São Paulo 2025

SECTIONS

INSTITUTO BRASILEIRO DE ATUÁRIA

Can General-Purpose Models Outperform Specifically Tailored Models? Evidence from Life Insurance Underwriting: A Case of Medical Examination Sampling

Chin-Jung Yeh

Cathay Life Insurance

Content

About Cathay Life Insurance & Speaker

About Cathay Life Insurance

國泰人壽
Cathay Life Insurance

Company Snapshot

- Largest insurance company in Taiwan
- Offers individual life, health, unit-linked, and group insurance products

Number of Customers	Num. of In-force Policies
8 million	20 million

Total Assets (US$)	Number of Tied Agents
280 billion	24,000

Awards & Recognition

IIC ASIA 2025

Insur-Innovator Connect Awards 2025

Digital Transformation Trailblazer Award

ITC ASIA — 2025 & 2024

ITC Asia Insurer Awards

(2025) Digital Transformation Trailblazer Award
(2024) Data and Analytics Master Award

Celent. Model Insurer 2024 — 2024 & 2023

Celent Model Insurer Awards

Data, Analytics and AI

Asia Insurance Review — Member: Beacon International Group, Ltd.

AIIA 2022
Digital Insurer of the Year

About the Speakers

Chin-Jung Yeh

Data Analyst,
Data and AI Development Department

Chia-An Wang

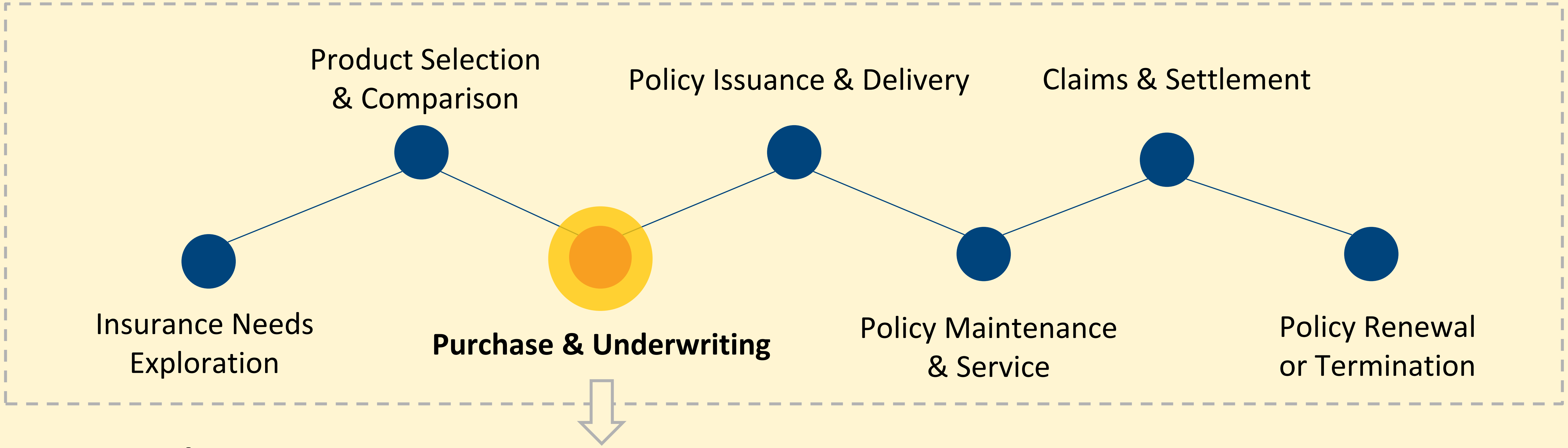
Senior Business Analyst,
Data and AI Development Department

Introduction & Research Motivation

São Paulo 2025

Case Study: Medical Examination Sampling

Insurance Journey



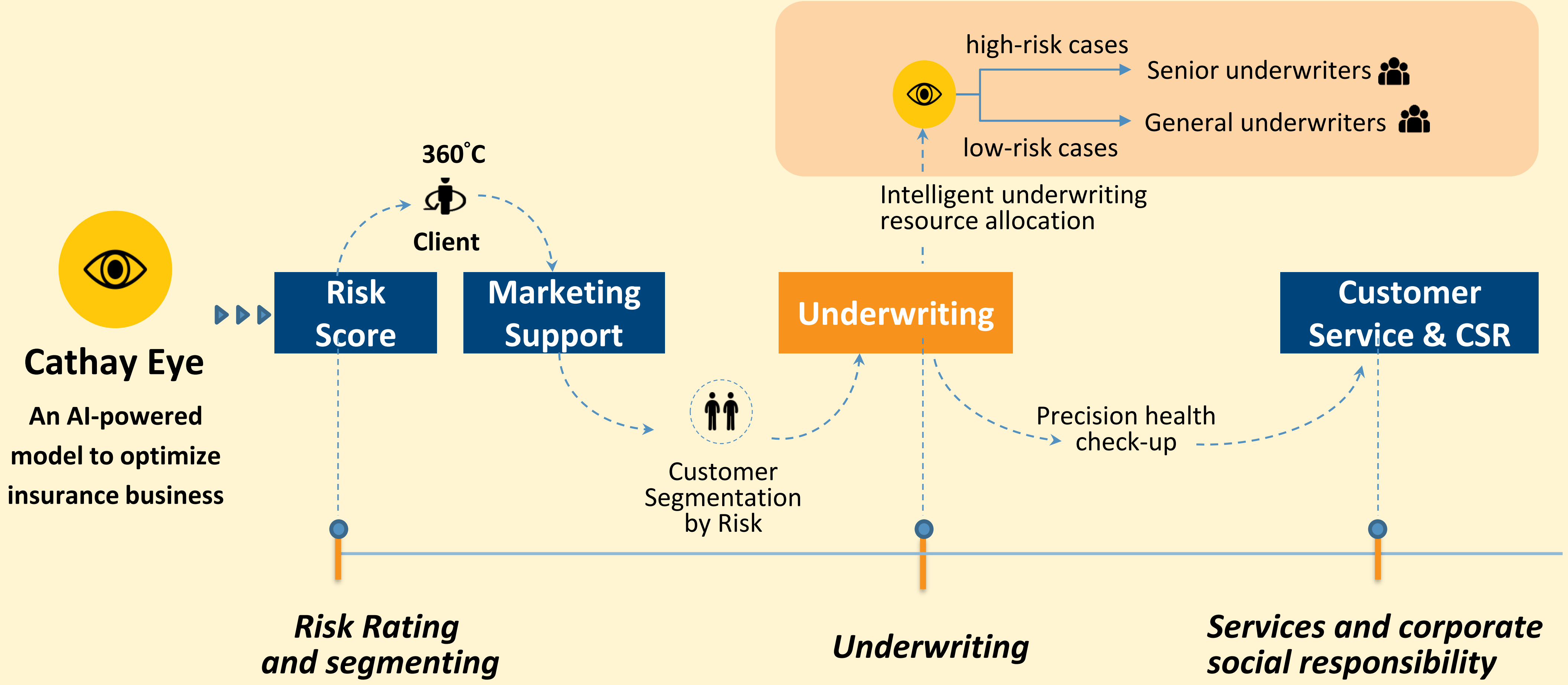
Medical Examination Sampling in Insurance

WHAT	WHY	HOW
Medical examination sampling is an essential risk control measure used by insurance companies to verify that policyholders' health status aligns with the provided insurance information.	This approach deters the concealment of medical history or health conditions by policyholders, thereby reducing potential claim risks and moral hazards, while improving actuarial and underwriting accuracy.	Insurance companies randomly select a certain percentage of policyholders for health examinations based on criteria such as age, insured amount, and health status.

Research Objectives

Our company has developed a predictive model (Cathay Eye) aimed at optimizing the entire insurance process.

【Applications】
- Risk assessment within the underwriting process
- Screening for high-risk medical examination sampling
- Allocation decisions for underwriting cases



Research Motivation

In practical modeling, we often start from specific business needs. Choosing whether to use the policy or the policyholder as the prediction unit directly impacts the model's ability to accurately reflect the insured's health risk—and ultimately determines how well it can support risk management and underwriting decisions.

Research Purpose

To explore and validate whether the specialized model delivers superior predictive performance in Medical examination sampling, supported by empirical data to address these concerns.

Can General-Purpose Models Outperform Specifically Tailored Models?

Experiment Overview

Models Comparison

	Specifically tailored models (STM)	General-purpose model (GPM)
Key Characteristics	Focus on the attributes and risk characteristics of **the contract** itself.	Focus on the attributes and risk characteristics of **the customer.**
Description	Contracts exhibit relatively high volatility. Based on this characteristic, a Rolling Window Training model is designed to ensure that the model continuously reflects the latest changes in data distribution.	Customers generally exhibit relatively low volatility. The model features include historical customer data and behavioral patterns, allowing the model to maintain robust performance and stability without requiring annual retraining.
Feature Selection	Features are selected from the contract perspective.	Features are selected from a customer perspective.
Feature Explanation	Based on the contract, using contract-related risk factors.	Based on the customer, the model utilizes customer-related risk factors to provide a comprehensive view of the overall risk profile.
Application Scope	Single contract risk assessment, applicable to underwriting claims and similar areas.	Broad customer risk segmentation, applicable in underwriting, claims, marketing, and related scenarios.

Experiment Overview

Model Training Framework

1. Raw Data (X variable)

Past policyholder data on new contracts, policy maintenance, and claims.

2. Feature Engineering

3. Raw Data (Y variable)

Average hospitalization days over the next three years.

4. Processing Policy Exposure

5. Modeling

- LGBM
- Model Training Structure

6. Risk Segmentation

- Model labeling
- High-risk and low-risk

7. Model Validation

- Models' RMSE

Real-World Evaluation

8. Application Validation

- Relative ratios in hospitalization days
- Health abnormality rate
- Claim ratio

Models : Data & Training Comparison

Specifically tailored models (STM)

	2015	2016	2017
N	≈ 113K	≈ 113K	≈ 95K
X (Features)	142		
Y (Target **(days)**)	Male: 1.74 Female: 1.57	Male: 1.71 Female: 1.54	Male: 1.68 Female: 1.65
RMSE	Male: 2.50 Female: 2.39	Male: 2.50 Female: 2.35	Male: 2.50 Female: 2.47

From 2015 to 2017, each year collects around one hundred thousand insurance contracts.

Average hospitalization days per policy projected over the next three years.

General-purpose model (GPM)

	2015/01/01
N	≈ 4.35M
X (Features)	349
Y (Target **(days)**)	Male: 2.48 Female: 2.10
RMSE	Male: 2.80 Female: 2.63

Data from 4 million policyholders as of 2015/01/01.

Predicted average hospitalization days per customer over the next three years.

JOCO São Paulo 2025

Model Training Structure

2015 2016 2017 2018 2019 2020 2021 2022 Present

Specifically tailored models (STM)

Train the model using the contracts from the entire year of 2015 along with the observable average number of hospitalization days over the next three years for each contract.

STM 1: Policies in 2015 — Y Tracked Over 3 Years (2015 Issue) — Policies in 2018 — Y Tracked Over 3 Years (2018 Issue)

STM 2: Policies in 2016 — Y Tracked Over 3 Years (2016 Issue) — Policies in 2019 — Y Tracked Over 3 Years (2019 Issue)

STM 3: Policies in 2017 — Y Tracked Over 3 Years (2017 Issue) — Policies in 2020 — Y Tracked Over 3 Years (2020 Issue)

Training data

A **Rolling Window training strategy** is used in STM, where the model is trained using a moving time window that updates annually.

In this study, we validate the model using three consecutive years of data, **resulting in three STM models: STM 1, STM 2, and STM 3**.

General-purpose model (GPM)

Train the model using the contracts from the entire year of 2015 along with the observable average number of hospitalization days over the next three years for each contract.

Policyholders as of 2015/01/01 — Y Tracked Over 3 Years

Policyholders as of 2018/01/01 | Policyholders as of 2019/01/01 | Policyholders as of 2020/01/01 — Y Tracked Over 3 Years

Training data

Training

Forecasting

Experiment Results & Performance Evaluation

São Paulo 2025

Models: Model Output

Model Output

Risk Label

1. Predicts the average hospitalization days over the next three years for each contract or customer.
2. Groups the data based on gender and age.
3. Sorts the predicted values within each group and divides them into 1000 segments.
4. Converts the predictions into risk labels using threshold values from each segment.
5. **Risk labels range from 1 to 1000, with higher values indicating greater risk.**

Gender	Age	Risk Score	Risk Label
Male	12	2.34	380
Female	36	3.18	120
Male	58	4.60	970

Male/12yrs

Percentage	Threshold (RS)
1	0.42
2	0.47
3	0.49
…	…
999	28.99

Models: Validation Dataset

2015 2016 2017 2018 2019 2020 2021 2022 Present

Specifically tailored models (STM)

During the model validation phase, risk labels from 2018, 2019, and 2020 are used to ensure no overlap between the training and testing data. This approach allows the model to fully assess its generalization capability on unseen data.

STM 1: Policies in 2015 — Y Tracked Over 3 Years (2015 Issue) — Policies in 2018 — Y Tracked Over 3 Years (2018 Issue)

STM 2: Policies in 2016 — Y Tracked Over 3 Years (2016 Issue) — Policies in 2019 — Y Tracked Over 3 Years (2019 Issue)

STM 3: Policies in 2017 — Y Tracked Over 3 Years (2017 Issue) — Policies in 2020 — Y Tracked Over 3 Years (2020 Issue)

validation risk label

General-purpose model (GPM)

The model validation phase utilizes the risk labels from 2018, 2019, and 2020. 。

Policyholders as of 2015/01/01 — Y Tracked Over 3 Years

Policyholders as of 2018/01/01 | Policyholders as of 2019/01/01 | Policyholders as of 2020/01/01 — Y Tracked Over 3 Years

validation risk label

Training

Forecasting

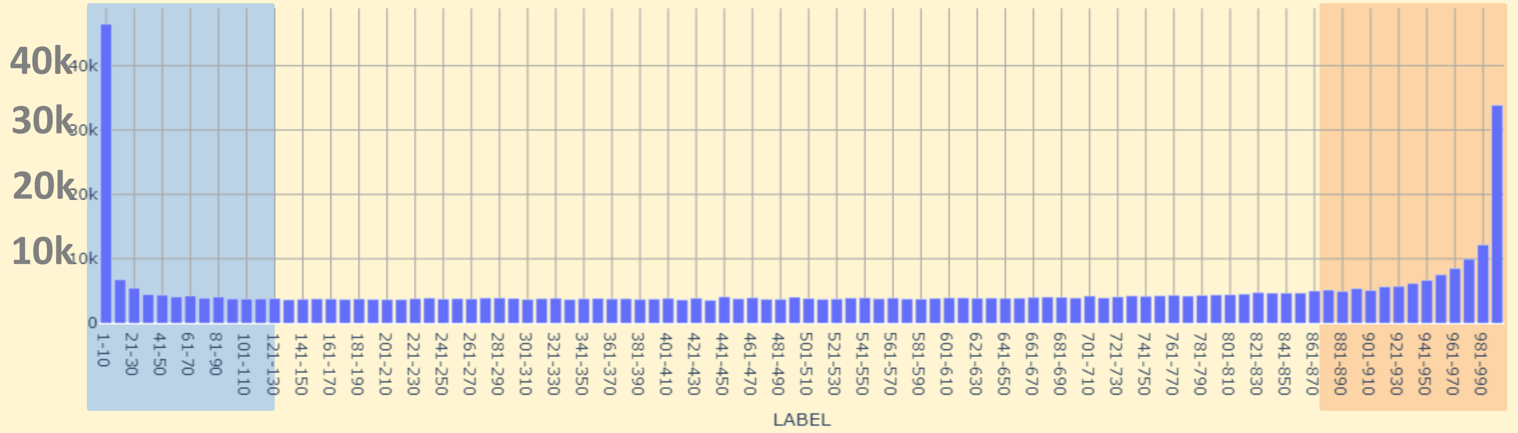
Risk Groups Definition

- Differences in risk distribution between the two models.
- We focus on their ability to differentiate between high-risk and low-risk groups.

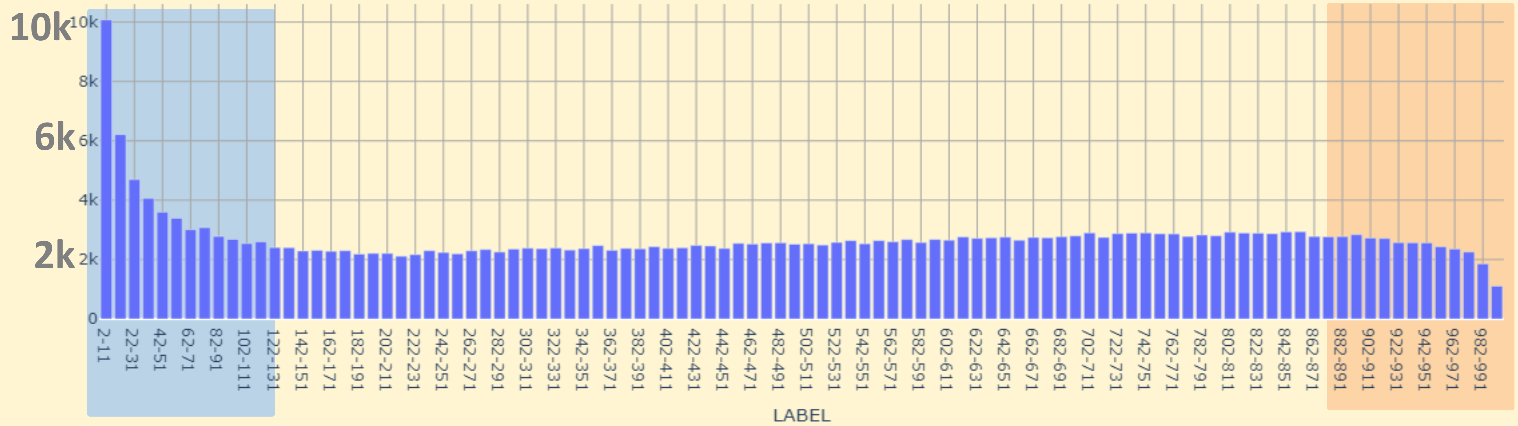
High and Low Risk Grouping Method

- The top 20% in risk ranking are designated as the high-risk group.
- The bottom 20% in risk ranking are designated as the low-risk group.
- This grouping method is applied to subsequent indicator validation.

Specifically tailored models (STM)



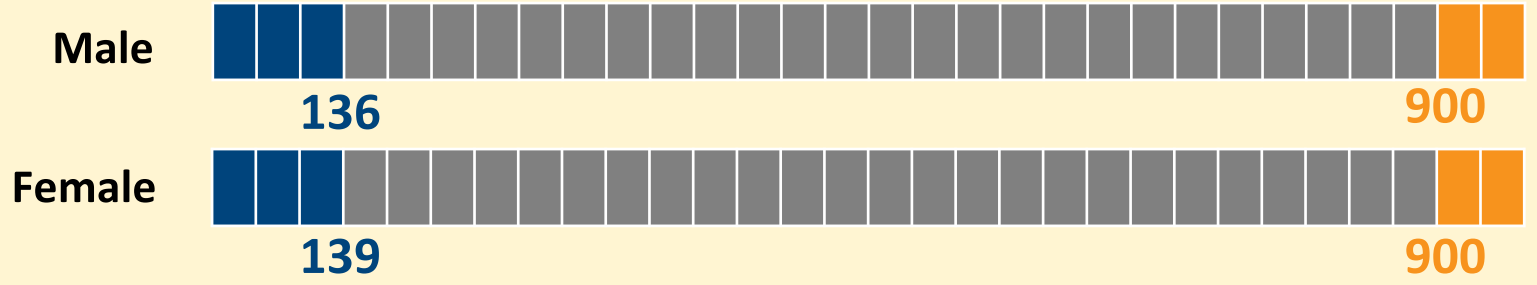
General-purpose model (GPM)



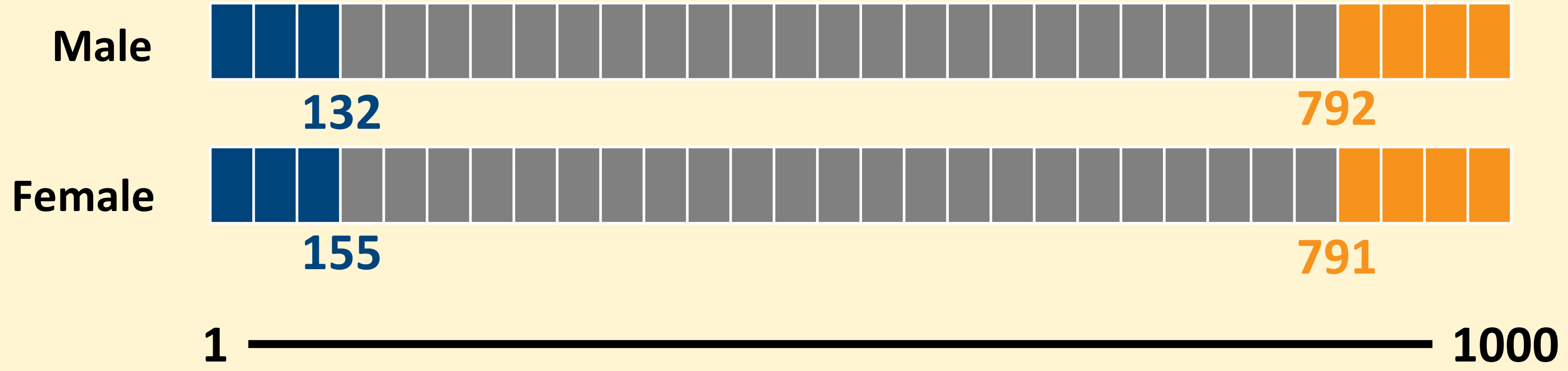
STM - H/L Threshold

Low-Risk Group (Bottom 20% Risk Ranking)

High-Risk Group (Top 20% Risk Ranking)



GPM - H/L Threshold



Application Validation Metrics

Relative Ratios in hospitalization days

$$\frac{(High\ Risk)Average\ Hospitalization\ Days\ over\ the\ Next\ 3\ Years}{(All)\ Average\ Hospitalization\ Days\ over\ the\ Next\ 3\ Years}$$

Assess how much higher the average hospitalization days are for contracts/ customers predicted as high risk compared to the overall contracts/ customers.

Health Abnormality Rate

$$\frac{(High\ Risk)\ Post-Screening\ Outcomes:\ Rejection\ or\ Special\ Approval}{(All)\ Post-Screening\ Outcomes:\ Rejection\ or\ Special\ Approval}$$

Evaluate how much higher the abnormal screening hit rate is for contracts/customers predicted as high risk compared to the overall contracts/ customers.

Claim Ratio

$$\frac{(High\ Risk)Claim\ Ratio\ over\ the\ Next\ 3\ Years}{(All)Claim\ Ratio\ over\ the\ Next\ 3\ Years}$$

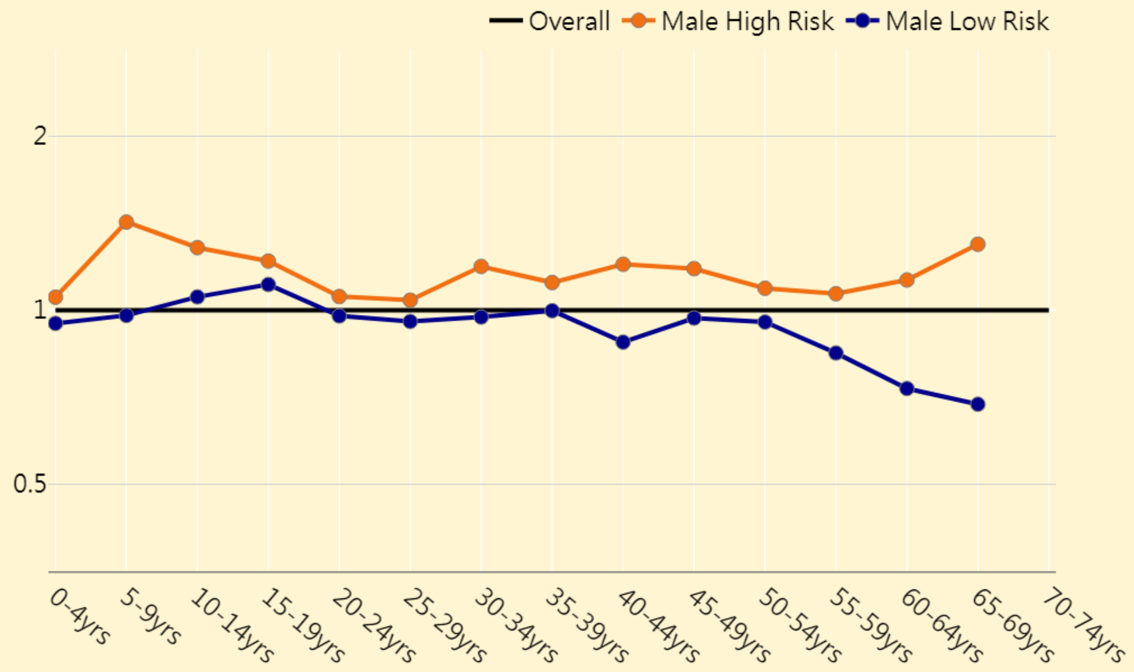
Examine how much higher the claim ratio is for contracts/customers predicted as high risk compared to the overall contracts/ customers.

Performance Comparison
- Relative Ratios in hospitalization days

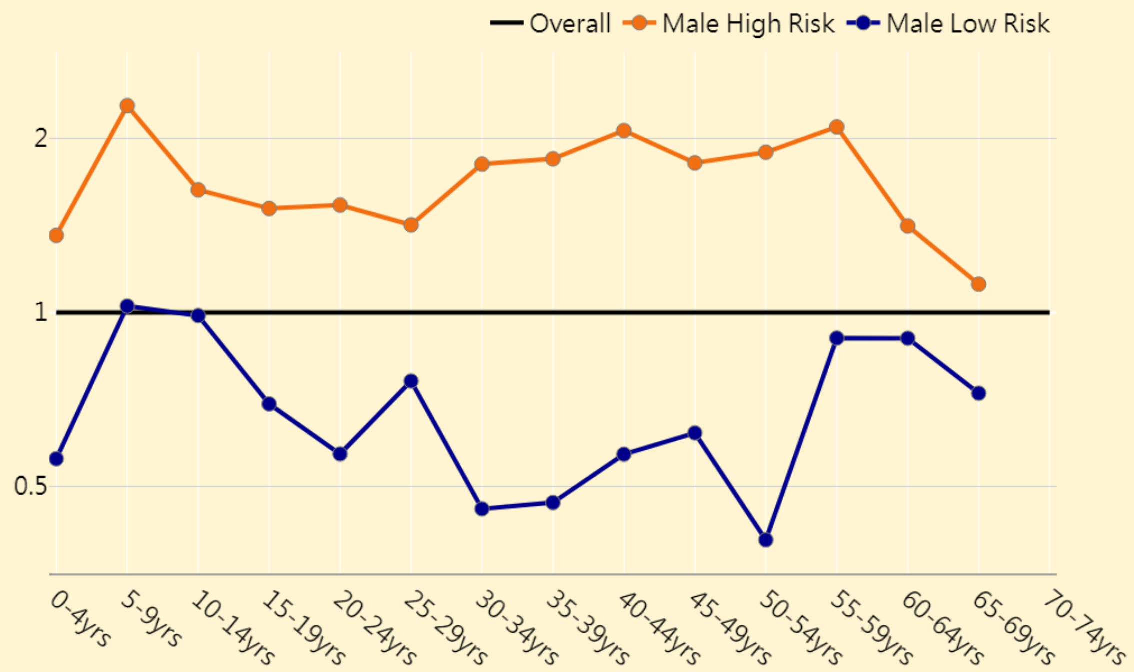
- **GPM demonstrates better hospitalization risk differentiation between the high-risk and low-risk groups.**
- The high-to-low risk ratio for STM is around 1.2 times.
- The high-to-low risk ratio for GPM is around 2.2 times.

Gender	Male		Female	
Model	**STM**	**GPM**	**STM**	**GPM**
High Risk	1.19	1.17	1.08	1.49
Low Risk	0.97	0.53	0.99	0.69
H/L Ratio	1.23	2.21	1.09	2.16

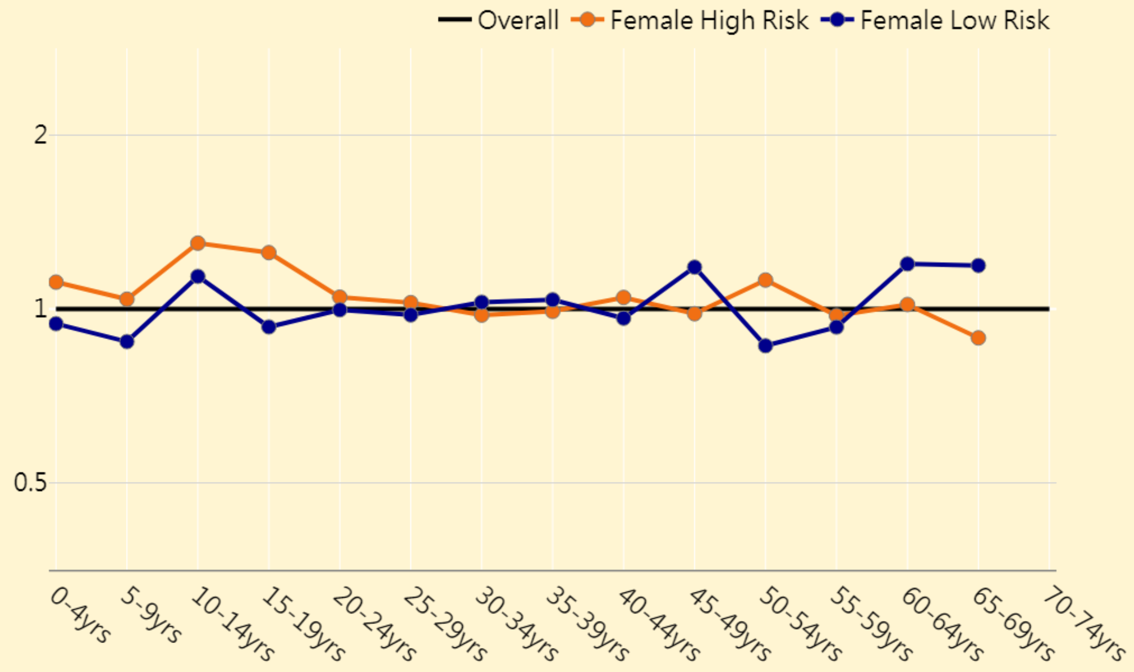
STM - male



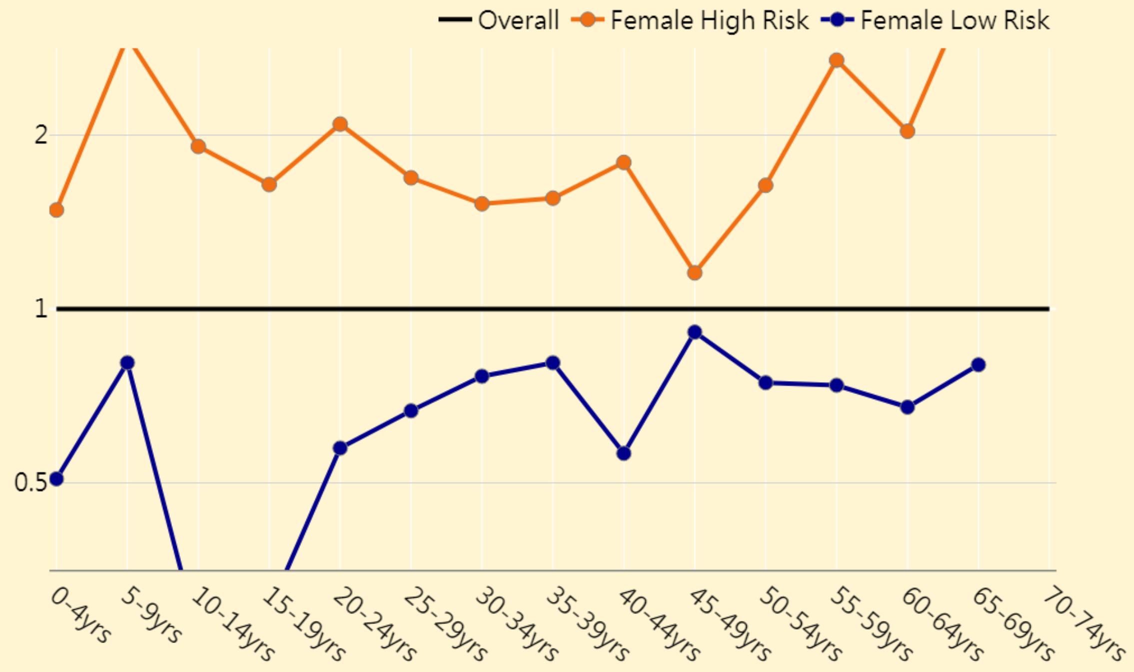
GPM - male



STM - female

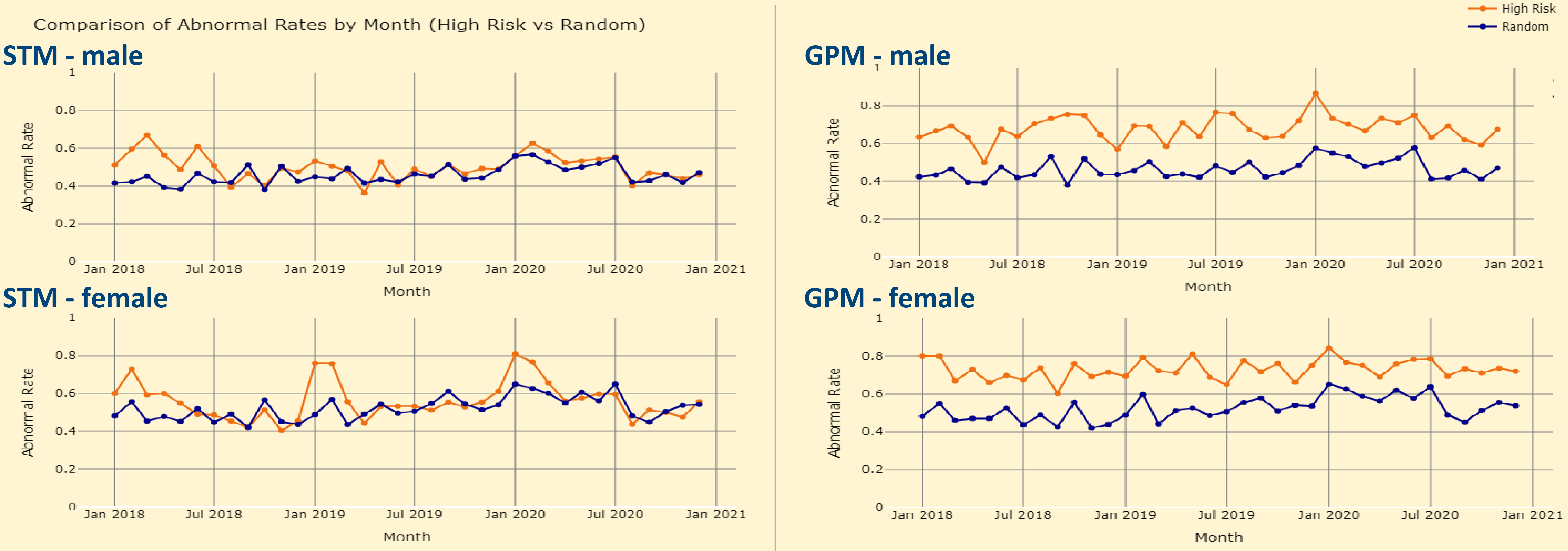


GPM - female

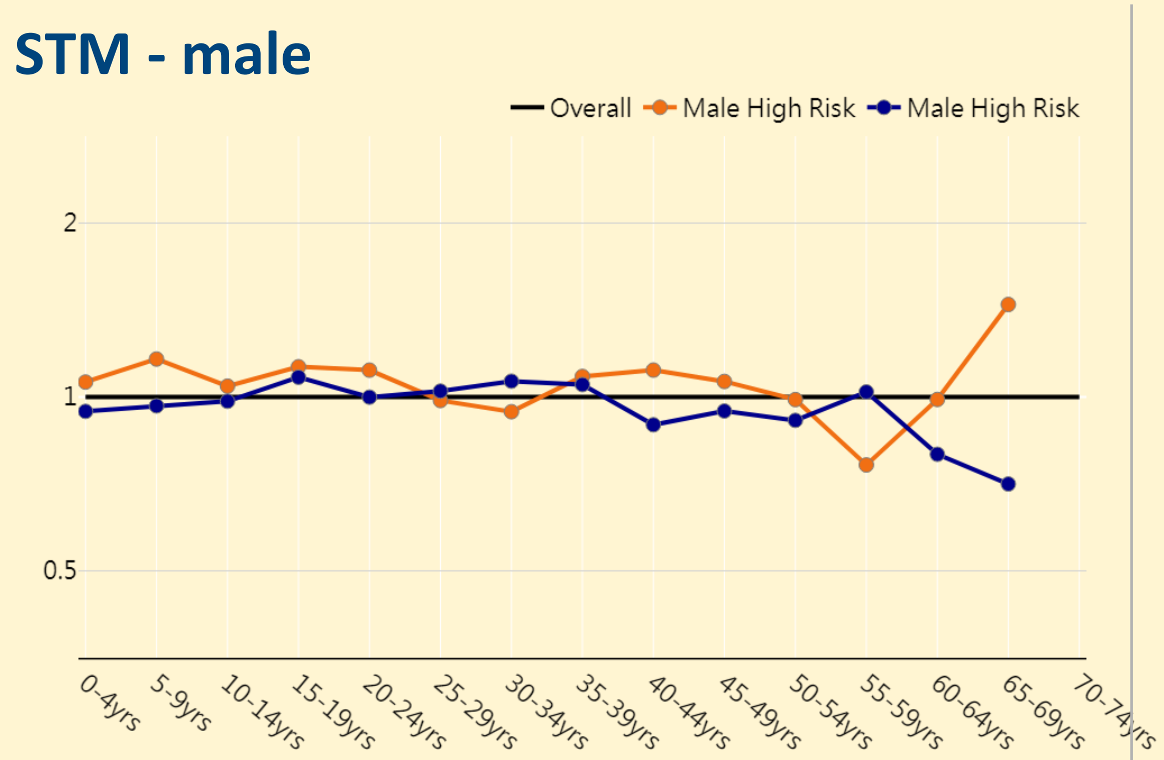


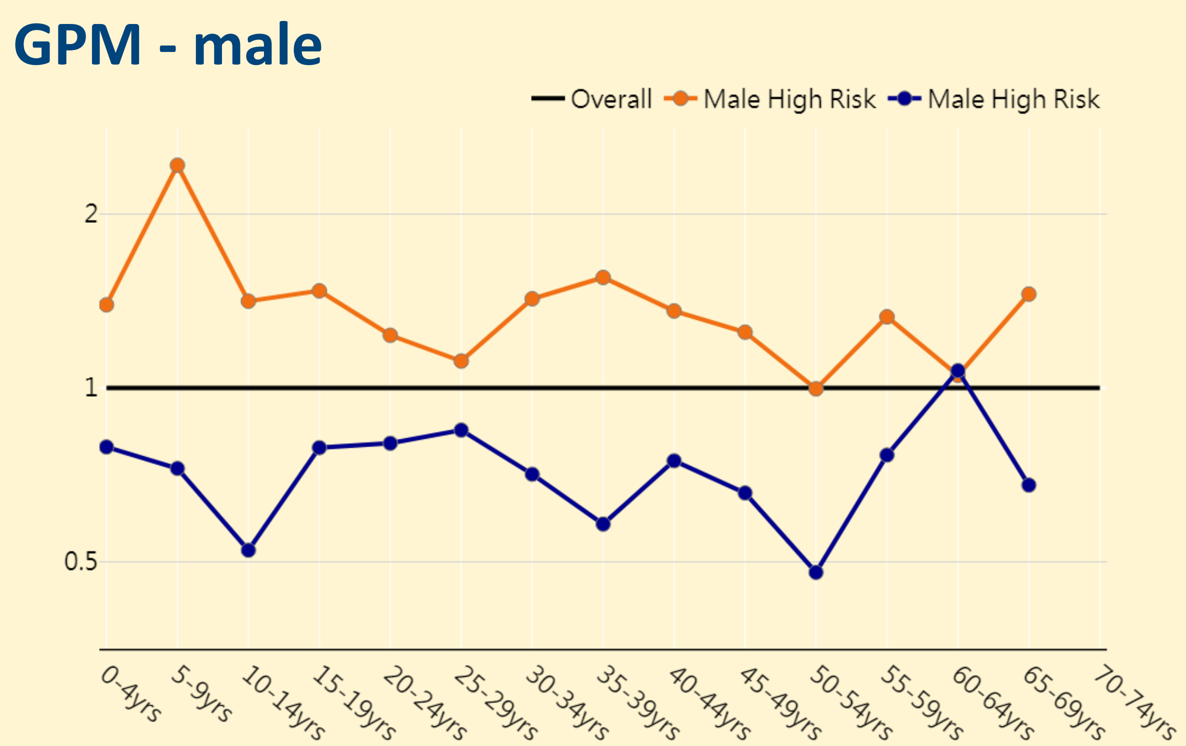
Performance Comparison - Health Abnormality Rate

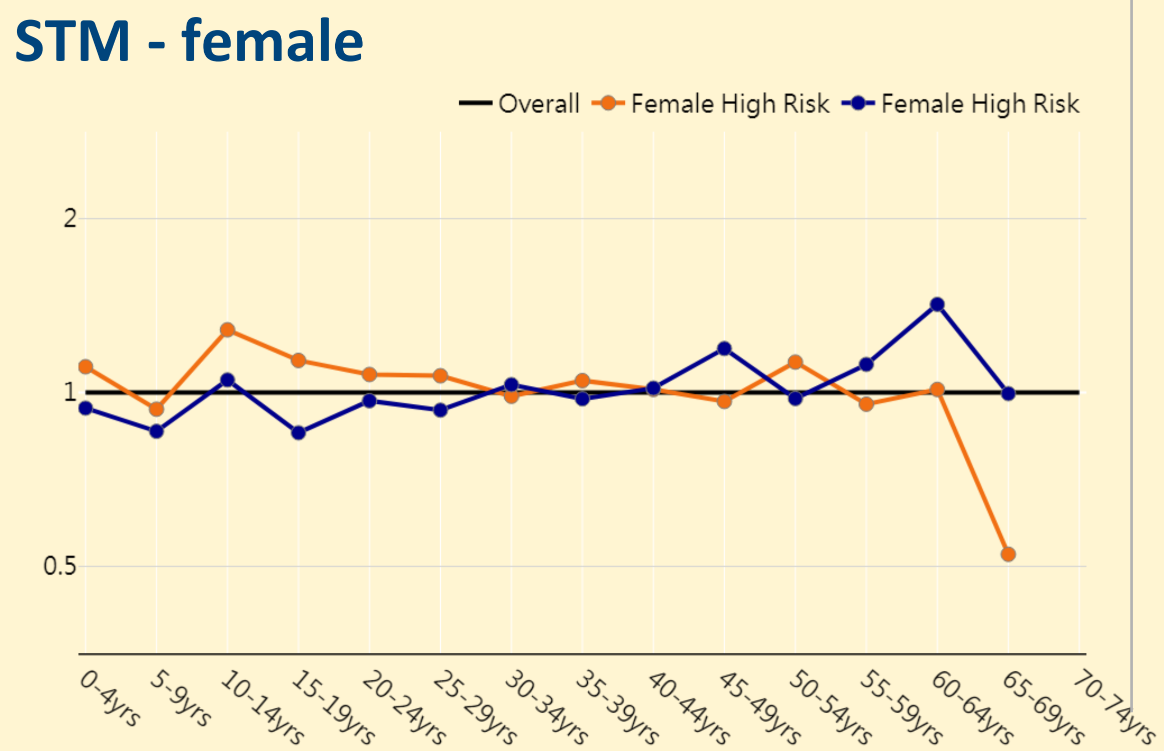
- **Based on data from 2018 to 2020, the health abnormality rate in the high-risk group identified by GPM has consistently been higher than that in the random sampling group.**
- The monthly fluctuations in the health abnormal rate are primarily due to the randomness introduced by sampling at the validation data level.

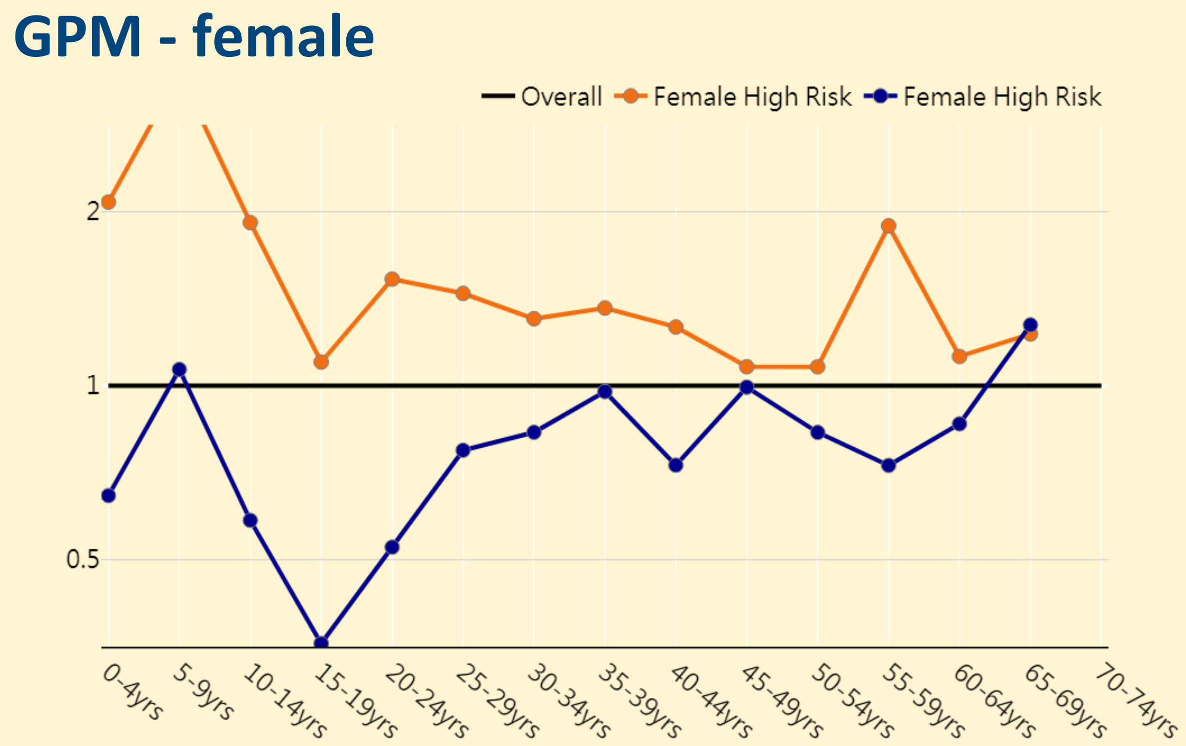


Performance Comparison - Claim Ratio









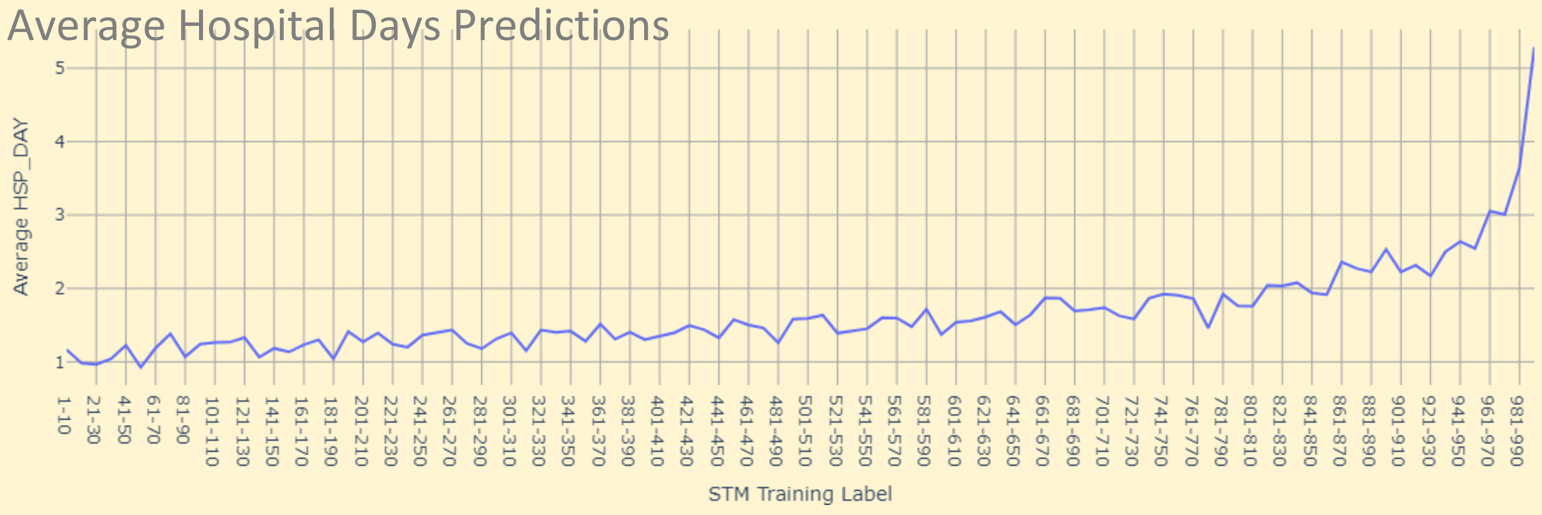
- **GPM demonstrates better claim risk differentiation between the high-risk and low-risk groups.**
- The high-to-low risk ratio for STM is around 1.1 times.
- The high-to-low risk ratio for GPM is around 2 times.

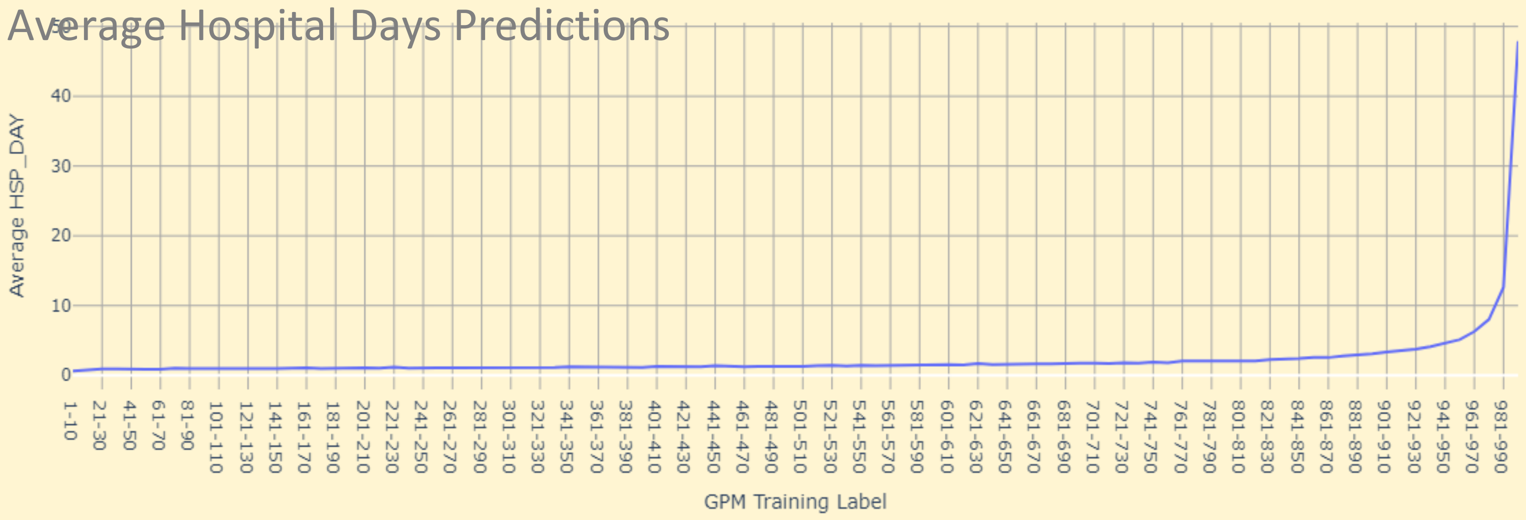
Gender	Male		Female	
Model	STM	GPM	STM	GPM
High Risk	1.08	1.52	1.08	1.48
Low Risk	0.96	0.69	0.99	0.76
H/L Ratio	1.13	2.2	1.09	1.94

Comparative Analysis of Model Predictions and Generalization Ability

In the training data, both the STM and GPM models exhibit comprehensive and smooth prediction curves, indicating that they achieve good fitting within the range of the training dataset.

Training Data's Label

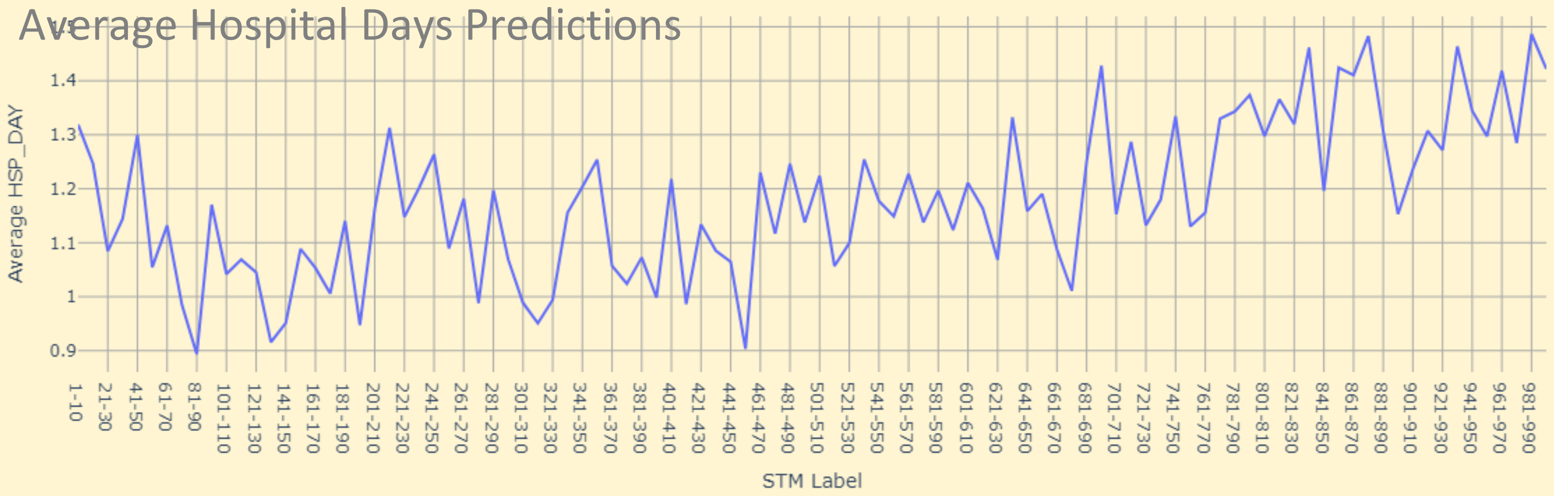


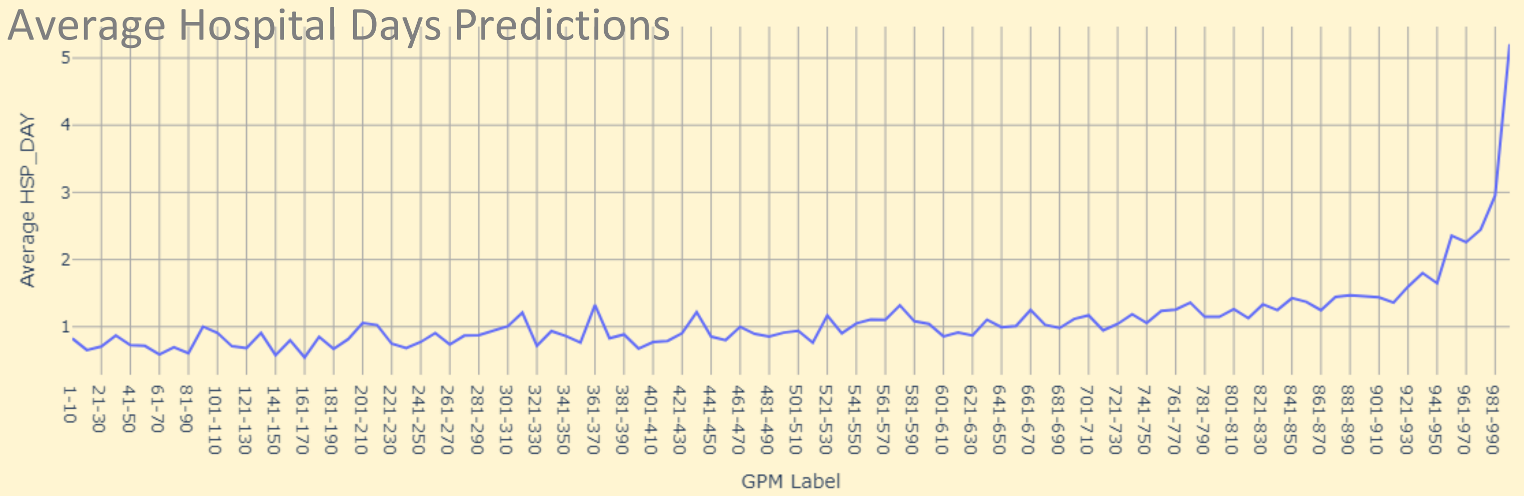


However, in the validation data, we observe significantly increased fluctuations in the prediction curves, **which reflects a decline in the prediction accuracy of the STM model on unseen data.**

Validation Data's Label

Unstable performance due to limited training scale.





Conclusion

Conclusions

GPM outperforms STM across all evaluation metrics.

Relative Ratios in hospitalization days

STM

H

1.2 times

L

GPM

H

2.2 times

L

Claim Ratio

STM

H

1.1 times

L

GPM

H

2 times

L

Health Abnormality Rate

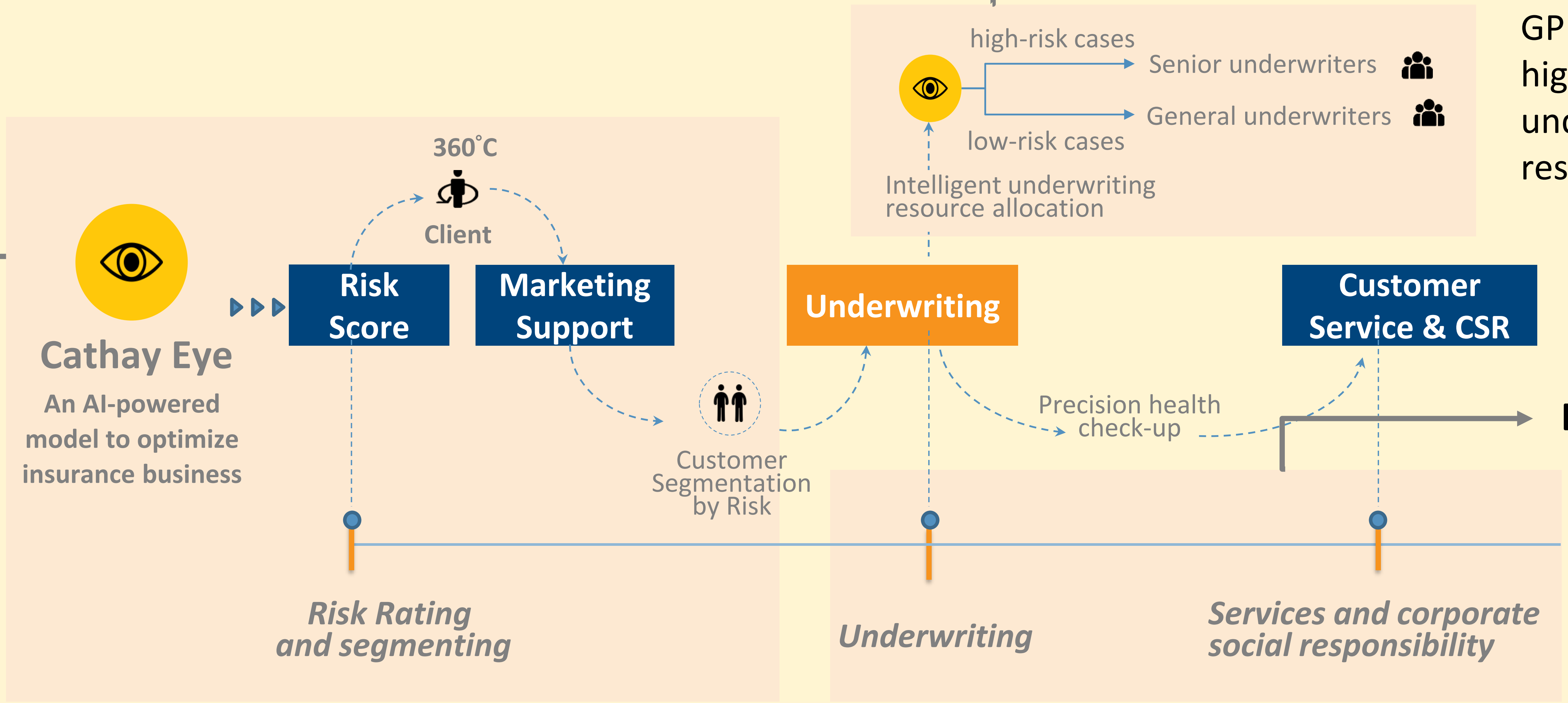
The health abnormality rate for the high-risk group in GPM is consistently 15% to 20% higher than that of the random group.

Business Implications

In practice, we deployed the GPM model architecture across the entire insurance process, achieving remarkable results.

GPM model architecture

By adopting the GPM model architecture, there is no need for annual updates due to highly volatile training data, significantly reducing deployment and maintenance costs.



Underwriter resource allocation

GPM effectively differentiates high-risk and low-risk groups in underwriting, optimizing resource allocation.

Diverse business applications

GPM more easily extends to diverse business applications, providing a universal risk assessment framework that supports precision marketing and long-term risk management strategies.

JOCO São Paulo 2025

Thank you! Obrigado!

Questions?

IAA AAI SECTIONS

IBA INSTITUTO BRASILEIRO DE ATUÁRIA