Proposed Database for the Analysis of Various Heart Disease treatments and their Effectiveness

Report by Team #3

Javier Celetstino
Mitchell Dahmer
Fabian Mcghee
Marcio Amaral

Table of Contents

Abstract

The motivation behind this project is the current state of cardiovascular health in the United States. "Heart disease is the leading cause of death for men, women, and people of most racial and ethnic groups in the United States" ("Heart Disease Facts"). According to the CDC, One person dies every 33 seconds in the United States from cardiovascular disease, and 1 in 5 deaths since 2021 were caused by heart disease. The economic impact on the nation is estimated to be $239.9 billion annually from 2018 to 2019. Currently, many organizations are developing their database systems. They are paying close attention to what information would help them create datasets worth studying. We believe that the more widespread these data sets are, the better research studies can be made. In order to contribute to a solution to this problem, we created a database that tracks patients' basic information, cardiovascular disease history, and treatments. The data for this database can be compiled from existing datasets that are free for the public to use or by partnering with medical providers to capture their patients' data. We aim to provide a tool researchers can use to find macro trends in treatments and their effectiveness.

Introduction/Background

Cardiovascular health is the most significant medical challenge America has ever faced. Since the year 2021, one in five deaths in America has been linked to cardiovascular diseases. It is estimated that every 33 seconds, an American dies from heart-related complications. The toll this epidemic places on our society is enormous. Between 2018 and 2019, the annual cost of dealing with cardiovascular diseases was more than $319 billion ("Heart Disease Facts"). Presented with this enormous problem, the question of how to find a solution is of the utmost importance. As the field of data science matures and becomes prominent in our lives, data scientists have begun to contribute to solving this epidemic. Efforts have been made to better predict cardiovascular diseases. Using key indicators, scientists have found a correlation between different attributes and heart complications. The goal of this type of research is to diagnose health diseases earlier. This work is significant as an early diagnosis could lead to more options being available to correct or control cardiovascular diseases among patients (Nadakinamani et al.). The datasets used by studies like the one previously mentioned are significant and perhaps the biggest hurdle for researchers to overcome. Adequate datasets must be created from data collected by researchers or adapted from existing third-party research. The quality of that data will significantly affect the effectiveness of any machine learning algorithm. The focus of our project is to develop a database that can be used to collect critical indicators of patients with an emphasis on the type of illness, treatment, and outcome of treatment. If a database like this is widely distributed for contribution and analysis, researchers will have another dataset that might help them make more insightful observations. This report includes the methods we use to structure our database along with how we implemented the addition, update, and deletion of records.

For this database we prioritized easily understood and well organized data. It was important that we had a schema that allowed for both quick access to well categorized data but also that the data be easily appended to so that long term historical data could be accessed. We also focused heavily on the practical and realistic elements of the database. The realities of storing medical data can often be different from other, less personal data. Knowing the value of someone's blood pressure reading may be useful, but a blood pressure reading can be inaccurate as soon as 1 day later. The main utility of medical data comes from long term study and care. If we want to be able to build

Literature review

During our initial review of filing systems and databases for storing cardiovascular medical information, we broke our goals for research into 3 different topics to better suit our needs for this project. First, what medical indicators could we use to get an accurate understanding of someone's heart health? Second, what were the primary diseases that we could expect regular people to be encountering and needing treatment for? Third, what treatments are commonly used for treating these diseases? It was important that we gained a proper understanding of the broad realities of cardiovascular treatment so that we can make a database and reports that fit the potential needs of researchers correctly.

Our targets for research were large, overreaching studies about cardiovascular treatment and causes rather than individually related studies. We felt that a broader approach to understanding the issue would help us tailor a database to fit the situation we were designing for. Additionally, very little scholarly work could be found about healthcare databases specifically so finding the conditions relating to the disease's treatment and constructing a model with those in mind allowed us to tailor the database to its intended purpose.

During our initial planning phase for the project, "Risk factors of deaths related to cardiovascular diseases in World Health Organization (WHO) member countries" by Bayram Şahin and Gülnur İlgün helped set the boundaries of the area our report would explore. This report written in late 2020 takes a look at the general state of cardiovascular care in the majority of WHO member countries and breaks down the methodology used to measure and record this information. After looking over their choices we were able to create a loose framework for how we would handle recording the illnesses and treatments. The paper also discusses the criteria used by researchers to measure the effectiveness of treatments on patients which allowed us to choose what attributes we would prioritize recording for our patients ongoing records.

As we developed the contents of and use cases for the database we found great inspiration from "Modernizing the World Health Organization List of Essential Medicines for Preventing and Controlling Cardiovascular Diseases" by Kishore, S, Blank, E, Heller, D. et al. This paper proposes an updated list for the WHO's current list of medications considered to be essential for treat cardiovascular disease. There are several of these lists for various specialties that give smaller clinics with less access to cutting edge medical supplies and knowledge a baseline standard to operate with. This allows for the recommendations presented in this paper to act as a good starting point for the kinds of diseases and treatments we should be expecting to have our database record information about. By choosing a broad selection from this report we developed a list of diseases and treatments that we shape the database around to give it medical accuracy without overshadowing the database design aspect that is the focus of this paper. For treatments we elected to have the list include Losartan (antihypertensive), Spironolactone (heart failure), NDACs (Anticoagulants), Bisoprolol (multipurpose), Amiodarone (Antiarrhythmics), Nicotine Replacement Agents (substance dependence), Simvastin (cholesterol lowering), Streptokinase (antithrombotics), Low Cholesterol Diet, Low Carbohydrate Diet, Light cardiovascular exercise (2 hours in a day), Moderate Cardiovascular exercise (1 hour), Intense cardiovascular exercise (1 hour), Coronary Stent and Coronary Angioplasty. We also developed a list of cardiovascular conditions that we would focus on. This list includes Arrhythmia, Valve disease, Coronary artery disease, Heart Failure, Peripheral Artery Disease, Aortic Disease, Cerebrovascular Disease, Deep Vein Thrombosis, Congenital Heart Disease and Pericardial Disease. Developing both of these lists helped us also choose what information about patient's lifestyles and vitals we would expect the database to be storing and checking.

As we began creating temporary data to fill the database we made use of several articles from the Mayo clinic about diabetes, hypertension and cholesterol. The purpose of the data was to allow reports and queries to be tested with real information during their development and to allow any demonstration of the database to be more easily understood. Each Mayo clinic article gave a summarized explanation of the conditions they were about including expected symptoms, treatments, and expected test results. These expected test results were used to develop realistically scaled values for every patient's blood pressure, cholesterol and glucose levels. By following the template of results expected we could make sure any outputs of our applications would be scaled to the expected output.

Problem description and definitions

A. How can we best store patient data to determine the effectiveness of treatment?

To optimize the storage of patient data for assessing treatment effectiveness, we propose a structured relational database model that captures essential patient information, treatment details, and corresponding outcomes. By organizing patient data into distinct tables, such as demographics, medical history, treatments, and results, we can establish clear relationships between different data elements. Utilizing a relational database management system enables efficient data retrieval and analysis, facilitating the identification of patterns and trends in treatment responses. additionally, implementing normalization techniques ensures data consistency and integrity, reducing redundancy and minimizing the risk of errors. By adopting this comprehensive database architecture, researchers can efficiently track patient progress, evaluate the impact of various treatments, and derive valuable insights into cardiovascular disease management.

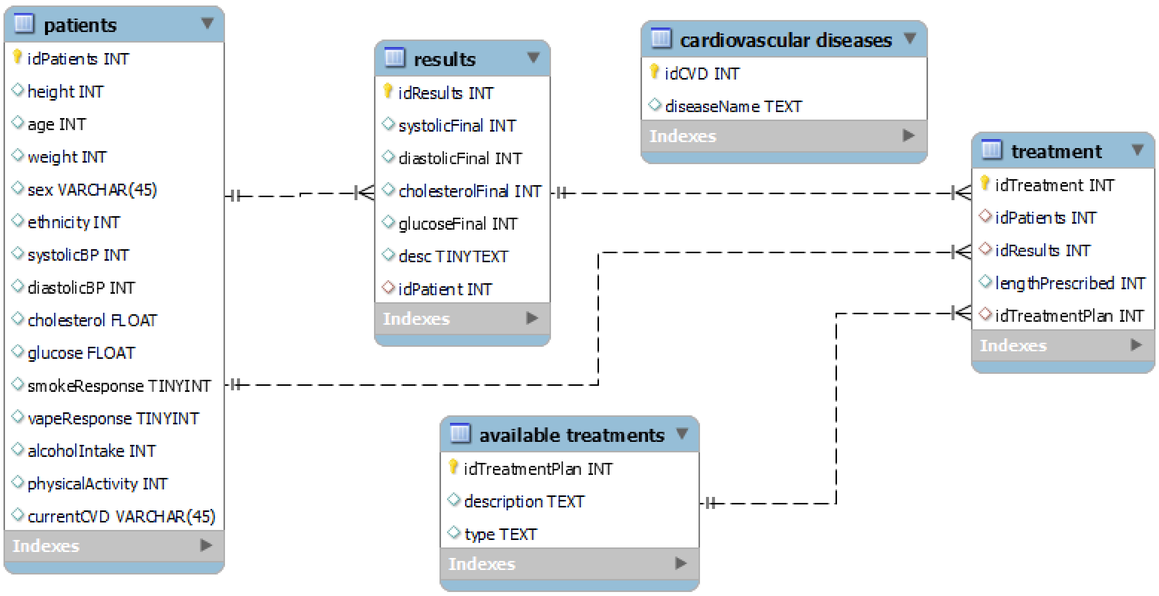
B. What treatments work best for what groups and situations?

Identifying optimal treatments for diverse patient groups and scenarios requires leveraging existing research and patient data. By analyzing treatment outcomes across demographic and clinical conditions, we aim to tailor treatment approaches effectively. Additionally employing machine learning algorithms may enable predictive modeling to anticipate treatment outcomes based on individual patient characteristics. Through a combination of empirical data analysis and clinical expertise, we aim to tailor treatment approaches to the unique needs of the patient, maximizing therapeutic effectiveness and improving cardiovascular health outcomes.

C. What is the best way to present this information to researchers?

The best way of presenting information to researchers requires a user-friendly and informative interface that facilitates data exploration and analysis. initially, our proposed solution involves developing intuitive data visualization tools(charts, graphs, heatmaps, etc) and an interactive dashboard that enables researchers to interact with the database effortlessly. moreover, implementing customizability query functionalities allows the researcher to retrieve specific subsets of data based on their research interests and hypotheses. Additionally, providing descriptive statistics and or summary metrics enhances data interpretation and facilitates comparative analyses across treatment modalities and patient cohorts. Collaborating with domain experts and end-users throughout the interface design process ensures that the platform meets the diverse needs of researchers and promotes efficient knowledge discovery. By prioritizing usability and accessibility, we aim to empower researchers to extract meaningful insights from the database, advancing our understanding of cardiovascular disease management.

Methods (Design at every level)



In order to better visualize the entire design project, it was decided that the database would be originally based off of two paper forms with the design moving through the steps needed to refine the raw input. Since the focus of this project is on creating this database for use by 3rd party researchers, these forms were designed only as a form that would pass between the clinics and the researchers and left more detailed patient and treatment info out for efficiency. The first form was the treatment log. This was a simplified record of treatments prescribed by the doctors at each clinic. This form lists the patient receiving treatment, the treatment prescribed (with a description), the length of the prescribed treatment and the effect on the patient. The second form was a simplified patient file that included the patient's demographic information, the current CVD's the patient is requesting treatment for, information about a patient's habits that may affect their health, and the information about the patient's initial blood pressure, cholesterol, and glucose levels. All of this together allowed for tracking a patient's progress in response to various treatments and finding trends of success that can hopefully be used to help other patients. Originally this was a process that would be done by hand, charted out over a series of readings with all updating and tracking done by hand. Trying to overlay patient data would require you either calculate averages for each case or make hundreds of transparent graphs for finding trends in how patients progressed. Our aim was to streamline the data gathering portion of this process and present the info in specific or unspecific terms as needed.

For the first step in the process both forms were translated directly into two tables, one for the patients and one for the treatment log. The treatment log was turned into a table called ‘treatment’ with the patient’s name (‘name’), the treatment given (‘treatment’), the description (‘desc’), the length of the treatment (‘length’) and a column for percent change in systolic blood pressure(‘systolchange’), diastolic blood pressure(‘distolchange’), ‘cholesterol’ and glucose levels (‘glucose’). Patient name was soon replaced with ‘idPatient’. This was done to create a relation with the table that the patient form was made into. For the patient form table called ‘patients’, columns were first made for patient’s name (‘name’), ‘height’, ‘weight’, ‘sex’ and ‘ethnicity’ in order to contain the demographic information of the patient. Next medical information was put into 5 columns: systolic blood pressure (‘systolicBP’), diastolic blood pressure (‘distolicBP’), ‘cholesterol’, ‘glucose’ and the patient’s current CVD’s (‘currentCVD’). Lastly 4 columns were made to contain information about the patient’s responses to a questionnaire. These questions were about: whether or not the patient smokes or vapes (‘smokeResponse’, ‘vapeResponse’) and how many days out of the week the patient drinks or works out (‘alcoholIntake’, ‘physicalActivity’). The first two accept either a 0 for no and a 1 for yes and the second two except a number between 0 and 7. After this, a ‘idPatient’ column was included to link entries in the treatment table to the patient table. All of these served as the foundation for the changes going forward.

The first major improvement was to decompose the treatment table into three different tables. We took the list of treatment names and descriptions and broke it out into a separate table to avoid potential update anomalies as treatments were updated. This new table contains three columns. One is the ‘idTreatment’ which also now take the place of the other treatment related columns. Every available treatment for prescription is in the ‘availableTreatment’ database with a description (‘desc’), associated ID (‘idResult’) and a user defined category for easier sorting(‘type’). This allows the treatments to still be logged while removing potentially unneeded information into a linked table for reference. The second improvement was a similar decomposition for the results area. The ‘results’ table was made to keep track of all the results entered in for each treatment. The same columns relating to medical information like blood pressure were included in the new ‘results’ table. Three additional columns were added: a column called ‘description’ to include any special notes regarding the results, ‘idResults” and the associated ‘idPatient’. This allowed the results table to be used for calculating a patient’s current CVD related information in response to recent readings from the patient along with quick reference for any recent anomalies in a patient’s health. Both of these changes served to make the treatment log a more functional storage of information for the reports and views that would be created as part of the user interface for this project. Since the treatment log was no longer a document that would be referenced directly by a doctor or researcher it was critical that we do what we can to make the tables act less as charts to be read and more as a web of information about individual people and their continued treatment.

The last change we made out of convenience for the user experience. In the same way the ‘availableTreatments’ table helped to standardize the treatments that a patient could be expected to undergo in this data, the ‘cardiovascular disease’ table helps to standardize and better organize the conditions faced by patients. Originally the current condition of the patient was left as a string to be entered by the doctor. By replacing all of these with an id and linking that id to a table featuring names (‘name’) and descriptions (‘desc’) of disease the conditions affecting a patient can be more easily listed and tied to other patients experiencing the same conditions. This can help us find better options for treatment by requesting all patient treatment records for each ‘idCVD’. We also included with it a list of common symptoms (‘symptoms’) for reference in reports and other functions. Overall this change may be minor but it does significantly help increase the ease of use associated with this research data.

After finishing this stage of the design, a passover was done to verify that proper normalization had been achieved.. Overall, the database would be free from most common anomalies. The database no longer had the feeling of manually hunting down logs from a list of treatments to find the logs tied to patients. Now patients could be narrowed down from the crowd and tied directly to their history. Additionally, this schema would allow researchers to cut out large samples from an enormous population to narrow down treatment options based on how well previous treatments had worked on similar patients. The treatment log may have lost most of its original information but that allowed it to work as the central element that the rest of the schema worked around.

Implementation

In setting up the CVD program, we used the MySQL Workbench IDE tool for managing our MySQL database due to its feature set and interface. It has visual database modeling capabilities that we can use to easily design and modify complex database schemas, ensuring efficient data organization and structure. Several key programming practices were implemented to ensure an efficient and user-friendly database system. The project began with two physical data structures translated directly into database tables. This approach facilitated the breakdown of complex forms into manageable database structures, laying the foundation for effective data management. Additionally, a strong emphasis was placed on normalization, By adhering to normalization principles, the database was structured to minimize redundancy and maintain data integrity, enhancing its reliability and scalability. Our program prioritized the standardization of data by creating separate tables for treatment ID numbers/names and cardiovascular disease ID numbers/conditions. This practice organized and categorized data and improved usability as well as consistency across entries. To optimize the user experience, changes were made to streamline data gathering and presentation, such as standardizing treatments and disease conditions.

The model for the database was first created in MySql Workbench. The first version of this was in line with our conceptual schema, with distinct tables and foreign keys linking relations. The forward engineering tool allowed for the model to be implemented into a database so that work could begin on the user interface and internal queries. To help assist with query design and troubleshooting and to allow demonstrations of the database to be better understood we also created a set of simulated data to add to the database. We elected to only create a small sample of data (30 patients, 15 Available Treatments, 10 CVDs and 50 Treatments with associated results) so that we could test the system without having to devote an overly long time to the creation of data. This data was created using Excel by creating workbooks representing each table with their attribute and creating formulas to simulate random values. Each formula applied to an attribute’s column was given a set with values that would be considered reasonable for the attribute. For example, a patient’s height in centimeters could be anywhere between 140 to 200 (from about 4’5” to 6’5”). For attributes relating to the heart condition’s of the patients we used information from our collected sources about heart disease to determine the reasonable values for these, with a broader upward range to partially simulate patient’s experiencing complications of CVDs. This was also done for the treatment results.

In our database implementation, we utilized CSV (Comma-Separated Values) files as a means to efficiently import and export data for various entities within the database. Excel lists were initially created for available treatments, cardiovascular diseases, patients, results, and treatments. These lists served as convenient areas for organizing and managing data before being imported into the database. In terms of good programming practices, we adhered to principles of modularity and reusability by encapsulating CSV reading and writing functionalities into separate modules or functions. This allows for easy modification or extension of the data import/export process as requirements evolve. While the database schema and structure were custom-designed for the specific requirements of the heart disease treatment analysis project, certain functions from the code can be reused in other future database projects with similar objectives as this. For example, the data loading scripts for importing CSV files into MySQL tables can be adapted for use in other database projects involving data migration or integration tasks. Additionally, the modular design of the database schema allows for easy extension and modification to handle future changes in data requirements.

Version control for the project was managed using Google Drive. Due to complications with attempts to host the database in a more accessible way, we elected to host our documents in a shared Google Drive folder. This contained all generated data, working documents involved in research and planning the schema, and a sql file that could be run to perform all database setup. Anytime this setup file was modified or improved, the group was notified of the change and we verified that no other work was done. Manual version control proved to be a mostly trivial issue with no major hindrances.

Reporting and Analytics

Patient by Attribute

This query would retrieve patients based on specific attributes such as age, gender, ethnicity, or smoking status. For example, it could generate a report listing all patients who are smokers or patients with a BMI above a certain level (Received using height and weight attributes).

Effective Treatment by CVD

This query would analyze the effectiveness of different treatments for specific cardiovascular diseases (CVDs). It would involve joining the treatment, results, and cardiovascular disease tables to identify which treatments return the best outcomes for each CVD. For instance, it could produce a report with a query that joins the treatment, results, and cardiovascular disease tables to analyze treatment outcomes.

Expected CVD by Existing Condition

This query would predict the likelihood of developing certain CVD based on existing patient conditions and demographics. It would involve analyzing previous patient data to identify correlations between risk factors and the development of CVDs. For example, by examining the systolic and diastolic pressure measurements you can check for existing conditions or susceptibility of hypertension.

View of joined tables in SQL workbench

idTreatment	lengthPrescribed	idPatients	height	age	weight	sex	ethnicity	systolicBP	diastolicBP	cholesterol	glucose	smokeResponse	vapeResponse	alcoholIntake	physicalActivity	idResults	systolicFinal	diastolicFinal	cholesterolFin
TR04	72	P022	192	62	252	F	Black/Africanamerican	102	97	260	101	1	1	7	7	R004	142	65	237
TR16	58	P028	178	64	217	M	Multiracial/Biracial	106	76	216	152	1	1	4	2	R016	128	78	214
TR25	26	P009	189	38	232	F	White/Caucasian	147	65	189	113	0	0	1	5	R025	104	78	215
TR31	41	P016	150	54	269	F	Hispanic/Latino	121	73	182	229	1	1	4	7	R031	157	79	257
TR33	82	P025	156	47	159	M	Black/Africanamerican	149	84	249	96	1	0	6	0	R033	126	95	256
TR45	14	P012	140	38	207	M	Hispanic/Latino	147	98	201	96	1	1	5	3	R045	117	64	214
TR47	30	P024	167	55	290	F	White/Caucasian	137	72	245	165	1	1	7	2	R047	136	80	258
TR02	96	P001	146	29	149	F	Hispanic/Latino	102	64	217	120	1	1	2	3	R002	113	98	225
TR36	57	P018	155	67	287	F	White/Caucasian	144	70	235	152	0	1	0	5	R036	134	80	206
TR44	72	P027	191	28	172	M	Multiracial/Biracial	124	91	257	159	1	0	6	5	R044	119	84	190
TR15	109	P024	167	55	290	F	White/Caucasian	137	72	245	165	1	1	7	2	R015	105	62	191
TR35	32	P008	179	34	215	F	Hispanic/Latino	137	87	220	236	0	1	7	4	R035	100	97	242
TR01	158	P005	164	24	179	M	White/Caucasian	141	100	246	126	0	1	3	0	R001	139	63	241
TR14	45	P025	156	47	159	M	Black/Africanamerican	149	84	249	96	1	0	6	0	R014	149	97	253
TR37	34	P008	179	34	215	F	Hispanic/Latino	137	87	220	236	0	1	7	4	R037	115	64	216
TR38	174	P006	193	53	146	F	Asian/Pacific Islander	131	91	260	121	1	1	3	4	R038	104	74	243
TR39	177	P012	140	38	207	M	Hispanic/Latino	147	98	201	96	1	1	5	3	R039	146	79	193
TR08	141	P013	153	68	274	F	Native American/Alas...	146	90	191	147	1	0	0	7	R008	153	76	254
TR12	147	P008	179	34	215	F	Hispanic/Latino	137	87	220	236	0	1	7	4	R012	156	88	227
TR29	138	P016	150	54	269	F	Hispanic/Latino	121	73	182	229	1	1	4	7	R029	126	92	241
TR21	18	P030	184	41	247	F	Asian/Pacific Islander	109	73	222	144	0	1	5	5	R021	118	69	181
TR46	40	P009	189	38	232	F	White/Caucasian	147	65	189	113	0	0	1	5	R046	134	75	211
TR50	127	P021	176	57	183	M	Black/Africanamerican	154	84	217	126	0	0	5	0	R050	152	74	198
TR10	121	P023	199	29	267	M	Native American/Alas...	113	60	200	95	1	0	4	6	R010	103	98	186
TR17	80	P028	178	64	217	M	Multiracial/Biracial	106	76	216	152	1	1	4	2	R017	121	82	244
TR22	98	P021	176	57	183	M	Black/Africanamerican	154	84	217	126	0	0	5	0	R022	153	90	229
TR40	173	P006	193	53	146	F	Asian/Pacific Islander	131	91	260	121	1	1	3	4	R040	141	80	219
TR48	35	P016	150	54	269	F	Hispanic/Latino	121	73	182	229	1	1	4	7	R048	121	96	244
TR24	114	P008	179	34	215	F	Hispanic/Latino	137	87	220	236	0	1	7	4	R024	113	99	186
TR41	176	P020	191	56	147	F	Native American/Alas...	140	69	247	223	0	1	2	7	R041	103	93	232
TR06	102	P016	150	54	269	F	Hispanic/Latino	121	73	182	229	1	1	4	7	R006	116	73	221
TR18	156	P023	199	29	267	M	Native American/Alas...	113	60	200	95	1	0	4	6	R018	103	95	240
TR07	124	P028	178	64	217	M	Multiracial/Biracial	106	76	216	152	1	1	4	2	R007	100	65	183
TR27	66	P009	189	38	232	F	White/Caucasian	147	65	189	113	0	0	1	5	R027	126	100	201
TR30	157	P030	184	41	247	F	Asian/Pacific Islander	109	73	222	144	0	1	5	5	R030	136	73	197
TR42	91	P027	191	28	172	M	Multiracial/Biracial	124	91	257	159	1	0	6	5	R042	157	72	210

P	cholesterol	glucose	smokeResponse	vapeResponse	alcoholIntake	physicalActivity	idResults	systolicFinal	diastolicFinal	cholesterolFinal	glucoseFinal	idTreatmentPlan	description	type	idCVD	diseaseName
	246	126	0	1	3	0	R001	139	63	241	213	TRP05	Amiodarone (Antiarrhythmics)	Perscription Medicine	CV06	Aortic Disease
	249	96	1	0	6	0	R014	149	97	253	131	TRP05	Amiodarone (Antiarrhythmics)	Perscription Medicine	CV05	Peripheral Artery Disease
	220	236	0	1	7	4	R037	115	64	216	224	TRP05	Amiodarone (Antiarrhythmics)	Perscription Medicine	CV10	Pericardial Disease
	260	121	1	1	3	4	R038	104	74	243	209	TRP05	Amiodarone (Antiarrhythmics)	Perscription Medicine	CV02	Valve disease
	201	96	1	1	5	3	R039	146	79	193	102	TRP05	Amiodarone (Antiarrhythmics)	Perscription Medicine	CV04	Heart Failure
	245	165	1	1	7	2	R015	105	62	191	137	TRP04	Bisoprolol (multipurpose)	Perscription Medicine	CV07	Cerebrovascular Disease
	220	236	0	1	7	4	R035	100	97	242	105	TRP04	Bisoprolol (multipurpose)	Perscription Medicine	CV10	Pericardial Disease
	235	152	0	1	0	5	R003	134	75	244	238	TRP15	Cornoary Angioplasty	Surgery	CV10	Pericardial Disease
	201	149	1	0	4	0	R023	107	69	253	146	TRP15	Cornoary Angioplasty	Surgery	CV04	Heart Failure
	210	228	0	0	3	3	R043	132	85	218	148	TRP15	Cornoary Angioplasty	Surgery	CV10	Pericardial Disease
	182	229	1	1	4	7	R049	112	76	230	181	TRP15	Cornoary Angioplasty	Surgery	CV05	Peripheral Artery Disease
	201	149	1	0	4	0	R011	100	82	196	131	TRP14	Cornonary Stent	Surgery	CV04	Heart Failure
	222	144	0	1	5	5	R019	110	98	233	150	TRP14	Cornonary Stent	Surgery	CV05	Peripheral Artery Disease
	240	207	1	1	6	0	R020	160	96	217	219	TRP14	Cornonary Stent	Surgery	CV03	Coronary artery disease
	201	96	1	1	5	3	R026	159	98	184	200	TRP14	Cornonary Stent	Surgery	CV04	Heart Failure
	235	152	0	1	0	5	R028	149	70	239	141	TRP14	Cornonary Stent	Surgery	CV10	Pericardial Disease
	260	101	1	1	7	7	R034	105	67	216	176	TRP14	Cornonary Stent	Surgery	CV01	Arrhythmia
	217	120	1	1	2	3	R005	153	77	249	144	TRP13	Intense cardiovascular excer...	Exercise	CV09	Congenital Heart Disease
	260	121	1	1	3	4	R009	149	60	201	212	TRP13	Intense cardiovascular excer...	Exercise	CV02	Valve disease
	201	96	1	1	5	3	R013	120	97	252	143	TRP13	Intense cardiovascular excer...	Exercise	CV04	Heart Failure
	213	83	0	1	4	6	R032	102	96	211	200	TRP13	Intense cardiovascular excer...	Exercise	CV07	Cerebrovascular Disease
	182	229	1	1	4	7	R006	116	73	221	159	TRP11	Light cardiovascular exercise...	Exercise	CV05	Peripheral Artery Disease
	200	95	1	0	4	6	R018	103	95	240	116	TRP11	Light cardiovascular exercise...	Exercise	CV05	Peripheral Artery Disease
	260	101	1	1	7	7	R004	142	65	237	175	TRP01	Losartan (antihypertensive)	Perscription Medicine	CV01	Arrhythmia
	216	152	1	1	4	2	R016	128	78	214	196	TRP01	Losartan (antihypertensive)	Perscription Medicine	CV06	Aortic Disease
	189	113	0	0	1	5	R025	104	78	215	210	TRP01	Losartan (antihypertensive)	Perscription Medicine	CV03	Coronary artery disease
	182	229	1	1	4	7	R031	157	79	257	115	TRP01	Losartan (antihypertensive)	Perscription Medicine	CV05	Peripheral Artery Disease
	249	96	1	0	6	0	R033	126	95	256	109	TRP01	Losartan (antihypertensive)	Perscription Medicine	CV05	Peripheral Artery Disease
	201	96	1	1	5	3	R045	117	64	214	103	TRP01	Losartan (antihypertensive)	Perscription Medicine	CV04	Heart Failure
	220	236	0	1	7	4	R024	113	99	186	142	TRP10	Low Charbohydrate Diet	Diet	CV10	Pericardial Disease
	247	223	0	1	2	7	R041	103	93	232	190	TRP10	Low Charbohydrate Diet	Diet	CV10	Pericardial Disease
	200	95	1	0	4	6	R010	103	98	186	167	TRP09	Low Cholesterol Diet	Diet	CV05	Peripheral Artery Disease
	216	152	1	1	4	2	R017	121	82	244	237	TRP09	Low Cholesterol Diet	Diet	CV06	Aortic Disease
	217	126	0	0	5	0	R022	153	90	229	208	TRP09	Low Cholesterol Diet	Diet	CV01	Arrhythmia
	260	121	1	1	3	4	R040	141	80	219	124	TRP09	Low Cholesterol Diet	Diet	CV02	Valve disease
	182	229	1	1	4	7	R048	121	96	244	226	TRP09	Low Cholesterol Diet	Diet	CV05	Peripheral Artery Disease

Here we have an SQL query that retrieves comprehensive data from multiple tables within the database. The SELECT statement gathers information from various tables, including treatment details, patient demographics, medical results, available treatments, and cardiovascular disease. Each table is assigned an alias for clarity and brevity. The join clause establishes relationships between tables based on the common key column, such as patient IDs, and treatment IDs. By linking these tables together, the query effectively combines related data, allowing for a holistic view of patients' treatment histories, demographics, and medical outcomes.

Our database can be used for a

This database could be used for various purposes related to healthcare management, specifically patient management. The database can store patient demographic information, medical history, and cardiovascular risk factors. Healthcare providers can use this information to track patient health over time and tailor treatment plans according to what's needed. Treatment plan healthcare professionals can use our database to track treatment plans for patients with cardiovascular diseases. This includes recording prescribed medications and lifestyle changes. Clinical researchers can analyze the data stored in the database to study trends in cardiovascular health, identify risk factors, and evaluate the effectiveness of different treatments. Disease public health agencies can use the database to monitor the occurrences of cardiovascular diseases within specific populations. This information can inform the public of disease outbreaks. Health providers can use the database to identify high-risk patients and provide prevention strategies. This can help empower patients to make informed decisions about their cardiovascular health. Overall, this database serves as a valuable tool for managing and monitoring cardiovascular health, facilitating evidence based on recorded data, and improving patient outcomes.

Discussion and future work

Our project provides an adequate foundation for a much larger open-source project. Ideally, medical records like the ones simulated in our projects would be easily accessible by researchers and students. To encourage the digitization of usable data, we may focus future research on three key areas: expanding categories observed, creating views, merging existing datasets, and web hosting.

First, expanding the functionality of our database is essential to promoting wide adoption. By collecting and studying different data sets already available, we could learn what researchers are looking for. Another option would be to directly contact several researchers for their input. They could tell us what makes a good dataset and what data has little value in collecting.

Secondly, we focus on creating views and documentation to help others merge their datasets with ours. Creating views would help researchers and students understand how to use the database for research. Each view would highlight how the dataset could be used to derive insightful information, which would also help researchers find new uses for the dataset. Collecting the existing data is also very important in promoting its use by third parties. We will need to use large existing datasets and add them to our database. If we create documentation explaining how we prepared the data for migration, we could help third parties migrate their own datasets.

Finally, researching how to create web applications and web hosting will be necessary if we want private practices or hospitals to actively contribute to the database. A web application will help users correctly enter and retrieve data. It will also lower the barrier to entry by allowing users to use any combination of hardware by offloading computation to a cloud server. We will also need to learn how to host a web page. This will be challenging as we must determine the most cost-effective way to host our web application.

Concluding remarks

In conclusion this project represents a comprehensive solution for managing CVD health data efficiently by capturing patient info, risk factors, treatments, and patients documented responses to medications. It's reach extends beyond patient management to include clinical research, surveying diseases, and health initiatives potentially being a big help in the future. While this project serves as a small scale setting, its impact on a larger scale could be substantial. This database lays a foundation for scalable solutions that can benefit larger systems and populations.

References

Centers for Disease Control and Prevention. "Heart Disease Facts." *Centers for Disease Control and Prevention*, 15 May 2023, www.cdc.gov/ Heart disease /facts.htm.

"Diabetes." https://www.mayoclinic.org/. Accessed April 30, 2024. https://www.mayoclinic.org/diseases-conditions/diabetes/diagnosis-treatment/drc-20371451

Gregory A. Roth, George A. Mensah, Valentin Fuster,The Global Burden of Cardiovascular Diseases and Risks: A Compass for Global Action, Journal of the American College of Cardiology, Volume 76, Issue 25, 2020, Pages 2980-2981, ISSN 0735-1097, https://doi.org/10.1016/j.jacc.2020.11.021. (https://www.sciencedirect.com/science/article/pii/S0735109720378037)

"High blood pressure." https://www.mayoclinic.org/. Accessed April 30, 2024. https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/symptoms-causes/syc-20373410

"High Cholesterol." https://www.mayoclinic.org/. Accessed April 30, 2024. https://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/diagnosis-treatment/drc-20350806.

Kishore, S, Blank, E, Heller, D. et al. Modernizing the World Health Organization List of Essential Medicines for Preventing and Controlling Cardiovascular Diseases. J Am Coll Cardiol. 2018 Feb, 71 (5) 564–574. https://doi.org/10.1016/j.jacc.2017.11.056

Nadakinamani, Rajkumar Gangappa, et al. "Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques." *Computational Intelligence and Neuroscience*, vol. 2022, 11 Jan. 2022, pp. 1–13, https://doi.org/10.1155/2022/2973324.

Petrella, R. J., Lattanzio, C. N., Demeray, A., Varallo, V., & Blore, R. (2005). Can adoption of regular exercise later in life prevent metabolic risk for cardiovascular disease? Diabetes Care, 28(3), 694-701. doi:https://doi.org/10.2337/diacare.28.3.694

Şahin, Bayram, and Gülnur İlgün. "Risk factors of deaths related to cardiovascular diseases in World Health Organization (WHO) member countries." Health & Social Care in the Community 30.1 (2022): 73-80.