Exploring Gender-Specific Symptoms in Coronary Heart Disease Diagnosis

Connor C.J. Hryhoruk, Carson K. Leung ✉, Aila Nik, Kelly Villamayor
University of Manitoba, Winnipeg, MB, Canada
✉ Carson.Leung@UManitoba.ca

Abstract. Coronary heart disease (CHD), characterized by reduced blood flow to the heart due to the narrowing of coronary arteries, remains the leading cause of death globally. Despite advancements in early detection, disparities in the recognition and treatment outcomes of CHD persist worldwide. Previous research has shown that women with CHD are less likely than men to receive timely and accurate diagnoses, particularly when presenting critical risk factors and symptoms. While initiatives to raise awareness of these disparities have been implemented, it remains unclear whether current clinical practices have adapted to incorporate this knowledge. In this paper, we present a big data analytic solution for analyzing previously identified markers of CHD in women and assessing the association between gender-specific symptoms and diagnosis outcomes. The solution makes good use of data mining techniques in network-enabled health informatics and biomedicine domains. The implications of these findings are critical for evaluating whether healthcare practices have begun to address the nuances in CHD manifestation across genders.

Keywords: Health informatics, Biomedicine, Data mining, Data analytics, Coronary heart disease (CHD), Gender analysis

1 Introduction

With advances in technology, big data are everywhere. Embedded in these big data is valuable information and knowledge. Data science—which makes good use of data management [1], data mining [2-7], and machine learning (ML) [8] techniques—analyzes big data and discovers knowledge and useful information for various real-world application areas such as bioinformatics [9], financial analysis [10-13], healthcare informatics and biomedicine [14-16], music analytics [17], social network analysis [18-25], and transportation analytics [26]. In this paper, we focus on a fusion of *health informatics* and *gender analysis*. More specifically, we present a big data analytic solution for analyzing previously identified markers of *coronary heart disease* (*CHD*) in women and assessing the association between gender-specific symptoms and diagnosis outcomes.

CHD [27] is a condition characterized as the narrowing or blockage of coronary arteries, primarily caused by atherosclerosis, reducing blood flow to the heart. Though studied since the 15th century early diagnosis was difficult due to its often "silent" progression. Despite advances in diagnostics by the mid-20th century, ischemic heart disease—including CHD—remains the leading global cause of death, with about 9.1 million deaths in 2021, according to the World Health Organization (WHO)[1].

[1] https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

While CHD is widely recognized as a global health issue, significant gender disparities in diagnosis and outcomes remain. Despite advances in cardiovascular research, women are still under-diagnosed and under-treated compared to men, particularly when presenting with atypical symptoms (e.g., fatigue, nausea). As Wenger [28] noted, CHD has long been viewed as a "man's disease," with diagnostic tools and clinical trials largely focused on male populations. This bias contributes to delayed diagnoses and inadequate treatment for women, especially those with non-obstructive coronary artery disease or microvascular dysfunction. Hiremath et al. [29] further found that women undergoing non-primary percutaneous coronary intervention (PCI) experienced worse outcomes and faced treatment disparities. These findings underscore the need to address both biological differences and sociocultural biases in clinical care to improve CHD outcomes for women.

Key contributions of this paper include our big data analytics solution for analyzing previously identified markers of CHD in women and assessing the association between gender-specific symptoms and diagnosis outcomes. The solution makes good use of data mining techniques in network-enabled health informatics and biomedicine domains.

The remainder of this paper is organized as follows. The next section provides background information and discusses related works. Section 3 describes our big data analytics solution. Sections 4 and 5 show evaluation results and draw conclusions.

2 Background and Related Work

Gender disparities in CHD diagnosis stem largely from differences in symptom presentation between men and women. While men often display "classic" symptoms like chest pain, women frequently experience atypical symptoms—shortness of breath, nausea, fatigue, or jaw/back pain—that are often misattributed to non-cardiac causes, resulting in misdiagnoses or delays.

Maas et al. [30] noted that many women with CHD exhibit ischemia with non-obstructive coronary arteries (INOCA), involving microvascular dysfunction or vasospasms rather than visible blockages. These subtler forms of CHD—more common in women—are harder to detect with standard tools developed for obstructive coronary artery disease (CAD), typically seen in men. Solola et al. [31] emphasized that microvascular dysfunction is especially prevalent among women and often escapes detection with tests designed for large-vessel disease. Moreover, women's symptom descriptions often emphasize emotional or diffuse discomfort, which clinicians may misinterpret. de Marvao et al. [32] further observed that the absence of chest pain in many women reinforces diagnostic bias, with symptoms like fatigue or anxiety often overlooked. This underrepresentation in research has led to diagnostic models centered on male symptom profiles, perpetuating the misconception that CHD is a "man's disease."

In terms of *limitations of current diagnostic practices* for CHD, standard diagnostic tools—such as electrocardiograms (ECGs), stress tests, and angiograms—were largely developed based on male physiology and symptoms, limiting their effectiveness for women. For example, women are more prone to false positives on stress ECGs due to physiological differences. While newer imaging methods like computed tomography (CT) angiography and cardiac magnetic resonance imaging

(MRI) offer improvements, they still face challenges: breast tissue and heart rate variability can reduce CT accuracy, and MRIs may be impractical for some women.

Beyond technology, systemic clinical bias compounds the issue. Atypical symptoms in women are often dismissed, and a male-centered research legacy continues to shape diagnostic approaches. These factors collectively delay diagnosis and worsen outcomes for women with CHD.

3 Our Big Data Analytics Solution

Recall from previous sections that CHD remains the leading cause of death among women of all ages, with high morbidity largely stemming from a lack of understanding of gender-specific pathophysiology, resulting in inequitable treatment and inadequate diagnostics. Prior research has often focused on male-centric "classic" symptoms or highlighted gender disparities without fully examining their collective impact on diagnosis rates. Despite awareness efforts, it remains unclear whether clinical practices have improved in addressing these issues.

Hence, in this paper, we aim to assess whether advancements in CHD diagnosis have effectively reduced gender-specific disparities, particularly for women presenting with high-risk or "female-specific" symptoms. In particular, we design and develop a big data analytics solution to analyze gender-specific markers of CHD in women and assess how female-associated symptoms relate to diagnosis outcomes. Our solution uses pattern discovery to evaluate the association between symptoms, risk factors, and CHD diagnosis. The goal is to determine whether these gender-specific variables contribute to higher—or still lower—diagnosis rates among women.

Recognizing that predictive performance depends on data quality and feature selection, we begin with a thorough dataset assessment to uncover biases, confounders, and underlying relationships. This includes descriptive statistics and pattern mining to examine age and gender distributions of CHD risk factors and symptoms. More specifically, our big data analytics solution consists of two key phases:

1. The first key phase focuses on exploratory analysis, including (a) demographic breakdowns and (b) identification of major risk factors (e.g., hypertension, diabetes, smoking, alcohol use, cholesterol, chronic inflammation).

2. The second key phase uses (a) frequent pattern mining to uncover common symptom and behavior patterns, and (b) association rule mining to identify relationships among factors like symptom profiles, medication use, and lifestyle. This helps assess whether women with high-risk, female-associated symptoms continue to face underdiagnosis relative to men.

4 Evaluation

To evaluate our big data analytics solution, we applied it to a real-world open dataset from the 2015-2016 National Health and Nutrition Examination Survey (NHANES)[2], an annual program of interviews and physical examinations conducted by the National Center for Health Statistics (NCHS). This cycle was chosen for its

[2] https://wwwn.cdc.gov/nchs/nhanes/continuousnhanes/overviewquex.aspx?BeginYear=2015

comprehensive cardiovascular health assessment, making it highly relevant to our study. Each participant was assigned a unique anonymized identifier to protect privacy while allowing for individual-level analysis. The final dataset comprises 9971 records and 274 variables spanning demographics, clinical measures (e.g., blood pressure, cholesterol), medical history, and cardiovascular assessments.

Phase 1(a)—Demographic Breakdown. Our exploratory analysis of the 2015–2016 NHANES dataset reveals a nearly equal representation of genders and age groups, with women comprising 51% (n=5079) and men 49% (n=4892) of participants. The age distribution is similarly balanced across genders (Fig. 1a).

However, CHD prevalence differs markedly by gender. Of the 244 participants diagnosed with CHD, 35% (n=85) were women and 65% (n=159) were men. Among 5693 respondents reporting CHD status, 2.79% of men and only 1.49% of women had CHD—indicating that men had more than double the reported CHD rate. Notably, CHD incidence for both genders increases sharply after age 50, with a pronounced rise in women beginning around age 55, aligning with the average onset of menopause. This pattern supports evidence [33] linking reduced estrogen levels to elevated CHD risk in women (Fig. 1b). Moreover, despite the equitable distribution of gender and age within the dataset, the prevalence of CHD varies significantly between these categories. Out of 244 participants diagnosed with CHD, 35% (n=85) were women and 65% (n=159) were men. Among those who reported either an absence or presence of CHD (n=5693), 1.49% of women and 2.79% of men have CHD. This disparity highlights that male participants exhibit more than double the rate of CHD compared to their female counterparts. Nevertheless, the incidence of CHD diagnoses in both men and women begins to increase significantly from approximately 50-years of age, peaking at 80 years. The frequency of CHD in women starts to rise notably after the 55-year mark, coinciding with the average onset of menopause. See Fig. 1(b). This trend suggests a link between reduced estrogen levels and increased CHD risk in women [33].

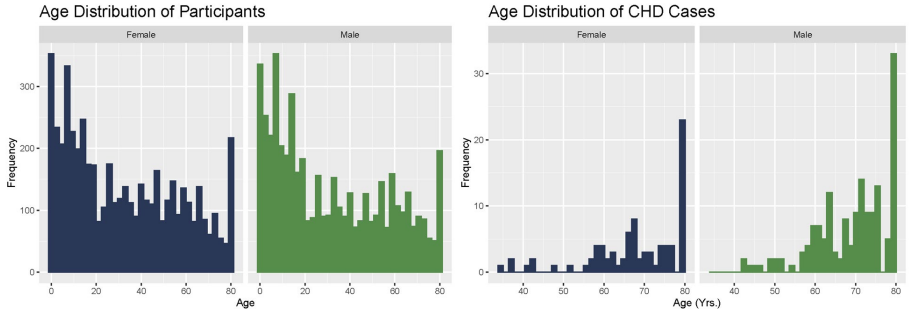

Fig. 1. (a) Both men and women show approximately the same age distribution, with minor differences in frequency. (b) Both men and women with CHD show approximately the same age distribution, again, with differences in frequency. An unforeseen fluctuation of female CHD cases can be seen at the 40-year-mark.

Previous studies [33] have shown that clinical manifestations of CHD typically appear earlier in men than in women, with rates converging later in life, particularly after menopause. This trend is also reflected in our dataset. Accordingly, we expect women with CHD to be generally older and to show a stronger presence of cardiovascular risk factors. However, our analysis reveals an unexpected rise in CHD

cases among women around age 40, which deviates from this anticipated pattern. This anomaly warrants further investigation, and we will keep it in consideration throughout the remainder of our analysis to explore whether additional factors might account for this early increase.

Phase 1(b)—Identification of Major Risk Factors. Although men and women share many CHD risk factors, certain conditions are stronger predictors for women. We explore the following key risk factors identified by Shah et al. [33], analyzing their impact on CHD risk alongside their age and gender distribution:

- *Hypertension* is more prevalent among male CHD patients—25% compared to 18% in women—suggesting it may play a more significant role in men's CHD risk profiles.
- *Diabetes* mellitus appears in 62% of women and 61% of men with CHD, confirming its critical role as a high-impact risk factor for both genders.
- Substance intake shows gender differences. *Smoking* rates are similar overall: 14% of women smoke 1–10 cigarettes per day (cf. 9% of men), while heavy smoking is less common in women (4%) than in men (8%). This suggests that heavy smoking poses a notable risk regardless of gender. *Alcohol intake*, however, shows a sharper gender disparity: 24% of female CHD patients report mild drinking (cf. 35% of men), and 1% of women report severe intake (cf. 9% of men) indicating alcohol may contribute more substantially to CHD risk in men.
- *Obesity* is another major factor, with 56% of women and 51% of men with CHD being categorized as overweight or obese, underscoring its importance in predictive modeling.
- Cholesterol abnormalities—total cholesterol, high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides—contribute to arterial plaque buildup, increasing the risk of atherosclerosis-related conditions such as coronary artery disease. We use ranges defined by the Cleveland Clinic[3, 4] to assess blood lipid levels (see Table 1), as these values help identify when cholesterol concentrations become harmful and elevate CHD risk. Analysis reveals notable gender differences. A smaller proportion of women (55%) maintain healthy *total cholesterol* levels (cf. 76% of men). This disparity extends to *LDL cholesterol*, where 24% of women fall within the moderate-to-severe risk range (cf. 8% of men)—highlighting a worse lipid profile among women. In contrast, *HDL cholesterol* levels are healthier in women, with 35% meeting the recommended range (cf. 14% of men). *Triglyceride* levels show smaller differences: 36% of women and 30% of men have healthy levels, and moderate-to-severe levels are relatively rare in both groups. These findings suggest that, while triglycerides are still relevant, other risk factors (e.g., LDL cholesterol, diabetes, hypertension) may carry more weight in predictive modeling (see Fig. 2a).
- *Chronic inflammatory diseases* (*CIDs*) exhibit a pronounced gender gap. Women consistently report higher rates of CIDs—except in the cases of cancer

[3] https://my.clevelandclinic.org/health/articles/11920-cholesterol-numbers-what-do-they-mean
[4] https://my.clevelandclinic.org/health/articles/11117-triglycerides

and gout—suggesting that inflammation-related conditions may be a more significant CHD risk factor in women than in men (see Fig. 2b).

Together, these analyses provide insights into how gender-specific expression of risk factors influences CHD development and will inform our predictive modeling approach.

Table 1. Classification of cholesterol levels

	Total cholesterol level (mg/dL)	LDL (mg/dL)	HDL for female (mg/dL)	HDL for male (mg/dL)	Triglyceride level (mg/dL)
Healthy	< 200	< 100	≥ 60	≥ 60	< 159
Mild					150-199
Moderate	200-239	100-159	50-59	40-59	200-499
Severe	≥ 240	≥ 160	< 50	< 40	≥ 500

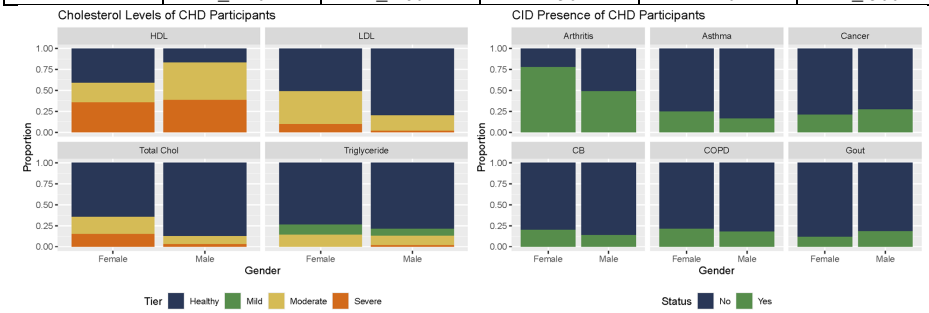

Fig. 2. (a) Distribution varies across the different cholesterol types. A higher portion of women tend to fall within the moderate-severe range compared to men in almost all cases except HDL. (b) The presence of chronic inflammatory diseases is higher in women than in men. Cancer and gout are the two conditions that affect a smaller portion of female participants with CHD than men.

Phase 2(a)—Frequent Pattern Mining. Using frequent pattern mining algorithms (e.g., Apriori, FP-growth) with a minimum support threshold of 0.6, we discovered frequent patterns that reveal key differences in CHD diagnosis patterns between men and women. These patterns highlight disparities in symptom recognition, diagnostic criteria, and the role of prior treatment in influencing diagnosis outcomes.

A consistent trend emerges: women diagnosed with CHD often present with a greater number and severity of comorbid conditions compared to men. This suggests that women are less likely to be diagnosed in the early stages of disease progression, with diagnoses typically occurring only after health deterioration becomes more pronounced.

Phase 2(b)—Association Rule Mining. Using a confidence level of 0.7, we formed interesting association rules that reveal key differences in CHD diagnosis patterns between men and women. These rules highlight disparities in symptom recognition, diagnostic criteria, and the role of prior treatment in influencing diagnosis outcomes. For women, the association rules often involve three or more interrelated health factors—including high blood pressure, use of prescribed medication, and arthritis. CHD diagnoses are more likely in women only when high blood pressure or cholesterol is both present and treated, indicating a systemic reliance on prior intervention before heart disease is recognized. Women with

untreated or less severe forms of these conditions are often overlooked, delaying both diagnosis and treatment. In contrast, men's CHD diagnoses appear to require fewer or less severe conditions. Association rules show that men are often diagnosed even when certain clinical indicators (e.g., cholesterol levels) fall within healthy ranges. Men with high blood pressure—even without prescribed medication—are still commonly diagnosed with CHD. This suggests a broader, more preventive diagnostic approach that facilitates earlier intervention in men.

Moreover, treatment history plays a critical role in CHD diagnosis for women. The presence of prescribed medications for high blood pressure or cholesterol is often a prerequisite for CHD recognition. For example, women with both a history of hypertension and a corresponding prescription are commonly diagnosed, whereas women without such treatment records—even if risk factors are present—are less likely to receive a diagnosis. In contrast, men are diagnosed independently of treatment history. Diagnostic models appear to weigh their risk factors—such as elevated blood pressure or lifestyle habits—more heavily, even in the absence of prior medical intervention.

Lifestyle factors further highlight gender-based diagnostic disparities. In men, behaviors like smoking and alcohol use are prominent contributors in association rules and are often tied to CHD diagnoses. These indicators seem to prompt early detection, even in men without advanced clinical symptoms or treatment history. For women, however, lifestyle risks are rarely central to diagnostic rules. Smoking and alcohol use appear underemphasized, leading to potential underdiagnosis in women who lack comorbidities or formal treatment records. This narrowed diagnostic scope may contribute to delayed recognition of CHD in women, especially those with relevant lifestyle risk factors but no documented clinical interventions.

In summary, the pattern mining results expose systemic biases in CHD diagnosis. Women are often diagnosed only after multiple severe conditions are present and treated, whereas men benefit from a broader, more proactive diagnostic framework. These findings underscore the need to revise diagnostic practices to better incorporate early risk indicators—such as lifestyle factors—for women, promoting earlier and more equitable detection of CHD.

5 Conclusions

In this paper, we presented a big data analytics solution for analyzing techniques, including frequent pattern and association rule mining, to uncover critical gender disparities in CHD diagnosis. Analysis of a real-world clinical dataset revealed that women are often diagnosed only after multiple severe conditions and documented treatments, while men are diagnosed earlier based on lifestyle risks or fewer symptoms. These findings highlight systemic biases in diagnostic practices and underscore the need to expand CHD criteria for women to include lifestyle factors and atypical symptoms. Our solution demonstrates the value of pattern mining in exposing healthcare disparities. Moreover, our solution is expected to be applicable to analyze other datasets. As *ongoing and future work*, we incorporate longitudinal data and/or other variables (e.g., place of residence, employment status) to evaluate the long-term impact of these diagnostic gaps and to inform more inclusive, equitable clinical guidelines.

Acknowledgment. This work is partially supported by NSERC (Canada) & U. Manitoba.

References

[1] Eom, C.S., et al.: Effective privacy preserving data publishing by vectorization. Inf. Sci. 527, 311-328 (2020)
[2] Chowdhury, M.E.S., et al.: A new approach for mining correlated frequent subgraphs. ACM TMIS 13(1), 9 (2022)
[3] Czubryt, T.J., et al.: Efficient vertical mining of frequent quantitative patterns. In: IEEE DASC-PiCom-CBDCom-CyberSciTech 2023, 950-955 (2023)
[4] Hryhoruk, C.C.J., et al.: Social network mining and analytics for quantitative patterns. In: IEEE/ACM ASONAM 2023, 717-724 (2023)
[5] Leung, C.K., et al.: Mining 'following' patterns from big sparse social networks. In: IEEE/ACM ASONAM 2016, 923-930 (2016)
[6] Leung, C.K., Hayduk, Y.: Mining frequent patterns from uncertain data with MapReduce for big data analytics. In: DASFAA 2013, Part I. LNCS, vol. 7825, 440-455 (2013)
[7] Roy, K.K., et al.: Mining sequential patterns in uncertain databases using hierarchical index structure. In: PAKDD 2021, Part II. LNCS (LNAI), vol. 12713, 29-41 (2021)
[8] Froese, R., et al.: The border k-means clustering algorithm for one dimensional data. In: IEEE BigComp 2022, 35-42 (2022)
[9] De Guia, J., et al.: DeepGx: deep learning using gene expression for cancer classification. In: IEEE/ACM ASONAM 2019, 913-920 (2019)
[10] Buttar, A.S., et al.: A real-time sentiment feedback system: binary categorization and context understanding based on product reviews. In: ASONAM 2024, Part IV. LNCS, vol. 15214, 205-220 (2024)
[11] Camara, R.C., et al.: Fuzzy logic-based data analytics on predicting the effect of hurricanes on the stock market. In: FUZZ-IEEE 2018, 576-583 (2018)
[12] Leung, C.K., Wen, Q.: Personalized privacy-preserving semi-centralized recommendation system in a social network. In: IEEE/ACM ASONAM 2023, 727-736 (2023)
[13] Madill, E., et al.: ScaleSFL: a sharding solution for blockchain-based federated learning. In: ACM BSCI 2022, 95-106 (2022)
[14] Knox, E.C.L., et al.: Gender differences in clinical practice regarding coronary heart disease: a systematic review. J. Clin. Med. 14(5), 1583 (2025)
[15] Leung, C.K., et al.: Big data science on COVID-19 data. In: IEEE BigDataSE 2020, 14-21 (2020)
[16] Leung, C.K., Zhao, C.: Big data intelligence solution for health analytics of COVID-19 data with spatial hierarchy. In: IEEE DataCom 2021, 13-20 (2021)
[17] Fan, C., et al.: Social network mining for recommendation of friends based on music interests. In: IEEE/ACM ASONAM 2018, 833-840 (2018)
[18] Baum, K., et al.: Investigating innovation diffusion in gender-specific medicine: insights from social network analysis. Bus. Inf. Syst. Eng. 66(3), 335-355 (2024)
[19] Hryhoruk, C.C.J., Leung, C.K.: Compressing and mining social network data. In: IEEE/ACM ASONAM 2021, 545-552 (2021)
[20] Hryhoruk, C.C.J., Leung, C.K.: Social network analysis on interpretable compressed sparse networks. In: IEEE/ACM ASONAM 2022, 324-331 (2022)
[21] Jiang, F., et al.: Big data mining of social networks for friend recommendation. In: IEEE/ACM ASONAM 2016, 921-922 (2016)
[22] Leung, C.K., et al.: Compression for very sparse big social data. In: IEEE/ACM ASONAM 2020, 659-666 (2020)
[23] Leung, C.K., Singh, S.P.: A mathematical model for friend discovery from dynamic social graphs. In: IEEE/ACM ASONAM 2021, 569-576 (2021)
[24] Madhumita, S., Jeyasekar, A.: Social network analysis in driver monitoring systems: understanding the impact of predictive models on driver safety. Soc. Netw. Anal. Min. 15(1), 34 (2025)
[25] Singh, S.P., Leung, C.K.: A theoretical approach for discovery of friends from directed social graphs. In: IEEE/ACM ASONAM 2020, 697-701 (2020)
[26] Kolisnyk, M., et al.: Analysis of multi-dimensional road accident data for disaster management in smart cities. In: IEEE IRI 2022, 43-48 (2022)
[27] Lichtlen, P.R.: History of coronary heart disease. Z. Kardiol. 91(Suppl. 4), 56–59 (2002)
[28] Wenger, N.K.: Women and coronary heart disease: a century after herrick: understudied, underdiagnosed, and undertreated. Circulation 126(5), 604–611 (2012)
[29] Hiremath, A., et al.: Sex differences in health status and clinical outcomes after non-primary percutaneous coronary intervention. Circ Cardiovasc Interv. 15(3), e011308 (2022)
[30] Maas, A.H.E.M.: Characteristic symptoms in women with ischemic heart disease. Curr. Cardiovasc. Risk Rep. 13, 17 (2019)
[31] Solola, S.N., et al.: Sex-specific considerations in the presentation, diagnosis, and management of ischemic heart disease. J. Am. Coll. Cardiol. 79(14), 1398–1406 (2022)
[32] de Marvao, A., et al.: Heart disease in women: a narrative review. Anaesthesia 76(Suppl. 4), 118-130 (2021)
[33] Shah, T., et al.: An update on gender disparities in coronary heart disease care. Curr. Atheroscler. Rep. 18, 28 (2016)