Transactions on Energy Systems and Engineering Applications

Article

Data analysis model to categorize the level of cardiovascular risk in the Department of Atlantico-Colombia

Luis Consuegra[1], Paul Sanmartin-Mendoza[1,*], Diana Heredia-Vizcaino[1] and Monica Huerta[2]

[1] Faculty of Engineering, Universidad Simón Bolívar, Barranquilla, Colombia.
[2] Facultad de Ciencias e Ingenierías, Universidad Estatal de Milagro, Milagro, Ecuador
[*] Correspondence: paul.sanmartin@unisimon.edu.co

Received: 19 January 2025; Accepted: 11 December 2025; Published: 14 January 2026

Abstract: This article proposes a model based on data analysis to categorize the level of cardiovascular risk in department of Atlantico-Colombia. In the process, the fundamental concepts on which this research proposal was based were identified. As a result, different risk estimation methods were identified according to the region where the studies were carried out; therefore, it was necessary to determine the most appropriate method for the target population of the proposed model, determining a cardiovascular risk prediction model validated in Colombia. To develop the model, we took into account the basis of the concepts and previous studies on cardiovascular risk estimation, identifying during the process the elements involved in the data analysis process, determining the components of the model and characterizing the functional relationships between them.

 https://doi.org/10.32397/tesea.vol6.n2.828

1. Introduction

Cardiovascular diseases (CVD) are the leading cause of death worldwide, according to the World Health Organization (WHO) [1]. The Ministry of Health in Colombia reports that ischemic heart disease, stroke, diabetes, and hypertensive disease rank 1°, 3°, 8° y 9°, respectively, among the top ten causes of mortality in Colombia [2].

CVD encompasses conditions affecting the heart and vascular system. The WHO classifies various pathologies under the umbrella term CVD, including stroke, peripheral arterial disease, deep vein thrombosis, and pulmonary embolism, as well as specific diseases such as coronary, congenital, and rheumatic heart diseases. These conditions belong to the group of non-communicable chronic diseases (NCDs), which are characterized by their chronic progression. Cardiovascular diseases require significant socio-health resources, representing a substantial expenditure in primary healthcare, far exceeding that of other diseases, as highlighted by Campos de Aldana [3,4].

How to cite this article: Lastname, Firstname; Lastname, Firstname; Lastname, Firstname. Data analysis model to categorize the level of cardiovascular risk in the Department of Atlantico-Colombia. *Transactions on Energy Systems and Engineering Applications*, 6(2): 828, 2026. DOI:10.32397/tesea.vol6.n2.828

According to the WHO, the incidence of NCDs is expected to increase by 55% to 59% by the 2030s, accounting for more than half of CVD-related deaths [1,5]. Notably, NCDs share modifiable behavioral risk factors, which can be treated or controlled, such as tobacco use, unhealthy diets, low intake of fruits and vegetables, physical inactivity, and harmful alcohol consumption. These behaviors contribute to overweight and obesity, elevated blood pressure and cholesterol levels, and, consequently, increase the likelihood of developing cardiovascular disease [6].

Globally, the number of deaths reported by the WHO [1] in 2016 reached 56.9 million. Among these, ischemic heart disease and stroke, classified as CVD, accounted for a combined total of 15.2 million deaths, representing a mortality rate of 26.7%. Similarly, Jiménez [7] emphasizes that cardiovascular diseases constitute the leading cause of mortality worldwide, accounting for 30% of all registered deaths, and confirms that in Colombia, they remain the primary cause of morbidity and mortality.

The challenges presented by cardiovascular diseases and their associated risk factors make it imperative to propose mechanisms and strategies, particularly through technological solutions, to strengthen the healthcare system's capacity for the identification, prevention, and control of cardiovascular diseases in populations or subgroups.

Table 1. High-risk criteria according to different societies and organizations [8].

Societies	Risk table	High risk
PAPPS-semFYC	Classic Framingham	$\geq 20\% 10 years$
European	European	$\geq 20\% 10 years$
British	British	$\geq 30\% 10 years$
New Zealand	New Zealand	$\geq 10 - 15\% 5 years$
	Sheffield	$\geq 30\% 10 years$

Historically, large datasets from epidemiological cohorts and clinical trials have resulted in risk models based on high-quality data. However, their external generalization can be limited for two main reasons [9]: the populations studied are not always representative of all adults, and the data collection methods used in these studies often differ from those in clinical settings. To address these limitations, recent studies have worked to generalize models further, as demonstrated by the Cohorts for Heart and Aging Research in Genomic Epidemiology – Atrial Fibrillation (CHARGE-AF) score [10, 11], which is based on pooled data from multiple observational cohort studies and follows the American Heart Association's 2013 cholesterol guidelines . In this context, the WHO developed a cardiovascular risk calculator based on an adaptation of the Framingham model for the AMR-B region of Latin America and the Caribbean. This tool, available for mobile devices and computers, estimates the risk of cardiovascular events such as myocardial infarction and stroke over a 10-year period. In addition to calculating body mass index (BMI), it includes practical features such as medication reminders to help patients avoid treatment interruptions. The tool aims to assist physicians in quickly estimating cardiovascular risk and discussing ways to mitigate it with their patients. It also serves as a resource for individuals concerned about their health, enabling them to assess whether a medical consultation may be necessary [12, 13]. Complementing these efforts, the Spanish Heart Foundation provides a simple tool that offers immediate recommendations to help patients reduce cardiovascular risk. This prevention-focused approach emphasizes proactive policies to improve patient health. However, despite its global recognition, the Framingham model has significant limitations: it is based on an American population, leading to biases when applied to other regions, such as Northern Europe. Additionally, it does not account for family history of early coronary disease or other risk factors like triglycerides (especially when combined with low HDL cholesterol levels), fibrinogen, or homocysteine. It also cannot be applied to patients with pre-existing cardiovascular disease. These limitations highlight the need for models tailored to specific populations, such as the Atlántico department in Colombia, which is the focus of this study [8, 14]. In line with these needs, Colombia's Ministry of Health implemented the "Know Your Risk" strategy, a mobile application designed to help adults over 18 identify their risk of conditions such as heart attack, stroke, or diabetes. The app evaluates risk factors such as obesity, hypertension, smoking, and physical inactivity, providing users with the opportunity to take timely actions to prevent these conditions. Finally, the evolution of cardiovascular risk estimation models reflects efforts to adapt them to specific contexts and populations. From the Framingham SCORE in the United States (1948) to more recent models like QRISK in the United Kingdom (2007), each tool has sought to overcome the limitations of its predecessors, demonstrating how research and innovation can contribute to better prevention and management of cardiovascular diseases (see Table 2).

Table 2. Main chronologies of cardiovascular risk estimators.

Baremo	Derivation Cohort	Validation cohort	Starting year
Framingham SCORE	United States, 30-62 years	Various	1948
PROCAM SCORE	Germany, 35-74 years	Germany	1979
	Europe, 45-65 years	Europe	2002
ASSIGN	Scotland, 30-74 years	Scotland	2006
Reynolds SCORE	United States, 45-80 years	United States	2007
QRISK	United Kingdom, 35-74 years	United Kingdom	2007

The proposed research integrates information technologies to address the pressing public health issue of cardiovascular diseases, focusing on prevention and control strategies tailored to the Atlántico department in Colombia. By leveraging a data-driven approach, the study identifies key elements for predicting cardiovascular risk and establishes a validated model suited to the region's population. This model incorporates fundamental concepts, prior studies, and region-specific conditions, ensuring its relevance and applicability. Validation was conducted using a representative sample from the Atlántico department, applying the model components and analyzing the results to support evidence-based decision-making in cardiovascular health management.

2. Methods and Materials

This section outlines the methodological framework and materials used to conduct the study, detailing the research approach, design, data collection methods, and technological integration. By providing a comprehensive overview of these components, it ensures clarity and coherence in understanding the processes and tools employed to achieve the research objectives.

2.1. Research Approach

This study employs a mixed-methods approach, integrating qualitative and quantitative elements with analytical considerations to achieve the research objectives. This methodology facilitates a comprehensive understanding of the characteristics and measurements of the studied phenomenon while engaging directly with relevant populations within the research context. By combining diverse perspectives, the mixed-methods approach enhances the depth and precision of the analysis. The selection of measurement instruments as viable solutions to the research problem further ensures the reliability of the obtained results, justifying the chosen methodology as an effective means to enrich the research process and outcomes.

2.2. Type of Research

The study is classified as applied research, focusing on determining cardiovascular risk through a mathematical data analysis model. The target population includes patients from the Atlántico department who meet predefined criteria. Grounded in a real-world contextual situation, the research adopts a formative focus, aiming to develop plural preventive actions that address the identified risks effectively.

2.3. Research Design

This study employs a non-experimental research design, characterized by the absence of variable manipulation or data alteration. It incorporates a longitudinal component, capturing data at a single point in time for the development of the analysis model.

2.4. Methods and Techniques for Data Collection

To ensure credibility and objectivity, this study employs a combination of data collection techniques. The gathered information includes variables relevant to the mathematical model designed for cardiovascular risk assessment. Quantitative data is systematically entered into the model and analyzed using numerical tools, such as tables and graphs. These techniques are applied to data from participants selected as the study sample, ensuring the reliability and legitimacy of the information. Tailored analytical tools enhance the precision and validity of the results, aligning them with the research objectives.

2.5. Population and Sample

The target population comprises 10,270 patients, while the sample includes 6,218 individuals selected based on specific criteria relevant to the model's variables. These variables include gender, age, smoking habits, family history (e.g., death from acute myocardial infarction or other cardiovascular diseases), body mass index (BMI), diabetes mellitus, systolic and diastolic blood pressure, HDL cholesterol, LDL cholesterol, and triglycerides. The sample was intentionally selected using a non-probabilistic convenience method to ensure alignment with the research objectives and the development of the analysis model.

2.6. Technological Integration for Cardiovascular Risk Analysis

The study integrates a technologically driven analysis model as a core component of its methodology. This model includes electronic health records (EHR), a category selector, a risk estimator, and a data presentation system. By embedding the model into EHR systems, it ensures immediate accessibility at the point of care, eliminating the need for manual risk calculations by healthcare professionals. The model's technical design, represented through detailed diagrams and well-defined specifications, enhances comprehension and operational efficiency, enabling effective analysis and the generation of valuable insights.

3. Introduction to the Analysis Model in E-Health

The proposed model aligns with current E-Health trends, emphasizing the integration of information and communication technologies to enhance clinical decision-making. According to Eysenbach (citation), E-Health, often referred to as "Internet medicine," transcends its digital dimension to include features such as efficiency and continuous improvement in medical process quality. Within this framework, the analysis model serves as a technical solution, leveraging interrelated components to facilitate robust data exploration. This approach generates accurate and actionable information essential for informed decision-making in cardiovascular risk prevention and treatment.

3.1. Medical Record

The medical record is a mandatory document that records a patient's health conditions and contains their detailed information, in accordance with the regulations outlined in [15], which govern medical ethics. According to Article 34 (Law 23 of the Congress of the Republic of Colombia), "The medical record is the

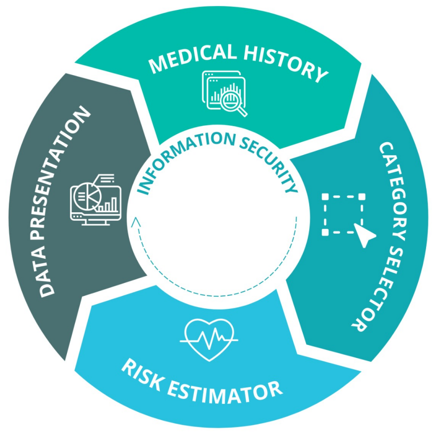

Figure 1. Proposed Data Analysis Model

mandatory documentation of a patient's health conditions." This record includes a set of documents related to the care process and the health status of an individual, prepared by a healthcare professional [16]. In this context, the central element of the medical record is the person or patient.

In the development of the proposed model, the medical record component establishes filters to exclude patients already diagnosed with cardiovascular events from the analysis population. Due to their condition, these patients are part of the population under established control and treatment. The model's objective is to identify patients with a significant probability of experiencing a cardiovascular event, categorizing them on a risk scale. This serves as a guide for determining the appropriate treatment for each case.

3.2. Category selector

This component aligns with the concept of Data Modeling, which focuses on presenting only the data that is valuable to the end user. In the context of cardiovascular risk estimation, this involves extracting and organizing medical characteristics and patient assessments based on specific information criteria. The category selection process leverages EHRs (Electronic Health Records) using Extract, Transform, and Load (ETL) methods and techniques. These methods enable the retrieval of both structured and unstructured data directly from the source, applying filters to extract relevant information, standardizing it, and preparing it for further analysis and implementation.

Within the category selector, informative, demographic, and clinical variables essential for estimating cardiovascular risk (see Table 3) are identified and selected. Additionally, this process examines the patient population requiring prioritization in prevention and control plans. An ETL process is applied to the EHRs to ensure the data is standardized and detailed, meeting the requirements for risk estimation. This approach facilitates the accurate interpretation of the resulting values and enhances decision-making processes.

After selecting the variables, the process continues with the transformation of medical records by categorizing the values and storing the information in a data warehouse for further analysis. A data warehouse is a specialized database that functions as a centralized repository, consolidating data from multiple sources within the existing information systems of an organization or company [18]. To ensure its effectiveness, a data warehouse must meet specific requirements, including the following:

- The data warehouse must make the information of a company/institution/organization easily accessible.
- The data warehouse must consistently provide information about the company/institution.
- The data warehouse must present data to be used as a foundation or guideline for decision-making.

Table 3. Attributes for the estimation of cardiovascular risk adjusted for Colombia [17].

Attribute	Description
Sex	Group of people (male - female)
Age	Quantitative variable in years of life of a person
Smoking	Tobacco use in the last 12-month period
Family history	Death due to acute myocardial infarction or other cardiovascular disease
BMI Body mass index	Quantitative variable according to body weight and height
Diabetes mellitus	Epidemiological variable of established diagnosis
Systolic and Diastolic blood pressure (mmhg)	Continuous quantitative variable that tracks the presence of intrinsic organic rhythms
HDL cholesterol (mg/dL)	Quantitative high-density lipoprotein variable
LDL cholesterol (mg/dL)	Quantitative low-density lipoprotein variable
Triglycerides (mg/dL)	Quantitative variable of a specific type of body fat

3.3. Risk estimator

Risk estimators are derived from cohort studies focused on cardiovascular risk. These studies, which are longitudinal, observational, and analytical in nature, involve cohorts—groups of individuals sharing a predefined characteristic—that are monitored over time [19]. These prospective studies, widely used in epidemiology, are conducted by professionals in the field who define hypotheses to identify risk factors, analyze the conditions that contribute to disease development, and calculate associated risks. Developing a cardiovascular risk estimator requires an epidemiological study that reviews past and current events related to disease progression, comparing and analyzing risk factors and the conditions affecting the population under study. This approach aims to determine whether observed conditions are driven by natural or social factors [20].

Table 4. Transformed values.

Order	Value
1	YES
2	NO

Table 5. HDL cholesterol.

Order	Name/Range	From	To	sex
1	Optimum	56	500	M
2	Standard Risk	35	55	M
3	High Risk	0	34	M
4	Optimal	66	500	F
5	Standard Risk	45	65	F
6	High Risk	0	44	F

Based on the results of these epidemiological cohort studies [19], it is possible to calculate the incidence of disease relative to risk factor exposure. This relationship is organized into calculation tables, also known as scales, which facilitate the classification of risk levels and the variables involved in estimating the likelihood that an individual within a specific population will develop a disease within a given timeframe. The quality of a scale depends on the representativeness of the reference sample. Selected values must not

Table 6. LDL cholesterol.

Order	Name/Range	From	To
1	Optimum	0	100
2	Near Optimum	101	129
3	High Limit	130	159
4	High	160	189
5	Very High	190	500

Table 7. Triglycerides.

Order	Name/Range	From	To
1	Normal	0	150
2	High Limit	151	199
3	High	200	499
4	Very High	500	2000

Table 8. Systolic blood pressure.

Order	Name/Range	From	To
1	Hypotension	0	79
2	Normal	80	120
3	Prehypertension	121	139
4	Hypertension grade 1 (HTA 1)	140	159
5	Grade 2 (HTA 2)	160	179
6	Hypertensive crisis	180	500

be arbitrary but should accurately reflect the characteristics of the group to which the individuals being evaluated belong.

Table 4 presents the options for establishing the ranges within the model in terms of smoking, diabetes mellitus and family history of acute myocardial infarction.

Tables 5 to 10 present the options for defining the ranges used in the model for various key health indicators:

- HDL Cholesterol (Table 5): Levels range from optimal (the ideal target for estimating and treating the patient) to high risk.
- LDL Cholesterol Table 6): Ranges go from optimal to very high as the upper level.
- Triglycerides (Table 7): Levels considered range from normal to very high.
- Systolic Blood Pressure (Table 8): Values span from hypotension, as the lower limit, to hypertensive crisis, as the highest range.
- Ages (Table 9): Patients are categorized within a range of 30 to 90 years, inclusive of both limits.
- Body Mass Index (BMI) (Table 10): Ranges include underweight (0 to 18.4) up to class III obesity (above 40).

In this way, the defined indicators allow for comprehensive categorization of patients according to their risk profiles and provide a foundation for personalized clinical interventions.

Table 9. Ages.

Order	Name/Range	From	To
1	From 30 to 40	30	40
2	From 41 to 50	41	50
3	From 51 to 60	516	60
4	From 61 to 70	61	70
5	From 71 to 80	71	80
6	From 81 to 90	81	90

Table 10. BMI.

Order	Name/Range	From	To
1	Underweight	0	18,4
2	Normal Weight	18,5	24,9
3	Overweight	25	29,9
4	Obesity Class I	30	34,9
5	Obesity Class II	35	39,9
6	Obesity Class III	40	100

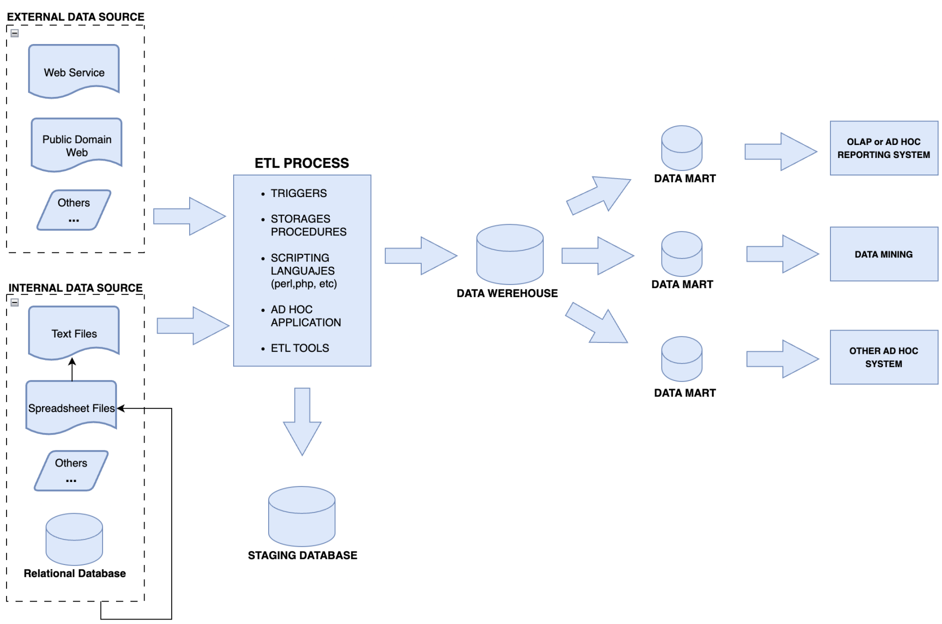

Figure 2. Data warehouse architecture [18].

3.4. Data presentation

In the analysis model, the data presentation component is linked to the concept of Data Analysis, which involves evaluating information and generating result reports. In this process, technological tools play a crucial role by enabling the manipulation of data and its representation through dashboards or scorecards, commonly referred to as Business Intelligence (BI) platforms.

These advanced analytical tools facilitate the modeling of query-based representations to create comprehensive dashboards that serve as the foundation for report presentation. They are essential in the field of business intelligence, which refers to the strategic use of data to support decision-making.

Furthermore, these tools not only enhance the understanding of an organization's current performance but also enable the anticipation of future events, providing critical insights to inform strategic decisions.

Figure 3. Magic quadrant for analitic and business intelligence platforms [21].

Modern analytics and business intelligence (BI) platforms are characterized by intuitive tools that support the entire analytical workflow, from data preparation and ingestion to visual exploration and actionable insights generation. These platforms differ significantly from traditional BI systems as they do not require extensive IT involvement to predefine data models or store information in conventional data warehouses [21]. Figure 3 presents the ranking of BI tool companies as classified by Gartner [21], a consulting and research firm specializing in emerging technologies, based on features and capabilities that distinguish the various products.

A notable feature of modern BI platforms is the integration of augmented analytics, which leverages machine learning (ML) algorithms to process and extract insights from large datasets. Additionally, augmented analytics incorporates natural language processing (NLP) to query data and generate narratives, making it easier to interpret key drivers and visualizations.

To optimize data interpretation, graph creation, secure ingestion of information sources, and presentation of dashboards to end users, the data presentation component must meet the following criteria:

- Administration, security, and BI platform architecture
- Cloud BI
- Data source connectivity and ingestion
- Metadata management
- Data storage and loading options
- Data preparation
- Scalability and complexity of data models
- Advanced analytics for citizen data scientists
- Analytical dashboards
- Interactive visual exploration
- Augmented data discovery

- Mobile exploration and creation
- Embedding analytical content
- Publishing, sharing, and collaborating on analytical content
- Ease of use, visual appeal, and workflow integration

4. Results

4.1. Medical Record

As a first step, the data management capabilities of a Health Service Provider in the Atlántico department - Health Services Provider Institution (HSPI), were analyzed. This entity operates a proprietary health management system, which records the clinical information of patients covered by its health services. The application of the model began by defining the patient population involved in the cardiovascular risk estimation analysis and determining the clinical values required to categorize the risk level. Electronic health records were accessed through the HSPI's information system, and anonymized and authorized data were retrieved via Structured Query Language (SQL) queries for this research.

During the execution of structured SQL queries, limitations related to the database design were identified. Each system maintains its own standards, resulting in non-homogeneous data across different health providers' systems, a recurring issue throughout Colombia. Law N°2015 (Congress of the Republic of Colombia, Bogotá) [22] aims to regulate interoperable electronic health records. This mechanism will facilitate information flow between health service providers and standardize data based on technical guidelines established by the Ministry of Information and Communication Technologies (MinTIC, 2020).

As a result, it was necessary to document the data dictionary of the HSPI's proprietary system to build the SQL queries and review its structure. Findings revealed unformatted values, diagnoses, and medical history recorded in admission observations. Data query restrictions included conditions such as age (adult population over 30 years), clinical laboratory results created within six months prior to the query, and patients diagnosed with cardiovascular events, in line with the criteria established by the model in the electronic health records.

The query results identified the attributes defined in the category selector, the patient identifier within the entity, and the demographic variable of location. Figure 4 presents the transactional model used to construct the queries within the Microsoft SQL Server Management Studio tool.

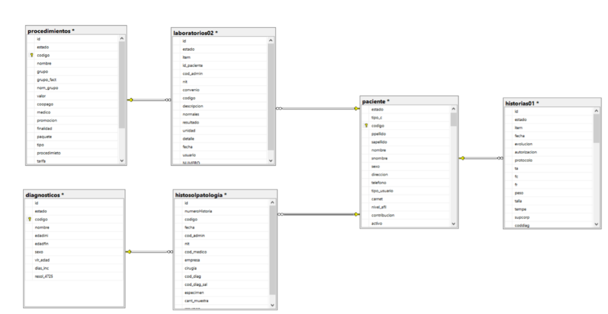

Figure 4. Transactional model.

4.2. Category selector

In the category selector stage, processes were carried out to extract and clean data from the Electronic Health Record (EHR). The data originated from queries performed on an SQL database, and the results were processed using the Pentaho Data Integration (PDI) Extract, Transform, and Load (ETL) tool. A data transformation and a scheduled task were created to ensure continuous synchronization with the original data source.

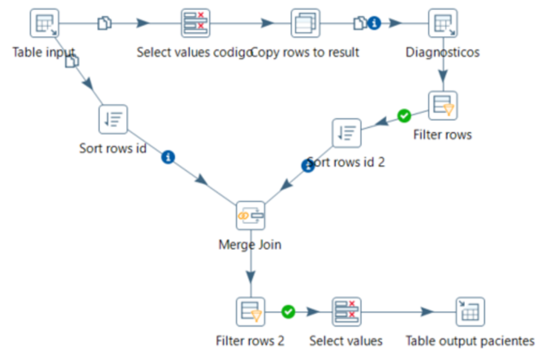

Figure 5. Tranformation Data Designe.

The process of data transformation and selection was designed using PDI's graphical components, organizing a sequence of steps to classify information according to predefined categories. The workflow began with the Table Input module, which retrieved 10,270 patient records from EHR databases. For each record, a second query was executed on a previously created view of clinical diagnoses, using parameters derived from the previous step. To achieve this, the Select Value and Copy Row to Result modules were employed, configuring the patient code column as the search parameter.

The transformation also included the Filter Rows module, which removed records with missing values or data types outside the ranges allowed by the category selector. The Sort Rows and Merge Join modules were used to unify the results from the two Table Input sources. Finally, the workflow concluded with the Table Output module, which stored 6,218 processed records in a PostgreSQL database management system used as a data warehouse.

This ETL tool not only served as an interface for input and output of EHR data but also facilitated the cleaning, refining, and storage of records, preparing them for subsequent analysis in the data warehouse.

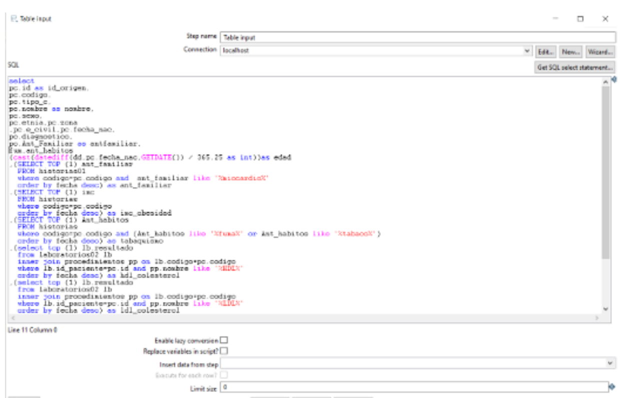

Figure 6. Transactional model.

Figure 6 shows the data input step from an SQL query in Pentaho Data Integration, used for transforming data from the clinical records. Meanwhile, Figure 6 illustrates the results of the category selector process, which were stored in a table within a PostgreSQL data warehouse, with the patients' attributes prepared for subsequent analysis.

4.3. Risk estimator

In the risk estimator component (Section 3.3), the calculation defined by the PROCAM model was implemented, following the recommendations for the Colombian population. This component was developed using the Power BI data analytics platform, selected for its ability to meet infrastructure requirements, role-based authentication and user permissions, network transport layer security, and data protection through gateway connections. These features ensure the security of the information being analyzed while meeting the requirements of a business intelligence (BI) platform for data presentation.

The first step in implementing the risk estimator was to establish a connection to the patients' data warehouse and create the tables defined in the category selector. Subsequently, a table containing the PROCAM scale, used in the risk estimator, was added to establish the necessary relationships within the data model in the platform (see Figure 7). This figure provides an overview of the data model visualized in Power BI after integrating the data into the tool. The system automatically associates cardiovascular risk levels based on the definitions of the PROCAM model, which serves as the foundational basis for creating this component.

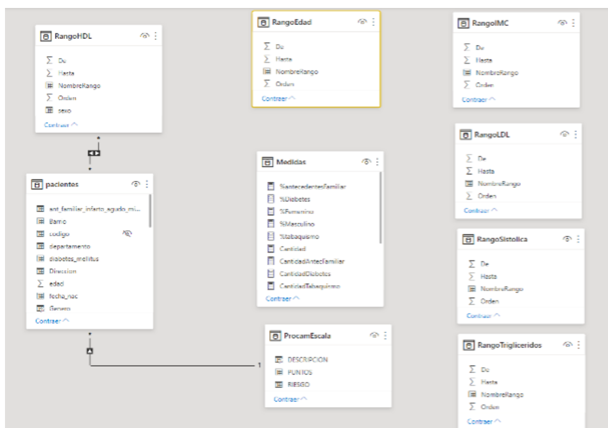

Figure 7. Power BI model.

The Power BI model view allows for the visualization of the dataset to facilitate data analysis, representing the tables and defining the relationships between them. Based on the established data model, the necessary metrics for the risk estimator component were created. Using DAX, the language provided by the tool, calculations were formulated based on the point values of the variables defined in the implemented PROCAM Score estimator.

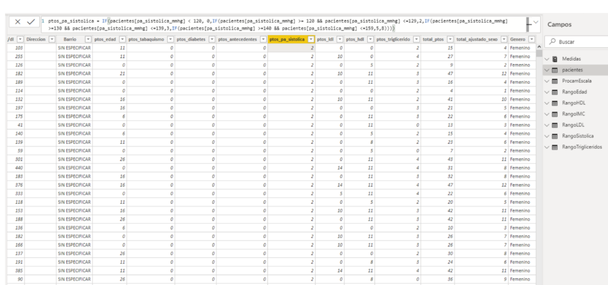

Figure 8. Formulation of PROCAM point calculation in POWER BI.

Using formulas, the score calculation for the implemented risk estimator is applied, and based on the obtained score, the 10-year cardiovascular risk probability is estimated. By leveraging the calculated risk score for each record and the relationship established between the patients and the PROCAM scale defined in the data model, the cardiovascular risk level for each patient is determined. This enables the creation of visualizations to facilitate data analysis.

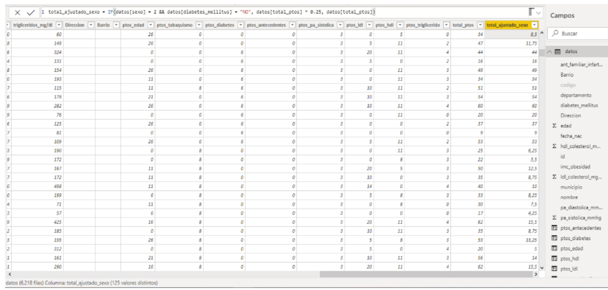

Figure 9. Formulation of PROCAM point calculation in POWER BI.

4.4. Data presentation

First, a design was proposed based on the available information about patient variables, the types of analyses that could be performed using this data, and a user experience assessment of the tool's various visualization options. Subsequently, a table was created in the data model to define the metrics used for constructing the dashboard. Building on this foundation, the dashboard was designed to analyze the data, incorporating various visualizations and filters to represent the population distribution according to the variables included in a population profile report and its risk estimation. Finally, the dashboards were published on the Power BI web platform for user analysis.

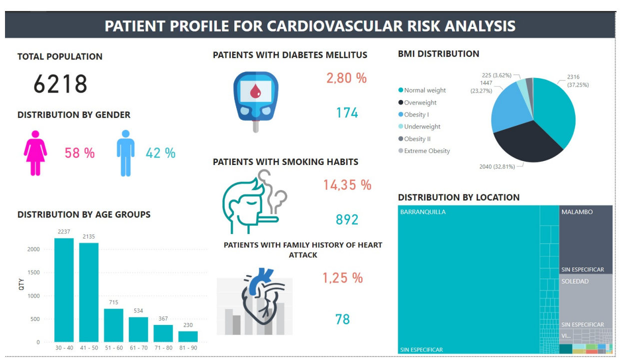

Figure 10. Presentation of patient profile data.

The data presentation was conducted through various visualizations provided by Power BI Desktop, illustrating the distribution of patients based on the variables selected for risk analysis. As a result, the population was classified into 3,599 women (58%) and 2,169 men (42%), totaling 6,218 individuals distributed across the Atlántico department (see Table 11).

The analysis identified risk factors in 174 patients diagnosed with diabetes, 892 smokers, and 78 with a family history of cardiovascular diseases. Another significant factor was the body mass index (BMI), which showed that 36.8% of the population, equivalent to 2,288 patients, had a normal weight, while 63.2%, or 3,970 patients, experienced nutritional issues (see Table 12).

The previous figure displays the dashboard illustrating the distribution of cardiovascular risk among patients based on the various variables included in the risk estimation calculation. Regarding risk categorization according to the PROCAM score, it was estimated that 163 patients (2.62%) are at high risk, 255 (4.10%) are at medium risk, and 5,800 (93.28%) are at low risk. This analysis indicates a low overall

Table 11. Distribution of patients in the municipalities of the Atlántico department.

Municipality	Women	Men	Total
BARANOA	7	6	13
BARRANQUILLA	3097	1571	4668
GALAPA	3	10 13	
MALAMBO	219	504	723
PALMAR DE VARELA	1	7	8
POLONUEVO	8	5	13
PONEDERA	0	8	8
PUERTO COLOMBIA	3	1	4
REPELON	0	2	2
SABANAGRANDE	10	16	26
SABANALARGA	1	3	4
SANTO TOMAS	4	6	10
SOLEDAD	246	477	723
SUAN	0	1	1
USIACURÍ	0	2	2
	3599	2619	6218

Table 12. Distribution of patients by BMI.

Rank Name	BMI
Underweight	233
Obesity Class I	1444
Obesity Class II	164
Obesity Class III	26
Normal Weight	2288
Overweight	2038
	6218

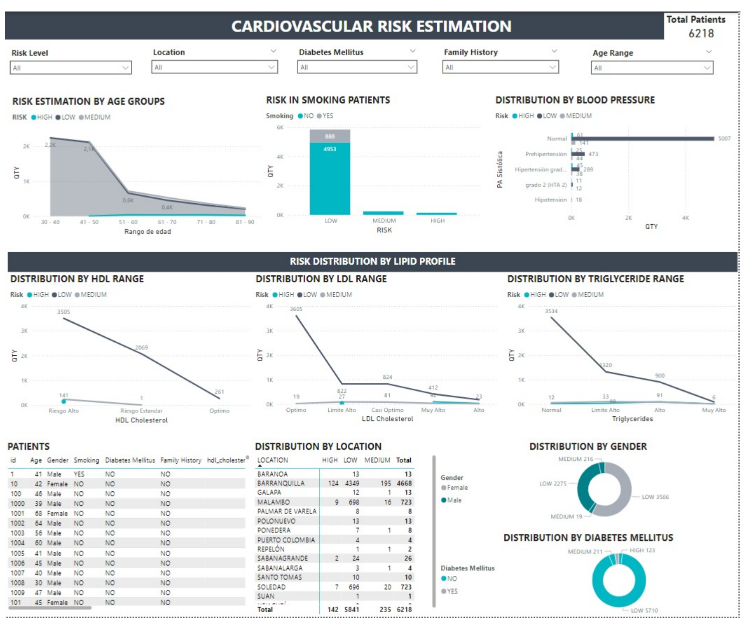

Figure 11. Presentation of estimated risk data.

cardiovascular risk in the population, as most individuals analyzed are middle-aged adults, as shown in Table 13.

Table 13. Risk distribution by age range.

Age Range	LOW	MEDIUM	HIGH	Total
From 30 to 40	2235	2	0	2237
From 41 to 50	2105	27	3	2135
From 51 to 60	606	65	44	715
From 61 to 70	411	80	43	534
From 71 to 80	272	49	46	367
From 81 to 90	171	32	27	230
	5800	255	163	6218

However, the analysis is limited by the over representation of patients from the city of Barranquilla. Table 14 presents the distribution of risk by municipality according to the model's application. This study aims to establish the relationship between age ranges and cardiovascular risk levels (low, medium, or high).

Table 14. Risk Distribution by Municipality.

Municipality	LOW	MEDIUM	HIGH	Total
BARANOA	13	0	0	13)
BARRANQUILLA	4309	215	144	4668
GALAPA	12	1	0	13
MALAMBO	697	16	10	723
PALMAR DE VARELA	8	0	0	8
POLONUEVO	13	0	0	13
PONEDERA	7	1	0	8
PUERTO COLOMBIA	4	0	0	4
REPELON	1	1	0	2
SABANAGRANDE	24	0	2	26
SABANALARGA	3	1	0	4
SANTO TOMAS	10	0	0	10
SOLEDAD	696	20	7	723
SUAN	1	0	0	1
USIACURÍ	2	0	0	2
	5800	255	163	6218

5. Conclusions

The research focused on proposing a data analysis-based model to categorize cardiovascular risk levels, aiming to support the development of prevention and control plans for these diseases in the Atlántico department (state) of Colombia.

The validation of the PROCAM scale conducted by Muñoz et al [] in the Colombian population provided an essential starting point for the development, proposal, and application of a cardiovascular risk estimation model tailored to the context of the Atlántico department. This model, specifically calibrated for the Colombian population, adjusted the original score—based on male populations—by applying the suggested factor of 0.25 for non-diabetic women.

The model's validation proved effective, as the sample used (Atlántico department) met the pre-established characteristics. However, its implementation requires significant effort in processing clinical data adapted to each healthcare institution. The adoption of an interoperable electronic health record is expected to streamline this stage, enabling uniform application across various institutions. Furthermore, this model represents a significant advancement, as no similar modeling had previously been conducted within the Atlántico department, integrating technological solutions into cardiovascular analysis strategies for the population.

Developing a CDSS (Clinical Decision Support System) is a complex process requiring a multidisciplinary approach. In this context, opportunities and challenges were identified concerning scope and requirements, infrastructure and interface, data quality, population, and model validity. Moreover, the regulatory environment, encompassing legislation, privacy, and scientific evaluation, must evolve to support this emerging field of eHealth.

The scope of this project is broad and applicable to other research requiring the use of predictive models, integrating the concepts within a systemic framework. However, limitations remain, such as the need for national studies to assess the model's adherence to similar or diverse contexts, in larger populations with varied age ranges. The COVID-19 pandemic is a relevant variable that could be incorporated into future research and developments, alongside the inclusion of new variables that would allow the model to adapt to different scenarios and populations.

Finally, the implementation of the model is expected to enable the continuous updating of data in electronic health records to identify and focus on preventive strategies for at-risk patients. Incorporating this model into healthcare management software will expedite risk estimation diagnoses for healthcare professionals, while web and mobile platforms will offer the general population tools to estimate their risk and seek timely medical care. Future work includes the publication of a scientific article summarizing the research findings and the continuation of the project, focusing on testing within traditional treatment frameworks.

Acknowledgments

We would like to thank the health service providers that supported the development of the research, as well as the Simon Bolivar University, which has been fundamental for this work, and the Ministry of Science, Technology and Innovation of Colombia for their support and funding. On the other hand, the grammar of this paper was corrected with IA tools.

Funding: This research was funded by Universidad Simón Bolivar and Minciencias (Ministerio de Ciencia Tecnología e Innovación - Colombia)

Disclosure statement: The authors declare no conflict of interest.

References

[1] World Health Organization. The top 10 causes of death, 2018. Retrieved from http://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death.

[2] Ministerio de salud. Principales causas de mortalidad en colombia, 2010. Retrieved from https://www.minsalud.gov.co/salud/Paginas/Enfermedades-cardiovasculares.aspx.

[3] José Canchila-Agámez, Santiago García-Cardona, Cesar Viloria-Nuñez, Jose Atilio Nuñez-Ramos, and David Rivas-Lalaleo. Proposed architecture for a patient management and classification system in the triage process. In *International Conference on Technological Innovation and AI Research (ICTIAIR 2025)*, volume 2025, pages 127–132. IET, 2025.

[4] M. S. Campos de Aldana, D. Moya Plata, J. D. Mendoza Matajira, and E. Y. Duran Niño. Chronic non-communicable diseases and use of information and communication technology: Systematic review. *Revista Cuidarte*, 5(1):661–669, 2014.

[5] World Health Organization. Enfermedades no transmisibles, 2018. Retrieved from https://www.who.int/es/news-room/fact-sheets/detail/noncommunicable-diseases.

[6] World Health Organization. Global status report on noncommunicable diseases 2014, 2014. Retrieved from https://www.who.int/nmh/publications/ncd-status-report-2014/en/.

[7] J. E. Jiménez Mesino. *Autopercepción del estado de salud en individuos con riesgo cardiovascular en Colombia*. PhD thesis, Master's thesis, 2019. Universidad del Norte.

[8] A. Álvarez Cosmea. Las tablas de riesgo cardiovascular: Una revisión crítica. *Medifam*, 11(3):20–51, 2001. Recuperado en 06 de enero de 2025.

[9] B. A. Goldstein, A. M. Navar, and M. J. Pencina. Risk prediction with electronic health records: the importance of model validation and clinical context. *JAMA cardiology*, 1(9):976–977, 2016.

[10] Yasuhiro Mukai, Koki Nakanishi, Masao Daimon, Kentaro Iwama, Yuriko Yoshida, Kazutoshi Hirose, Yuko Yamamoto, Hikari Seki, Tomoko Nakao, Tsukasa Oshima, Takumi Matsubara, Yu Shimizu, Gaku Oguri, Toshiya Kojima, Eriko Hasumi, Katsuhito Fujiu, Hiroyuki Morita, and Issei Komuro. Prevalence, associated factors, and echocardiographic estimation of left atrial hypertension in patients with atrial fibrillation. *Journal of the American Heart Association*, 12(18), 2023.

[11] VQ Méndez, PS Mendoza, and DH Vizcaino. Interoperabilidad: un reto que trasciende el intercambio de información y datos. *Datos, información, tendencias*, 11, 2018.

[12] Karen Avila, Paul Sanmartin, Daladier Jabba, and Miguel Jimeno. Applications based on service-oriented architecture (soa) in the field of home healthcare. *Sensors*, 17(8):1703, 2017.

[13] J. Orozco, L. Rodríguez, and C. Viloria-Nuñez. Medibot: Intelligent virtual assistant for personalized medical care and guidance. In *2024 IEEE Technology and Engineering Management Society (TEMSCON LATAM)*, pages 1–6. IEEE, 2024.

[14] Vladimir Quintero, Cesar Chevel, and Paul Sanmartin-Mendoza. Analysis on the interoperability of health information systems. *2024 IEEE Technology and Engineering Management Society (TEMSCON LATAM)*, pages 1–6, 2024.

[15] D. E. C. L. A. R. A. C. I. Ó. N. DE PRINCIPIOS. Ley 23 de 1981, 1981.

[16] J. E. Díaz-Realpe, J. Muñoz-Martínez, and C. H. Sierra-Torres. Factores de riesgo para enfermedad cardiovascular en trabajadores de una institución prestadora de servicios de salud, colombia. *Revista de Salud pública*, 9:64–75, 2007.

[17] P. Cullen, H. Schulte, and G. Assmann. The münster heart study (procam), 1997.

[18] R. Kimball and J. Caserta. *The data warehouse ETL toolkit*. John Wiley & Sons, 2004.

[19] P. Salazar, C. Manterola, G. Quiroz, N. García, T. Otzen, M. Mora, and G. Duque. Estudios de cohortes. 1 parte. descripción, metodología y aplicaciones. *Revista de cirugía*, 71(5):482–493, 2019.

[20] Gabriela González. Estudio prospectivo: características, ejemplos y limitaciones, 2021. Lifeder. Recuperado de https://www.lifeder.com/estudio-prospectivo/.

[21] B. Analysts, C. Howson, J. Richardson, R. Sallam, and A. Kronz. Magic quadrant for analytics and business intelligence platforms, 2020.

[22] República de Colombia. Ley 2015, historia clínica electrónica interoperable, 2020. Congreso de la República.