

# INFO-B512/556 BIOLOGICAL DATABASE MANAGEMENT

Instructor -

Amir Manzour, Assistant Professor, BioHealth Informatics

*“Deaths in heart failure and its contributing factors”*

## GROUP 13

Manusha Gadde

Sai Shivani Chiraag

Snigdha Mekarthy


Sudhanshu Dwivedi



**IUPUI**

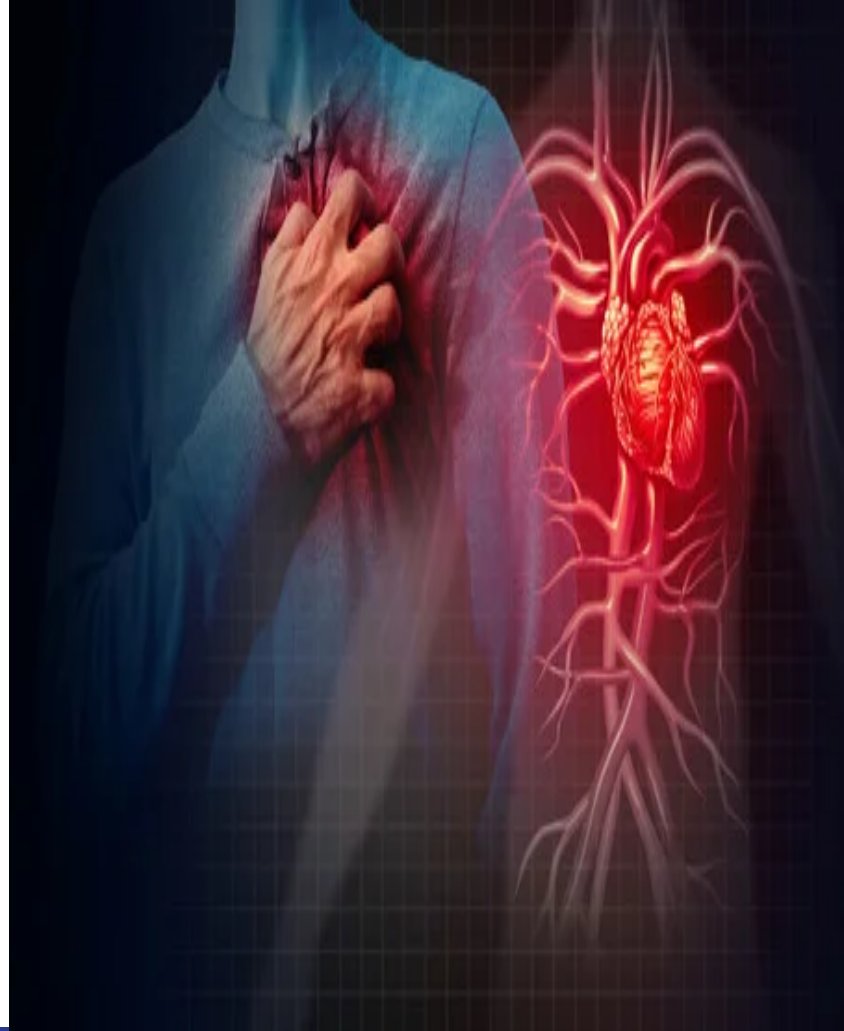
SCHOOL OF INFORMATICS AND COMPUTING

# Contents

- ❖ Introduction
  - ❖ Abstract
  - ❖ Methodology
  - ❖ ERD Diagram
  - ❖ Data Analysis
    - Table Creation
    - Data Cleaning
    - Research Questions
  - ❖ Data Visualization
  - ❖ Results
  - ❖ Discussion - Contributions
  - ❖ Conclusion
  - ❖ References
- 

# INTRODUCTION

- Heart failure has become one of the leading causes of human mortality.
- It's not only the disease but associated factors that lead to the mortality of humankind/people.
- Many people are unaware of the fact that risks can be reduced with early detection and management of established diseases.
- To aid healthcare practitioners with early diagnosis, various data analytics techniques are deployed for the accurate decision making and optimum therapy to address the cardiac risk.
- The creation of this database may increase understanding of opportunities to help fill knowledge gaps in cardiac care concerning cardiac incidents caused by numerous risk indicators.
- Additionally, the database created may be utilized to get real - time data inputs, which can subsequently be used to look at epidemiology and health implications.



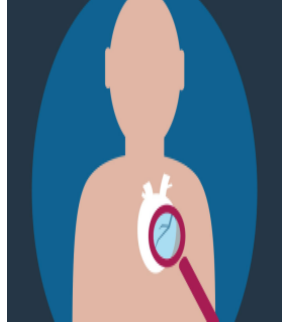
# ABSTRACT

**Motivation** - An accurate cardiac condition assessment can rescue a person's life, but an incorrect prognosis might be dangerous. As a result, we intended to establish a database that may help people become more aware of the various risk factors that contribute to cardiovascular events.

**Goal** - The purpose of this study is to explore the relation between risk factors and early-onset cardiac issues, as well as the relative contribution of risk variables to the development of heart failure based on age groups. In addition, we plan to use SQL queries to address research questions based on the dataset, summarize the final results, and identify the most influencing features among demographics, patient vitals, and environmental factors in causing cardiovascular problems.

**Proposed Methods** - Kaggle , Microsoft Excel , phpMyAdmin,, MySQL workbench and Tableau.

**Expected Results** - We plan to use the database to interpret the most influential factors in causing cardiovascular problems, such as demographic characteristics, patient vitals, and environmental elements, in order to aid in the development of initiatives which can be utilised to assess the risk factors connected with early detection and management of established disease.



## METHODOLOGY

**Data collection :** We have extracted the dataset from Kaggle. we decoded the variables and saved our data in CSV file format. Later, we created 3 separate .csv files based on the relevant attributes.

### **Data cleaning and import :**

The cleaning and removal of outliers were done in PHPMysqlAdmin and finally we had 30,000 observations for the data analysis. The data importing was done through import functions.

### **Data modeling**

We have established connections by adding constraints to the tables. Composition of Entity-relationship diagram (ERD), logical modeling design, relational schema of the given entities and their attributes for better understanding of their relationship and structure of the dataset.

### **Data Visualization :**

Establishing a connectivity between SQL server and Tableau in order to see the research outcomes created after running SQL queries to and also to identify the main affecting parameters in really causing cardiac problems.



# ERD (Entity-Relationship Diagram)

The 3 different tables that we created were named as follows - patientinformation , demographics and outcomes.

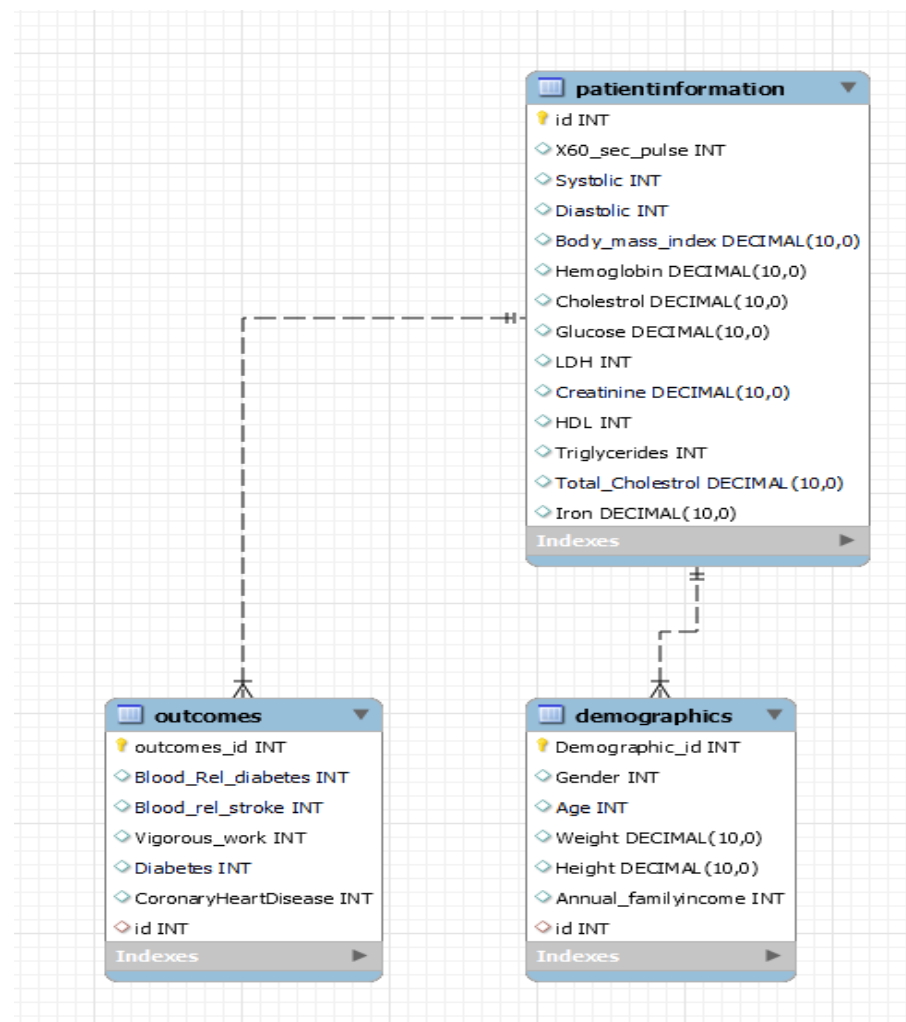
The first table that we created was – “patientinformation” with the “id” attribute as the primary key.

The 2nd table and 3rd tables that were designed are “Demographics” with “demographic\_id” as a primary key and “outcomes” with “outcome\_id” as a primary key respectively .

The “id” column from the patientinformation table serves as a foreign key in these 2 tables.

One patient can have many outcomes which indicate one to many Relationship between patientinformation and outcomes.

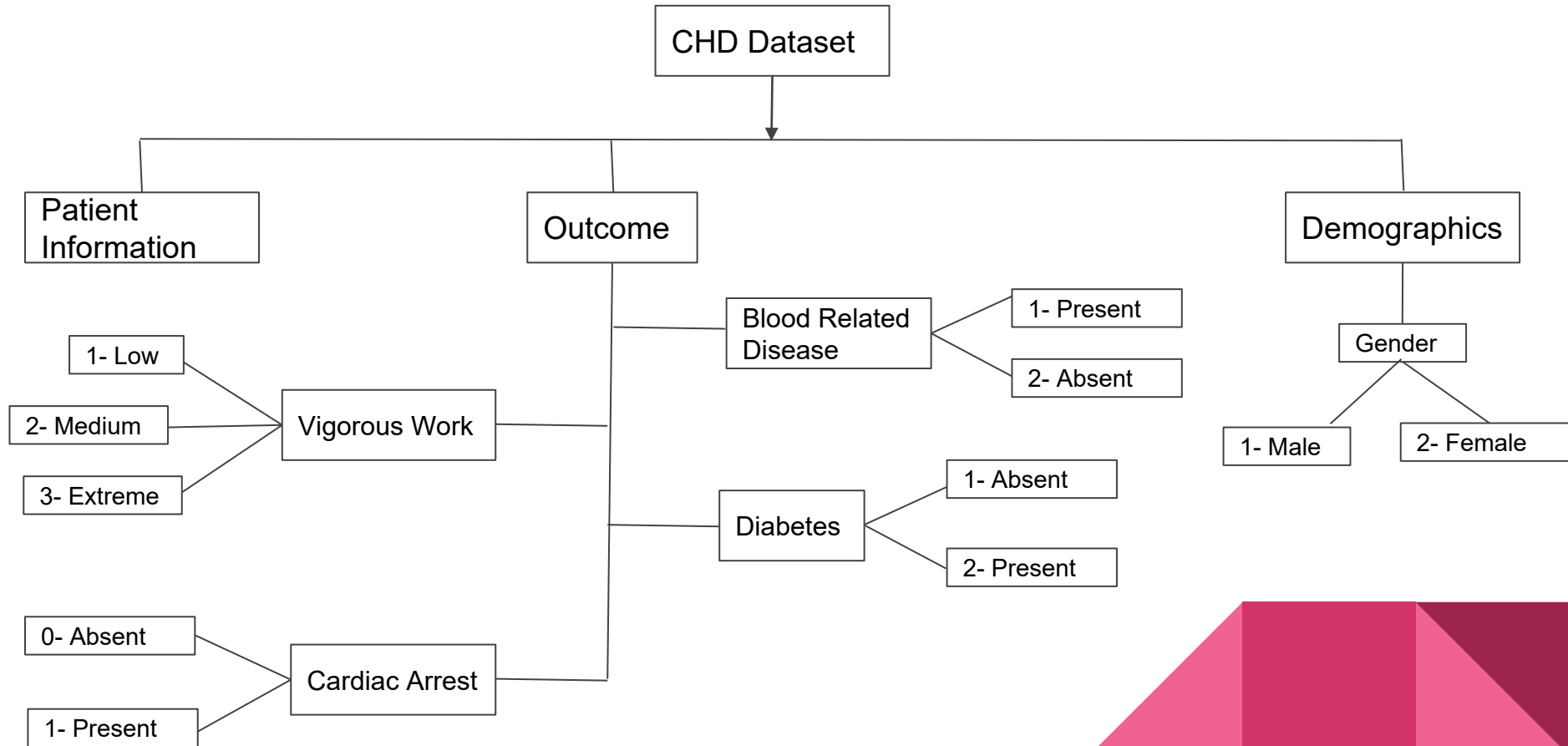
Every id will have individual demographic details that indicate their is a one to one relationship between patientinformation table and demographics table.



# Data Analysis



# Data Attributes





## Table Creation

- We divided the large dataset in three different tables which are Demographics, patient information, and the final outcome.
- The structure were created with primary and foreign key for all three tables using CREATE TABLE in PHPMyAdmin.
- Every tables has its own primary key. Also demographic and outcome tables is connected with the primary key of the main table which is patient information.
- All three CSVs were imported using import function in PHPMyAdmin.



# Table Creation

Run SQL query/queries on database `sdwivedi_db`: 

```
1 CREATE TABLE patientinformation(  
2 id INT NOT NULL PRIMARY KEY,  
3 X60_sec_pulse INT,  
4 Systolic INT,  
5 Diastolic INT,  
6 Body_mass_index decimal,  
7 Hemoglobin decimal,  
8 Cholestrol decimal,  
9 Glucose decimal,  
10 LDH INT,  
11 Creatinine decimal,  
12 HDL INT,  
13 Triglycerides INT,  
14 Total_Cholestrol decimal,  
15 Iron decimal);
```

```
1 CREATE TABLE demographics (  
2 Demographic_id INT NOT NULL PRIMARY KEY,  
3 Gender INT,  
4 Age INT,  
5 Weight decimal,  
6 Height decimal,  
7 Annual_familyincome INT,  
8 id INT,  
9 CONSTRAINT foreign key(id) REFERENCES patientinformation(id)  
10 );  
11  
12 CREATE TABLE outcomes(  
13 outcomes_id INT NOT NULL PRIMARY KEY,  
14 Blood_Rel_diabetes INT,  
15 Blood_rel_stroke INT,  
16 Vigorous_work INT,  
17 Diabetes INT,  
18 CoronaryHeartDisease INT,  
19 id INT,  
20 CONSTRAINT foreign key(id) REFERENCES patientinformation(id)  
21 );
```

# Data Cleaning

- After importing the dataset we first started with the cleaning of dataset before we proceed to the data analysis.
- To check the null or blank values we used SELECT statement

```
SELECT * FROM demographics  
WHERE GENDER = ' ' |
```

```
SELECT * FROM Height  
WHERE GENDER = 0 |
```

- Total number of observations we have for the final data analysis is 30,000.
- Instead of doing data imputation we dropped the observation because those were outliers which could also affect the final result of our analysis.



# Research Questions

## Research Question 1:

Which age group is highly affected with the Cardiac Heart Failure?

✓ Showing rows 0 - 1 (2 total, Query took 0.1715 seconds.)

```
SELECT SUM(CASE WHEN d.Age <=40 THEN 1 ELSE 0 END) AS '20-40', SUM(CASE WHEN d.Age BETWEEN 41 AND 60 THEN 1 ELSE 0 END) AS '41-60', SUM(CASE WHEN d.Age >=61 THEN 1 ELSE 0 END) AS '61-85'
FROM demographics d INNER JOIN outcomes o ON d.id=o.id GROUP BY o.CoronaryHeartDisease
```

☐ Profiling [\[Edit inline\]](#) [\[ Edit \]](#) [\[ Explain SQL \]](#) [\[ Create PHP code \]](#) [\[ Refresh \]](#)

☐ Show all

Number of rows: 

25

 ▾

Filter rows: 

Search this table

+ Options

20-40	41-60	61-85
11252	9362	8159
23	254	949

According to CDC 41-60 are highly vulnerable however the result shows 61-85 are most affected age group

## Research Question 2:

How is the annual family income of a person influencing the occurrence of heart failure?

✓ Showing rows 0 - 1 (2 total, Query took 0.1434 seconds.)

```
SELECT SUM(CASE WHEN d.Annual_familyincome <=5 THEN 1 ELSE 0 END) AS 'Less than 50k', SUM(CASE WHEN d.Annual_familyincome BETWEEN 6 AND 10 THEN 1 ELSE 0 END) AS '51K-100K', SUM(CASE WHEN d.Annual_familyincome >=11 THEN 1 ELSE 0 END) AS '100K-150K' FROM demographics d INNER JOIN outcomes o ON d.id=o.id GROUP BY o.CoronaryHeartDisease
```

☐ Profiling [Edit inline] [ Edit ] [ Explain SQL ] [ Create PHP code ] [ Refresh ]

☐ Show all | Number of rows: 25 ▾ Filter rows:


+ Options

Less than 50k	51K-100K	100K-150K
10570	12113	6090
534	495	197

The result state that people whose income is less than \$50k have higher chances of developing Cardiac Arrest


### Research Question 3:

What is the gender distribution of CHD?

 Showing rows 0 - 1 (2 total, Query took 0.1288 seconds.)

```
SELECT d.Gender, COUNT(*) AS patient_having_CHD FROM demographics d INNER
JOIN outcomes o ON d.id=o.id WHERE o.CoronaryHeartDisease = 1 GROUP BY
d.Gender
```

☐ Profiling [\[Edit inline\]](#) [\[ Edit \]](#) [\[ Explain SQL \]](#) [\[ Create PHP code \]](#) [\[ Refresh \]](#)

☐ Show all | Number of rows:  

+ Options

Gender	patient_having_CHD
1	841
2	385

- 1 - Male
- 2 - Female

Male have higher chances  
of developing heart arrest  
than female



## Research Question 4:

### Does Vigorous work affect CHD?

✓ Showing rows 0 - 2 (3 total, Query took 0.0142 seconds.)

```
SELECT Vigorous_work, COUNT(*) AS people_having_CHD FROM outcomes WHERE  
CoronaryHeartDisease = 1 GROUP BY Vigorous_work
```

☐ Profiling [\[Edit inline\]](#) [\[Edit\]](#) [\[Explain SQL\]](#) [\[Create PHP code\]](#) [\[Refresh\]](#)

☐ Show all | Number of rows: 25 ▾

+ Options

Vigorous_work	people_having_CHD
1	184
2	960
3	82

Here, we got surprising result because according to CDC individuals do more vigorous work have higher chances of developing CHD but according to the data someone who does mediocre work have higher chances

## Research Question 5:

How many people with comorbidities(Blood Related Stroke and Blood Related Diabetes) have CHD?

✓ Showing rows 0 - 1 (2 total, Query took 0.0185 seconds.)

```
SELECT CoronaryHeartDisease, COUNT(*) AS people_having_CHD FROM outcomes WHERE Blood_rel_stroke = 2 AND Blood_Rel_diabetes = 2 GROUP BY CoronaryHeartDisease
```

☐ Profiling [\[Edit inline\]](#) [\[ Edit \]](#) [\[ Explain SQL \]](#) [\[ Create PHP code \]](#) [\[ Refresh \]](#)

☐ Show all | Number of rows: 25  Filter rows:

+ Options

CoronaryHeartDisease	people_having_CHD
0	13447
1	466

Based on this result if an individual have underline or comorbidities such as Blood Related Stroke and Blood Related Diabetes have less chances of developing CHD

































**CREATING VIEW:** We have have created a view using outcome table to show Blood Related Diabetes, Blood Related Stroke, Vigorous work, Diabetes and Coronary Heart Disease.

✓ MySQL returned an empty result set (i.e. zero rows). (Query took 0.0055 seconds.)

```
CREATE VIEW Patient_with_Comorbidities_And_CHD AS SELECT * FROM outcomes WHERE Blood_Rel_diabetes = 2 AND Blood_rel_stroke = 2 AND CoronaryHeartDisease = 1
```

[\[Edit inline\]](#) [\[ Edit \]](#) [\[ Create PHP code \]](#)

		outcomes_id	Blood_Rel_diabetes	Blood_rel_stroke	Vigorous_work	Diabetes	CoronaryHeartDisease	id
dit	 Copy  Delete	246	2	2	1	2	1	239
dit	 Copy  Delete	264	2	2	2	2	1	257
dit	 Copy  Delete	649	2	2	2	2	1	642
dit	 Copy  Delete	886	2	2	1	2	1	879
dit	 Copy  Delete	892	2	2	2	2	1	885
dit	 Copy  Delete	1199	2	2	2	2	1	1192
dit	 Copy  Delete	1303	2	2	1	1	1	1296
dit	 Copy  Delete	1837	2	2	2	2	1	1830
dit	 Copy  Delete	1905	2	2	1	2	1	1898
dit	 Copy  Delete	2029	2	2	3	2	1	2022
dit	 Copy  Delete	2365	2	2	2	2	1	2358
dit	 Copy  Delete	2555	2	2	3	2	1	2548
dit	 Copy  Delete	2776	2	2	2	2	1	2769
dit	 Copy  Delete	2795	2	2	2	2	1	2788
	 Console  Delete	2805	2	2	2	2	1	2798

# Data Visualizations

# Establishing a connectivity between SQL server and Tableau for graphical representation of the research outcomes



outcomes — Custom S...

→ rows ⚙️ ▼

How do relationships differ from joins? [Learn more](#)

outcomes	Operator	Custom SQL Query1
# Coronary Heart I ▼	= ▼	# CoronaryHeartD ▼

⊕ Add more fields

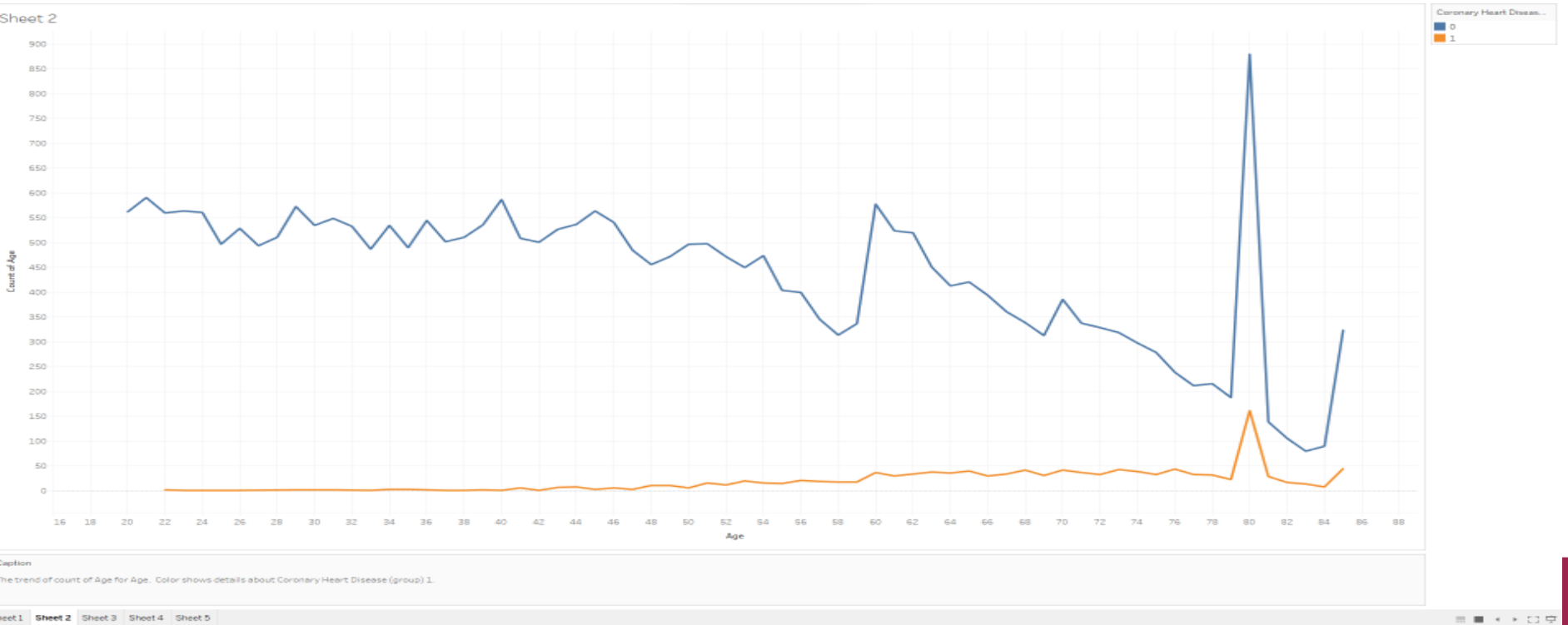
> Performance Options

# Custom SQL Query1 CoronaryHeartDisease (...)	# Custom SQL Query1 people_having_CHD (Cu...
---	---

Update Now

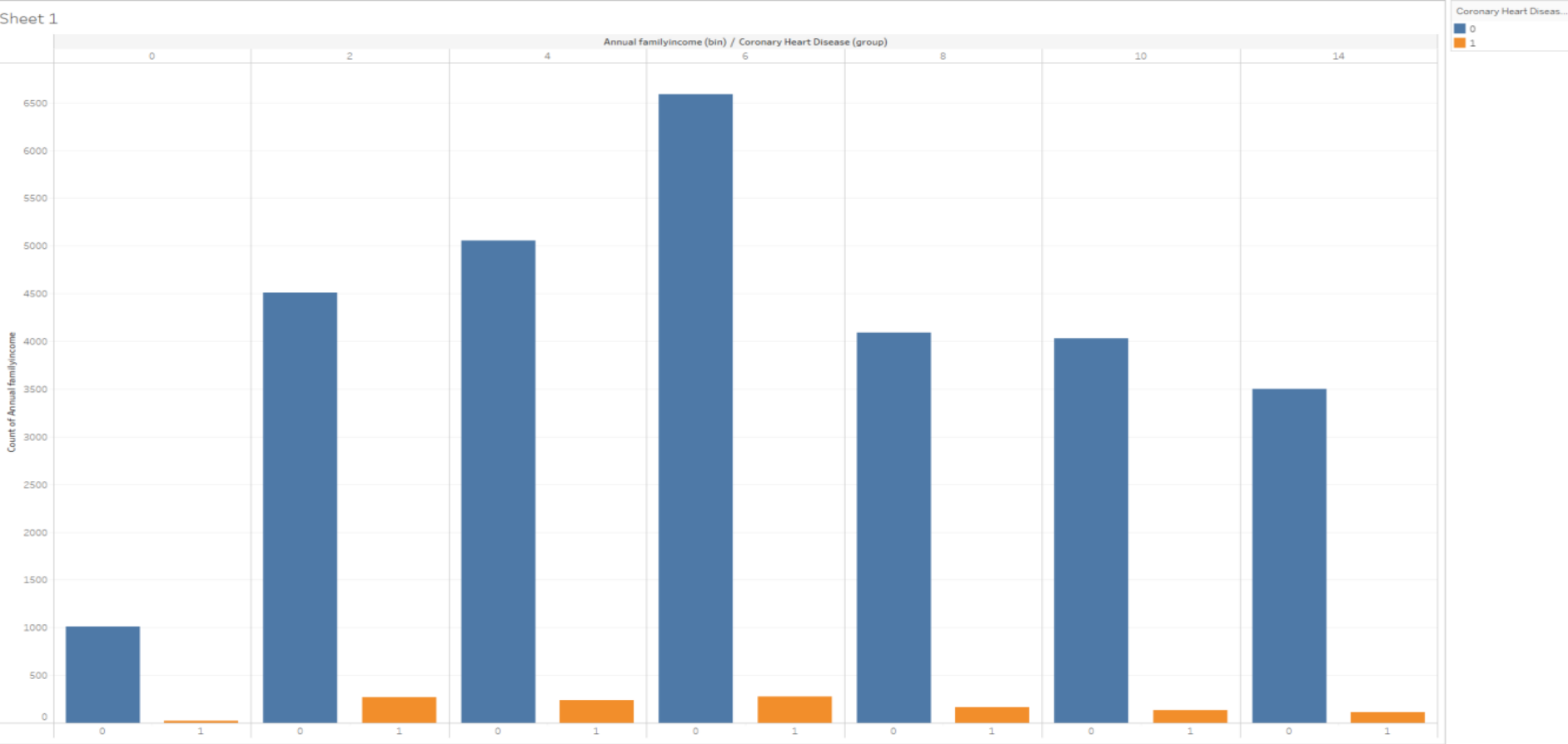
Update Automatically

# Age group and Cardiac Heart Failure relationship



# Annual family income of a person influencing the occurrence of heart failure

Sheet 1

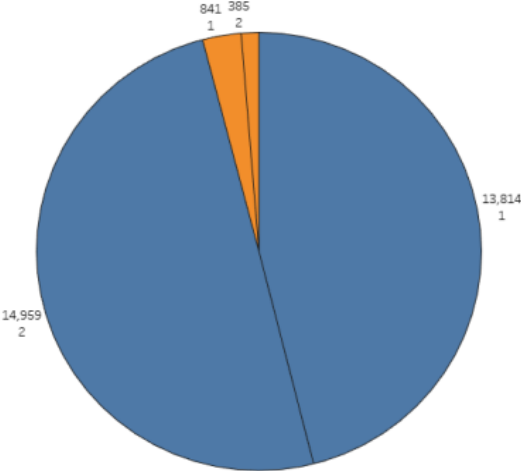


# Gender distribution of CHD

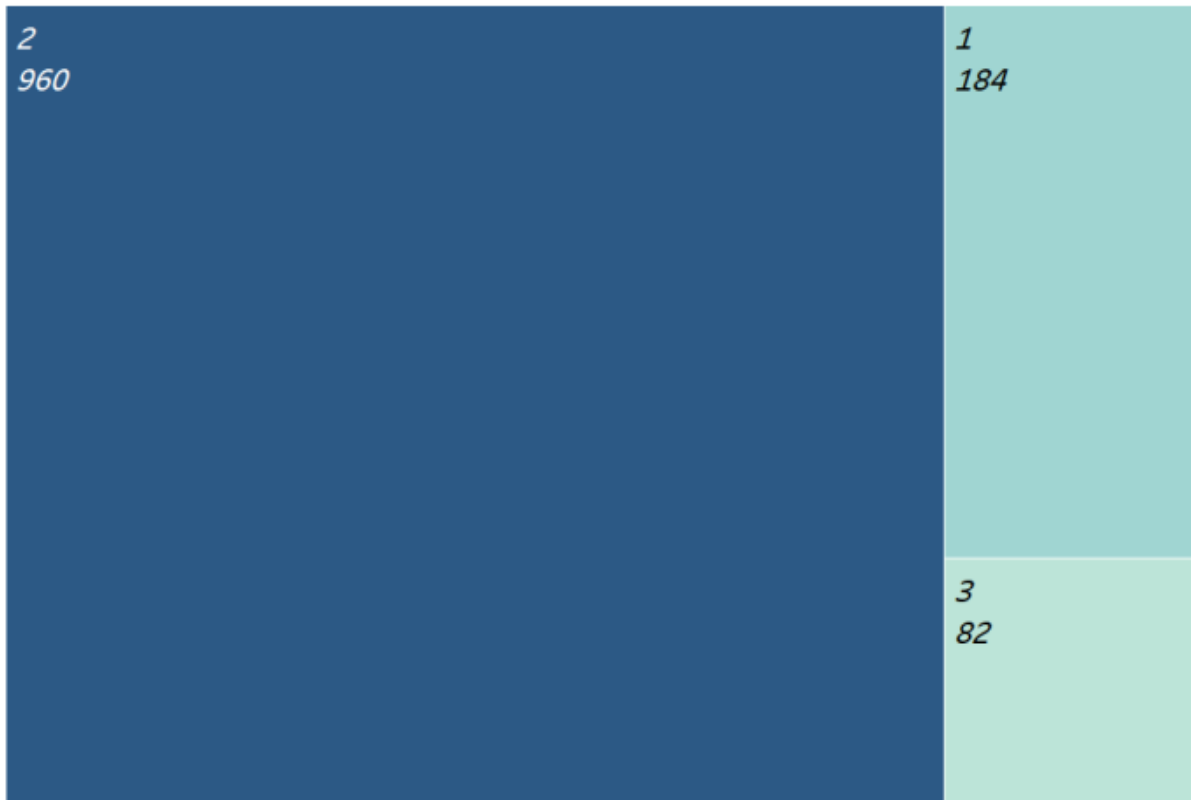
Sheet 3

CNT(Coronary Heart Di...  
29,999

Coronary Heart Diseas...  
0  
1



## Affect of vigorous work on incidence of Coronary Heart Disease

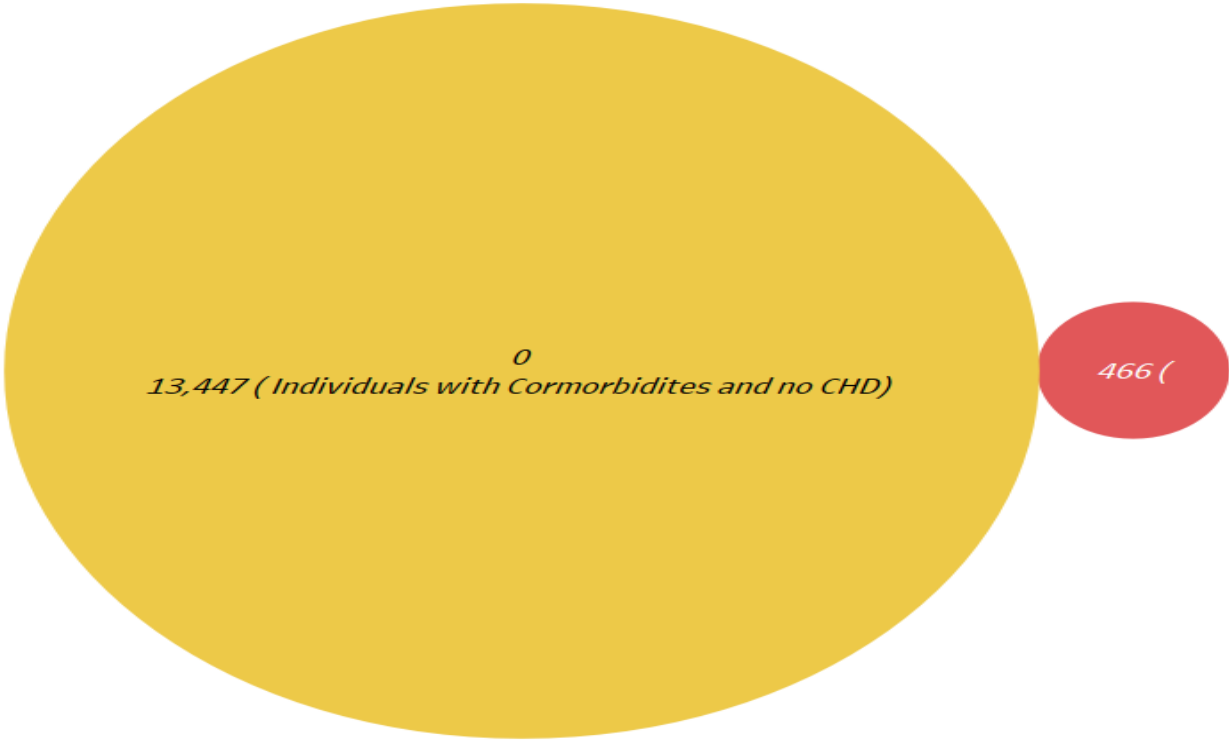


*Number of individuals prone to CHD with associated comorbidities.*

Incidence of CHD in people with Comorbidities such as Blood related ..

13,447 ( Individuals with Cormorbidites and no CHD)

466 ( Individuals with Cormorbidites and CHD)






# RESULTS



## **Contributions:**

- Based on the research we specifically looked for the contribution factors in occurrence of CHD.
- The demographic characteristics such as age, BMI, and gender distribution have a greater influence on the mortality rate caused by CHD incidence.
- This study also reveals the vital signs that increase the mortality rate amongst the CHD patient and this clinical dataset will aid various hospitals and public health departments in the surveillance of the disease and the preventive measure.

## **Prevention:**


- Stop smoking - It's a major factor in the arterial damage that can cause heart failure. Also steer clear of secondhand smoke.
  - Eat in heart-healthy ways - Foods that are low in saturated fat, trans fat, sugars, and salt are beneficial to the health.
  - Lose pounds if you're overweight - Along with diet, being physically active helps achieve this goal and is also great for your heart.
  - A close follow up is required if there is any other form of heart disease or any other condition.
- 

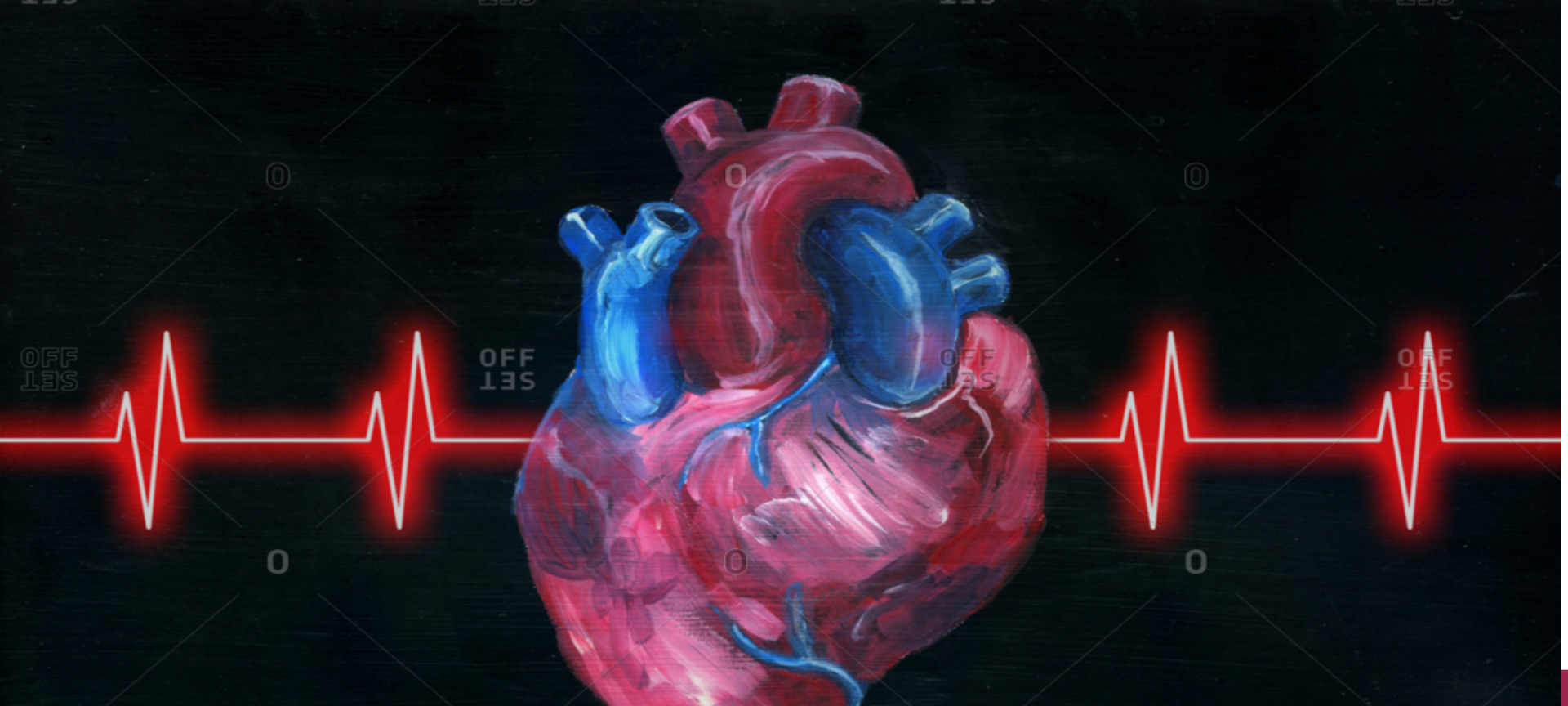
## Results and Conclusion:

- Genetic or lifestyle factors cause plaque to build up in your arteries as you age. Also, changes in the small blood vessels of the heart as you age raise the risk for coronary heart disease.
- Air pollution in the environment can put you at higher risk of coronary heart disease. The increase in risk may be higher in older adults, women, and people who have diabetes or obesity.
- Coronary heart disease affects both men and women. Obstructive coronary artery disease is seen more common in men.
- Based on this study we were able to see the relationship between various factors and outcome of coronary heart disease on the basis of query and data visualization we did in SQL and Tableau.
- To conclude, heart failure worsens over time, so you need to be familiar with changes in your body. Keep track of blood pressure, weight and other vital signs as your doctor advises. Get lab work done as recommended, as it gives key clues to your heart health and medication needs.
- People experiencing depression, anxiety, stress over a period of time may experience increase in the risk of an adverse cardiac event.



## References:

- Andersen, L. W., Holmberg, M. J., Berg, K. M., Donnino, M. W., & Granfeldt, A. (2019). In-Hospital Cardiac Arrest: A Review. *JAMA*, 321(12), 1200–1210. <https://doi.org/10.1001/jama.2019.1696>
- Bui, A. L., Horwich, T. B., & Fonarow, G. C. (2011). Epidemiology and risk profile of heart failure. *Nature reviews. Cardiology*, 8(1), 30–41. <https://doi.org/10.1038/nrcardio.2010.165>
- Dunlay, S. M., Weston, S. A., Jacobsen, S. J., & Roger, V. L. (2009). Risk factors for heart failure: a population-based case-control study. *The American journal of medicine*, 122(11), 1023–1028. <https://doi.org/10.1016/j.amjmed.2009.04.022>
- Gardner, A. (n.d.). Heart failure and joint conditions can occur with it. <https://www.webmd.com/heartdisease/heart-failure/heartfailure-common-conditions>
- Lawson, C. A., Zaccardi, F., Squire, I., Okhai, H., Davies, M., Huang, W., Mamas, M., Lam, C. S. P., Khunti, K., & Kadam, U. T. (2020). Risk factors for heart failure. *Circulation: Heart Failure*, 13(2). <https://doi.org/10.1161/circheartfailure.119.006472>
- 



**THANK YOU**