

A Trend of Eight-Years Big Data Analytics of Electronic Medical Records to Review and Study Diagnosis and Treatment of Coronary Artery Disease in Different Genders

Maryam Panahiazar (✉ Maryam.Panahiazar@ucsf.edu)

University of California, San Francisco

Andrew M. Bishara

University of California, San Francisco

Yorick Chern

University of California, Berkeley

Roohallah Alizadehsani

Deakin University

Jennifer Perri

University of California, San Francisco

Dexter Hadley

University of Central Florida

Rima Arnaout

University of California, San Francisco

Ramin Beygui



University of California, San Francisco

Research Article

Keywords: Cardiovascular Disease (CVD), Coronary Artery Disease (CAD), data analytics, gender discrepancies

Posted Date: April 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-366930/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Cardiovascular Disease (CVD) and Coronary Artery Disease (CAD) in particular, is one of the leading causes of death, morbidity, and mortality in the United States. Notably, women continue to have worse outcomes than men. The causes of these discrepancies have yet to be fully elucidated. The main objective of this study is to detect gender discrepancies in outcome using data analytics to risk stratify ~ 32,000 patients with CAD of the total 960,129 patients treated at UCSF Medical Center during an eight years. As an implementation of clinical care, this study's long-term goal is to improve precision diagnosis and ultimately management of CAD for both early detection and identification of patients at risk for rapid progression of the disease.

Methods: We designed and implemented a multidimensional framework to trace patients from admission through treatment as a path of events. The time between events for a similar set of paths was calculated. Then the average waiting time for each step of the treatment was calculated for men and women. Finally, we applied statistical analysis to determine differences in time between diagnosis and treatment steps for men and women.

Discussions: There were statistically significant gender-based differences in the common path of diagnosis and treatment of patients with CAD. The average time for women from the first visit to diagnostic Cardiac Catheterization was more than 2 months than for men (358.77 vs. 291.83 days). By contrast, the average time from diagnostic Cardiac Catheterization to treatment Cardiac Catheterization and Coronary Artery Bypass Grafting (CABG) was not significant. Women with CAD requiring revascularization have a significantly longer interval between their first physician encounter indicative of CVD and their first diagnostic cardiac catheterization compared to men. Avoiding the delay in diagnosis and treatment will provide a better outcome for patients at risk.

Introduction

Cardiovascular Disease (CVD) encompasses a broad range of conditions. Coronary Artery Disease (CAD), commonly referred to as Ischemic Heart Disease (IHD), is the leading cause of death, morbidity, and mortality in the United State and globally. Aggarwal et al.¹ in their study on sex differences in CVD suggested that despite advances in treatment and survival, it is still the leading cause of death among women. For example, compared to men, women are less likely to be accurately diagnosed. Several non-traditional health occurrences in women predispose them to CVD including early menopause and menarche, gestational diabetes mellitus, and hypertension. Gender, ethnic, racial, and age discrepancies within CVD diagnosis and treatment exist and have been well reported¹⁻⁴. Nearly half of all African American adults, 47.7 percent of women, and 46.0 percent of men have some form of CVD³. Although the overall guidelines and management of CVD are similar in most of the aspects for both genders, gender-based variations in the pathophysiology, symptomatology, presentation, efficacy of diagnostic tests, and response to pharmacological interventions do exist.

We summarized these differences in one of our previous works on gender based differences in CVD⁵. The etiology for the differences is less well understood. In a study by Ong et al.⁶, the authors suggested that sex

hormones affecting blood pressure could play a major role in disparate development of CVD. Even though there was higher diastolic blood pressure in men, higher systolic pressure was reported in women which is a greater risk factor for CVD. Regardless of etiology it is apparent women have a poor outcome compared with men when it comes to CAD in particular. A major reason may be delay in diagnosis or a different treatment algorithm as compared to their male counterparts.

The main objective of this study is to find these discrepancies using a multidimensional data analytics framework to risk stratify CAD as a subgroup of the CVD patient population at UCSF. We create a cohort selection that allows for simple manipulation and search of the data within the Clinical and Research Data Warehouse (CRDW). This facilitates rapid familiarization and hypothesis testing of the data set. As an example, we hypothesized that there are gender-based discrepancies in the diagnosis and treatment of CAD. With such a large patient database and the infrastructure for data abstraction in place at UCSF Bakar Computational Health Sciences Institute, we have been able to describe these discrepancies. We believe that specific studies for individual patients based on medical record profiles with demographic information are more accurate for improving health outcomes in patients with CAD. Our long-term goals are to translate the multi-dimensional big data that is generated at the University of California System to directly improve and assist clinical care decision making. This translation ultimately would improve outcomes for patients and reduce cost.

In previous work, we created a database as a comprehensive resource for research, comprising 126 papers and 68 datasets relevant to CAD diagnosis, extracted from the scientific literature from 1992 and 2018⁷. We showed significant research outcome on Mayo Clinic patient data for implementing a novel model on survival analysis⁸, recommendation and treatment plan for new patient based on patient similarity⁹, and developing a novel prediction model¹⁰. To date none of our prior research is specifically discussed the gender based differences. A systematic review of gender based studies of diagnosis and treatment of CAD in the last 20 years⁵ shows discrepancies in outcomes of CAD between men and women^{2,3,6,11-22}. However, the causes of these discrepancies have yet to be fully elucidated and require further detailed analysis to design interventions and structures to minimize bias. Studies suggest that knowledge and awareness of bias reduce discrimination and therefore our publication will aid in decreasing physician bias²³. Besides unique situations germane to women such as pregnancy and hormone therapy further make it challenging to diagnose female patients, sometimes young, with CAD in a timely manner.

In this study, we traced potential paths to shed light on the causes of gender discrepancies, and by using path analysis, we uncover delays and possible gender- based differences in diagnosis and treatment of CAD. These differences in healthcare delivery, methodology, diagnosis, procedure, and the time interval between diagnosis procedure and therapeutics may have a significant impact on patients' health and outcome.

Personalized treatment for individuals based on a particular EMR profile may significantly reduce unnecessary treatments and cost, and potentially, morbidity and mortality of downstream procedures associated with incorrect or late diagnosis. Moreover, improved precision may change the debate surrounding the standard guidelines based on gender and other individual characteristics of the patients. Suggesting new guidelines based on patient characteristics will help providers in both detection and management of patients

at risk of rapid progression of CAD and generally in CVD and it will be an innovation in clinical care. The findings of this study will allow us to better identify the systemic causes of discrepancies within CAD treatment and pinpoint the best methods for intervention to reduce them.

In the following sections, we describe the overview of the study design from the hypothesis definition to future work. We explain the study cohort following by data preprocessing, data dictionary, data processing, and data analytics. Next, we show the validation and results. In the last section we discuss the results, study limitations and next steps. And finally, we conclude the study with impact of the study, innovation, perspective with clinical competencies in medical record and competency in patient care.

Materials And Methods

Study Design and Overview

This study is designed around the basic workflow considering several steps including hypothesis definition, study cohort and population, data dictionary creation, data preprocessing, data processing, data analytics, validation and results and finally future steps. Figure 1 illustrates the major components of the study from hypothesis definition to future plan.

We define the existence of discrepancies across different genders in:

- diagnosis and the time of diagnosis.
- procedures including invasive and noninvasive procedures.
- time interval between diagnosis, medication order, and procedure.

Study Cohort and Population

Our data analytics built using EMR data on 960,129 patients admitted to UCSF between July 2011 and December of 2018. This study does not include any human subject and experimental protocol. All data-based De-Identified Clinical Data Warehouse (De-ID CDW) were authorized to access as “de-identified” by the University of California San Francisco and all IDs and metadata (e.g., location) have been removed. All methods were carried out in accordance with relevant guidelines and regulations at UCSF.

De-ID CDW is a de-identified database copy of high-value EHR data. Therefore, this data is not subject to HIPAA restrictions on research use and hence IRB approval or an honest broker intermediary and the need for informed consent was waived by the UCSF Research data team committee. The De-ID CDW system accelerates the research process by permitting UCSF investigators to locate research data and encourage an exploratory approach to hypothesis generation. The De-ID CDW is available to the UCSF research community.

After authorization to access “de-identified” EMR data for research, in consultation with cardiac, thoracic, and vascular surgeons, cardiologists and cardiovascular epidemiologists, the following cohort identification criteria were developed:

- Coronary Artery Disease (CAD), commonly referred to as Ischemic Heart Disease (IHD) based on the ICD10 code (I20–I25).
- Patients with missing value specifically for ICD10 code excluded.
- Patients defined as unknown, and unspecified definition excluded.

To be included in this cohort, patients needed to meet the above criteria, leading to a cohort size of 32,904 CAD patients. Vital such as cholesterol (HDL), cholesterol (LDA), cholesterol (TOTAL), systolic blood pressure, diastolic blood pressure, BMI, age have been considered. Demographic characteristics such as ethnicity have been considered. Smoking conditions including patients never smoked, current every day smoker, former smoker, passive smoke exposure are very important characteristics to be considered. Co-morbidities (e.g. hypertension, liver disease, hyperlipidemia, diabetes, dialysis) for patients with CAD for both genders are calculated. All vitals, characteristics, and co-morbidities are shown in Fig. 2. This data set consisted of de-identified patient ID, demographic information (e.g. gender), and diagnosis based on ICD10 code as shown in Table 1. The details of ICD codes are described in supplementary material Table S1 (ICD10details). For procedure code, we used Current Procedural Terminology (CPT) and date of procedure services for both invasive and non-invasive procedures. For medication, we used the medication code, medication name, and date of the orders.

Table 1

ICD10 I20-I25 for CAD. Table S1 (ICD10codes) in supplementary materials show all details for ICD10 Codes.

ICD10	Definition	Subgroups
I20	Angina pectoris	I20.0, I20.1, I20.8, I20.9
I21	Acute myocardial infarction	I21.0(I21.01,I21.02,I21.09), I21.1(I21.11,I21.19), I21.2(I21.21,I21.29), I21.3, I21.4, I21.9, I21.A(I21.A1,I21.A9)
I22	Subsequent ST elevation and non-ST elevation myocardial infarction	I22.0, I22.1, I22.2, I22.8, I22.9
I23	Certain current complications following ST elevation and non-ST elevation myocardial infarction	I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6, I23.7, I23.8
I24	Other acute ischemic heart diseases	I24.0, I24.1, I24.8, I24.9
I25	Chronic ischemic heart disease	I25.1, I25.2, I25.3, I25.4, I25.5, I25.6, I25.7, I25.8

Data Dictionary

We manually created data dictionary for procedures (e.g. CPT codes include diagnostic cardiac catheterization, treatment cardiac catheterization, cardiac CT scan, echo, EKG, myocardial lab, and stress test). Refer to supplementary material Table S2 (Procedure Dictionary) for the full list including all codes and names for procedures. To create a dictionary for medication, different medications were classified into main classes as anticoagulants, antiplatelet, aspirin, beta-blocker, calcium antagonist, cardiac drug, cardiovascular drug, nitrate, ranolazine, and statin. Refer to supplementary material Table S3 (Medication Dictionary) for the full dictionary including all codes and names for medications.

Data Preprocessing

The patients whose medical history does not include at least one element from the set of CPT codes were eliminated from the initial cohort patient. By doing so, the patient number was reduced to ~ 23,000. Before proceeding, the CPT codes were mapped and translated to procedure names (e.g. EKG, CABG for Coronary Artery Bypass Grafting) based on our dictionary. The medication history data set contains patient ID, medication name, medication code, therapeutic class, pharmaceutical class, pharmaceutical subclass, date of medication ordered and gender. A similar translation was done for medication based on medication dictionary and medications were assigned into main classes.

Data Processing and Statistical Analytics

Our approach was based on time series patients' data. Because of a big and diverse patient cohort at UCSF, we could follow each patient from initial interaction with the UCSF medical system following up any medication order and invasive/noninvasive CAD related procedures over months and years of treatment. For each patient the sequence of events was created from the time of initial presentation to the UCSF medical system to the last invasive procedure as the date of extracting data (e.g. CABG as one of the important targets). We have implemented methods to determine the first suspicion of CAD by providers (primary care and/or cardiologist). We measured the time between different events (e.g. time between prescribing of aspirin or and any other medications and ordering the EKG test, EKG test to CABG) and found the sequence of events for each patient and group of similar patients.

One of the novelties of this study is tracing a multidimensional aspect of patients' treatment over time. It means we look at both medications and procedures over time of treatment. We merged the sequence of medication orders and procedures over time as a time series sequence from the time of admission to the end of treatment as recorded in EMR. Event time was defined as the date of the first event (e.g. prescribing aspirin, ordering stress test) until the date of the next event (e.g. ordering EKG test) and the next event. All medications and procedures from the dictionary can count as the first event in the patient records. We explored all possible existing events (e.g. aspirin = > EKG test = > diagnostic Cardiac Catheterization = > CABG) paths for individual patients. Then we calculated the time interval between every two pairs of events and the number of days. Table 2 shows a few examples of the events. Table S4 (Time Intervals) and Table S5 (All Paths) in supplementary materials show all paths and time intervals for all possible sequences of events. The data set is divided into separate data sets for men and women. For each set, we grouped each row with the same "Path" and compiled the days spanned into a list containing different days from different patients. Upon the completion of the list of days for each different path, the mean, standard deviation, number of patients and essentially the length of the day is calculated for both men and women data set. As the very last step, both the men and women sets are merged, or concatenated, on the same paths. Then, 2 sample t-Tests are performed for each row to evaluate whether the differences between the average delay days for men and women are statistically significant or not. Differences in delay time between groups were assessed with the p-value. Table 3 shows a few examples of the results of data analytics. Table S5 (All Paths) in supplementary materials show data analytic results for all paths and gender-based differences for all patients.

Table 2

Example of the time interval and days between events for each patient. Table S4 (Time Intervals) in supplementary materials shows all paths and time intervals for all possible sequences of events.

deidentified Patient ID	Path of events	Time interval	Days
**1(deID PID)	Aspirin ⇒ EKG	[2011-09-05,2012-01-21]	76
**2(deID PID)	EKG ⇒ diagnostic Cardiac Catheterization	[2012-01-21,2012-04-11]	80
**3(deID PID)	diagnostic Cardiac Catheterization ⇒ CABG	[2012-04-11,2012-04-22]	11

Table 3

Example of gender-based time interval calculation for individual paths. It includes, statistical analysis (average day between events in the path as average days; median as MD; standard deviation as SD, number of patients as #n, and p-value) for patients who went through the path of interest. The full table is in supplementary materials Table S5 (All Paths).

Path	Men				Women				<i>p-value</i>
	average	MD	SD	#n	average	MD	SD	#n	
	days				days				
Aspirin ⇒ EKG	160.97	28.0	305.17	2457	178.03	33.0	317.29	1371	0.106062
EKG ⇒ diagnostic Cardiac Catheterization	304.70	51.0	471.16	2010	368.29	105.0	496.86	1033	0.000682
diagnostic Cardiac Catheterization ⇒ CABG	77.06	15.0	231.72	237	127.18	17.0	329.98	64	0.257025

Results

Upon possessing the data, the next step is to search for the evidence that there is a delay in definitive diagnosis and treatment of CAD in women compared to men. The first step to validate this hypothesis is to determine the first point encounter with a physician when a patient was suspected of having cardiovascular disease. We included the treatment path with suspicion of potential cardiovascular disease that combined both procedures and medications. Initially, our experts determined that aspirin is one of the drugs that is frequently ordered early upon the encountering a patient at risk of cardiovascular disease. Thus, as a first analytic step we calculated the time it took between the first time aspirin (other medications has been considered too) prescribed to the first diagnostic Cardiac Catheterization that occurred, and then from diagnostic Cardiac Catheterization to treatment procedures such as percutaneous coronary intervention (treatment Cardiac Catheterization) and coronary artery bypass graft (CABG). In medication data, 40 medication codes were considered as aspirin including 2 groups of therapeutic classes defined as analgesics and antiplatelet, which includes groups of medication pharmaceutical classes including analgesic antipyretics, salicylates, analgesics, salicylate and non-salicylate comb, bulk chemicals, and platelet aggregation inhibitors. These medications are under the medication pharmaceutical sub classes defined as

salicylate analgesics, salicylate analgesics with non-salicylate analgesics combinations, and salicylate analgesics buffered. Our dictionary for complete information about aspirin and classifications is in supplementary material Table S6 (aspirin). Table 4 shows data analytics for all different paths with aspirin as a starting point.

Table 4

Example of gender-based data analytics for all paths with aspirin as a starting Point. It includes 2 sample t-Tests to compare the average delay in men and women for each path of interest. A complete table is in supplementary materials Table S6 (Aspirin).

Path	Men			Women			<i>P-value</i>
	avg days	SD	#n	avg days	SD	#n	
aspirin ⇒ anticoagulants	113.56	300.95	2437	123.84	307.93	1294	0.328073423
aspirin ⇒ antiplatelet	108.51	263.92	1056	133.02	330.11	434	0.169014934
aspirin ⇒ beta-blockers	88.37	237.42	2492	105.97	258.35	1268	0.042529161
aspirin ⇒ calcium antagonist	231.83	412.29	328	236.84	413.31	230	0.887902392
aspirin ⇒ cardiac drugs	180.69	364.89	1944	193.70	373.32	1078	0.355236041
aspirin ⇒ cardiovascular drugs	106.42	257.81	2463	138.62	304.98	1255	0.001382782
aspirin ⇒ EKG	160.97	305.17	2457	178.03	317.29	1371	0.106061569
aspirin ⇒ nitrate	178.89	346.62	1378	179.45	338.49	757	0.971348372
aspirin ⇒ ranolazine	238.98	396.04	70	387.89	490.56	37	0.116452062
aspirin ⇒ statin	99.68	247.01	2549	115.03	264.83	1276	0.083961313
aspirin ⇒ CABG	276.96	418.96	94	358.07	508.62	28	0.446247762
aspirin ⇒ diagnostic Cardiac Catheterization	300.71	442.23	509	347.77	472.48	252	0.187327501
aspirin ⇒ treatment Cardiac Catheterization	267.83	427.68	212	347.72	516.78	69	0.248297118

Of patients who ultimately underwent a therapeutic intervention – ie CABG or PCI – there was a greater delay in time to diagnostic Cardiac Catheterization for women. Performing 2-sample t-Test on time to diagnostic Catheterization between men and women showed statistical significance, meaning that there is indeed a delay in women who eventually undergo therapeutic procedures for CAD to get diagnostic Cardiac

Catheterization. Suspecting that the reason for reaching insignificant statistical results is that the number of patients who had an order of aspirin as a first encounter recorded was not large enough to reach statistical significance. Since aspirin is an over-the-counter medication, there is a high possibility of aspirin not being recorded as a medication order in EMR. We decided to look at the time between the other first cardiovascular medication order and the first procedure recorded in EMR. This time, we are not limiting the starting point to aspirin. Any types of medication that belongs to the classes of cardiovascular and cardiac drugs, anticoagulants, anti-platelets, aspirin, beta-blockers, and statin are included as the starting point medication for patient suspected to be at risk for cardiovascular disease. With this new starting point the number of patients (both men and women) increased. The path of the all medications to diagnostic Cardiac Catheterization reached statistical significance, showing that the first medication prescribed to the first diagnostic Cardiac Catheterization ordered is significantly delayed in women compared with men who eventually end up undergoing treatment Cardiac Catheterization or CABG as shown in Table 5.

Table 5

Example of gender-based data analytics for all cardiovascular related medications as starting point of the treatment plan. The medications are belong to the classes of cardiovascular drugs, cardiac drugs, anticoagulants, anti-platelets, aspirin, beta-blockers, and statin. The complete table is in supplementary materials Table S5 (All Paths).

Path	Men		Women		<i>P-value</i>
	average interval days	# patient	average interval days	#patient	
all Cardiovascular Medications ⇒ CABG	414.25	157	436.28	47	0.8096
all Cardiovascular Medications ⇒ diagnostic Cardiac Catheterization	395.67	884	457.76	444	0.0482
all Cardiovascular Medications ⇒ treatment Cardiac Catheterization	419.69	298	539.53	95	0.0920

In the next step we considered procedures as the starting point plus medication to find the path between the very first event (any kind of related medications and procedures, because for some patients the first event in noninvasive procedure and not a medication order) for the patients at the time of admission and next steps such as diagnostic Cardiac Catheterization and treatment Cardiac Catheterization. As shown in Table 6, there is a significant time difference from first event to diagnostic Cardiac Catheterization between genders (p -value = 0.000119), while the p -value for diagnostic Cardiac Catheterization to CABG is not statistically significant. This result is a validation of the hypothesis that there are discrepancies within cardiovascular diagnosis and treatment. It shows that there is a delay to diagnostic Cardiac Catheterization in women who eventually undergo treatment Cardiac Catheterization. With clear results that show the discrepancies in diagnostic procedures in women vs. men, in the next section, we discuss the possible implications of this study on patient care.

Table 6

Statistical analysis. It shows the statistical analysis of very first event to diagnostic Cardiac Catheterization and from diagnostic Cardiac Catheterization to the treatment Cardiac Catheterization and CABG.

Path	Men			Women			<i>P-value</i>
	average	SD	#n	average	SD	#n	
	days			days			
first event \Rightarrow diagnostic Cardiac Catheterization	291.83	479.14	2532	358.77	514.28	1255	0.000119
diagnostic Cardiac Catheterization \Rightarrow treatment Cardiac Catheterization	108.77	309.07	481	89.59	285.52	160	0.471496
diagnostic Cardiac Catheterization \Rightarrow CABG	77.06	231.72	237	127.18	329.98	64	0.257025

Discussion

In this study, we explored the use of data analytics to reveal the gender-based discrepancies in the diagnosis and treatment of CAD. We hope by recognizing a clear delay in diagnosis (i.e. time to diagnostic catheterization in women) will make a change in practice and will result in improved outcomes for women with CAD with early detection. We have implemented methods to determine the first suspicion of CAD by providers (primary care and/or cardiologist). We measured the time interval between different events (e.g. time between prescribing a medication and ordering the cardiac stress test) and found the sequence of events for each patient and group of similar patients. We used statistical analyses to find the differences between women and men. Our results, based on the analysis of a subset of patients with CAD condition, support the hypothesis that there exist discrepancies in the diagnosis and treatment of CAD based on patient demographic characteristics such as gender.

As the first step, we use landmarks (e.g. aspirin initiation) as a trigger for early suspicion of CAD and follow up with other markers and medications (e.g. beta-blocker, statins) and we followed that by noninvasive and invasive cardiac procedures. As the next step we used all identifiable cardiovascular-related medications as a starting point instead of just aspirin to expand the patient cohort and found significant discrepancies. Finally we changed the starting point to include both medications and procedures.

We discovered that when women with the eventual diagnosis of severe CAD are started on aspirin it takes them longer to start beta-blockers, as a known drug to reduce cardiovascular risk, compared to men. Our analysis shows that women who have undergone CABG on the average have waited for 358 to get the “Gold Standard” diagnostic Cardiac Catheterization followed by an extra 127 days to undergo CABG for severe CAD. Men who have undergone CABG on average waited for 291 to get the “Gold Standard” diagnostic Cardiac Catheterization followed by extra 77 days to undergo CABG for severe CAD. From a starting point of any first event (e.g. aspirin order, cardiac stress test order), on average it takes over two months longer for women to undergo CABG compared to men. In the patients with left main and multi-vessel CAD or unstable angina, the risk of a CAD event is high. For example, if 50 percent are at risk of some event in 6 months (ACS,

STEMI, NSTEMI, or sudden cardiac death), then it can be extrapolated that a delay of 2 months would result in a 17 percent increased risk for women compared to men. Our goal was to simplify hypothesis testing as much as possible for the healthcare providers and researchers. We showed that the kind of data analytics, which has been used in this study, is sufficient to find the discrepancies within cardiovascular diagnosis and treatment. While our work focused on the UCSF data, we anticipate that our approach can be applied to other databases of patient data with similar levels of success (e.g. UC System-wide data). Based on our analysis, the difference in the interval from the first event to diagnostic Cardiac Catheterization is the intervention with a significant p-value (p-value = 0.000119), while the p-value for diagnostic Cardiac Catheterization to CABG and diagnostic Cardiac Catheterization to treatment Cardiac Catheterization is not statistically significant.

In summary our research has important implications for initiatives aimed for improving the use of EMR to find the possible reasons for a different outcome in women versus men or based on differences in other patient characteristics. Several efforts are devoted to finding the different outcome and risk factors in different genders - but the reasons for the differences are not yet fully identified.

Our work is not without limitations. First, a key limitation in this study is the lack of reliable medication history before admission to UCSF. Because of this limitation, there is a possibility of not capturing some of the medications that patients had been taking over the years prior to first admission to UCSF. We are planning to overcome this limitation by considering unstructured clinical notes in our future study. With that, we will have access to the history of patients before 2011, which is the starting point of data collection in our study cohort. Because UCSF Medical Center is a tertiary referral center for CABG, some patients are admitted when they have already had a diagnostic Cardiac Catheterization. For this group of patients who arrived from other institution, sometimes the code for diagnostic Catheterization is not entered. As a result of this limitation, we have a decrease in the number of patients with the path from diagnostic Cardiac Catheterization to CABG. Although the number of patients with CABG is 752, a subset of the patients have the starting point of CAD on admission to UCSF before undergoing therapeutic procedures at UCSF (CABG or therapeutic Catheterization).

We are planning to find a patient profile that describes rapidly progressive CAD and flag these patients for frequent and regular cardiovascular assessment. We will develop interactive visualization tools for providers, payers, and researchers to assist the personalized treatment plans for individual patients with specific characteristics based on new guidelines and suggestions as EMR order sets. Our long-term goal is to translate the multi-dimensional big data including EMR that is generated at the University of California System, to directly improve and assist clinical care decision making that ultimately would improve outcomes for patients and reduce cost. Moreover, this study lays the foundation to develop novel translational interventions through powerful big data-driven analytics that leverage the wide availability of UC System patient data.

As an implementation of clinical care, this study's goal is to improve precision diagnosis and ultimately, management of CVD for both early detection and identification of patients at risk for rapid progression of the disease. As a clinical care outcome, we will provide the protocol in an EMR order sets format for early detection of severe CAD in patients at risk for rapid progression. As an example, for a woman with a history of hormone therapy, pregnancy with hypertension in early age, family history, and increased BMI, we can expedite the more sensitive testing (stratified and varied order sets depending on that patient's risk profile)

instead of long-term therapy with medications (e.g. aspirin, statins, beta-blockers) and diagnose the CAD expeditiously. As an assistant tools for providers, payers, and researchers, we are planning to deliver Interactive visualization Tools, EMR order sets, and recommendation system to access data to search and reuse and guidelines for the treatment for individual patients with specific characteristics. The outcome of this research lays the foundation to develop novel translational interventions through powerful big data-driven analytics that leverage the wide availability of UC System patient data.

Conclusion

Although the overall guidelines and management of CAD are similar for both genders, gender-based variations in the pathophysiology, symptomatology, presentation, efficacy of diagnostic tests, and response to pharmacological interventions do exist. When features and predictive variables are different in men and women, decision making based on the unified platforms and guidelines for diagnosis and treatment of the patients appears to lead to the poor outcomes in women in comparison with men. Therefore studies on CAD based on individual characteristics (e.g. demographics) will have a big impact on the diagnosis and treatment of CAD.

There are discrepancies in the delivery of healthcare in general across different genders. Women with severe CAD requiring revascularization have a significantly longer interval between their first physician encounter indicative of cardiovascular disease to their first diagnostic cardiac catheterization compared to men. These differences in healthcare delivery, methodology, diagnosis procedure, the time interval between diagnosis procedure and therapeutics may have a significant impact on patients' health and outcome.

Personalized treatment for individuals based on specific EMR profile and demographic characteristics may significantly reduce unnecessary treatments and costs, and potentially, morbidity and mortality of downstream procedures associated with wrong or late diagnosis. Moreover, improved precision may change the debate surrounding the standard guidelines based on gender and other individual characteristics of the patients. Developing updated gender based guidelines will help provider for both early detection and manage individual patients at risk of rapid progression of CAD and generally in CVD will be an innovation in clinical care.

Declarations

Contributors

M.P., R.B., D.H., A.B. designed the research studies. M.P., R.B. A.B. defined and select data for study cohort. M.P., R.B., A.B., Y.C., D.H implemented big data analytics platform. M.P., R.B., A.B., D.H., J.P. discussed and analyzed the results. M.P., R.B., Y.C., A.B. wrote the manuscript. R.A., M.P. provided the systematic review of related works. M.P., R.B., J.P., R.A. edited and revised the manuscript.

Declaration of interests

All other authors declare no competing interests.

Abbreviations

Cardiovascular Disease (CVD), Coronary Artery Disease (CAD), Ischemic Heart Disease (IHD), Precision Cardiovascular Medicine (PCM), Electronic Medical Record (EMR), Clinical and Research Data Warehouse (CRDW), Current Procedural Terminology (CPT), Research Data Browser (RDB), Coronary Artery Bypass Graft (CABG)

References

1. Aggarwal, N. R. *et al.* Sex Differences in Ischemic Heart Disease. *Circulation: Cardiovascular Quality and Outcomes*. **11** (2), e004437 (2018).
2. ReinEMR, T. *et al.* Cardiovascular risk factors in overweight German children and adolescents: Relation to gender, age and degree of overweight. *Nutrition, Metabolism and Cardiovascular Diseases*. **15** (3), 181–187 (2005).
3. Johnston, A., Mesana, T. G., Lee, D. S., Eddeen, A. B. & Sun, L. Y. Sex Differences in Long-Term Survival After Major Cardiac Surgery: A Population-Based Cohort Study. *Journal of the American Heart Association*. **8** (17), e013260 (2019).
4. Zhao, M. *et al.* Sex differences in risk factor management of coronary heart disease across three regions. *Heart*. **103** (20), 1587 (2017).
5. Panahiazar, M., Alizadehsani, R., Bishara, A. M. & Chern, Y. D. H. Systematic Review of Gender Based Studies of Diagnosis and Treatment of Cardiovascular Disease in Last 20 Years. *Adv Card Res*. **2** (4), 192–194 (2019).
6. Ong, K. L., Tso, A. W. K., Lam, K. S. L. & Cheung, B. M. Y. Gender Difference in Blood Pressure Control and Cardiovascular Risk Factors in Americans With Diagnosed Hypertension. *Hypertension*. **51** (4), 1142–1148 (2008).
7. Alizadehsani, R. *et al.* A database for using machine learning and data mining techniques for coronary artery disease diagnosis. *Scientific Data*. **6** (1), 227 (2019).
8. Panahiazar, M., TaslimitEMRani, V., Pereira, N. & Pathak, J. Using EMRs and Machine Learning for Heart Failure Survival Analysis. *Stud Health Technol Inform*. **216**, 40–44 (2015).
9. Panahiazar, M., TaslimitEMRani, V., Pereira, N. L. & Pathak, J. Using EMRs for Heart Failure Therapy Recommendation Using Multidimensional Patient Similarity Analytics. *Stud Health Technol Inform*. **210**, 369–373 (2015).
10. TaslimitEMRani, V., Dong, G., Pereira, N. L., Panahiazar, M. & Pathak, J. Developing EMR-driven heart failure risk prediction models using CPXR(Log) with the probabilistic loss function. *Journal of Biomedical Informatics*. **60**, 260–269 (2016).
11. Maas, A. H. E. M. & Appelman, Y. E. A. Gender differences in coronary heart disease. *Netherlands Heart Journal*. **18** (12), 598–603 (2010).

12. Penno, G. *et al.* Gender differences in cardiovascular disease risk factors, treatments and complications in patients with type 2 diabetes: the RIACE Italian multicentre study. *Journal of Internal Medicine*. **274** (2), 176–191 (2013).
13. Virani, S. S. *et al.* Gender Disparities in Evidence-Based Statin Therapy in Patients With Cardiovascular Disease. *The American Journal of Cardiology*. **115** (1), 21–26 (2015).
14. Appelman, Y., van Rijn, B. B., ten Haaf, M. E., Boersma, E. & Peters, S. A. E. Sex differences in cardiovascular risk factors and disease prevention. *Atherosclerosis*. **241** (1), 211–218 (2015).
15. Khamis, R. Y., Ammari, T. & Mikhail, G. W. Gender differences in coronary heart disease. *Heart*. **102** (14), 1142 (2016).
16. Regitz-Zagrosek, V., Lehmkuhl, E. & Mahmoodzadeh, S. Gender Aspects of the Role of the Metabolic Syndrome as a Risk Factor for Cardiovascular Disease. *Gend. Med.* **4**, S162–S77 (2007).
17. Hollier, L. M. J. J. *et al.* ACOG practice bulletin no. 212: pregnancy and heart disease. *Obstetrics and gynecology*. **133** (5), e320–e56 (2019).
18. Kameneva, M. V., Watach, M. J. & Borovetz, H. S. Gender difference in rheologic properties of blood and risk of cardiovascular diseases. *Clinical Hemorheology and Microcirculation*. **21**, 357–363 (1999).
19. van Roeters, J. E., Westerveld, H. T., Erkelens, D. W. & van der Wall, E. E. Risk factors for coronary heart disease: implications of gender. *Cardiovascular. Res.* **53** (3), 538–549 (2002).
20. Argulian, E. *et al.* Gender Differences in Short-Term Cardiovascular Outcomes After Percutaneous Coronary Interventions. *The American Journal of Cardiology*. **98** (1), 48–53 (2006).
21. Welch, B. M., Kawamoto, K., Drohan, B. & Hughes, K. S. Chapter 14 - Clinical Decision Support for Personalized Medicine. In: Greenes RA, ed. *Clinical Decision Support* (Second Edition). Oxford: Academic Press; 2014: 383–413.
22. Mercuro, G. *et al.* Gender determinants of cardiovascular risk factors and diseases. *Journal of Cardiovascular Medicine* 2010; **11**(3).
23. Pope, D. G., Price, J. & Wolfers, J. *Awareness reduces racial bias* (National Bureau of Economic Research, 2013).

Figures

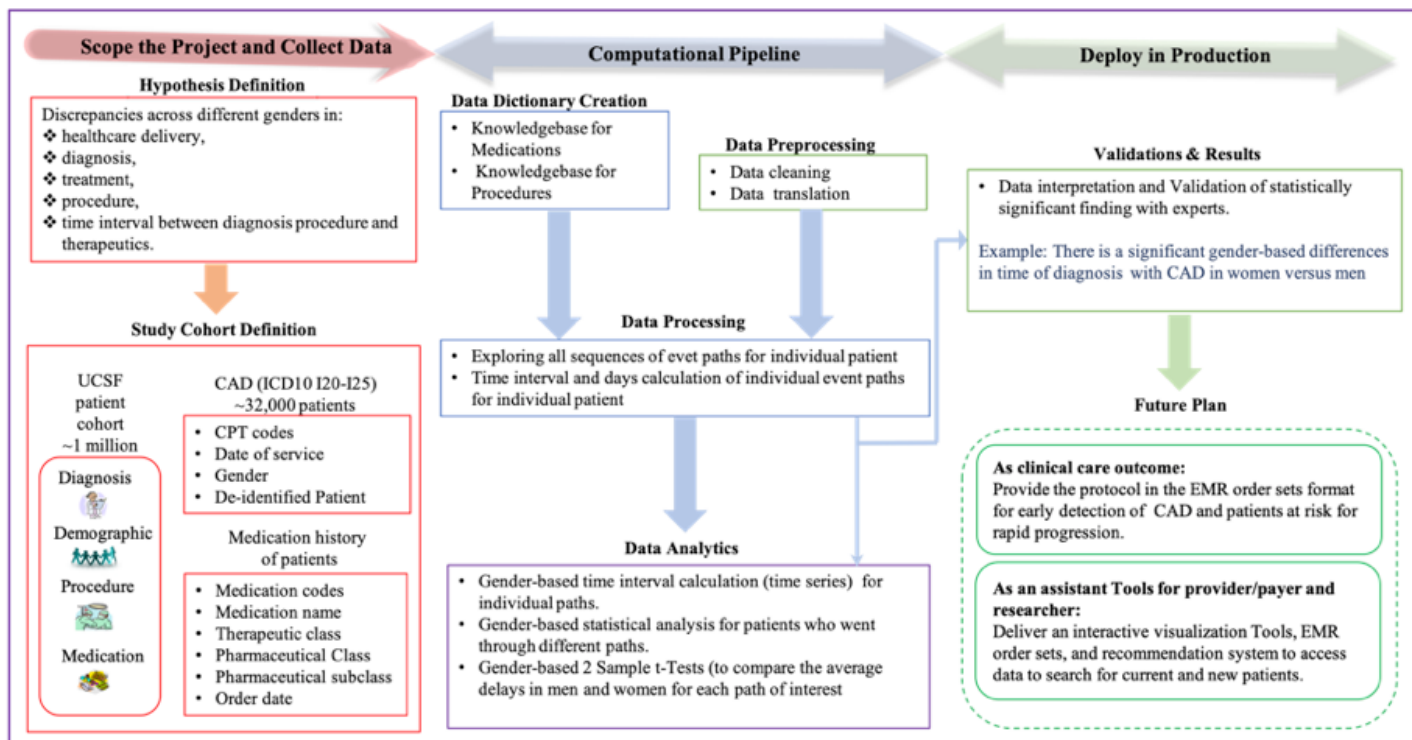


Figure 1

Study Overview and Architecture. This figure illustrates the major components of the study from hypothesis definition to the results and future plan.

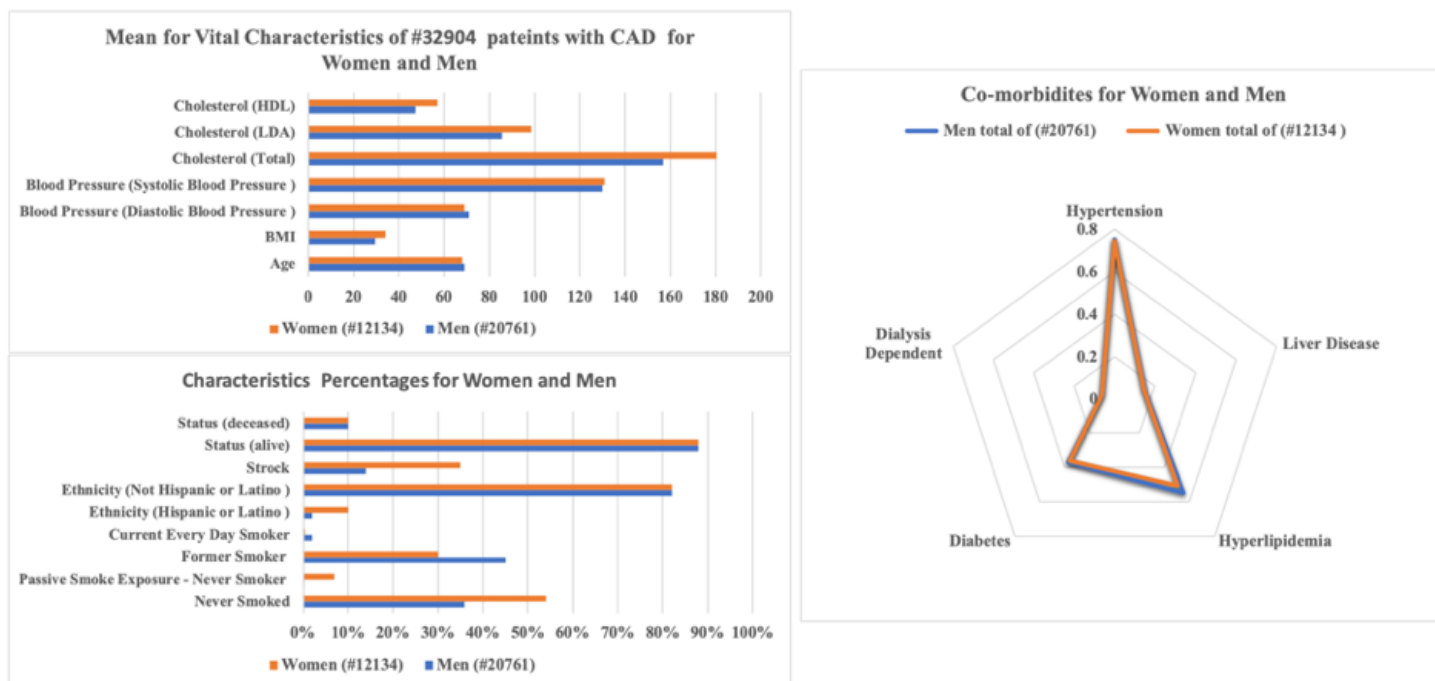


Figure 2

Vital, Demographic Characteristics and Co-Morbidities for Patients with CAD. It illustrates the mean for vital characteristics (e.g. Cholesterols, Blood Pressure, BMI, Age), Characteristics percentage for both gender (e.g. Status, Ethnicity, Smoking Status), and Co-morbidities (e.g. Hypertension, Dialysis, Diabetes, Hyperlipidemia)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1ICD10codes.docx](#)
- [TableS2CPTCodeDictionary.csv](#)
- [TableS3MedicationDictionary.csv](#)
- [TableS4TimeIntervals.csv](#)
- [TableS5AllPaths.csv](#)
- [TableS6Aspirin.csv](#)