When Less is More: Optimizing Prescription ALERTS under Fatigue

Hossein Piri
Assistant Professor
Haskayne School of Business
University of Calgary

Michael Lingzhi Li
Assistant Professor
Harvard Business School
Harvard University

Feb 2026

Research Motivation

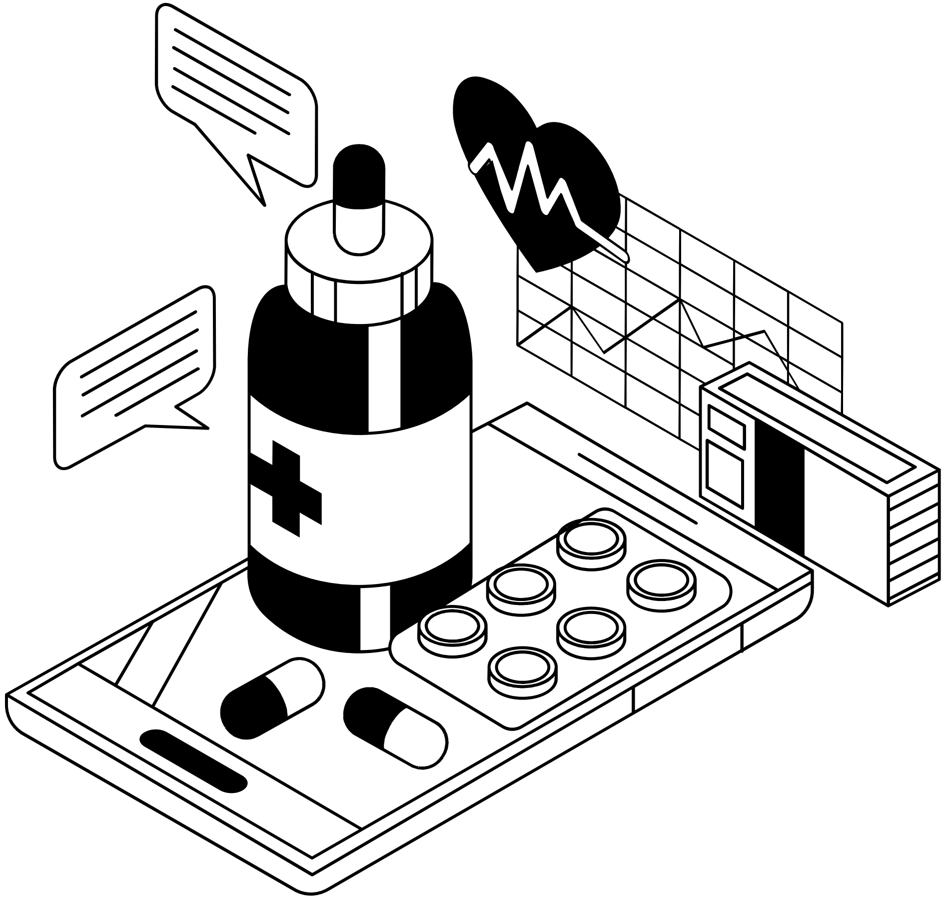
Literature Review

Model Framework & Structure

Special Case in the Real World

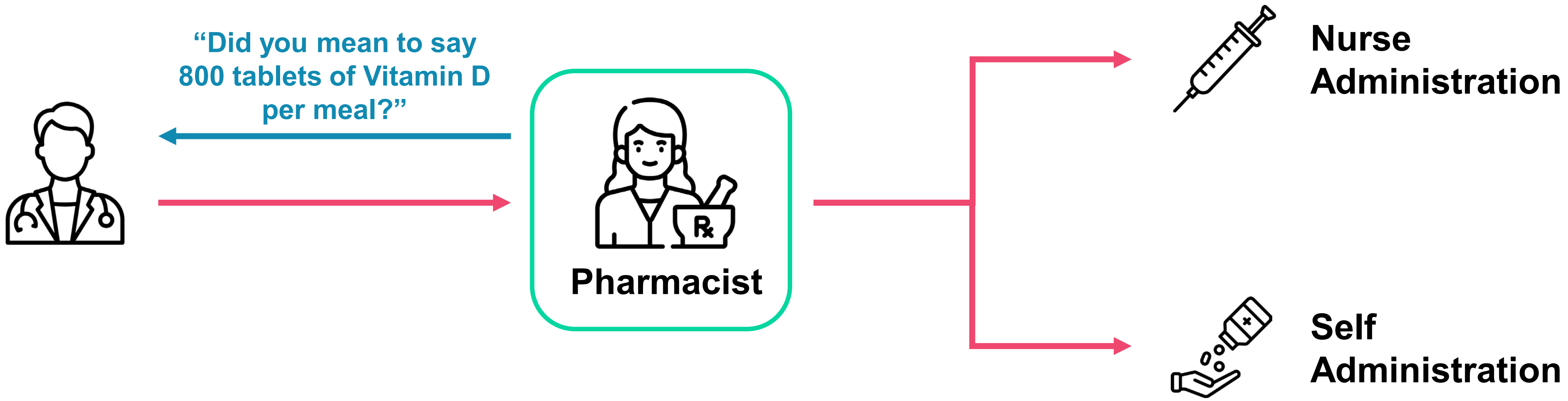
Case Study

Conclusion

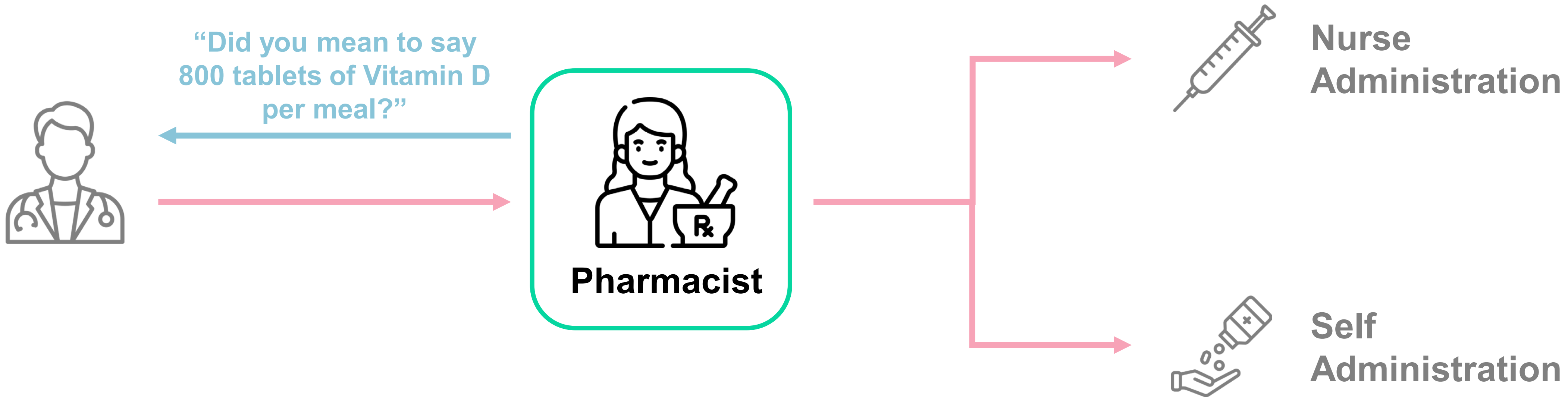


Research Motivation

Overview of the Prescription Workflow



Overview of the Prescription Workflow



Pharmacists act as the **last line of defence**

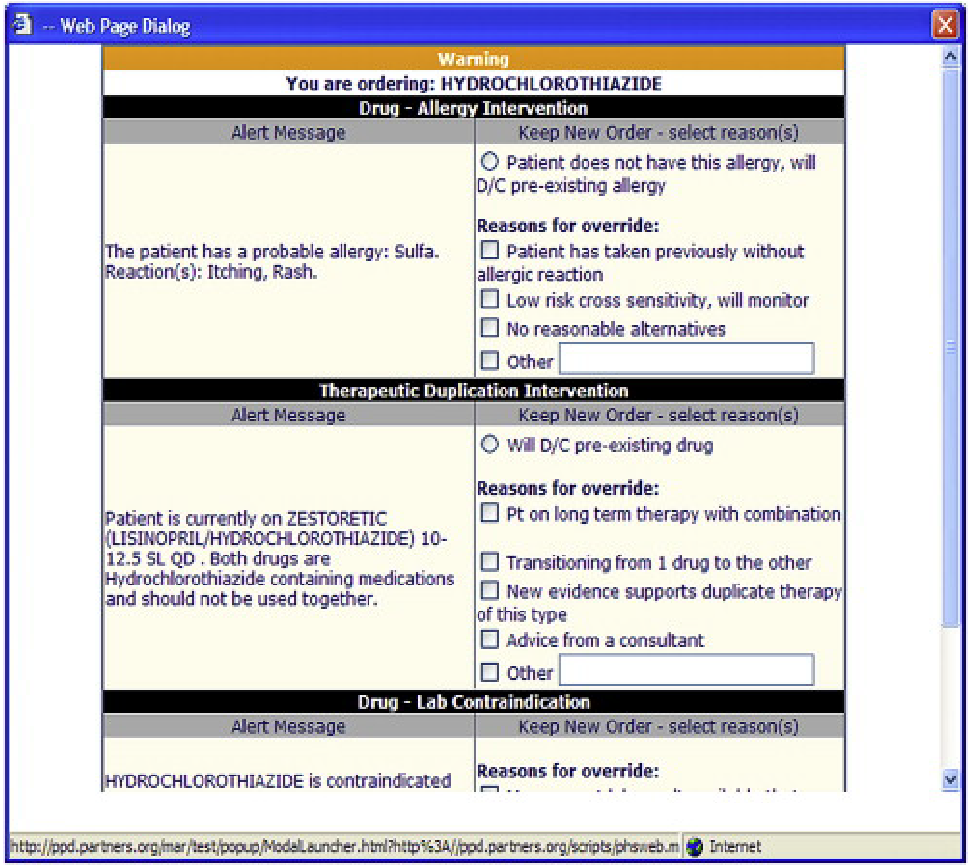
3%
Prescription Error Rate

(Naseralallah et al. 2020)

50K+
Annual Deaths due to Prescription Error

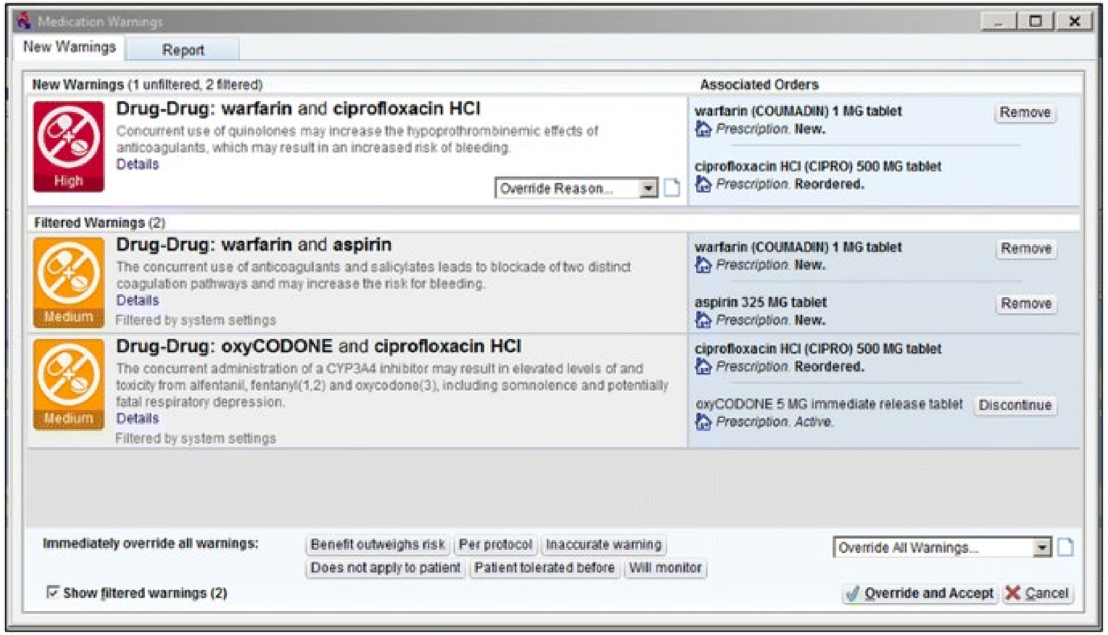
(AHRQ 2015)

Computer Provider Order Entry (CPOE) Systems



Ⓐ Tier 1 (most severe)

Ⓑ Tiers 2 (moderate) and 3 (least severe)



Error Reduction by 50%+

How much of a good thing is TOO much?

7-15

Average # of Alerts per Prescription

100-1000

Average # of Alerts per Shift

90%+

Time Spent Dealing with Alerts

96%

Override

*"A hospitalized teenager received **38x** overdose of antibiotic because clinician advised to 'ignore all alerts'."*
-AHRQ Patient Network

How much of a good thing is TOO much?

7-15
Average # of Alerts per Prescription

100-1000
Average # of Alerts per Shift

90%+
Time Spent Dealing with Alerts

96%
Override

*"A hospitalized teenager received **38x** overdose of antibiotic because clinician advised to 'ignore all alerts'."*
-AHRQ Patient Network

GOAL: Understand optimal alerting strategy under fatigue.

Research Question

How to optimize CPOE alerts to balance between providing essential alerts and minimizing unnecessary interruptions?

Literature Review

Literature Review

Behavioral Dimension

- **Fatigue ↓ compliance in clinicians**
- **Repeated interruptions → lower responsiveness**
  - Nanji et al. 2014
  - Dai et al. 2015
  - Ancker et al. 2017
  - Schwartz et al. 2021
  - Joseph et al. 2021
  - Bavafa and Jonasson 2024
- **Probability of a favorable parole among judges: 65% → ~0%**
  - Danziger et al. 2011

Literature Review

Behavioral Dimension

- **Fatigue ↓ compliance in clinicians**
- **Repeated interruptions → lower responsiveness**
  - Nanji et al. 2014
  - Dai et al. 2015
  - Ancker et al. 2017
  - Schwartz et al. 2021
  - Joseph et al. 2021
  - Bavafa and Jonasson 2024
- **Probability of a favorable parole among judges: 65% → ~0%**
  - Danziger et al. 2011

Clinical Approaches

- **Tiering/context filters; adjust severity/frequency**
- **Rule-based threshold tuning can reduce burden (often manual/static)**
  - Cash 2009
  - Del Beccaro et al. 2010
  - Khreis et al. 2019
  - Naeem et al. 2022
  - Chan 2022

Literature Review

Behavioral Dimension

- **Fatigue ↓ compliance in clinicians**
- **Repeated interruptions → lower responsiveness**
  - Nanji et al. 2014
  - Dai et al. 2015
  - Ancker et al. 2017
  - Schwartz et al. 2021
  - Joseph et al. 2021
  - Bavafa and Jonasson 2024
- **Probability of a favorable parole among judges: 65% → ~0%**
  - Danziger et al. 2011

Clinical Approaches

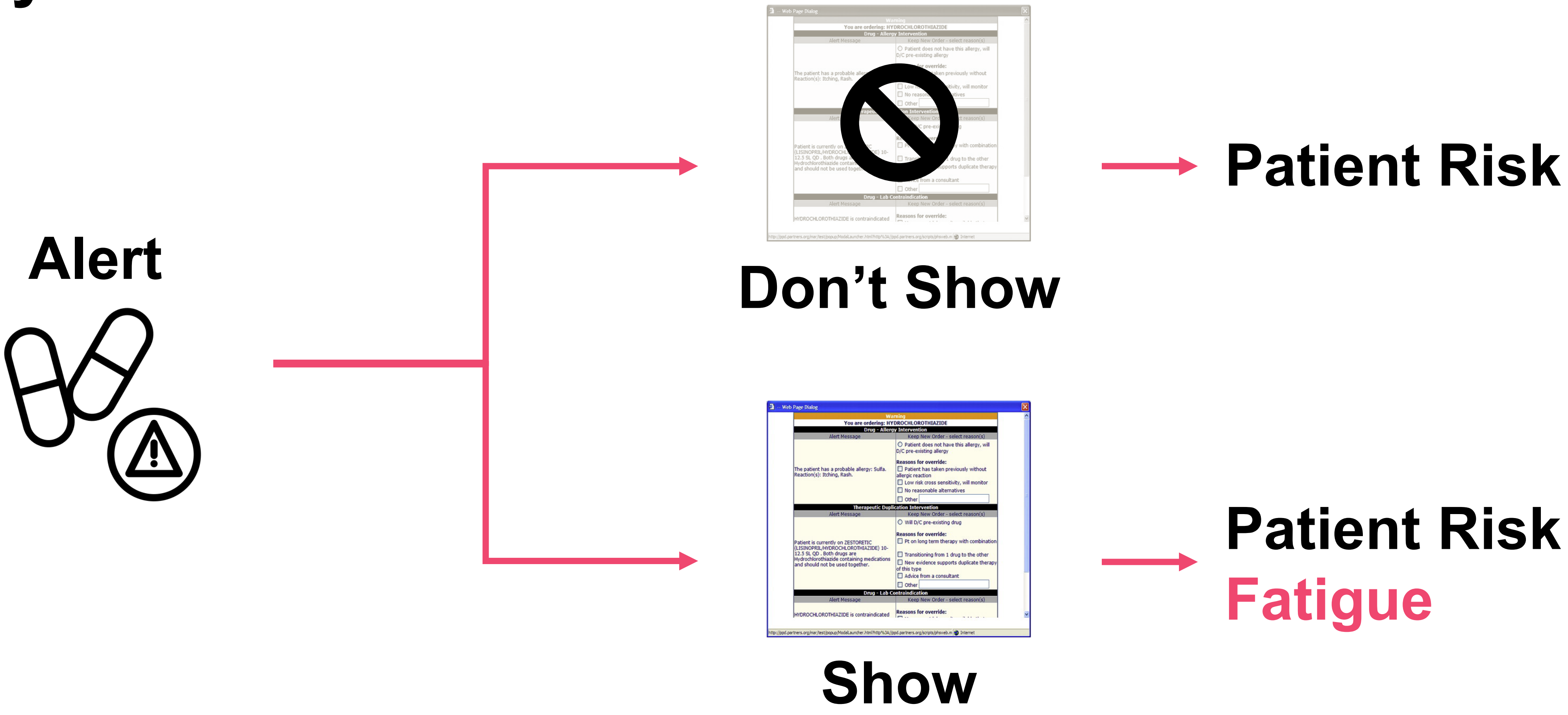
- **Tiering/context filters; adjust severity/frequency**
- **Rule-based threshold tuning can reduce burden (often manual/static)**
  - Cash 2009
  - Del Beccaro et al. 2010
  - Khreis et al. 2019
  - Naeem et al. 2022
  - Chan 2022

Machine Learning / Statistical Approaches

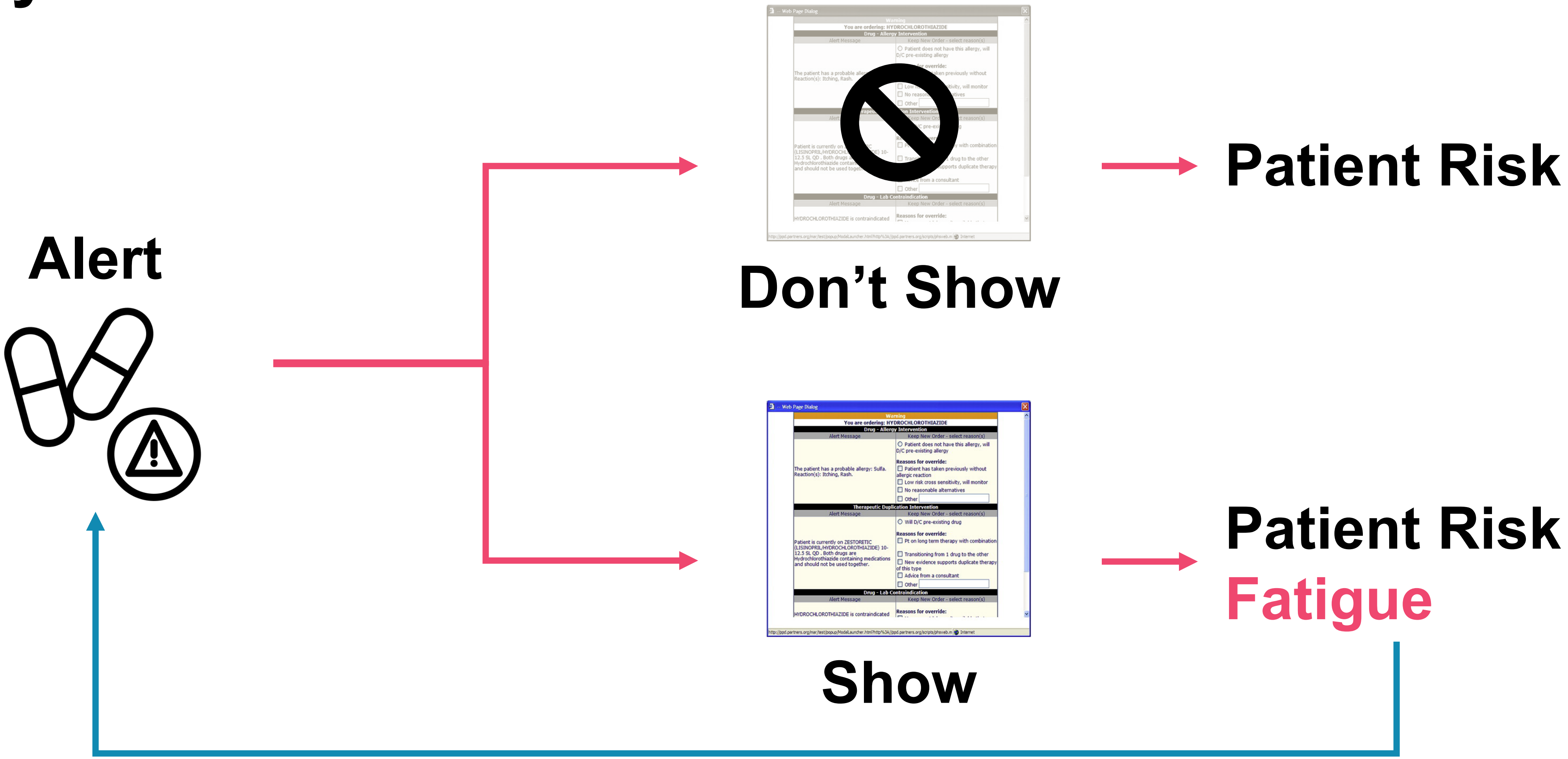
- **Predict/suppress low-value alerts (overrides/outcomes)**
- **Personalize relevance (patient/clinician-specific models)**
  - Malone et al. 2023
  - Chen et al. 2024
  - Martell 2019; Poly et al. 2020
  - Graafsma et al. 2024

Model Framework & Structure

Why is this difficult?

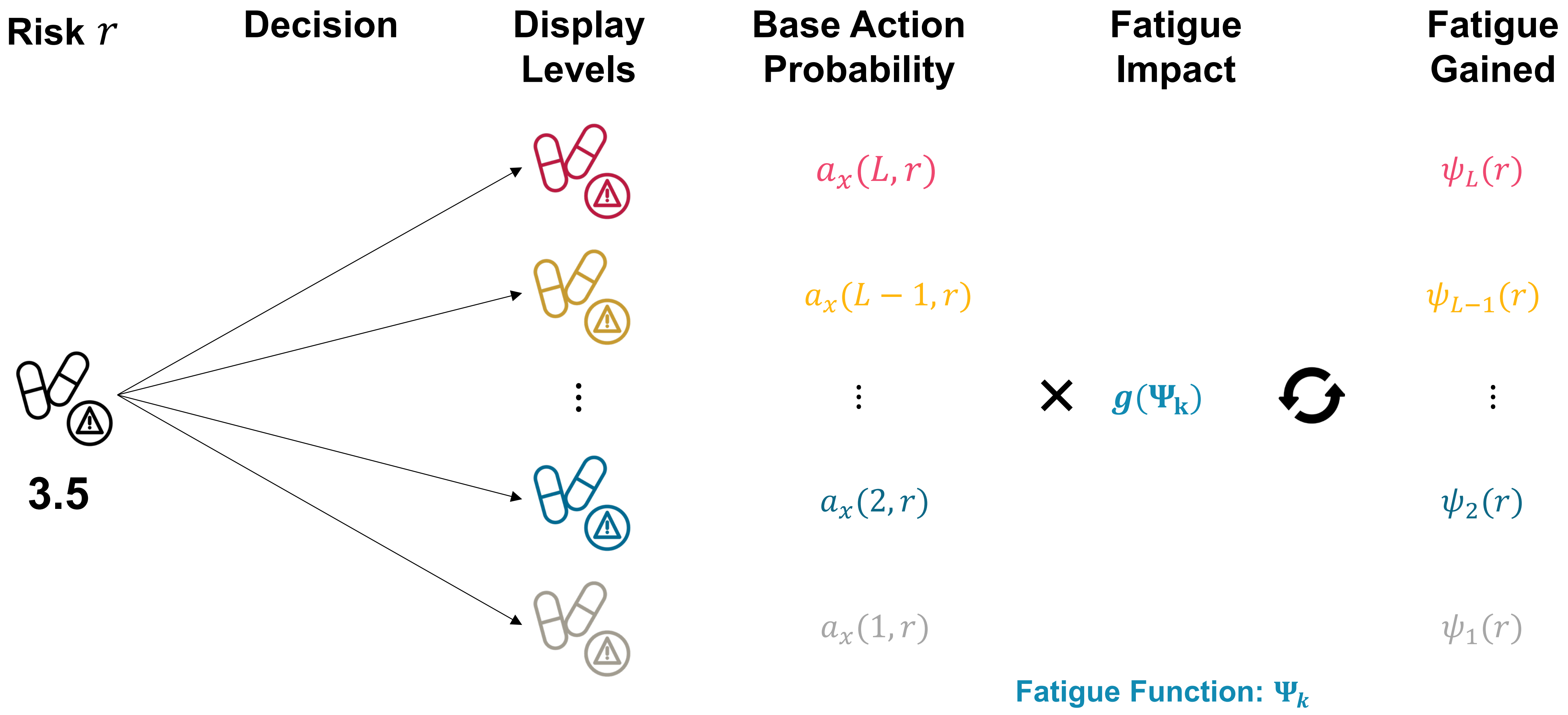


Why is this difficult?

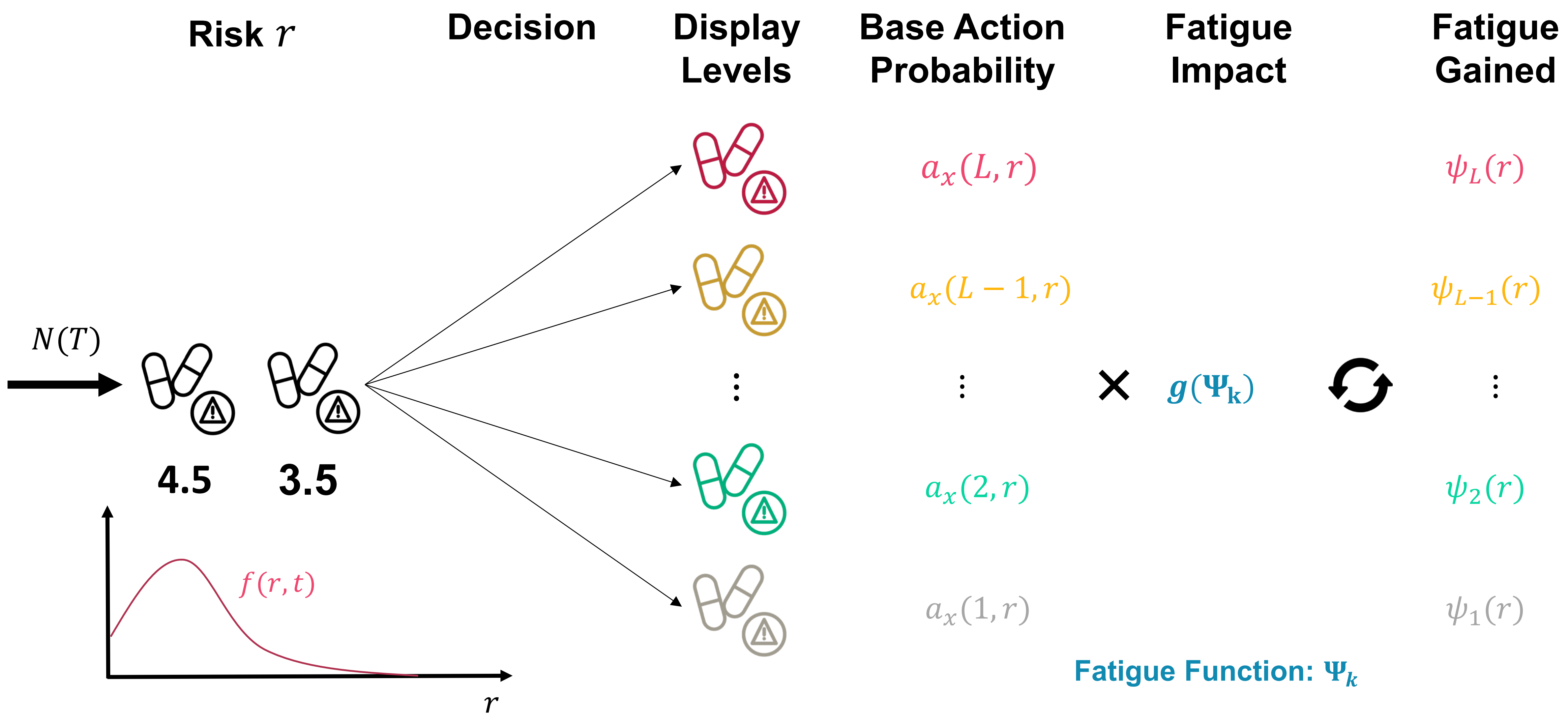


Embi and Leonard (2012),
Ancker et al. (2017)

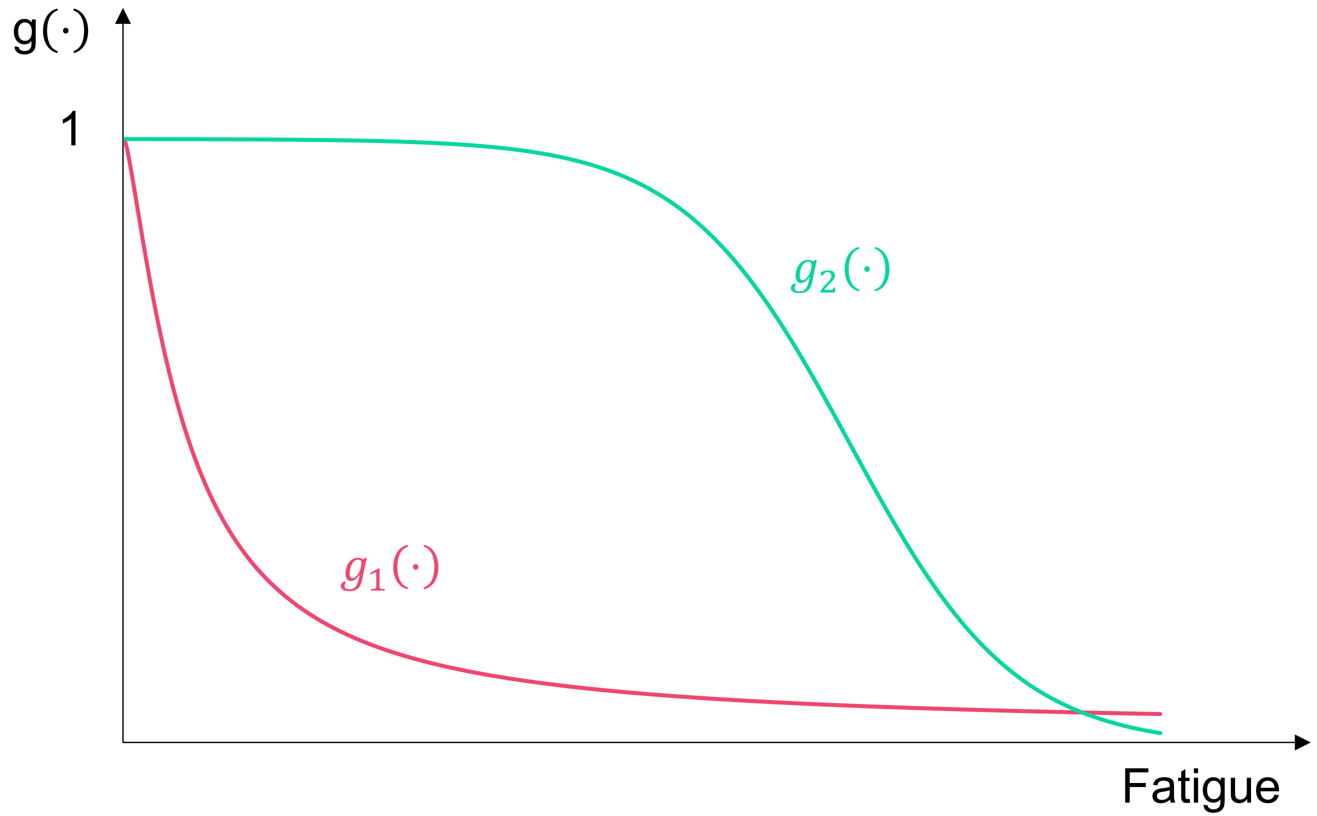
Modeling Framework



Modeling Framework



The Fatigue Linking Function

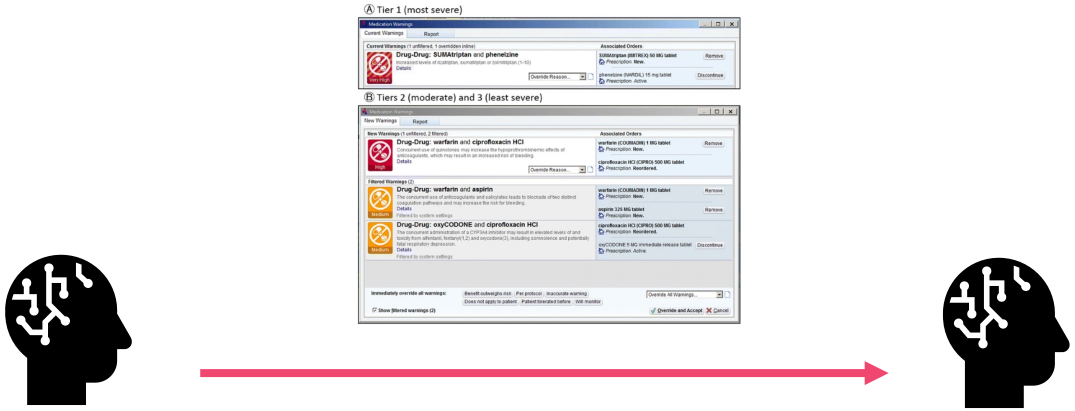


$g_1(\cdot)$: Pharmacist/clinician performance quickly deteriorates with fatigue, but maintains residual efficiency even under high fatigue

$g_2(\cdot)$: Pharmacist/clinician maintains high level of performance until fatigue is very high, and drops significantly

Fatigue Transition Dynamics

Ψ_k : Pharmacist's fatigue after processing k^{th} alert



Ψ_k Ψ_{k+1}

Display Levels

Assumption 1:

$U_{l,r}: [0,\infty) \rightarrow [0,\infty)$: Fatigue transition map

- Monotone
- Cumulative
- Identifiability

$$\Psi_{k+1} = U_{l+1,r+1}(\Psi_k)$$

Technical Result

Theorem 1

Under assumption 1, ∃ a strictly increasing transformation $\phi: [0, \infty) \to R$ and constants $\psi_{l,r} \geq 0$ such that defining the rescaled fatigue state $\widetilde{\Psi} = \phi(\Psi)$, the fatigue dynamic satisfies

$$\widetilde{\Psi}_{k+1} = \widetilde{\Psi}_k + \psi_{l_{k+1}, r_{k+1}}$$

Technical Result

Theorem 1

Under assumption 1, ∃ a strictly increasing transformation $\phi\colon [0, \infty) \to R$ and constants $\psi_{l,r} \geq 0$ such that defining the rescaled fatigue state $\widetilde{\Psi} = \phi(\Psi)$, the fatigue dynamic satisfies

$$\widetilde{\Psi}_{k+1} = \widetilde{\Psi}_k + \psi_{l_{k+1}, r_{k+1}}$$

Translation

Any fatigue transition dynamics that satisfies Assumption 1 admits an equivalent representation in which fatigue transition is additive.

Dynamic Program

Decisions: $\pi_k(l|\Psi, r)$

Probability of a risk r alert
arriving at epoch k
will be shown at level l

Dynamic Program

Decisions: $\pi_k(l|\Psi, r)$

Probability of a risk r alert
arriving at epoch k
will be shown at level l

Objective:

$$\max_{\pi_1,\pi_2,\ldots} \mathbb{E}\left[\sum_{k=0}^{N(T)} \sum_{l=1}^{L} r_k \times \pi_k(l|\Psi_{k-1}, r_k) \times a_x(l, r_k).\, g(\Psi_{k-1}, r_k)\right]$$

Risk avoided
for K alerts

Dynamic Program

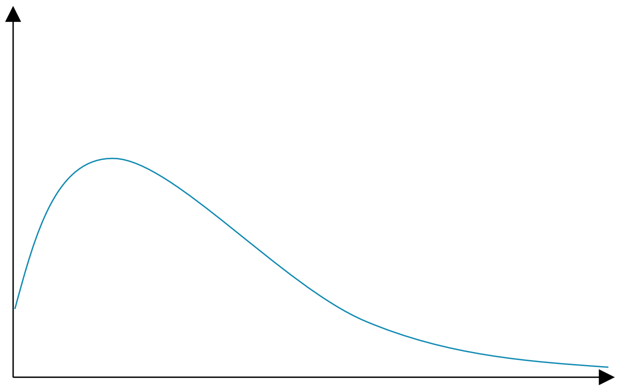
Decisions: $\pi_k(l|\Psi, r)$

Probability of a risk r alert
arriving at epoch k
will be shown at level l

Objective:

$$\max_{\pi_1,\pi_2,\dots} \mathbb{E}\left[\sum_{k=0}^{N(T)} \sum_{l=1}^{L} \boxed{r_k} \times \pi_k(l|\Psi_{k-1}, r_k) \times a_x(l, rk).\, g(\Psi_{k-1}, r_k)\right]$$

Risk avoided
for K alerts



Dynamic Program

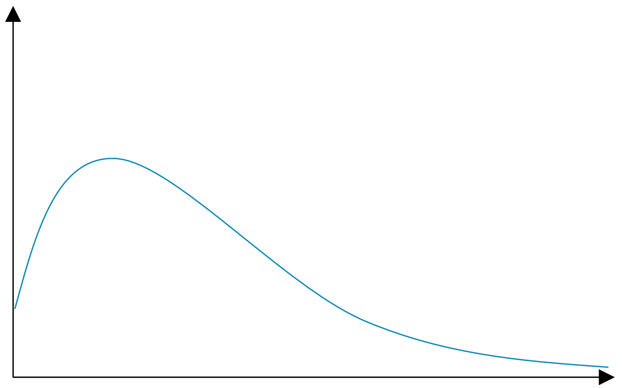
Decisions: $\pi_k(l|\Psi, r)$

Probability of a risk r alert
arriving at epoch k
will be shown at level l

Objective:

$$\max_{\pi_1,\pi_2,\dots} \mathbb{E}\left[\sum_{k=0}^{N(T)} \sum_{l=1}^{L} r_k \times \underbrace{\boxed{\pi_k(l|\Psi_{k-1}, r_k)}}_{\text{risk } r_k \text{ alert shown with level } l} \times \; a_x(l, rk).\, g(\Psi_{k-1}, r_k)\right]$$

Risk avoided
for K alerts



Dynamic Program

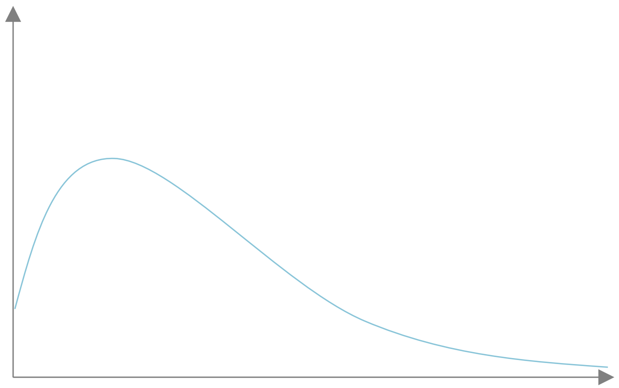
Decisions: $\pi_k(l|\Psi, r)$

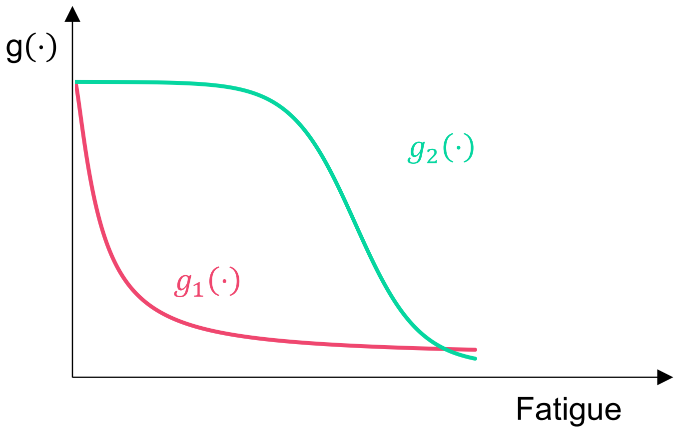
Probability of a risk r alert arriving at epoch k will be shown at level l

Objective:

$$\max_{\pi_1,\pi_2,\ldots} \mathbb{E}\left[\sum_{k=0}^{N(T)} \sum_{l=1}^{L} r_k \times \underbrace{\pi_k(l|\Psi_{k-1}, r_k)}_{\text{risk } r_k \text{ alert shown with level } l} \times \underbrace{\boxed{a_x(l, rk).\, g(\Psi_{k-1}, r_k}}_{\text{Fatigue-impacted action probability}}\right]$$

Risk avoided for K alerts





Dynamic Program

Decisions: $\pi_k(l|\Psi, r)$ 

Probability of a risk r alert arriving at epoch k will be shown at level l

Objective:

$$\max_{\pi_1,\pi_2,\dots} \mathbb{E}\left[\sum_{k=0}^{N(T)} \sum_{l=1}^{L} r_k \times \underbrace{\pi_k(l|\Psi_{k-1}, r_k)}_{\text{risk } r_k \text{ alert shown with level } l} \times \underbrace{a_x(l, rk).\, g(\Psi_{k-1}, r_k)}_{\text{Fatigue-impacted action probability}}\right]$$

Risk avoided for K alerts

Constraints: $l_k \sim \pi_k(.|\Psi_{k-1}, r_k) \qquad r_k \sim f(r, t_k)$

Fatigue depends on all past decisions

Dynamic Program

Decisions: $\pi_k(l|\Psi, r)$

Probability of a risk r alert arriving at epoch k will be shown at level l

Objective:

$$\max_{\pi_1,\pi_2,\ldots} \mathbb{E}\left[\sum_{k=0}^{N(T)} \sum_{l=1}^{L} r_k \times \underbrace{\pi_k(l|\Psi_{k-1}, r_k)}_{\text{risk } r_k \text{ alert shown with level } l} \times \underbrace{a_x(l, rk).\, g(\Psi_{k-1}, r_k)}_{\text{Fatigue-impacted action probability}}\right]$$

Risk avoided for K alerts

Constraints: $l_k \sim \pi_k(.\,|\Psi_{k-1}, r_k)$ $\quad r_k \sim f(r, t_k)$

Fatigue depends on all past decisions

Fatigue Dynamics:

$$\Psi_{k+1} = \Psi_k + \psi_{l_{k+1}, r_{k+1}} \qquad \Psi_1 = 0$$

Fluid Approximation

Decisions: $p_l(r,t)$

Probability of a risk r alert arriving at time t will be shown at level l

Objective:

$$\sum_{l=1}^{L}\int_0^{\infty}\int_0^{T} r \times \underbrace{\lambda(t)\, f(r,t) p_l(r,t)}_{\text{Instantaneous rate of risk } r \text{ level } l \text{ alerts}} \times \underbrace{a_x(l,r) g(\psi(t), r)}_{\text{Fatigue-impacted action probability}}\, dt\, dr$$

Risk avoided through alerts within time $[0, T]$

Constraints: $\psi(t) = \sum_{l=1}^{L}\int_0^{\infty}\int_0^{t} \psi_l(r)\, \lambda(s) f(r,s) p_l(r,s) ds\, dr$

Fatigue depends on all past decisions

Infinite-dimensional functional optimization problem

Technical Result

Theorem 2

Under the following assumptions, fluid-approximation is an asymptotically optimal approximation to the dynamic programming problem.

- Bounded Mean Risk and Arrival
- Bunded Fatigue
- Lipschitz Fatigue: $\exists L_g < \infty$ s.t. $\forall r, \Psi, \Psi', |g(\Psi, r) - g(\Psi', r)| \leq L_g |\Psi - \Psi'|$

Special Case in the Real World

A Special Case: Alert or Not?

Two Levels

Alert No Alert

Stationary Risk Distribution

$$f(r,t) = f(r)$$

Exponential Linking Function

$$g(\Psi, r) = \eta(r)\exp(-\beta\ \Psi)$$

"Rate of attention decrement"

Mackworth (1948), Warm et al. (2008)

A Special Case: Alert or Not?

Two Levels

Alert

No Alert

Stationary Risk Distribution

$$f(r,t) = f(r)$$

Exponential Linking Function

$$g(\Psi, r) = \eta(r)\exp(-\beta\ \Psi)$$

Fatigue rate

Mackworth (1948), Warm et al. (2008)

Theorem

Define: $H(y) = \int_0^{\infty} \left(r\, a_x(r) - y\, \psi(r)\right)^+ f(r) dr\,,$ $\qquad G(y) = \int_0^y \frac{1}{H(x)}\, dx$

Screening index: $S(r) = \frac{r\, a_x(r)}{\psi(r)}$

Show an alert iff $S(r) \geq G^{-1}(\beta(T - t))$

A Special Case: Alert or Not?

Theorem

Define: $H(y) = \int_0^\infty \left(r\, a_x(r) - y\, \psi(r)\right)^+ f(r) dr\,,$ $\quad G(y) = \int_0^y \frac{1}{H(x)}\, dx$

Screening index: $S(r) = \frac{r\, a_x(r)}{\psi(r)}$

Show an alert iff $S(r) \geq G^{-1}(\beta(T-t))$

$$S(r) = \frac{r\, a_x(r)}{\psi(r)}$$

Baseline Expected Risk Avoidance

Fatigued Induced

A Special Case: Alert or Not?

Theorem

Define: $H(y) = \int_0^\infty \left(r\, a_x(r) - y\, \psi(r)\right)^+ f(r) dr\,,$ $\quad G(y) = \int_0^y \frac{1}{H(x)}\, dx$

Screening index: $S(r) = \frac{r\, a_x(r)}{\psi(r)}$

Show an alert iff $S(r) \geq G^{-1}(\beta(T-t))$

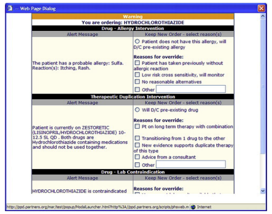
$$H(y) = \int_0^\infty \left(r\, a_x(r) - y\, \psi(r)\right)^+ f(r) dr\,,$$

Baseline Expected Risk Avoidance

Threshold y

Fatigued Induced

Illustration of Screening Threshold



$S(r)$

G^{-1}

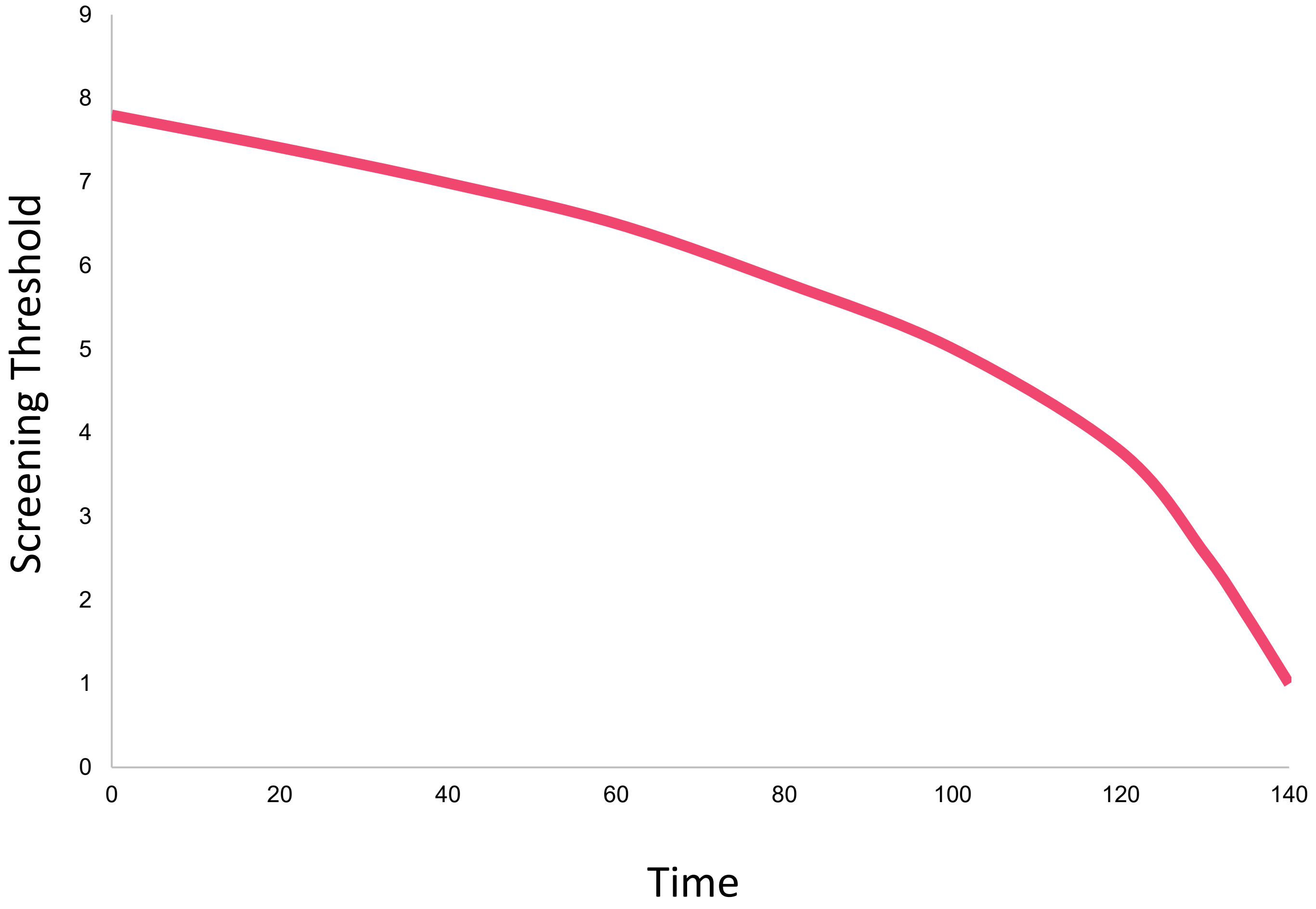


Illustration of Screening Threshold

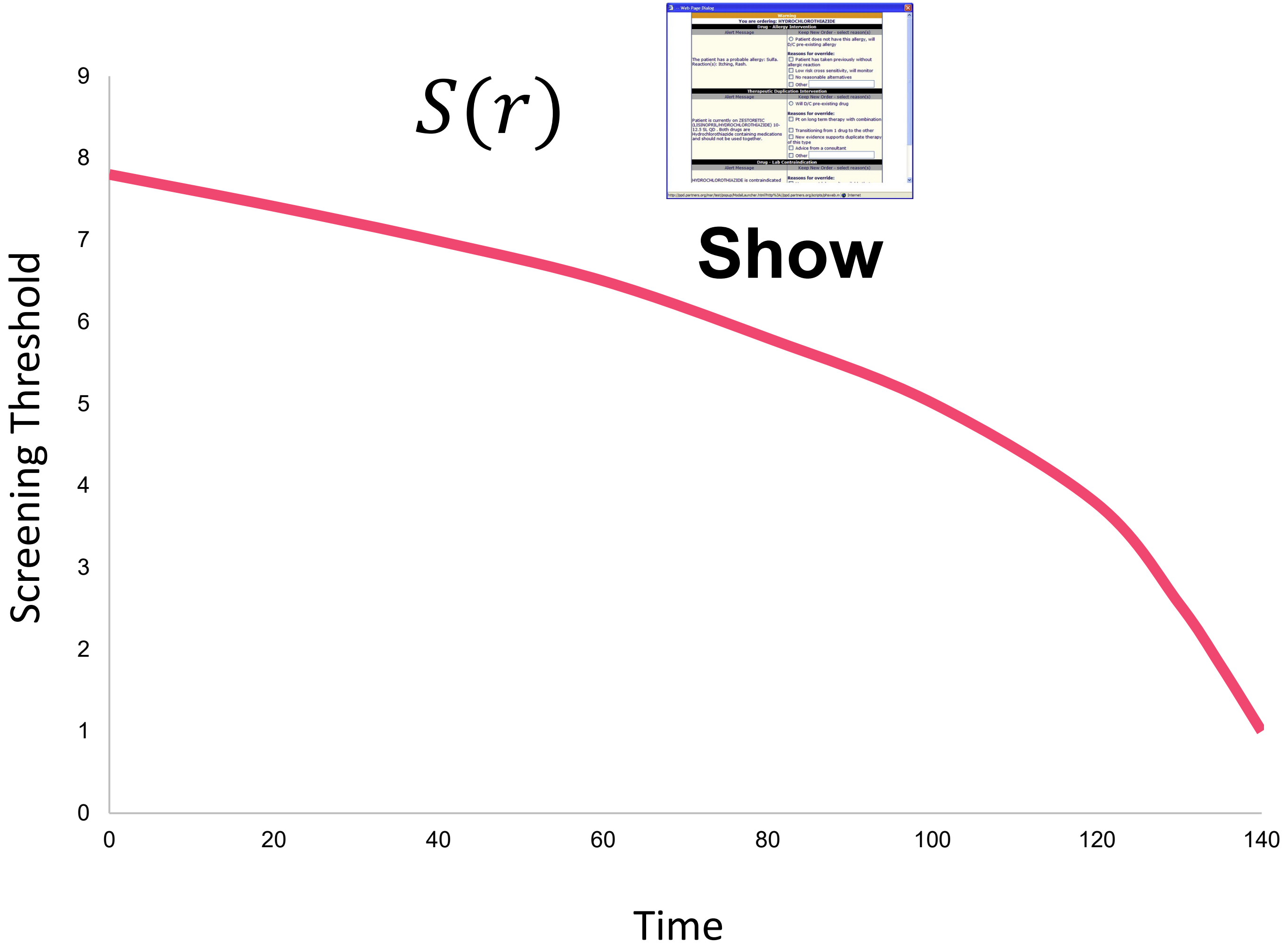


Illustration of Screening Threshold

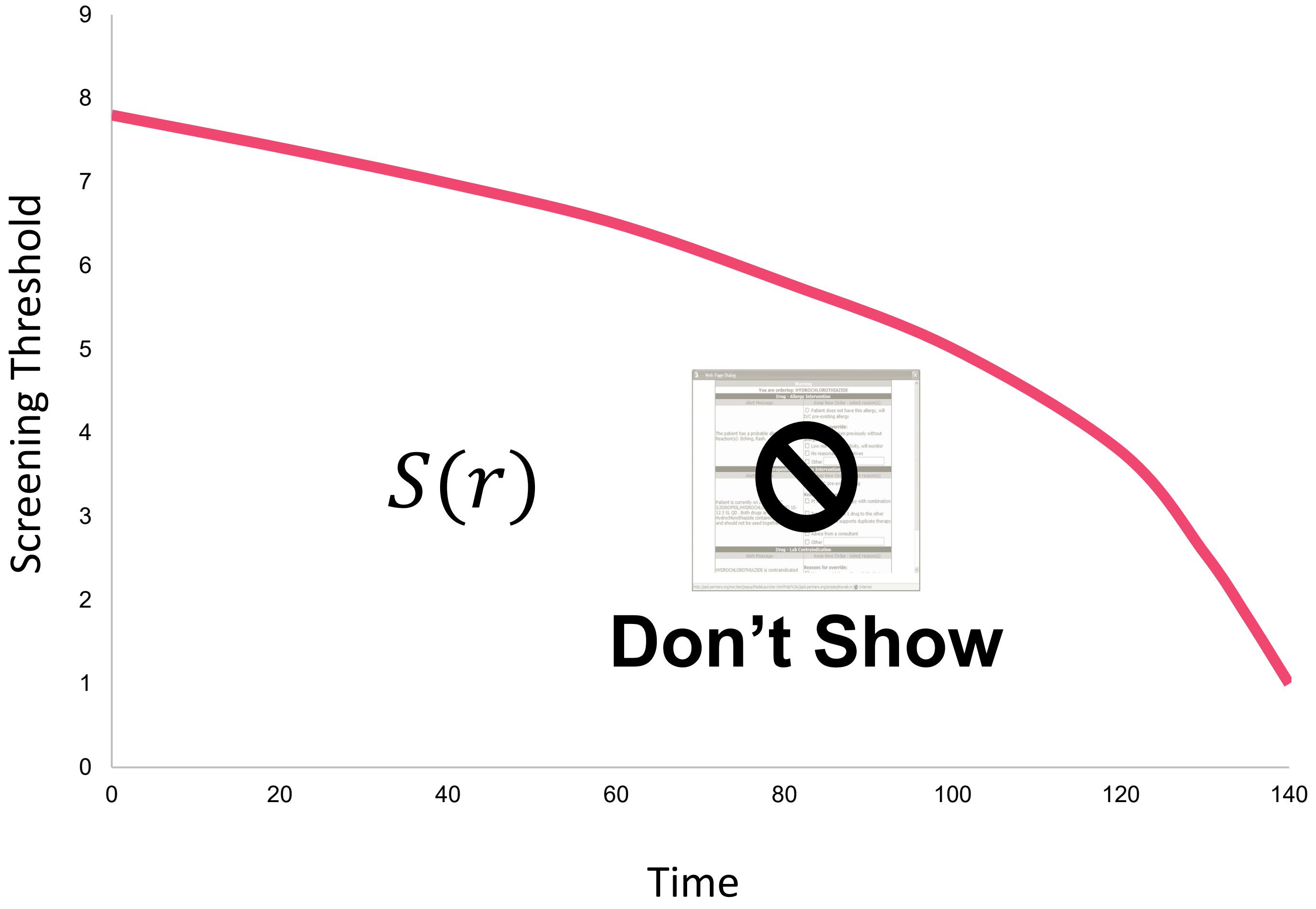
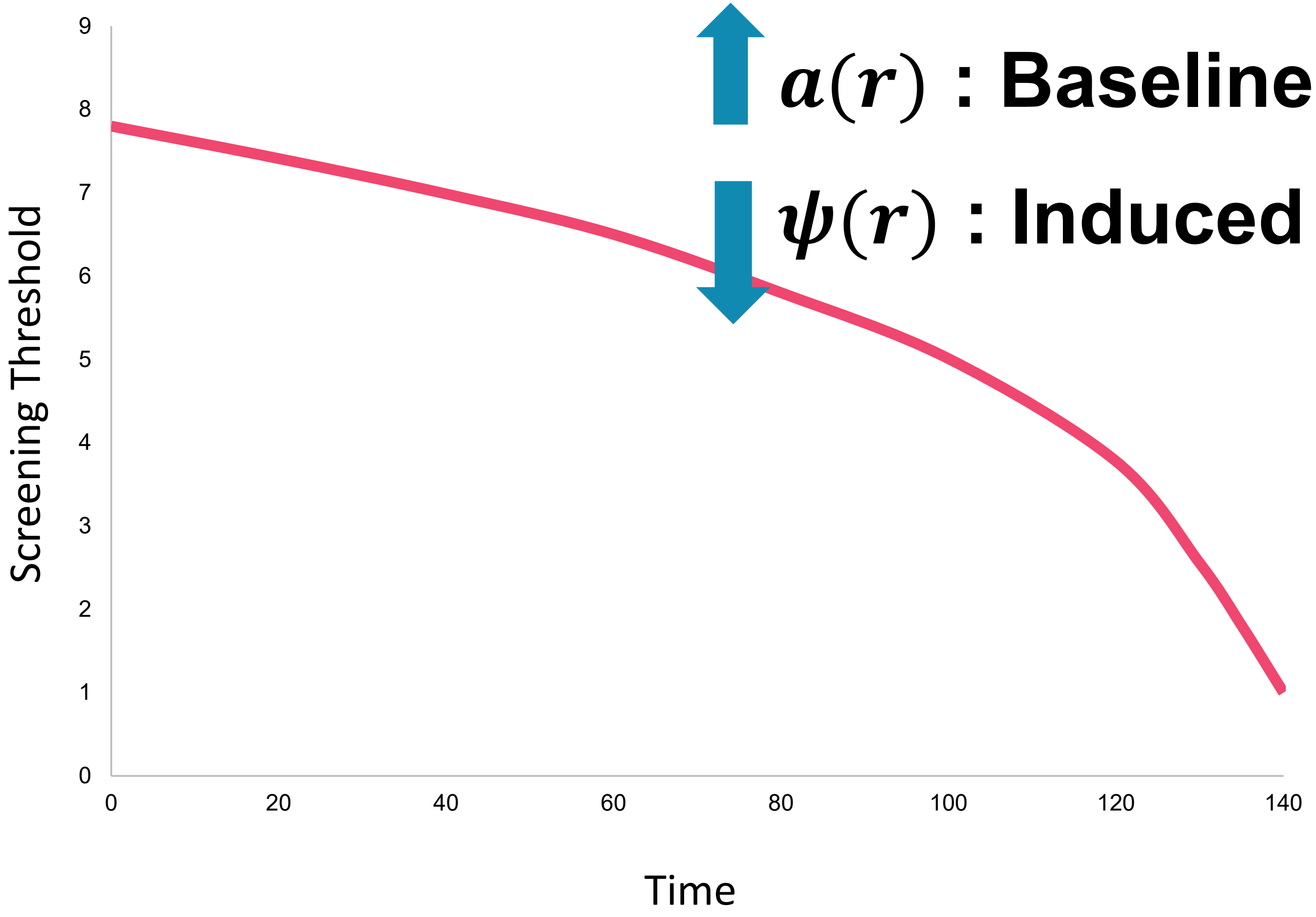


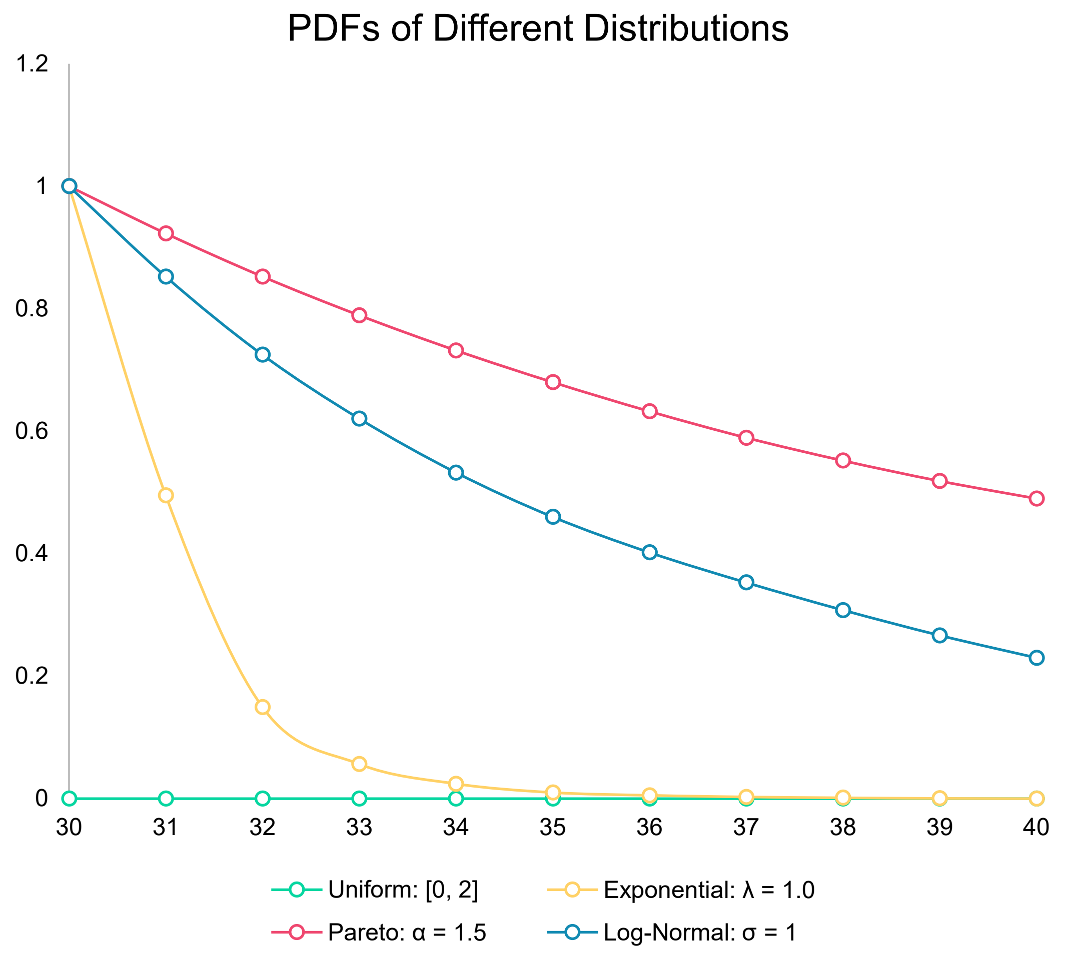
Illustration of Screening Threshold

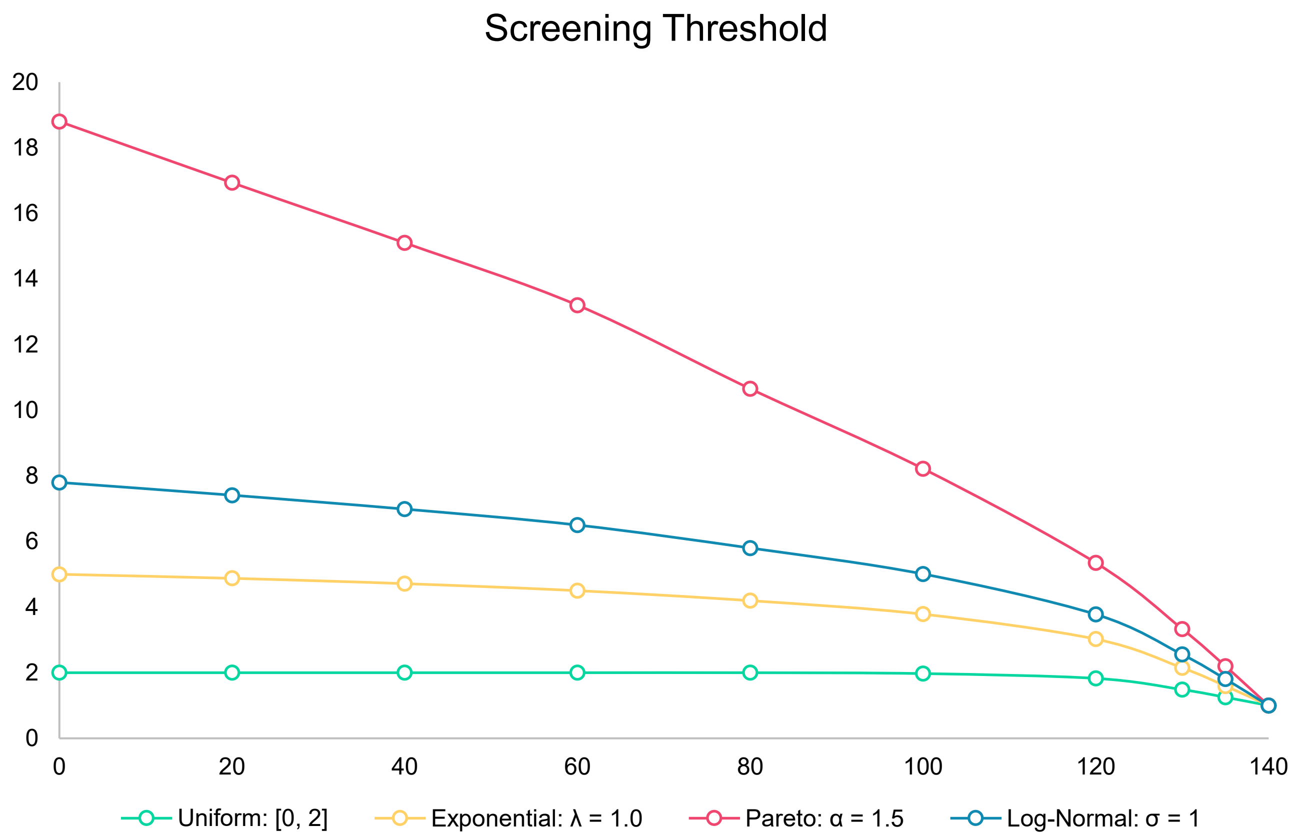


$a(r)$: **Baseline Acceptance**

$\psi(r)$: **Induced Fatigue**

Illustration of the Tail Risk Effect





Comparison with Baseline Policies

$\mathcal{P}^0$	**No Alert Strategy**
$\mathcal{P}^*$	**Optimal Strategy:** Decreasing Threshold Policy
$\mathcal{P}^A$	**All Alerts Policy:** Common default policy

Comparison with Baseline Policies

$\mathcal{P}^0$	**No Alert Strategy**
$\mathcal{P}^*$	**Optimal Strategy:** Decreasing Threshold Policy
$\mathcal{P}^A$	**All Alerts Policy:** Common default policy

$$\frac{B_2^0}{B_2^*} = \frac{\mu_A}{\mu_\psi} \times \frac{1}{S_{max}}$$

Comparison with Baseline Policies

$\mathcal{P}^0$ **No Alert Strategy**

$\mathcal{P}^*$ **Optimal Strategy:** Decreasing Threshold Policy

$\mathcal{P}^A$ **All Alerts Policy:** Common default policy

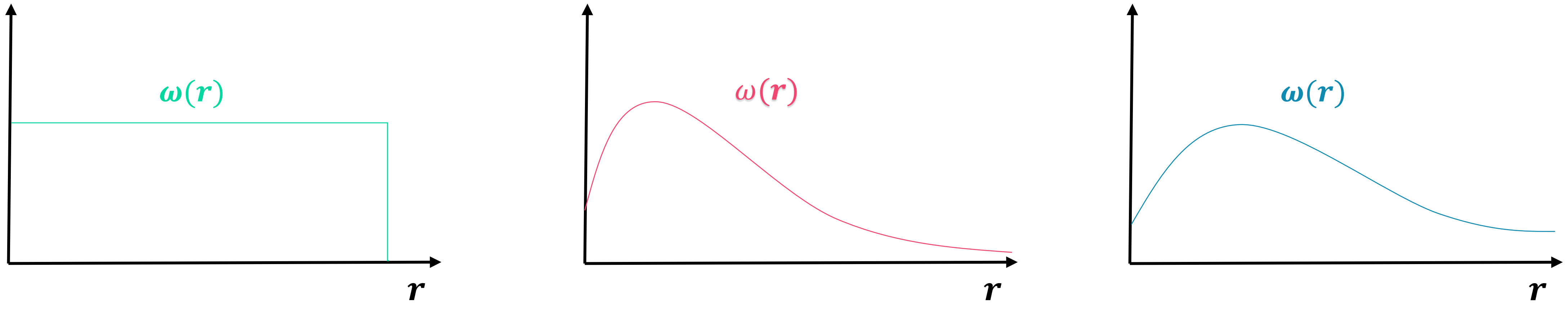
$\mathcal{P}^\dagger$ **Optimal *Fixed* Policy:** $r_1^*(t) = k$

The Effectiveness of Fixed Policies

Theorem

Normalized Expected Fatigue: $\omega(y) = \frac{1}{\mu\omega}\int_{S(r)=y}^{\infty} \psi(y) f(r)\, dr$

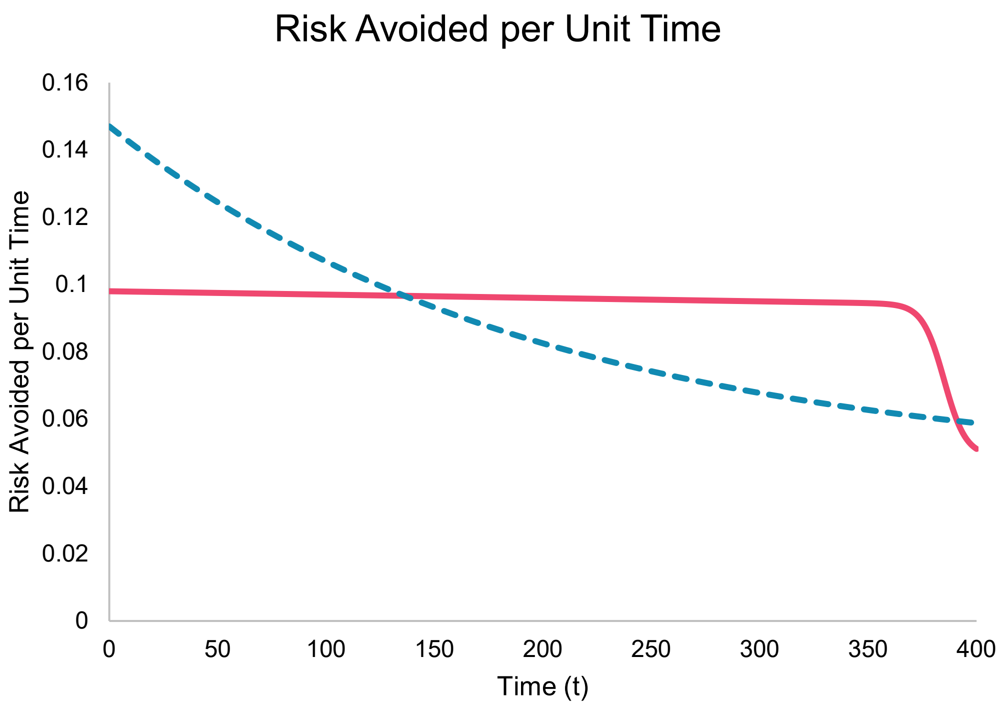
1. (Light-tailed $\omega(y)$) Tail of $\omega(y)$ belongs to the Weibull or Gumbel → $\lim_{T\to\infty} \frac{B_2^{Fix}(T)}{B_2^{*}(T)} = 1$
2. (heavy-tailed $\omega(y)$) Tail of $\omega(y)$ belongs to the Frechet → $\lim_{T\to\infty} \frac{B_2^{Fix}(T)}{B_2^{*}(T)} < 1$



Constant Optimal ✓ **Constant Optimal** ✓ **Constant Near-Optimal** ✓

Near-optimality of optimal fixed policy.

Optimal Fixed vs Optimal Policies



Optimal Solution: Decreasing

Fixed Solution: Constant

Optimal Solution: (Near) Constant

Fixed Solution: Decreasing

Tradeoff between alert vs risk fairness

Case Study

Real-world Data Calibration

University of Alberta Hospital

Pharmacist alert data from the University of Alberta Hospital System

Study Period
Jan 2021 – Dec 2024

Total Alerts
583,313

Avg Alerts / Shift
142

90th Percentile
> 449 Alerts / Shift

Override Rate
78%

Risk Estimation

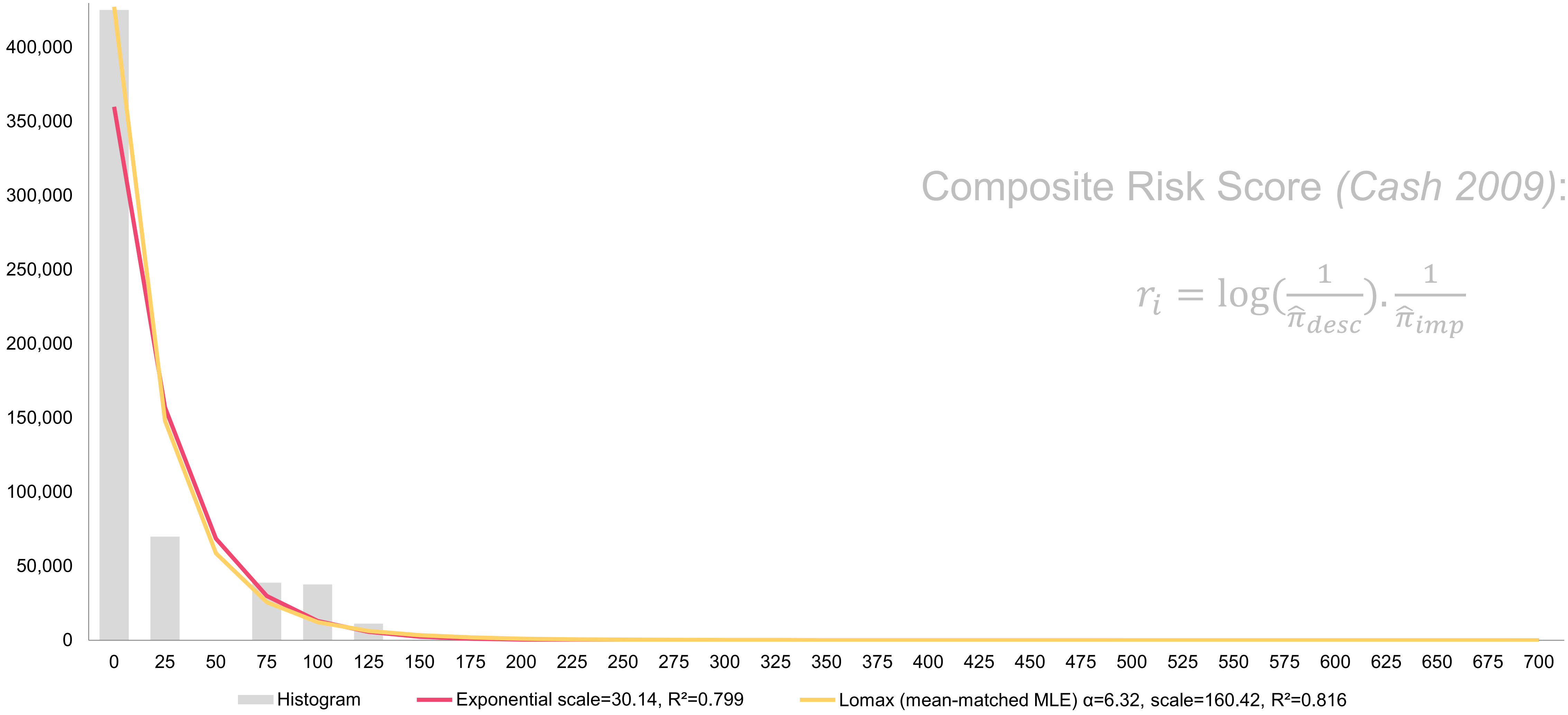
Alert Description	**Importance Level**
Moderate Interaction	**Low**
Severe Warning	**Medium**
Absolute Contraindication	**High**
Moderate Warning	**Very High**

Description → ***Templates***
Lowercasing, removing markup, digits, etc.

Composite Risk Score *(Cash 2009)*:

$$r_i = \log(\frac{1}{\hat{\pi}_{desc}}).\frac{1}{\hat{\pi}_{imp}}$$

Risk Estimation

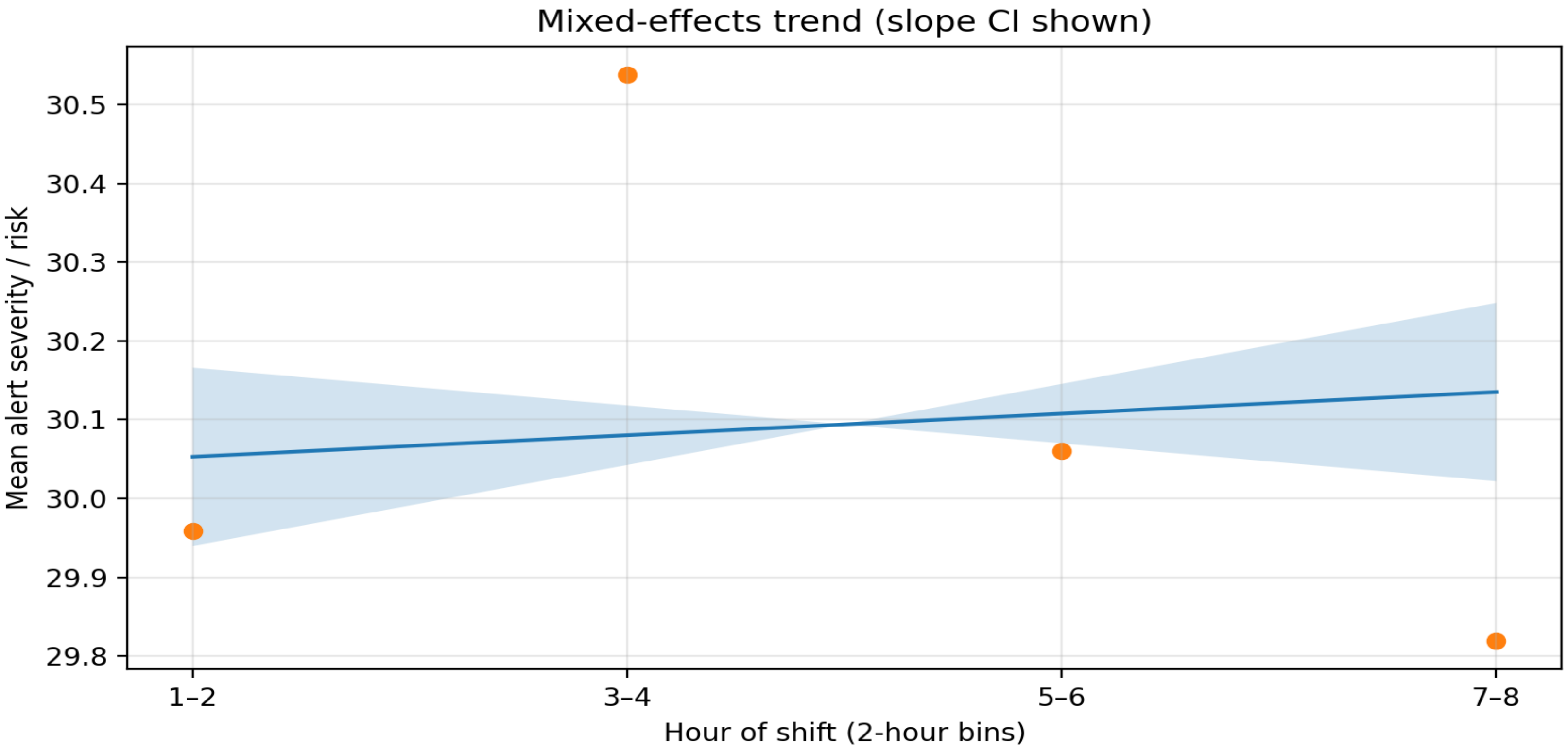


Composite Risk Score *(Cash 2009)*:

$$r_i = \log(\frac{1}{\hat{\pi}_{desc}}).\frac{1}{\hat{\pi}_{imp}}$$

Risk Stationarity

Risk is stable across shift hours (no meaningful drift)



Behavioral Model

Exponential Decay:

Mackworth (1948), Warm et al. (2008)

$$P_i(\Psi, r, x) = \exp(-\hat{\eta}_i)$$

$$\hat{\eta}_i = Fatigue + Risk + Cumaltive\ Risk + Recovery + Exposure + Pharmacist + Shift$$

Behavioral Model

Exponential Decay:

Mackworth (1948), Warm et al. (2008)

$$P_i(\Psi, r, x) = \exp(-\,\hat{\eta}_i)$$

$$\hat{\eta}_i \;= \boxed{Fatigue} + Risk + Cumaltive\ Risk + Recovery + Exposure + Pharmacist + Shift$$

Behavioral Model

Exponential Decay:

Mackworth (1948), Warm et al. (2008)

$$P_i(\Psi, r, x) = \exp(-\,\hat{\eta}_i)$$

$$\hat{\eta}_i \; = Fatigue \boxed{+\,Risk} + Cumaltive\ Risk + Recovery + Exposure + Pharmacist + Shift$$

Behavioral Model

Exponential Decay:

Mackworth (1948), Warm et al. (2008)

$$P_i(\Psi, r, x) = \exp(-\,\hat{\eta}_i)$$

$$\hat{\eta}_i \;= Fatigue + Risk + Cumaltive\ Risk + Recovery + Exposure + Pharmacist + Shift$$

Behavioral Model

Exponential Decay:

Mackworth (1948), Warm et al. (2008)

$$P_i(\Psi, r, x) = \exp(-\hat{\eta}_i)$$

$$\hat{\eta}_i = Fatigue + Risk + Cumaltive\ Risk + Recovery + Exposure + Pharmacist + Shift$$

Behavioral Model

Exponential Decay:

Mackworth (1948), Warm et al. (2008)

$$P_i(\Psi, r, x) = \exp(-\hat{\eta}_i)$$

$$\hat{\eta}_i = Fatigue + Risk + Cumaltive\ Risk + Recovery + Exposure + Pharmacist + Shift$$

Behavioral Model

Exponential Decay:

Mackworth (1948), Warm et al. (2008)

$$P_i(\Psi, r, x) = \exp(-\hat{\eta}_i)$$

$$\hat{\eta}_i = Fatigue + Risk + Cumaltive\ Risk + Recovery + Exposure + Pharmacist + Shift$$

Behavioral Model

Exponential Decay:

Mackworth (1948), Warm et al. (2008)

$$P_i(\Psi, r, x) = \exp(-\hat{\eta}_i)$$

$$\hat{\eta}_i = Fatigue + Risk + Cumaltive\ Risk + Recovery + Exposure + Pharmacist + Shift$$

Behavioral Model

5-fold CV AUC: 0.76

↓27% by 100 alerts

Stationary risk

Alert Fatigue

Term	Coef. Estimate	Feature Importance * $\times 1e-4$
Intercept	0.696318^{***} (0.063483)	NA
Alert index	$\boxed{0.003159^{***}\ (0.000133)}$	446
Alert Risk	-0.003197^{***} (0.000065)	68
Cumulative Alert index	-0.000043^{**} (0.000001)	9
Exposure before shift (learning)	-0.000189^{***} (0.000003)	32
Time between prescriptions (recovery)	0.0000007 (0.006239)	78
Shift Fixed Effect	Included not significant	<5
Pharmacist fixed effect	Included not significant	<5

** Feature importance is LOFO ΔNLL*

Policy Evaluation

Life-threatening alerts: *High-importance + contraindication*

Policy	Risk avoided (%)	Alerts shown	LT accepted (%)
Current	18.98	583,313	9.86
Fixed	25.62	389,717	53.31
Dynamic	26.20	415,923	55.10
Adaptive	27.95	429,668	57.72

Near-optimality of optimal fixed policy

Vs All Alerts	**Alerts**	**LT Reduction**
Optimal Dynamic\ Fixed Policies	~ -33%	~ +40%

2000-4000 lives saved

0.5 Billion in cost across the U.S.

Sensitivity Analysis

Misspecification	Policy	Parameter
Average Risk	Overestimation	$\hat{\lambda}_{exp} = 0.5\ \lambda_{exp}^{true}$
	Underestimation	$\hat{\lambda}_{exp} = 2\ \lambda_{exp}^{true}$
Risk Model	Model Mismatch	-
Fatigue	Overestimation	$\hat{\beta}^{glob} = 2\ \beta^{glob,\,true}$
	Underestimation	$\hat{\beta}^{glob} = 0.5\ \beta^{glob,\,true}$

<1% performance decay

No significant compromise of safety

Our proposed policies **still** perform significantly better

Reduce alerts by **50%**

Summary

Analytical Model A novel fluid approximation model for alert decision-making under operator fatigue

Summary

Analytical Model A novel fluid approximation model for alert decision-making under operator fatigue

Theoretical Results Dynamic vs. Fixed: Alerting is not just a filter. It is a fair attention-allocation policy over time.

Summary

Analytical Model — A novel fluid approximation model for alert decision-making under operator fatigue

Theoretical Results — Dynamic vs. Fixed: Alerting is not just a filter. It is a fair attention-allocation policy over time.

Impact — 2000-4000 lives **and** 0.5 billion in cost saved across the U.S.

Conclusion

Takeaway!

Healthcare staff have a limited ***Cognitive Budget***!

Takeaway!

Healthcare staff have a limited ***Cognitive Budget***!

Use it wisely by optimizing **human-AI interaction**

Hossein Piri
hossein.piri@ucalgary.ca

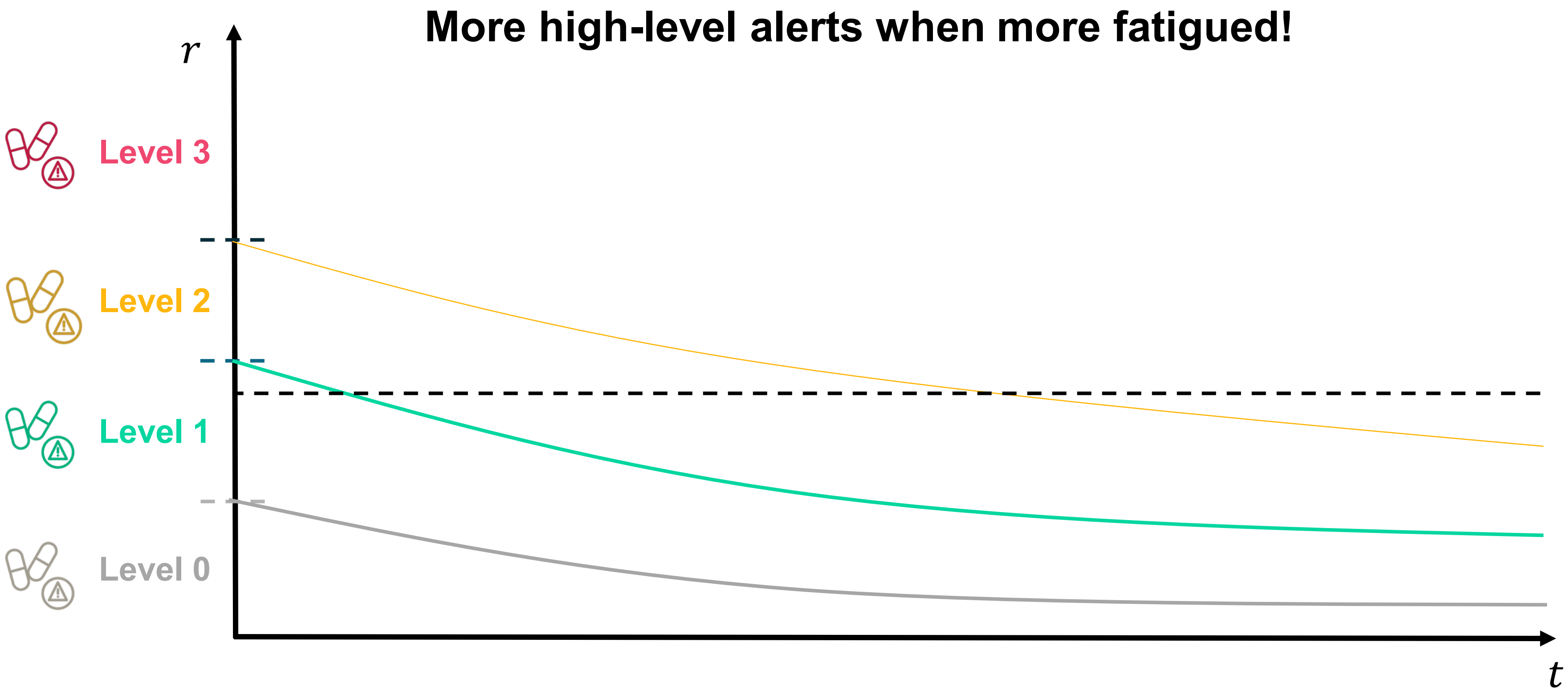
Michael Lingzhi Li
mili@hbs.edu

Thank You! Questions?

Structure of the Optimal Solution

Theorem

Under very mild conditions on the linking function $g(\cdot)$, the optimal solution to the alert fatigue problem always follows a threshold policy, in which the threshold decays over time.



Impact of Model Parameters

Corollary

The optimal risk threshold, $r_1^*(t)$, under Pareto, exponential and uniform risk distribution:

$$r_{1p}^*(t) = \lambda_p[(1 + \frac{\alpha\beta c}{\alpha - 1}(T - t))^{\frac{1}{\alpha}} - 1]$$

$$r_{1e}^*(t) = \frac{1}{\lambda_e}\ln(\beta c(T - t) + 1)$$

$$r_{1u}^*(t) = \frac{b\, \beta c\, (T - t)}{\beta c\, (T - t) + 2}$$

Mild Dependence on Costs (c, β) **Linear Dependence on Scale (b, λ_e, λ_p)**