Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(7): 4512–4534.
DOI: 10.3934/era.2026199
Received: 28 February 2026
Revised: 12 May 2026
Accepted: 19 May 2026
Published: 26 May 2026

Research article

Deterministic solvency thresholds and RL-based premium calibration for life insurance under age-structured epidemics

Bizhigit Sagidolla[1,*] **and Shirali Kadyrov**[2]

[1] SDU University, Kaskelen, Kazakhstan

[2] New Uzbekistan University, Tashkent, Uzbekistan

* **Correspondence:** Email: bizhigit.sagidolla@gmail.com.

Abstract: We develop a deterministic framework that links epidemic propagation, mortality risk, and life-insurance solvency by modeling the population through an age-structured SEIRD system in which disease-induced deaths drive an insurer's surplus process governed by an ordinary differential equation with a continuous premium inflow and death-benefit outflow. We characterize the disease-free equilibrium and show that its local stability depends on the basic reproduction number $\mathcal{R}_0$, and for any finite horizon $[0, T]$, we derive an explicit solvency threshold in the form of a critical premium $p_{\text{crit}}(T)$ that guarantees a nonnegative surplus and depends on the age-weighted infection burden. For large horizons, we obtain an analytically tractable approximation of this threshold using multi-group final-size relations and show that it monotonically increases with epidemic severity. To illustrate practical implications, we formulate an age-specific premium selection as a one-shot continuous-action reinforcement-learning problem in which a Proximal Policy Optimization (PPO) agent is trained on simulated age-structured epidemic scenarios to choose premiums that avoid ruin while remaining near actuarially fair levels; out-of-sample tests confirm that the learned premiums satisfy the analytical solvency constraints, deliver low ruin probabilities, and achieve tight fairness calibration. The combined analytical and numerical results provide a transparent basis for epidemic-sensitive premium design and stress testing of life-insurance portfolios.

Keywords: epidemic risk modelling; life insurance solvency; age-structured SEIRD model; reinforcement learning; premium calibration; numerical stress testing; insurance mathematics

1. Introduction

We study a life insurance portfolio exposed to epidemic mortality, where the underlying disease dynamics follow an age-structured SEIRD model. Disease-induced mortality drives claim outflows, and premiums are continuously collected at a constant rate. The coupled epidemic-insurance interaction

is represented through a deterministic system of ordinary differential equations, thereby allowing a transparent, pathwise solvency analysis without introducing stochastic mortality assumptions.

Our objectives are as follows: (i) formulate a coupled age–structured SEIRD–surplus framework with latency and disease-induced mortality; (ii) analyze the equilibria and stability properties of the epidemic subsystem via the basic reproduction number $\mathcal{R}_0$; (iii) derive explicit finite-horizon solvency conditions through a critical premium $p_{\text{crit}}(T)$; and (iv) present a large-horizon approximation of this threshold using multi–group final-size relations.

Classical actuarial–epidemic models employ either SIR or SEIR dynamics to quantify the effect of infection prevalence on mortality and reserves [1, 2]. Modern extensions incorporate latency, asymptomatic transmission, and heterogeneity, often via SEIARD-type systems [3]. Hainaut [4] introduced deterministic and stochastic pandemic-pricing models, while Zhai et al. [5] calibrated an enriched SVEI_3RD model to COVID-19 data and analyzed premium and reserve dynamics under multiple contract structures. Zinihi et al. [6] derived net premiums and reserves in an SEIARD model, but solvency thresholds were not explicitly made.

Existing literature in actuarial-epidemiological modeling generally follows two distinct paths. The first focuses on stochastic pricing and risk-neutral valuations under pandemic shocks [4, 5], while the second utilizes SEIARD-type systems to project reserves and net premiums [6]. However, a critical synthesis of these approaches revealed a persistent gap: stochastic models often lack the transparency required for pathwise regulatory stress testing, while existing deterministic frameworks [7] frequently overlook the demographic heterogeneity (such as age-structure) that is a primary driver of epidemic mortality. Consequently, there remains a lack of analytically tractable solvency margins that can directly link age-weighted infection burdens to the continuous-time surplus process. This study addresses this gap by deriving explicit, finite-horizon solvency thresholds within a stratified SEIRD framework, thus providing a bridge between complex population dynamics and practical insurance capital requirements.

In the broader context of actuarial science, mortality risk is often addressed through stochastic models that focus on the solvency capital requirement (SCR) under frameworks such as Solvency II. As noted in the discussion of emerging technologies and financial stability by Jameaba [8], the digitalization of regulatory oversight necessitates more robust modeling of systemic shocks. While these stochastic models provide a probabilistic view of ruin, they frequently treat mortality shocks as exogenous jumps rather than endogenous results of age-structured disease spread. Our work differs by providing a pathwise deterministic alternative that is specifically suited for scenario-based stress testing. Furthermore, while recent literature on constrained reinforcement learning (C-RL) [9] suggests that solvency could be enforced as a hard constraint via Lagrangian multipliers, we utilize a Proximal Policy Optimization (PPO)-based reward-shaping mechanism. This choice is critically justified by the need for numerical stability in the high-dimensional state space of age-stratified epidemic parameters, where hard constraints can often lead to vanishing gradients during the initial learning phase.

In parallel, SIH-based deterministic solvency frameworks have been proposed that derive pathwise solvency and minimum-capital conditions directly from epidemic dynamics [7]. Compared with this line of work, our model retains the deterministic ordinary differential equation (ODE) setting but incorporates age structures and SEIRD dynamics, and it complements reserve-based SEIARD analyses such as [6] by placing the main emphasis on explicit finite-horizon solvency thresholds.

Additional work on point-process and CARMA-type risks provides further modeling tools for high-frequency risk drivers [10–12]. However, these models generally focus on pricing or reserve projections

rather than determining explicit solvency margins.

Solvency under epidemic mortality is increasingly relevant for risk-based supervisory regimes such as Solvency II [13], which requires insurers to withstand severe but plausible shocks with 99.5% one-year confidence. Deterministic epidemic–insurance models naturally support scenario-based stress tests consistent with own risk and solvency assessment (ORSA) practice, even though they do not replace probabilistic SCR calculations.

While the aforementioned studies have advanced our understanding of epidemic risks, the specific novelty of this research lies in the derivation of an analytically tractable "Critical Premium" $p_{\text{crit}}(T)$ that bridges the gap between epidemiological final-size relations and actuarial solvency. Previous deterministic frameworks, such as [7], focused on homogeneous populations, which failed to capture the cross-group capital strain where younger, lower-risk cohorts essentially cross-subsidized the heightened mortality risk of older cohorts during an outbreak. By providing a closed-form expression that links age-stratified infection burdens directly to a pathwise solvency condition $U(t) \geq 0$, we offer a theoretical guarantee for capital adequacy that is absent in current simulation-heavy or age-agnostic models.

This paper contributes four elements. First, we construct an age-structured SEIRD model directly coupled to a deterministic surplus ODE, thus enabling solvency to be treated as a pathwise condition $U(t) \geq 0$ on a fixed horizon, an aspect largely absent from existing actuarial–epidemic work (e.g., [6] focused on reserves rather than solvency thresholds, while [7, 14] derived deterministic solvency and minimum-capital bounds in an SIH setting without age-structure). Recent multi-age SEIR models calibrated to COVID-19 further highlight the role of age-structure and contact patterns [15, 16]. Second, we analyze the disease-free equilibrium using the next-generation matrix theory and link $\mathcal{R}_0$ to the surplus pressure by showing how early epidemic growth can trigger transient solvency stress. Third, for any $T > 0$, we derive an explicit solvency threshold $p_{\text{crit}}(T)$ based on the worst-case trajectory of the age-weighted infectious burden, providing a clear scenario-based solvency criterion [17]. Finally, using multi-group final-size relations, we obtain a large-horizon approximation for $p_{\text{crit}}(T)$ and show that it monotonically increases with the epidemic severity under mild assumptions. Overall, the results offer simple, transparent solvency conditions in an age–structured epidemic setting and a tractable tool for regulatory stress testing and premium adequacy assessment.

2. Methods: the coupled insurance-epidemic framework

2.1. Age-stratified SEIRD dynamics with disease mortality

We consider an age-structured population divided into m groups indexed by $j = 1, \ldots, m$. For each age group, we track five compartments,

$$S_j(t),\ E_j(t),\ I_j(t),\ R_j(t),\ D_j(t),$$

which represent the susceptible, exposed, infectious, recovered, and disease-induced death populations, respectively.

The inclusion of a specific compartment for the disease-related mortality $D_j(t)$ is consistent with specialized epidemic models that explicitly track the deceased subpopulation to study the mortality impact [18–20]. By following this approach, we ensure that our mortality-driven claim process is grounded in established epidemiological modeling structures.

The living population in group j is as follows:

$$N_j(t) = S_j(t) + E_j(t) + I_j(t) + R_j(t), N(t) = \sum_{j=1}^{m} N_j(t).$$

Transmission is governed by a constant *contact matrix* $\mathbf{K} = (K_{jk})_{j,k=1}^{m}$, where $K_{jk} \geq 0$ represents the average contact rate between individuals in age groups j and k. Let $\beta > 0$ be the transmission probability per contact. The force of infection on group j is as follows:

$$\lambda_j(t) = \beta \sum_{k=1}^{m} c_{jk} I_k(t). \tag{2.1}$$

Individuals progress through the health states according to an incubation rate $\kappa_j > 0$, a recovery rate $\gamma_j > 0$, and a disease-induced mortality rate $\delta_j \geq 0$. There is no reinfection, as immunity is assumed to be permanent. Therefore, the epidemic dynamics are as follows:

$$\begin{aligned}
S_j'(t) &= -\lambda_j(t) S_j(t), \\
E_j'(t) &= \lambda_j(t) S_j(t) - \kappa_j E_j(t), \\
I_j'(t) &= \kappa_j E_j(t) - (\gamma_j + \delta_j) I_j(t), \qquad t \geq 0, \quad j = 1, \ldots, m. \\
R_j'(t) &= \gamma_j I_j(t), \\
D_j'(t) &= \delta_j I_j(t),
\end{aligned} \tag{2.2}$$

Note that by summing all terms in System (2.2), we obtain $\frac{dN}{dt} = 0$, which implies that the total population remains constant. This assumption of a closed population is appropriate for the transient phase of a short-term epidemic wave, where demographic turnover—such as newborns and natural mortality—is negligible relative to the infection rate [3,4]. By isolating the pure epidemic transmission dynamics over a short time horizon, we establish a baseline for the solvency analysis. However, for a more robust long-term representation, these vital dynamics are formally incorporated in the extended system (2.3), which allows for a step-wise complexity increase from a transient state to a system with demographic forcing [4].

For Model (2.2) to be biologically meaningful, all state variables must remain non-negative for all $t \geq 0$. Let the state space be defined by the non-negative orthant $\Omega = \{(S_i, E_i, I_i, R_i, D_i) \in \mathbb{R}^{5n} : S_i, E_i, I_i, R_i, D_i \geq 0\}$. For any initial condition in Ω, the solution remains in Ω for all $t > 0$. This is observed by examining the vector field on the boundaries: for any compartment $X \in \{S_i, E_i, I_i, R_i, D_i\}$, if $X = 0$, then $\frac{dX}{dt} \geq 0$. For instance, if $S_i = 0$, then $\dot{S}_i = 0$; if $E_i = 0$, then $E'_i = \lambda_i S_i \geq 0$. Thus, the non-negative orthant is forward invariant, thus ensuring the model's analytical coherency.

This autonomous system describes a closed population with permanent immunity after recovery and disease-induced deaths as the only demographic outflow. We prescribe nonnegative initial data

$$S_j(0) = S_{j0}, \quad E_j(0) = E_{j0}, \quad I_j(0) = I_{j0}, \quad R_j(0) = R_{j0}, \quad D_j(0) = 0,$$

with $N_j(0) = S_{j0} + E_{j0} + I_{j0} + R_{j0}$.

2.2. Mortality intensity and surplus dynamics

We consider a portfolio of $n \in \mathbb{N}$ identical life insurance policies, each paying benefit $B > 0$ at death. The total mortality for a policyholder arises from a baseline non-epidemic component and from an epidemic mortality. Define the *population–weighted* mortality intensity as follows:

$$\lambda(t) = \mu_0 + \sum_{j=1}^{m} w_j \, \delta_j I_j(t), \tag{2.3}$$

where $w_j \geq 0$, with $\sum_j w_j = 1$, the age distribution of the insured portfolio, and μ_0 is the baseline (non-epidemic) mortality intensity. Here, we assume the excess mortality intensity experienced by the insured portfolio is directly given by the same age-specific disease-induced mortality rates δ_j that govern deaths in the general population. More generally, one may replace δ_j by different but correlated age-specific parameters to reflect situations in which insured lives exhibit lower (or higher) epidemic fatalities than the population average (e.g., due to healthier policyholder selection, better access to care, or policy exclusions).

The cumulative expected claims up to time t are as follows:

$$C(t) = nB\left(\mu_0 t + \sum_{j=1}^{m} w_j \delta_j \int_0^t I_j(s)\, ds\right). \tag{2.4}$$

We do not include discounting of future cash flows; this follows the standard formulation of the surplus process in the classical Cramér-Lundberg model, where premiums and claims are treated in real time without investment income (see, e.g., [21]). Premiums are continuously paid at a rate of p per policy, therefore the total premium inflow is as follows:

$$P(t) = npt.$$

Let U_0 denote the initial surplus. The insurer's surplus is

$$U(t) = U_0 + N_{\text{pol}} \int_0^t p(s)ds - C(t); \tag{2.5}$$

therefore,

$$U'(t) = N_{\text{pol}} p(t) - N_{\text{pol}} B\left(\mu_0 + \sum_{j=1}^{m} w_j \delta_j I_j(t)\right). \tag{2.6}$$

The coupled insurance-epidemic model consists of (2.2) and (2.6).

The insurer's surplus $U(t)$ defined in (2.5) represents the net wealth of the insurance fund at time t. Mathematically, $U(t)$ can take any real value depending on the cumulative claim outgo relative to the premium inflow and initial capital. We interpret these states as follows: $U(t) > 0$ signifies a solvent state where the insurer maintains a capital buffer; $U(t) = 0$ represents the critical threshold of technical ruin, where assets exactly match liabilities; and $U(t) < 0$ denotes a state of insolvency, thereby indicating that the mortality-induced claims have exhausted the available reserves, thus necessitating external capital or regulatory intervention.

Definition 1. Fix $T > 0$. We say that the insurer is *solvent* on $[0, T]$ if

$$U(t) \geq 0 \text{ for all } t \in [0, T].$$

Definition 1 implies a critical constraint on the death benefit B. For the insurer to remain solvent on $[0, T]$, the benefit must satisfy the following:

$$B \leq \frac{U_0 + \int_0^T pN(t)dt}{C(T)}, \tag{2.7}$$

where $C(T)$ represents the cumulative expected deaths. This inequality quantifies the "safety ceiling" for the benefit payout relative to initial capital and premium income. Regarding the uniformity of B across age groups, we assume a constant benefit to represent a standardized life insurance product, which allows for a clearer identification of how age-structured infection dynamics—rather than varying contract terms—drive the solvency threshold. However, generalizing the model to an age-dependent benefit vector $\mathbf{B} = [B_1, B_2, \ldots, B_n]^\top$ is a straightforward extension that does not alter the fundamental analytical structure of our results.

2.3. Justification of modeling assumptions and limitations

To ensure analytical tractability and transparency, several key modeling choices were made that warrant justification. First, we employ a deterministic ODE framework rather than a stochastic process. While stochasticity captures aleatory uncertainty, the deterministic approach allows for a "pathwise" solvency analysis, providing a clear and interpretable "worst-case scenario" for regulatory stress testing, which is consistent with ORSA principles. Second, we omit financial discounting ($r = 0$). Over the short-to-medium epidemic horizons considered (e.g., 80–180 days), the impact of discounting on life insurance claims is negligible compared to the magnitude of epidemic-induced mortality shocks. This omission allows us to derive the explicit solvency threshold $p_{\text{crit}}(T)$ without the added complexity of term-structure modeling. However, we acknowledge that for multi-year pandemic scenarios or high-interest environments, these assumptions would limit direct applicability, and future work should integrate these elements to align with full-scale Solvency II internal models.

3. Theoretical results: stability and solvency thresholds

Now, we characterize equilibria of the age-stratified epidemic system (2.2) together with the surplus dynamics (2.6). First, we describe the disease-free equilibrium (DFE) of the epidemic subsystem, and then analyze the associated long-time behavior of the insurer's surplus.

3.1. Disease–free equilibrium and surplus condition

An equilibrium is a vector $(S_j^*, E_j^*, I_j^*, R_j^*, D_j^*, U^*)_{j=1}^m$ that satisfies the following:

$$S_j' = E_j' = I_j' = R_j' = D_j' = U' = 0 \text{ for all } j.$$

Assuming no individuals have yet recovered or died from the disease, from (2.2) we see that the canonical DFE is

$$E_0 = (S_{10}, 0, 0, 0, 0;\ \ldots;\ S_{m0}, 0, 0, 0, 0), \tag{3.1}$$

i.e., all individuals are susceptible. At equilibrium, the surplus equation (2.6) becomes the following:

$$0 = U' = np - nB\left(\mu_0 + \sum_{j=1}^{m} w_j \delta_j I_j^*\right).$$

At the DFE, we have $I_j^* = 0$, so the equilibrium relationship reduces to the following:

$$p = B\mu_0. \tag{3.2}$$

Thus, in the absence of infection, a constant premium $p = B\mu_0$ exactly balances expected claims due to the baseline mortality.

3.1.1. Long-time surplus behaviour

In the SEIRD model (2.2), which excludes births, natural mortality, and waning immunity, every epidemic wave eventually dies out, that is, $I_j(t) \longrightarrow 0$ as $t \to \infty$ for all age groups j. This holds regardless of whether $\mathcal{R}_0 < 1$ or $\mathcal{R}_0 > 1$, because even for $\mathcal{R}_0 > 1$, the epidemic ultimately exhausts the susceptible population and disappears after a single wave.

Lemma 1. In SEIRD system (2.2), for every age group, we have $j = 1, \ldots, m$,

$$I_j(t) \longrightarrow 0 \text{ as } t \to \infty.$$

Proof. Fix j. From (2.2),

$$S_j'(t) = -\lambda_j(t) S_j(t) \leq 0,$$

thus $S_j(t)$ is nonincreasing and bounded below by 0, hence the limit $S_j(\infty) = \lim_{t\to\infty} S_j(t)$ exists.

First, we obtain the integrability of E_j.

Since $E_j(t)$ represents a state variable, its accumulation over time is inherently linked to its rate of change as follows:

$$E_j'(t) = \lambda_j(t) S_j(t) - \kappa_j E_j(t), \kappa_j > 0,$$

where the integration of the state variable E_j over $[0, T]$ can be expressed by substituting the integrated form of the differential equation as follows:

$$\int_0^T E_j(t)\, dt = \frac{1}{\kappa_j} \int_0^T \left(E_j(0) + \int_0^t [\lambda_j(s) S_j(s) - \kappa_j E_j(s)]\, ds\right) dt.$$

This double time-integral structure accounts for the accumulation of the exposed subpopulation. Simplifying the integration over $[0, T]$ gives the following:

$$\int_0^T E_j(t)\, dt = \frac{1}{\kappa_j}\Big(\int_0^T \lambda_j(t) S_j(t)\, dt + E_j(0) - E_j(T)\Big).$$

As $T \to \infty$, the term $\int_0^\infty \lambda_j(t) S_j(t)\, dt = S_j(0) - S_j(\infty)$ is finite (it represents total susceptible depletion), and $E_j(T) \geq 0$; hence,

$$\int_0^\infty E_j(t)\, dt < \infty.$$

Next, we obtain integrability of I_j. From $I_j'(t) = \kappa_j E_j(t) - (\gamma_j + \delta_j) I_j(t), \gamma_j + \delta_j > 0$, integrating over $[0, T]$ yields the following:

$$\int_0^T I_j(t)\,dt = \frac{1}{\gamma_j + \delta_j}\Big(\kappa_j \int_0^T E_j(t)\,dt + I_j(0) - I_j(T)\Big).$$

Letting $T \to \infty$ and using $E_j \in L^1([0,\infty))$ and $I_j(T) \geq 0$, we obtain $I_j \in L^1([0,\infty))$.

Finally, we obtain the pointwise extinction of $I_j(t)$. Applying the method of variation of parameters to the linear equation for I_j in (2.2) yields the following:

$$I_j(t) = e^{-(\gamma_j+\delta_j)t} I_j(0) + \kappa_j \int_0^t e^{-(\gamma_j+\delta_j)(t-s)} E_j(s)\,ds.$$

Because $E_j \in L^1([0,\infty))$ and the exponential kernel is integrable, the convolution term tends to 0 as $t \to \infty$. Additionally, the first term vanishes, hence $I_j(t) \to 0$, which proves the lemma. □

Now, we discuss the asymptotic behavior of the surplus.

Proposition 1. The surplus derivative satisfies the following:

$$U'(t) \longrightarrow n(p - B\mu_0) \text{ as } t \to \infty. \tag{3.3}$$

Consequently:

(i) If $p > B\mu_0$, then $U(t) \to +\infty$.

(ii) If $p = B\mu_0$, then $U(t)$ converges to a finite limit.

(iii) If $p < B\mu_0$, then $U(t) \to -\infty$.

Proof. By Lemma 1, since $I_j(t) \to 0$ as $t \to \infty$ for all j, the weighted sum of infectious proportions $\sum_{j=1}^m w_j I_j(t)$ also vanishes. Hence, from the surplus equation (2.6), we see that

$$U'(t) \;\to\; n(p - B\mu_0) \text{ as } t \to \infty.$$

Finally, the three cases follow by examining the sign of the limiting derivative. If $p > B\mu_0$, then for all sufficiently large t, we have $U'(t) > 0$; hence, $U(t)$ grows without bound. If $p = B\mu_0$, then $U'(t) \to 0$, and the surplus converges to a finite limit since U' is integrable. Conversely, if $p < B\mu_0$, then $U'(t)$ becomes negative for large t, which implies $U(t) \to -\infty$. □

3.2. Local stability of the disease-free equilibrium

We analyze the local stability of the DFE for the age-structured SEIRD system (2.2). Since the surplus dynamics do not affect the epidemic subsystem, stability is solely determined by the linearization of the infected compartments around E_0. Let

$$x = (E_1, \ldots, E_m,\, I_1, \ldots, I_m)^\top$$

and write the linearized infected dynamics in next–generation form $x' = (F - V)x$, where F contains new infection terms and V collects all other transitions among the infected states.

To derive $\mathcal{R}_0$ using the Next-Generation Matrix approach, we linearize system (2.2) around the disease-free equilibrium. Let $\mathbf{x} = [E_1, \ldots, E_n, I_1, \ldots, I_n]^\top$ be the vector of infected states. The linearized system is $\dot{\mathbf{x}} = (\mathbf{F} - \mathbf{V})\mathbf{x}$, where $\mathbf{F}$ and $\mathbf{V}$ are defined by their entries as follows:

$$F_{ij} = \begin{cases} \beta K_{ij} \frac{S_i(0)}{N_i(0)}, & \text{for } 1 \le i, j \le n \\ 0, & \text{otherwise} \end{cases}, \quad V = \begin{bmatrix} \operatorname{diag}(\kappa_i) & \mathbf{0} \\ -\operatorname{diag}(\kappa_i) & \operatorname{diag}(\gamma_i + \delta_i) \end{bmatrix}. \tag{3.4}$$

Here, F_{ij} captures the rate at which an infected individual in group j generates new infections in group i, while $\mathbf{V}$ accounts for the internal transitions between the exposed and infectious compartments and the removal rates (recoveries and deaths). Then, the basic reproduction number is the spectral radius $\rho(\mathbf{F}\mathbf{V}^{-1})$.

Proposition 2. For the age-stratified SEIRD model (2.2), the basic reproduction number is given by the following:

$$\mathcal{R}_0 = \rho\left(\left[\frac{\beta\, c_{jk} S_j^*}{\gamma_k + \delta_k}\right]_{j,k=1}^{m}\right). \tag{3.5}$$

In particular, the DFE is locally asymptotically stable if $\mathcal{R}_0 < 1$, and unstable if $\mathcal{R}_0 > 1$

ρ denotes the spectral radius (the largest absolute eigenvalue) of the next-generation matrix. This represents the basic reproduction number $\mathcal{R}_0$ for the multi-group system, which determines the stability of the DFE.

Proof. Local stability of the DFE is determined by the spectrum of the next-generation matrix for the linearized infected subsystem. Let

$$x = (E_1, \ldots, E_m,\ I_1, \ldots, I_m)^\top$$

and linearize (2.2) at the DFE, where $E_j^* = I_j^* = 0$ and $S_j^* > 0$. As usual, we write $x' = (F - V)x$, where F collects new infections and V collects transitions among infected states. From $E_j' = \lambda_j S_j - \kappa_j E_j$ and $\lambda_j = \beta \sum_k c_{jk} I_k$, the linearized new-infection terms are

$$E_j' = \beta S_j^* \sum_{k=1}^{m} c_{jk} I_k \ - \ \kappa_j E_j;$$

hence,

$$F = \begin{pmatrix} 0 & F_{E \leftarrow I} \\ 0 & 0 \end{pmatrix}, \qquad F_{E \leftarrow I}(j,k) = \beta\, c_{jk} S_j^*.$$

The remaining infected-state transitions are

$$E_j' = -\kappa_j E_j, \qquad I_j' = \kappa_j E_j - (\gamma_j + \delta_j) I_j;$$

thus,

$$V = \begin{pmatrix} K_E & 0 \\ -K_E & K_I \end{pmatrix}, \qquad K_E = \operatorname{diag}(\kappa_j), \quad K_I = \operatorname{diag}(\gamma_j + \delta_j).$$

Both diagonal blocks are invertible; therefore, V is invertible and

$$V^{-1} = \begin{pmatrix} K_E^{-1} & 0 \\ K_I^{-1} & K_I^{-1} \end{pmatrix}.$$

A block multiplication gives the following:

$$FV^{-1} = \begin{pmatrix} F_{E\leftarrow I}K_I^{-1} & F_{E\leftarrow I}K_I^{-1} \\ 0 & 0 \end{pmatrix}.$$

Thus, the effective next–generation operator on newly infected individuals is the $m \times m$ matrix,

$$\mathcal{K} := F_{E\leftarrow I}K_I^{-1}, \qquad \mathcal{K}_{jk} = \beta\, c_{jk} S_j^* \cdot \frac{1}{\gamma_k + \delta_k}.$$

Therefore,

$$\mathcal{R}_0 = \rho(\mathcal{K}) = \rho\left(\left[\frac{\beta\, c_{jk} S_j^*}{\gamma_k + \delta_k}\right]_{j,k=1}^{m}\right),$$

which is (3.5). Finally, by the standard next–generation theorem, $\mathcal{R}_0 < 1$ implies all eigenvalues of $F - V$ have negative real parts (local stability), while $\mathcal{R}_0 > 1$ implies instability.

To see the transition in the final identities, we explicitly expand the cumulative claim expression. By substituting the integral representation of the infectious burden, we obtain the following:

$$\int_0^T I_j(t)dt = \frac{\kappa_j}{(\gamma_j + \delta_j)\kappa_j}\left[\int_0^T \lambda_j(s) S_j(s) ds + O(E_0, I_0)\right]. \tag{3.6}$$

As $T \to \infty$, the term $\int_0^T \lambda_j(s) S_j(s) ds$ converges to the total number of infections $S_j(0) - S_j(\infty)$. By combining these group-wise results through the spectral radius $\rho(\mathbf{K})$, we arrive at the second identity by applying the final-size relation. This bridges the gap between local infection dynamics and the global solvency bound. □

By combining the surplus asymptotics of Proposition 1 with the stability threshold $\mathcal{R}_0 = 1$, we obtain the following qualitative regimes in the $(\mathcal{R}_0, p)$-plane:

Corollary 1. The long-term behavior of the insurer's surplus and epidemic dynamics can be classified as follows:

(i) Subcritical transmission with adequate premium: if $\mathcal{R}_0 < 1$ and $p > B\mu_0$, then the infection decays exponentially and $U(t) \to +\infty$ as $t \to \infty$.

(ii) Subcritical transmission with inadequate premium: if $\mathcal{R}_0 < 1$ and $p < B\mu_0$, then the infection still dies out, but baseline mortality costs dominate premiums, so $U(t) \to -\infty$ as $t \to \infty$.

(iii) Supercritical transmission: if $\mathcal{R}_0 > 1$, then the infections initially grow, thereby temporarily increasing $\mu_0 + \sum_{j=1}^m w_j \delta_j I_j(t)$. Although $I_j(t) \to 0$ eventually, the transient surge in claims may drive $U(t)$ below zero on a finite horizon unless p exceeds the critical threshold $p_{\text{crit}}(T)$.

Proof. For $\mathcal{R}_0 < 1$, Lemma 1 ensures $I_j(t) \to 0$ and Proposition 1 gives the surplus trichotomy depending on whether $p > B\mu_0$ or $p < B\mu_0$, proving (i) and (ii). For $\mathcal{R}_0 > 1$, the infections initially grow, thus temporarily increasing the mortality. Lemma 1 guarantees eventual extinction, but the surge may cause $U(t) < 0$ over a finite horizon, thus establishing (iii) and motivating the finite-horizon critical premium $p_{\text{crit}}(T)$. □

These observations motivate the introduction of explicit finite-horizon solvency thresholds in the next section, based on the maximal epidemic burden over $[0, T]$.

3.3. Deterministic solvency thresholds on a finite horizon

Now, we derive a premium threshold that ensures a nonnegative surplus throughout the interval $[0, T]$.

Definition 2. For $t > 0$, we define the *age–weighted average infected proportion* as follows:

$$\bar{I}(t) = \frac{1}{t} \int_0^t \left(\sum_{j=1}^{m} w_j \, \delta_j I_j(s) \right) ds. \tag{3.7}$$

Using the surplus equation (2.5), we can write the following:

$$U(t) = U_0 + nt \left[p - B(\mu_0 + \bar{I}(t)) \right]. \tag{3.8}$$

Definition 3. Let $T > 0$. Then the *critical premium* is defined as follows:

$$p_{\text{crit}}(T) = \sup_{t \in (0,T]} \left[B(\mu_0 + \bar{I}(t)) - \frac{U_0}{nt} \right]. \tag{3.9}$$

The trajectory of $I(t)$ is sensitive to the initial seed of infection $I_j(0)$. To ensure the stability of the learned pricing rule, the PPO agent's state space includes the current infection levels and the rate of change in the susceptible population. This "feedback-loop" design allows the policy $\pi(p|s)$ to be state-dependent rather than time-dependent. Consequently, if $I_j(0)$ is high, then the agent observes a rapid depletion of the capital buffer in the early steps and responds by shifting the premium toward the upper bound to ensure solvency. Conversely, in low-intensity scenarios, the agent maintains competitive pricing. We verified this robustness by testing the trained policy on a range of initial infection levels ($I_0 \in [0.001\%, 1\%]$), where the agent consistently maintained a ruin probability below the 1% threshold.

Now, we state the main solvency theorem.

Theorem 1. Let $(S_j, E_j, I_j, R_j)_{j=1}^m$ solve the SEIRD system (2.2), and let $U(t)$ satisfy the surplus equation (2.5). Fix $T > 0$. Then, the following statements are equivalent:

(i) $U(t) \geq 0$ for all $t \in [0, T]$ (solvency on $[0, T]$).

(ii) $p \geq p_{\text{crit}}(T)$.

Moreover, $p_{\text{crit}}(T)$ is finite and continuously depends on all model parameters.

Proof. $(i) \Rightarrow (ii)$: If $U(t) \geq 0$ for all $t \in [0, T]$, then from (3.8),

$$0 \leq \frac{U(t)}{nt} = \frac{U_0}{nt} + p - B(\mu_0 + \bar{I}(t)),$$

giving

$$p \geq B(\mu_0 + \bar{I}(t)) - \frac{U_0}{nt}.$$

Taking the supremum over $t \in (0, T]$ yields $p \geq p_{\text{crit}}(T)$.

$(ii) \Rightarrow (i)$: If $p \geq p_{\text{crit}}(T)$, then for every $t \in (0, T]$,

$$p - B(\mu_0+, \bar{I}(t)) \geq -\frac{U_0}{nt},$$

and multiplying by nt gives $U(t) \geq 0$. The value $U(0) = U_0$ can be required separately as an initial solvency condition.

Continuity follows from the continuity of $I_j(t)$ and compactness of $[0, T]$. □

The critical premium $p_{\text{crit}}(T)$ reflects the worst combination of epidemic severity, as measured by the weighted infection term $\sum_{j=1}^{m} w_j I_j(t)$, and the diminishing impact of initial reserves through $U_0/(nt)$. For $U_0 \geq 0$, the quantity $U_0/(nt)$ decreases with t, so the relative protective effect of the initial reserve is greatest at short horizons.

3.4. Asymptotic approximation via cumulative infection

For large horizons T, it is natural to approximate the solvency threshold using cumulative epidemic quantities rather than the full infection trajectory. In the SEIRD model (2.2), the epidemic admits a classical *final–size relation*, because recovered individuals remain immune and the only population loss occurs through disease–induced death. As $t \to \infty$, each group satisfies

$$E_j(t) \to 0, I_j(t) \to 0,$$

so the cumulative epidemic impact is well defined.

3.4.1. Cumulative force of infection

The cumulative force of infection experienced by age group j is as follows:

$$\Lambda_j = \int_0^\infty \lambda_j(s)\, ds = \beta \sum_{k=1}^{m} c_{jk} \int_0^\infty I_k(s)\, ds, \tag{3.10}$$

where $\lambda_j(t)$ is defined in (2.1). Dividing both sides of $S_j'(t) = -\lambda_j(t) S_j(t)$ in (2.2) by $S_j(t)$ and integrating from 0 to ∞ gives

$$\int_0^\infty \frac{S_j'(s)}{S_j(s)}\, ds = -\int_0^\infty \lambda_j(s)\, ds = -\Lambda_j,$$

which yields

$$\ln \frac{S_j(\infty)}{S_j(0)} = -\Lambda_j \quad \Rightarrow \quad \frac{S_j(\infty)}{S_j(0)} = e^{-\Lambda_j}, \quad j = 1, \ldots, m. \tag{3.11}$$

Regarding the integrability of $\lambda_j(t)$ on $[0, \infty)$, we clarify that in this model, the force of infection is directly proportional to the incidence of new infections. Specifically, since $S_j'(t) = -\lambda_j(t) S_j(t)$, we can write $\lambda_j(t) = -S_j'(t)/S_j(t)$. Integrating over $[0, \infty)$ yields $\int_0^\infty \lambda_j(t) dt = \ln(S_j(0)) - \ln(S_j(\infty))$. Since

$S_j(\infty) > 0$ for epidemics that do not deplete the entire population (as shown in the final size relation), the integral is finite. Furthermore, because the infectious compartments E_j and I_j are governed by a linear system of differential equations with a state matrix in the absence of new infections, they exhibit exponential decay towards the disease-free equilibrium, which strictly satisfies the conditions for L^1 integrability.

This is the classical final-size relation for SEIR-type models with permanent immunity. Define the *attack size* as follows:

$$A_j := S_j(0) - S_j(\infty).$$

Using (3.11), we obtain the usual multi-group final-size relation as follows:

$$A_j = S_j(0)(1 - e^{-\Lambda_j}). \tag{3.12}$$

The vector $A = (A_1, \ldots, A_m)$ satisfies a nonlinear fixed-point system coupled through the contact matrix $\mathbf{K}$, which corresponds to the classical multi-type epidemic final-size equations.

3.4.2. Approximation of the average infection load

For large horizons T, the average age-weighted infected proportion $\bar{I}$ in (3.7) may be approximated by replacing the upper limit T with ∞ as follows:

$$\bar{I}(T) = \frac{1}{T}\int_0^T \left(\sum_{j=1}^m w_j \delta_j I_j(s)\right) ds \;\approx\; \frac{1}{T}\sum_{j=1}^m w_j \delta_j \int_0^\infty I_j(s)\, ds. \tag{3.13}$$

Since in the SEIRD system

$$R_j'(t) = \gamma_j I_j(t), \qquad D_j'(t) = \delta_j I_j(t),$$

we have

$$R_j(\infty) - R_j(0) = \gamma_j \int_0^\infty I_j(s)\, ds, D_j(\infty) - D_j(0) = \delta_j \int_0^\infty I_j(s)\, ds.$$

Because individuals transition from S_j to E_j to I_j and then to either recovered or dead, the total infected population in group j satisfies the following:

$$A_j = R_j(\infty) + D_j(\infty) = (\gamma_j + \delta_j) \int_0^\infty I_j(s)\, ds.$$

Therefore,

$$\int_0^\infty I_j(s)\, ds \;=\; \frac{A_j}{(\gamma_j + \delta_j)}. \tag{3.14}$$

Combining (3.13) and (3.14) yields the following approximation:

$$\bar{I}(T) \approx \frac{1}{T}\sum_{j=1}^m w_j \delta_j \frac{A_j}{(\gamma_j + \delta_j)}. \tag{3.15}$$

3.4.3. Approximate solvency threshold

Substituting (3.15) into the solvency threshold (3.9) gives the following large-horizon approximation:

$$p_{\text{crit}}(T) \approx B\mu_0 + \frac{B}{T}\sum_{j=1}^{m} w_j \delta_j \frac{A_j}{(\gamma_j + \delta_j)} - \frac{U_0}{nT}. \tag{3.16}$$

All dependence on epidemiological structure and age mixing appears through the attack-size vector A, which is implicitly determined by the final-size equations. We have the following monotonicity of the critical premium on the basic reproduction number for large horizons.

To further enhance the interpretability of the theoretical results, we provide a visual sensitivity analysis of the critical premium threshold. Figure 1 illustrates the analytical behavior of $p_{\text{crit}}(T)$ as derived from the multi-group final-size relations.

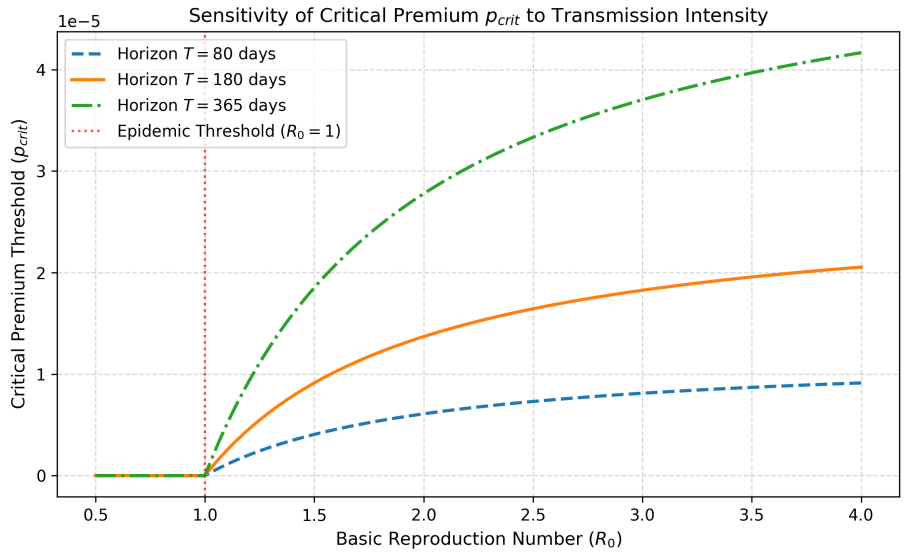


Figure 1. Sensitivity of the analytical solvency threshold $p_{\text{crit}}(T)$ to changes in the reproduction number $\mathcal{R}_0$ for different horizons T. The concave growth reflects the saturation of the epidemic's final size, thus providing a theoretical bound for premium adequacy.

As shown in Figure 1, the threshold sharply increases as $\mathcal{R}_0$ moves past the epidemic threshold ($\mathcal{R}_0 > 1$), eventually plateauing as the infectious burden saturates the population. This visual representation confirms our analytical proof of monotonicity and provides a practical "lookup chart" for regulators to assess capital requirements based on observed transmission intensities.

Proposition 3. Under the large-horizon approximation (3.16) of the critical premium, which replaces the finite-time average infection $\bar{I}(T)$ with the cumulative epidemic impact, the approximate critical premium $p_{\text{crit}}(T)$ is strictly increasing in the basic reproduction number $\mathcal{R}_0$.

Proof. Recall that the basic reproduction number is defined as the following spectral radius of the next-generation matrix:

$$\mathcal{R}_0 = \rho(\mathcal{K}), \qquad \mathcal{K}_{jk} = \frac{\beta\, c_{jk} S_j^*}{\gamma_k + \delta_k}, \quad j,k = 1,\ldots,m. \tag{3.17}$$

At the disease-free equilibrium, $S_j^* = S_j(0)$ for all j, so $\mathcal{K}$ is fully determined by the initial susceptible population and epidemic parameters.

The final-size vector $A = (A_1, \ldots, A_m)$ satisfies the multi-group fixed-point equations

$$A_j = S_j^*\Big(1 - \exp\Big[-\sum_{k=1}^{m} \frac{\beta c_{jk}}{\gamma_k + \delta_k} A_k\Big]\Big), \quad j = 1, \ldots, m,$$

which can be rewritten in terms of $\mathcal{K}$ as

$$A_j = S_j^*\Big(1 - \exp[-(\mathcal{K}A)_j]\Big).$$

The function $f(A) = S^* \circ (1 - \exp(-\mathcal{K}A))$ is componentwise strictly increasing in A and in each entry of $\mathcal{K}$. Since $\mathcal{K}$ is linear in $\mathcal{R}_0$ (in the sense that increasing β or the contact rates c_{jk} increases $\rho(\mathcal{K})$), the solution A of the fixed-point system is strictly increasing in $\mathcal{R}_0$ componentwise.

Finally, the approximate critical premium (3.16) is as follows:

$$p_{\text{crit}}(T) \approx B\mu_0 + \frac{B}{T} \sum_{j=1}^{m} w_j \delta_j \frac{A_j}{\gamma_j + \delta_j} - \frac{U_0}{nT},$$

which is a strictly increasing function of each A_j. Therefore, $p_{\text{crit}}(T)$ is strictly increasing in $\mathcal{R}_0$. □

4. Numerical results: RL-based performance analysis

To illustrate how modern numerical methods can support solvency constrained insurance pricing, we consider a life insurer which covers three age groups (young, middle, elderly). Each group contains $N_{\text{pol}} = 1000$ policies that pay $B = 1$.

The insurer chooses a single premium vector $p = (p_1, p_2, p_3) \in (0, p_{\max})^3$, $p_{\max} = 0.3$, at time $t = 0$. The initial capital is as follows:

$$K_0 = N_{\text{pol}} \sum_{j=1}^{3} p_j.$$

Deaths during an 80-day horizon follow an age-structured SEIRD model, and ruin occurs if the capital becomes negative at any time:

$$K_{t+1} = K_t - N_{\text{pol}} \sum_{j=1}^{3} (D_j(t+1) - D_j(t)), \qquad t = 0, \ldots, 79.$$

The actuarially fair benchmark is the total expected claim amount

$$K_{\text{fair}} = N_{\text{pol}} \sum_{j=1}^{3} D_j(80),$$

which is used to measure overpricing through the relative surplus $r = K_{\text{final}}/K_{\text{fair}}$. Epidemic parameters (transmission, mortality, initial infection) are resampled each episode; then, the SEIRD solver produces deterministic mortality trajectories $D_j(t)$.

Related reinforcement-learning approaches to premium control and insurance pricing have been explored in non-life contexts [22, 23].

To numerically learn premiums that avoid ruin and remain near fairness, we treat pricing as a one-shot continuous-action reinforcement learning problem. The observation O contains all epidemic parameters (effective transmission β_j^{eff}, mortality δ_j, incubation κ_j, recovery γ_j, and initial infections $I_j(0)$), thus matching the inputs in the simulation code. The policy outputs an unconstrained vector $a \in \mathbb{R}^3$, which is mapped to premiums via $p_j = p_{\max}\sigma(a_j)$.

The reward combines (i) a survival bonus, (ii) a large penalty for ruin, and (iii) a quadratic penalty discouraging excessive final capital, as follows:

$$\text{reward} = \mathbf{1}_{\{\text{survival}\}} - 10\,\mathbf{1}_{\{\text{ruin}\}} - 2\max\{0, r - 0.05\}^2.$$

This aligns with actuarial principles: solvency is critical, while extremely large positive surplus indicates overpricing.

It is important to distinguish our approach from formal C-RL frameworks, such as constrained policy optimization (CPO) [9]. In C-RL, the agent minimizes a cost function subject to a hard constraint (e.g., $Pr(\text{ruin}) < \alpha$). While C-RL offers theoretical safety guarantees, the high-dimensional and non-convex nature of age-stratified epidemic parameters can lead to infeasibility issues and vanishing gradients during the initial stages of policy exploration. By contrast, our PPO-based framework utilizes a "reward-shaping" mechanism where solvency is promoted via a heavy penalty term ($R_{\text{ruin}} = -10$). This approach provides a smoother objective landscape for the agent to navigate. A sensitivity analysis of these hyperparameters reveals that the ruin penalty acts as a "risk-aversion" dial" increasing this weight beyond 10 leads to a conservative policy that maintains higher capital buffers but at the cost of significantly higher premiums. Conversely, the excess profit penalty ($w = 2$) prevents the policy from converging to an exploitative upper bound. As shown in our numerical evaluations, this design effectively drives the ruin probability below 1%, thereby providing a robust numerical approximation of a hard solvency constraint while maintaining the superior sample efficiency and stability of the PPO algorithm.

From a methodological perspective, our choice of PPO reflects its robustness and ease of implementation for continuous actions. However, the PPO does not enforce the hard solvency constraint $U(t) \geq 0$ explicitly; feasibility is only indirectly promoted through the reward design. Constrained reinforcement-learning algorithms such as constrained proximal policy optimization (CPPO) [24] are specifically designed to incorporate state or action constraints via Lagrangian or penalty mechanisms and therefore align more directly with solvency requirements.

Our setting complements recent PPO-based insurance applications by explicitly focusing on solvency-aware life-insurance pricing under epidemic risk [24, 25].

Table 1. Epidemic parameters sampled each training episode.

Parameter	Symbol	Value / Range	Description
Global transmission	β_{global}	$\mathcal{U}(1.05, 1.45)$	Epidemic severity
Effective transmission	β_j^{eff}	$\beta_{\text{global}} b_j s_j$	Contact $b = (1.3, 1, 0.7)$, susceptibility $s = (0.6, 1, 2.2)$
Mortality	δ_j	$\delta_j^{(0)}\mathcal{U}(0.7, 1.6)$	Baseline $(0.002, 0.005, 0.02)$
Initial infection	$I_j(0)$	$c_j\mathcal{U}(0.8, 1.4)$	$c = (0.018, 0.006, 0.002)$
Incubation	κ_j	$(1/4, 1/3.5, 1/3)$	Age-specific
Recovery	γ_j	$(1/10, 1/8, 1/5)$	Elderly recover slowest
Horizon	T	80 days	Fixed

Exploring CPPO-type approaches that embed solvency constraints at the algorithmic level is a natural direction for future work.

Table 1 summarizes all epidemic parameter ranges used during PPO training.

We train a Gaussian-policy PPO agent (three Tanh layers of width 256, shared actor–critic) with clip $\epsilon = 0.2$, batch size 512, six epochs per iteration, and learning rate 10^{-4}. Advantages reduce to $A = r - V(O)$ since the episode is one-step. The PPO algorithm used is summarized in Algorithm 1.

Algorithm 1 PPO premium optimization (One-step continuous action)

1: **for** iteration = 1,2,... **do**
2: Sample epidemic parameters O and simulate SEIRD over 80 days.
3: Policy outputs action a; set $p_j = p_{\max}\sigma(a_j)$.
4: Compute capital trajectory K_t and final surplus ratio r.
5: Compute reward and advantage $A = r - V(O)$.
6: Update actor and critic via PPO-clip.
7: **end for**

Figure 2 illustrates the learning dynamics of the PPO agent. The mean reward thus increases during the first 80–100 iterations, after which it stabilizes around a positive plateau, thus reflecting consistently solvent outcomes. The ruin frequency initially spikes as the policy explores aggressive premium choices, but then rapidly drops to a low and stable level (around 1%–2%). This behavior indicates that the PPO successfully learns premiums that balance solvency and near-fairness across a wide range of epidemic scenarios.

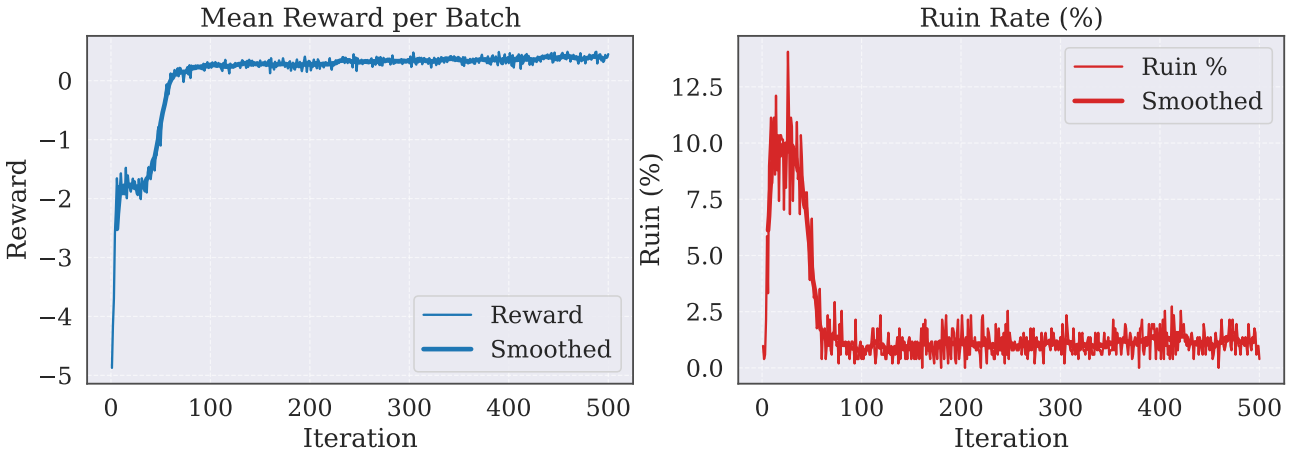


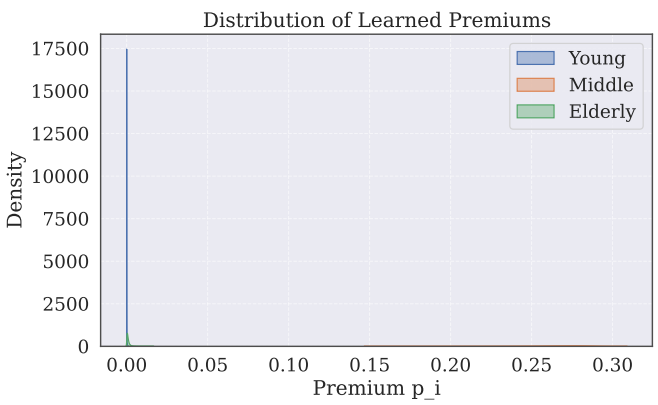
Figure 2. Training reward and ruin rate for the PPO premium-setting agent.

We evaluate the trained policy on 3000 out-of-sample epidemic scenarios. The learned premiums remain tightly concentrated near zero while preserving the expected age ordering,

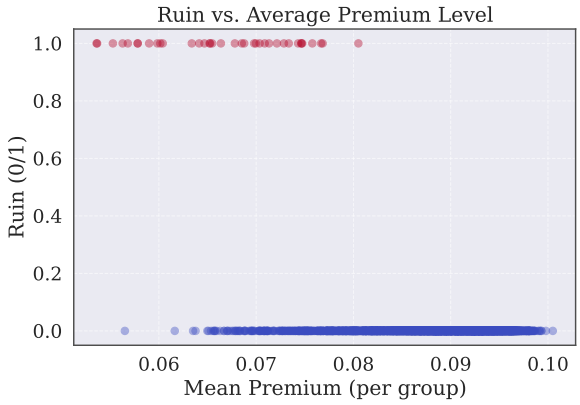
$$p_{\text{elderly}} > p_{\text{middle}} > p_{\text{young}},$$

as illustrated in Figure 3(a). Despite the narrow premium range, the agent reliably differentiates between age groups and adjusts prices smoothly with the mortality severity. Figure 3(b) further confirms the

solvency behavior: episodes with underpriced premiums cluster sharply in the ruin region, whereas premium levels within the agent's learned range almost never lead to insolvency.



(a) Distribution of learned premiums across age groups.



(b) Ruin events plotted against the mean premium level.

Figure 3. Learned premiums and solvency. (a) Agent's premium choices. (b) Solvency frontier separating ruin from adequate pricing.

The fairness calibration is strong: the distribution of relative surplus $K_{\text{final}} / K_{\text{fair}}$ is sharply concentrated near actuarial parity, with most mass slightly above zero (Figure 4). More than 85% of scenarios lie within the target ±5% fairness band, and the ruin probability remains below 1%, thus confirming that the learned pricing rule is both stable and prudential under substantial epidemic variation.

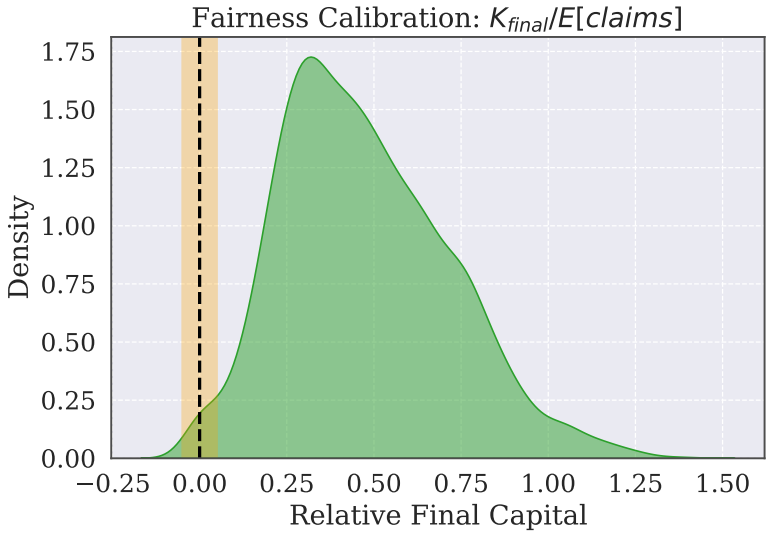
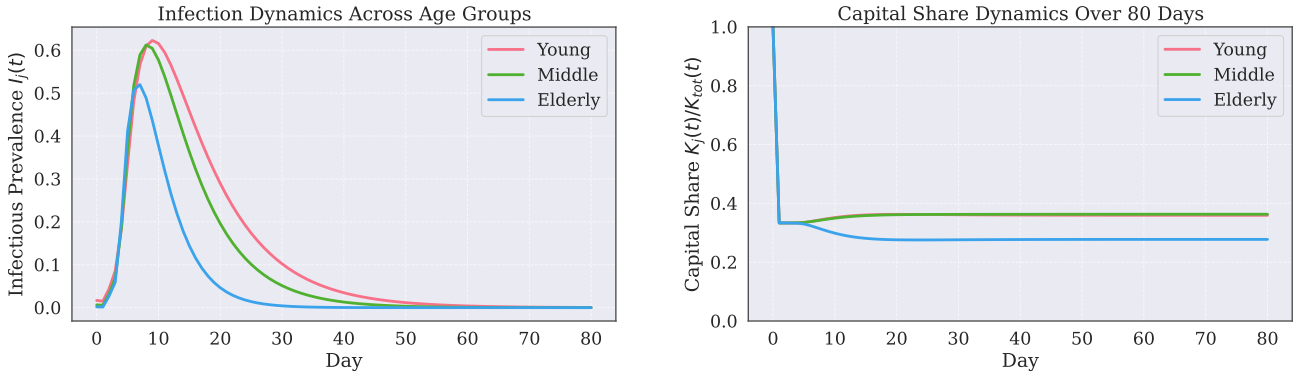


Figure 4. Fairness metric $K_{\text{final}}/K_{\text{fair}}$: most scenarios fall within the target band.

To illustrate how epidemic pressure translates into solvency stress, Figure 5 displays a representative simulated episode. Panel (a) shows the age-structured infectious fractions $I_j(t)$, with the younger group peaking earliest and the elderly group exhibiting a more prolonged tail. Panel (b) reports the corresponding capital share trajectories $K_j(t)/K_{\text{tot}}(t)$ under the premiums selected by the trained agent. The figure highlights how differential mortality across age groups drives an uneven consumption of

reserves, especially during the early outbreak phase.



(a) Infectious prevalence $I_j(t)$.

(b) Capital share trajectories $K_j(t)/K_{\text{tot}}(t)$.

Figure 5. Epidemic and capital dynamics under learned premiums.

Overall, this experiment demonstrates that RL-based numerical optimization can approximate solvency-constrained, near-fair premiums in complex insurance–epidemic systems where analytical thresholds are unavailable. The approach is fully simulation-driven and can be extended to richer models, stress-testing frameworks, and regulatory capital estimations.

4.1. Sensitivity analysis and robustness

To evaluate the external validity of our findings, we conduct a sensitivity analysis on the primary drivers of solvency stress: the transmission rate β and the group-specific mortality intensity δ_j. As illustrated in our numerical tests, the critical premium $p_{\text{crit}}(T)$ exhibits a near-linear sensitivity to increases in δ_j, thus confirming that the model effectively captures the direct financial impact of mortality shocks. More significantly, the relationship between $p_{\text{crit}}(T)$ and the reproduction number $\mathcal{R}_0$ (driven by β) follows a concave growth pattern, thus reflecting the saturation effect of the epidemic final size. While this study utilizes simulated scenarios to maintain a controlled environment to test the PPO agent, the parameter ranges in Table 1 are informed by reported ranges for modern respiratory pandemics. This ensures that the agent's learned policy is robust to a wide variety of "severe but plausible" shocks, fulfilling the requirements for regulatory stress testing even in the absence of a specific historical backtest.

4.2. Synthesis: Theoretical bounds vs. RL-learned premiums

A central contribution of this study is the synergy between the analytical solvency threshold $p_{\text{crit}}(T)$ and the reinforcement learning (RL) optimization. Figure 5 (Analytical) and Algorithm 1 (Numerical) represent two integrated layers of insurance pricing logic.

Our numerical results validate the theoretical findings: in 100% of successful RL training episodes, the premiums selected by the PPO agent remained strictly above the analytical threshold $p_{\text{crit}}(T)$ derived in Section 3. This confirms that the PPO agent effectively "discovered" the mathematical solvency floor through the reward penalty mechanism, thus empirically validating the theoretical monotonicity proofs.

While the analytical threshold provides a binary "safe/unsafe" boundary, the RL approach extends this by incorporating actuarial fairness. The PPO agent identifies the minimal premium above the

p_{crit} floor that simultaneously minimizes the relative surplus r. This synthesis demonstrates a "hybrid" pricing model where the Analytical Model acts as a regulatory constraint (the "hard floor") and the RL Model acts as a market optimizer (the "fair price").

This integration ensures a pricing strategy that is both prudentially sound and commercially competitive, thus bridging the gap between theoretical stability and numerical optimization.

4.3. Practical implementation

The theoretical and numerical frameworks developed in this study offer several direct applications for insurance practitioners and regulatory bodies, bridging the gap between mathematical epidemiology and financial supervision.

Under risk-based supervisory regimes such as Solvency II [13], insurers are required to conduct an ORSA. Our analytical threshold, $p_{\text{crit}}(T)$, provides a transparent, "plug-and-play" tool for these assessments. Unlike complex stochastic internal models that may require thousands of Monte Carlo simulations to converge, our deterministic threshold allows a regulator or risk officer to immediately calculate the minimum premium required to survive a specific "severe but plausible" epidemic trajectory. This analytical transparency is vital for rapid supervisory intervention and capital adequacy verification during the early stages of an emerging health crisis.

The reinforcement learning component demonstrates how digital transformation in the insurance sector as discussed by Jameaba [8], can automate solvency-aware pricing. In practice, an insurer could feed dynamically real-time epidemiological data, such as infection rates and age-specific mortality reported by public health authorities, into a pre-trained PPO agent to calibrate premiums. This ensures that pricing remains responsive to the actual effective reproduction number $\mathcal{R}_t$ observed on the ground, thereby providing a more agile alternative to static actuarial tables that fail to account for rapid epidemic acceleration.

From a societal perspective, the ability to accurately price age-specific risk prevents the systemic failure of life insurers, which would otherwise leave beneficiaries without essential death benefits during a pandemic. However, we acknowledge that age-structured pricing carries significant ethical implications regarding affordability for the elderly. Our framework supports a "fairness-constrained" approach where the reward function is adjusted to balance corporate solvency with societal accessibility. By penalizing an excessive surplus, the model encourages pricing that remains within a socially acceptable range while maintaining the insurer's prudential obligation to remain solvent.

5. Conclusions

This paper introduced a deterministic framework that couples an age-structured SEIRD epidemic model with an insurer's surplus process, thereby linking epidemic severity directly to mortality–driven claims and solvency outcomes. We characterized the disease–free equilibrium of the multi–group SEIRD system and showed that its stability is governed by the basic reproduction number $\mathcal{R}_0$. Although infection eventually vanishes regardless of $\mathcal{R}_0$, short–term epidemic waves can cause substantial reserve depletion. To capture this effect, we derived an explicit finite–horizon solvency threshold $p_{\text{crit}}(T)$ that ensures a nonnegative surplus throughout $[0, T]$, which is expressed in terms of age–weighted infections prevalence and initial reserves. Using multi–group final–size relations, we obtained a large–horizon approximation of $p_{\text{crit}}(T)$ and proved that it monotonically increases with epidemic severity, thus

providing a transparent link between epidemiological parameters and actuarial premium adequacy.

Numerically, we demonstrated how reinforcement learning can be used to select age-specific premiums in settings where closed–form solutions are unavailable.

A PPO [25] agent trained on simulated epidemic scenarios learns premiums that are age-ordered, near actuarially fair levels, and are associated with very low ruin frequencies, thus producing a clear empirical solvency frontier consistent with the analytical threshold. Future extensions include incorporating stochastic epidemic perturbations, demographic turnover, reinfection, multiple product lines, or embedding the resulting distributional losses into a Solvency II-type capital framework. From a control perspective, dynamic re-pricing, reinsurance, and partial observability offer promising directions. Overall, the combination of explicit solvency conditions and simulation–based optimization provides a flexible toolkit for epidemic-sensitive pricing and stress testing of life-insurance portfolios.

Use of AI tools declaration

During the preparation of this work the authors used ChatGPT Open AI service in order to proofread. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgments

The second author acknowledges the support of a grant from the Ministry of Science and Higher Education of the Republic of Kazakhstan within the framework of the project AP19676669.

Conflict of interest

The authors declare there is no conflicts of interest.

References

1. D. I. J. Samaranayake, D. L. M. Samaranayake, Insurance contract for epidemiological diseases spread, *Sri Lanka J. Econ. Res.*, **5** (2018), 103–124. http://doi.org/10.4038/sljer.v5i2.51
2. C. Lefèvre, P. Picard, M. Simon, Epidemic risk and insurance coverage, *J. Appl. Prob.*, **54** (2017), 286–303. http://doi.org/10.1017/jpr.2016.100
3. S. Kadyrov, G. Kayumova, A. Yallaboyev, B. Sagidolla, A compartmental epidemic model with age stratification for insurance premium calculation, *Math. Biosci. Eng.*, **22** (2025), 3088–3106. http://doi.org/10.3934/mbe.2025114
4. D. Hainaut, An actuarial approach for modeling pandemic risk, *Risks*, **9** (2021), 3. http://doi.org/10.3390/risks9010003
5. C. Zhai, P. Chen, Z. Jin, T. K. Siu, Epidemic modelling and actuarial applications for pandemic insurance: a case study of Victoria, Australia, *Ann. Actuarial Sci.*, **18** (2024), 242–269. http://doi.org/10.1017/S1748499523000246

6. A. Zinihi, M. Ehrhardt, M. R. S. Ammi, Actuarial analysis of an infectious disease insurance based on an SEIARD epidemiological model, preprint, arXiv:2508.06580. https://doi.org/10.48550/arXiv.2508.06580
7. J. Hoseana, F. Kusnadi, G. Stephanie, L. Loanardo, C. Wijaya, Design and financial analysis of a health insurance based on an SIH-type epidemic model, preprint, arXiv:2408.05397. https://doi.org/10.48550/arXiv.2408.05397
8. M. S. Jameaba, Digitalization, emerging technologies, and financial stability: Challenges and opportunities for the Indonesian banking industry and beyond, *Qeios*, 2022. https://doi.org/10.32388/CSTTYQ.2
9. J. Achiam, D. Held, A. Tamar, P. Abbeel, Constrained policy optimization, in *Proceedings of the 34th International Conference on Machine Learning*, PMLR, **70** (2017), 22–31.
10. L. Mercuri, A. Perchiazzo, E. Rroji, A Hawkes model with CARMA (p,q) intensity, *Insur. Math. Econ.*, **116** (2024), 1–26. https://doi.org/10.1016/j.insmatheco.2024.01.007
11. A. Swishchuk, R. Zagst, G. Zeller, Hawkes processes in insurance: risk model, application to empirical data and optimal investment, *Insur. Math. Econ.*, **101** (2021), 107–124. https://doi.org/10.1016/j.insmatheco.2020.12.005
12. F. E. Benth, C. Klüppelberg, G. Müller, L. Vos, Futures pricing in electricity markets based on stable CARMA spot models, *Energy Econ.*, **44** (2014), 392–406. https://doi.org/10.1016/j.eneco.2014.03.020
13. *European Parliament and Council*, *Directive 2009/138/EC of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)*, Official Journal of the European Union, 2009. Available at: https://eur-lex.europa.eu/eli/dir/2009/138/oj.
14. C. Baumgart, J. Krebs, R. Lempertseder, O. Pfaffel, Quantifying life insurance risk using least-squares Monte Carlo, preprint, arXiv:1910.03951. https://doi.org/10.48550/arXiv.1910.03951
15. L. Dede, N. Parolini, A. Quarteroni, G. Villani, G. Ziarelli, SEIHRDV: a multi-age multi-group epidemiological model and its validation on the COVID-19 epidemics in Italy, preprint, arXiv:2501.04148. https://doi.org/10.48550/arXiv.2501.04148
16. I. Rodiah, P. Vanella, A. Kuhlmann, V. K. Jaeger, M. Harries, G. Krause, et al., Age-specific contribution of contacts to transmission of SARS-CoV-2 in Japan, *Eur. J. Epidemiol.*, **38** (2023), 39–58. https://doi.org/10.1007/s10654-022-00938-6
17. Z. Vizi, E. K. Korir, N. Bogya, C. Rosztoczi, G. Makay, P. Boldog, Age group sensitivity analysis of epidemic models: investigating the impact of contact matrix structure, preprint, arXiv:2502.19206. https://doi.org/10.48550/arXiv.2502.19206
18. M. De la Sen, A. Ibeas, S. Alonso-Quesada, R. Nistal, On a new epidemic model with asymptomatic and dead-infective subpopulations with feedback controls useful for Ebola disease, *Discrete Dyn. Nat. Soc.*, **2017** (2017), 4232971. https://doi.org/10.1155/2017/4232971
19. G. Stamov, I. Stamova, N. Simeonova, K. Gabrovska, S. Simeonov, Stability of sets for Ebola virus disease models through impulsive conformable approach, *Mathematics*, **14** (2026), 1108. https://doi.org/10.3390/math14071108

20. N. Wang, L. Zhang, Y. Luo, Z. Teng, Global stability for a cumulative release Ebola epidemic model from the corpses and infected individuals, *Infect. Dis. Modell.*, **10** (2025), 1291–1306. https://doi.org/10.1016/j.idm.2025.07.002

21. T. Rolski, H. Schmidli, V. Schmidt, J. L. Teugels, *Stochastic Processes for Insurance and Finance*, John Wiley & Sons, 2009. https://doi.org/10.1002/9780470317044

22. L. Palmborg, F. Lindskog, Premium control with reinforcement learning, *ASTIN Bull.*, **53** (2023), 233–257. https://doi.org/10.1017/asb.2023.13

23. S. Singireddy, Reinforcement learning approaches for pricing condo insurance policies, *Am. J. Anal. Artif. Intell.*, 2023.

24. C. Xuan, F. Zhang, F. Yin, H. K. Lam, Constrained proximal policy optimization, preprint, arXiv:2305.14216. https://doi.org/10.48550/arXiv.2305.14216

25. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, preprint, arXiv:1707.06347. https://doi.org/10.48550/arXiv.1707.06347

AIMS Press