



Article

# Major Adverse Cardiovascular Events in Coronary Type 2 Diabetic Patients: Identification of Associated Factors Using Electronic Health Records and Natural Language Processing

Carlos González-Juanatey <sup>1,\*</sup>, Manuel Anguita-Sánchez <sup>2</sup>, Vivencio Barrios <sup>3</sup>, Iván Núñez-Gil <sup>4</sup>, Juan José Gómez-Doblas <sup>5</sup>, Xavier García-Moll <sup>6</sup>, Carlos Lafuente-Gormaz <sup>7</sup>, María Jesús Rollán-Gómez <sup>8</sup>, Vicente Peral-Disdier <sup>9</sup>, Luis Martínez-Dolz <sup>10</sup>, Miguel Rodríguez-Santamarta <sup>11</sup>, Xavier Viñolas-Prat <sup>6</sup>, Toni Soriano-Colomé <sup>12</sup>, Roberto Muñoz-Aguilera <sup>13</sup>, Ignacio Plaza <sup>14</sup>, Alejandro Curcio-Ruigómez <sup>15</sup>, Ernesto Orts-Soler <sup>16</sup>, Javier Segovia <sup>17</sup>, Víctor Fanjul <sup>18</sup>, Ángel Cequier <sup>19</sup> and SAVANA Research Group <sup>†</sup>

- <sup>1</sup> Hospital Universitario Lucus Augusti, 27003 Lugo, Spain
  - <sup>2</sup> Hospital Universitario Reina Sofía, Instituto Maimonides de Investigación Biomédica de Córdoba (IMIBIC), Universidad de Córdoba, 14014 Córdoba, Spain
  - <sup>3</sup> Hospital Universitario Ramón y Cajal, 28034 Madrid, Spain
  - <sup>4</sup> Hospital Clínico Universitario San Carlos, 28040 Madrid, Spain
  - <sup>5</sup> Hospital Universitario Virgen de la Victoria CIBERCV (Centro de Investigación Biomédica en Red Enfermedades Cardiovasculares) and IBIMA (Instituto de Investigación Biomédica de Málaga), 29010 Málaga, Spain
  - <sup>6</sup> Hospital Universitario Santa Creu i Sant Pau, 08041 Barcelona, Spain
  - <sup>7</sup> Hospital Universitario de Albacete, 02006 Albacete, Spain
  - <sup>8</sup> Hospital Universitario Río Hortega, 47012 Valladolid, Spain
  - <sup>9</sup> Hospital Universitario Son Espases, 07120 Palma de Mallorca, Spain
  - <sup>10</sup> Hospital Universitario y Politécnico La Fe, CIBERCV, IIS La Fe, 46026 Valencia, Spain
  - <sup>11</sup> Hospital Universitario de León, 24071 León, Spain
  - <sup>12</sup> Hospital Vall d'Hebron, CIBERCV, 08035 Barcelona, Spain
  - <sup>13</sup> Hospital Infanta Leonor, 28031 Madrid, Spain
  - <sup>14</sup> Hospital Infanta Sofía, 28703 Madrid, Spain
  - <sup>15</sup> Hospital Universitario de Fuenlabrada, 28942 Madrid, Spain
  - <sup>16</sup> Hospital General Universitario de Castellón, 12004 Castellón de la Plana, Spain
  - <sup>17</sup> Hospital Universitario Puerta de Hierro, 28222 Madrid, Spain
  - <sup>18</sup> Savana Research, 28013 Madrid, Spain
  - <sup>19</sup> Hospital Universitario de Bellvitge, IDIBELL (Instituto de Investigación Biomédica de Bellvitge), Universidad de Barcelona, 08007 Barcelona, Spain
- \* Correspondence: carlos.gonzalez.juanatey@sergas.es; Tel.: +34-982-296-809  
† SAVANA Research Group are shown in the Acknowledgements.



**Citation:** González-Juanatey, C.; Anguita-Sánchez, M.; Barrios, V.; Núñez-Gil, I.; Gómez-Doblas, J.J.; García-Moll, X.; Lafuente-Gormaz, C.; Rollán-Gómez, M.J.; Peral-Disdier, V.; Martínez-Dolz, L.; et al. Major Adverse Cardiovascular Events in Coronary Type 2 Diabetic Patients: Identification of Associated Factors Using Electronic Health Records and Natural Language Processing. *J. Clin. Med.* **2022**, *11*, 6004. <https://doi.org/10.3390/jcm11206004>

Academic Editors: Alessandro Delitala and Francesco Giallauria

Received: 8 September 2022

Accepted: 6 October 2022

Published: 11 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Patients with Type 2 diabetes mellitus (T2DM) and coronary artery disease (CAD) are at high risk of developing major adverse cardiovascular events (MACE). This is a multicenter, retrospective, and observational study performed in Spain aimed to characterize these patients in a real-world setting. Unstructured data from the Electronic Health Records were extracted by EHRRead<sup>®</sup>, a technology based on Natural Language Processing and machine learning. The association between new MACE and the variables of interest were investigated by univariable and multivariable analyses. From a source population of 2,184,662 patients, we identified 4072 adults diagnosed with T2DM and CAD (62.2% male, mean age 70 ± 11). The main comorbidities observed included arterial hypertension, hyperlipidemia, and obesity, with metformin and statins being the treatments most frequently prescribed. MACE development was associated with multivessel (Hazard Ratio (HR) = 2.49) and single coronary vessel disease (HR = 1.71), transient ischemic attack (HR = 2.01), heart failure (HR = 1.32), insulin treatment (HR = 1.40), and percutaneous coronary intervention (PCI) (HR = 2.27), whilst statins (HR = 0.73) were associated with a lower risk of MACE occurrence. In conclusion, we found six risk factors associated with the development of MACE which were related with cardiovascular diseases and T2DM severity, and treatment with statins was identified as a protective factor for new MACE in this study.

**Keywords:** diabetes mellitus; coronary artery disease; MACE; risk factors; electronic health records; natural language processing

## 1. Introduction

Diabetes mellitus (DM) is a key public health problem worldwide associated with the development of complications such as cardiovascular comorbidities, which include coronary artery disease (CAD) [1–3]. It was listed as the underlying cause of mortality in >85,000 deaths in the United States in 2019 [2], consisting mainly of cardiovascular complications and CAD development, the main causes of death and disability in patients with type 2 diabetes mellitus (T2DM) [4]. In addition, DM has been associated with a 2 to 4-fold increased risk of death due to heart disease [5].

Given the clinical burden that cardiovascular complications have on patients with T2DM, there has been an increased focus on the joint management of both entities, T2DM and cardiovascular disease (CVD). Furthermore, during the last decade, there has been increasing pressure from regulatory agencies for antidiabetic treatments to demonstrate cardiovascular safety and benefits, especially for major adverse cardiovascular events (MACE) [6]. However, patients included in controlled trials are generally selected on the basis of protocol eligibility and, therefore, do not fully represent the entire population encountered in clinical practice [7]. In this regard, real-world-based studies are needed to improve diagnostic and therapeutic approaches to palliate the burden of disease by reducing cardiovascular events.

The present study aimed to use population-based data to provide real-world insights into the clinical characterization and treatment management of diabetic patients who develop CAD in Spain. Moreover, this study investigated the probability of developing MACE after presentation of CAD in T2DM patients and the potential risk or protective factors associated with its occurrence within this population. To do so, we used EHRead<sup>®</sup> (Madrid, Spain), a technology that applies Natural Language Processing (NLP) and machine learning, to extract, organize, and analyze the unstructured clinical information that health professionals register in patients' electronic health records (EHRs) [8–13].

## 2. Materials and Methods

### 2.1. Study Design

This was a real-world, multicenter, retrospective, and observational study (ACORDE: "Assessment of medical management in CORonary Diabetic Type 2 patients at high risk of cardiovascular events") based on the secondary use of the unstructured data captured in the EHRs. A cross-sectional analysis of all patients was performed at the index date, defined as the first time in the study period when the patient fulfilled all inclusion and no exclusion criteria. At this point, demographics, comorbidities, vital signs, general characteristics of T2DM and CAD and treatments were evaluated. The cumulative incidence of MACE, defined as previously described by the presence of myocardial infarction, stroke, hospitalization for unstable angina, and urgent coronary revascularization [14], was analyzed during the follow-up. Although included in this definition, all-cause or cardiovascular death was not analyzed as, due to the nature of the data source, their occurrence was not accurately identifiable. The follow-up period was defined as the period of time spanned from index date to the last EHR available for each patient within the study period.

### 2.2. Data Source

The data source was the free-text information within the EHRs of 12 representative hospitals from six major regions in Spain: Madrid (Hospital Universitario de Fuenlabrada, Hospital Universitario Infanta Sofía, Hospital Universitario Infanta Leonor, Hospital Universitario Puerta de Hierro), Catalonia (Hospital Universitari Vall d'Hebron, Hospital de la Santa Creu i Sant Pau), Valencia (Hospital Universitari i Politècnic La Fe, Hospital General

Universitario de Castellón), Balearic Islands (Hospital Universitari Son Espases), Castilla La Mancha (Complejo Hospitalario Universitario de Albacete), and Castilla y León (Hospital Universitario Río Hortega, Hospital Universitario de León). Data entry in outpatient clinical reports, discharge reports, emergency reports, prescriptions, and other medical reports, were collected from all available services and departments in each participating site, including inpatient hospital, outpatient hospital, and emergency room, between 1 January 2014 and 31 December 2018.

### 2.3. Study Subjects

The source population of the study comprised all adult patients with available EHRs in the participating hospitals within the study period with at least 6 months of follow-up. Patients with T2DM and a diagnosis of CAD were included. These two entities were considered when they were contained in the unstructured, free-text information in the EHRs based on clinical diagnosis. T2DM were also considered when documented ongoing use of glucose-lowering drugs (oral hypoglycemic agents) was found for at least 6 months. CAD was considered if there was evidence of stenosis  $\geq 50\%$  of at least 1 coronary artery but without a history of having a previous myocardial infarction or stroke, and without planned coronary, cerebrovascular, or peripheral arterial revascularization. Patients with prior myocardial infarction or stroke, history of liver cirrhosis or liver cancer, intracranial bleeding, renal failure requiring dialysis, or ongoing treatment with anticoagulant medication at index date or non-available follow-up information spanning at least 6 months were excluded from the study.

### 2.4. Extraction of the Unstructured Free-Text from EHRs

Unstructured clinical data were extracted using the EHRead<sup>®</sup> technology, which uses NLP, machine learning, and deep learning techniques for extracting free text from de-identified and processed EHR and translates it into a study database. EHRead<sup>®</sup> technology have been described elsewhere [9,10,12,13,15–19].

Using the information obtained from this process, a statistical model was generated to describe the demographic and clinical characteristics of the population with T2DM and CAD. Information was analyzed and expressed by means of concepts that contain the most significant information in the text. The terminology considered by EHRead<sup>®</sup> includes codes, concepts, synonyms, and definitions used in clinical documentation and is based on SNOMED CT [20,21].

### 2.5. External Validation of EHRead<sup>®</sup> Performance

EHRead<sup>®</sup> performance was assessed to evaluate the ability of the technology to identify mentions of coronary disease and its related clinical variables within patients' EHRs [9,12,13,22]. First, a comparison between a physician-annotated set of EHRs (i.e., "gold standard") and the "output" of EHRead<sup>®</sup> upon reading that same set of EHRs was carried out independently at each hospital site [9]. Thus, two designated expert physicians from each hospital annotated a set of randomly selected records while a third physician reviewed the annotations made and resolved any possible discrepancies among the two site physicians. The "output vs. gold standard" comparison was reflected in the standard evaluation metrics, including precision, recall, and their harmonic mean (F1-score). As previously described, we demonstrated a high performance of EHRead<sup>®</sup> at identifying records that contain mentions of "Coronary disease" in the target population with a precision, recall, and F1-score of 0.89, 0.73, and 0.80, respectively. Moreover, the evaluation yielded a F1-score of above 90% in most of the variables analyzed [9,10].

### 2.6. Statistical Data Analyses

Categorical variables were described via frequency tables; whereas numerical variables were presented using summary tables that included the mean, standard deviation (SD), median, and quartiles (Q1, Q3). Missing data were handled according to the nature of the

data collection process, thus lack of information (i.e., unavailable data in patients' EHRs) was considered a "true zero" for binary variables (e.g., absence of a comorbidity) but was treated as missing data for numerical variables (e.g., laboratory values). No missing data imputation was performed. Cumulative incidence of new cases of MACE was estimated through the Kaplan–Meier approach. Both univariable and multivariable analyses were performed to investigate the association between the occurrence of new MACE and the variables of interest. Variables with >20% missing values or with zero variance were excluded. In the univariable analysis, a Cox proportional hazard (PH) regression model was fitted to the study population for each variable at baseline. In the multivariable approach, a data-driven variable and model selection was performed. First, a Cox PH model was fitted to the study population using all eligible study variables. Then, we carried out backward stepwise selection (i.e., the least informative variables were excluded one at a time, and new models were fitted in a recursive manner). We used the Bayesian information criterion (BIC) to reach a model with the optimal explanatory variables. Significant differences were considered when  $p < 0.05$ . All analyses were performed using "R" software, version 4.0.2 (The R Foundation for Statistical Computing, Vienna, Austria).

### 2.7. Ethical Considerations and Study Approval

This study was classified as a "non-post-authorization study" by the Spanish Agency of Medicines and Health Products (AEMPS) and was approved by the Institutional Review Board (IRB) of each participating hospital. All methods and analyses were conducted in compliance with local legal and regulatory requirements, as well as generally accepted research practices described in the Helsinki Declaration in its latest edition, and Good Pharmacoepidemiology Practices. Data were analyzed from de-identified EHRs, which were aggregated in an irreversibly dissociated manner. Thus, individual patient consent was not required in the study.

## 3. Results

### 3.1. Study Population

The source population of the study comprised a total of 2,184,662 adult patients with available EHRs in the hospitals of interest including 217,632 (10%) patients with a confirmed diagnosis of T2DM. The study population only included adult patients with T2DM and stable CAD diagnosis with no previous history of myocardial infarction and stroke, which yielded to a final population of 4072 patients. In total, 37.4% of the patients presented insulin-dependent diabetes. Type of CAD was identified in 60% of the patients. Of which, 833 of the patients (34.21% of the patients with CAD type identified within the EHR) had single coronary vessel disease and 1,602 (65.79%) multivessel coronary disease. Patients were followed up for a median of 33.6 months (IQ1-3: 19.5–47.1).

### 3.2. Demographic and Clinical Characteristics

The sociodemographic and comorbidities of the study population at index date are summarized in Table 1 and Figure 1. As shown, the mean age of the patients was  $70 \pm 11$  years, with two thirds being male (62.2%) and 17.5% current smokers (Table 1).

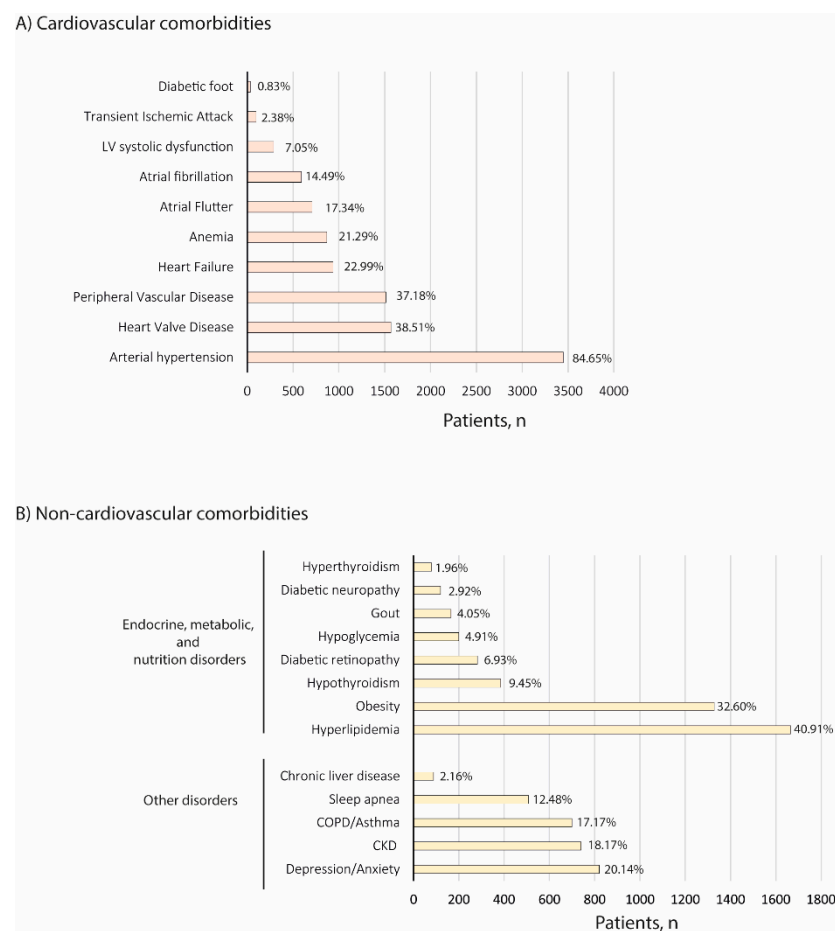
**Table 1.** Sociodemographic of the study population at index.

	Study Population (N = 4072)
Age, years	
Mean (SD)	70 (11)
Median (Q1,Q3)	71 (63,78)

**Table 1.** Cont.

	Study Population (N = 4072)
Gender, n (%)	
Female	1533 (37.6)
Male	2531 (62.2)
Tobacco use	2208 (54.2)
Current smoker	713 (17.5)
Former smoker	1495 (36.7)

SD: Standard deviation.



**Figure 1.** Comorbidities of the study population at index. (A) Cardiovascular and (B) non-cardiovascular comorbidities are shown. CKD: Chronic kidney disease; COPD: Chronic obstructive pulmonary disease.

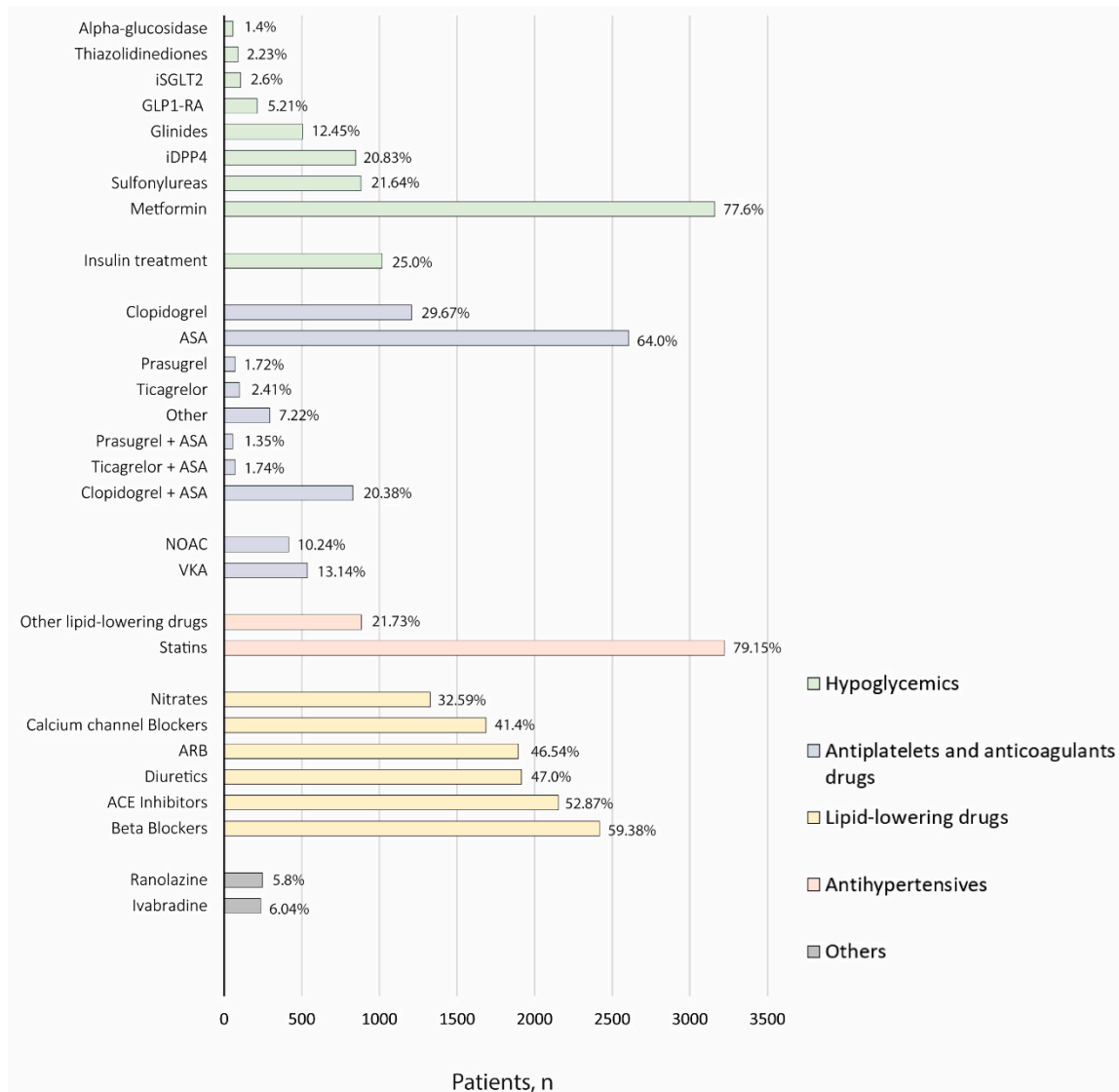
The most common cardiovascular comorbidities reported in the target population included arterial hypertension (84.65%), and heart valve disease (38.51%) (Figure 1A). In regard to other comorbidities observed, the most frequent disorder was hyperlipidemia (40.91%), followed by obesity (32.60%) (Figure 1B).

As shown in Table S1, our study population had clinical parameters consistent with metabolic abnormalities including elevated glucose, body mass index, lipids parameters and hemoglobin A1c values.

### 3.3. Disease Management

Treatment strategies in our study population at index date included pharmacological management and interventional approaches.

Data regarding pharmacological treatment for the management of patients diagnosed with T2DM and CAD are shown in Figure 2.



**Figure 2.** Pharmacological treatment of the study population at index. ACE: Angiotensin-converting enzyme; ARB: Angiotensin II receptor blockers; GLP1-RA: Glucagon-like peptide-1 receptor agonists; iDPP4: Dipeptidyl peptidase-4 inhibitors; iSGLT2: Sodium-glucose cotransporter-2; NOAC: Non-vitamin K antagonist oral anticoagulants; VKA: Vitamin K antagonists.

Metformin was identified as the most frequently prescribed antidiabetic drug (77.6%), followed by insulin/analogs (25%), sulfonylureas (21.64%), dipeptidyl peptidase-4 inhibitors (iDPP4) (20.83%), and glinides (12.45%). Other treatments such as statins (79.15%) and antiplatelet therapy, mainly ASA (64.0%), were commonly prescribed according to the available information within the EHRs.

Regarding interventional procedures, we observed that a total of 1574 patients (38.8%) were revascularized with percutaneous coronary interventions (PCI) and 585 patients (14.4%) were treated with coronary artery bypass graft (CABG).

### 3.4. Major Adverse Cardiovascular Events (MACE)

We evaluated the cumulative incidence of MACE, during the study follow-up.

As shown in Table 2, the probability of developing at least one MACE after 4-years follow-up was 27.09%, with it being more probable to undergo myocardial infarction (13.89%) than developing ischemic stroke (6.94%), unstable angina (4.54%), or urgent coronary revascularization (8.69%).

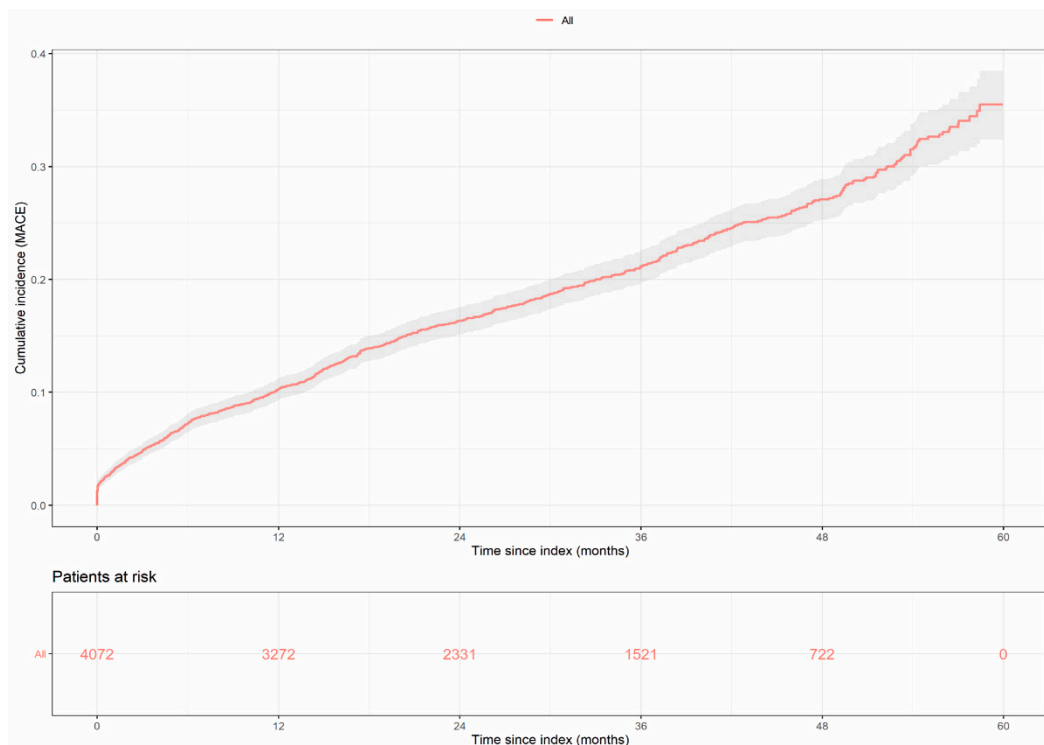
**Table 2.** Cumulative incidence of MACE.

	1-Year FU	2-Year FU	3-Year FU	4-Year FU
MACE *	10.28	16.32	21.18	27.09
Myocardial infarction	4.05	7.8	10.79	13.89
Ischemic stroke	1.77	3.41	5.05	6.94
Unstable angina	1.67	2.49	3.45	4.54
Urgent revascularization	4.04	5.57	6.82	8.69

FU: Follow-up; MACE: Major adverse cardiovascular events.

\* Cumulative incidence of new cases of MACE was estimated through the Kaplan-Meier approach. Each value represents the probability of MACE at different FU time points.

Cumulative incidence curves of the first MACE occurrence after study entry and detailed probability of individual components of interest over study follow-up are shown in Figure 3 and Figure S1, respectively.

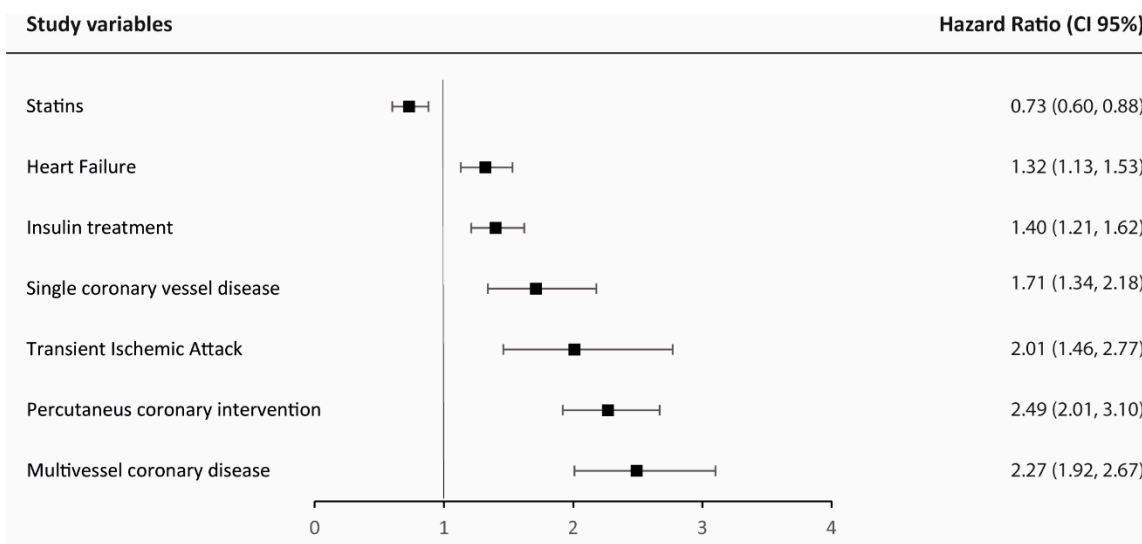


**Figure 3.** Cumulative incidence of MACE over time during the follow-up period. Probability of any new MACE (myocardial infarction, stroke, unstable angina, and urgent revascularization) is shown. The number of patients at risk across the follow-up period is indicated below.

The univariable analysis using Cox PH models showed that age, history of smoking, and the time of first mention of T2DM and CAD disease in the EHRs were significantly associated with the occurrence of new MACE during the follow-up period, while female gender was inversely associated with the development of MACE ( $p < 0.001$ ) (Table S2).

Regarding comorbidities, in general, the presence of all cardiovascular disorders was associated with the occurrence of the first MACE detected during the follow-up. Likewise, diabetic retinopathy, hypothyroidism, and COPD/asthma were found to be associated with the appearance of new MACE in the study (Table S2). In addition, new MACE were associated with prescription of insulin ( $p < 0.001$ ), some hypoglycemic agents including sulfonylureas ( $p = 0.008$ ) and alpha-glucosidase ( $p = 0.049$ ), anticoagulant and antiplatelet therapy ( $p = 0.021$  and  $p < 0.001$ , respectively), beta blockers, calcium channel, nitrates, ranolazines, lipid lowering drugs ( $p < 0.001$  each), diuretics ( $p = 0.012$ ), and statins ( $p = 0.003$ ) (Table S2). Revascularization treatment, including PCI and CABG were also found to be related the occurrence of new MACE during the study period (both  $p < 0.001$ ).

In a multivariable approach using backward stepwise selection based on BIC, we found 7 independent variables to be associated with MACE during follow-up, including coronary involvement with multivessel and single coronary vessel disease (HR = 2.49; 95% CI = 2.01–3.10;  $p < 0.001$ ), and (HR 1.71; 95% CI = 1.34–2.18;  $p < 0.001$ ), transient ischemic attack (HR = 2.01; 95% CI = 1.46–2.77;  $p < 0.001$ ), and heart failure (HR = 1.32; CI 95% = 1.13–1.53;  $p < 0.001$ ). Pharmacological treatment with insulin and interventional therapies as PCI were also found to be independently associated with the occurrence of new MACE (HR = 1.40; CI 95% = 1.21–1.62;  $p < 0.001$  and HR = 2.27; CI 95% = 1.92–2.67;  $p < 0.001$ , respectively). On the contrary, treatment with statins was found to be inversely associated with MACE during follow-up (Figure 4).



**Figure 4.** Forest plot of the 7 independent variables associated with the occurrence of MACE during the follow-up in the multivariable model.

#### 4. Discussion

The present study uses NLP and machine learning to characterize patients with T2DM and CAD in a real-world setting in Spain. Our results showed that: (1) T2DM patients who develop CAD are predominantly male with arterial hypertension, hyperlipidemia and obesity as the most prevalent comorbidities; (2) these patients present an increasing probability of developing MACE during the study follow-up; and (3) in this population the risk factors associated with the development of MACE are related with cardiovascular diseases (multivessel or single vessel disease, PCI, transient ischemic attack, heart failure) and T2DM severity (insulin treatment), whilst lipid lowering control strategies (treatment with statins) seems to be a protective factor. The described methodology has been recently used in the same cohort of patients to describe medical management in coronary T2DM patients with previous percutaneous coronary intervention in Spain, where the EHRead<sup>®</sup> ability to correctly identify patients' records containing key variables associated with the study disease is shown [10].



Prior studies have described that around 7.5–15% of the patients with T2DM also had a diagnosis of CAD [23,24], and only 1.78% had a dual diagnosis and no history of stroke, myocardial infarction, or severe liver disease [7]. These patients showed a higher prevalence of comorbidities, with the consequent elevated burden of the disease [7,23,24]. In our study, and in line with previous publications, 1.87% of the patients were identified with both pathologies after applying the described exclusion criteria. Furthermore, we observed arterial hypertension as the main cardiovascular comorbidity present in the most of included patients, followed by heart valve disease, and peripheral vascular disease. Moreover, hyperlipidemia and obesity were also frequently detected. These data correlate with the ones described previously in ATHENA study [25] and using the Ricerca e Salute (ReS) database [7], in which arterial hypertension was detected as the most common comorbidity observed, followed by dyslipidemia.

Due to the chronicity of the diseases and the different comorbidities associated, the management of the patients with diabetes and CAD is challenging, and usually requires a multidisciplinary approach. In 2012, the ACCF/AHA/ACP/AATS/PCNA/SCAI/STS published their guideline for the diagnosis and management of stable CAD with a few specific recommendations for those patients with diabetes [26]. Since then, several advancements in the diagnosis and management of CAD in diabetic patients have occurred. In order to update management evidence, the American Heart Association released in 2020 a scientific statement on the clinical management of stable CAD in patients with T2DM, emphasizing not only recommendations for diabetes or CAD, but also for hyperlipidemia, hypertension and lifestyle management in this group of patients [27]. We found that metformin, the recommended first-line glucose-lowering treatment [28], was the most frequently antidiabetic prescribed drug in our studied population. Moreover, despite clopidogrel, either alone or in combination with ASA, being preferable as antiplatelet therapy to prevent the generalized prothrombotic state options [29], our results showed that ASA alone were most frequently prescribed. On the other hand, we identified that statins were frequently prescribed in these patients in agreement with actual recommendations, which specifies that when diabetes and CAD occur simultaneously, statin therapy is strongly recommended in the primary and secondary prevention of CAD [27]. Additionally, we found that antiplatelet therapy, insulin and b-blockers were frequently prescribed in clinical practice in this population. Finally, although we did not explore therapeutic strategies to improve lifestyle management, the presence of obesity and tobacco use suggest that there is still a need to intervene in this field.

Despite disease management, patients with T2DM and CAD are at high risk of experiencing MACE [30]. In order to address reliable tools to estimate the absolute cardiovascular risk in these patients, several studies have attempted to identify different factors involved [24,31]. Our results showed that coronary disease and T2DM severity, along with other cardiovascular diseases, were independently associated with the development of MACE. Coronary revascularization has been previously described as a protective factor for the development of MACE in patents after the first acute coronary syndrome [32]. Our results in the subpopulation of T2DM with CAD could suggest that DM is not just a comorbidity that affects the development and progression of CAD, but it also has an important role increasing cardiovascular risk per se, even when revascularization occurs.

In line with previous publications, we found that insulin treatment was associated with MACE occurrence [10,33]. Although randomized trials have shown that insulin reduces adverse cardiovascular events [27], in these patients insulin tends to be prescribed in more advanced phases of the disease, which could explain our findings. Finally, other cardiovascular diseases such as heart failure or transient ischemic attack have also been previously associated with the development of MACE [32], although this has not been evaluated in patients with T2DM and CAD. On the other hand, our results confirmed that the use of statins was a protective factor for MACE in the study population. This is in agreement with the large body of evidence, which indicates that statin-based strategies are beneficial lipid-lowering drugs associated with a relative cardiovascular risk reduction

in patients with T2DM and CAD [27]. These results, including risk and protective factors, could have an important clinical impact since they could help develop and implement preventive measures for MACE when CAD occurs in diabetic patients, especially those that are modifiable.

## 5. Limitations

There are several limitations to this study. First, our results are based on data captured directly from the unstructured, free-text narratives in patients' EHRs. Thus, the findings are limited by the availability and accuracy of EHRs, and by the information reported by physicians in their routine practice. Second, because this is a retrospective study based on real-world data, the findings are only hypothesis generating. Moreover, it was not designed to collect variables in a strict, a priori fashion. Thus, there were some interesting variables that were not properly documented and therefore not analyzed. In this regard, the presence of cardiomyopathy at baseline was not analyzed and patients with an implantable cardioverter defibrillator (ICD) were not considered, both possibly related with MACE development [34–37]. Yet, the effect of some risk factors such as tobacco use may be underreported within the EHRs. Finally, although known confounders were accounted for, unknown confounders might have influenced the findings of the study.

## 6. Conclusions

In conclusion, this study showed that CAD appears in mainly male T2DM patients with other cardiovascular related comorbidities. When it happens, these patients need to be closely monitored because there exists an increasing risk of MACE occurrence. While independent risk factors associated with MACE are difficult to modify since they are intrinsic to the underlying diseases (coronary disease, insulin treatment, transient ischemic attack and heart failure), here we identified statins used as a protective factor that could be used by physicians in clinical practice. Further research is warranted to confirm these findings in future studies.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm11206004/s1>, Figure S1: Cumulative incidence of MACE over time during the follow-up period; Table S1: Vital signs and clinical parameters of the study population at index; Table S2: Factors associated with the occurrence of MACE during the follow-up.

**Author Contributions:** Conceptualization, C.G.-J.; methodology, V.F. and SAVANA Research Group; software, V.F. and SAVANA Research Group; validation, C.G.-J., M.A.-S., V.B., I.N.-G., J.J.G.-D., X.G.-M., C.L.-G., M.J.R.-G., V.P.-D., L.M.-D., M.R.-S., X.V.-P., T.S.-C., R.M.-A., I.P., A.C.-R., E.O.-S., J.S., V.F., SAVANA members and Á.C.; formal analysis, V.F.; investigation, C.G.-J.; data curation, V.F. and SAVANA Research Group; writing—original draft preparation, SAVANA Research Group; writing—review and editing, all authors; visualization, V.F.; supervision, C.G.-J. and Á.C.; funding acquisition, C.G.-J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by AstraZeneca Spain (Externally Sponsored Scientific Research, ESR-18-13815) and sponsored by the Spanish Society of Cardiology.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Drug Institutional Review Board (or Ethics Committee) of the Community of Madrid (CEIm-R, protocol code ESR-18-13815-ACORDE and date of approval October 2018).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** SAVANA Research Group are: Alberto Porras, Miren Taberna, Stephanie Marchesseau, Rosa Bretón-Romero, Ignacio Salcedo, Enrique Josué Álvarez, Claudia Maté, Judith Marín-Corral, Florinda Meléndez, Sebastian Menke, José Aquino and Natalia Polo.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Saeedi, P.; Petersohn, I.; Salpea, P.; Malanda, B.; Karuranga, S.; Unwin, N.; Colagiuri, S.; Guariguata, L.; Motala, A.A.; Ogurtsova, K.; et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. *Diabetes Res. Clin. Pract.* **2019**, *157*, 107843. [[CrossRef](#)] [[PubMed](#)]
2. Tsao, C.W.; Aday, A.W.; Almarzooq, Z.I.; Alonso, A.; Beaton, A.Z.; Bittencourt, M.S.; Boehme, A.K.; Buxton, A.E.; Carson, A.P.; Commodore-Mensah, Y.; et al. Heart Disease and Stroke Statistics-2022 Update: A Report from the American Heart Association. *Circulation* **2022**, *145*, e153–e639. [[CrossRef](#)]
3. Zamora, A.; Marrugata, J. Prognosis of Diabetic Patients with Coronary Heart Disease. *Rev. Española Cardiol.* **2002**, *55*, 751–762. [[CrossRef](#)]
4. DeFronzo, R.A.; Ferrannini, E.; Groop, L.; Henry, R.R.; Herman, W.H.; Holst, J.J.; Hu, F.B.; Kahn, C.R.; Raz, I.; Shulman, G.I.; et al. Type 2 diabetes mellitus. *Nat. Rev. Dis. Prim.* **2015**, *1*, 15019. [[CrossRef](#)] [[PubMed](#)]
5. Aronson, D.; Edelman, E.R. Coronary artery disease and diabetes mellitus. *Cardiol. Clin.* **2014**, *32*, 439–455. [[CrossRef](#)] [[PubMed](#)]
6. Einarson, T.R.; Acs, A.; Ludwig, C.; Pantou, U.H. Prevalence of cardiovascular disease in type 2 diabetes: A systematic literature review of scientific evidence from across the world in 2007–2017. *Cardiovasc. Diabetol.* **2018**, *17*, 83. [[CrossRef](#)] [[PubMed](#)]
7. Maggioni, A.P.; Dondi, L.; Andreotti, F.; Ronconi, G.; Calabria, S.; Piccinni, C.; Pedrini, A.; Esposito, I.; Martini, N. Prevalence, prescriptions, outcomes and costs of type 2 diabetes patients with or without prior coronary artery disease or stroke: A longitudinal 5-year claims-data analysis of over 7 million inhabitants. *Ther. Adv. Chronic Dis.* **2021**, *12*, 20406223211026390. [[CrossRef](#)]
8. Del Rio-Bermudez, C.; Medrano, I.H.; Yebes, L.; Poveda, J.L. Towards a symbiotic relationship between big data, artificial intelligence, and hospital pharmacy. *J. Pharm. Policy Pract.* **2020**, *13*, 75. [[CrossRef](#)]
9. Canales, L.; Menke, S.; Marchesseau, S.; D'Agostino, A.; Del Rio-Bermudez, C.; Taberna, M.; Tello, J. Assessing the Performance of Clinical Natural Language Processing Systems: Development of an Evaluation Methodology. *JMIR Med. Inform.* **2021**, *9*, e20492. [[CrossRef](#)]
10. González-Juanatey, C.; Anguita-Sánchez, M.; Barrios, V.; Núñez-Gil, I.; Gómez-Doblas, J.J.; García-Moll, X.; Lafuente-Gormaz, C.; Rollán-Gómez, M.J.; Peral-Disdie, V.; Martínez-Dolz, L.; et al. Assessment of medical management in Coronary Type 2 Diabetic patients with previous percutaneous coronary intervention in Spain: A retrospective analysis of electronic health records using Natural Language Processing. *PLoS ONE* **2022**, *17*, e0263277. [[CrossRef](#)]
11. Gomollon, F.; Gisbert, J.P.; Guerra, I.; Plaza, R.; Pajares Villarroja, R.; Moreno Almazan, L.; Lopez Martin, M.C.; Dominguez Antonaya, M.; Vera Mendoza, M.I.; Aparicio, J.; et al. Clinical characteristics and prognostic factors for Crohn's disease relapses using natural language processing and machine learning: A pilot study. *Eur. J. Gastroenterol. Hepatol.* **2021**, *34*, 389–397. [[CrossRef](#)]
12. Izquierdo, J.L.; Almonacid, C.; Gonzalez, Y.; Del Rio-Bermudez, C.; Ancochea, J.; Cardenas, R.; Lumbreras, S.; Soriano, J.B. The Impact of COVID-19 on Patients with Asthma. *Eur. Respir. J.* **2020**, *57*, 2003142. [[CrossRef](#)]
13. Izquierdo, J.L.; Ancochea, J.; Savana, C.-R.G.; Soriano, J.B. Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients with COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing. *J. Med. Internet Res.* **2020**, *22*, e21801. [[CrossRef](#)]
14. Choi, B.; Rha, S.; Yoon, S.G.; Choi, C.; Lee, M.; Kim, S. Association of Major Adverse Cardiac Events up to 5 Years in Patients with Chest Pain Without Significant Coronary Artery Disease in the Korean Population. *J. Am. Heart Assoc.* **2019**, *8*, e010541. [[CrossRef](#)]
15. Sheikhalishahi, S.; Miotto, R.; Dudley, J.T.; Lavelli, A.; Rinaldi, F.; Osmani, V. Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review. *JMIR Med. Inform.* **2019**, *7*, e12239. [[CrossRef](#)]
16. Goldstein, B.A.; Navar, A.M.; Pencina, M.J.; Ioannidis, J.P. Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. *J. Am. Med. Assoc.* **2017**, *318*, 198–208. [[CrossRef](#)]
17. Luo, Y.; Thompson, W.K.; Herr, T.M.; Zeng, Z.; Berendsen, M.A.; Jonnalagadda, S.R.; Carson, M.B.; Starren, J. Natural Language Processing for EHR-Based Pharmacovigilance: A Structured Review. *Drug Saf.* **2017**, *40*, 1075–1089. [[CrossRef](#)]
18. Ancochea, J.; Izquierdo, J.L.; Medrano, I.H.; Porras, A.; Serrano, M.; Lumbreras, S.; Del Rio-Bermudez, C.; Marchesseau, S.; Salcedo, I.; Zubizarreta, I.; et al. Evidence of gender differences in the diagnosis and management of COVID-19 patients: An analysis of Electronic Health Records using Natural Language Processing and machine learning. *J. Womens Health* **2021**, *30*, 393–404. [[CrossRef](#)]
19. Graziani, D.; Soriano, J.B.; Del Rio-Bermudez, C.; Morena, D.; Díaz, T.; Castillo, M.; Alonso, M.; Ancochea, J.; Lumbreras, S.; Izquierdo, J.L. Characteristics and Prognosis of COVID-19 in Patients with COPD. *J. Clin. Med.* **2020**, *9*, 3259. [[CrossRef](#)]
20. Espinosa-Anke, L.; Tello, J.; Pardo, A.; Medrano, I.; Ureña, A.; Salcedo, I.; Saggion, H. Savana: A Global Information Extraction and Terminology Expansion Framework in the Medical Domain. *Proces. Leng. Nat.* **2016**, *57*, 23–30.
21. Benson, T. *Principles of Health Interoperability HL7 and SNOMED*; Springer: London, UK, 2012.
22. Izquierdo, J.L.; Morena, D.; Gonzalez, Y.; Paredero, J.M.; Perez, B.; Graziani, D.; Gutierrez, M.; Rodriguez, J.M. Clinical Management of COPD in a Real-World Setting. A Big Data Analysis. *Arch. Bronconeumol.* **2021**, *57*, 94–100. [[CrossRef](#)]
23. Al-Ozairi, E.; Jallo, M.K.; Hafidh, K.; Alhajer, D.M.; Ashour, T.; Mahmoud, E.F.N.; Abd ElAal, Z.; Loulou, M. Prevalence of Cardiovascular and Renal Co-morbidities in Patients with Type 2 Diabetes in the Gulf, a Cross-sectional Observational Study. *Diabetes Ther.* **2021**, *12*, 1193–1207. [[CrossRef](#)]

24. Jelinek, H.F.; Osman, W.M.; Khandoker, A.H.; Khalaf, K.; Lee, S.; Almahmeed, W.; Alsafar, H.S. Clinical profiles, comorbidities and complications of type 2 diabetes mellitus in patients from United Arab Emirates. *BMJ Open Diabetes Res. Care* **2017**, *5*, e000427. [[CrossRef](#)]
25. Wittbrodt, E.; Bhalla, N.; Andersson Sundell, K.; Gao, Q.; Dong, L.; Cavender, M.A.; Hunt, P.; Wong, N.D.; Mellstrom, C. Assessment of the high risk and unmet need in patients with CAD and type 2 diabetes (ATHENA): US healthcare resource utilization, cost and burden of illness in the Diabetes Collaborative Registry. *Endocrinol. Diabetes Metab.* **2020**, *3*, e00133. [[CrossRef](#)]
26. Fihn, S.D.; Gardin, J.M.; Abrams, J.; Berra, K.; Blankenship, J.C.; Dallas, A.P.; Douglas, P.S.; Foody, J.M.; Gerber, T.C.; Hinderliter, A.L.; et al. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: A report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *Circulation* **2012**, *126*, e354–e471. [[CrossRef](#)]
27. Arnold, S.V.; Bhatt, D.L.; Barsness, G.W.; Beatty, A.L.; Deedwania, P.C.; Inzucchi, S.E.; Kosiborod, M.; Leiter, L.A.; Lipska, K.J.; Newman, J.D.; et al. Clinical Management of Stable Coronary Artery Disease in Patients with Type 2 Diabetes Mellitus: A Scientific Statement from the American Heart Association. *Circulation* **2020**, *141*, e779–e806. [[CrossRef](#)]
28. Gnesin, F.; Thuesen, A.C.B.; Kahler, L.K.A.; Madsbad, S.; Hemmingsen, B. Metformin monotherapy for adults with type 2 diabetes mellitus. *Cochrane Database Syst. Rev.* **2020**, *6*, CD012906. [[CrossRef](#)]
29. CAPRIE Steering Committee. *A Randomised, Blinded, Trial of Clopidogrel versus Aspirin in Patients at Risk of Ischaemic Events (CAPRIE)*; CAPRIE Steering Committee: Amsterdam, The Netherlands, 1996; pp. 1329–1339.
30. Berger, A. Incidence of Major Adverse Cardiovascular and/or Limb Events among Patients Using Aspirin for Secondary Prevention. *Adv. Cardiovasc. Res.* **2020**, *3*, 158. [[CrossRef](#)]
31. Kengne, A.P. The ADVANCE cardiovascular risk model and current strategies for cardiovascular disease risk evaluation in people with diabetes. *Cardiovasc. J. Afr.* **2013**, *24*, 376–381. [[CrossRef](#)]
32. Okkonen, M.; Havulinna, A.; Ukkola, O.; Huikuri, H.; Pietilä, A.; Koukkunen, H.; Lehto, S.; Mustonen, J.; Ketonen, M.; Airaksinen, J.; et al. Risk factors for major adverse cardiovascular events after the first acute coronary syndrome. *Ann. Med.* **2021**, *53*, 817–823. [[CrossRef](#)]
33. Schwartz, G.G.; Nicholls, S.J.; Toth, P.P.; Sweeney, M.; Halliday, C.; Johansson, J.O.; Wong, N.C.W.; Kulikowski, E.; Kalantar-Zadeh, K.; Ginsberg, H.N.; et al. Relation of insulin treatment for type 2 diabetes to the risk of major adverse cardiovascular events after acute coronary syndrome: An analysis of the BETonMACE randomized clinical trial. *Cardiovasc. Diabetol.* **2021**, *20*, 125. [[CrossRef](#)] [[PubMed](#)]
34. Van der Velde, N.; Huurman, R.; Yamasaki, Y.; Kardys, I.; Galema, T.W.; Budde, R.P.; Zijlstra, F.; Krestin, G.P.; Schinkel, A.F.; Michels, M.; et al. Frequency and Significance of Coronary Artery Disease and Myocardial Bridging in Patients with Hypertrophic Cardiomyopathy. *Am. J. Cardiol.* **2020**, *125*, 1404–1412. [[CrossRef](#)] [[PubMed](#)]
35. Mascia, G.; Crotti, L.; Groppelli, A.; Canepa, M.; Merlo, A.C.; Benenati, S.; Di Donna, P.; Della Bona, R.; Soranna, D.; Zambon, A.; et al. Syncope in hypertrophic cardiomyopathy (part I): An updated systematic review and meta-analysis. *Int. J. Cardiol.* **2022**, *357*, 88–94. [[CrossRef](#)] [[PubMed](#)]
36. Fumagalli, S.; Pieragnoli, P.; Haugaa, K.H.; Potpara, T.S.; Rasero, L.; Ramacciati, N.; Ricciardi, G.; Solimene, F.; Mascia, G.; Mascioli, G.; et al. The influence of age on the psychological profile of patients with cardiac implantable electronic devices: Results from the Italian population in a multicenter study conducted by the European Heart Rhythm Association. *Aging Clin. Exp. Res.* **2019**, *31*, 1219–1226. [[CrossRef](#)]
37. Shahreyar, M.; Mupiddi, V.; Choudhuri, I.; Sra, J.; Tajik, A.J.; Jahangir, A. Implantable cardioverter defibrillators in diabetics: Efficacy and safety in patients at risk of sudden cardiac death. *Expert Rev. Cardiovasc. Ther.* **2015**, *13*, 897–906. [[CrossRef](#)]