

BC2406

# DIABETES PREDICTION MODEL

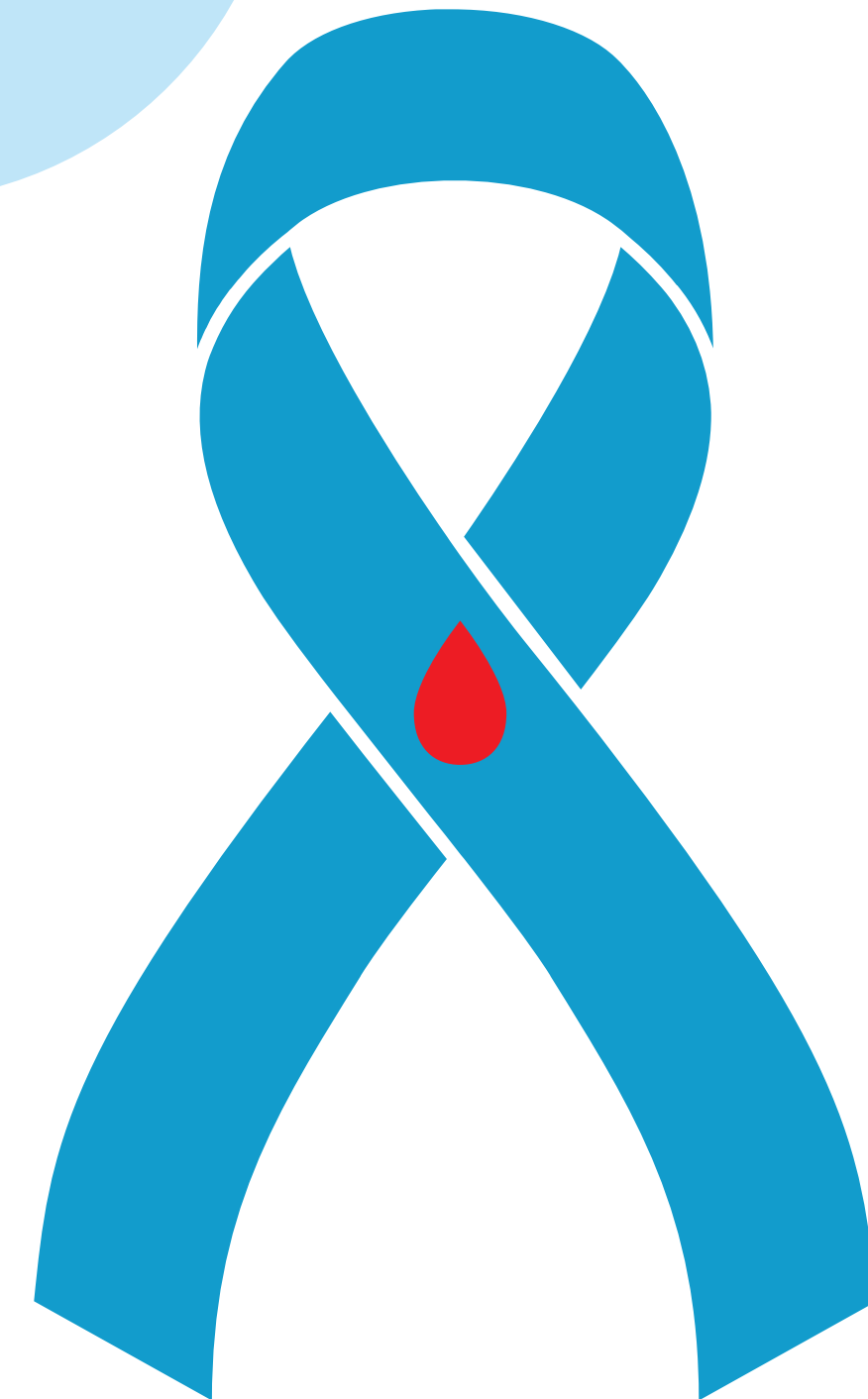
Group 7

ALOYSIOUS LAW JIA JIAN (U2420162J)

BRYAN LIM HOW MENG (U2420547L)

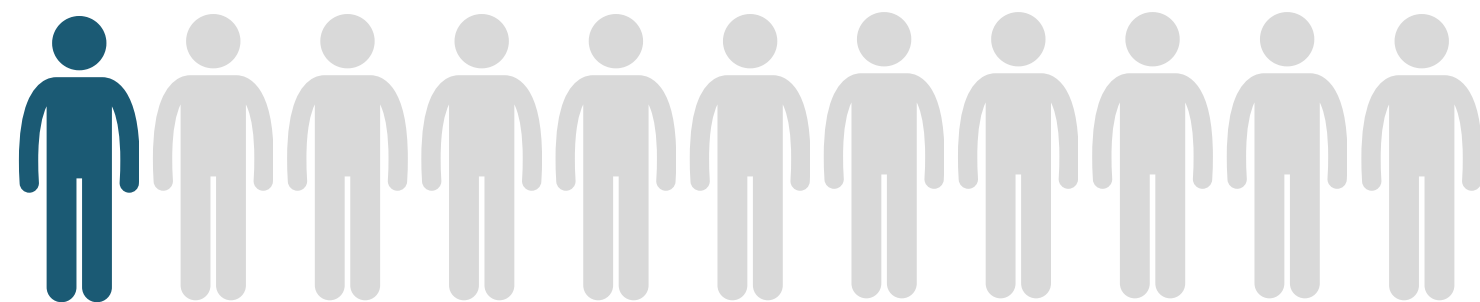
DE SOUZA ALYSSA ANNE (U2210429D)

NG YONG WEI (U2440621H)



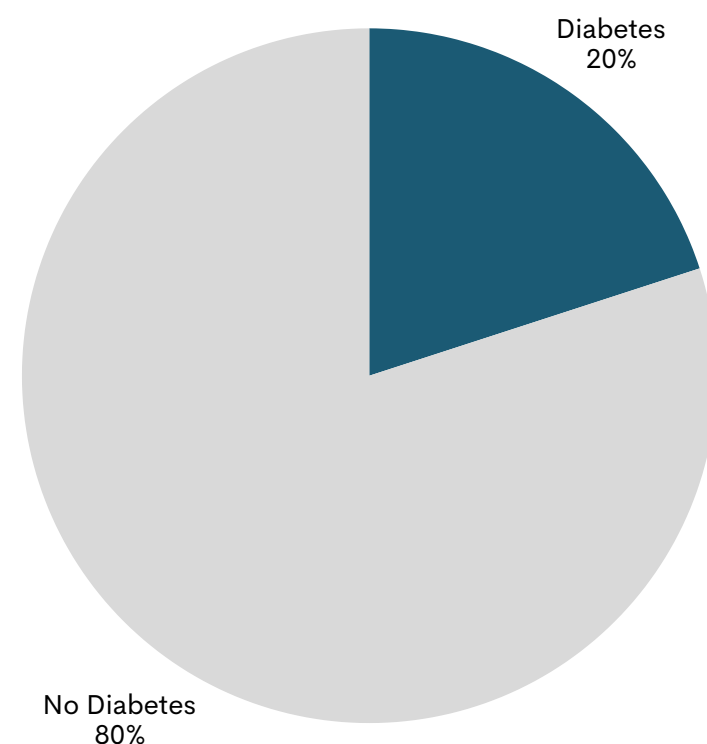
# Diabetes in Singapore

## SOCIAL IMPACT



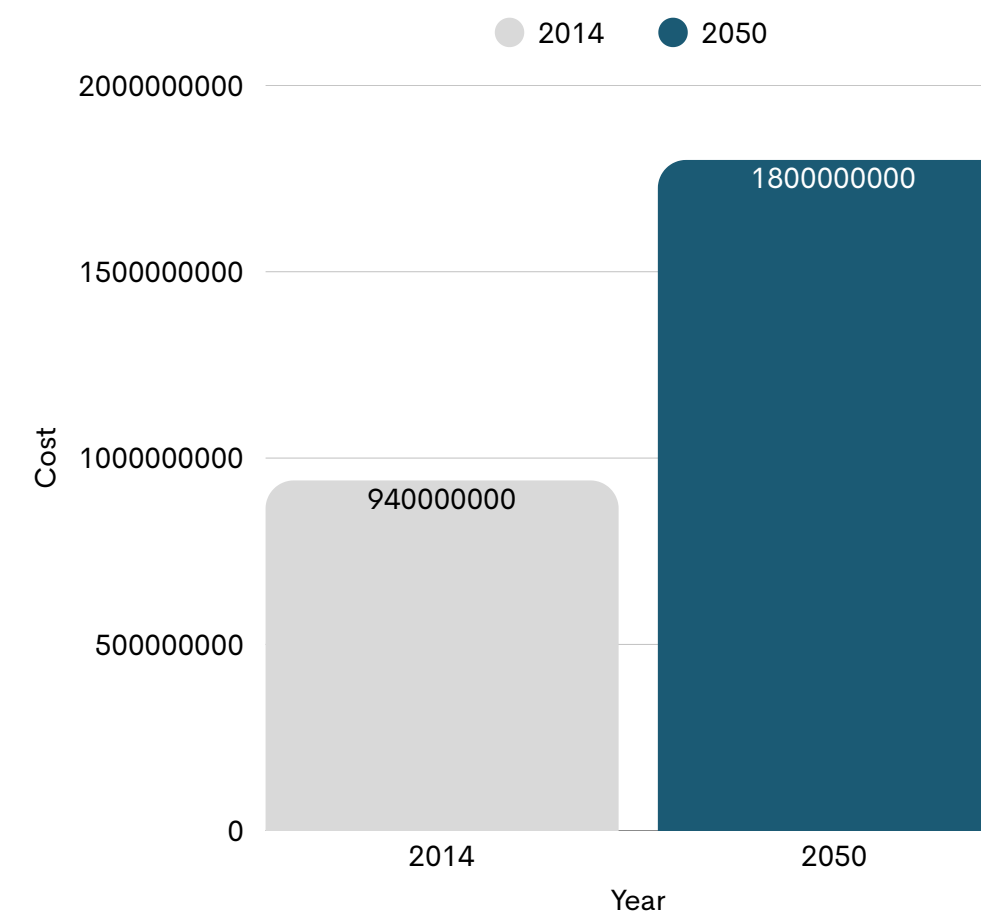
**1 in 11**  
suffer from diabetes in Singapore

Diabetes in older adults (60 - 74)



Among older Singaporeans aged 60 - 74, **20% are diagnosed with diabetes**

## FINANCIAL IMPACT



The financial impact of diabetes, including medical costs and lost productivity, was projected to increase from over \$940 million in 2014 to **\$1.8 billion by 2050.**

# Diabetes in Singapore

## THE PROBLEM

**Many Singaporeans are unaware of their  
personal risk of developing diabetes**

Thus, no preventive measures are taken to lower their risk of developing diabetes

# Dataset Overview

- ✔ Source: **CDC Diabetes Health Indicators**  
→ 253,680 survey responses  
21 features related to chronic disease and lifestyle.
- ✔ Target variable: **Diabetes\_binary**  
→ "0" = Non-diabetic; "1" = Prediabetic/Diabetic

**Balanced Subset Used:**  
70,692 responses, **50-50 class split**  
→ Ensures fair model training

diabetes\_binary\_5050split\_health\_indicators\_BRFSS2015.csv

Diabetes_binary	HighBP	HighChol	CholCheck	BMI	Age	Smoker	HvyAlcoholConsump	DiffWalk	GenHlth
0	1	0	1	26	4	0	0	0	3
0	1	1	1	26	12	1	0	0	3
0	0	0	1	26	13	0	0	0	1
0	1	1	1	28	11	1	0	0	3
0	0	0	1	29	8	1	0	0	2

# Variable Analysis

## 1. Correlation analysis (heatmap)

- Top correlated features:  
**General Health** ( $r=0.294$ ), **HighBP** ( $r=0.263$ ), **BMI** ( $r=0.218$ )
- Heatmaps confirmed **no multicollinearity** (all  $r < 0.6$ ).

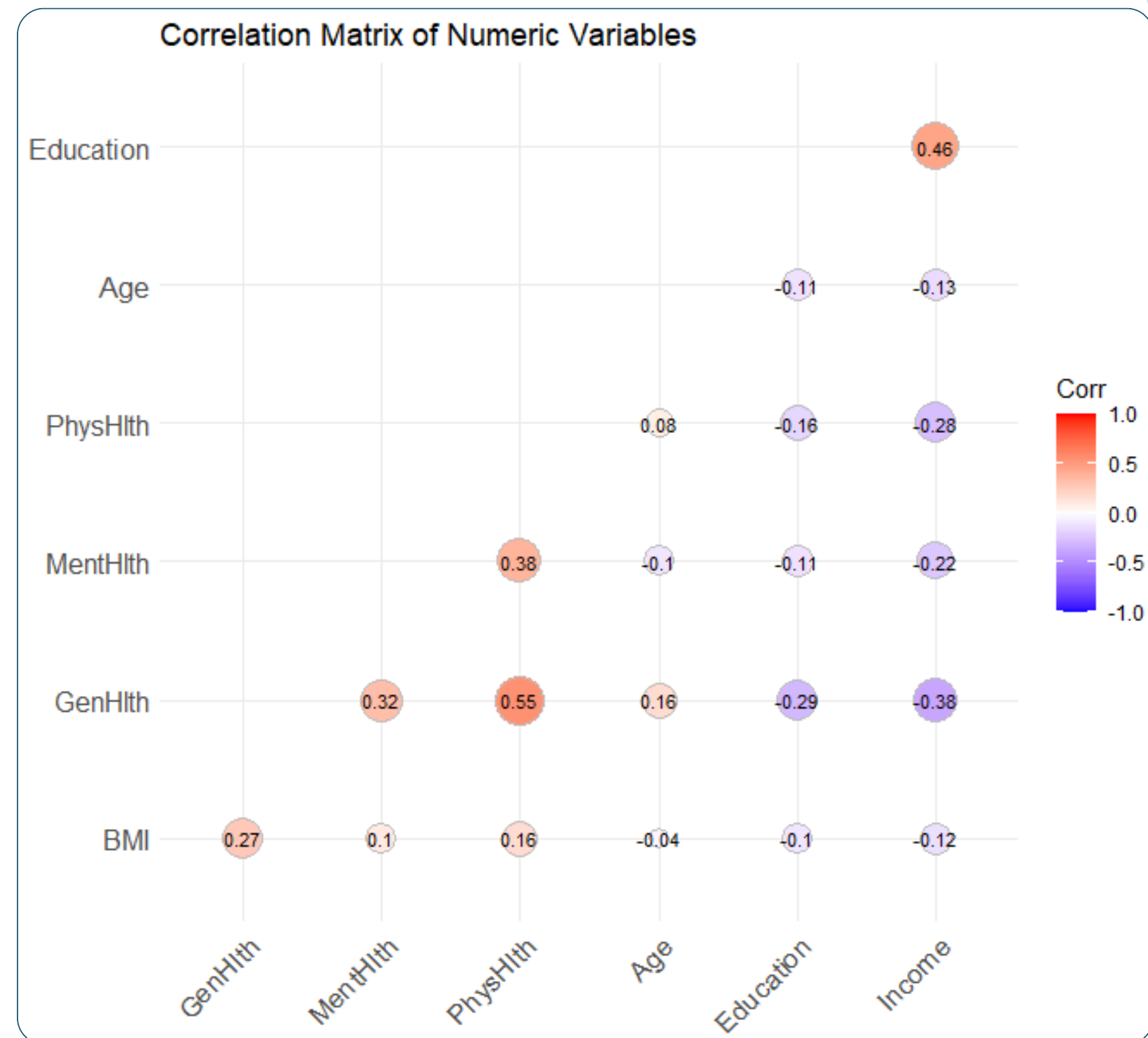
## 2. Statistical tests

### a) Chi-square test:

- Used on categorical variables
- Strongest categorical predictors → **HighBP, DiffWalk, and HighChol**

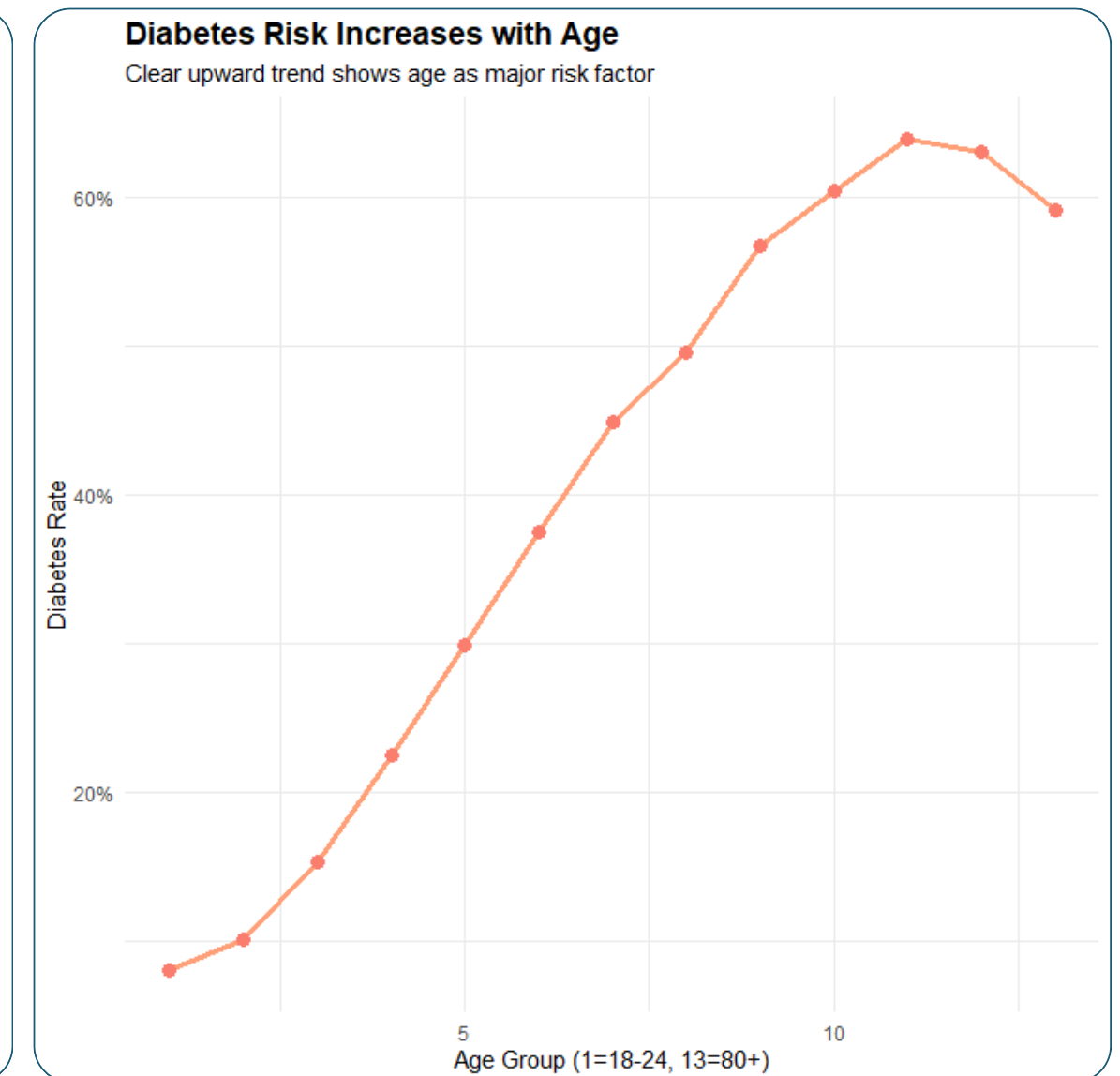
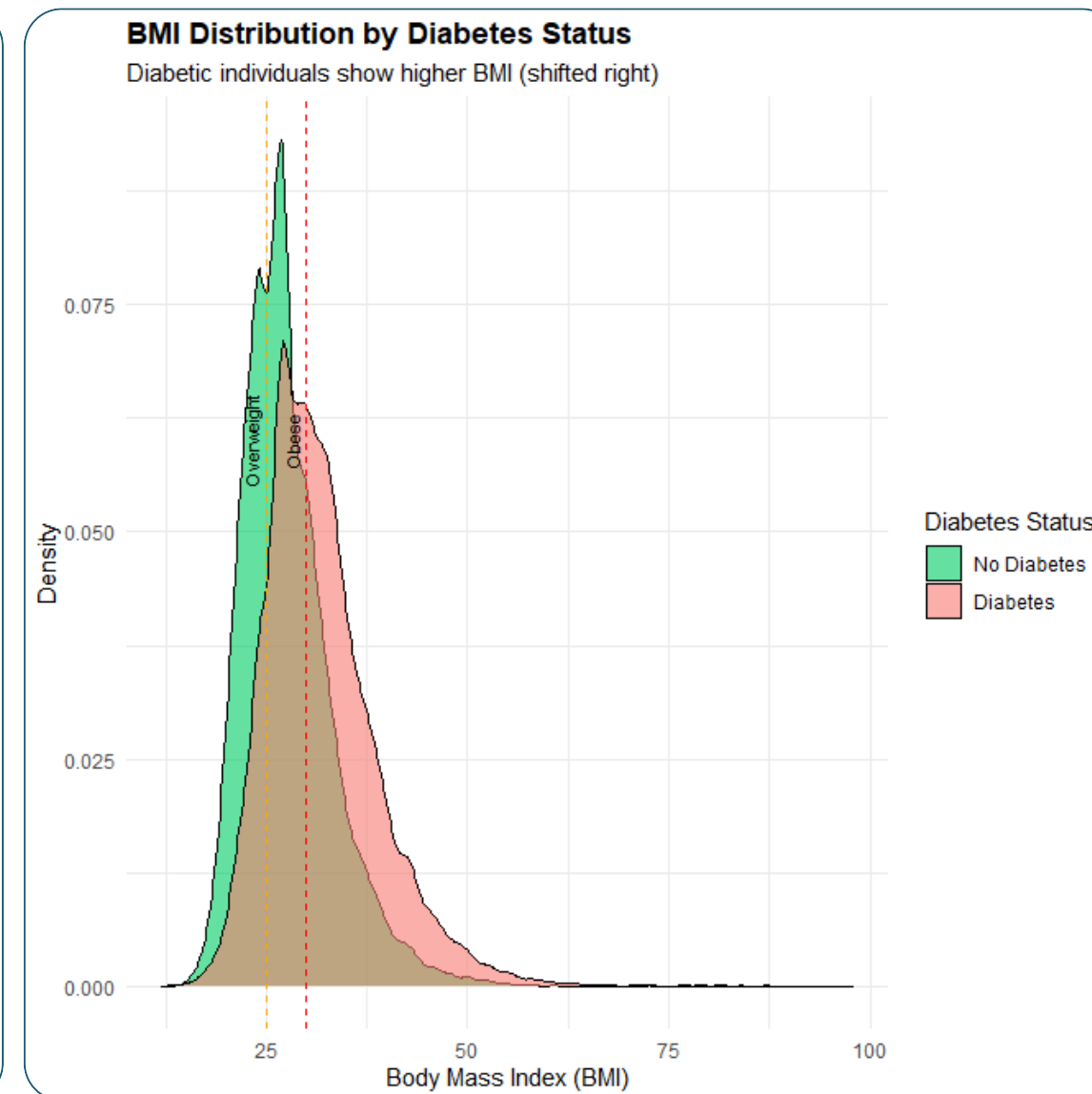
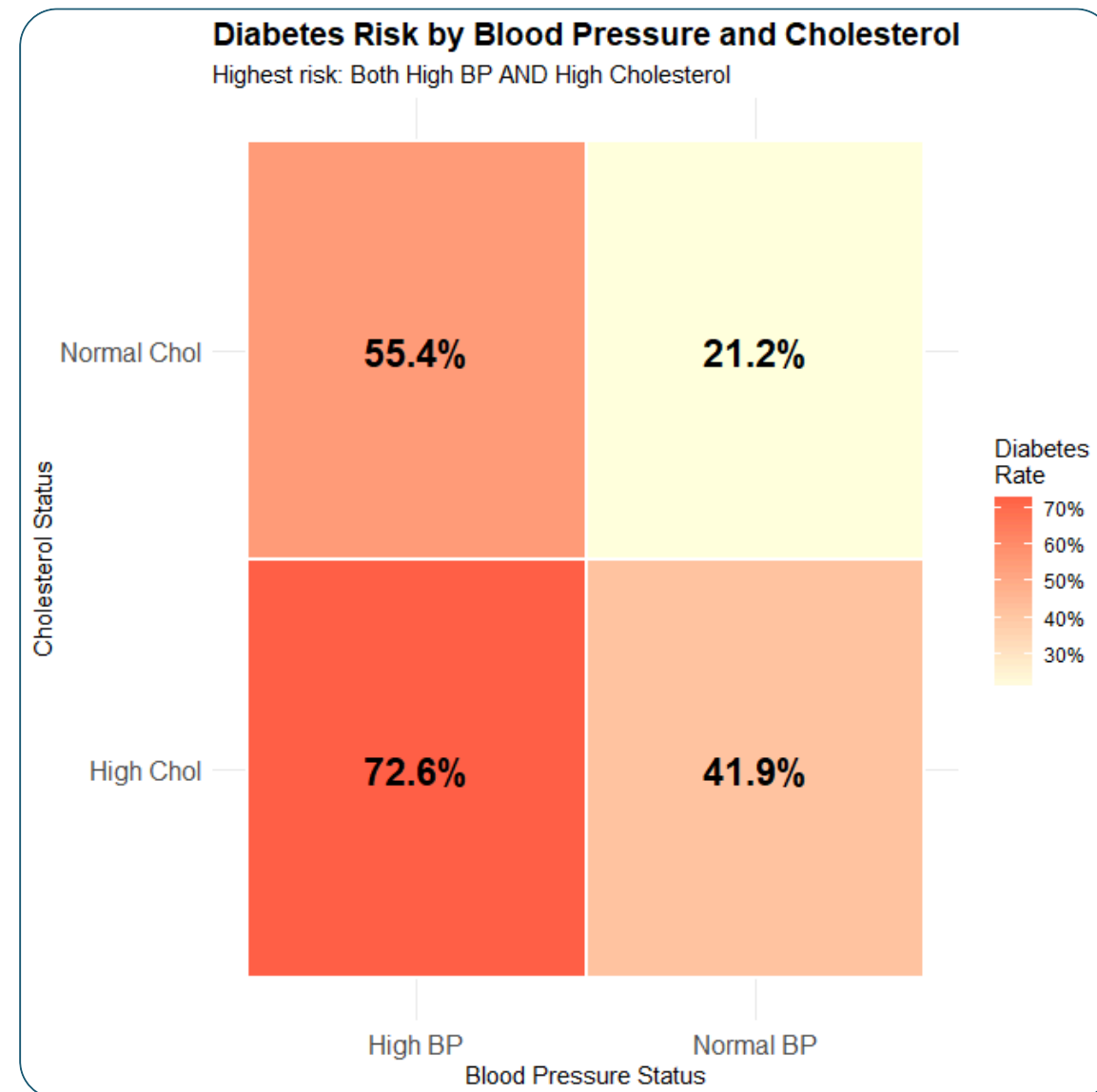
### b) T-test:

- Used on continuous variables
- Factors with  $p < 0.001$  = highly significant  
→ 7 significant continuous predictors



### c) Visualisations for top predictors:

- Strong diabetes risk patterns with BMI, Age, Blood Pressure, and Cholesterol



- Highest risk: **High BP + High Cholesterol**
- Normal levels → lowest risk
- Shows compounding effects of cardiovascular factors
- Diabetic group has **higher BMI** overall
- Most fall into Obese ( $\text{BMI} \geq 30$ )
- Reinforces link between weight and diabetes risk
- Risk **increases steadily** with age
- **Slight dip after 75–80** due to survival bias & underdiagnosis

# Predictive Models

## Logistic Regression Model

### The problem we faced...

For every 100 prediabetic/diabetic patients:



Model with 0.5 threshold

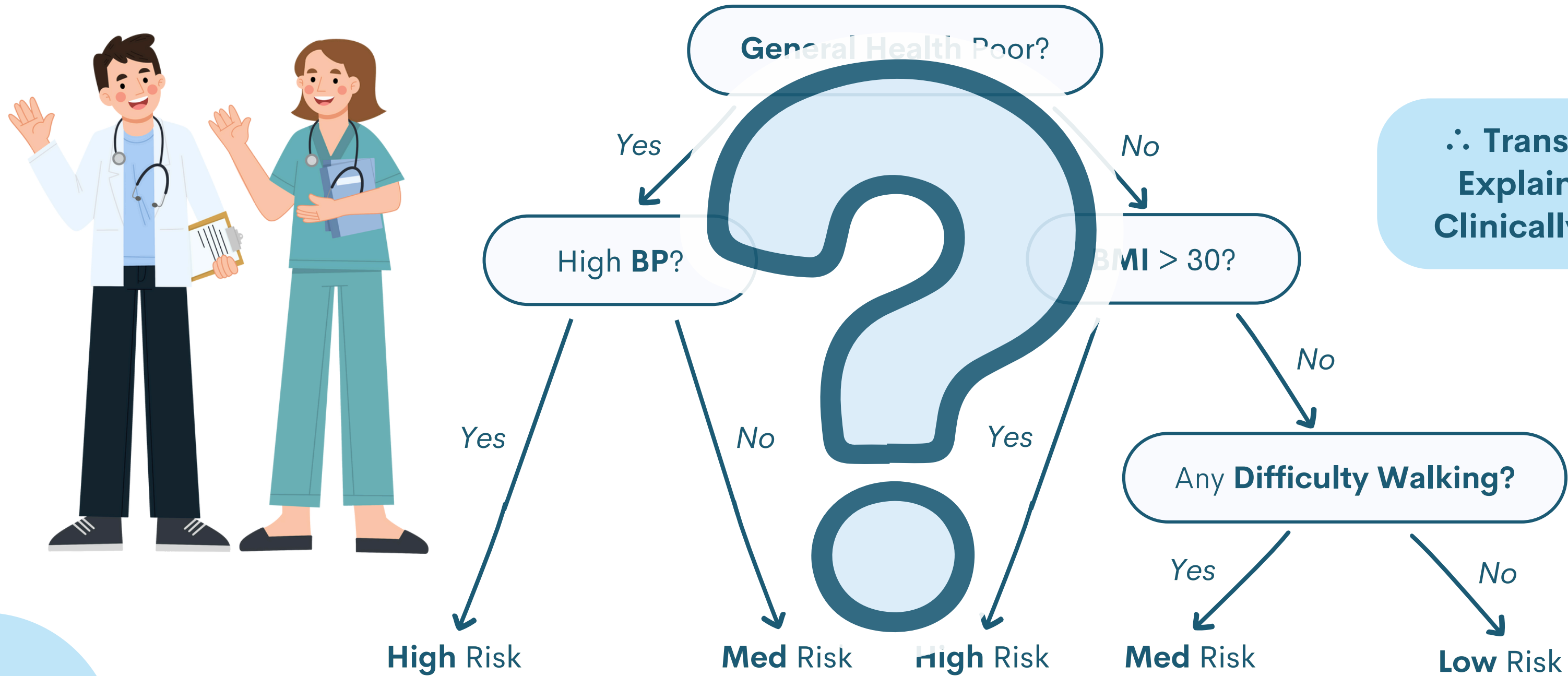
Missing **23% of prediabetic/diabetic cases** is unacceptable in healthcare!

	Before (Threshold 0.5)	After (Threshold 0.35)
<b>Missed Cases</b>	2488	1151 ↓
<b>Detection Rate</b>	76.5%	89.2% ↑

∴ **12.7% increase** in succesful detection  
**46.3% reduction** in missed cases

# CART Model

→ Automated Clinical Reasoning

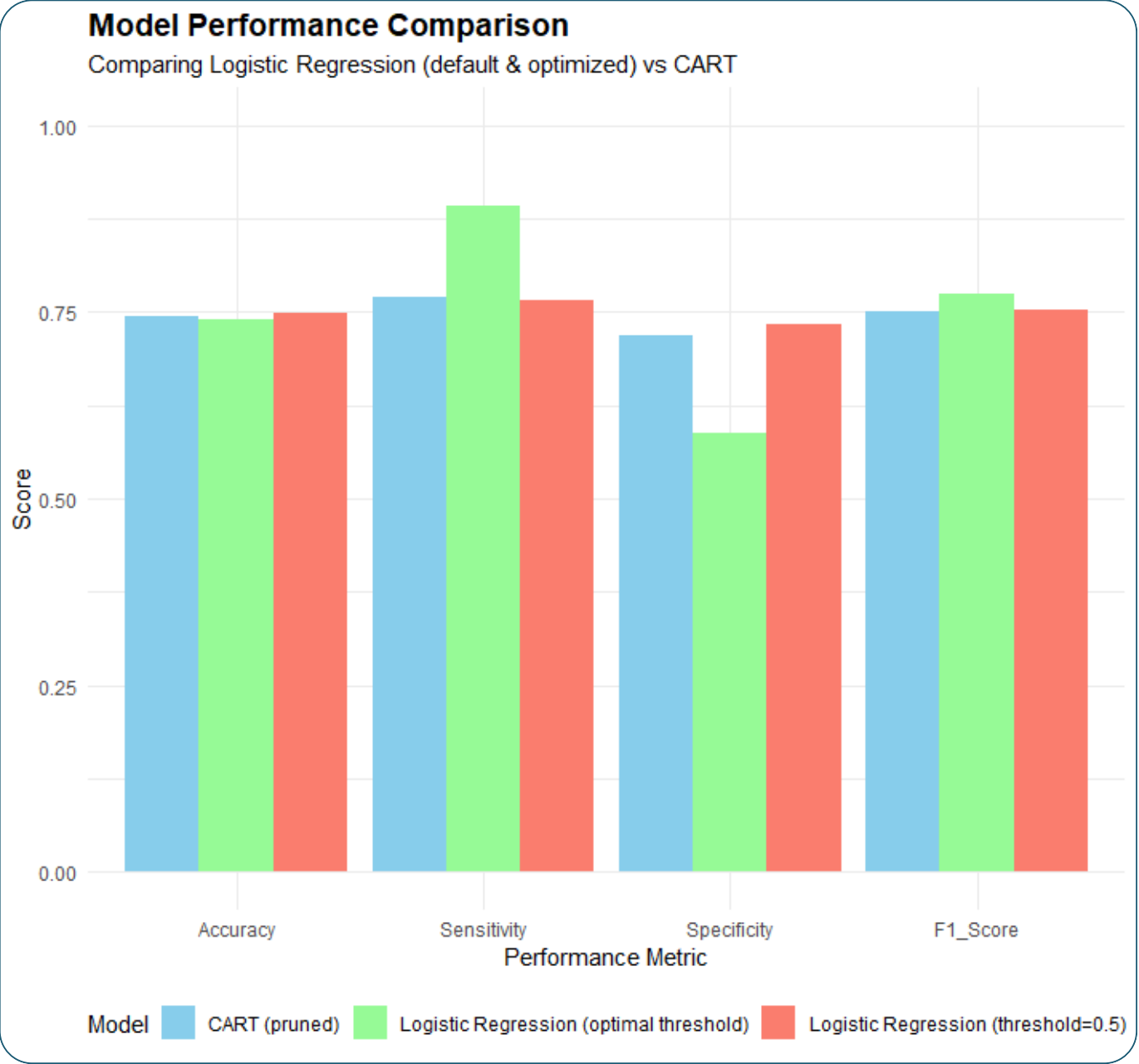




# Models Comparison

Model	Accuracy	Sensitivity	<u>F1 Score*</u>	False Negatives
Logistic Regression (threshold 0.5)	<u>0.7491</u>	0.7654	0.7531	2488
Logistic Regression (threshold 0.35)	0.7401	<u>0.8915</u>	<u>0.7742</u>	<u>1151</u>
CART	0.7444	0.7700	0.7508	2439

\* F1 Score = Mean of Precision and Sensitivity



∴ The Optimised Logistic Regression model has the highest F1 score → **optimal balance**,  
best AUC score → **strong discriminative ability**

## RISK SCORING

- Model coefficients converted into a 0–100 risk score
- Based on all 21 variables from the dataset
- Higher score → higher likelihood of diabetes
- Simple, interpretable, and ready for integration into LifeSG

### Risk Scoring Formula:

$$\text{Risk Score} = \frac{1}{1 + e^{-(\beta_0 + \sum_i \beta_i x_i)}} \times 100$$

- Low Risk (0–40)
- Moderate (40–70)
- High Risk (70–100)

# Implementation Concept

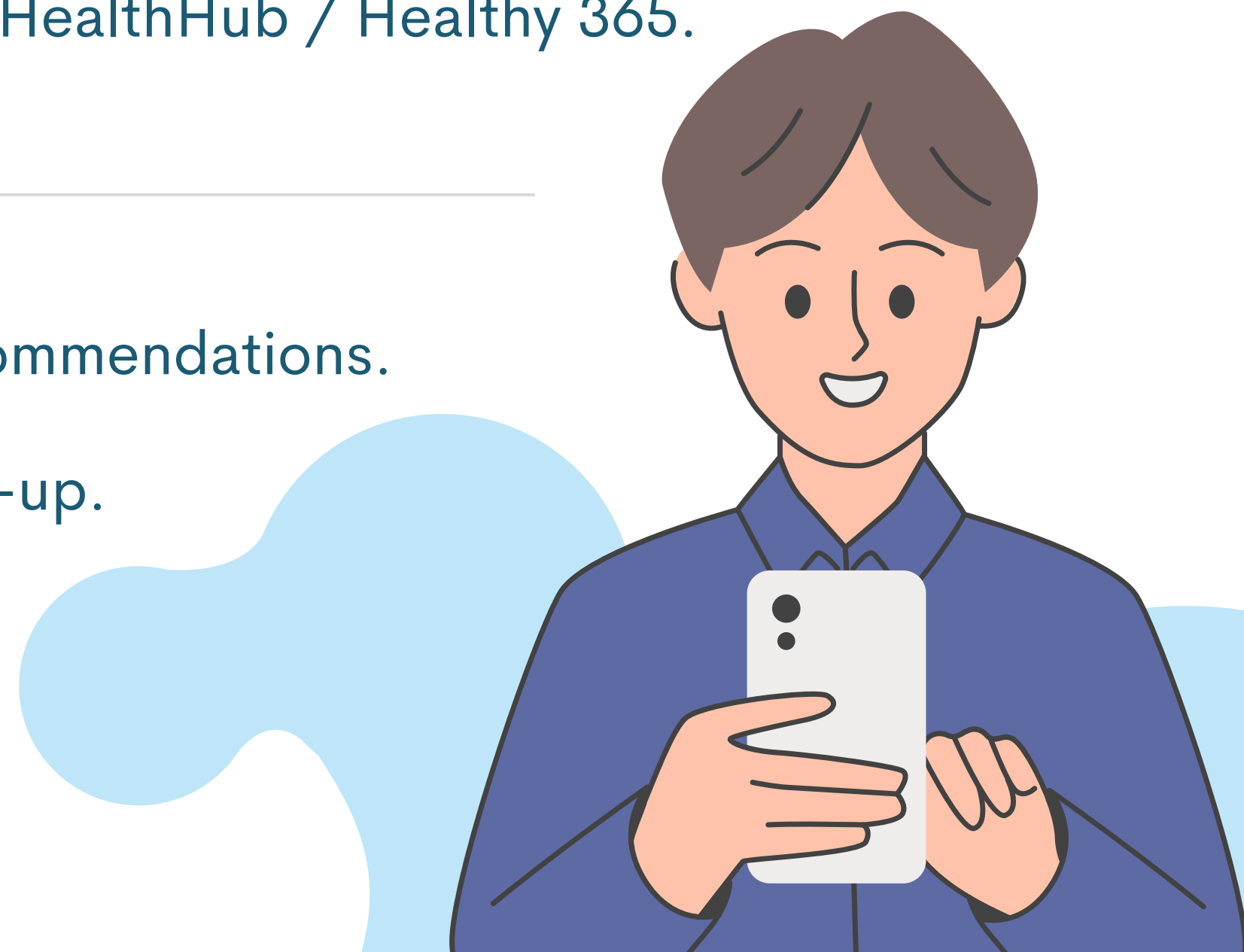


Embedded under LifeSG's Health & Wellness section.



Users can input or sync data from HealthHub / Healthy 365.

- 
- Instant feedback: personalised risk score + recommendations.
  - Shared (securely) with family doctors for follow-up.



# Implementation Concept

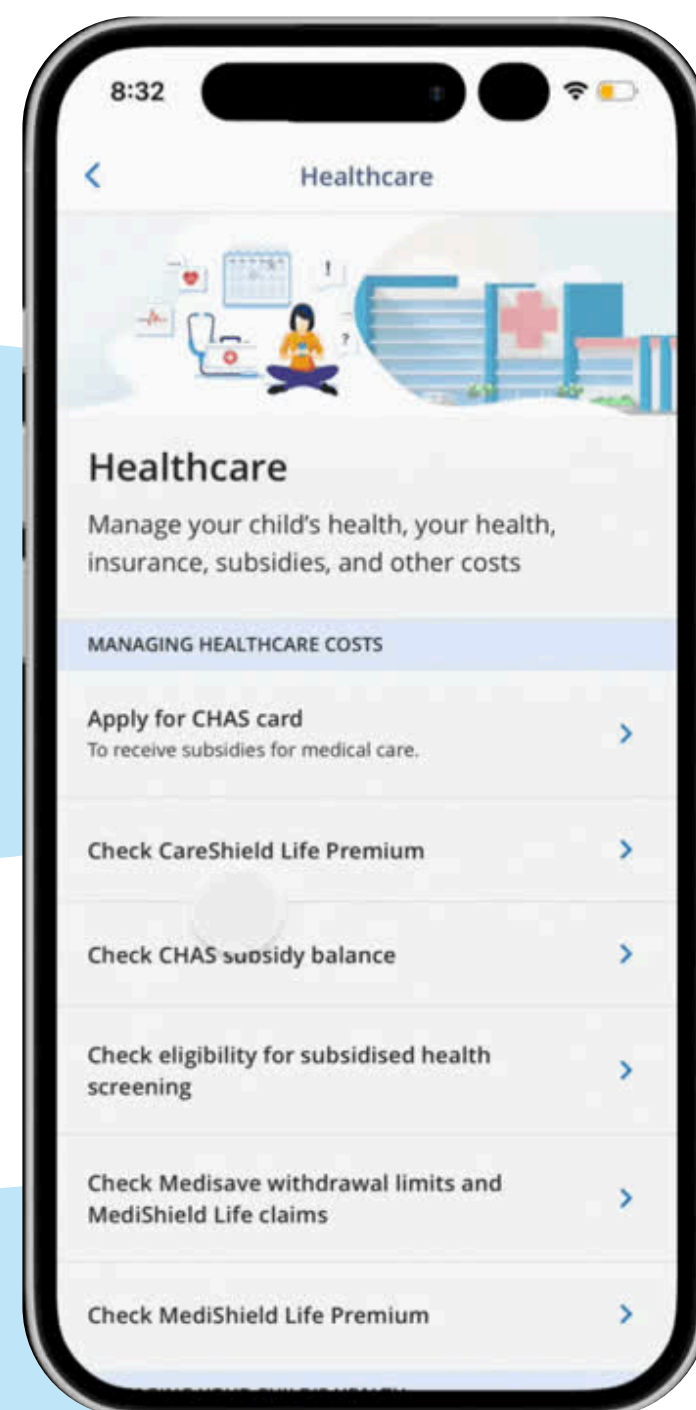
- Encourages early detection and health ownership.
- Provides aggregated, anonymised data to identify high-risk areas.
- Enables targeted national campaigns and resource planning.
- Supports Healthier SG's shift from reactive care to proactive prevention.



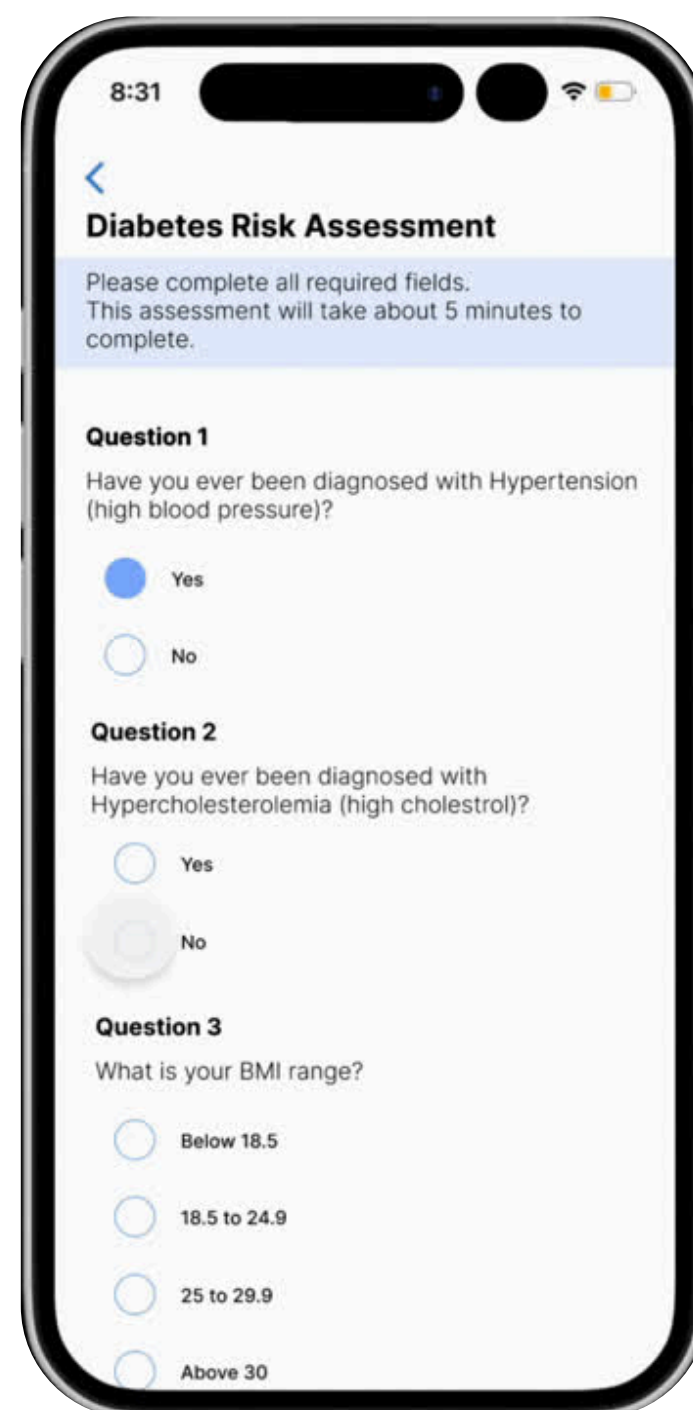
# Implementation Output

## USER FLOW DEMONSTRATION

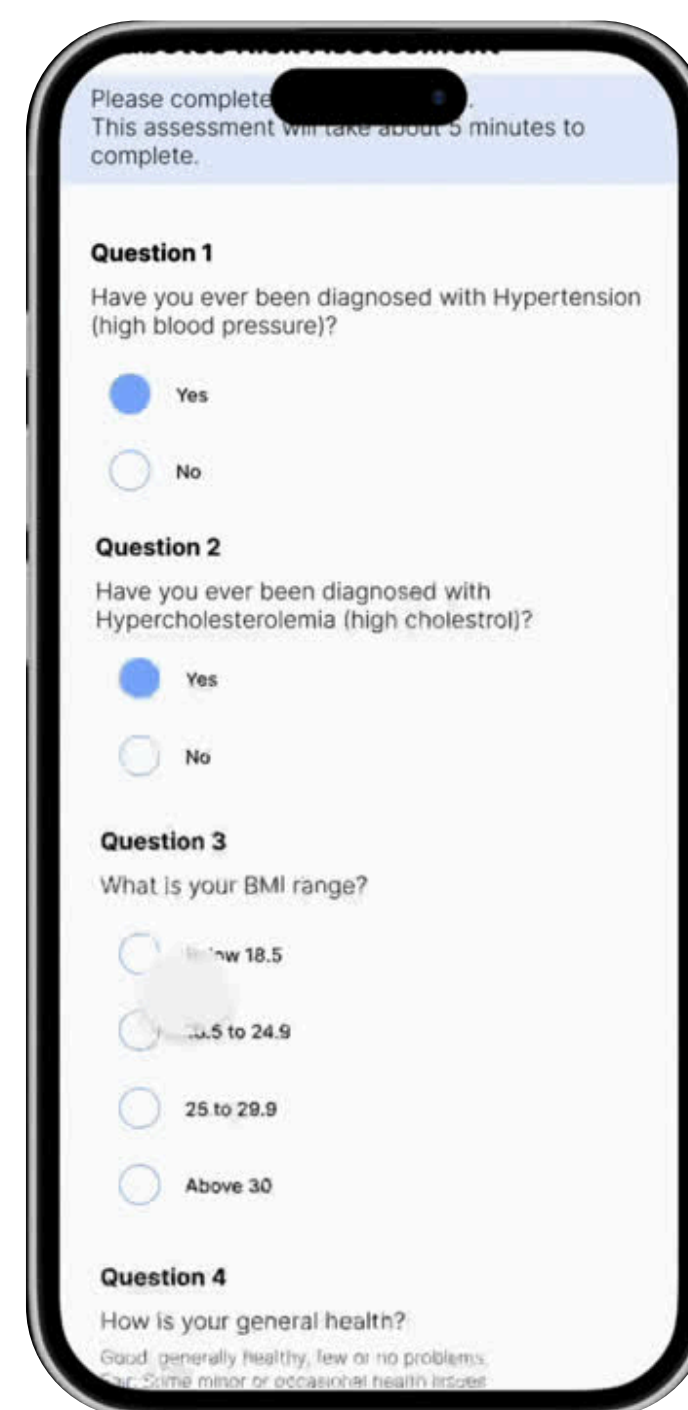
### LOW RISK



### MEDIUM RISK



### HIGH RISK



Users will access the Diabetes Risk Assessment through the "Healthcare" section on LifeSG.

The risk assessment is in the form of a short questionnaire. Once completed, users will be grouped into "Low", "Medium" and "High" diabetes risk based on their personal risk score.



# Implementation Output

## RISK ASSESSMENT - QUESTIONNAIRE

8:31

<

Diabetes Risk Assessment

Please complete all required fields.

This assessment will take about 5 minutes to complete.

Question 1

Have you ever been diagnosed with Hypertension (high blood pressure)?

☐ Yes

☐ No

Question 2

Have you ever been diagnosed with Hypercholesterolemia (high cholesterol)?

☐ Yes

☐ No

Question 3

What is your BMI range?

☐ Below 18.5

☐ 18.5 to 24.9

☐ 25 to 29.9

☐ Above 30

Question 4

How is your general health?

Good: generally healthy, few or no problems

Fair: Some minor or occasional health issues

Poor: Frequent or chronic health problems

☐ Good

☐ Fair

☐ Poor

Question 5

What is your age range?

☐ Below 18 years old

☐ 19 to 30 years old

☐ 31 to 50 years old

☐ Above 50 years old

Submit Form

Question 1

Have you ever been diagnosed with Hypertension (high blood pressure)?

- ☐ Yes
- ☐ No

Question 2

Have you ever been diagnosed with Hypercholesterolemia (high cholesterol)?

- ☐ Yes
- ☐ No

Question 3

What is your BMI range?

- ☐ Below 18.5
- ☐ 18.5 to 24.9
- ☐ 25 to 29.9
- ☐ Above 30

Question 4

How is your general health?

Good: generally healthy, few or no problems  
Fair: Some minor or occasional health issues  
Poor: Frequent or chronic health problems

- ☐ Good
- ☐ Fair
- ☐ Poor

Question 5

What is your age range?

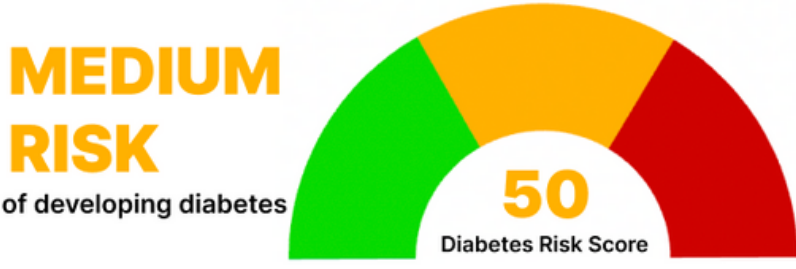
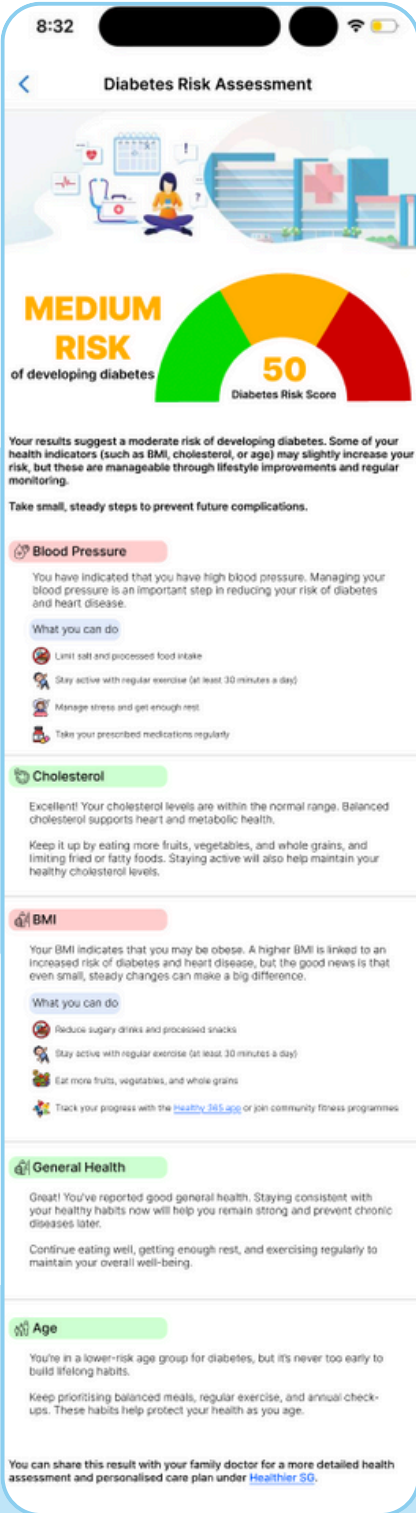
- ☐ Below 18 years old
- ☐ 19 to 30 years old
- ☐ 31 to 50 years old
- ☐ Above 50 years old

Submit Form

- **User-friendly UI/UX:** 5 simple MCQ questions to select from, with additional clarification for questions that need them
- **Data handling:** All responses are securely stored and synced with HealthHub/Healthy 365 upon user consent

# Implementation Output

## RISK ASSESSMENT - RESULTS



Your results suggest a moderate risk of developing diabetes. Some of your health indicators (such as BMI, cholesterol, or age) may slightly increase your risk, but these are manageable through lifestyle improvements and regular monitoring.

Take small, steady steps to prevent future complications.

**Blood Pressure**

You have indicated that you have high blood pressure. Managing your blood pressure is an important step in reducing your risk of diabetes and heart disease.

What you can do

- Limit salt and processed food intake
- Stay active with regular exercise (at least 30 minutes a day)
- Manage stress and get enough rest
- Take your prescribed medications regularly

**BMI**

Your BMI indicates that you may be obese. A higher BMI is linked to an increased risk of diabetes and heart disease, but the good news is that even small, steady changes can make a big difference.

What you can do

- Reduce sugary drinks and processed snacks
- Stay active with regular exercise (at least 30 minutes a day)
- Eat more fruits, vegetables, and whole grains
- Track your progress with the [Healthy 365 app](#) or join community fitness programmes

**Cholesterol**

Excellent! Your cholesterol levels are within the normal range. Balanced cholesterol supports heart and metabolic health.

Keep it up by eating more fruits, vegetables, and whole grains, and limiting fried or fatty foods. Staying active will also help maintain your healthy cholesterol levels.

**General Health**

Great! You've reported good general health. Staying consistent with your healthy habits now will help you remain strong and prevent chronic diseases later.

Continue eating well, getting enough rest, and exercising regularly to maintain your overall well-being.

**Age**

You're in a lower-risk age group for diabetes, but it's never too early to build lifelong habits.

Keep prioritising balanced meals, regular exercise, and annual check-ups. These habits help protect your health as you age.

Risk score is displayed using a traffic light system, and each result is paired with targeted, Healthier SG-aligned advice.



# Conclusion

## KEY TAKEAWAYS

**Predictive Diabetes Risk Model:** empowers Singaporeans to understand and manage their health proactively and independently

**Integration with LifeSG:** allowing preventive healthcare into a personal, accessible, and data-driven experience

**Actionable Insights:** individuals can be more independent with their health as data insights are translated into simple and actionable insights

## HEALTHIER SG ALIGNMENT

The early detection and preventive action encourage engagement with national health initiatives and enables data-driven policy-making through population-level insights.

## BROADER SOCIAL IMPACT

Shifts Singapore's healthcare focus from reactive treatment to proactive prevention, and contributes to building a healthier and more sustainable nation.