran, G., Varatharajan, R., Lopez, D., Kumar, P. M., Sundarasekar, R., & Thota, C. (2018). A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. *Future Generation Computer Systems*, *82*, 375–387.

[45] Nelson, S. T. (2025). Private information and price regulation in the US credit card market. *Econometrica*, *93*(4), 1371–1410.

[46] Paravisini, D., Rappoport, V., & Ravina, E. (2017). Risk aversion and wealth: Evidence from person-to-person lending portfolios. *Management science*, *63*(2), 279–297.

[47] Peter, R. (2021). A fresh look at primary prevention for health risks. *Health Economics*, *30*(5), 1247–1254.

[48] Peter, R., Richter, A., & Thistle, P. (2017). Endogenous information, adverse selection, and prevention: Implications for genetic testing policy. *Journal of Health Economics*, *55*, 95–107.

[49] Picard, P. (2008). Natural disaster insurance and the equity-efficiency trade-off. *Journal of Risk and Insurance*, *75*(1), 17–38.

[50] Rashidi, P., Cook, D. J., Holder, L. B., & Schmitter-Edgecombe, M. (2011). Discovering activities to recognize and track in a smart environment. *IEEE Transactions on Knowledge and Data Engineering*, *23*(4), 527-539.

[51] Rothschild, C. (2011). The efficiency of categorical discrimination in insurance markets. *Journal of Risk and Insurance*, *78*(2), 267–285.

[52] Rothschild, M., & Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics*, *90*(4), 629–649.

[53] Strohmenger, R., & Wambach, A. (2000). Adverse selection and categorical discrimination in the health insurance markets: The effects of genetic tests. *Journal of Health Economics*, *19*(2), 197–218.

[54] Sydnor, J. (2010). (Over)insuring modest risks. *American Economic Journal: Applied Economics*, *2*(4), 177–199.

[55] Thornton, R. L. (2008). The demand for, and impact of, learning HIV status. *American Economic Review*, *98*(5), 1829–1863.

[56] Wilson, C. (1977). A model of insurance markets with incomplete information. *Journal of Economic Theory*, *16*(2), 167–207.

Appendixes

A Proof of Proposition 1

For lower-risk groups with $\theta_i \leq \theta_0$, we have

$$U_i(s_i^*) \geq U_i(s_0^*) \geq U_0(s_0^*). \tag{A-1}$$

Therefore, lower-risk groups are never worse off, regardless of their loss prevention efficiency.

Solving the optimization problem in Equations (2) and (6) yield the first order derivative as follows:

$$\frac{dU_i 0(s)}{ds} = u'(W_0 - p_0(s)D - s) \times (-p_0'(s_0)D - 1) = 0, \tag{A-2}$$

which is equivalent to $\theta_0\gamma_0 D = 1$, and

$$\frac{dU_i(s)}{ds} = u'(W_0 - p_i(s)D - s) \times (\theta_i\gamma_i \exp(-\gamma_i(s_i - s_0^*))D - 1) = 0. \tag{A-3}$$

With some algebra, we have

$$s_i^* = s_0^* + \frac{1}{\gamma_i}\ln(\theta_i\gamma_i D), \tag{A-4}$$

and thus

$$g(\gamma_i) \equiv p_i(s_i^*)D + s_i^* - (p_0(s_0^*)D + s_0^*) = \frac{1}{\gamma_i} + \frac{1}{\gamma_i}\ln(\theta_i\gamma_i D) - \theta_0 D. \tag{A-5}$$

It is straightforward that

$$U_i(s_i^*) \geq U_0(s_0^*) \Leftrightarrow g(\gamma_i) \leq 0. \tag{A-6}$$

Thus, we focus on the sign of $g(\gamma_i)$ for higher-risk groups with $\theta_i > \theta_0$. Given $\gamma_i > 0$, we have

$$\lim_{\gamma_i \to 0} g(\gamma_i) = -\infty < 0 \text{ and } \lim_{\gamma_i \to \infty} g(\gamma_i) = -\theta_0 D < 0. \tag{A-7}$$

Besides,

$$g(\frac{1}{\theta_i D}) = (\theta_i - \theta_0)D > 0. \tag{A-8}$$

Taking derivative with respect to γ_i yields

$$g'(\gamma_i) = \frac{1-\ln(\theta_i\gamma_i D)}{\gamma_i^2} \begin{cases} \geq 0, & \text{if } \gamma_i \leq \frac{1}{\theta_i D}, \\ < 0, & \text{if } \gamma_i > \frac{1}{\theta_i D}. \end{cases} \tag{A-9}$$

Equations (A-7) to (A-9) implies that there exists two cutoffs $\underline{\gamma}(\theta_i) < \frac{1}{\theta_i D} < \overline{\gamma}(\theta_i)$ that satisfy

$$g(\underline{\gamma}(\theta_i)) = g(\overline{\gamma}(\theta_i)) = 0,\ g'(\underline{\gamma}(\theta_i)) > 0, \text{ and } g'(\overline{\gamma}(\theta_i)) < 0, \tag{A-10}$$

such that

$$g(\gamma_i) \begin{cases} \leq 0, & \text{if } \gamma_i \leq \underline{\gamma}(\theta_i) \text{ or } \gamma_i \geq \overline{\gamma}(\theta_i), \\ > 0, & \text{if } \underline{\gamma}(\theta_i) < \gamma_i < \overline{\gamma}(\theta_i), \end{cases} \tag{A-11}$$

which, together with Equation (A-6), completes the proof of Proposition 1. ■

B Proof of Proposition 2

Following the expression of $g(\gamma_i)$ in Equation (A-5), we have

$$\lim_{\theta_i \to \theta_0} g(\gamma_0) = 0 \text{ and } \lim_{\theta_i \to \theta_0} g'(\gamma_0) = 0. \tag{B-1}$$

That is, γ_0 is the only solution to $g(\gamma_i) = 0$ as $\theta_i \to \theta_0$. Together with Equation (A-10), we have

$$\lim_{\theta_i \to \theta_0} \underline{\gamma}(\theta_i) = \lim_{\theta_i \to \theta_0} \overline{\gamma}(\theta_i) = \gamma_0. \tag{B-2}$$

In Equation (A-10), taking derivative with respect to θ_i yields

$$\frac{\partial \underline{\gamma}(\theta_i)}{\partial \theta_i} = -\frac{\partial g(\gamma_i)}{\partial \theta_i}\Big|_{\gamma_i = \underline{\gamma}(\theta_i)} \times \frac{1}{g'(\underline{\gamma}(\theta_i))} = -\frac{1}{\theta_i \underline{\gamma}(\theta_i) g'(\underline{\gamma}(\theta_i))} < 0, \tag{B-3}$$

and

$$\frac{\partial \overline{\gamma}(\theta_i)}{\partial \theta_i} = -\frac{\partial g(\gamma_i)}{\partial \theta_i}\Big|_{\gamma_i = \overline{\gamma}(\theta_i)} \times \frac{1}{g'(\overline{\gamma}(\theta_i))} = -\frac{1}{\theta_i \overline{\gamma}(\theta_i) g'(\overline{\gamma}(\theta_i))} > 0, \tag{B-4}$$

which completes the proof of Proposition 2. ■

C Proof of Proposition 3

Given $\theta_i > \theta_0$, it is straightforward to have

$$p_i(s_0^*) > p_0(s_0^*). \tag{C-1}$$

Given that a Pareto improvement occurs, we have

$$U_i(s_i^*) \geq U_0(s_0^*) > U_0(s_i^*). \tag{C-2}$$

That is,

$$W_0 - p_i(s_i^*)D - s_i^* > W_0 - p_0(s_i^*)D - s_i^*. \tag{C-3}$$

With some algebra, we have

$$p_i(s_i^*) < p_0(s_i^*), \tag{C-4}$$

which completes the proof of Proposition 3. ■

D Proof of Proposition 4

Denote the final wealth in the uninformed state and any state after risk classification as

$$V_i(s_i) = W_0 - p_i(s_i)D - s_i,\ i = 0,1,2,...,N. \tag{D-1}$$

Following Equations (2) and (6), given $u'(\cdot) > 0$, we have $s_i^* = argmax_{s_i} V_i(s_i)$, $i = 0,1,2,...,N$. For simplification, let $V_i^* = V_i(s_i^*)$. Therefore, we have

$$\sum_{i=1}^{N} \lambda_i V_i^* > \sum_{i=1}^{N} \lambda_i V_i(s_0^*) = V_0^*. \tag{D-2}$$

Without loss of generality, we consider that

$$V_1^* \leq V_2^* \leq ... \leq V_N^*. \tag{D-3}$$

Therefore, we have

$$V_0^* < V_N^*. \tag{D-4}$$

Equation (8) can be rewritten as

$$\sum_{i=1}^{N} \lambda_i u(V_i^*) > u(V_0^*). \tag{D-5}$$

Inserting CARA utility function, this is equivalent to

$$h(\tau_A) \equiv \sum_{i=1}^{N} \lambda_i \exp(-\tau_A(V_i^* - V_0^*)) < 1. \tag{D-6}$$

That is, we need to solve the condition for $h(\tau_A) < 1$.

Taking derivative with respect to τ_A yields

$$h'(\tau_A) = -\sum_{i=1}^{N} \lambda_i (V_i^* - V_0^*) \exp(-\tau_A(V_i^* - V_0^*)). \tag{D-7}$$

It is straightforward to have that

$$h(0) = \sum_{i=1}^{N} \lambda_i = 1 \text{ and } h'(0) = V_0^* - \sum_{i=1}^{N} \lambda_i V_i^* < 0. \tag{D-8}$$

First, if a Pareto improvement occurs, i.e., $\forall i = 1, 2, ..., N$, $V_i^* \geq V_0^*$, then we have $V_i - V_0^* \geq 0$ with the equality holds strictly for some i, and thus

$$h(\tau_A) < \sum_{i=1}^{N} \lambda_i = 1, \forall \tau_A > 0. \tag{D-9}$$

That is, if a Pareto improvement occurs, then an ex-ante welfare improvement occurs for individuals with any $\tau_A > 0$, or equivalently, $\tau_{Amax} = \infty$.

If no Pareto improvement occurs, individuals end up with lower final wealth at least in one state after risk classification, i.e.,

$$V_1^* < V_0^*. \tag{D-10}$$

Therefore,

$$\lim_{\tau_A \to \infty} h(\tau_A) > \lim_{\tau_A \to \infty} \lambda_1 \exp(-\tau_A(V_1^* - V_0^*)) = \infty > 1. \tag{D-11}$$

Together with Equation (D-8), we have that there should exists a $\tau_{Amax} > 0$ such that

$$h(\tau_{Amax}) = 1. \tag{D-12}$$

Following Equation (D-7), we have

$$\begin{aligned} h'(\tau_{Amax}) &= -\sum_{i=1}^{N} \lambda_i (V_i^* - V_0^*) \exp(-\tau_{Amax}(V_i^* - V_0^*)) \\ &> -\sum_{i=1}^{N} \lambda_i (V_N^* - V_0^*) \exp(-\tau_{Amax}(V_i^* - V_0^*)) \\ &= (V_N^* - V_0^*) h(\tau_{Amax}) \\ &= V_N^* - V_0^* \\ &> 0, \end{aligned} \tag{D-13}$$

where the first inequality follows Equation (D-4). Therefore, τ_{Amax} is the only solution to $h(\tau_A) = 1$. Otherwise, if $h(\tau_A) = 1$ has multiple solutions, then there should exists $\hat{\tau}_A \neq \tau_{Amax}$ such that $h(\hat{\tau}_A) = 1$ and $h'(\hat{\tau}_A) < 0$, which contradicts Equation (D-7). Given τ_{Amax} is the only solution to $h(\tau_A) = 1$ and Equations (D-8) and (D-11), we have

$$h(\tau_A) \begin{cases} < 1, & \text{if } 0 < \tau_A < \tau_{Amax}, \\ \geq 1, & \text{if } \tau_A \geq \tau_{Amax}, \end{cases} \tag{D-14}$$

which completes the proof of Proposition (4). ■

E Proof of Proposition 5

For Proposition 5, we need to prove that if an ex-ante welfare improvement occurs for any individual with $\tau_A > 0$, then a Pareto improvement occurs. We prove this by contradiction. Suppose that an ex-ante welfare improvement occurs for any individual with $\tau_A > 0$, but no Pareto improvement occurs. That is,

$$h(\tau_A) < 1, \forall \tau_A > 0, \text{ but } V_1^* < V_0^*. \tag{E-1}$$

Then we have $\lim_{\tau_A \to \infty} h(\tau_A) > 1$ as shown in Equation (D-11), which is contradictory. Thus, the supposition should not hold, which completes the proof of Proposition 5. ■

F Proof of Remark 6

Consider the optimization problem of Group 1; that is

$$\max_s U_1(s; \lambda_{H1}, \lambda_{M1}) = u(W_0 - p_1(s)D - s), \tag{F-1}$$

where $p_1(s) = \lambda_{H1}\theta_H \exp(-\gamma_H(s - s_0^*)) + \lambda_{M1}\theta_M \exp(-\gamma_M(s - s_0^*)) + (1 - \lambda_{H1} - \lambda_{M1})\theta_L \exp(-\gamma_L(s - s_0^*))$. The first order condition is

$$\frac{dU_1(s; \lambda_{H1}, \lambda_{M1})}{ds}|_{s=s_1^*} = u'(W_0 - p_1(s_1^*)D - s_1^*) \times (-p_1'(s_1^*)D - 1) = 0. \tag{F-2}$$

Given the optimal s_1^*, we denote

$$U_1^*(\lambda_{H1}, \lambda_{M1}) = U_1(s_1^*; \lambda_{H1}, \lambda_{M1}). \tag{F-3}$$

It is straightforward to have

$$U_1^*(\lambda_H, \lambda_M) = U_0^*. \tag{F-4}$$

Given Equation (**??**), we have

$$\frac{\theta_0 - \theta_L}{\theta_H - \theta_L} - \frac{\lambda_{M1}(\theta_M - \theta_L)}{\theta_H - \theta_0} \leq \lambda_{H1} \leq 1 - \lambda_{M1}. \tag{F-5}$$

Given any λ_{M1}, we have

$$U_1^*\left(\frac{\theta_0 - \theta_L}{\theta_H - \theta_L} - \frac{\lambda_{M1}(\theta_M - \theta_L)}{\theta_H - \theta_0}, \lambda_{M1}\right) \geq U_1\left(s_0^*; \frac{\theta_0 - \theta_L}{\theta_H - \theta_L} - \frac{\lambda_{M1}(\theta_M - \theta_L)}{\theta_H - \theta_0}, \lambda_{M1}\right) = U_0^*, \tag{F-6}$$

where the inequality strictly holds when $\lambda_{M1} \neq \lambda_M$. In $U_1^*(\lambda_{H1}, \lambda_{M1})$, taking derivative with respect to λ_{H1} yields

$$\frac{\partial U_1^*(\lambda_{H1}, \lambda_{M1})}{\partial - \lambda_{H1}} = u'(W_0 - p_1(s_1^*)D - s_1^*)(p_H(s_1^*) - p_L(s_1^*))D < 0. \tag{F-7}$$

Therefore, for any λ_{M1}, there exists $f(\lambda_{M1}) > \frac{\theta_0-\theta_L}{\theta_H-\theta_L} - \frac{\lambda_{M1}(\theta_M-\theta_L)}{\theta_H-\theta_0}, \lambda_{M1}$ such that

$$U_1^*(\lambda_{H1}, \lambda_{M1}) \geq U_0^*, \forall \lambda_{H1} \leq f(\lambda_{M1}), \tag{F-8}$$

which completes the proof of Remark 6. ■

G Risk Aversion Level

In this section, we first prove Propositions 5 and 6 with a constant relative risk aversion (CRRA) utility function $u(x) = \frac{x^{1-\tau_R}}{1-\tau_R}$. Second, Table G1 summarizes estimations of absolute and relative risk aversions in literature.

Proposition G1. *There exists a threshold $\tau_{Rmax} > 0$ such that an ex-ante welfare improvement occurs for all individuals with relative risk aversion level τ_R lower than τ_{Rmax}.*

Proof: Inserting CRRA utility function, Equation (8) equivalent to

$$h_2(\tau_R) \equiv \sum_{i=1}^{N} \lambda_i \left(\frac{V_i^*}{V_0^*}\right)^{1-\tau_R} \begin{cases} > 1, & \text{if} \tau_R < 1, \\ < 1, & \text{if} \tau_R > 1. \end{cases} \tag{G-1}$$

Taking derivative with respect to τ_R yields

$$h_2'(\tau_R) = -\sum_{i=1}^{N} \lambda_i \left(\frac{V_i^*}{V_0^*}\right)^{1-\tau_R} \ln\left(\frac{V_i^*}{V_0^*}\right), \tag{G-2}$$

and

$$h_2''(\tau_R) = \sum_{i=1}^{N} \lambda_i \left(\frac{V_i^*}{V_0^*}\right)^{1-\tau_R} \ln^2\left(\frac{V_i^*}{V_0^*}\right) > 0. \tag{G-3}$$

Following Equation (D-2), it is straightforward to have that

$$h_2(0) = \frac{\sum_{i=1}^{N} \lambda_i V_i^*}{V_0^*} > 1 \text{ and } h_2(1) = \sum_{i=1}^{N} \lambda_i = 1. \tag{G-4}$$

First, if a Pareto improvement occurs, i.e., $\forall i = 1, 2, ..., N$, $V_i^* \geq V_0^*$, then we have $\frac{V_i^*}{V_0^*} \geq 1$ with

the equality holds strictly for some i, and thus

$$h_2(\tau_R)\begin{cases} > \sum_{i=1}^{N} \lambda_i = 1, \forall \tau_R < 1, \\ < \sum_{i=1}^{N} \lambda_i = 1, \forall \tau_R > 1. \end{cases} \tag{G-5}$$

That is, if a Pareto improvement occurs, then an ex-ante welfare improvement occurs for individuals with any $\tau_R > 0$, or equivalently, $\tau_{Rmax} = \infty$.

If no Pareto improvement occurs, individuals end up with lower final wealth at least in one state after risk classification, i.e.,

$$V_1^* < V_0^*. \tag{G-6}$$

Therefore,

$$\lim_{\tau_R \to \infty} h_2(\tau_R) > \lim_{\tau_R \to \infty} \lambda_1 \left(\frac{V_0^*}{V_1^*}\right)^{\tau_R - 1} = \infty > 1. \tag{G-7}$$

Given Equations (G-2) and (G-3), we have if $h_2'(1) = 0$ then

$$h_2(\tau_R) > 1, \forall \tau_R > 0 \text{ and } \tau_R \neq 1. \tag{G-8}$$

Therefore, Equation (G-1) holds for $\tau_R < 1$; that is $\tau_{Rmax} = 1$.

If $h_2'(1) < 0$, together with Equation (G-3), we have that $h_2(\tau_R)$ decreases in τ_R for $\tau_R \in (0,1)$. Therefore,

$$h_2(\tau_R) > h_2(1) = 1, \ \forall \tau_R \in (0,1). \tag{G-9}$$

Given $h_2(1) = 1$, $h_2'(1) < 0$, Equations (G-7) and (G-3), there exists $\tau_{Rmax} > 1$ such that

$$h_2(\tau_R)\begin{cases} < 1, & \text{if } 1 < \tau_R < \tau_{Rmax}, \\ \geq 1, & \text{if } \tau_R \geq \tau_{Rmax}. \end{cases} \tag{G-10}$$

Equations (G-9) and (G-10) implies that Equation (G-1) holds for $\tau_R < \tau_{Rmax}$, where $\tau_{Rmax} > 1$ if $h_2'(1) < 0$.

If $h_2'(1) > 0$, together with Equation (G-3), we have that $h_2(\tau_R)$ increases in τ_R for $\tau_R > 1$.

Therefore,

$$h_2(\tau_R) > h_2(1) = 1,\ \forall \tau_R > 1. \tag{G-11}$$

Given $h_2(0) > 1$, $h_2(1) = 1$, $h_2'(1) > 0$, and Equation (G-3), there exists $\tau_{Rmax} < 1$ such that

$$h_2(\tau_R) \begin{cases} > 1, & \text{if } 0 < \tau_R < \tau_{Rmax}, \\ \leq 1, & \text{if } \tau_{Rmax} \leq \tau_R < 1. \end{cases} \tag{G-12}$$

Equations (**??**) and (G-12) implies that Equation (G-1) holds for $\tau_R < \tau_{Rmax}$, where $\tau_{Rmax} < 1$ if $h_2'(1) > 0$, which completes the proof of Proposition (G1). ■

Proposition G2. *A Pareto improvement is equivalent to an ex-ante welfare improvement for individuals with any risk aversion level* $\tau_R \in (0, \infty)$.

Proof: For Proposition G2, we need to prove that if an ex-ante welfare improvement occurs for any individual with $\tau_R > 0$, then a Pareto improvement occurs. We prove this by contradiction. Suppose that an ex-ante welfare improvement occurs for any individual with $\tau_R > 0$, but no Pareto improvement occurs. That is, Equation (G-1) always hold, but

$$V_1^* < V_0^*. \tag{G-13}$$

Then we have $\lim_{\tau_R \to \infty} h(\tau_R) > 1$ as shown in Equation (G-7), which is contradictory. Thus, the supposition should not hold, which completes the proof of Proposition G2. ■

Table G1 summarizes the estimations of absolute (ARA) and relative risk aversion (RRA) in literature across different contexts, where the rule that RRA $\approx$ ARA $\times$ Average Wealth Level is always applied. The estimated absolute risk aversion level is relatively stable across different contexts, supporting that the upper bound 0.1 selected in this paper is reasonable. However, estimated relative risk aversion level varies. The reason lies in the great difference in average wealth level across difference contexts.

Table G1: Estimations of Absolute and Relative Risk Aversion in Literature

	Context	Estimated Absolute Risk Aversion		Estimated Relative Risk Aversion		Average Wealth Level
Gertner (1993)	television game show	lower bound	0.00031	lower bound	4.79	15,437
Barsky et al. (1997)	survey on gambles over lifetime income	–	–	mean	4.15	–
				$\pi_{5\%}$	1.30	
				$\pi_{95\%}$	33.33	
Chetty (2006)	labor supply	–	–	mean	0.71	–
		–	–	range	0.15-1.78	
Cohen and Einav (2007)	auto insurance	$\pi_{25\%}$	$2.3 * 10^{-6}$	$\pi_{25\%}$	0.03	14,536
		$\pi_{95\%}$	0.0099	$\pi_{95\%}$	143.27	
Sydnor (2010)	homeowner insurance	range*	0.0018-0.014	range	$1,839 - 14,043$	1 million
Barseghyan et al. (2011)	auto and home insurance	range	0.002-0.0082	–	–	–
Einav et al. (2013)	health insurance	mean	0.0019	mean*	59.45	31,292
Paravisini, Rappoport, and Ravina (2017)	person-to-person lending	mean	0.037	mean	2.81	125.6
		$\pi_{1\%}$	-0.00812	$\pi_{1\%}$	-0.16	4.1
		$\pi_{99\%}$	0.08675	$\pi_{99\%}$	16.79	1,226.4
Lau and Yoo (2025)	laboratory experiments on gambling	mean	0.287	mean	0.99	–

Notes: * indicates estimations that are not provided by the specific studies, but calculated from the relationship between absolute and relative risk aversion, i.e., RRA $\approx$ ARA $\times$ Average Wealth Level.